International Journal of Computer Applications Technology and Research
Volume 4– Issue 1, 18 - 23, 2015, ISSN:- 2319–8656

Cluster as a Service (CaaS) in Secure Deduplication System

R.Nalla Kumar
Department of CSE
Regional Centre
Anna University
Coimbatore.India

A.Kumari savitha sree
Department of CSE
Regional Centre
Anna University
Coimbatore, India

X.Alphonseinbaraj
Department of CSE
Regional Centre
Anna University,
Coimbatore, India

Abstract: Data deduplication is one of compression method of data for eliminating duplicate copies of repeating data that is mainly used to reduced the amount of storage space and bandwidth. Cloud computing systems have been made possible through the use of large –scale clusters,service –oriented architecture (SOA),web services ad virtualization.While the idea offering resources via web services is common place in cloud computing ,little attention has been paid to clients themselves specifically ,human operators.Despite that clouds host a variety of resources which in turn are accessible to variety of clients ,support for human users is minimal .To provide better security ,this paper makes the first attempt to formally address the problem of authentication ,integrity and availability.By using Tag generation , moreover one of additional cloud storage service such that Cluster as a Service (CaaS) can make secure deduplication possible and reduced cloud storage space.With out key generation ,attribute based encryption makes secure data deduplication in the cluster of computers.

Keywords: Deduplication,PoW,Cluster as a Service ,Identification Protocol

1 INTRODUCTION

Cloud computing provide unlimited “ **virtualized resources** ”to users as services across the whole Internet ,while hiding platform and implementation today. Cloud Infrastructure providers are establishing cloud centers to host a variety of ICT services and platforms of world wide individuals ,innovators and institutions. Cloud Service Providers(CSP) are very aggressive in experimenting and embracing the cool cloud ideas and today every business and technical services are being hosted in cloud to be delivered to global customers ,clients and consumers over the Internet communication infrastructure .For example Security as a service (SaaS) is a prominent cloud –hosted security service that can be subscribed by spectrum of users of any connected device and users just pay for the exact amount or time of usage .Besides the modernization of legacy applications and positing the updated and upgraded in clouds ,fresh applications are being implemented and deployed on clouds to be delivered to millions of global users simultaneously affordably.While web services have simplified resource access and management ,it is not possible to know if the resource(s) behind the webservice is (are) ready for request.Clients need to exchange numerous message with required Web services to learn the current activity of resources and thus face significant overhead loss if most of the web services prove ineffective.Furthermore ,even in ideal circumstances where all resources behind Web services are the best choice,clients still have to locate the services themselves . Finally ,the Web services have to be stateful so they are able to best reflect the current state of their resources.

Although data deduplication provide lots of benefits and advantages ,security and privacy concerns sensitive data are susceptible to both insider and outsider attacks . Normal traditional deduplication is incompatible with encryption. Specifically different users produces different ciphertext makes data deduplication ineffective and not feasible .Convergent encryption [1],[2] has been proposed to enforce data confidentiality while data deduplication is feasible .

It encrypt and decrypts a data copy with *convergent key* and further to avoid unauthorized entry in system, a secure proof of ownership[5] is also needed to provide the proof that the user indeed owns the same file while duplicate is found . In additionally tag was generated by attribute based encryption .

This method is different from traditional techniques . In traditional method,each time user can access the file with their own private key ,but here attribute based encryption done such that generating tag and by which privileges[6][4] is associated with that. Each file is uploaded to the cloud is also bounded by set of privileges to specify which kind of users is allow to perform the duplicate check and access the file.The user can find the duplicating of the file if it is stored in cloud .In our system, we need to consider the three things as follows. 1. Cloud Management System 2. Virtual cluster administrations 3.user .

1.1 CONTRIBUTIONS

In this paper,

1) Performing convergent encryption with differential privileges and tag generation which is used to avoid duplicate without generate key in client /user side.

2) The users without corresponding privileges cannot perform the duplicate check .For example ,in a

company , many privileges will be assigned to employees.In otherwords ,no differential privileges have been considered in the deduplication based on convergent encryption techniques in traditional deduplication methods.

3) Introduce the first provably –secure deduplication method in Cluster based Service

1.2 Organizations

The rest of this paper proceeds as follows. Section 2 briefly revisit some preliminaries of this paper. Section 3 propose the system model for secure data protection. In section 4, implementation progress of the secure duplication system is described. In section 5, some other related work regarding this system is described.Some experimental result are shown in the section 6. In section 7, some future work and some other ideas to enhance security is described. Finally conclusions are drawn in section 8.

2 PRELIMINARIES

In this section, we are going to consider attribute based encryption and review some secure primitives used in our secure deduplication

2.1 Cluster as a Service (CaaS) implementation:

The main role of CaaS is to (i) provide easy and intuitive file transfer so clients can upload file(s),(ii) offer an easy to use interface for clients to monitor their process.The CaaS does this by allowing clients to upload files as they would any web page while carrying out the required data transfer to the cluster transparently.

By hiding hardware and software features of cluster ,the CaaS provides higher level abstraction. Clients only receive the minimal amount of the data and provide the web pages to deploy,run and control execution of process.Because clients to the cluster cannot know how the data storage is managed,the CaaS offes a simple transfer interface to clients while addressing the transfer specified.

In some other experiments,CaaS was implemented by Windows Communication Foundation of .Net using web services .The problem is client(s) exchanges the numerous messages in case of activation of resources.

2.2 Proof of Ownership(PoW):

Halevi et al [10] proposed the notion of "Proof of Ownership(PoW)" for deduplication system.Such that client prove their authentication without uploading their files .Several PoW constructions based on the Merkle-Hash Tree proposed [10] to enable client side deduplication which include the bounded leakage setting and another schemes such that Pietro and Sorniotti[7] proposed some bit positions based file proof.Now recently Ng et al [8] proposed some PoW method but that not address that how to minimize the key management overhead etc.,

In our system PoW is used to enable users to prove their ownership in order to found deduplication occur in system.Generation of Tag is mainly used to detect the duplication .Virtual machine technology makes it very flexible and easy to mange resources in cloud computing environments,because they improve the utilization of such resources by multiplexing many virtual machines on one physical host(server consolidation).These machines can be scaled up and down on demand with a high level of resources abstraction .

2.3 Proof & Verify protocol:

There are identification protocols vailable.Such that in literature,including certificate –based ,identity based identification[3][9]. According to that ,there are two phase,Proof and verify.In the stage of Proof ,a prover can prove their authorized identity to verifier.The verifier verify that prover identity based information and then proceed by either accept or reject.This protocol is mainly used in our system because user uploading file if user found some duplication on this particular file .This is most important protocol for secure user identity .

To provide data integrity ,the Azure stroage service stores the uploaded data MD5 checksum in the database and returns it to the user when user wants to retrieve the data.Amazon AWS computes the data MD5 checksum and e-mails it to the user for integrity checking .

3 SYSTEM MODEL

This section contains the three different entities: . 1. Cloud Management System 2. Virtual cluster administrations 3.user . Fig1. Depicts the architecture view of proposed deduplication system

1) Cloud Management System(CMS):This is entity that provides the storage service .This is main management system which control all Virtual Cluster Administration and user which is connected under the Virtual Cluster Administration.Here each user (s) tag only stored instead of store entire file . This is maintaining each users tag for whole file management system.

2) Virtual Cluster Administration (VCA):Its an entity which has the expertise and capabilities that client(s)/user(s) do not have and trusted to assess and expose the risk of cloud storage services on behalf the client request.

Here Webserver and Shibboleth [13] available.Shibboleth perform Single- Sign-On(SSO) for authorized user entry in this phase.Webserver will redirect unauthorized user (s) to shibboleth for verification.Tag was generated for each user based on attribute based .Through tag shibboleth allow the SSO.That generated tag was shared by CMS for verification purpose only.

3) User :Its also an entity who will access the massive data and controlled by central management system and Virtual Cluster Administration.User upload the file to VCA.According uploading the file ,tag generated which is stored in VCA as well as CMS for perfect secure file management system. Tag generation phase occurred here only.

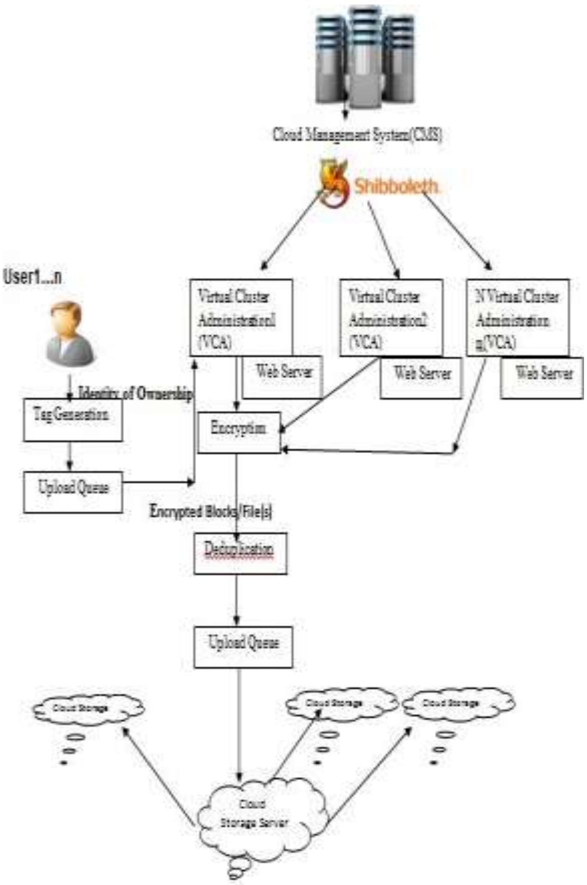

Fig1. Architecture view of proposed deduplication system.Optionally ,data on the back server can be replicated to multiple cloud storage in the back ground.

3.1 Adversary

Typically ,We assume that public and private cloud are both "honest but -curious" they will follow our proposed system.but users would try to access data either within or out off scopes of their privileges .

To illustrate this ,lets consider the following two scenarios. First ,assume that Alice , a company CFO ,stores the company financial data at a cloud storage service provided by Eve .And then Bob ,the company administration chairman ,downloads the data from the cloud .There are three important concerns in this procedure: Confidentiality,Integrity and Repudiation

Confidentiality : Eve is considered as an untrustworthy third party.Alice and Bob do not want reveal the data to Eve .

Integrity: As the administrator of the storage service ,Eve has capability to play with the data in hand.How can Bob be confident that the data he fetched from Eve are the same what was sent by Alice ? Are there any measures to guarantee that the data have not been tampered by Eve?

Repudiation: If Bob finds that the data have been tampered with,is there any evidence for him to demonstrate that it is Eve who should be responsible for the fault? Similarly ,Eve also needs certain evidence to prove her innocence .

Recently some reply from developer was **"*We wont lose your data ….we have a robust back up and recovery strategy but we are not responsible for you losing your own data* -"**Obviously ,it is not persuasive to the potential customer to be confident with the service .

Confidentiality can not achieved by without adopting robust encryption schemes . But however ,the integrity and repudiation issues not handled well in the current cloud service platform . There are some linking missing between uploading and downloading sessions .

That leads some following questions:

Repudiation between users and VCA and Upload and Download sessions integrity

Repudiation between users and VCA: In case some data errors are occurred without transmission error means ,how can users and VCA prove their innocence and authorization?

Uploading and Downloading sessions integrity:

Since integrity in uploading and downloading phase are handled separately, how users and service providers download the same data content which is previously uploaded in system?

Is there mistake or error occurred means how to solve the sessions in uploading and downloading session?

4 SECURE DEDUPLICATION SYSTEM

There is a Authority Certificate (AC) issued by user and VCA and That AC copy is stored in CMS.The user and

VCA are using Key Sharing (KS),there are four solutions to bridge the mission link of data integrity between uploading and downloading procedures .

1) Neither AC nor KS
2) With KS without AC
3) With AC without KS
4) With both AC and KS

4.1 Neither AC nor KS:

Uploading file in our secure deduplication system :

1) User :sends the data to VCA with Tag generation and this known as Tag Generation by User(TGU)
2) VCA:verifies the data with Tag generation .If it is valid ,VCA sends back Tag Generation and this known as Tag Generation byVCA (TGVCA).

TGU is stored at the user side and TGVCA is stored at the VCA side .Then it is stored in cloud service provider without any problem.

Once uploading is finished ,both side agreed on the integrity of the uploaded data ,and each side owns TGU and TGVCA generated by opposite site.

Downloading file in our secure deduplication system:

1) User:Send request to VCA with Authorized identity proof (PoW)
2) VCA: Verifies the Authorized identity proof (PoW) .if it is valid ,the VCA send back the data with TGVCA to user .

User then verifies the data with TAG Generation. This things will update in CMS

4.2 With KS without AC

Uploading file in our secure deduplication sytem:

1) User: Sends the data to VCA with Tag Generation.
2) VCA:verifies the data with Tag Generation.If it is valid ,the VCA send back the Tag Generation.

VCA and user share Tag Generation with KS.

Then both sides agree on the integrity of the uploading data ,and they share the agreed Tag Generation ,which is used when disputation happens.

Downloading file in our secure deduplication system:

1) User :Send request to the VCA with Tag Generation
2) VCA:Verifies the request identity,if it is valid ,the VCA send back the data with Tag Generation
3) User verifies the data through Tag Generation.

When disputation happens the user or VCA can take the shared Tag Generation together recover it and prove his /her innocence.

4.3 With AC without KS

Uploading file in our secure deduplication system:

1) User : Sends the data to VCA along with Tag Generation by User (TGU).
2) VCA: Verifies the data with Tag Genearation,if it is valid ,the VCA send back the Tag Generation and TGVCA

TGU and TGVCA are send to AC

On finishing the uploading phase ,both sides agree on the integrity of the uploaded data , and AC owns their agreed Tag Generation.

Downloading file in our deduplication system:

1) User: Send request to VCA with Authorized identity proof (PoW).
2) VCA:verifies the request with PoW ,if it is valid ,the VCA send back the data with Tag Genration.

User verifies the data through the Tag Generation.

When disputation happens , the user or VCA can prove their innocence by presenting the TGU and TGVCA which are stored at the AC.

Similarly ,there are some special cases .when VCA is trustworthy ,only Tag Generation is needed.When user is trustworthy,only the TGVCA is needed;

4.4 With both AC and KS:

Uploading file in our secure deduplication system:

1) User : sends the data to VCA with Tag Generation.
2) VCA:Verifies the data with Tag Generation.

Both the user and VCA send Tag to AC.

AC verifies the two Tag .If they match ,the AC distributes Tag to the user and VCA by KS.

Both side agree on the integrity of the uploaded data and share the same Tag by KS and AC own their agreed Tag Generation.

Downloading file in our deduplication system:

1) User :Sends request to the VCA with PoW;
2) VCA:verifies the request identity ,if it is valid ,the VCA send back the data with Tag .

User verifies the data through Tag

Here are some special cases .when the VCA is trustworthy ,only the user needs Tag.When the user is trustworthy,only the VCA need Tag.

5 RELATED WORK

Yuan et al.[15] proposed a deduplication system in cloud storage to reduce the storage size of the tag for integrity check.Bellare [2]showed how to protect the data confidentiality by transforming the predictable message into unpredictable message.Stanek et al[14] proposed some techniques that is for popular data and unpopular data .Li et al[12]addressed some key management by distributing keys across multiple servers after encryption the files .

Convergent Encryption:Xu et al [11] address the problem and showed a secure convergent encryption without considering key-management .It is known that some commercial cloud storage providers ,such as Bitcasa ,also deploy convergent encryption

Proof of Ownership(PoW): Halevi et al[10]proposed Proof of user authentication identity.Similarly Ng et al.[16]extended this same PoW but address the problem and how to management the key overhead. In our secure deuplication system based on attribute encryption .Therefore with out key overhead. And consider to be secured one by VCA and CMS.

6 EXPERIMENTAL RESULTS:

In this experimental result ,Fig 2 shows that how virtual machines node use the memory for storing the file .There are three categories available

1) Maximum
2) Average
3) And current

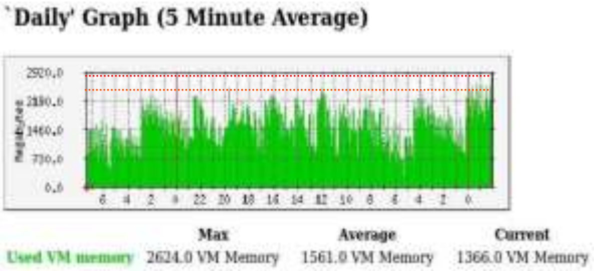

Figure 2. Timeline of the memory usage of the virtual machine.

Here there are some file are stored on daily basis.According maximum used Virtual Memory is 1460 MB to 2624.0 ,Average used Virtual memory is 730 MB to 1561 and current used virtual memory is 2190MB .Daily storing file and viewing file take memory usage is shown in experimental results.

7 FUTURE WORK

In our secure dedupliocation system,we consider three different categories such that CMS,VCA and user .Here group deduplication work and inter and intra group are considered.therefore generating Tag consume some amount of bandwidth .but not much more in cloud storage .In future reduce the generating tag space and bandwidth .Because tag usage is varied according transfer the file format such that .doc,.pdf,.jpeg,etc., Therefore we enhance fast transfer of any file from one group to another of user.

8 CONCLUSION

Our secure deduplication system is provided to protect the data from attack. Its make attacker job very complicated .because of generating Tag and PoW is making data more secure. In our system Cloud Management System (CMS) and Virtual Cloud Administration (VCA) having each user’s file tag copy .Therefore transferring and storing file in cloud is very secure.Not visible to attack which means attack cannot find the path of target file.and CMS having CA for verifying owner identify by performing PoW.We showed that our secure deduplication system incurs minimal overhead compared to previous deduplication system.

9 ACKNOWLEDGMENTS

My Sincere Thanks to Mr.M.Newlin kumar sir and Mr. R.Nalla kumar sir ,mentors for providing their guidance and cooperation in this Research.

REFERENCES:

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

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

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

[4]. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman.Role-based access control models. IEEE Computer, 29:38–47, Feb 1996.

[5]. 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.

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

[7]. R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y.

Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.

[8]. W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012.

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

[10]. 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.

[11]. J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In ASIACCS ,pages 195–206, 2013.

[12] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

[13]. Shibboleth, http://shibboleth.internet2.edu, 2010.

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

[15]] J. Yuan and S. Yu. Secure and constant cost public cloud Storage auditing with deduplication. IACR Cryptology ePrint Archive, 2013:149, 2013

[16]. W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012