© 2017 IJSRSET | Volume 3 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A De-Duplication Scheme and Distributed Key Generation for Achieving the Strongest Privacy in Cloud

L. Sankaran, D.Ashok Kumar, S. Dushyanth, E. Ayyappan
Computer Science and Engineering, Anna University, SKP Engineering College, Tiruvannamalai, Tamil Nadu, India

ABSTRACT

In this paper we study about hybrid cloud approach for secure authorized deduplication. Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been rapidly used in clouds to reduce the amount of storage space. To protect the privacy of sensitive data while supporting deduplication, the convergent encryption technique has been used to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication. This technique is 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 that have been supporting the authorized deduplication in a hybrid cloud environment. Security analysis demonstrates that our deduplication scheme is secure by the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized deduplication scheme and conduct tested experiments using our prototype. We show that our proposed authorized deduplication scheme incurs minimal overhead compared to normal operations.

Keywords : Security, Data Protection, Deduplication, Convergent Encryption, Cloud Storage.

I. INTRODUCTION

Nowadays,several trends are opening up in the area of Cloud Computing, which is an Internet-based development and use of computing technology. The ever cheaper and most powerful processors, along with the SaaS computing architecture, are transforming data centers into pools of computing service on a large scale. The increasing bandwidth of the network and reliable yet flexible network connections make it ever possible that users can be able to subscribe high quality services from data and software that reside on remote data centers.

Moving data into the cloud provides great opportunity to users since they don't have to care about the complexities of direct hardware management.

The pioneer of Cloud Computing vendors, Amazon Simple Storage Service (S3) and Amazon's Elastic Compute Cloud (EC2) are well known examples. While these internet-based online services provide huge amount of storage space and customizable computing resources, this computing platform shifts, eliminates the responsibility of local systems for data maintenance simultaneously. As a result, users are at the mercy of their CSP's for their data availability and data integrity . Recent trends of Amazon's S3 is an example . From the view of data security, which always has been an important principle of quality of service. Cloud Computing unavoidably has new challenging security threats for various reasons.

Initially, traditional cryptographic primitives for data security protection purpose can't be directly embraced because the user's may loss control on data under Cloud Computing. Therefore, the verification of correct cloud data storage should be handled without clear knowledge of whole data in cloud. Considering different data for every user stored in the cloud storage and the requirement of long term continuous surity of their data protection, the problem of checking correctness of data storage in the cloud becomes more challenging. Secondly, Cloud Computing is not only a third party

data warehouse. The data stored in the cloud should be continuously updated by the users, including insertion of data, deletion of data and modification of data.

To ensure storage correctness under real time data update is too important. However, this dynamic feature also makes conventional integrity insurance techniques futile and entails new results. Finally, the deployment of Cloud Computing is powered by data centers running in a simultaneous, cooperated and distributed way. Individual user's data is duplicately stored in multiple physical locations of the cloud server to further compress the data integrity issues. Therefore, distributed protocols for storage correctness assurance will be of most importance in achieving a powerful and secure cloud data storage system in the real world. However, such important field remains to be fully explored in the literature.

II. METHODS AND MATERIAL

A. Existing System

According to the perspective of data security, which has always been a major aspect of quality of service, Cloud Computing inevitably have new challenging security threats for many reasons.

1. Firstly, traditional cryptographic technique for the purpose of data protection can not be directly adopted due to the user's loss their control of data under Cloud Computing. Therefore, authorization of correct data storage in the cloud should be done without the deep knowledge of the whole data. Considering ,different data for each user stored in the cloud servers and the requirement of long term continuous assurance of their data protection, the problem of checking the correctness of data storage in the cloud servers was very challenging.
2. Secondly, Cloud Computing is not only a third party data warehouse. The data which is stored in the cloud may be rapidly updated by the different users, including insertion of data, deletion of data, modifications of data, etc. To ensure storage correctness under real time data update is hence of very huge importance.

Disadvantages:

- The user needs to know private key.
- Less protect security.
- This deduplication systems cannot support differential authorization duplicate check.

B. Proposed System

In this paper, we are proposing an effective and resourceful distributed scheme with clear-cut dynamic data support to ensure the correctness of users data in the cloud. We rely on erasure correcting code in the file distribution preparation to provide duplicate copies and guarantee the data dependability. This construction often reduces the communication and storage overhead as compared to the conventional replication-based file distribution techniques. By using the similar tokens with distributed verification of erasure-coded data, our scheme achieves the storage purity as well as cloud data error localization: whenever corruption on data has been detected during the storage correctness verification, our scheme can almost guarantee the simultaneous localization of data errors, i.e., the identification of the misbehaving server.

Aditionally we are proposing to use convergent encryption method,which initiates to encrypts the end users document before they stored in the cloud.While the user's retrival,the cloud decrypts the stored data using the end users convergent key.

When a user request a file to the CSP,it sends such file to the requested user.If some other user request the same file to the CSP, will automatically send to the user.
Advantages:

- The user dont needs to know private key
- Better protect security
- This deduplication systems can support differential authorization duplicate check

C. System Architecture

Fig. shows the system architecture and describes the following modules. This architecture contains data owners, users, cloud service provider and granted applications. Cloud service provider provides the secret key to users. Users want to upload the file means management server check the secret key and allow that function. If it same file means service provider gives the notification like file already exits.

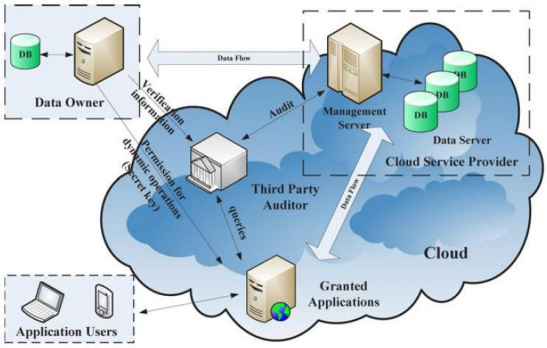


Figure 1. System Architecture

Processing steps:

Step 1:
The users have to register with the cloud . if they are already registered , then they should login to the cloud by using their user name and password. Every users will be verified by their username and password.if the username and password are valid , then the user will be allowed into the cloud for uploading and accessing of their files .otherwise they will be rejected.

Step 2:
The secret key is provided by the CSP for every granted user when they are uploading the files. Using that convergent key , the files that has to be uploaded by the user , will be encrypted. The keys will be generated by CSP.

Step 3:
After the encryption is over, the encrypted files will be handed over to the TPA to check for the duplication. The duplication check is done by comparing the encrypted text of that files. If the file is already present in any of the cloud server , then the user won't be upload that file into the cloud.

Multi-layered Cryptosystem:

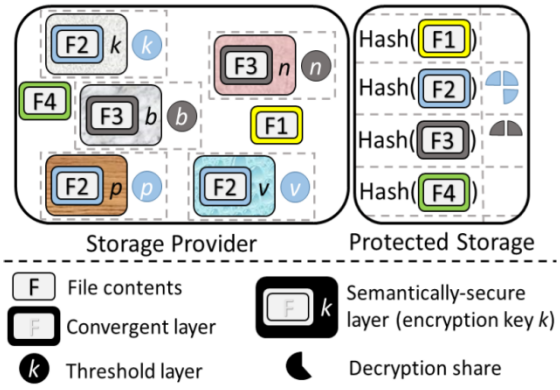


Figure 2. The cryptosystem with multilayer

D. Modules

- V base and VDFS Consumption
- Cloud Admittance
- Secure deduplication System
- Access of Deduplicated system using Secure Check Token
- Amputation of Cloud files

E. Module Description

1. V base and VDFS Consumption

The client sends the request queries to the server. Based on the request queries the server sends the response file corresponding to the client's request. Before this process, the client verification step is involved. In the server side, it checks the client name and its password for security purpose. If it is valid and then received the queries form the clients and make search for the corresponding files in the database. Finally, find that matching file and send to the client. If the server finds the any other intruder means, it set the different path to those intruder.

2. Cloud Admittance

Three different network entities can be identified as follows:

• User:
Users, who have data to be stored in the cloud and depends on the cloud for data computation, may be individual consumers or from organizations.

• Cloud Service Provider (CSP):

A CSP, who has significant resources and expertise in creating and maintaining distributed servers of cloud storage, owns and operates live Cloud Computing systems.

• Third Party Auditor (TPA):

An optional TPA, who has different capabilities and expertise that users may not have, isTrusted to assess and avoid the risks of cloud storage services on behalf of the users upon request.

3. Secure deduplication System

Fig. describes the cloud data storage, a user stores his data by using a CSP into a set of servers located in cloud, which are running in a simultaneous, the user

communicates with the cloud servers via CSP to retrieve his data. In some cases, the user may need to perform block level tasks on his data stored in the cloud. users should be equipped with security because they can correct assurance of their stored data continuously, even without the presence of local copies. In case that users do not have the time or resources to monitor their data, they can handover the tasks to an optional trusted TPA of their choices. In our model, we assume that the point-to-point communication channels between each server of the cloud and the user is verified and reliable, which can be achieved in practical with minimal overhead.

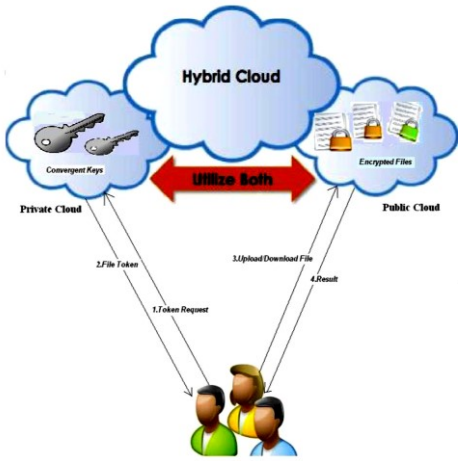


Figure 3. Secure deduplication System

4. Access of Deduplicated system using Secure Check Token

Fig. 3 describes the Authentication Server (AS) functions as any AS would with some extra behaviors added to the typical client-verification protocol. The first new behaviour is the sending of the client's authentication information to the impersonate router. The AS in this model also functions as a ticketing control, controlling requests on the application network. The other function that should be supported by the AS is the updation of client lists, reducing the authentication time or may be the removal of the client as a authorized client depending upon the permissions.

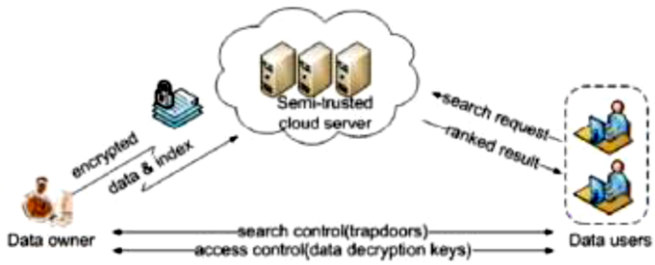
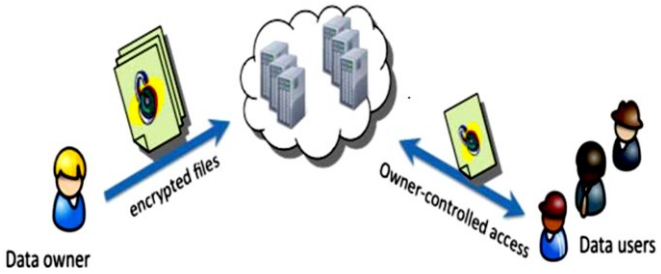


Figure 5. Access of Deduplicated System using SecureCheck Token

5. Amputation of Cloud files

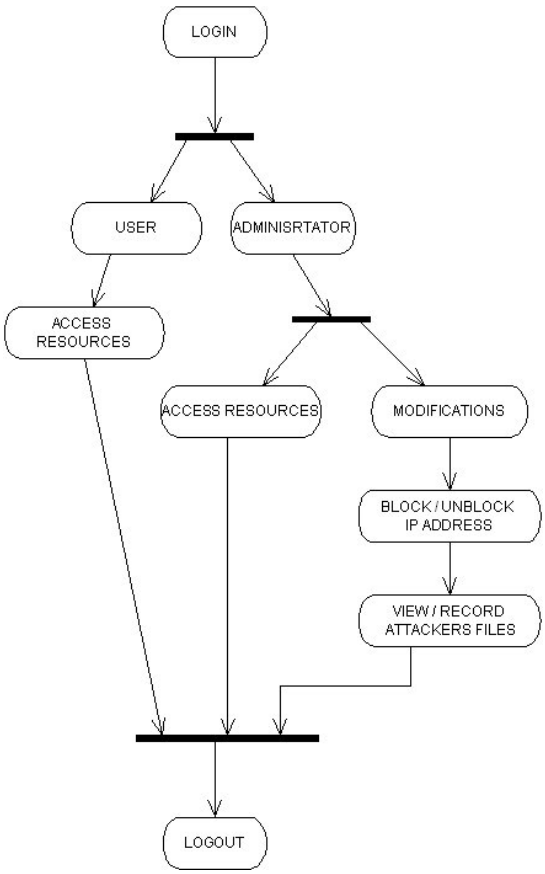
One of The major issues is to detect the modification or corruption of data, which is happening in an unauthorized manner.it may be happening because of the server compromise with the users or the random byzantine failures. besides, if such inconsistencies are detected, to find the data error of such servers.

The security threats of the cloud storage can be of two types. The first one is , a CSP will be self-interested or malicious.it does move the unassessed data to the lower tier of the storage, but it may leads to data loss incident due to the management errors.



F. Unified Modeling Language(UML)

Activity Diagram



III. CONCLUSION

In proposed system investigated the problem of data security in cloud data storage, which is essentially a distributed cloud storage . To verify the pureness of user's data in cloud data storage, we proposed an effective and resourceful distributed scheme with clear-cut dynamic data support, including block update, delete, and append. We depend on erasure-correcting code in the file distribution preparation to provide duplicate copies of parity vectors and guarantee the data dependability. By utilizing the similar token with distributed verification of erasure-coded data, our scheme achieves storage purity and cloud data error localization, i.e., whenever the corruption on data has been detected during the storage correctness verification across the distributed cloud servers, we can identify the misbehaving server(s) simultaneously for sure. Through clear security and performance analysis, we prove that our scheme is more efficient and can recover from Byzantine failures, corrupted data modification attack, and also server colluding attacks.

IV. REFERENCES

[1]. J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," in Financial Cryptography and Data Security - 18th International Conference, Christ Church, Barbados, March 3-7, 2014, Revised Selected Papers, pp. 99–118.

[2]. D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage," IEEE Security & Privacy, vol. 8, no. 6, pp. 40–47, 2010.

[3]. S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proceedings of the 22th USENIX Security Symposium, Washington, DC, USA, August 14-16, 2013, pp. 179–194.

[4]. D. Meister and A. Brinkmann, "Multi-level comparison of data deduplication in a backup scenario," in Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference 2009, Haifa, Israel, May 4-6, 2009, p. 8.

[5]. N. Mandagere, P. Zhou, M. A. Smith, and S. Uttamchandani, "Demystifying data deduplication," in Middleware 2008, ACM/IFIP/USENIX 9th International Middleware Conference, Leuven, Belgium, December 1-5, 2008, Companion Proceedings, pp. 12–17.

[6]. L. Aronovich, R. Asher, E. Bachmat, H. Bitner, M. Hirsch, and S. T. Klein, "The design of a similarity based deduplication system," in Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference 2009, Haifa, Israel, May 4-6, 2009, p. 6.

[7]. M. Dutch and L. Freeman, "Understanding data de-duplication ratios," http://www.snia.org/sites/default/files/Understanding Data Deduplication Ratios-20080718.pdf, accessed: 2016-03-07.

[8]. D. Harnik, O. Margalit, D. Naor, D. Sotnikov, and G. Vernik, "Estimation of deduplication ratios in large data sets," in IEEE 28th Symposium on Mass Storage Systems and Technologies, MSST 2012, April 16-20, 2012, CA, USA, pp. 1–11.

[9]. S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS 2011, Chicago, IL, USA, October 17-21, 2011, pp. 491–500.

[10]. R. D. Pietro and A. Sorniotti, "Boosting efficiency and security in proof of ownership for deduplication," in 7th ACM Symposium on Information, Compuer and Communications Security, ASIACCS '12, Seoul, Korea, May 2-4, 2012, pp. 81–82.

[11]. 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, 2002, pp. 617–624.

[12]. M. W. Storer, K. M. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in Proceedings of the 2008 ACM Workshop On Storage Security And Survivability, StorageSS 2008, Alexandria, VA, USA, October 31, 2008, pp. 1–10.

[13]. M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings, pp. 296–312.

[14]. M. Bellare and S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," in Public-Key Cryptography - PKC 2015 - 18th IACR International Conference on Practice and Theory in Public-Key Cryptography, Gaithersburg, MD, USA, March 30 - April 1, 2015, Proceedings, pp. 516–538.

[15]. J. Xu, E. Chang, and J. Zhou, "Weak leakage-resilient client-side deduplication of encrypted data in cloud storage," in 8th ACM Symposium on Information, Computer and Communications Security, ASIA CCS '13, Hangzhou, China - May 08 - 10, 2013, pp. 195–206.

[16]. F. Armknecht, J.-M. Bohli, G. O. Karame, and F. Youssef, "Transparent data deduplication in the cloud," in Proceedings of the 22nd ACM SIGSAC Conference

on Computer and Communications Security. ACM, 2015, pp. 886–900.
[17]. J. Li, X. Chen, M. Li, J. Li, P. P. C. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 6, pp. 1615–1625.
[18]. P. Meye, P. R. Parv´edy, F. Tronel, and E. Anceaume, "A secure two-phase data deduplication scheme," in 6th IEEE International Symposium on Cyberspace Safety and Security CSS 2014, Paris, France, August 20-22, 2014, pp. 802–809.
[19]. Y. Duan, "Distributed key generation for encrypted deduplication: Achieving the strongest privacy," in Proceedings of the 6th edition of the ACM Workshop on Cloud Computing Security, CCSW '14, Scottsdale, AZ, USA, November 7, 2014, pp. 57–68.
[20]. P. Puzio, R. Molva, M. O¨ nen, and S. Loureiro, "Cloudedup: Secure deduplication with encrypted data for cloud storage," in IEEE 5th International Conference on Cloud Computing Technology and Science, Bristol, UK, December 2-5, 2013, Volume 1, pp. 363–370.
[21]. J. Liu, N. Asokan, and B. Pinkas, "Secure deduplication of encrypted data without additional independent servers," in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, Oct. 2015, pp. 874–885.
[22]. X. Zhao, Y. Zhang, Y. Wu, K. Chen, J. Jiang, and K. Li, "Liquid: A scalable deduplication file system for virtual machine images," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 5, pp. 1257–1266.
[23]. J. R. Douceur, "The sybil attack," in Peer-to-Peer Systems, First International Workshop, IPTPS 2002, Cambridge, MA, USA, March 7-8, 2002, Revised Papers, pp. 251–260.
[24]. S. Goldwasser and S. Micali, "Probabilistic encryption," J. Comput. Syst. Sci., vol. 28, no. 2, pp. 270–299, 1984.
[25]. A. Shamir, "How to share a secret," Commun. ACM, vol. 22, no. 11, pp. 612–613, 1979.
[26]. G. Ateniese, J. Camenisch, S. Hohenberger, and B. de Medeiros, "Practical group signatures without random oracles," IACR Cryptology ePrint Archive 2005/385.
[27]. P. S. L. M. Barreto, B. Lynn, and M. Scott, "Efficient implementation of pairing-based cryptosystems," J. Cryptology, vol. 17, no. 4, pp. 321–334, 2004.
[28]. B. Lynn, "The pairing-based crypto. library," http://crypto. stanford.edu/pbc/, accessed: 2016-03-07.
[29]. V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proceedings of the 13th ACM Conference on Computer and Communications Security, Alexandria, VA, USA, Oct. 2006, pp. 89–98.
[30]. "The Pirate Bay 2008-12 Dataset," University Of Zurich, Department of Informatics webpage, http://www.csg.uzh.ch/ publications/data/piratebay.html, accessed: 2016-03-07.
[31]. "Ubuntu Popularity Contest," http://popcon.ubuntu.com/, accessed: 2016-03-15.
[32]. D. F. Aranha, K. Karabina, P. Longa, C. H. Gebotys, and J. L´opez, "Faster explicit formulas for computing pairings over ordinary curves," in Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings, pp. 48–68.
[33]. J. Camenisch, S. Hohenberger, and A. Lysyanskaya, "Balancing accountability and privacy using e-cash (extended abstract)," in Security and Cryptography for Networks, 5th International Conference, SCN 2006, Maiori, Italy, Sept. 6-8, 2006, Proceedings, pp. 141–155.
[34]. A. Lysyanskaya, R. L. Rivest, A. Sahai, and S. Wolf, "Pseudonym systems," in Selected Areas in Cryptography, 6th Annual International Workshop, SAC'99, Kingston, ON, Canada, August 9-10, 1999, Proceedings, pp. 184–199