Secure Deduplication and Data Security with Efficient and Reliable Convergent Key Management

Nikhil O. Agrawal[1], Prof.S.S.Kulkarni[2]

Student, Information Technology, PRMIT&R, Badnera [1]

Professor, Information Technology, PRMIT&R, Badnera[2]

Abstract: Secure deduplication is a technique for eliminating duplicate copies of storage data, and provides security to them. To reduce storage space and upload bandwidth in cloud storage deduplication has been a well-known technique. For that purpose convergent encryption has been extensively adopt for secure deduplication, critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. The basic idea in this paper is that we can eliminate duplicate copies of storage data and limit the damage of stolen data if we decrease the value of that stolen information to the attacker. This paper makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. We first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose Dekey, User Behavior Profiling and Decoys technology. Dekey new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers for insider attacker. As a proof of concept, we implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments. User profiling and decoys, then, serve two purposes. First one is validating whether data access is authorized when abnormal information access is detected, and second one is that confusing the attacker with bogus information. We posit that the combination of these security features will provide unprecedented levels of security for the deduplication in insider and outsider attacker.

Keywords: Secure deduplication, Dekey, User Behavior Profiling, Decoy Technology.

I. INTRODUCTION

- ### DATA DEDUPLICATION

Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. One critical challenge of today's cloud storage services is the management of the ever-increasing volume of data. To make data management scalable deduplication we are use convergent Encryption for secure deduplication services.

- ### User Behavior Profiling

By monitoring data access in the cloud and detect abnormal data access patterns User profiling is a well-known Technique that can be applied here to model how, when, and how much a user accesses their information in the Cloud. Such „normal user‟ behavior can be continuously checked to determine whether abnormal access to a user's information is occurring. This method of behavior-based security is commonly used in fraud detection applications. Such profiles would naturally include volumetric information, how many documents are typically read and how often. We monitor for abnormal search behaviors that exhibit deviations from the user baseline the correlation of search behavior anomaly detection with trap-based decoy files should provide stronger evidence of malfeasance, and therefore improve a detector's accuracy.

- ### Decoy Technology:

Decoy technology is the technology which is providing the decoy information to the unauthorized user or the attacker. Decoy technologies for example honeypot, or the generating the useless data files on the demand of the system to do attack against the attacker. Using this technique the original information gets changed in unexpected format so that the ex-filtering of the document or information is becomes impossible. This technology may be integrated with user behavior profiling technology to secure a user's information in the Cloud. Whenever abnormal access to a cloud service is noticed, decoy information may be returned by the Cloud and delivered in such a way as to appear completely legitimate and normal. The true user, who is the owner of the information, would readily identify when decoy information is being returned by the Cloud, and hence could alter the Cloud's responses through a variety of means, such as challenge questions, to

inform the Cloud security system that it has inaccurately detected an unauthorized access.
In the case where the access is correctly identified as an unauthorized access, the Cloud security system would deliver unbounded amounts of bogus information to the adversary, thus securing the user's true data from unauthorized disclosure.

Objective: In this dissertation we aim to achieve, we can eliminate duplicate copies of storage data and limit the damage of stolen data if we decrease the value of that stolen information to the attacker. Validating whether data access is authorized when abnormal information access is detected, and Confusing the attacker with bogus information.

Advantage: The detection of masquerade activity. The confusion of the attacker and the additional costs incurred to distinguish real from bogus information, and The deterrence effect which, although hard to measure, plays a significant role in preventing masquerade activity by risk-averse attackers.

Problem Definition: Many proposals have been made to secure remote data in the Cloud using encryption and standard access controls. It is fair to say all of the standard approaches have been demonstrated to fail from time to time for a variety of reasons, including insider attacks, mis-configured services, faulty implementations, buggy code, and the creative construction of effective and sophisticated attacks not envisioned by the implementers of security procedures. Building a trustworthy cloud computing environment is not enough, because accidents continue to happen, and when they do, and information gets lost, there is no way to get it back. One needs to prepare for such accidents.

II. LITERATURE REVIEW/SURVEY

We defined the notions used in based paper, review some secure primitives used secure deduplication. Symmetric Encryption, Convergent Encryption, Proofs of Ownership (pows), Ramp Secret Sharing, Secure Deduplication.

Symmetric Encryption : In Symmetric Encryption [1,3] explain that notion of security and scheme for Symmetric encryption in concentrate security framework. They give several differ notion of security and analyses the concrete complexity of reduction among them. Then they provide concrete security analyses of various method of encryption using a block cipher, including two most popular methods, Cipher block chaining and counter Mode.

Convergent Encryption : In this Convergent Encryption [4,8] explain mechanism to reclaim space from this incidental duplication to make it available for controlled file replication. Their mechanism includes First one convergent encryption, which enables duplicate files to coalesced into the space of a single file, even if the files are encrypted with different users' keys, and second one SALAD, a Self- Arranging, Lossy, Associative Database for aggregating file content and location information in a decentralized, scalable, fault-tolerant manner. Addresses the problems of identifying and coalescing identical files in the Farsite [5] distributed file system, for the purpose of reclaiming storage space consumed by incidentally redundant content. Farsite is a secure, scalable, server less file system that logically functions as a centralized file server but that is physically distributed among a networked collection of desktop workstations. This paper addresses the problems of identifying and coalescing identical files in the Farsite [6,7] distributed file system, for the purpose of reclaiming storage space consumed by incidentally redundant content. Farsite is a secure, scalable, server less file system that logically functions as a centralized file server but that is physically distributed among a networked collection of desktop workstations. Since desktop machines are not always on, not centrally managed, and not physically secured, the space reclamation process must tolerate a high rate of system failure, operate without central coordination, and function in tandem with cryptographic security.

Proof of Ownership: In proof of ownership [9,11] defined and explore proofs of retrievability (PORs). A POR scheme enables an archive or back-up service (prover) to produce a concise proof that a user (verifier) can retrieve a target file F, that is, that the archive retains and reliably transmits file data sufficient for the user to recover F in its entirety. A POR may be viewed as a kind of cryptographic proof of knowledge (POK), but one specially designed to handle a large file (or bitstring) F. To overcome attacks, they introduce the notion of proofs-of ownership (PoWs), which lets a client efficiently prove to a server that that the client holds a file, rather than just some short information about it. They formalize the concept of proof-of-ownership, under rigorous security definitions, and rigorous efficiency requirements of Petabyte scale storage systems.

Ramp Secret Sharing: In that [11,18]explained Dekey technique by using the Ramp secret sharing scheme (RSSS) [12] to store convergent keys. Specifically, the (n, k, r) RSSS(where n > k > r>= 0) generates n shares from a secret such that First the secret can be recovered from any k shares but cannot be recovered from fewer than k shares, and second no information about the secret can be deduced from any r shares. It is known that when r = 0, the (n, k, O) RSSS becomes the (n, k) Rabin's Information Dispersal Algorithm (IDA) [13]; when r =k-1, the (n, k, k – 1)-RSSS becomes the (n, k) Shamir's Secret Sharing Scheme (SSSS) [14].

Secure Deduplication : In 2008 Mark W. Storer et al. [19,27] developed two models for secure deduplicated storage: authenticated and anonymous. These two designs demonstrate that security can be combined with deduplication in a way that provides a diverse range of security characteristics. In the models they present, security is provided through the use of convergent encryption. This technique, first introduced in the context of the Farsite system [5, 6], provides a deterministic way of generating an encryption key, such that two different users can encrypt data to the same cipher text. In both the authenticated and anonymous models, a map is created for each file that describes how to reconstruct a file from chunks. This file is itself encrypted using a unique key. To enhance the security of deduplication and protect the data confidentiality, Bellare et al. [1] showed how to protect the data confidentiality by transforming the predictable message into unpredictable message. In their system, another third party called key server is introduced to generate the file tag for duplicate check. Q. Wang et al. [21] presented a novel encryption scheme that provides differential security for popular data and unpopular data.

Modeling user behaviors and Decoy Technology :
In 2011Malek Ben Salem et al. [28,32] defined Masquerade attacks (such as identity theft and fraud) are a serious computer security problem. They conjecture that individual users have unique computer search behavior which can be profiled and used to detect masquerade attacks. The behavior captures the types of activities that a user performs on a computer and when they perform them. The use of search behavior profiling for masquerade attack detection permits limiting the range and scope of the profiles they compute about a user, thus limiting potentially large sources of error in predicting user behavior that would be likely in a far more general setting. In 2012 Salvatore J. Stolfo et.al [29] explained a novel approach to securing personal and business data in the Cloud. They propose monitoring data access patterns by profiling user behavior to determine if and when a malicious insider illegitimately accesses someone's documents in a Cloud service.

III. SYSTEM IMPLEMENTATION

After careful analysis the system has been identified to have the following modules:
1. Secure Deduplication
2. User Behavior Profiling
3 .Decoy documents.

Secure Deduplication: Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent (data) compression and single-instance (data) storage. This technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. As the analysis continues, other chunks are compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk. Given that the same byte pattern may occur dozens, hundreds, or even thousands of times (the match frequency is dependent on the chunk size), the amount of data that must be stored or transferred can be greatly reduced.

This type of deduplication is different from that performed by standard file-compression tools, such as LZ77 and LZ78. Whereas these tools identify short repeated substrings inside individual files, the intent of storage-based data deduplication is to inspect large volumes of data and identify large sections – such as entire files or large sections of files – that are identical, in order to store only one copy of it. This copy may be additionally compressed by single-file compression techniques. For example a typical email system might contain 100 instances of the same 1 MB (megabyte) file attachment. Each time the email platform is backed up, all 100 instances of the attachment are saved, requiring 100 MB storage space.

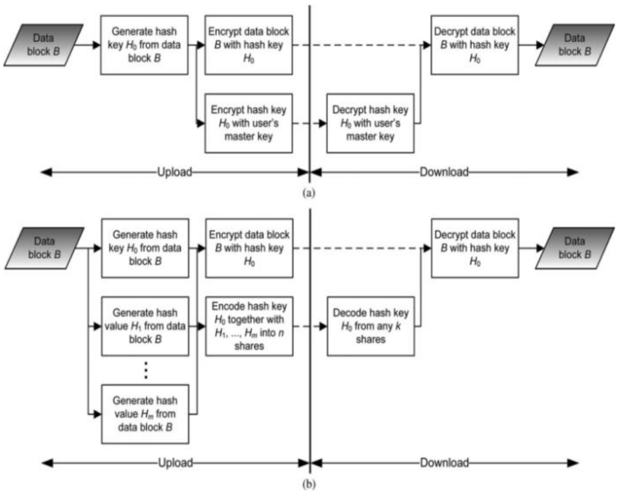
Data block B
Generate hash key H_0 from data block B
Encrypt data block B with hash key H_0
Encrypt hash key H_0 with user's master key
Decrypt hash key H_0 with user's master key
Decrypt data block B with hash key H_0
Data block B
Upload
Download
(a)
Data block B
Generate hash key H_0 from data block B
Encrypt data block B with hash key H_0
Generate hash value H_1 from data block B
Encode hash key H_0 together with H_1, ..., H_m into n shares
Generate hash value H_m from data block B
Decode hash key H_0 from any k shares
Decrypt data block B with hash key H_0
Data block B
Upload
Download
(b)

Fig 1: Secure deduplication
(a)Flow diagram keeping hash key
(b) Flow diagram of Dekey keeping hash key with RSSS.

User Behavior Profiling : We monitor data access in the cloud and detect abnormal data access patterns. User profiling is a well known Technique that can be applied here to model how, when, and how much a user accesses their information in the Cloud. Such 'normal user' behavior can be continuously checked to determine whether abnormal access to a user's information is occurring. This method of behavior-based security is commonly used in fraud detection applications. Such profiles would naturally include volumetric information, how many documents are typically read and how often. We monitor for abnormal search behaviors that exhibit deviations from the user baseline the correlation of search behavior anomaly detection with trap-based decoy files should provide stronger evidence of malfeasance, and therefore improve a detector's accuracy.

Decoy documents : We propose a different approach for securing data in the cloud using offensive decoy technology. We monitor data access in the cloud and detect abnormal data access patterns. We launch a disinformation attack by returning large amounts of decoy information to the attacker. This protects against the misuse of the user's real data. We use this technology to launch disinformation attacks against malicious insiders, preventing them from distinguishing the real sensitive customer data from fake worthless data the decoys, then, serve two purposes: (1) Validating whether data access is authorized when abnormal information access is detected, and (2) Confusing the attacker with bogus information.

DOI 10.17148/IJARCCE.2015.4747

IV. RESULT ANALYSIS

In order to verify the performance of our approach, we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can achieve this through a 'preventive' disinformation attack. We can store Secure Deduplication in the cloud through Convergent Encryption Key Management for insider attacker and monitor them with providing additional security in the previous base paper model by using user behavior profiling and Decoy Technology for outsider attacker. The graph result as shown that the number of unauthorized user access denied and try to upload duplicate file on the cloud.

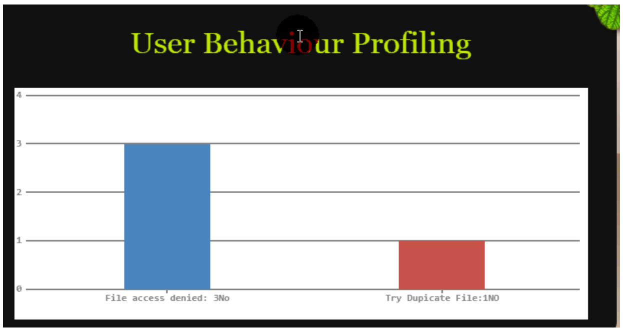

Fig 2: Graph of user behavior profiing and try for deduplication

Secure deduplication services can be implemented given additional security features insider attacker on Deduplication and outsider attacker by using the detection of masquerade activity.

V. CONCLUSION

The basic idea is that we posit that secure deduplication services can be implemented given additional security features insider attacker on Deduplication and outsider attacker by using the detection of masquerade activity. The confusion of the attacker and the additional costs incurred to distinguish real from bogus information, and the deterrence effect which, although hard to measure, plays a significant role in preventing masquerade activity by risk-averse attackers. We posit that the combination of these security features will provide unprecedented levels of security for the deduplication.

ACKNOWLEDGMENT

It is pleasant endeavor to present final report on "Secure Deduplication and Data Security with Efficient and Reliable Convergent Key Management". I avail this opportunity to express my deep sense of gratitude and whole hearted thanks to my guide **Prof. S. S. Kulkarni** (Professor Department of Information Technology) of **Prof. Ram Meghe** Institute of Technology & Research, Badnera for substantial guidance and cooperation in the project work. He has provided all the facilities whenever I need and mostly for his gracious encouragement, advice and guidance to make this project a success.

I also express my gratitude to **Prof. A. P .Bodhke** Principal, Prof. Ram Meghe Institute of Technology & Research, Badnera and **Prof. V. M. Deshmukh** Head of Department of Information Technology Prof. Ram Meghe Institute of Technology & Research, Badnera for constant inspiration and valuable advice.

REFERENCES

[1]M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation. Proceedings of the 38th Symposium on Foundations of Computer Science, IEEE, 1997.

[2] Ayushi "A Symmetric Key Cryptographic Algorithm " International Journal of Computer Applications (0975 - 8887) ©2010 Volume 1 – No. 15

[3] Abdul Wahid Soomro, Nizamuddin, Arif Iqbal Umar, Noorul Amin." Secured Symmetric Key Cryptographic Algorithm for Small Amount of Data" 3rd International Conference on Computer & Emerging Technologies (ICCET 2013)

[4] 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 Proc. ICDCS, 2002, pp. 617-624.

[5] W. J. Bolosky, J. R. Douceur, D. Ely, and M. Theimer, "Feasibility of a Serverless Distributed File System Deployed on an Existing Set of Desktop PCs", SIGMETRICS 2000, ACM, 2000, pp.34-43.

[6] A. Adya, W. J. Bolosky, M. Castro, R. Chaiken, G. Cermak, J. R. Douceur, J. Howell, J. R. Lorch, M. Theimer, and R. Wattenhofer. FARSITE: Federated, available, and reliable storage for an incompletely trusted environment. In Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI), Boston, MA, Dec.2002. USENIX.

[7]R. Anderson and E. Biham, "Two Practical and Provably Secure Block Ciphers: BEAR and LION", 3rd International Workshop on Fast Software Encryption, 1996, pp. 113-120.

[8] P. Golle, S. Jarecki, and I. Mironov. Cryptographic primitives enforcing communication and storage complexity. In "Financial Cryptography '02", volume 2357 of LNCS, pages 120–135. Springer, 2003.

[9] A. Juels and B. S. Kaliski, Jr. Pors: proofs of retrievability for large files. In ACM CCS '07, pages 584–597. ACM, 2007

[10]H. Shacham and B. Waters. Compact proofs of retrievability. In ASIACRYPT '08, pages 90–107. Springer-Verlag, 2008.

[11] A.D. Santis and B. Masucci, ''Multiple Ramp Schemes,'' IEEE Trans. Inf. Theory, vol. 45, no. 5, pp. 1720-1728, July 1999.

[12] G.R. Blakley and C. Meadows, ''Security of Ramp Schemes,'' in Proc. Adv. CRYPTO, vol. 196, Lecture Notes in Computer Science,G.R. Blakley and D. Chaum, Eds., 1985, pp. 242-268.

[13] M.O. Rabin, ''Efficient Dispersal of Information for Security, Load Balancing, Fault Tolerance,'' J. ACM, vol. 36, no. 2, pp. 335- 348, Apr. 1989.

[14] A. Shamir, ''How to Share a Secret,'' Commun. ACM, vol. 22, no. 11, pp. 612-613, 1979.

[15] J. Gantz and D. Reinsel, The Digital Universe in 2020: Big Data, Bigger Digital Shadows, Biggest Growth in the Far East, Dec. 2012. [Online]. Available: http://www.emc.com/collateral/analystreports/idc-the-digital-universe-in-2020.pdf.

[16] A. Yun, C. Shi, and Y. Kim, ''On Protecting Integrity and Confidentiality of Cryptographic File System for Outsourced Storage,'' in Proc. ACM CCSW, Nov. 2009, pp. 67-76.

[17] P. Anderson and L. Zhang, ''Fast and Secure Laptop Backups with Encrypted De-Duplication,'' in Proc. USENIX LISA, 2010,pp. 1-8.

[18] AmazonCase Studies. [Online]. Available: https://aws.amazon.com/solutions/case-studies/#backup.

[19] M.W. Storer, K. Greenan, D.D.E. Long, and E.L. Miller, ''Secure Data Deduplication,'' in Proc. StorageSS, 2008, pp. 1-10.

[20]G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song. Provable data possession at untrusted stores. In ACM CCS '07, pages 598–609. ACM, 2007.

[21] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou. Enabling public verifiability and data dynamics for storage security in cloud computing. In ESORICS'09, pages 355–370. Springer-Verlag, 2009.

[22] M. Bellare, S. Keelveedhi, and T. Ristenpart, ''Message-Locked Encryption and Secure Deduplication,'' in Proc. IACR Cryptology ePrint Archive, 2012, pp. 296-3122012:631.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

[23] Ciphertite data backup. http://www.ciphertite.com/. (Cited on page 3.)

[24] A. Rahumed, H. Chen, Y. Tang, P. Lee, and J. Lui. A secure cloud backup system with assured deletion andversion control. In Parallel Processing Workshops (ICPPW), 2011 40th International Conference on, pages160-167 IEEE, 2011.

[25]Z. Wilcox-O'Hearn and B. Warner. Tahoe: The least-authority _lesystem. In Proceedings of the 4th ACM international workshop on Storage security and survivability, pages 21-26. ACM, 2008.

[26]S. P. Vadhan. On constructing locally computable extractors and cryptosystems in the bounded storage model. In D. Boneh, editor, CRYPTO 2003, volume 2729 of LNCS, pages 61-77. Springer, Aug. 2003.

[27].Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou" A Hybrid Cloud Approach for Secure Authorized Deduplication" IEEE Transactions On Parallel And Distributed System VOL:PP NO:99 YEAR 2013.

[28]M. Ben-Salem and S. J. Stolfo, "Modeling user search-behavior for masquerade detection," in Proceedings of the 14th International Symposium on Recent Advances in Intrusion Detection . Heidelberg: Springer, September 2011, pp. 1–20.

[29] Salvatore J. Stolfo, Malek Ben Salem and Angelos D. Keromytis "Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud" IEEE Symposium On Security And Privacy Workshop (SPW) YEAR 2012

[30] I.Sudha1, A.Kannaki2, S.Jeevidha3" Alleviating Internal Data Theft Attacks by Decoy Technology in Cloud", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.3, March-2014, pg. 217-222.

[31] B. M. Bowen and S. Hershkop, "Decoy Document Distributor: http://sneakers.cs.columbia.edu/ids/fog/," 2009. [Online]. Available: http://sneakers.cs.columbia.edu/ids/FOG/

[32] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou "Secure Deduplication with Efficient and Reliable Convergent Key Management" IEEE Transactions On Parallel And Distributed Systems, VOL. 25, NO. 6, JUNE 2014..

[33] Mr N.O.Agrawal, Prof. S.S.Kulkarni"Secure Deduplication and Data Security with efficient and reliable CEKM" IJAIEM Transition On parallel And Distributed System,VOL.3,Issue. 11,November 2014.

DOI 10.17148/IJARCCE.2015.4747