International Engineering Journal For Research & Development

Vol.5
Issue 5

E-ISSN NO:-2349-0721

Impact factor : 6.549

SECURE DEDUPLICATION WITH CONVERGENT ENCRYPTION KEY MANAGEMENT

[1]Kalyani R. Gawande, [2]Dr. Roshan. R. Bhure

M.E Student, Dept. of Computer Engineering, P. R. Pote Amravati,India[1], Ph.D. Scholar, Dept. of Computer Engineering, P. R. Pote Amravati, India[2]

ABSTRACT

Secure deduplication is a technique to remove duplicate copies of storage data, and provides security to them. To reduce storage space and upload bandwidth in cloud storage deduplication has been a excellent technique. For that purpose convergent encryption has been extensively adopt for secure deduplication, condemn issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. The fundamental idea in this, is that eliminate duplicate copies of storage data and limit the damage of stolen data if decrease the value of that thieved information to the attacker. Here it makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. Here, it first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them. Such a baseline key handling scheme generates an enormous number of keys with the expand number of users and requires users to dedicatedly protect the master keys. To this end, here offeredDekey, User Behaviour Profiling and Decoys technology. New construction ,Dekeyin 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 feasibility studies, implementation ofDekey by using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments. Two technology,User profiling and decoys, then, serve the 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. The combination of these security features will provide unprecedented excellent levels of security for the deduplication in insider and outsider attacker.

Keywords: Deduplication, Convergent encryption key management, Dekey, User behaviour profiling, Decoy Technology

INTRODUCTION

The advent of cloud storage motivates enterprises and organizations tooutsource data storage to third-party cloud providers, as evidenced by many real-life case studies. One significant challenge of today's cloud storage services is the management of the ever-increasing volume of data. Making data management scalable, deduplication has been a well-known technique to reduce storage space and upload bandwidth in cloud storage. Keeping multiplex data copies with the equivalent content, deduplication discard redundant data by keeping only one physical copy and referring other redundant data to that copy. Extracting duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. It is, an arising challenge to execute secure deduplication in cloud storage. Also convergent encryption has been extensively adopted for secure deduplication, a vital issue of making convergent encryption practical is to efficiently and reliably control

a huge number of convergent keys. Here the challenge of today's cloud storage services is the management of the ever-increasing volume of data. For scalable data management deduplication we use convergent Encryption for secure deduplication services.

From a user's point of view, data outsourcing raises security and privacy concerns. Trusting to the third-party cloud providers to properly enforce confidentiality, integrity checking, and access control mechanisms against any insider and outsider attacks. How However, deduplication, while improving storage and bandwidth efficiency, is compatible with Convergent key management. Here, traditional encryption requires different users to encrypt their data with their own keys. Many deal have been made to secure remote data in the Cloud using encryption and standard access controls. Actually,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 creativeconstruction of effective and sophisticated attacks not envisioned by the implementers of security procedures. Cloud computing environment is not enough, because accidents continue to happen, and for that, information gets lost, there is no way to get it back. One needs to prepare for such accidents.

Related Work

The prevalent data for each user stored in the cloud and the demand of long term continuous assurance of their data safety, the problem of monitoring correctness of data storage in the cloud becomes even more challenging. Cloud Computing is not just a third party data warehouse. The data which is stored in the cloud may be habitually updated by the users, including insertion, deletion, modification, appending, reordering, etc. One of the critical challenge of today's cloud storage services is the management of the ever-increasing volume of data.The mechanism to reclaim space from this incidental duplication to make it available for controlled file replication. Their mechanism includes that, the 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 gathering file content and location information in a decentralized, scalable, fault-tolerant manner addresses the problems of identifying and coalescing identical files in the Farsite[4] distributed file system, for the purpose of reclaiming storage space consumed by incidentally redundant content. Farsite is a strong, scalable, serverless file system that logically functions as a centralized file server but that is physically distributed among a networked collection of desktop workstations.

Proposed System:

We offer security in insider attacker as well as outsider attacker and monitoring them we use for that Dekey, user behaviour profiling and Decoy Technology. A new construction, Dekey in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers.Here the Dekey Technology using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments we propose a new construction called Dekey, which provides efficiency and reliability guarantees for convergent key management on both user and cloud storage sides.Dekey is presenting to provide efficient and reliable convergent key management through convergent key Deduplication and secret sharing. This technology supports both file-level and block level Deduplication. Security analysis demonstrates that Dekey is secure in terms of the definitions specified in the proposed security system model. Here, Dekey remains secure even the adversary controls a limited number of key servers. By implementingDekey using the Ramp secret sharing scheme that enables the key management to adapt to

different reliability and confidentiality levels. Our rating demonstrates that Dekey incurs limited overhead in normal upload/download operations in realistic cloud environments.

We will add this system also in project as mentioned below

The basic idea is that we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can reach this through a 'preventive' disinformation attack. We posit that secure deduplication services can be implemented given two additionalsecurity features: User Behaviour Profiling & Decoy.

SECURE DEDUPLICATION

Convergence encryption

Convergent encryption is a deterministic encryption algorithm, which can be essentially a special symmetric encryption scheme, the special point is that the encryption key is not a user's private decision, but rather is determined by the text to be encrypted, using the hash value of the text as the encryption key, thus guaranteeing that, although in different user encryption, as long as the same text is to be encrypted, the same cipher can be obtained. Based on this property, the symmetric encryption method of convergent encryption is widely used in the heavy scheme of all kinds of encrypted data.

Formally, a convergent encryption scheme can be defined with four primitive functions:

- KeyGenSE (M) :- K is the key generation algorithm that maps a data copy M to a convergent key K;
- EncryptSE(k,M)_:- C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs aciphertext C;
- DecryptSE (k,C):-M is the decryption algorithm that takes both the ciphertext C and the convergent key K as inputs and then outputs the original data copy M; and
- TagGenCE (M) :- T(M) is the tag generation algorithmthat maps the original data copyM and outputs a tag T(M). We allow TagGenCE to generate a tag from the corresponding ciphertext as in [5], by using T(M) = TagGenCE (C), where C= EncryptSE(k,M)

Message Authentication Code

The Message Authentication Code (MAC) is a short message that validates the message and provides integrity and authenticity to the message. Here, we apply the message authentication code to realize the integrity of the cloud medical data. Also it is easy to construct a hash function; by entering a user's private key and an arbitrary length of the file, you can get a MAC. After that, a hash is obtained by the downloader based on its own private key and the received file, and if the hash is equal to the MAC of the file, the integrity and authenticity of the file is guaranteed. Else, it indicates that the file has been adjusted.

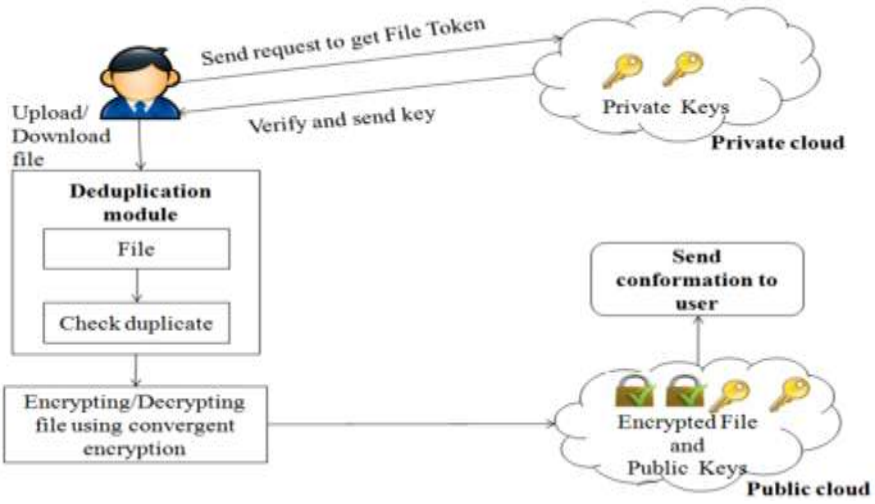

Fig no 1: Architecture of Secure Deduplication

User Behaviour and Decoy Technology

Kindly, it is expect that access to a user's information in the Cloud will exhibit a normal means of access. A well-known User profiling technique that can be apply here to model how, when, and how much a user accesses their information in the Cloud. The 'normal user' behaviour can be continuously check to determine whether abnormal access to a user's information is occurring. For safety, this method of behaviour-base security is commonly used in fraud detection applications. And this such profiles would naturally include volumetric information, how many documents are typically read and how often. This kind of simple user specific features can serve to detect abnormal Cloud access based partially upon the scale and scope of data transfer .

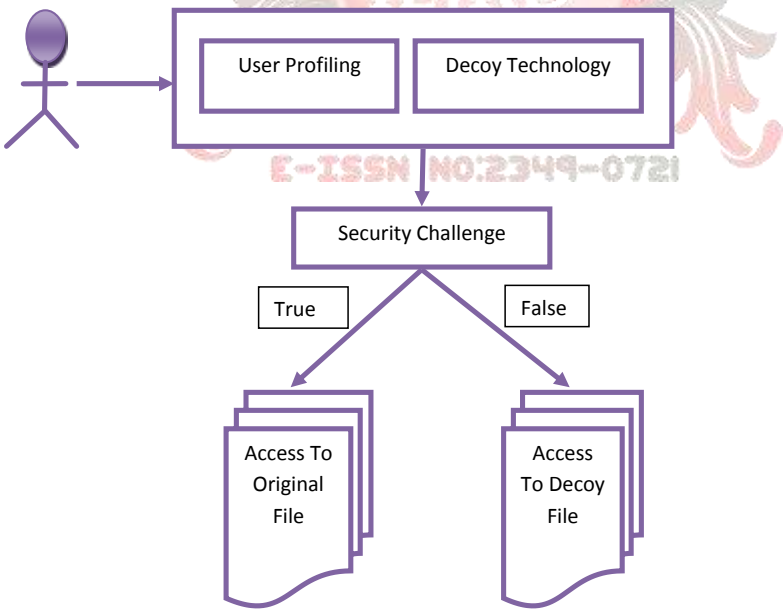

Fig no.2: Outsider Attacker Security

RESULT ANALYSIS

We analyzed our datasets around multiple aspects for dedup space savings and used those findings to design our primary data deduplication system. In this section, we evaluate some other aspects of our data deduplication system that are related to post-processing deduplication. Post-processing deduplication throughput. Using the dataset, we examined post-processing deduplication throughput, calculated as the amount of original data

processed per second. An entry-level HP ProLiant SE326M1 system with one quad-core Intel Xeon 2.27 GHz L5520 and 4 GB of RAM was used, with a 3- way RAID-0 dynamic volume on top of three 1TB SATA 7200 RPM drives. To perform an windows-to-windows comparison, we ran a post-processing deduplication session multiple times with different indexing options in a controlled environment. Each deduplication session uses a single thread. Moreover, we ensured that there were no CPU-intensive tasks running in parallel for increased measurement accuracy. The baseline case uses a regular index where the full hash (SHA-256, 32 bytes) and location information (16 bytes) for each unique chunk is stored in RAM. The optimized index uses the RAM space efficient design.

Upload and Download: we estimated the time utilization of arbitrary records upload in the cloud, or records download from remote servers. Our method consists at first, to generate a random data file of a fixed size.

Average Time in "s"		
Size	Upload	Download
10	0.338	0.192
10^2	0.329	0.191
10^3	0.337	0.187
10^4	0.325	0.192

0.4

Table 1: Average time to upload and download file

Second approach for Data Security we used JFreeChart tool for the detection of masquerade activity in graphical format of user profiling behavior and decoy.

Table 2: User Profiling Detection

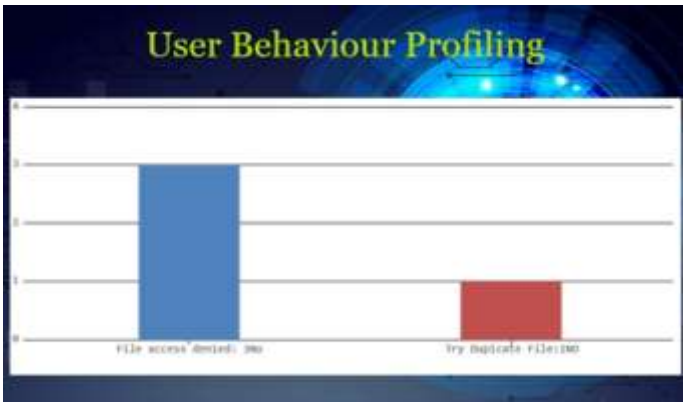

Table 3: Graphical User profiling detection

CONCLUSION

We can limit the damage of stolen information if we decrease the value of that stolen information to the attacker. We can achieving this through a 'preventive' disinformation attack. 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. By the combination of these security features will provide unprecedented levels of security for the deduplication.

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, ArifIqbal 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.