De Duplication and Encryption of Data on Cloud

Rishabh Singh, S.Ganesh Kumar

Abstract: ***The duplication of information on the importance of data compression techniques to eliminate duplicate copies of repetitive information. It is widely used in the cloud save storage space and storage to reduce the amount of bandwidth. In the hold, as long as the secret of the duplication of the data supporting the sensitive, flocked to the encryption, the knowledge of the art is the divine plan, is to encrypt the front of the outsourcing. For more information security protection, this is the first project tries to address the issue formally authorized the duplication of data. Unlike traditional systems, doubling the double differential privileges of users are considered, and patterned in addition to that information. The new duplication and many buildings are present and we allowed supports duplication of control in a hybrid cloud architecture. The analysis showed that the security system is about defining a secure certain terms in the proposed security model. Proof of concept, as with any kind of work implement copy control system proposed by our test bench for experiments on our prototype. In that they show the work of our normal like very little, proposed from the above is the cause of the system proposed by the authority of the command of the head of the copy compared to it.***

I. INTRODUCTION

Provides users with a cloud computing resources "virtualized" there seems to be no end to all the services in the Internet while hiding the details of the platform and implementation. Or cloud computing service providers today offer highly available storage resources, and massively parallel resources at relatively low cost. And by a rivulet he sendeth forth And just as cloud computing, more and more data stored in the cloud, and it is shared by the users on the proper to the privileges of which they are the rights of access to the data storage. One of the major challenges of cloud storage data management services is a steadily increasing volume. The scalable data management in a cloud computing, replication technique is well known and has attracted more attention.

Data replication is a specialized technical information such as the pressure to eliminate duplicate copies of data storage end. This technique is used, and can also be applied to the network to improve storage utilization passes the data to reduce the number of bytes to send. Again, instead of maintaining multiple copies of data content, data deduplication eliminates redundant data in addition to the one to be independent of others and physical copy copies.

Revised Manuscript Received on April 15, 2019.

Rishabh Singh, Computer Science and Engineering, SRM Institute of Engineering and Technology, Kattankulathur, Chennai

S.Ganesh Kumar, Associate Professor, Department of Computer Science and Engineering, SRM Institute of Engineering and Technology, Kattankulathur, Chennai

Possible duplication of file level or block level. In fact, the duplication of the files, eliminates duplicate is to be of the same files. Take the square is a block level, which eliminates duplicate data blocks are to be non-identical files. Although deduplication brings many benefits information, data security and privacy sensitive users, but problems arise because it is exposed to internal and external attacks. The traditional encryption, data deduplication should not be running secret information. It encrypts / decrypts data and the model would meet a key obtained by calculating a cryptographic relay information content copy. Encryption key information after generating, users have in the encrypted key and send it to the cloud. And is one of the encryption of the data of the deterministic because the deed of the contents of the same thing, knowledge, therefore, the pattern indeed of the key to generate ciphertext is the same as the same, and the battalions came to the same thing. To prevent unauthorized access to, which is also the property is required by the secure protocol, such as by way of proving this is to provide to the user is already a copy of the file is the same as was found to be a double. After the verification, subsequent users to the servant, assigned to it is of the same file rule without having to download the same file.

A user can download the file is encrypted with a pointer to a server that can be decrypted only by the converging given keys to their owners. Thus, encryption Converge allows the cloud to perform a duplication of encrypted texts and proof of ownership of preventing unauthorized user access to the files. We will not deal with the differential of the duplication of the power of authority prior to the duplication of the systems, however, which is of major importance in many applications. So great an abundance of each one of the privileges with the duplicate ratio of the user account is permitted to boot. Every file uploaded to the cloud and is linked to the right to specify a set of users allowed access to the square to perform check in files.

Subject to certain files duplicate to request verification, the user has to use its own brand for the file and inputs. A user cannot find the files there is a twin of the privileges, if and only if there is a copy of the file is stored in the cloud, and in like proportion. For example, in the company, as many as possible to use the privileges of my kingdom, it shall be performed.

In order to reduce costs and optimize management and data storage will be moved to the server provider (SSP) in the ratio of public cloud with certain rights and for the technology to be applied to replace one copy of the same file. For confidentiality reasons, there are certain files that use the check is encrypted copy right it should be allowed access control.

So they converge, according to the traditional the duplication of the systems of encryption, while ensuring the secret of as much as to say, do not check for duplicates by means of the difference of the support it has. In other words, is not included in the square of the differential benefit based on encryption technology converge. And this does not seem to be contrary to the trust, if this is at the same time, and by the authority of the duplication of the difference of the power of the power of the duplication of the.

II. SCOPE OF THE PROJECT

A data deduplication technique widely used to store data and storage data network, and to minimize costs by detecting and eliminating the redundant data.

Objective

To principal object of the developer undergraduate and distributed through the duplication of many servers and storage.

a. Existing System

Data deduplication systems, private cloud as a proxy used to allow owners/ users to safely perform data in duplicate checking the difference. This is practical architecture and attention than he can help researchers. The clouds, to meet the public storage by making use of the knowledge of them were able to be removed from the content owners, so long as, in the private cloud is given to the operation of the work is carried on.

Data deduplication is a specialized technical information such as the pressure to eliminate duplicate copies of data storage end. This technique is used, and can also be applied to the network to improve storage utilization passes the data to reduce the number of bytes to send. For maintaining multiple copies of the same information content with redundant eliminates duplication of data retention is only one physical copy, and others referring to this unacceptable given copies.

Possible duplication of file level or block level. That with the eyes you have to double the levels of a tablet, a copy of the file it to a double of the same is excluded. Take the square is a block level, which eliminates duplicate data blocks are to be non-identical files. Identical copies of the data of the distinct subjects, so that they wearied themselves to the eyes of the ciphertext it you have to double the lead.

Disadvantages

- The traditional encryption, secret information should not be running duplication of information.
- The copy of the identical copies of the data of the distinct subjects, so that they wearied themselves to the eyes of the ciphertext it you have to double the lead.

b. Proposed System

At this proposed work better security system. Specifically, it has the support enhanced the security of the file from the encrypting it so that it is the key to the things that the differentials in years, he has. This way, users without correspondence, not only to check in to perform the copy. In addition, users could be able to decrypt ciphertext and requires that the intimate union of the S-CSP. Our analysis showed that the security system is secure in the definitions of certain terms of the proposed security model.

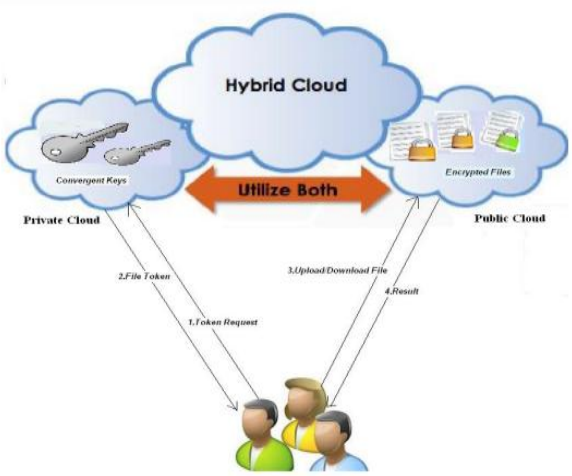
From the converging of encryption before them a royal statute, and to the secret of their knowledge, while allowing the duplication of the. It encrypts / decrypts data and the model would meet a key obtained by calculating a cryptographic relay information content copy. Encryption key information after generating, users have in the encrypted key and send it to the cloud. And is one of the encryption of the data of the deterministic because the deed of the contents of the same thing, knowledge, therefore, the pattern indeed of the key to generate ciphertext is the same as the same, and the battalions came to the same thing. To prevent unauthorized access to, which is also the property is required by the secure protocol, such as by way of proving this is to provide to the user is already a copy of the file is the same as was found to be a double.

Advantages

- The user is not allowed to listen to copy files marked with its due rights.
- Behave in the present schema to support the key differentials enhanced security by encrypting the file command.
- Reduce the size and integrity checking tags for storage. The security of a strong garrison, and the proceeds from the secret from the duplication of the data.

→System Architecture

The system architecture and system tests basis for defining the essential features of the design and basic elements forming part of the system. The system architecture and provides users with the vision of architects vision of what needs to be a reason and what to do, and because of the way in which evolve and strives to maintain integrity of that vision by the manufacturer. .



→Proposed Algorithm

Blow fish algorithm

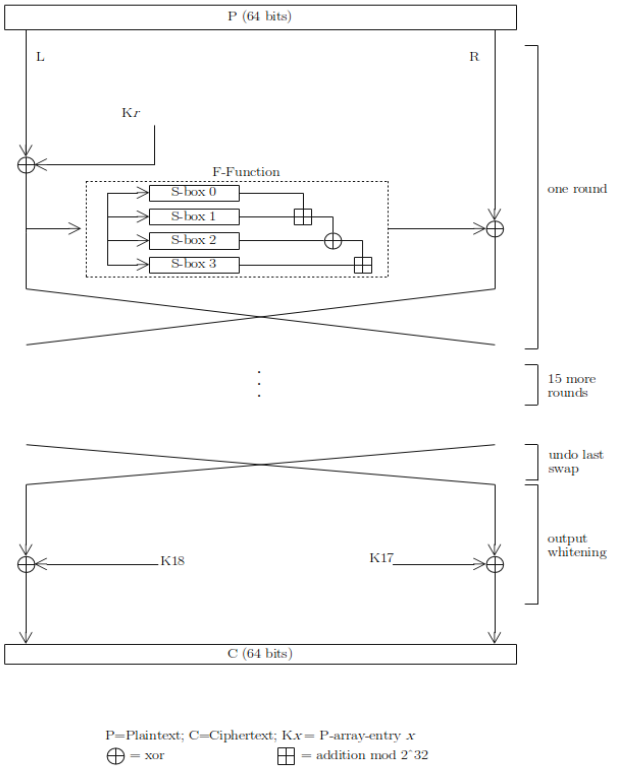
The map shows the Blowfish encryption routine. Each row represents 32 bits. There are five sub-arrays, array entries at P 18 (x k to a diagram to prevent confusion with the plaintext) and four S-boxes with 256 entries (S0 S1, S2 and S3).

Each 4 R, again, depends on the actions, first, the XOR left half (50) LONGINA given r through entry th table by the second use that information XORed approach to the role Blowfish, F, third, XOR as a present right in the middle of the F output (R) data In brief, swap L and R.

The function F 32 bits, an approach divides into four quarters bits in eight quarters, and uses an approach to the flocks S, 8-bit to 32-bit mounting exitibusque produce outputs. At every discutiendæ outputs are added to produce XORed 232 and the end of the 32-bit output (see picture in the upper right corner).

After 16 round pick last exchange manifold, with the K18 and K17 50 to the XOR R (output of washing).

Encryption and decryption is not exactly the same as P1, P2, ..., P18 are in reverse order. It is not the custom is not so, it is clear that, commutative and xor. Error contrary to the common order of encryption as decryption algorithm (i.e.first XORing P17 and P18 to the ciphertext block, then using the P-entries in reverse order).

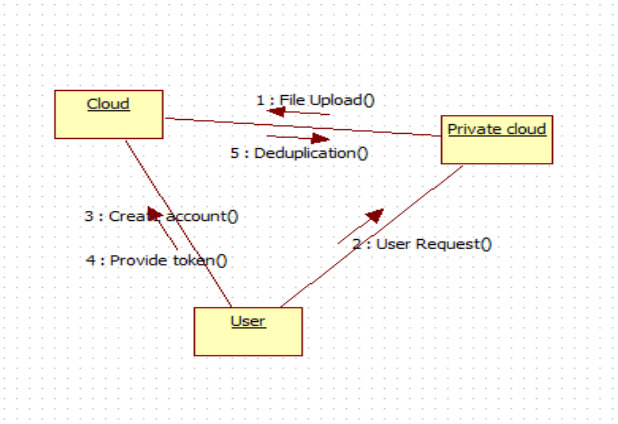


a. Chunking Technique for Deduplication

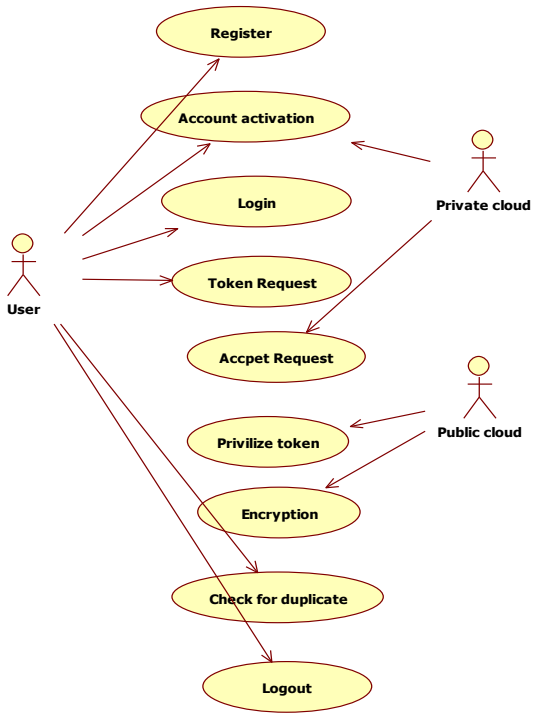
Cutting into pieces to process many files in a file called into smaller pieces. However, in some applications, to interact with the compression of the information, the data, and the data synchronization by way of replication, the ratio of the detection performance also out of that which his decision is copy the career option. 500 (left Content-left) to the files is split into pieces of varying length where the break points are defined by certain internal features of the file. Sure dissimilar parts, much more lumps resist byte various recipes. Thus, these songs are from the more also to me, and that of the probability to find and in the duplicate is to be among the files in the files. However, as the locust swarm calculations require additional 500 points algorithms, which can be computationally expensive for some applications. In previous work (Widode et al., 2016), according to the algorithm of the relay 500 is in the system and the system of others delayed replication process. This shows the need for the proposed switching speed is hashing. Instead of using hashes, RAM bytes in the value uses to declare break points. A window algorithm uses certain variable size average and maximum window to get a point byte value that corresponds to the cutoff. Byte maximum value is located in the block included in the block, and the boundary. Of Ram, that sealing admits that, and, as was said in the minority of cases, without prejudice to the 500 to the similitudes of things. We compared the ram deduplication systems based on existing relay is hashing. Experiment and the results of our algorithm is proposed from the other, to show that it is higher than the second storage bytes throughput, and the algorithms by means of therapy.

→UML Diagram

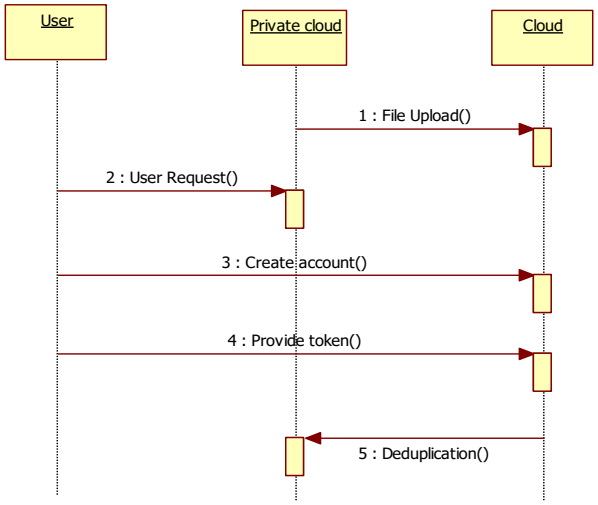
A. Collabration Diagram



B. Usecase Diagram



C. Sequence Diagram



III. MODULE DESCRIPTION

Modules

- User Module
- Server start up and Upload file
- Secure De duplicate System
- Download file

User Module

In this module, Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first. At the very least, you to provide an emails address, username, password, display name, and whatever profile fields you have set to required. The display name is what will be used when the system needs to display the proper name of the us

USER LOGIN→ CHECK VALID OR NOT→ DATABASE

Server start up and upload file

The user can start up the server after cloud environment is opened. Then the user can upload the file to the cloud.

START →AFTER USER UPLOADS DATA →DATABASE

Secure de duplication system

To support authorized de duplication the tag of a file F will be determined by the file F and the privilege. To show the difference with traditional notation of tag, we call is file token instead. To support authorized access a secret key KP will be bounded with a privilege p to generate a file Token. De duplication exploits identical content, while encryption attempts to make all content appear random; the same content encrypted with two different keys results in very different ciphertext. Thus, combining the space efficiency of de duplication with the secrecy aspects of encryption is problematic.

UPLOAD DATA → CHUNK PROCESS →DE DUPLICATION PROCESS

Download File

After the cloud storage, the user can download the file based on key or token. Once the key request was received, the sender can send the key or he can decline it. With this key and request id which was generated at the time of sending key request the receiver can decrypt the message.

DATABASE → USER CAN DOWNLOAD OR VIEW THE FILE

IV. LITERATURE SURVEY

Title 1: Avoiding the disk bottleneck in the data domain deduplication file system (Author:B. Zhu, K. Li, and R.H. Patterson)

Deduplication disk storage has become the next-generation storage system for the protection of corporate data. Replace the tape library. Deduplication removes redundant data segments to compress data in a very compact size than a mechanical one of the world yet to be passed, and in the back of trucks. With A having the data seemed to smile on the crucial the protection rate is high, for the most part is greater than 100 MB / s, let, out of which is perfected in the back of the shortly come to pass.

This article describes the three technique in a given system Domain deduplication document production bottleneck in the dish. These techniques include: (1) In summary, the compact structure of the data in the memory identify new segments; (2) The arrangement informed by the flow segment, is to place the data in the prescribed manner to improve She is accessed segments; and (3) the protection orders may be given cache bringing localized, keep the things that are in these parts, fingerprints duplicate parts to achieve a high hit rate. At the same time are able to remove 99% of the world -Access given deduplication workload of the real world. These techniques can create a modern, two-door system in two minds 90% to 15 bands act as the process used to innate drives and 100 MB / s flow rate is one and 210 MB / s multi-stream through the ages.

Title 2: Message-Locked encryption and secure deduplication (Author: M. Bellare, S.Keelveedhi, and T.Ristenpart)

We formalize the new cryptographic teach the people MLE (Message-Locked encryption) where the key they have a decryption encryption graduated. Provides MLE and the means of achieving secure deduplication (outsourced obtained compact storage) currently being objective targeted by many cloud storage providers. We also provide privacy and definitions. They call consistency and integrity of the tags. We are accomplished theoretical contributions. In practice, we provide for the safety analyzes of a natural family ROMMLE is deployed between the devices.

Title 3: Dupless: Server aided encryption for deduplicated storage (Author: S.Keelveedhi, M. Bellare, and T.Ristenpart)

The storage service providers in the cloud as Dropbox, Mozy; to make the space of deduplication He saved others; hide themselves in the example of the downloaded files. to customers The

savings of those files, however, are typically encrypt the grandchild. Message encryption locked (most He concurred encryption) solves The tension. However, it is a subject in itself brute force attack to recover the files, which falls Known set. It is our purpose that it provides to society is secure storage deduplicated resistant to brute force attacks and make it into a system called DupLESS. With DupLESS, one who has got the keys of his host, and one messenger to customers a key way to Protocol Server PRF perceived by the senses. He will not suffer him to be the things derived from these to store data encrypted on the server. office, contact the office to perform deduplication strong and manages the privacy of their guarantees. Please show us the encryption, which is broken for the storage of deduplicated saving money and space, and performance close to him, and follow to be able to storage service to use the data in plain text.

Title 4: Interactive message-locked encryption and secure deduplication (Author: M. Bellare, S.Keelveedhi)

This paper examines the question of the safe storage of data outsourced. To enable the duplication of information. We are the first only then does as she can, a letter, which was then one to another, and is dependent in the hidden depths Simulacrum parameters public. A new formula can do Can quarter. She is worried that his mind locked Message encryption (MLE) Message encryption interactive locked in a first-past work (IMLE) where they upload and download protocols. It is our diet, providing although they are not only linked to information security the image of the parameters, standard and was lying in the public or the banner on to an example. We there will be a decrease in the speed, may be made explicit, it is not that they are in the practice of the assumption of the additional systems of non-existing deduplication is now interactive.

Title 5: Fast and Secure laptop backups with encrypted de-duplication (Author: P. Anderson, and L. Zhang)

Now a great multitude of men, the supply of proper and personal data storage. Laptops or corporate data on home computers. Often these wear and low or intermittently Conectividad theft or hardware failure. Conventional backup for this, neither of the solutions of a neat and investors often emergency plans to cope office. This paper receives from a knowledge of the algorithm described by the above speed issues, which is common among the users he sold; and the needs of the back of the store. This algorithm Talk of encryption which is a part of such support, need for personal privacy information. It also supports a unique feature that allows the detection common to all, under the trees, eliminating the need to go to the back of the query in each file system. We describe any implementation

V. CONCLUSION

In this paper, it is the concept of the security of the information by the authority of the duplication of the data and proposed to protect, for it was for the check in the differential between the user privileges, copy the. The duplication of the duplication sustains the power of the new buildings, We, have followed many had been intrusted, in the hybrid cloud architecture, in which there are signs of a personal server in the cloud over the duplication of the keys of the power generated are in the private. Our analysis shows that the security systems are not secure in terms of certain internal and external security attacks on the proposed model. For example, we implemented the system, and the model for our proof of the conception of the authority, and power, and set on an exemplary prototype in a laboratory in either of two experiments. It has been shown by us under our system results in a copy of the same distance from the small overhead, and compared to the collision encryption transit network.

REFERENCES

1. B.Zhu,K.Li, and R.H. Patterson, ”Avoiding the disk bottleneck in the data domain deduplication file system,” in 6th USENIX Conference on file and storage Technologies, FAST 2008,February 26-29,2008,San Jose,CA,USA.USENIX,2008,pp. 269-282.
2. 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, ser. Lecture Notes in Computer Science,Vol. 7881.Springer, 2013,pp. 296-312.
3. S.Keelveedhi, M.Bellare, and T.Ristenpart, “Dupless:Server aided encryption for deduplicated storage,” in Proceedings of the 22th USENIX Security Symposium,Washington, DC, USA, August 14-16, 2013. USENIX Association, 2013,pp. 179-194.
4. M.Bellare and S.Keelveedhi, “Interactive message-locked encryption and secure deduplication,” in Public-Key Crptography - PKC 2015 – 18th IACR International Conference on Practice and Theory in Public-Key Cryptography, Gaithersburg, MD, USA, March 30- April 1,2015, Proceedings, Ser.Lecture Notes in Computer Science, vol.9020. Springer, 2015,pp. 516-538.
5. P.Anderson and L. Zhang, “Fast and secure laptop backups with encrypted de-duplication,” in Uncovering the Secrets of System Administration: Proceedings of the 24th large Installation System Administration Conference, LISA 2010, San Jose, CA, USA, November 7-12, 2010. USENIX Association, 2010.
6. 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. ACM, 2008,pp.1-10.
7. A.Rahumed, H.C.H.Chen, Y.Tang, P.P.C.Lee, and J.C.S.Lui, ‘A secure cloud backup system with assured deletion and version control,” in 2011 International Conference on Parallel Processing Workshops, ICPPW 2011, Taipei, Taiwan, Sept.13-16, 2011. IEEE Computer Society, 2011,pp.160-167.
8. P.Puzio, R.Molva, M.Onen, and S.Loureiro,”Cloudedup: Secure Deduplication with encrypted data for cloud storage,” in IEEE 5th International Conference on Cloud Computing Technology and science, Cloud com 2013, Bristol, United Kingdom, December 2-5, 2013, Volume 1.IEEE Computer Society, 2013,pp.363-370.
9. 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, FC 2014, Christ church, Barbados, March 3-7,2014, Revised Selected Papers,ser.Lecture Notes in Computer Science, vol.8437. Springer, 2014,pp.99-118.
10. E. Fujisaki and T.Okamoto, “Secure integration of Asymmetric and symmetric encryption schemes,” J.Cryptology,vol. 26, no. 1, pp. 80-101, 2013.