International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 7, July-2018

Achieving Data Consistency in the Bigdata Systems by Providing a Secure Deduplication Scheme on Encrypted Data

Bala Krishna Veerala, Professor

ABSTRACT—*In the cloud environment, if we want store or upload any data into the cloud, we require minimum security to our data. Presently, we can encrypt our data based on some access policies which are related to the data owner as well as data user. If we add attribute set to the access policies to access the data from cloud, then we can say this is the Attribute-Based Encryption (ABE). However we used the ABE scheme, we are not able to provide secure deduplication to the cloud data. To improve the secure deduplication, in this paper we are implementing an attribute-based storage system and it can useciphertext-policy attribute-based encryption (CP-ABE) and supports secure deduplication. By using this proposed scheme, we can save the cloud storage space and also we can provide security to the cloud data.*

Keywords: Cloud Storage, Atribute-Based Encryption, CP-ABE, Deduplication

1. INTRODUCTION

Cloud computing is a distributed computing architecture in which the computing assets along with hardware, software program, processing strength are introduced as a service over a community infrastructure. The cloud computing model lets in the users to access facts and different assets from anywhere that a network connection is to be had. Cloud offerings and packages can also require all preferred protection capabilities which include information confidentiality, integrity, privateness, robustness and access control. Hence securing the ought to and its facts is a tough assignment. There are numerous cryptographic techniques to at ease the records stored in cloud storage systems.

More concretely, the agreement between the cloud and its customers must be verifiable (at least to a point) and the capacity to discover failures, without depending completely on the cloud company's document, may be useful to the customers. For example, it could be important to directly discover that a service does no longer understand its beneficial specification; or that it generously stocks personal facts with the world; or that it's far down, underperforms, or if its number one safety controls appear to be failing. These records may be mainly beneficial for vital applications which consist of medicinal drug or banking and facilitate their procedure of adopting cloud technologies. In addition, verification equipment to check the ones additives can assist purchasers pick out and pick a particular cloud provider.

Identity-Based Encryption (IBE) permits for a sender to encrypt a message to identification without access to public key certificates. The ability to do public key encryption without certificates has many realistic packages. One not unusual function of all preceding Identity-Based Encryption systems is that they view identities as a string of characters. In this paper we suggest a new type of Identity-Based Encryption that we call Fuzzy Identity-Based Encryption wherein we view identities as a set of descriptive attributes.

With the swiftly growing quantities of records produced international, networked and multi-consumer storage systems have become very famous. However, worries over statistics protection nevertheless save you many customers from migrating facts to remote storage. The traditional answer is to encrypt the data before it leaves the proprietor's premises. While sound from a protection attitude, this approach prevents the storage company from correctly making use of storage performance capabilities, inclusive of compression and deduplication, which could permit top-quality utilization of the resources and consequently lower service fee. Client-side records deduplication in particular guarantees that more than one uploads of the same content material simplest consume network bandwidth and storage area of a single add. Deduplication is actively used by some of cloud backup companies (e.g.Bitcasa) in addition to numerous cloud offerings (e.g.Dropbox).

2. RELATED WORK

The advanced industrial deduplication answers can't manage with encrypted facts. Existing answers for deduplication are at risk of brute-force assaults and can't flexibly support information access manage and revocation. It raises issues relating to safety and ownership. Many users are probable to encrypt their facts before outsourcing them to the cloud storage to preserve privacy, but this hampers de duplication due to the randomization property of encryption.

S. Keelveedhi, M. Bellare, and T. Ristenpart studied the trouble of presenting comfy outsourced storage that each helps deduplication and resists brute-pressure assaults. They composed a framework, DupLESS that joins a CE-type base MLE conspire with the ability to accomplish message-determined keys with the assistance of a key server (KS) shared among an accumulation of clients. The customers have connection with the KS by methods for a convention for unaware PRFs, ensuring that the KS can cryptographically mix in puzzle material to the in step with-message keys while mastering nothing about documents saved with the aid of customers. These mechanisms ensure that DupLESS presents sturdy security towards external assaults which compromise the Storage Service (SS) and communique channels (nothing is leaked past document lengths, equality, and access styles), and that the safety of DupLESS gracefully degrades within the face of comprised structures. Should a consumer be compromised, learning the plaintext underlying every other customer's ciphertext calls for mounting a web bruteforce assaults (which can be slowed by using a fee-restrained KS). Should the KS be compromised, the attacker needs to still attempt an offline brute-pressure assault, matching the guarantees of conventional MLE schemes.

V. Goyal, O. Pandey, A. Sahai, and B. Waters created a machine for Ciphertext-Policy Attribute Based Encryption. Their device lets in for a shiny new sort of scrambled inspire section to oversee in which client's non-open keys are particular by utilizing an arrangement of properties and a festival encoding information can determine a scope over these qualities indicating which clients are equipped for decode. Their contraption lets in approaches to be communicated as any monotonic tree inspire passage to structure and is impervious to conspiracy assaults wherein an assailant may achieve two or three private keys. At long last, they outfitted a usage in their framework, which incorporated a few enhancement methods.

Farsite is a dispensed record machine which turned into proposed by using J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer that offers protection and reliability with the aid of storing encrypted replicas of every report on multiple computer machines. To unfastened area for storing those replicas, the gadget consists of by the way duplicated files, which include shared documents amongst workgroups or more than one customer's copies of not unusual utility packages. This includes a cryptosystem that allows same documents to be coalesced although encrypted with unique keys, a scalable disbursed database to perceive same files, a document-relocation machine that co-locates identical files at the equal machines, and a single-example store that contains equal documents while maintaining separate-report semantics.

Message-Locked Encryption furnished by M. Bellare and S. Keelveedhi to acquire cozy deduplication, goal currently targeted through numerous cloud storage companies. Message-Locke Encryption wherein the key under which encryption and decryption are carried out is itself derived from the message. It is used in a huge form of industrial and studies storage service structures. A patron first computes a key K H(M) by means of applying a cryptographic hash function H() to M, and then computes the ciphertext C E(K, M) thru a deterministic symmetric encryption scheme. A 2nd consumer encrypting the equal report M will produce the identical C, permitting deduplication.

3. FRAMEWORK

A.Overview of the Proposed System

To provide the secure data deduplication in the cloud environment, we are implementing a unique technique to realize an attribute-based totally storage device assisting relaxed deduplication. Our storage machine is built under a hybrid cloud architecture, where a non-public cloud manipulates the computation and a public cloud manages the storage. The personal cloud is provided with a trapdoor key associated with the corresponding ciphertext, with which it is able to switch the ciphertext over one get right of entry to policy into ciphertexts of the equal plaintext underneath some other get admission to rules without being aware of the underlying plaintext. After receiving a storage request, the non-public cloud first assessments the validity of the uploaded item via the connected proof.

B. Architecture of the Proposed Framework

The architecture of our attribute-based storage device with secure deduplication is shown in Fig. 1 wherein 4 entities are worried: information provider, attribute authority (AA), cloud and users.

Data Provider:

A data issuer desires to outsource his/her information to the cloud and percentage it with users possessing positive credentials.

Attribute Authority:

The AA issues each person a decryption key associated with his/her set of attributes.

Cloud:

The cloud includes a public cloud that is in rate of statistics storage and a personal cloud which performs sure computation along with tag checking. When sending a file storage request, each statistics issuer first off creates a tag T and a label L associated with the facts, after which encrypt the statistics underneath an get entry to shape over a fixed of attributes. Also, each information company generates a proof pf on the relationship of the tag T, the label L and the encrypted message ct3, but this evidence will no longer be stored everywhere inside the cloud and is best used throughout the checking phase for any newly generated storage request. After receiving a storage request, the private cloud first assesments the validity of the proof pf, and then assesments the equality of the new tag T with existing tags in the device. If there is no fit for this new tag T, the non-public cloud adds the tag T and the label L to a tag-label listing, and forwards the label and the encrypted records, (L, ct) to the general public cloud for storage.

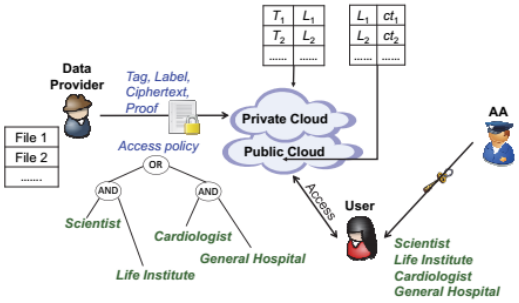


Fig1. System Architecture

User:

At the user aspect, every person can download an object, and decrypt the cipher textual content with the attribute-based non-public key generated by using the AA if this person's attribute set satisfies the get right of entry to shape. Each consumer assesments the correctness of the decrypted message the use of the label, and accepts the message if it's miles constant with the label.

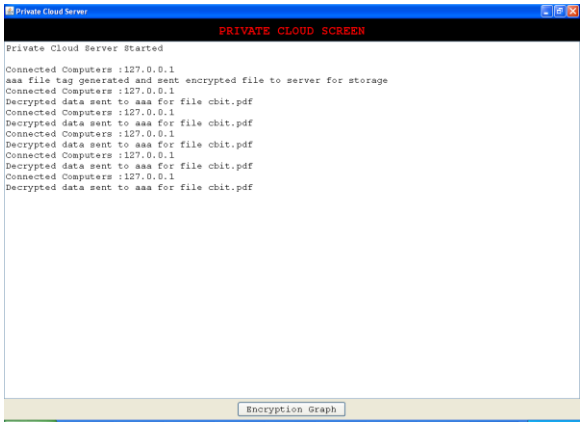
C. Workflow of the Proposed System

Our ciphertext-coverage attribute-based totally storage machine with securededuplication consists of the subsequent algorithms: setup set of rules Setup, characteristic-based totally personal key era algorithm KeyGen, encryption set of rules Encrypt, validity testing set of rules Validity-Test, equality checking out set of rules Equality-Test, re-encryption algorithm Re-encrypt and decryption algorithm Decrypt.

The proposed storagesystem enjoys important blessings. Firstly, it can be used to confidentially share records with different users via specifying access policies in preference to sharing the decryption key. Secondly, it achieves the usual notion of semantic security at the same time as current deduplication schemes only gain it beneath a weaker safety belief.

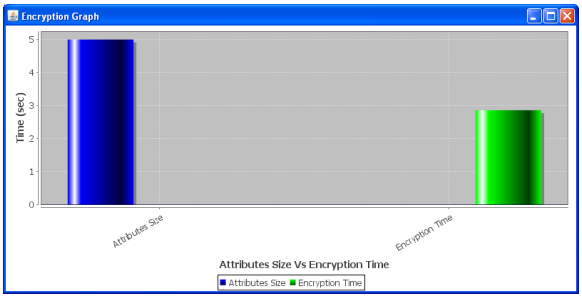
3. EXPERIMENTAL RESULTS

In this experiment, we have start the private cloud which is used to encrypt the files in the application and the cloud users need to register into the application and they need to login. If any user may upload any file on to the cloud as data owner and file which is uploaded by the data owner will be accessed by authorized user only who has access permission.



If any non access permission user may try to access the file, he won't get the file because of access permission.

Finally, we can view the encryption time as well as attribute size in the graph.



4. CONCLUSION

We conclude in this paper is that, we proposed a CP-ABE scheme to provide secure deduplication on encrypted cloud data in cloud environments. Here, we used two types of clouds such as private cloud and public cloud and this scenario called as hybrid cloud architecture. Here, we used private cloud for encryption keys computations and public cloud used for to store the encrypted data of the data owners.The non-public cloud is provided with a trapdoor key related to the corresponding ciphertext, with which it may transfer the ciphertext over one get right of entry to policy into ciphertexts of the same plaintext under some other get admission to regulations without being aware about the underlying plaintext. After receiving a storage request, the personal cloud first assessments the validity of the uploaded item via the attached proof.

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, 2002, pp. 617–624.

[2] 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. USENIX Association, 2013, pp. 179–194

[3] 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, CCS 2006, Alexandria, VA, USA, Ioctober 30 - November 3, 2006, ser. Lecture Notes in Computer Science, vol. 5126. Springer, 2006, pp. 89–98.

[4] 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, ser. Lecture Notes in Computer Science, vol. 9020. Springer, 2015, pp. 516–538.

[5] K. R. Choo, J. Domingo-Ferrer, and L. Zhang, "Cloud cryptography: Theory, practice and future research directions," Future Generation Comp. Syst., vol. 62, pp. 51–53, 2016.

[6] K. R. Choo, M. Herman, M. Iorga, and B. Martini, "Cloud forensics: State-of-the-art and future directions," Digital Investigation, vol. 18, pp. 77–78, 2016.

[7] 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.

[8] 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.

[9] 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.

[10] 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, CloudCom 2013, Bristol, United Kingdom, December 2-5, 2013, Volume 1. IEEE Computer Society, 2013, pp. 363–370.