International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

DEDUPLICATION OF ENCRYPTED DATA IN CLOUD

D VISWASAHITHYA[1], K VENKATA LAKSHMI[2], M KARTHIKA[3],
BALAJI SHANMUGA SRINIVAS[4], K GOVARDHAN[5]
[1]Assitant professor, [2,3,4,5] Students, Dept of CSIT
[1,2,3,4,5] Siddartha Institute of Science and Technology, Puttur

---

Abstract

Data is crucial for both individuals and organizations in the modern digital environment. Duplicate data contents cannot be kept since the amount of data generated is growing exponentially over time. As a result, using storage optimization strategies is a must for huge storage spaces like cloud storage. One such storage optimization method that prevents storing multiple copies of data is deduplication. Data is currently encrypted in order to ensure security and is stored in the cloud as well as other huge storage locations. This presents a challenge because deduplication cannot be applied to encrypted data. Thus, performing deduplication securely over the encrypted data in cloud appears to be a challenging task

Keywords: Deduplication, Cloud storage , Convergent Encryption

Introduction

Deep learning is widely employed in picture recognition, image processing, and particularly facial expression detection with the advent of the information age. In the area of human-computer interaction, face recognition has emerged as a research hotspot, yet there are still restrictions on how image processing results can be used. The data in the image is not used for secondary processing, which means that it has not been fully and effectively utilised in the actual production and life process, and image research frequently focuses on increasing the accuracy of recognition. In this study, a convolutional neural network expression recognition model is created and trained using a deep learning technique. When the outcomes of image processing are integrated with a music recommendation system, the music that best suits the person's mood is suggested. Major music websites' playlists and manually annotated tracks are crawled to build music data collections. The outputs of image processing have a suitably broader range of applications.

2. SYSTEM ANALYSIS

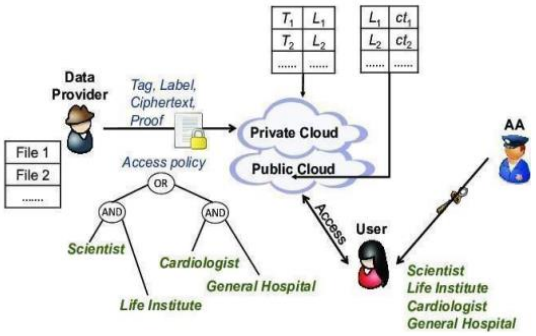
2.1 Proposed System

Cloud computing has made extensive use of attribute-based encryption (ABE), where a data provider outsources his or her encrypted data to a cloud service provider and can share the data with users who have certain credentials (or attributes). Secure deduplication, which is essential for removing duplicate copies of identical data in order to conserve storage space and network bandwidth, Cloud computing has made

extensive use of attribute-based encryption (ABE), where a data provider outsources his or her encrypted data to a cloud service provider and can share the data with users who have certain credentials (or attributes). Secure deduplication, which is essential for removing duplicate copies of identical data in order to conserve storage space and network bandwidth, is not supported by the basic ABE system. In this research, we offer an attribute-based storage system with safe deduplication in a hybrid cloud environment,where the storage is managed by a public cloud and duplicate detection is handled by a private cloud. Our approach has two benefits over earlier data deduplication solutions.

First off, rather than distributing decryption keys, it may be used to set access controls and exchange data with users in a private manner. Second, it defines a stronger security notion than existing systems, achieving the standard notion of semantic security for data confidentiality. Additionally, we present a method for converting ciphertexts that are

protected by one access policy into ciphertexts that are protected by another access policy without exposing the underlying plaintext.



2.3 System Design

2.3.1 System Architecture

A system architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. The architecture of our attribute-based storage system with secure deduplication is shown in Fig. 4.1 in which four entities are involved: data providers, attribute authority (AA), cloud and users

2.3.2 Input Design

A data provider wants to outsource his/her data to the cloud and share it with users possessing certain credentials. The AA issues every user a decryption key associated with his/her set of attributes. The cloud consists of a public cloud which is in charge of data storage and a private cloud which performs certain computation such as tag checking. When sending a file storage request, each data provider firstly creates a tag T and a label L associated with the data, and then encrypts the data under an access structure over a set of attributes. Also, each data provider generates a proof pf on the relationship of the tag T, the label L and the encrypted message ct3, but this proof will not be stored anywhere in the cloud and is only used during the checking phase for any newly generated storage request.

The objectives of input design are

- To design data entry and input procesures
- To reduce input volume
- To design source documents for data capture or devise other data capture methods
- To design input data records, data entry screens, user interface screens, etc..,
- To use validation checks and develop effective input controls

2.3.3 Output Design

At the user side, each user can download an item, and decrypt the ciphertext with the attribute-based private key generated by the AA if this user's attribute set satisfies the access structure. Each user checks the correctness of the decrypted message using the label, and accepts the message if it is consistent with the label. Objectives of Output Design: The objectives of output design are:

• To design a smart agent that has contextual information about the user and helps in managing and planning tasks.

• Virtual assistant helps you to save time and focus attention on what matters most

• To deliver the appropriate quantity of output.

• Creating input data records, data entry screens, user interface screens, etc. to provide the right amount of output.

• To provide the results on time so that they can be used to make wise judgements.

2.4 Modules

Scientist: An individual who does scientific study to increase knowledge in a particular field is considered to as a scientist.

Life Institute: The Future of Life Institute is a nonprofit organisation that strives to lessen the existential and global cataclysmic dangers that humanity faces, notably the existential risk posed by highly developed artificial intelligence . **Cardiologist:** A cardiologist is a medical professional who focuses on treating conditions that affect the cardiovascular system, primarily the heart and blood vessels. A doctor must complete four years of medical school as well as an extra six to eight years of internal medicine and specialist cardiology training in order to become a cardiologist.

General hospital: A general hospital is a hospital that Hospital does not specialize in the treatment of particular illnesses or patients.

2.5 Framework

The following algorithms make up our safe deduplication, attribute-based, ciphertext-policy storage system: the setup process Algorithm for configuring an attribute-based algorithm for generating encryption keys KeyGen Algorithmic validity testing and encryption Valid-testing and equalitytesting algorithms Algorithms for re-encryption and decryption, as well as equality checking Decrypt. • Setup $(1\lambda) \rightarrow$ (pars, msk). This setup algorithm outputs the master private key msk and the public parameter pars for the system after taking the security parameter as an input. The AA is in charge of this algorithm.

2.6 Implementation

We implemented our storage system's algorithms inside Charm [35]7, a framework designed to speed up the development of cryptographic protocols and schemes. All Charm routines are created in asymmetric groups, thus before implementation, our construction is changed to an asymmetric configuration. In other words, there are three groups employed, G, G, and G1, and the pairing e is a function from G, G, and G1. Observe that it was claimed in [22] that the security proofs and assumptions can be generalised to the asymmetric context. In our implementation, we make use of Python 3.4 and Charm 0.43. For the underlying cryptographic operations, we install the PBC library together with Charm-0.43. Our tests are conducted on a laptop

running 64-bit Ubuntu 16.04 on an Intel Core i5-4210U Processor clocked at 1.70GHz and 4.00 GB of RAM. Over four different elliptic curves—SS512, MNT159, MNT201, and MNT224—we simulate the proposed attribute-based storage system with secure deduplication. The pairings on the other three curves are asymmetric Type 3 pairings, while SS512 is a super singular elliptic curve with the symmetric Type 1 pairing on it. These four curves offer, respectively, security levels of 80 bits, 100 bits, and 112 bits. The suggested attribute-based storage system allowing secure deduplication is depicted in Fig. 5 in terms of the computing complexity of four algorithms: key generation algorithm KeyGen (Fig. 5-(a)), encryption algorithm Encrypt (Fig. 5-(b)), and reencryption algorithm. Decryption algorithm and re-encryption (Fig. 5-(c)). The most effective curve is SS512, whereas MNT224 has the highest computational cost of all the curves.

2.6.1 Deduplication in hybrid cloud

The inability of the current methods to achieve safe deduplication (e.g., [8], [23]) to satisfy the normative security definition for secrecy, such as semantic security, is one of their intrinsic drawbacks (See Section 3.3 for the reason). A less-secure security concept known as privacy under chosen-distribution attacks [8] was proposed as a solution to this issue under the presumption that the input message is sufficiently unpredictable. Our storage system introduces a hybrid cloud architecture, which consists of a pair of public and private clouds, in contrast to the 5 traditional way of establishing a weaker security idea for the cloud storage system with safe deduplication.

2.6.2 Use case Diagram

A collection of use cases, actors, and their relationships are shown in use case diagrams. They represent a system's use case perspective. A use case illustrates a certain system capability. The links between the functionalities and their internal/external controllers are therefore described using a use case diagram. Actors are the name given to these controllers. In the Unified Modeling Language (UML), a use case diagram is a specific kind of behavioural diagram that results from and is defined by usecasestudy. Its objective is to offer a graphical picture of a system's functionality in terms of actors, their objectives (expressed as use cases), and any relationships among those use cases. A use case diagram's primary objective is to identify which system functions are carried out for which actor. The system's actors can be represented by their roles.

Related Work

Encryption based on attributes. Attribute-based encryption (ABE) was first described by Sahai and Waters [6], and then key-policy ABE (KPABE) and ciphertext-policy ABE (CP-ABE) were developed by Goyal et al. [16] as two variants that complement each other. The first KP-ABE system supporting the expression of non-monotone formulas was presented in [17] to enable more workable access policies, the first large class KP-ABE system was presented by in the standard model in [18], and the first KP-ABE construction was given in [16] that realised the monotonic access structures. However, since the access policy is set once the user's attribute private key is issued, we think KP-ABE is less adaptable than CP-ABE.

Secure Deduplication With the intention of conserving storage space for cloud storage services, Douceur et al. [23] introduced the first method for achieving convergent encryption, which encrypts a message using a messagederived key to ensure that identical plaintexts are encrypted to the same ciphertexts. If two users submit the same file in this scenario, the cloud server may identify the identical ciphertexts and keep just one instance of them. Convergent encryption implementations and variations have been used in [24], [25], [26], [27], and [28]. Bellare, Keelveedhi, and Ristenpart [8] proposed a cryptographic primitive called messagelocked encryption and provided numerous definitions that adequately capture various security needs in order to codify the precise security definition for convergent encryption.

5. Conclusion:

Deduplication is a technique for reducing bandwidth and storage capacity available in cloud storage. But, deduplication is less doable with translated data since, different crucial encryptions convert same data into different formats. In this project colorful methods are bandied where deduplication methods are carried out on translated data in a large storage area. utmost of the methods studied then work on the basis of coincident encryption, which is a simple approach that makes deduplication compatible with translated data. In this information thick world, we cannot compromise on both security and duplication of data across storage areas. A strategy needs to be formulated which will enhance storage optimization without negotiating on encryption method; by furnishing deduplication technique in data storage servers where the available data is translated.

6. References:

[1] D. Quick, B. Martini, and K. R. Choo, Cloud Storage Forensics. Syngress Publishing / Elsevier, 2014.[Online].Available: http://www.elsevier.com/books/cloud-storage forensics/quick/978-0-12-419970-5

[2] 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.

[3] 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

[4] Y. Yang, H. Zhu, H. Lu, J. Weng, Y. Zhang, and K. R. Choo, “Cloud based data sharing with finegrained proxy re-encryption,” Pervasive and Mobile Computing, vol. 28, pp. 122–134, 2016.

[5] D. Quick and K. R. Choo, “Google drive: Forensic analysis of data remnants,” J. Network and Computer Applications, vol. 40, pp. 179– 193, 2014

[6] A. Sahai and B. Waters, “Fuzzy identity-based encryption,” in Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings, ser. Lecture Notes in Computer Science, vol. 3494. Springer, 2005, pp. 457–473.

[7] 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.

[8] 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.

[9] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, "Message-locked encryption for lockdependent messages," in Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I, ser. Lecture Notes in Computer Science, vol. 8042. Springer, 2013, pp. 374–391. [10] 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 7 [11] 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. [12] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin ¨ clouds: Secure cloud computing with low latency - (full version)," in Communications and Multimedia Security, 12th IFIP TC 6 / TC 11 International Conference, CMS 2011, Ghent, Belgium, October 19- 21,2011. Proceedings, ser. Lecture Notes in Computer Science, vol. 7025. Springer, 2011, pp. 32–44.