Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud

Maria Roofina

Abstract— **Attribute-based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud service provider, and can share the data with users possessing specific credentials (or attributes). However, the standard ABE system does not support secure deduplication, which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth. In this paper, we present an attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage. Compared with the prior data deduplication systems, our system has two advantages. Firstly, it can be used to confidentially share data with users by specifying access policies rather than sharing decryption keys. Secondly, it achieves the standard notion of semantic security for data confidentiality while existing systems only achieve it by defining a weaker security notion. In addition, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under other access policies without revealing the underlying plaintext.**

Keywords— ABE, Storage, Deduplication.

I. INTRODUCTION

This Cloud computing greatly facilitates data providers who want to outsource their data to the cloud without disclosing their sensitive data to external parties and would like users with certain credentials to be able to access the data [1], [2], [3], [4], [5]. This requires data to be stored in encrypted forms with access control policies such that no one except users with attributes (or credentials) of specific forms can decrypt the encrypted data. An encryption technique that meets this requirement is called attribute-based encryption (ABE) [6], where a user's private key is associated with an attribute set, a message is encrypted under an access policy (or access structure) over a set of attributes, and a user can decrypt a ciphertext with his/her private key if his/her set of attributes satisfies the access policy associated with this ciphertext. However, the standard ABE system fails to achieve secure deduplication [7], which is a technique to save storage space and network bandwidth by eliminating redundant copies of the encrypted data stored in the cloud. On the other hand, to the best of our knowledge, existing constructions [8], [9], [10], [11] for secure deduplication are not built on attribute-based encryption. Nevertheless, since ABE and secure deduplication have been widely applied in cloud computing, it would be desirable to design a cloud storage system possessing both properties.

We consider the following scenario in the design of an attribute-based storage system supporting secure deduplication of encrypted data in the cloud, in which the cloud will not store a file more than once even though it may receive multiple copies of the same file encrypted under different access policies. A data provider, Bob, intends to upload a fileM to the cloud, and shareM with users having certain credentials. In order to do so, Bob encrypts M under an access policy A over a set of attributes, and uploads the corresponding ciphertext to the cloud, such that only users whose sets of attributes satisfying the access policy can decrypt the ciphertext. Later, another data provider, Alice, uploads a ciphertext for the same underlying file M but ascribed to a different access policy A0. Since the file is uploaded in an encrypted form, the cloud is not able to discern that the plaintext corresponding to Alice's ciphertext is the same as that corresponding to Bob's, and will store M twice. Obviously, such duplicated storage wastes storage space and communication bandwidth.

A. Our Contributions

In this paper, we present an attribute-based storage system which employs ciphertext-policy attribute-based encryption (CP-ABE) and supports secure deduplication. Our main contributions can be summarized as follows.

- Firstly, the system is the first that achieves the standard notion of semantic security for data confidentiality in attribute-based deduplication systems by resorting to the hybrid cloud architecture [12].
- Secondly, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under any other access policies without revealing the underlying plaintext. This technique might be of independent interest in addition to the application in the proposed storage system.
- Thirdly, we propose an approach based on two cryptographic primitives, including a zero-knowledge proof of knowledge [13] and a commitment scheme

B. Related Work

The template Attribute-Based Encryption. Sahai and Waters [6] introduced the notion of attribute-based encryption (ABE), and then Goyal et al. [16] formulated key-policy ABE (KP-ABE) and ciphertext-policy ABE (CP-ABE) as two complimentary forms of ABE. The first KP-ABE construction given in [16] realized the monotonic access structures, the first KP-ABE system supporting the expression of non-monotone formulas was presented in [17] to enable more viable access policies, and the first large class KP-ABE system was presented by in the standard model in [18]. Nevertheless, we believe that KP-ABE is less flexible than CP-ABE because the access policy is determined once the user's attribute private key is issued. Bethencourt, Sahai and Waters [19] proposed the first CP-ABE construction, but it is secure under the generic group model. Cheung and Newport [20] presented a CPABE scheme that is proved to be secure under the standard model, but it only supports the AND access structures. A CP-ABE system under more advanced access structures is proposed by Goyal et al. [21] based on the number theoretic assumption. In order to overcome the limitation that the size of the attribute space is polynomially bounded in the security parameter and the attributes are fixed ahead, Rouselakis andWaters [22] built a large universe CP-ABE system under the prime-order group. In this paper, the Rouselakis-Waters system is taken as the underlying scheme for the concrete construction.

Secure Deduplication.With the goal of saving storage space for cloud storage services, Douceur et al. [23] proposed the first solution for balancing confidentiality and efficiency in performing deduplication called convergent encryption, where a message is encrypted under a message-derived key so that identical plaintexts are encrypted to the same ciphertexts. In this case, if two users upload the same file, the cloud server can discern the equal ciphertexts and store only one copy of them. Implementations and variants of convergent encryption were deployed in [24], [25], [26], [27], [28]. In order to formalize the precise security definition for convergent encryption, Bellare, Keelveedhi and Ristenpart [8] introduced a cryptographic primitive named message locked encryption, and detailed several definitions to capture various security requirements. Abadi et al. [9] then strengthened the security definition in [8] by considering the plaintext distributions depending on the public parameters of the schemes. This model was later extended by Bellare and Keelveedhi [11] by providing privacy for messages that are both correlated and dependent on the public system parameters. Since message-locked encryption cannot resist to brute-force attacks where files falling into a known set will be recovered, an architecture that provides secure deduplicated storage resisting brute-force attacks was put forward by Keelveedhi, Bellare and Ristenpart [10] and realized in a system called server-aided encryption for deduplicated storage. In this paper, a similar technique to that [9] is used to achieve secure deduplication with regard to the private cloud in the concrete construction.

C. Security Definitionss

Traditionally, an encryption system is required to provide privacy of the encrypted data, which is captured by ndistinguishability under either chosen plaintext attacks (INDCPA) or chosen ciphertext attacks (IND-CCA). However, neither IND-CPA nor IND-CCA is feasible in an encrypted storage system with secure deduplication, since it can be easily broken by an adversary in either IND-CPA or INDCCA security game as follows. An adversary, given a challenge CT——————————————— for a plaintext mb with b 2 f0; 1g where m0, m1 are chosen by the adversary, can output the correct b by creating a tag T for mb and running the equality testing algorithm to see whether T matches the tag T of CT Noticeably, it is impossible to design an encryption scheme with an equality-checking tag to satisfy the standard notions of confidentiality [9]. Thus, we alternatively aim to achieve IND-CPA security at the public cloud side, whilst preserving a security notion called PRV-CDA security (privacy under chosen distribution attacks) [8] at the private cloud side under the assumption that the message space M() is sufficiently large such that the plaintexts in the system are npredictable (i.e., given the public parameter and encryption of a randomly selected plaintext in the message space M(), it is infeasible for any polynominal time algorithm A to obtain the plaintext). IND-CPA Security. Denote our attribute-based storage system with secure deduplication . The definition of selective IND-CPA security with respect to the public cloud in is given in Fig. 3, where we restrain algorithm A to issuing queries to the key generation oracle on attribute sets satisfying the access structures A0 and A1. An attribute-based storage system with secure deduplication is IND-CPA secure if the advantage function referring to the security game GameIND

;A

AdvIND

;A()

def

= Pr[b0 = b]

is negligible in the security parameter for any probabilistic polynomial-time (PPT) adversary algorithm A. PRV-CDA Security. Based on the definition of PRV-CDA given in [8], the definition of PRV-CDA for is shown in Fig. , where the adversary is given an additional trapdoor key for he challenge ciphertext but is not given access to any attribute-based private keys (as the private cloud is not allowed to collude with users). An attribute-based storage system with secure deduplication is PRV-CDA secure if the advantage function referring to the security game GamePRV

;A

AdvPRV-CDA

;A ()

def

= Pr[b0 = b]

is negligible in the security parameter for any PPT adversary algorithm A. With regard to a storage system, it is crucial to ensure consistency [9] to resist duplicate faking attacks such that a legitimate message will not be unnoticeably replaced by a fake one. Consistency in our attribute-based storage system with secure deduplication can be divided into ciphertext consistency, tag and label consistency. Ciphertext consistency guarantees that given a ciphertext outsourced by an honest data provider, an adversary who has no idea about the encrypted data can not generate another valid ciphertext with the same tag but under a different plaintext to cheat the private cloud. Tag/Label consistency ensures consistency of the data used in the tag/label derivation and the ciphertext generation such that an adversary is not able to create a tag/label that does not match the underlying data to cheat a user having access to the encrypted data. Consistency. Ciphertext consistency for our attributebased storage system with secure deduplication is given in Fig. 4, in which given a ciphertext (T, L, ct, pf) and the public parameter, an adversary wins the game if it outputsanother ciphertext (T, L0, ct0, pf 0) such that pf 0 is valid for (T, L0, ct0). This game prevents an adversary from capturing an outsourcing request from an honest data provider and replacing the corresponding ciphertext to another ciphertext without being detected by the private cloud. Taking the definition for consistency in [9] into consideration, we depict the security game for tag/label consistency for our system in Fig. 4, which provides security against duplicate faking attacks where a legitimate message is replaced by a fake one without being discovered. Specifically, assume that an adversary creates and uploads a ciphertext ct0 of M0 associated with a tag and label pair for M. Later, an honest data provider, holding M computes and uploads thefine abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

D. Conclusions

Comp Attribute-based encryption (ABE) has been widely used in cloud computing where data providers outsource their encrypted data to the cloud and can share the data with users possessing specified credentials. On the other hand, deduplication is an important technique to save the storage space and network bandwidth, which eliminates duplicate copies of identical data. However, the standard ABE systems do not support secure deduplication, which makes them costly to be applied in some commercial storage services. In this paper, we presented a novel approach to realize an attribute-based storage system supporting secure deduplication. Our storage system is built under a hybrid cloud architecture, where a private cloud manipulates the computation and a public cloud manages the storage. The private cloud is provided with a trapdoor key associated with the corresponding ciphertext, with which it can transfer the ciphertext over one access policy into ciphertexts of the same plaintext under any other access policies without being aware of the underlying plaintext. After receiving a storage request, the private cloud first checks the validity of the uploaded item through the attached proof. If the proof is valid, the private cloud runs a tag matching algorithm to see whether the same data underlying the ciphertext has been stored. If so, whenever it is necessary, it regenerates the ciphertext into a ciphertext of the same plaintext over an access policy which is the union set of both access policies. The proposed storage system enjoys two major advantages.

Firstly, it can be used to confidentially share data with other users by specifying an access policy rather than sharing the decryption key.

Secondly, it achieves the standard notion of semantic security while existing deduplication schemes only achieve it under a weaker security notion.

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- storageforensics/ 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 fine-grained 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 lock-dependent 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.

[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. N¨ urnberger, 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.

[13]. S. Goldwasser, S. Micali, and C. Rackoff, “The knowledge complexity of interactive proof-systems (extended abstract),” in Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 6-8, 1985, Providence, Rhode Island, USA. ACM, 1985, pp. 291– 304.

[14]. M. Fischlin and R. Fischlin, “Efficient non-malleable commitment schemes,” in Advances in Cryptology - CRYPTO 2000, 20th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2000, Proceedings, ser. Lecture Notes in Computer Science, vol. 1880. Springer, 2000, pp. 413–431.

[15]. S. Goldwasser and S. Micali, “Probabilistic encryption,” J. Comput. Syst. Sci., vol. 28, no. 2, pp. 270–299, 1984.

[16]. 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.

[17]. R. Ostrovsky, A. Sahai, and B.Waters, “Attribute-based encryption with non-monotonic access structures,” in Proceedings of the 2007

ACM Conference on Computer and Communications Security, CCS 2007, Alexandria, Virginia, USA, October 28-31, 2007. ACM, 2007, pp. 195–203.

[18]. A. B. Lewko and B. Waters, “Unbounded HIBE and attributebased encryption,” in Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 201 1. Proceedings, ser. Lecture Notes in Computer Science, vol. 6632. Springer, 2011, pp. 547–567.

[19]. J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute-based encryption,” in 2007 IEEE Symposium on Security and Privacy (S&P 2007), 20-23 May 2007, Oakland, California, USA. IEEE Computer Society, 2007, pp. 321–334.

[20]. L. Cheung and C. C. Newport, “Provably secure ciphertext policy ABE,” in Proceedings of the 2007 ACM Conference on Computer and Communications Security, CCS 2007, Alexandria, Virginia, USA, October 28-31, 2007. ACM, 2007, pp. 456–465.

[21]. V. Goyal, A. Jain, O. Pandey, and A. Sahai, “Bounded ciphertext policy attribute based encryption,” in Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik,