ISSN 2321-743X
International Journal of Research in
Engineering and Bioscience
Volume 6 Issue 3 (Pages 35- 42)
IJREB
Journal home page: www.ijreb.org

ATTRIBUTE BASED DE-DUPLICATION OF ENCRYPTED DATA IN CLOUD

Muthu Kumar P[1], Hursley J[1], Mohamed Shajoon[1], Venifa Mini G[2]

[1]UG Student, [2]Assistant Professor, Department of Computer Science & Engineering, Noorul Islam Centre for Higher Education, Kumaracoil- 629 180, India.

ABSTRACT

ABE is a type of public-key encryption in which the secret key of a user and the cipher text are dependent upon attributes. In this project secure data de-duplication is deployed which is a technique for eliminating duplicate copies of data, in cloud storage to reduce storage space. To achieve the goal a hybrid cloud which is an integrated cloud service utilizing both private and public clouds to perform distinct functions within the same organization is used. Data confidentiality is provided in this project by using an authority and data authentication process. Cipher-text is the method used for encrypting or encoding the information because it contains a form of the original plaintext that is unreadable by a human or computer without the proper cipher to decrypt it.

KEYWORDS: Attribute based encryption, Secure de-duplication, Hybrid cloud, Data Confidentiality, Cipher Text

INTRODUCTION

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. 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), 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, 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 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.

RELATED WORK

A. *Review On Data De-duplication In Cloud Computing*

Data deduplication is a technique to improve the storage utilization. De-duplication technologies can be designed to work on primary storage as well as on secondary storage.[1,4] Data deduplication technique allows the cloud users to manage their cloud storage space effectively by avoiding storage of repeated data's and save bandwidth.[1,5,6] The data are finally stored in cloud server. To ensure data confidentiality the data are stored in an encrypted type using DES algorithm.[1,2,3]

B. *A survey on Attribute Based Encryption Scheme in Cloud Computing*

A secure user enforced data access control mechanism must be provided before cloud users have the liberty to outsource sensitive data to the cloud for storage with the emergence of sharing confidential corporate data on cloud servers.[2,3] Attribute-based encryption is a public key based encryption that enables access control over encrypted data using access policies and ascribed attributes.[2,3,4] In this paper, they have discussed Attribute based encryption (ABE),KP-ABE, CP-ABE.

C. *Data Deduplication on Secured Cloud Storage*

The cloud storage server performs data compression i.e. data deduplication which eliminates the duplicate data and also saves the storage space[3,5]. Also, cloud storage service poses challenges with respect to privacy and confidentiality of the data stored by the different users[4]. We proposed a Data deduplication scheme, which can be used to eliminate the duplicate copies of data on cloud as well as preserves privacy and confidentiality of the data[6,7].

D. *Time Based Proxy Re-encryption Scheme for Secure Data Sharing in a Cloud Environment*

The CSP mainly provides two services: data storage and re-encryption. After obtaining the encrypted data from the data owner, the CSP will store the data on several cloud servers, which can be chosen by the consistent hash function, where the input of the consistent hash function is the key of the data, and the outputs of the consistent hash function are the IDs of the servers that store the data [4,8]. On receiving a data access request from a user, the CSP will re-encrypt the cipher text based on its own time, and return the re-encrypted cipher text [4].

SYSTEM ARCHITECTURE

The proposed architecture consists of four entities, Data providers, Attribute authority, Cloud and Users. A data provider wants to outsources his data to the cloud and share it with users possessing certain credentials. The AA issues every user a

decryption key associated with his set of attributes. The cloud consists of public cloud which in charge of data storage and private cloud which performs certain computation such as tag checking.

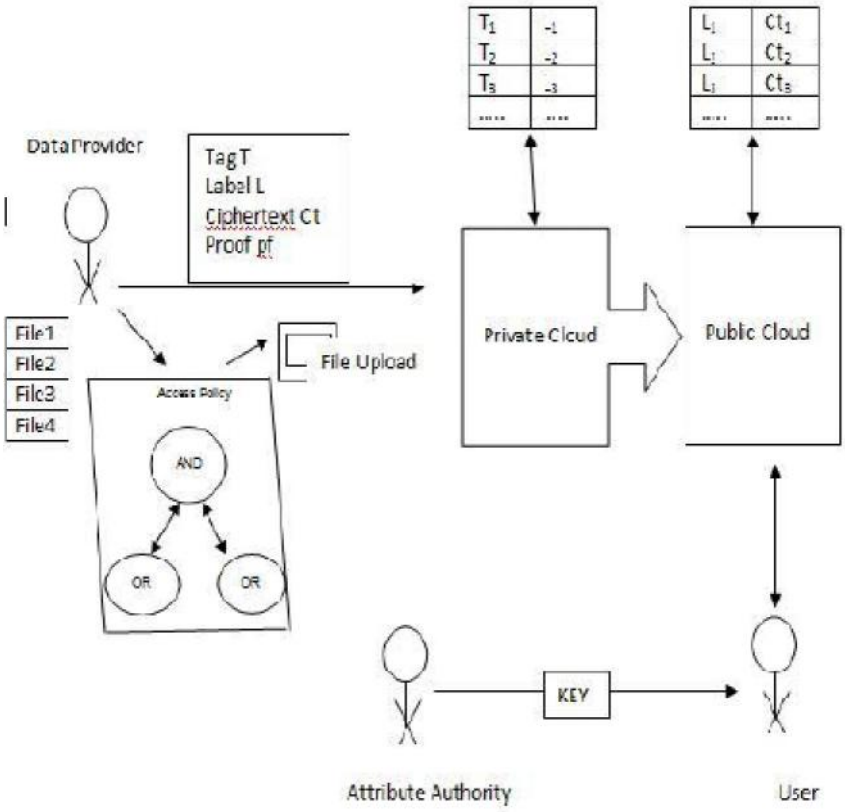


Fig.1: Architectural Diagram

1. Data Provider

Data provider has to register with the attributes. Then Login with the username and password. He/She has the capability to upload the files with the public parameter(pars). Before uploading the file into the cloud, Owner have to encrypt the data for security purposes. Here AES Encryption is used for Encrypting the data. Finally Provider will sent the request for storage to private cloud.

2. USER

User needs to register with his/her attributes. After registration, user have to wait for activation. After activated by AA, then login with username and password. View all the files and decrypt the file. But user should have the same attribute as owner, then only he can download the files.

3. PUBLIC CLOUD

Public cloud firstly login by giving the username and password, all the non duplicate files will store in public cloud.

4. PRIVATE CLOUD

Private cloud firstly login by giving the username and password, then validity test will done on the uploaded files and then Equality

test will be applied to the file for checking the duplication. If it is a duplicate file, public cloud will not allow the file to store in public cloud and eliminates all the duplicate files.

5. ATTRIBUTE AUTHORITY

AA should login by giving the username and password, AA has the authority to activate both the Owner and user. It will generate 3 keys like public parameter, Master key and private key.

IMPLEMENTATION

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. After receiving a storage request, the private cloud first checks the validity of the proof pf, and then tests the equality of the new tag T with existing tags in the system. If there is no match for this new tag T, the private cloud adds the tag T and the label L to a tag-label list, and forwards the label and the encrypted data, (L, ct) to the public cloud for storage. Otherwise, let ct` be the ciphertext whose tag matches the new tag and L` be the label associated with ct`, and then the private cloud executes as follows.

1. If the access policy in ct is a subset of that in ct`, the private cloud simply discards the new storage request; else, if the access policy in ct` is a subset of that in ct, the private cloud asks the public cloud to replace the stored pair (L`, ct`) with the new pair (L, ct) where L = L`.
2. If the access policies in ct and ct` are not mutually contained, the private cloud runs the ciphertext regeneration algorithm to yield a new ciphertext for the same underlying plaintext file and associated with an access structure which is the union of the two access 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.

ALGORITHMS AND TESTING

A. Bilinear Pairings and Complexity Assumptions

Let G_1, G_2 be two additive cyclic groups of prime order q, and G_T another cyclic group of order q written multiplicatively. A pairing is a map: $G_1 * G_2 \quad G_T$, which satisfies the following properties:

$$\forall a, b \in F_q^*, \forall P \in G_1, Q \in G_2 : e(aP, bQ) = e(P, Q)^{ab}$$

B. Symmetric Encryption

A symmetric encryption (SE) scheme SE with a key space K and a message space M [30] is composed of two algorithms: an encryption algorithm SE. Enc(K, m) which outputs a ciphertext CT on input a key K 2

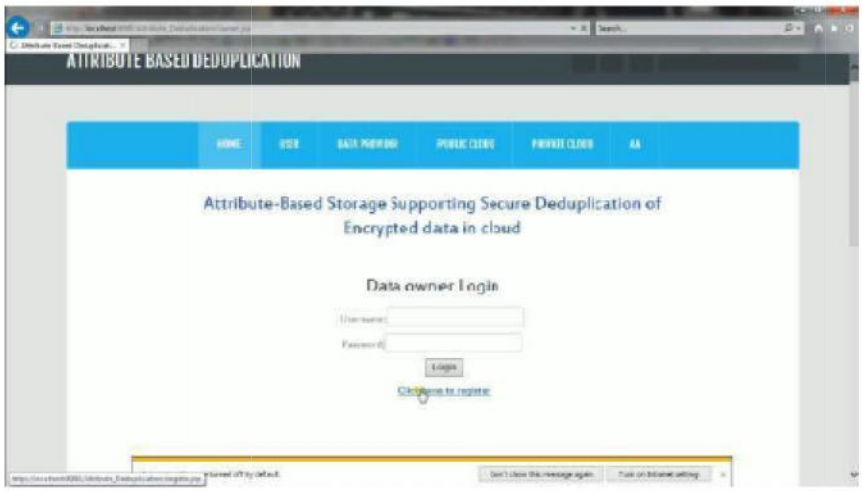
K and a message m € M, and a decryption

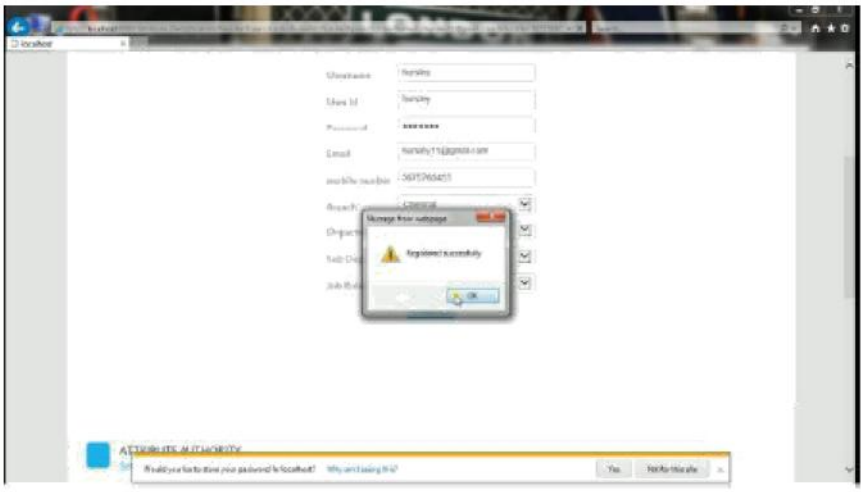
algorithm SE. Dec(K, CT) which outputs a message m or a failure symbol ‡ on input a key K € K and a ciphertext CT.

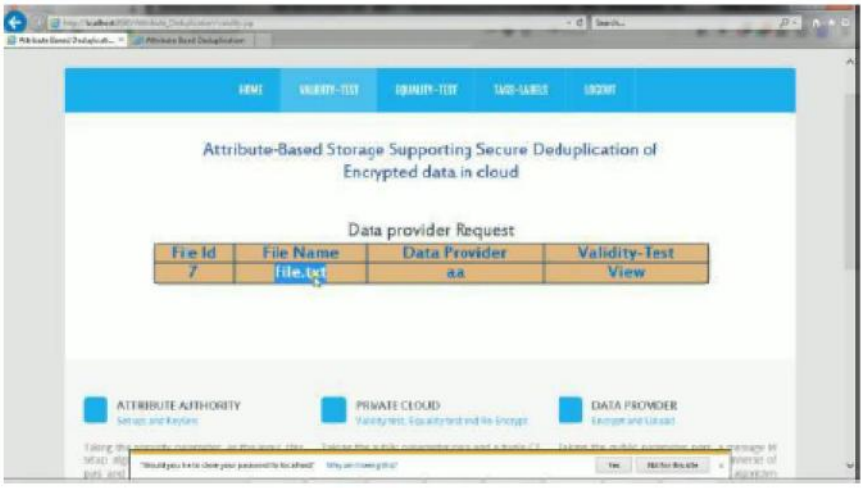
$$\mathbf{Adv}^{\text{IND-CPA}}_{\mathcal{SE},\mathcal{A}}(\lambda) = \Pr\left[b' = b \;\middle|\; \begin{array}{l} K \leftarrow \mathcal{K}; b \leftarrow \{0,1\} \\ (m_0, m_1, st) \leftarrow \mathcal{A}_1(1^\lambda) \\ \mathrm{CT}^* \leftarrow \mathcal{SE}.\mathrm{Enc}(K, m_b) \\ b' \leftarrow \mathcal{A}_2(par, m_0, m_1, st, \mathrm{CT}^*) \end{array} \right] - 1/2$$

Let st be the state information. A symmetric encryption scheme SE is secure under chosen plaintext attacks (IND-CPA secure), if for any PPT adversary A = (A1, A2), the advantage function is negligible in the security parameter ‡, where |m0| = |m1|.

RESULTS AND DISCUSSION







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-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 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 onference 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 CryptographicTechniques, Tallinn, Estonia, May 15-19, 2011. 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.