Secure Deduplication of Data ABE with Hybrid Cloud

Safa Ashraf[1] Bhavya K Bharathan[2]
[2]Assistant Professor
[1,2]Department of Computer Engineering
[1,2]Cochin College of Engineering and Technology, India

Abstract— The technique Attribute-based encryption has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud, and can share the data with users who satisfies the possessing credentials. Unfortunately the standard ABE system does not support secure deduplication, which is important for eliminating duplicate copies of identical data in order to save cloud storage and the network bandwidth. The term deduplication of data in cloud refers to the elimination of duplicate or redundant data or file especially from the cloud. The paper is a variant of the paper proposed by Hui Cui and Robert Hdeng that focuses on the concept of ABE. Though method is secure for data and save storage space in the cloud, it is not efficient and effective in its performance. In the proposed system, encryption is performed using the combination of AES and CPABE. It enhances the performance of secure storage using hybrid cloud.

Keywords: ABE, Deduplication, CP-ABE, AES

I. INTRODUCTION

Security is one of the most crucial aspects among those prohibiting the wide-spread adoption of cloud computing [21]. Cloud security issues may stem due to the core technology's implementation, cloud service offerings and arising from cloud characteristics. For a cloud to be secure, all of the participating entities must be secure. In any given system with multiple units, the highest level of the system's security is equal to the security level of the weakest entity. Data security plays a prominent role in cloud computing where confidentiality, integrity and availability are considered as its key objectives. The data outsourced to a public cloud must be secured. Unauthorized data access by other users and processes must be prevented [21]. The most common technique for achieving data security is cryptography. Cloud computing help the data providers to outsource their data to the cloud without disclosing their original data to external parties and users with certain property can access the data. This requires data to be stored in encrypted forms with access control policies such that no one except users with attributes. The method used here is ABE where user's private key is associated with an attribute set, and the message is encrypted over a set of attributes, and a user can decrypt a cipher text with their private key if their set of attributes satisfies the access policy associated with this cipher text. But 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 [2].In a deduplication system cloud will not be allow to store a file more than once even though it may receive multiple copies of the same file encrypted under different access policies. Hence it prevents wastage of storage space and communication bandwidth.

[Figure-1] is an example of the data on which deduplication is performed. When deduplication applies, it automatically deletes all the replicated data as shown in figure 1.

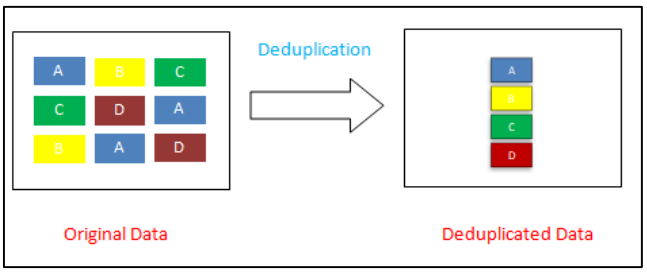

Fig. 1: Deduplication

II. RELATED WORKS

Attribute-Based Encryption: The concept of Fuzzy Identity-Based Encryption (FIBE) introduced by Sahai and Waters in [1] is a just a notion of identity-based cryptosystems where users can access secret data based on their attributes. Its drawback is construction limited to fixed threshold defined in the system. Later XiMing Li, Ming Wu Zhang [2] overcome this problem by constructing fuzzy identity based encryption with dynamic threshold. In addition Goyal et al [3] proposed key policy based encryption where cryptosystem combines the secret key and access structure. The idea of cipher text-ABE construction is explained by Bethencourt, Sahai and Waters [4] is secure only under the generic group model. The concept of Broadcast encryption introduced by A. Fiat and M. Naor [5] is an encryption technique which allows a sender to send a message to a dynamically chosen subset of users such that users in that subset can only decrypt it. A. B. Lewko and B. Waters [6] explained key policy ABE in the system proposed where user's private key is issued as soon as the access policy is determined. Later Chang and Newport [7] proposed cipher text policy system which is much flexible than KP ABE but that was secure under standard models only. M. Abadi, D. Boneh [8] proposed an encryption scheme with equality checking tag to meet standard notion of security. J. Lai, R. H. Deng, Y. Yang, and J. Weng [9] introduced adaptable CP-ABE with semi-trusted prox. The proxy introduced into the CP-ABE, given the system a trapdoor key to transform any cipher text under one access policy into cipher texts of the same plaintext under any other access policies without learning any information about the plaintext during the process of transformation. M. Pirretti [10] proposed an attribute based system for securely managing information in large and distributed system. Melissa Chase [18] proposed system with Multi-Authority Attribute Based Encryption.

AES Algorithm: It is a symmetric block cipher, a symmetric encryption algorithm which is found to be six

times faster than triple DES. AES is an iterative rather than Feistel cipher. It is based on substitution–permutation network. It comprises of a series of connected operations, some of which involve replacing inputs with specific outputs (substitutions) where as others involve shuffling bits around (permutations).AES performs all its computations on bytes rather than bits. Unlike DES, the number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. Each of these rounds uses a different 128-bit round key, which is calculated from the original AES key. Symmetric (also known as secret-key) ciphers use the same key for encrypting and decrypting, so the sender and the receiver must both know -- and use -- the same secret key. All key lengths are deemed sufficient to protect classified information up to the "Secret" level with "Top Secret" information requiring either 192- or 256-bit key lengths. There are 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys - a round consists of several processing steps that include substitution, transposition and mixing of the input plaintext and transform it into the final output of cipher text. The AES encryption algorithm defines a number of transformations that are to be performed on data stored in an array. The first step of the cipher is to put the data into an array; after which the cipher transformations are repeated over a number of encryption rounds. The number of rounds is determined by the key length, with 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. The first transformation in the AES encryption cipher is substitution of data using a substitution table; the second transformation shifts data rows, the third mixes columns. The last transformation is a simple exclusive or (XOR) operation performed on each column using a different part of the encryption key - longer keys need more rounds to complete.

Secure Deduplication: B. Zhu, K. Li, and R. H. Patterson [11] proposed a standard ABE system with no deduplication. J. R. Douceur [12] proposed solution to remove duplication called convergent encryption, where a data is encrypted under a data-derived key so that identical plaintexts are encrypted to the same cipher texts. Sukanya Gunjal [13] proposed deduplication with access control. Later R Shobana [14] introduced a method for data deduplication using AES. Pasquale et al [15] proposed system with combined advantages of deduplication and convergent encryption. N.Vaishnavi Moorthy [16] proposed system with automatic deduplication process with high confidentiality and security of the data. Jan Stanek [17] introduced a secure deduplication scheme for encrypted data that has dynamic ownership management capability. Sandeep Kaur [19] proposed a method which uses hashing algorithms to remove duplication. R. Anitha Rani [20] proposed a data deduplication scheme using Convergent key encryption algorithm for Data confidentiality.

III. Our Contribution

In this paper, we present a secure system which employs combination of advanced encryption standard (AES) and cipher text-policy attribute-based encryption (CP-ABE) and supports secure deduplication using hybrid cloud. Our main contributions can be summarized as follows. Firstly, the system achieves the standard notion of security for data confidentiality in storage systems using the hybrid cloud. Secondly, we use a methodology to modify the cipher text over one access policy into cipher texts of the same plaintext but under any other access policies without revealing the underlying plaintext. Thirdly, we use an approach based on zero-knowledge proof of knowledge and commitment scheme.

IV. System Architecture

The architecture of the proposed method contains entities named data providers, attribute authority (AA), cloud and users. A data provider is the person who outsource his/her data to the cloud and share it with users possessing certain properties (attributes).The attribute authority issues every user a key associated with a set of attributes. The cloud is a hybrid cloud with public cloud for data storage and a private cloud which performs certain computation such as tag checking. Before sending a file storage request, each data provider need to create a tag and a label associated with the data, and then encrypt the data over a set of attributes. The message will be first encrypted using AES algorithm followed by ABE algorithm. Each data provider must generates a proof on the relationship of the tag, the label and the encrypted message, 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 using validity checking algorithm, then tests the equality of the new tag with existing tags in the system using equality based algorithm. If there is no match for this new tag, the private cloud adds the tag and the label to a tag-label list, and forwards the label and the encrypted data, (Label, encrypted data) to the public cloud for storage. Otherwise, discard the old and store new. At the user side, each user can download an item, and decrypt the cipher-text with the attribute-based private key generated by the AA if this user's attribute set satisfies the access structure. The valid user will get the private key for decrypting the message from the Attribute authority.

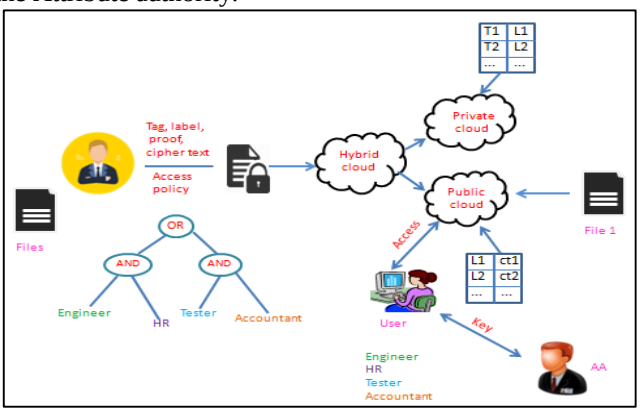

Fig. 2: System Architecture

V. Conclusion

In this system cloud will not be allow to store a file more than once. It can be used to share confidential data with other users by specifying an access policy. It achieves the standard notion

of semantic security. It provides security, and also of the efficient storage system backup facility. It can be extended to pdf files too.

VI. FUTURE WORK

It can be extended to files such as pdf, docs etc. It can be implemented as an android application. For providing more security and for facilitating easy management of the files, files can be split and store into the cloud.

REFERENCES

[1] Sahai and B. Waters, "Fuzzy identity-based encryption," in EUROCRYPT, 2005, pp. 457–473.
[2] XiMing Li, Bo Yang, Ming Wu Zhang," New Construction of Fuzzy Identity-Based Encryption," in WASE International Conference on Information Engineering, 2009.
[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, October 30 -November 3, 2006, ser. Lecture Notes in Computer Science, vol.5126. Springer, 2006, pp. 89–98.
[4] Fiat and M. Naor, "Broadcast encryption," in CRYPTO, 1993, pp. 480–491.
[5] J. Bethencourt, A. Sahai, and B. Waters, "Cipher text-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.
[6] B. Lewko and B. Waters, "Unbounded HIBE and attribute based encryption," in Advances in Cryptology - EUROCRYPT 2011-30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings, ser. Lecture Notes in Computer Science, vol. 6632. Springer, 2011, pp. 547–567.
[7] L. Cheung and C. C. Newport, "Provably secure cipher text 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.
[8] 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.
[9] J. Lai, R. H. Deng, Y. Yang, and J. Weng, "Adaptable cipher text policy attribute-based encryption," in Pairing-Based Cryptography -Pairing 2013 - 6th International Conference, Beijing, China, November 22-24, 2013, Revised Selected Papers, ser. Lecture Notes in Computer Science, vol. 8365. Springer, 2013, pp. 199–214.
[10] M. Pirretti, P. Traynor, P. McDaniel, and B. Waters, "Secure attribute-based systems," in ACM Conference on Computer and Communications Security, 2006, pp. 99–112.
[11] 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.
[12] 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.
[13] Sukanya Gunjal,"Secure Deduplication On Encrypted Big Data In Cloud Computing Environment," in IJIRSET, 2017.
[14] R Shobana,K Shanta Shalini,S.Leelavathy And V Sridevi "De-Duplication Od Data In Cloud," In TSIJOURNALS,2016.
[15] Pasquale Puzio, RefikMolva, Melek O nen, Sergio Loureiro.,"ClouDedup: Secure Deduplication with Encrypted Data for Cloud Storage" IEEE Cloud Com, 2013.
[16] Vaishnavi Moorthy, Arpit Parwal and Udit Rou," Deduplication In Cloud Storage Using Hashing Technique For Encrypted Data,"in ARPN Journal of Engineering and Applied Sciences, vol. 13, No. 5, March 2018.
[17] Jan Stanek, Alessandro Sorniottiy, Elli Androulakiy, and Lukas Kencl, "A Secure Data De-duplication Scheme for Cloud Storage", IBM Research, Zurich, May 1994.
[18] Melissa Chase," Multi-Authority Attribute Based Encryption," Computer Science Department Brown University.
[19] Sandeep Kaur ,"A Novel Technique to Remove Duplicancy of Files and Encrypt the Data Files Symmetrically In Cloud Environment," in International Journal of Advance research, Ideas and Innovations in Technology,2017.
[20] Anita N, Kumar, S. R., & Kumar, P. P. (2016). A survey on data redundancy check in a hybrid cloud by using convergent encryption. Indian Journal of Science and Technology, 9(4).

