International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 03 Issue: 06 | June -2019 ISSN: 2590-1892

Implementation of Data Deduplication and Heterogeneous Storage for Cloud

Akshay Jambholkar[1], Deepak Koradkar[2], Nikhil Patil[3] , Shubham Patil[4], Prof.Suvrna Ghule[5]

[1]Akshay Jambholkar Computer Engineering JSPM's Jayawantrao Sawant College of Engineering ,Pune-411028

[2]Deepak Koradkar Computer Engineering JSPM's Jayawantrao Sawant College of Engineering ,Pune-411028

[3]Nikhil Patil Computer Engineering JSPM's Jayawantrao Sawant College of Engineering ,Pune-411028

[4]Shubham Patil Computer Engineering JSPM's Jayawantrao Sawant College of Engineering,Pune-411028

[5]Prof.Suvrna Ghule Computer Engineering JSPM's Jayawantrao Sawant College of Engineering ,Pune-411028

Abstract - Here we are discuss about deduplication techniques using files are stored on cloud server. But in this section with elaborately discuss with how to check the particular file is duplicate or not. Suppose file is not duplicated, then only auditor is allow the file is stored on cloud server with using different methods based file storage. Here we are also additionally to talking about attribute based encryption method based data processing and re-encryption process is stored on sips server with Self Proxy. We are facing challenges on encrypted data storage and management with deduplication. Deduplication schemes continuously target specific application eventualities, during which the deduplication is totally controlled by either knowledge house owners or cloud servers. Then also access control mechanism is included in this section. Here overall checking deduplication with heterogeneous data storage management process across multiple cloud service providers with respect to the analyzing performance of cloud server. Key Words: de-duplication , Heterogeneous data, Encryption, Decryption, Proxy server

1.INTRODUCTION

Cloud computing permits centralized information storage and on-line access to pc services or resources. It offers a new way of Information Technology (IT) services by re-arranging various resources and providing them to users based on their demands. Cloud computing has greatly enriched pervasive services and become a promising service platform due to a number of desirable properties , such as scalability, elasticity, fault-tolerance, and pay-per-useData storage service is one of the most widely consumed cloud services. Cloud users have greatly benefited from cloud storage since they can store huge volume of data without upgrading their devices and access them at any time and in any place. However, cloud data storage offered by Cloud Service Providers (CSPs) still incurs some problems.First of all, various data stored at the cloud may request different ways of protection due to different data sensitivity. The data stored at the cloud include sensitive personal information, publicly shared data, data shared Within a group, and so on. Obviously, crucial data should be protected at the cloud to prevent from any access of Unauthorized parties. Some unimportant data, however, have no such a requirement. As outsourced data could disclose personal or even sensitive information, data owners sometimes would like to control their data by themselves, while on some occasion, they prefer to delegate their control to a third party since they cannot be always online or have no idea how to perform such . How to make cloud data access control adapt to various scenarios and satisfy different user demands becomes a practically important issue. Access control on encrypted data has been widely studied in the literature. However, few of them can flexibly support various requirements on cloud data protection in a uniform way, especially with economic deduplication management. Second, flexible cloud data deduplication with data access control is still an open issue. Duplicated data could be stored at the cloud.

2. PROBLEM STATEMENT

Here we are discuss about Deduplication techniques based secure data stored on cloud server. So in this section with we are facing some challenges and problems to be

overcome here. Data ownership proof is an essential process of data deduplication, especially for encrypted data. But this scheme does not consider flexible deduplication control across multiple CSPs. Here we are processing as proxy server with re-encrypted data stored on SPS(self-proxy server). Then facing attributes based encryption as different user attributes are converted to keys and transfer the data for user attributes keys received with ciphering data context on cloud. Here also set the access controls for download user or data consumer is given access permission from cloud service providers.

3. PROPOSED SYSTEM

We propose a scheme to deduplication encrypted data at CSP by applying PRE to issue keys to different authorized data holders based on data ownership challenge. It is applicable in scenarios where data holders are not available for deduplication control. We ar still facing challenges on encrypted information storage and management with deduplication.Traditional deduplication schemes always focus on specific application scenarios, in which the deduplication is completely controlled by either data owners or cloud servers. `We propose a heterogeneous information storage management theme, which flexibly offers both deduplication management and access management at an equivalent time across multiple Cloud Service Providers (CSPs).

4. SYSTEM ARCHITECTURE

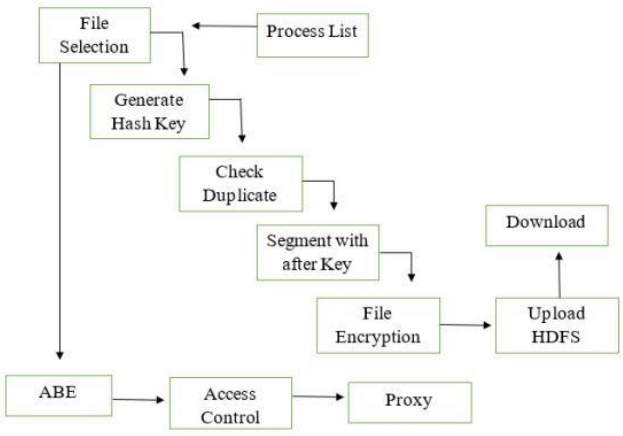


Fig-1: System Architecture

5.ALGORITHMS

5.1 ATTRIBUTE BASED ENCRYPTION (ABE)

On focussed cryptography, the system create a modification specified the focussed key of file is generated and controlled by a secret “seed”, specified any someone couldn't directlyderive the merging key from the content of file and thus the lexicon attack is prevented.In deduplication focussed cryptography provides information confidentiality. From {the information|the info|the information} content and encrypts the information copy with the focussed key a user (or data owner) derives a focussed key.In addition, the user derives a tag for the information copy, specified the tag are going to be wont to discover duplicates.Here, we have a tendency to assume that the tag correctness property holds i.e. if a pair of info copies ar the same, then their tags area unit constant.Attribute based mostly cryptography with files transferring with takes user attributes.

Attribute-based writing (ABE) is also a relatively recent approach that reconsiders the construct of public-key cryptography.In ancient public-key cryptography, a message is encrypted for a selected receiver mistreatment the receiver’s public-key.Identity-based cryptography associated above all identity-based cryptography (IBE) modified the standard understanding of public-key cryptography by permitting the public-key to be an arbitrary string, e.g., the e-mail address of the receiver.ABE goes one step extra and defines the identity not atomic but as a group of attributes, e.g., roles, and messages is encrypted with regard to subsets ofattributes (key-policy ABE - KP-ABE) or policies printed over a group of attributes (ciphertext-policy ABE - CP-ABE).The key issue is, that someone have to be compelled to only be able to decipher a ciphertext if the person holds a key for "matching attributes" (more below) where user keys are always issued by some trusty party.

Ciphertext-Policy ABE

In ciphertext-policy attribute-based secret writing (CP-ABE) a user’s private-key is related to a collection of attributes associate degreed a ciphertext specifies an access policy over an outlined universe of attributes among the system. A user are going to be beer to rewrite a ciphertext, if and provided that his attributes satisfy the policy of the

several ciphertext. Policies may be defined over attributes using conjunctions, disjunctions and (k,n)(k,n)-threshold gates, i.e., kk out of nn attributes have to be present (there may also be non-monotone access policies with extra negations and meantime there are constructions for policies outlined as impulsive circuits). For instance, let us assume that the universe of attributes is defined to be and user 1 receives a key to attributes and user 2 to attribute . If a ciphertext is encrypted with respect to the policy (A∧C)∨D(A∧C)∨D, then user 2 will be able to decrypt, while user 1 will not be able to decrypt.CP-ABE thus allows to comprehend implicit authorization, i.e., authorization is included into the encrypted data and only people who satisfy the associated policy can decrypt data. Another nice features is, that users can obtain their private keys after data has been encrypted with respect to policies. So data can be encrypted without knowledge of the actual set of users that will be able to decrypt, but only specifying the policy which allows to decrypt. Any future users can|which will|that may} lean a key with regard to attributes such the policy are often happy will then be ready to rewrite the information.

Key-Policy ABE

KP-ABE is the dual to CP-ABE in the sense that an access policy is encoded into the users secret key e.g. (A∧C)∨D(A∧C)∨D, and a ciphertext is computed with respect to a set of attributes, e.g., . In this example the user would not be able to decrypt the ciphertext but would for instance be able to decrypt a ciphertext with respect to .An important property which must be achieved by each, CP- and KP-ABE is called collusion resistance. This primarily means it shouldn't be attainable for distinct users to "pool" their secret keys specified they may along decipher a ciphertext that neither of them could decrypt on their own (which is achieved by independently randomizing users' secret keys).

Beyond ABE

ABE is just one type of the more general concept of functional encryption (FE) covering IBE, ABE and many other concepts such as inner product or hidden vector encryption (yielding e.g., searchable encryption) etc. It is a very active and young field of research and has many interesting applications (in particular in the field of cloud computing).

Attribute Based Encryption Flow :

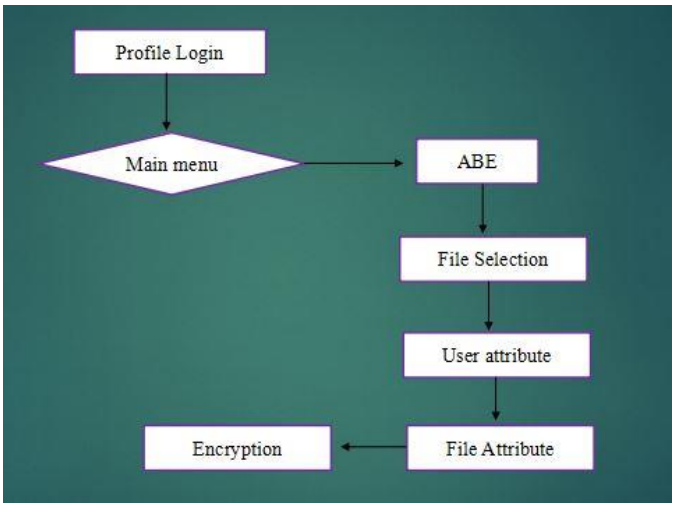


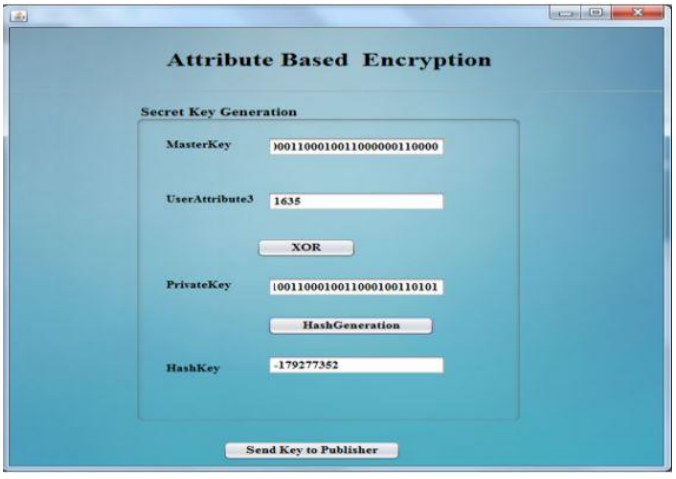
Fig-2: Attribute-based encryption flow

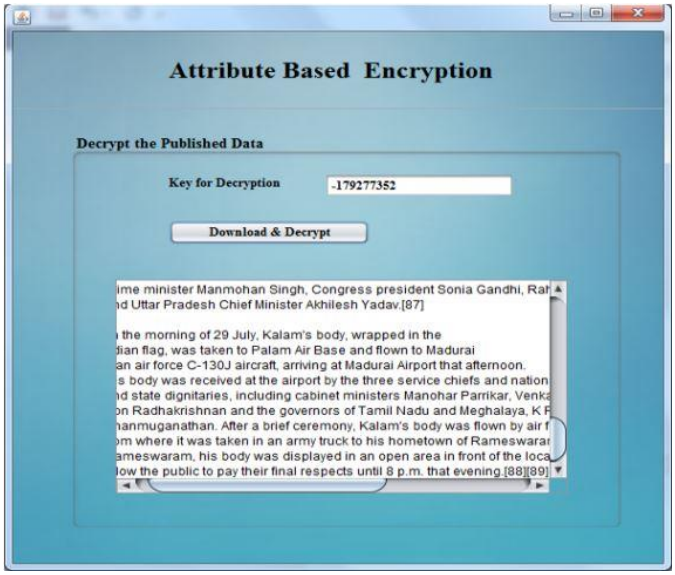
5.2 SECURE HASH ALGORITHM(SHA):

A secure hash algorithmic rule is truly a collection of algorithms developed by the National Institutes of Standards and Technology (NIST) and different government and personal parties. These secure encryption or "file check" functions have arisen to meet some of the top cyber security challenges of the 21st century, as a number of public service groups work with federal government agencies to provide better online security standards for organizations and the public. Within the family of secure hash algorithms, there square measure many instances of those tools that were got wind of to facilitate higher digital security. The first one, SHA-0, was developed in 1993. Like its successor, SHA-1, SHA-0 features 16-bit hashing.functions with 256-bit and 512-bit technologies, respectively. There is also a top-level secure hash algorithm known as SHA-3 or "Keccak" that developed from a crowd sourcing contest to see who could design another new algorithm for cybersecurity. All of those secure hash algorithms square measure a part of new secret writing standards to stay sensitive knowledge safe and forestall differing kinds of attacks. Although a number of these were developed by agencies just like the National Security Agency, and some by independent developers, all of them are related to the general functions of hash encryption

that shields knowledge in bound info and network situations, serving to to evolve cybersecurity within the digital age.

6.RESULTS





7.CONCLUSION

Data deduplication is important and significant in the practice of cloud data storage, especially for big data storage management. In this paper, we proposed a heterogeneous data storage management scheme, which offers flexible cloud data deduplication and access control. Our scheme can adapt to various application scenarios and demands and offer economic big data storage management across multiple CSPs. It can achieve data deduplication and access control with different security requirements. Security analysis, comparison with existing work and implementation based performance evaluation showed that our scheme is secure, advanced and efficient. Here also discuss with how to perform deduplication and proxy server, set for access controls on multi-users and attributes based encryption techniques.

8.REFERENCES

1]R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, and J.
Molina, ”Controlling information within the cloud: outsourcing computation whereas not outsourcing management,” in Proc. 2009 ACM Workshop Cloud Comput. Secur., pp. 85-90, 2009.

2]S. Kamara, and K. Lauter, ”Cryptographic cloud storage,” Financ. Crypto. Data Secur., pp. 136-149, Springer, 2010.

3] Q. Liu, C. C. Tan, J. Wu, and G. Wang, ”Efficient data retrieval for hierarchical queries in economical cloud environments,” in Proc.

2012 IEEE INFOCOM, pp. 2581-2585, 2012.

4] M. Kallahalla, E. Riedel, R. Swaminathan, Q. Wang, and K.

Fu, ”Plutus: scalable secure file sharing on untrusted storage,” in Proc. USENIX Conf. File Storage Technol., pp. 2942, 2003.

5]E.-J. Goh, H. Shacham, N. Modadugu, and D. Boneh, ”SiRiUS: securing remote untrusted storage,” in Proc. Netw. Distrib. Syst. Secur. Symp., pp. 131-145, 2003.

6]V. Goyal, O. Pandey, A. Sahai, and B. Waters, Attribute-based writing for fine-grained access management of encrypted data, in Proc. of thirteenth ACM Comput. Commun. Secur., pp. 8998, 2006.

7]G. J. Wang, Q. Liu, J. Wu, and M. Y. Guo, stratified attribute-based writing and scalable user revocation for sharing data in cloud servers, Comput. Secur., vol. 30, no. 5, pp. 320331, 2011.

8] Z. G. Wan, J. E. Liu, and R. H. Deng, HASBE: a stratified attribute-based resolution for versatile and scalable access management in cloud computing, IEEE Trans. Inf. Forensics Secur., vol. 7, no. 2, pp. 743-754, 2012.

9] Dropbox, ”A file-storage and sharing service,” http://www.dropbox.com/. 10 G. J. Wang, Q. Liu, and J. Wu, ”Hierarchical attribute-based writing for fine-grained access management in cloud storage services,” in Proc. of seventeenth ACM Comput. Commun. Secur., pp. 735-737, 2010.