ISSN 2229-3531 (Print)

JCIT Vol. 9(2), 26-32 (2018). Periodicity-2-Monthly

ISSN 2455-9997 (Online)

Estd. 2010

JOURNAL OF COMPUTER & INFORMATION TECHNOLOGY

An International Open Free Access Peer Reviewed Research Journal of Computer Science Engineering & Information Technology

website:- www.compitjournal.org

A Review and Survey of Various Attribute based Data Storage & Deduplication in Cloud

DEEKSHA CHOUKSEY[1] and MA RIZVI[2]

[1]PG Scholar, Department of Computer Engineering and Applications
National Institute of Technical Teachers Training & Research, Bhopal (M.P)
[2] Associate Prof. & HOD ,Department of Computer Engineering and Applications
National Institute of Technical Teachers Training & Research, Bhopal (M.P)
Corresponding Author Email: deeksha.chouksey1992@gmail.com

http://dx.doi.org/10.22147/jucit/090202

Acceptance Date 20th April 2018, Online Publication Date 24th April, 2018

Abstract

Data deduplication implies to the method that reduces the unnecessary data on the data centres and the send it on the network and it will be one of the most-enabling storage technologies that presents well-organized reserve development in the cloud computing. On the additional hand, be appropriating data deduplication acquires protection vulnerabilities in the cloud storage method so that unauthorized attributes including a cloud server or unknown users may break data confidentiality, privacy and integrity on the outsourced data. It is difficult to resolve the difficulty of data defense and confidentiality regarding data de-duplication, but positively essential for presenting amature and established cloud storage service. Here in this paper survey of all the existing techniques that are implemented based on Attribute based Encryption Data Storage and Deduplication over Cloud Computing, hence on the basis of various limitations a new and efficient technique is implemented in future.

Key words: Data Deduplication, Cloud Computing, Attribute based Encryption, Public Clouds.

Introduction
CLOUD COMPUTING

Cloud Computing is the emerging technology where we can get platform as a service, software as a service and infrastructure as a service. Once it originates to storing as a facility, data privacy and data utilization are the primary issues to be dealt with. To handle the transaction of files to and from the cloud waitperson, the archives are encrypted before being subcontracted to the profitable public mist. Cloud calculating is an emergent model offering outsourced services to enterprises for storing and processing a huge amount of data at very competitive costs. However, they do not sustenance admittance regulator strategies to normalize admittance to a certain subgroup of the deposited statistics. State-of-the-art policy based mechanisms can effort only once they are organized and functioned within an important sphere[1].

Today, our statistics is wandering yonder the limitations of our individual computers and all our statistics would still safely exist in on the web, available from any Internet-connected computer, anywhere in the ecosphere since of cloud computing.

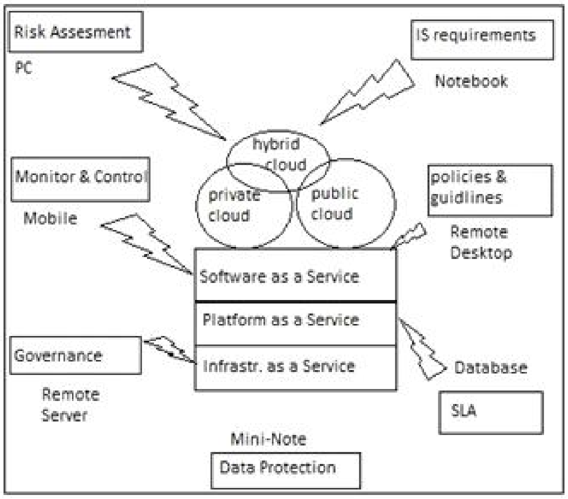

Figure 1: Cloud Computing

ATTRIBUTE BASED ENCRYPTION

Protecting networks from computer security attacks is a vital apprehension of computer security. Since the large amount of text is usually uploaded into many sites and thus it need to be secured especially when sensitive information is uploaded. Protecting networks from computer security attacks is a vital apprehension of computer security.

Currently network and computing are more developed technologies. These enable many users to easily share and update their sensitive information with others via online storage. Peoples are usually sharing their life proceedings with natives by uploading photographs and other sensitive information in to social networking sites like Facebook, Orkut, twitter etc. As people enjoy the advantages of these new technologies & services, their main concerns about data security and efficient access control also take place. Inappropriate access of data over data centers or unknown access by external users could be potential threats to their data[2]. Since the data is private and needs to be made secure from un-authorized users in the system. Also the information refuge is made probable by provided that diverse admittance policies in the complex based n the attributes or uniqueness of the users[3].

It is essential to secluded data that is uploaded in to different social sites or stored online. Attribute-based encryption (ABE) is a talented cryptographic method that achieves a fine-grained statistics admittance organize[4,5,6,7]. It gives a way of crucial admission policies based on different attributes of the requested user, scenario, or the data object. Particularly, cipher text policy attribute-based encryption (CP-ABE) enables an encryptor (who encrypt data) to identify the characteristic set over a creation of attributes that a decryptor (that decrypt data) require to acquire in order to decrypt the ciphertext, and implement it on the stuffing[4]. Thus, each user with a unusual set of entities is allowed to decrypt an unusual block of information per the refuge policy.

Though there are a variety of schemes implemented for the encoding of the information in the network, one such method is IBE. individuality based encryption is a method which is based on the encryption of the information using distinctiveness of users. The idea is to engender a key pairs which is based on the individuality of the users and the encoding and decoding of the information is probable using these identities. Fuzzy IBE is a system which is more competent as compared to the accessible IBE, since fuzzy IBE generates confirmation using biometric system and encryption based on the biometric individuality of the client[4].

In Attribute-Based Encryption an encryptor will correlate encrypted information with a set of attribute. An influence will apprehension users miscellaneous private keys, where a user's classified key is connected with an admittance configuration over entities and reflects the admittance policy recognized to the client[7]. In an ABE organization, certain keys are developed based on the attribute of the clients and also the individual cipher text. Such type of techniques needs to be more precise and competent and blunder free as compared to other identity based encryptions and is also constructive for outsized systems[7].

Ciphertext-policy attribute based encryption (CP-ABE)[5] is a public-key cryptography primitive that is used to resolution the precise concern of fine-grained admission manage on communal information in one-to-many connections. In CP-ABE, each user is allotted a set of attributes which are surrounded into the user's clandestine key. A public key constituent is distinct for each user quality. When encrypting the communication, the encryptor chooses an admission organization on attributes, and encrypts the significance under the admittance configuration via encrypting with the equivalent public key mechanism[3]. The enforcement of admittance policies and the sustain of policy updates are significant demanding issues in the information distribution systems.

Here the Data Owner can send the data to the user by providing the attributes. On each bloc of the data an attribute is generated and encrypted data using these attributes to the data storing center. The Key generation center is used for the generation of keys with the attributes generation from the data owner. It is an authority which is responsible for the generation of keys from the various public parameters[2].

Since both of the key managers, the Key Generation Center (KGC) and the data storing center are semi-trusted, it should be prevented from accessing plaintext of the data to be shared; mean while, they should be still capable to concern secret keys to users. Collusion resistance is one of the most important security property required in ABE systems[4,5,6]. If multiple users get together, they may be capable to decrypt a ciphertext by combining their attributes even if each of the users cannot decrypt the ciphertext alone.

DATA DEDUPLICATION

In computing, **data deduplication** is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are **intelligent (data) compression** and **single-instance (data) storage**. This method is used to recover storage operation and can also be practical to complex data transfers to diminish the amount of bytes that must be sent. In the deduplication procedure, exceptional chunks of data, or byte patterns, are recognized and stored throughout a procedure of examination. As the investigation continues, other chunks are compared to the stored copy and each time a match occurs, the unneeded chunk is replaced with a small orientation that points to the stored chunk[8].

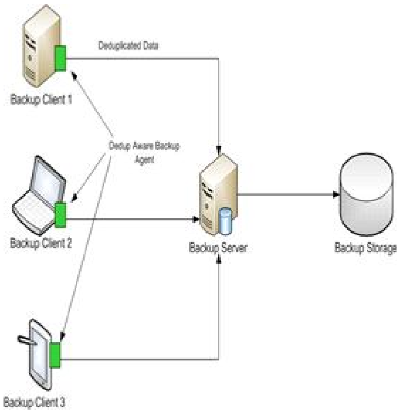

Figure 2. Target based Deduplication

Target-based deduplication acts on the target data storage media. In this case, the client is unmodified and is not aware of any deduplication. The deduplication engine can be embedded in the hardware array, which can be used as NAS/SAN device with deduplication capabilities.. Alternatively, the engine can be offered as an independent software or hardware appliance which acts as intermediary between backup server and storage arrays. In both cases, it improves only the storage utilization.

The obvious advantages are:

1. Increase in overall efficiency as data is only passed and processed once.
2. The processed data is instantaneously available for post storage processes, such as recovery and replication, reducing the RPO and RTO window.

Literature Review :

Hur, Junbeom proposed Improving security and efficiency in attribute-based data sharing. They offered a novel CP-ABE scheme for a secure data sharing system, which features the following achievements. Here in this paper various problems such as Escrow based problem and proxy encryption based problems are resolved using attribute based encryption technique[2].

Yu, Shucheng *et al.*[3] suggested Attribute based data sharing with attribute revocation. They explore a feasible solution based on novel cryptographic methods. Fig. 2 shows semi-trustable proxy servers that are always available for providing various types of content services. The scenario provided here in this technique is based on the semi-trusted servers where the data are shared among various users and the authentication is provided using the attribute policies provided to each user in the network[3].

A. Sahai and B. Waters proposed Fuzzy Identity-Based Encryption. They present two constructions of Fuzzy IBE schemes. This construction can be viewed as an Identity-Based Encryption of a message under several attributes that compose a (fuzzy) identity. This IBE schemes are both error-tolerant and secure against collusion attacks. Additionally, our basic construction does not use random oracles. They prove the security of this scheme under the Selective-ID security model. They first introduced attribute based encryption (ABE) for encrypted access control. In an ABE system, both the user secret key and the ciphertext are associated with a set of attributes. Only if at least a threshold number of attributes overlap between the ciphertext and his secret key, can the user decrypt the ciphertext[4].

V. Goyal *et al.*[5] first introduced the concept of

CP-ABE based on ABE. The main idea is to develop a much richer and secure type of attribute-based encryption cryptosystem. In this system each ciphertext is labeled by the encryptor with a set of expressive attributes. Each private key is connected with an access construction that specifies which type of ciphertexts the key can decrypt. They call such a idea a Key-Policy Attribute-Based Encryption (KP-ABE), since the access structure is specified in the private key, while the ciphertexts are simply labeled with a set of descriptive attributes. A user is able to decrypt a ciphertext if the attributes associated with a ciphertext satisfy the key's access structure. Their construction supports delegation of private keys which subsumes Hierarchical Identity-Based Encryption (HIBE)[5].

Bethencourt *et al.*[6] suggested Ciphertext-Policy Attribute Based Encryption. They provide the first construction of a ciphertext-policy attribute-based encryption (CP-ABE) to address this problem, and give the first construction of such a scheme. In this system, a user's private key will be associated with an arbitrary number of attributes expressed as strings. On the other hand, when a party encrypts a message in this system, they specify an associated access structure over attributes. A user will only be able to decrypt a ciphertext if that user's attributes pass through the ciphertexts access structure. At a mathematical level, access structures in our system are described by a monotonic "access tree", where nodes of the access structure are composed of threshold gates and the leaves describe attributes. They created a system for Ciphertext-Policy Attribute Based Encryption. Our system allows for a new type of encrypted access control where user's private keys are specified by a set of attributes and a party encrypting data can specify a policy over these attributes specifying which users are able to decrypt. This system allows policies to be expressed as any monotonic tree access structure and is resistant to collusion attacks in which an attacker might obtain multiple private keys. Finally, they provided an implementation of this system, which included several optimization techniques[6].

R. Ostrovsky *et al.*[7] proposed Attribute-Based Encryption with Non-Monotonic Access Structures. They present a new Attribute-Based Encryption scheme where private keys can represent any access formula over attributes, including non-monotone ones. In particular, our construction can handle any access structure that can be represented by a Boolean formula involving AND, OR, NOT, and threshold operations. At a high level, the technical novelty in our work lies in finding a way to (implicitly) make a share "available" to the decryptor only if a given attribute is not present among the attributes of the ciphertext. In designing this construction several challenges arise from adapting these negation techniques while preserving the collusion resistance features that are necessary for Attribute-Based Encryption systems. They achieved this through a novel application of revocation methods into existing ABE schemes. In addition, the performance of our scheme compares very favorably to that of existing, less-expressive ABE systems. An important goal in ABE systems is to create even more expressive systems[7].

Alfin Abraham *et al.*[9] proposed survey of Identity-based encryption with efficient revocation. They propose a new way to mitigate the limitation of IBE with regard to revocation, and improve efficiency of the previous solution. They want to remove interaction form the process of key update, as keeping the PKG online can be a bottleneck, especially if the number of users is very large[9].

Di Vimercati *et al.*[10] proposed Over-encryption: management of access control evolution on outsourced data. They propose a solution that removes these issues, facilitating the successful development of outsourcing data in emerging scenarios. in scenarios involving potentially huge sets of resources of considerable size, re-encryption and re-transmission by the owner may not be acceptable. The advantage compared with a solution requiring re-sending a novel encrypted version of the resource is typically huge and arbitrarily large. An important strength of this solution is that it does not substitute the current proposals, rather it complements them, enabling them to support encryption in a selective form and easily enforce dynamic policy changes[10].

Ibraimi *et al.*[11] suggested Mediated ciphertext-policy attribute-based encryption and its application. They propose a new scheme for attribute revocation in CP-ABE called mediated Ciphertext-Policy Attribute-Based Encryption (mCP-ABE). In this scheme the secret key is divided into two shares, one share for the mediator and the other for the user. To decrypt the data, the user must contact the mediator to receive a decryption token. The mediator keeps an attribute revocation list (ARL) and refuses to issue the decryption token for revoked attributes. They also define a security model for the proposed scheme which formalizes the security attacks and provide a security proof under the generic group model. The mCP-ABE scheme consists of three entities: a trusted authority (TA), a mediator and users. The TA uses the master key to generate a user secret key, which is then divided into two shares such that the first share of the user secret key is sent to the mediator and the second share of the user secret key is sent to a user. The mediator has to stay online all the time, while the TA can be put offline once it has generated secret keys for all users[11].

Gentry-Waters broadcast encryption scheme[13] to achieve attribute collusion resistance and to support complex Boolean access policies, the attribute collusion attack being likely the principal reason why broadcast encryption primitives cannot be directly used to build ABE and ABBE schemes. The security of our scheme is proven in the generic model of groups with pairings. Finally, they compare their scheme to several other ABBE designs, both in terms of bandwidth requirements and implementation costs[12]. In this paper author[14] has try to assess how can cloud providers earn their customer's trust and provide the security, privacy and reliability, when a third party is meting out sensitive data in a remote machine established in various countries. A thought of utility cloud has been characterized to provide a variety of services to the users. Various technologies can help to concentrate on the challenges of security, privacy and trust in cloud computing. Unfortunately, the implementation of cloud computing came before the suitable technologies become visible to deal with the supplementary confronts of trust. This opening between implementation and improvement is so extensive that cloud computing consumers don't fully expectation this innovative way of computing. To close this opening, we require identifying with the trust issues join together with cloud computing from both a technology and business perception. Then we'll be able to establish which up-and-coming technologies could best address these problems.

In particular, their approach[15] is more flexible than hierarchical key assignment which can only save spaces if all key-holders share a similar set of privileges methods give the first public-key patient-controlled encryption for flexible hierarchy, which was until now to be known. The difficult trouble is how to efficiently share encrypted data. Obviously users can download the encrypted data from the storage, decrypt them then send them to others for sharing, but it loses the value of cloud storage. Users should be able to delegate the access rights of the sharing data to others so that they can access these data from the server directly. However, finding an efficient and make safe way to share unfinished data in cloud storage is not insignificant. An inadequacy of their work is the predefined bounce of the number of maximum ciphertext classes. In cloud storage, the number of ciphertexts more often than not produces quickly. So we have to hold back an adequate amount of ciphertext classes for the upcoming expansion. The table shown below is the analysis and comparison of Storage Complexity between two existing Systems based on various parameters. In analysis given below 'l' shows the number of attributes in the access structure and 'k' shows size of attribute.

Table 1. Comparison of Storage Complexity

	System public paramet er|par s|	System master private key|msk|	Public Cloud label and ciphert-ext |ct| + |L|	Private Cloud tag and label |T| + |L||	User private key|sk|
CP-ABE [16]	6	1	3l+2+|A|	-	2k+2
Cui et.al's [17]	10	1	3l+5+|A|	3	2k+2

The Table shown below is the analysis and comparison of computational Costs. The analysis done here is on the basis of various parameters such as Data Provider and Private Cloud and User level on the criteria of Expo and Pairing.

Table 2. Comparison of Computational Costs

		Data Provider	Private Cloud	User <=k
CP-ABE [16]	Expo	5l+2	-	<=3k+1
	Pairing	0		
Cui et. al's [17]	Expo	5l+6	5+(6l+2)	<=k+2
	Pairing	0	2y	<=3k+1

The table shown below is the analysis of Efficiency of various techniques based on their CipherText Size and Rekeying message and Private Key Size.

Table 3. Analysis of Efficiency

System	Cipher Text Size	Rekeying message	Private key Size
BSW[6]	(2t+1)C0+C1+Ct	mC0	(2k+1)C0
BCP-ABE2 [18]	(t+2r+1)C0+ C1+Ct	0	(k+4)C0
YWRL [19]	(u+1)C0+C1+Ct	2umC0+ 2uCp	Ck+ (2u+1)C0
Hur et. al's [2]	(2t+1)C0+C1+Ct	(m+2)C0	(2k+2)C0

S. No.	Paper	Author	Issues and challenges
1.	Improving security and efficiency in attribute-based data sharing[2].	Hur, Junbeom	Efficiency is low when compared with various existing Attribute based Encryption techniques. The methodology is complex to implement and the chances of various attacks are possible due to escrow problems in the methodology.
2.	Attribute based data sharing with attribute revocation[3].	Yu, Shucheng, Cong Wang, Kui Ren, and Wenjing Lou	The number of user revocation increases with the number of number of attributes. The methodology also provides high rate of Computational cost and Storage Complexity due the number of attributes generated during Encryption.
3.	Ciphertext-Policy Attribute Based Encryption[6].	J. Bethencourt, A. Sahai, and B. Waters	This system allows policies to be expressed as any monotonic tree access structure and is resistant to collusion attacks in which an attacker might obtain multiple private keys. Finally, they provided an implementation of this system, which included several optimization techniques
4.	Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud[17].	Hui Cui	The Methodology implemented doesn't provides Security in Cloud Data Storage from various attacks and has low Utilization of Computational Cost at Data Provider and User and takes more Time for Secure Deduplication for Key Generation and Encryption and Decryption with high Computational Overhead in the Storage System.
5.	Conjunctive Broadcast and Attribute-Based Encryption[18]	N. Attrapadung and H. Imai	The System provides less efficiency with the number of attributes. The methodology generated increased size of public and private keys as well as provides increased Rekeying Size.
6.	Attribute Based Data Sharing with Attribute Revocation[19].	S. Yu, C. Wang, K. Ren, and W. Lou	The number of user revocation increases with the number of number of attributes. The methodology also provides high rate of Computational cost and Storage Complexity due the number of attributes generated during Encryption.

Conclusion

Data privacy has been a major concern in cloud storage since users have to trust the cloud service providers for security and privacy. Here we also review on various existing method for new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture to analyzed almost every security threat various de-duplication method for both the cloud models, in which the duplicate-check tokens of files are generated by the private cloud server with private keys.

Here in this paper survey of all the existing techniques that are implemented based on Attribute based Encryption Data Storage and Deduplication over Cloud Computing, hence on the basis of various limitations a new and efficient technique is implemented in future.

References

1. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-Based Access Control Models," Computer, vol. *29,* no. 2, pp. 38–47 (1996).
2. Hur, Junbeom "Improving security and efficiency in attribute-based data sharing", *IEEE Transactions On Knowledge And Data Engineering,* Vol. *25,* No. 10, pp. 2271 – 2282, October (2013).
3. Yu, Shucheng, Cong Wang, Kui Ren, and Wenjing Lou.,"Attribute based data sharing with attribute revocation", *In Proceedings of the 5th ACM Symposium on Information Computer and Communications Security,* pp. 261-270. ACM, (2010).
4. A. Sahai and B. Waters, "Fuzzy Identity-Based Encryption", *Proceedings International Conference on Theory and Applications of Cryptographic Techniques (Eurocrypt '05),* pp. 457-473 (2005).
5. V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-BasedEncryption for Fine-Grained Access Control of Encrypted Data", *Proceedings of ACM Conference on Computer and Communication Security*, pp. 89-98 (2006).
6. J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-PolicyAttribute Based Encryption", *Proceedings IEEE Symposium Security and Privacy,* pp. 321-334, (2007).
7. R. Ostrovsky, A. Sahai, and B. Waters, "Attribute-Based Encryptionwith Non-Monotonic Access Structures", *Proceedings ACM Conference Computer and Comm. Security*, pp. 195-203 (2007).
8. https://www.druva.com/blog/understanding-data-deduplication/
9. Alfin Abraham, Vinodh Ewards, Harlay Maria Mathew ,"A Survey on Optimistic Fair Digital Signature Exchange Protocols", *International Journal on Computer Science and Engineering (IJCSE)*, ISSN: 0975-3397, Vol. 3, No. 2, pp. 821 – 825, Feb (2011).
10. Di Vimercati, Sabrina De Capitani, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati Over-encryption: management of access control evolution on outsourced data." *In Proceedings of the 33rd international conference on Very large data bases*, pp. 123-134. VLDB endowment (2007).
11. Ibraimi, Luan, Milan Petkovic, Svetla Nikova, Pieter Hartel, and Willem Jonker ,"Mediated ciphertext-policy attribute-based encryption and its application", *In Information Security Applications*, pp. 309-323. Springer Berlin Heidelberg, (2009).
12. Junod, Pascal, and Alexandre Karlov ,"An efficient public-key attribute-based broadcast encryption scheme allowing arbitrary access policies", *In Proceedings of the tenth annual ACM workshop on Digital rights management,* pp. 13-24. ACM, (2010).
13. Boneh, Dan, Craig Gentry, and Brent Waters "Collusion resistant broadcast encryption with short ciphertexts and private keys", *In Advances in Cryptology–CRYPTO 2005*, pp. 258-275. Springer Berlin Heidelberg, (2005).
    Pardeep Kumar, Vivek Kumar Sehgal, Durg Singh Chauhan,
14. P. K. Gupta and Manoj Diwakar, "Effective Ways of Secure, Private and Trusted Cloud Computing "JCSI International Journal of Computer Science Issues, Vol. *8,* Issue 3, No. 2, May (2011).
15. Gade Swati,Prof.Prashant Kumbharkar, "Cryptosystem For Secure Data Sharing In Cloud Storage" IJIRT Volume 1 Issue 6 (2014).
16. Y. Rouselakis and B. Waters, "Practical constructions and new proof methods for large universe attribute-based encryption," in 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013. ACM, 2013, pp. 463–474.
17. Hui Cui, "Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud", Journal of Latex Class Files, (2016).
18. N. Attrapadung and H. Imai, "Conjunctive Broadcast and Attribute-Based Encryption," Proc. Int'l Conf. Palo Altoon Pairing Based Cryptography (Pairing), pp. 248-265 (2009).
19. S. Yu, C. Wang, K. Ren, and W. Lou, "Attribute Based Data Sharing with Attribute Revocation," Proc. ACM Symp. Information, Computer and Comm. Security (ASIACCS '10), (2010).