Secure Data Deduplication over hybrid cloud using Fine-grained access control Scheme

[1]N.Diana,[2]Bushra Tahseen
[1] Pursuing M.Tech [CSE],Dr.K.V Subbareddy Institute of Technology.
[2]Associate Professor,Department of CSE, Dr.K.V Subbareddy Institute of Technology.

Abstract: - Cloud computing plays vital role in several sector like education, banking, medical, business and several other areas. Cloud-based services for large scale content storage, processing, and distribution. Security and privacy are among prime concerns. In this regards secure data storage and sharing among cloud servers is not trusted due to cloud managed as remote service. In cloud computing there is no assurance for the data privacy, security and data integrity due to semi honest framework in nature. As our presented system huge storage of data duplication over cloud servers effect on storage space and bandwidth. As previous research work concerned secure data deduplication with access policy control is still a challenging issue in order to address this issue, we proposed a Secure CP-ABE scheme over Hybrid cloud framework which address the issues over presented system effectively.

Keywords: Cloud Computing, Hybrid cloud, CP-ABE.

1. Introduction

Cloud computing, often referred to as simply “the cloud,” is the delivery of on-demand computing resources everything from applications to data centers over the internet on a pay-for-use basis. It is common to categorize cloud computing services as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) or Software as a Service (SaaS). Essential Characteristics of Cloud Computing are

On-demand self-service: Means a consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

Broad network access: Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

Resource pooling: Merging of resources or integrating the resources towards to deliver the services

Rapid elasticity: It means scalable provisioning or ability to provide scalable services

Measured service: Cloud systems automatically control and optimize resource.

The most important and popular cloud service is data storage service. Cloud users upload personal or confidential data to the data center of a Cloud Service Provider (CSP) and allow it to maintain these data. Since intrusions and attacks towards sensitive data at CSP are not avoidable [2], it is prudent to assume that CSP cannot be fully trusted by cloud users. Moreover, the loss of control over their own personal data [1] leads to high data security risks, especially data privacy leakages [3]. Due to the rapid development of data mining and other analysis technologies, the privacy issue becomes serious. Hence, a good practice is to only outsource encrypted data to the cloud in order to ensure data security and user privacy. But the same or different users may upload duplicated data in encrypted form to CSP, especially for scenarios where data are shared among many users. Although cloud storage space is huge, data duplication greatly wastes network resources, consumes a lot of energy, and complicates data management. The development of numerous services further makes it urgent to deploy efficient resource management mechanisms [4]. Consequently, deduplication becomes critical for big data storage and processing in the cloud. Deduplication has proved to achieve high cost savings, e.g., reducing up to 96 percent storage needs for backup applications and up to 72 percent in standard file systems . Obviously, the savings, which can be passed back directly or indirectly to cloud users, are significant to the economics of cloud business. How to manage encrypted data storage with deduplication in an efficient way is a practical issue. However, current industrial deduplication solutions cannot handle encrypted data. Existing solutions for deduplication suffer from brute-force attacks [5], [6].They cannot flexibly support data access control and revocation at the same time [7], [8]. Most existing solutions cannot ensure reliability, security and privacy with sound performance.

In this paper, we propose a scheme based on data ownership challenge and Proxy Re-Encryption (PRE) to manage encrypted data storage with deduplication. We aim to solve the issue of deduplication in the situation where the data holder is not available or difficult to get involved. Meanwhile, the performance of data deduplication in our scheme

The rest of the paper is organized as follows. Section 2 gives a brief overview of related work. Section 3 introduces system and security models, Section 4 give the detailed description of our algorithm, followed by Result analysis in Section 5. Finally, a conclusion is presented in the last section.

2. Literature review

Y. Yang et.al [9] In this paper the author has proposed encryption mechanisms I.e Encryption of cloud data by the data owner is now commonly accepted as necessary to achieve data privacy against a third-party (semi-honest) cloud provider. ABE is a viable solution, as it allows fine-grained cryptographic access control, only if there is a satisfactory solution to the user revocation issue. Therefore, to address the user revocation challenge, we proposed a ciphertext-policy attribute based CPRE scheme, which was utilized in the implementing of the cloud-enabled user revocation approach. This application achieves access control over cloud data with granularity comparable to ABE, yet without incurring any cost for user revocation. Their proposal has applications in many real-world situations.

S. Keelveedhi et.al[10] In this paper the author has been studied the problem of providing secure outsourced storage that both supports deduplication and resists brute-force attacks. They design a system, DupLESS, that combines a CE-type base MLE scheme with the ability to obtain message-derived keys with the help of a key server (KS) shared amongst a group of clients. The clients interact with the KS by a protocol for oblivious PRFs, ensuring that the KS can cryptographically mix in secret material to the per-message keys while learning nothing about files stored by clients. These mechanisms ensure that DupLESS provides strong security against external attacks which compromise the SS and communication channel and that the security of DupLESS gracefully degrades in the face of comprised systems.

V. Goyal et.al[11]In this paper, the author has been considered the setting where ciphertext are associated with sets of attributes, whereas user secret keys are associated with policies. As we have discussed, this setting has a number of natural applications. Another possibility is to have the reverse situation: user keys are associated with sets of attributes, whereas ciphertext are associated with policies. We call such systems Ciphertext-Policy Attribute-Based Encryption (CP-ABE) systems. We note that the construction of Sahai and Waters [34] was most naturally considered in this framework.

J. Bethencourt et.al[12] In this paper author has been present a system for realizing complex access control on encrypted data that we call ciphertext-policy attribute-based encryption. By using our techniques encrypted data can be kept confidential even if the storage server is untrusted; moreover, their methods are secure against collusion attacks. Previous attribute-based encryption systems used attributes to describe the encrypted data and built policies into user's keys; while in our system attributes are used to describe a user's credentials, and a party encrypting data determines a policy for who can decrypt. Thus, our methods are conceptually closer to traditional access control methods such as role-based access control (RBAC). In addition, we provide an implementation of our system and give performance measurements.

J. Stanek et.al [13]This work deals with the inherent tension between well-established storage optimization methods and end-to-end encryption. Differently from the approach of related works, that assume all files to be equally security-sensitive, we vary the security level of a file based on how popular that file is among the users of the system. We present a novel encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data, so that data deduplication can be applied for the (less sensitive) popular data. Files transition from one mode to the other in a seamless way as soon as they become popular. We show that our protocols are secure under the SXDH Assumption. In the future we plan to deploy and test the proposed solution and evaluate the practicality of the notion of popularity and whether the strict popular/unpopular classification can be made more fine-grained. Also, we plan to remove the assumption of a trusted indexing service and explore different means of securing the indexes of unpopular files.

B. Waters[14] In this paper the author has been presented a new methodology for realizing Ciphertext-Policy Attribute Encryption (CPABE) under concrete and no interactive cryptographic assumptions in the standard model. Our solutions allow any encryptor to specify access control in terms of any access formula over the attributes in the system. In our most efficient system, ciphertext size, encryption, and decryption time scales linearly with the complexity of the access formula. The only previous work to achieve these parameters was limited to a proof in the generic group model. They present three constructions within our framework. Our first system is proven selectively secure under a assumption that we call the decisional Parallel Bilinear Diffie-Hellman Exponent (PBDHE) assumption which can be viewed as a generalization of the BDHE assumption. Our next two constructions provide performance tradeoffs to achieve provable security respectively under the (weaker) decisional Bilinear-Diffie-Hellman Exponent and decisional Bilinear DiffieHellman assumptions.

J. Lai et.al [15] In this paper the author has been introduced a new cryptographic primitive, called adaptable ciphertext-policy attribute-based encryption (CP-ABE). Adaptable CP-ABE extends the traditional CP-ABE by allowing a semi-trusted proxy to modify a ciphertext under one access policy into ciphertext of the same plaintext under any other access policies; the proxy, however, learns nothing about the underlying plaintext. With such “adaptability” possessed by the proxy, adaptable CP-ABE has many real world applications, such as handling policy changes in CP-ABE encryption of cloud data and outsourcing of CP-ABE encryption. Specifically, they first specify a formal model of adaptable CP-ABE; then, based on the CP-ABE scheme by Waters, they propose a concrete adaptable CP-ABE scheme and further prove its security under their security model.

3. System Study

3.1 Presented system: In our presented system User deduplication on the client-side cannot generate a new tag when they update the file. In this situation, the dynamic Ownerships would fail. As a summary, existing dynamic Ownerships cannot be extended to the multi-user environment. Whenever data is transformed, concerns arise about potential loss of data. By definition, data deduplication systems store data differently from how it was written. As a result, users are concerned with the integrity of their data. One method for deduplication data relies on the use of cryptographic hash functions to identify duplicate segments of data. If two different pieces of information generate the same hash value, this is known as a collision. The probability of a collision depends upon the hash function used, and although the probabilities are small, they are always non zero.

Disadvantages of the presented system

- No data confidentiality while sharing data with users by without specifying access policies rather than sharing decryption keys.
- There is no secure system supports for secure data deduplication.
- The standard ABE systems do not support secure deduplication, which makes them costly to be applied in some commercial storage services.

3.2 Proposed system

3.2.1 Our proposed system proposes Adaptable Attribute-Based Encryption.

CP-ABE (Ciphertext policy –Attribute based encryption) which uses to protect the user privacy where only authorized access policy satisfied person access the data means it regulating from the unauthorized access controls.

3.2.2 Deduplication in Hybrid Cloud

Hybrid cloud architecture, consisting of a pair of public and private clouds, is introduced in our storage system such that the semantic security becomes achievable for the public cloud.

We believe that the hybrid cloud architecture is a promising approach to storage systems with deduplication, in which the encrypted data is outsourced to the public cloud whilst the deduplication checking is handled by the private cloud

3.2.3 Attribute-Based Storage with Secure Deduplication

We describe a concrete construction of an attribute-based storage system supporting secure deduplication, analyze its security, and show its performance from theoretical and experimental analysis.

Proposed system flow chart

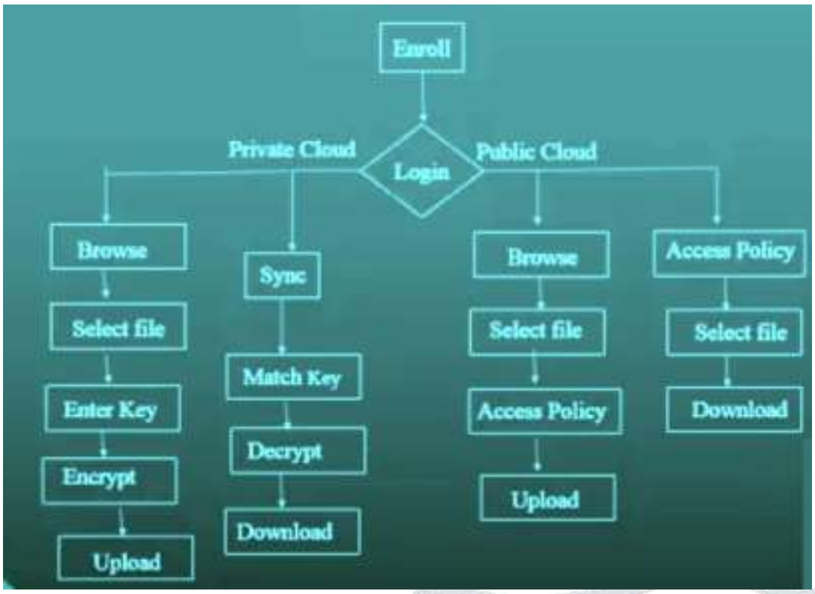

Figure 1. Proposed system flow chart

4. System Architecture

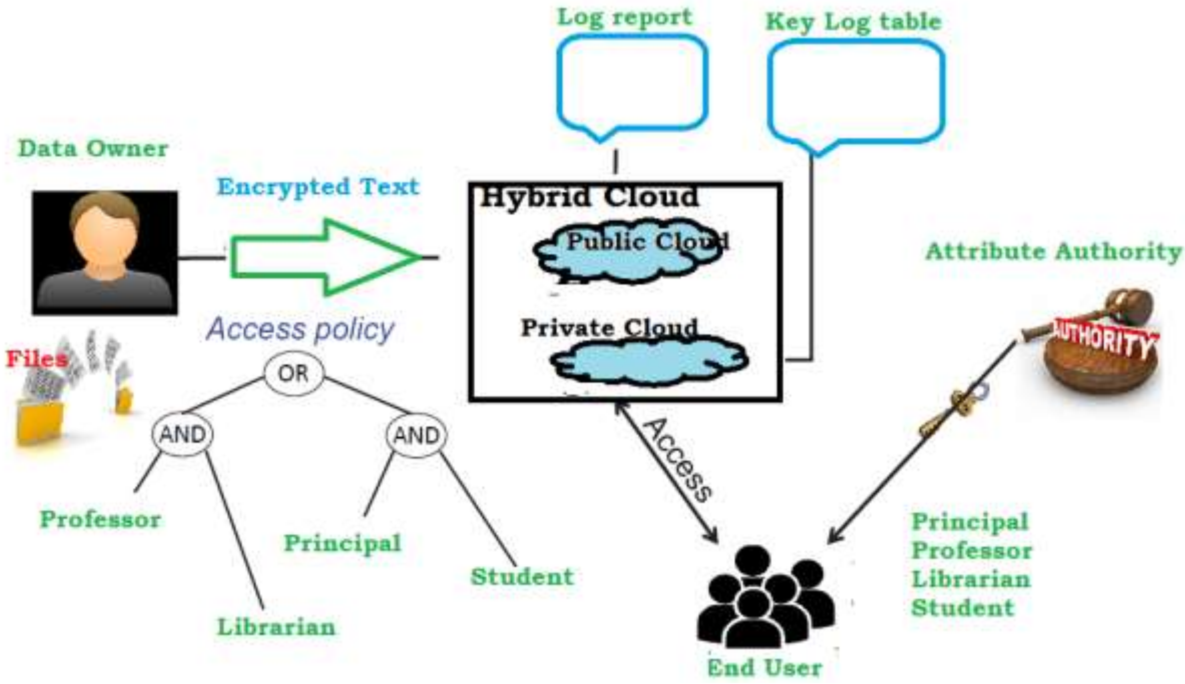

Figure 2. Proposed System architecture

In this project we have following four modules.

5. Data Owner
6. Hybrid Cloud
7. Deduplication
8. Attribute Authority

Data owner:- Data Owner uploading file to cloud with tag , label and security key , the proposed scheme guarantees data integrity against any tag inconsistency attack. Thus, security is enhanced in the proposed scheme.

Cloud Storage:-Secure Deduplication with the goal of saving storage space for cloud storage services, Douceur et al 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 cipher texts. In this case, if two users upload the same file, the cloud server can discern the equal ciphertext and store only one copy of them. Which may violate the privacy of the data if the cloud server cannot be fully trusted. This is a client who owns data, and wishes to upload it into the cloud storage to save costs. A data owner encrypts the data and outsources it to the cloud storage with its index information, that is, a tag.

Deduplication:-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 technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. Deduplication techniques take advantage of data similarity to identify the same data and reduce the storage space. In contrast, encryption algorithms randomize the encrypted files in order to make ciphertext indistinguishable from theoretically random data.

Attribute Authority:-The AA issues every user a decryption key associated with user set of attributes 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.

5. Algorithm Used

RSA Algorithm The original file to be encrypted is based on RSA algorithm. The user encrypts the original file by using the public key generated. The key generation process is carried out and the encrypted and the decrypted files are generated by RSA algorithm. The public key is known to all users to encrypt the file and the private key is kept secret and is stored in the private cloud server. The user also generates the tag while uploading the file and is stored in the server along with the private key. RSA produces 1024 bits Key.

Key generation Steps:

1. Choose two different large random prime numbers say "p" and "q".
2. Calculate n = p * q. Since "n" is the modulus for the public key and the private keys
3. Calculate the totient: Ø (n) = (p - 1)(q - 1)
4. Choose an integer "e" such that 1 < e < Ø (n) and "e" is co-prime to Ø (n) i.e. "e" and Ø (n) share no factors other than 1.

5. Find out decryption key “d” such that e * d = 1 mod (p - 1) (q - 1).
6. Encrypt the message “m” using encryption key e, c = m^e mod n.
7. Decrypt the message “m” using decryption key d, m = c^d mod n.

Algorithm for Encryption,

Step 1: Key Generation Q=d*P The key is generated for encryption and decryption purpose.

Step 2: Encryption C1=k*P, C2=M+k*Q Encryption is done using the above equation. Converting Plain text into cipher text.

Step 3: Decryption: M=C2-d*C1 Decryption is done using the above equation. Converting the cipher text into original form or plain text.

Proof: M= C2-d*C1 M can be represented as C2-d*C1

C2-d*C1=(M+K*Q)-d*(K*Q) (C2=M+K*Q and C1=K*P) =M+K*d*P-d*K*P

6. System performance

In below shown figure the performance of our attribute based storage supporting secure deduplication over encrypts and reencrypt and decryption performance based on number of consumers

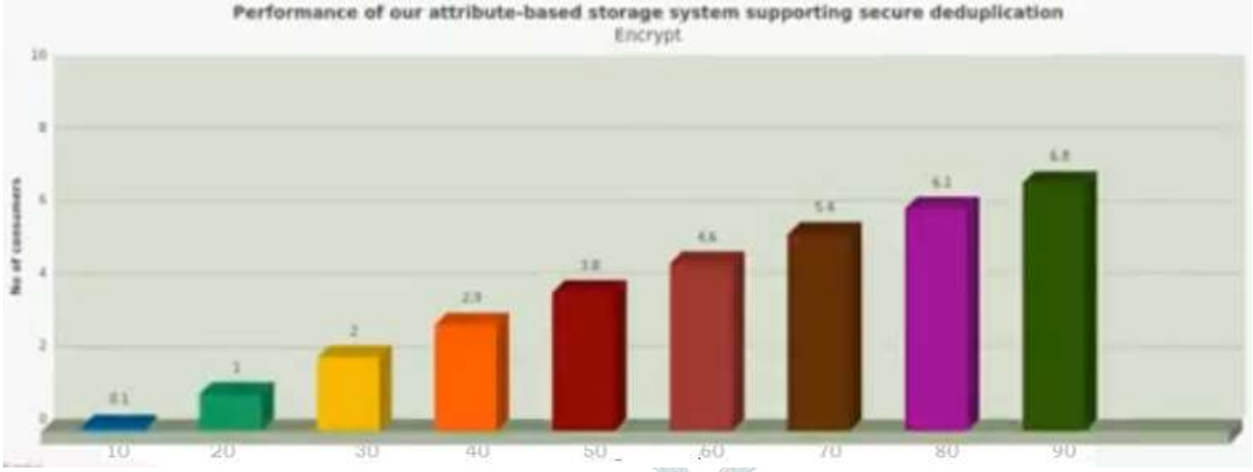

Figure 3. Performance of our attribute based storage supporting secure deduplication over encryption

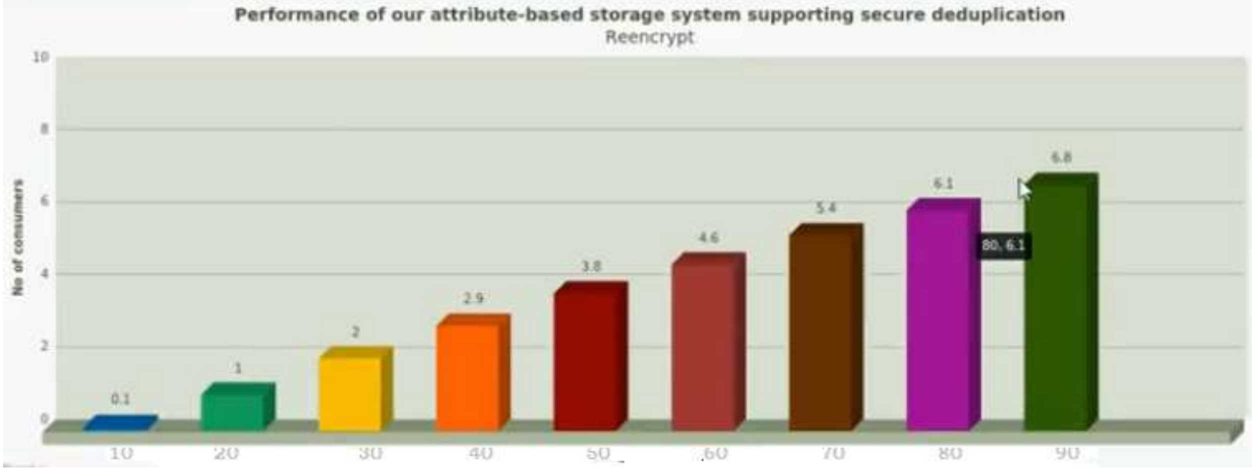

Figure 4. Performance of our attribute based storage supporting secure deduplication over Reencryption

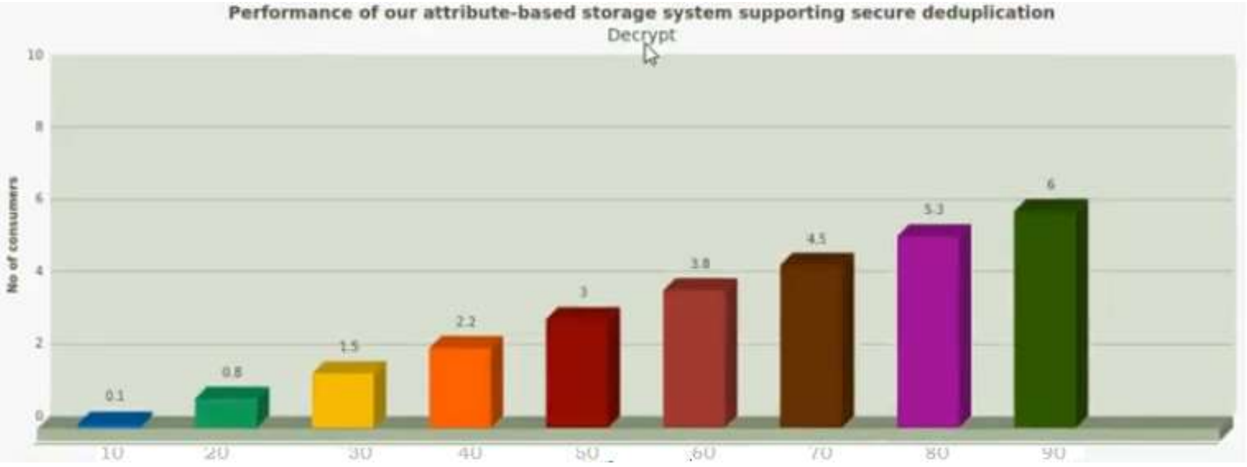

Figure 5. Performance of our attribute based storage supporting secure deduplication over Decrypt

7. Conclusion

In this paper, we proposed Secure Data Deduplication and supporting storage system using fine-grained access control Scheme, which address the privacy and security issue while deduplication from the public cloud .Our proposed framework is robust from unauthorized access controlling and datadeduplication.as future enhance of the work need to resolve the user revocation issues and its policy dynamically.

References

[1] M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," Inf. Sci., vol. 305, pp. 357– 383, 2015, doi:10.1016/j.ins.2015.01.025.

[2]T. T. Wu, W. C. Dou, C. H. Hu, and J. J. Chen, "Service mining for trusted service composition in cross-cloud environment," IEEE Systems Syst. J., vol. PP, no. 99, pp. 1–12, 2014, doi:10.1109/ JSYST.2014.2361841.

[3] L. F. Wei, et al., "Security and privacy for storage and computation in cloud computing," Inf. Sci., vol. 258, pp. 371–386, 2014, doi:10.1016/j.ins.2013.04.028.

[4] Y. Z. Zhou, Y. X. Zhang, H. Liu, N. X. Xiong, and A. V. Vasilakos, "A bare-metal and asymmetric partitioning approach to client virtualization," IEEE Trans. Serv. Comput., vol. 7, no. 1, pp. 40–53, Jan.-Mar. 2014, doi:10.1109/TSC.2012.32.

[5] Rohit Shere , Sonika Shrivastava , R.K. Pateriya," CloudSim Framework for Federation of identity management in Cloud Computing", International Journal of Computer Engineering In Research Trends.,Vol.4,no.6,pp.269-276.2017.

[6] K.Naga Maha Lakshmi, A.Shiva Kumar," Secure Data Deduplication and Data accessing among Multi-cloud Framework". International Journal of Computer Engineering In Research Trends., Vol.2, no.10, pp.687-693.2015.

[7] Ashok Nagasai Manchalla, D.Ravikiran,” Revocable Data Access Control in Public Cloud”. International Journal of Computer Engineering In Research Trends., Vol.2, no.1, pp.820-825.2015.

[8] B.Natraj Kumar, M.Sri Lakshmi,Dr S.Prem Kumar,” Investigation on Revocable Fine-grained

Access Control Scheme for Multi-Authority Cloud Storage Systems”. International Journal of Computer Engineering In Research Trends., Vol.2, no.8, pp.486-491.2015.

[9]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.

[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]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.

[12]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.

[13] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, “A secure data deduplication scheme for cloud storage,” in Financial Cryptography and Data Security - 18th International Conference, FC 2014, Christ Church, Barbados, March 3-7, 2014, Revised Selected Papers, ser.Lecture Notes in Computer Science, vol. 8437. Springer, 2014.

[14] B. Waters, “Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization,” in Public Key Cryptography - PKC 2011 - 14th International Conference on Practice and Theory in Public Key Cryptography, Taormina, Italy, March 6-9,2011. Proceedings, ser. Lecture Notes in Computer Science, vol.6571. Springer, 2011, pp. 53–70.

[15]J. Lai, R. H. Deng, Y. Yang, and J. Weng, “Adaptable ciphertextpolicy 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.