Flexible Big Data Storage Management scheme for De-duplication and access control in Cloud

[1]Shweta Pattankar [2] V Mallesi
[1]M.Tech II year, [2]Associat Professor & Head
[1]Computer Science & Engineering
[1]Bheema Institute of Technology & Science, Adoni, Dist: Kurnool, India

Abstract: In Cloud computing the most significant services are cloud big data storage and cloud computing Facilitates cloud users to expand the data storage without upgrading their devices. In this paper, we propose a flexible heterogeneous big data storage management scheme to offer both de-duplication management and access control simultaneously across multiple Cloud Service Providers (CSPs). We evaluate the performance of proposed scheme with security analysis, comparison and implementation. The implementation and analysis results show its effectiveness security and efficiency towards potential practical usage.

IndexTerms – De-duplication, Cloud computing, Access Control, Storage Management, big data.

I. INTRODUCTION

Cloud computing permits centralized storage of data and online access to resources. It presents the services by re-arranging the different resources and provides the re-arranged resources on demand to users. Cloud computing has services which are scalable, elasticity, fault tolerance and pay per use [1]. The most widely consumed services are data storage. The benefit to cloud user is that they can store large volume of data without devices up gradation and access the resources any time and place. However, cloud service providers still has problem with cloud data storage. Data de-duplication is vital and significant in the practice of big data storage management. The proposed scheme can adapt to different application scenarios and offer economic big data storage management across multiple Cloud Service Providers (CSPs)[2]. It can attain big data de-duplication and access control using different security requirements. Due to different data sensitivity there is a requirement of different ways of protection of various data stored at cloud. The various data stored in cloud are sensitive personal information, publicly shared data and data shared within a group [3]. There should be protection of crucial data stored in cloud to stop from any unauthorized parties accessing [4]. The outsourced data can be disclosed personal information, sometimes data owners want to control data by their own and sometime they prefer third party to control because they cannot be always online or have no idea how to perform such a control. The practical issue is adaptation of cloud data access control for various scenarios and fulfilling various user demands. The study of various literature gives the access control on encrypted data. But few of them will flexibly support various requirement on protection of cloud big data in uniform way i.e. especially with economic de-duplication management [5, 6, 7, 8].
Another open issue in cloud is flexibility in cloud data de-duplication with data access control. Obviously, cloud data de-duplication is mainly important for big data storage and management. However, the literature still requires studies on flexible cloud big data de-duplication across many CSPs. Existing work cannot presents a generic solution to carry both de-duplication and access control in a flexible and uniform way over the cloud system. In this paper, we propose scheme called heterogeneous big data storage management scheme in order to solve the above problems. This scheme is flexible for data owner or trusted third party or both for data storage management for performing operation like both de-duplication of data and access control. Also this proposed scheme give satisfaction about miscellaneous data security demands and simultaneously saves storage spaces across multiple CSPs with de-duplication. The proposed scheme is generic to realize encrypted cloud big data de-duplication with access control and supports the cooperation among multiple CSPs. The rest of the paper is organized as Section 2: presents the details literature survey of proposed system design Section 3: presents the existing system and details design of proposed system details Section 4: presents about system design and implementation details, section 5 presents results and discussion section 6 present about conclusion, future work and references.

II. RELATED WORK

In 2009, R. Chow et al[9], has proposed the data controlling in the cloud: outsourcing computation without outsourcing control. In this work data is encrypted before outsourcing it to the cloud in order to prevent data privacy from being invaded at Cloud Service Providers (CSPs). Access control on encrypted data requests that only authorized entities can decrypt the encrypted data. An ideal approach is to encrypt each data once and issue relevant keys to authorized entities only once. However, due to the change ability of trust relationships, key management becomes complicated due to frequent key update.
In 2010, S. Kamara et al [10], has proposed cryptographic cloud storage. The major aims of this technique a secure multi-owner information sharing theme.
In 2012, Q. Liu et al [11], have proposed the efficient information retrieval for ranked queries in cost-effective cloud environments. In this work privacy and efficiency issues are addressed.
In 2003, M. Kallahalla et al[12], have proposed the new secure file system, Plautus, which strives to provide strong security even with an untrusted server.

In 2003, E.-J. Goh et al [13], has proposed Sirius: securing remote untrusted storage, a secure file system designed to be layered over insecure network and P2P file systems such as NFS, CIFS, Ocean Store, and Yahoo! Briefcase.
In 2006, V. Goyal et al [14], have proposed attribute-based encryption for fine-grained access control of encrypted data.
M. Zhou et al [15], has proposed privacy-preserved access control for cloud computing. This technique is based on two layers of encryption that targets such requirement.
S. C. Yu, C. Wang et al [16], in 2010 proposed a secure, scalable, and fine-grained data access control in cloud computing. They developed a new cryptosystem for NE-grained sharing of encrypted data that is called Key-Policy Attribute-Based Encryption (KP-ABE). In 2012, Z. G. Wan et al [17], have proposed "HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing[18]
From the above review of literature we see that they did not considered how to solve the issue of duplicated data storage in cloud computing in a holisticand comprehensive manner, especially for encrypted data in various data storage scenarios. This issue is practically significant for big data secure storage over the cloud.
Existing industrial solutions fail to perform de-duplication on encrypted data, e.g., Dropbox [19], Google Drive [20], and Mozy [21]. Message-Locked Encryption (MLE) was proposed to resolve this tension [22]. Convergent Encryption (CE), the most prominent manifestation of MLE, was introduced [23, 24].
Bellare et al. proposed DupLESS to resist the above-mentioned brute-force attacks [26].
Wen et al. constructed a session-key-based convergent key management scheme and a convergent key sharing scheme to solve the issue that encrypted data blocks and data ownership are frequently changed [26].
Liu et al. proposed a secure cross-user de-duplication scheme that supports client-side encryption without requiring any additional independent servers by applying a password authenticated key exchange protocol[27]
Existing schemes realized de-duplication in either server-side or owner-side. Seldom, a hybrid solution was proposed to gain advantages of both approaches.
In [28], the authors proposed a method to solve de-duplication controlled by data owner only.
Hur et al. proposed a novel server-side de-duplication scheme for encrypted data[29].
Yan et al. [30, 31] proposed a de-duplication scheme based on PRE, but it completely relied on an authorized party to control data de-duplication.
In another line of our previous work [32], we applied ABE to realize de-duplicated data access control managed by data owners.
Yang et al. proposed a scheme called Provable Ownership of the File (POF)[33], which allows a user to prove to a server that it really possesses a file without the need to upload the entire file. Yuan and Yu proposed a scheme to achieve data de-duplication and secure data integrity auditing at the same time[34]. It supports both public and batch auditing. A hybrid data de-duplication mechanism was proposed by Fan et al. [35]. It can de-duplicate both plaintext and cipher-text. Li et al. formally addressed the problem of authorized data de-duplication [36].

III. PROPOSED SYSTEM

The aim of proposed work is to design a heterogeneous data management scheme to support both de-duplication and access control according to the demands of data owners.

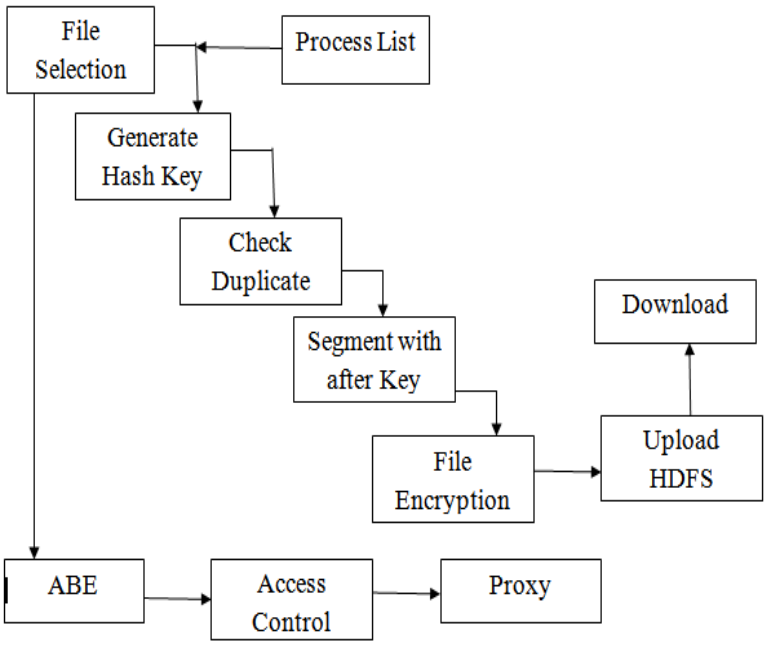

Fig.1 shows the architecture of proposed system

The proposed system is implemented language called JAVA, IDE: Net Beans7.3.1 and data base used is MySQ. The advantages of proposed system are

- Allows the client to perform the checking of duplicate copy for records selected with the particular subject.
- The distinct privilege keys are used to encode the record to ensure stronger security.
- For the purpose of checking reliability the storage space of tags are decreased. To strengthening of the security of de-duplication increased and ensure the data privacy

The objectives of proposed work are

- To save cloud storage across multiple CSPs and preserve data security and privacy by managing encrypted data storage with de-duplication in various situations.
- To support data sharing among eligible users in a flexible way, which can be controlled by either the data owners or other trusted parties or both of them.
- To justify the performance of the proposed scheme through security analysis, comparison with existing work and implementation based performance evaluation.

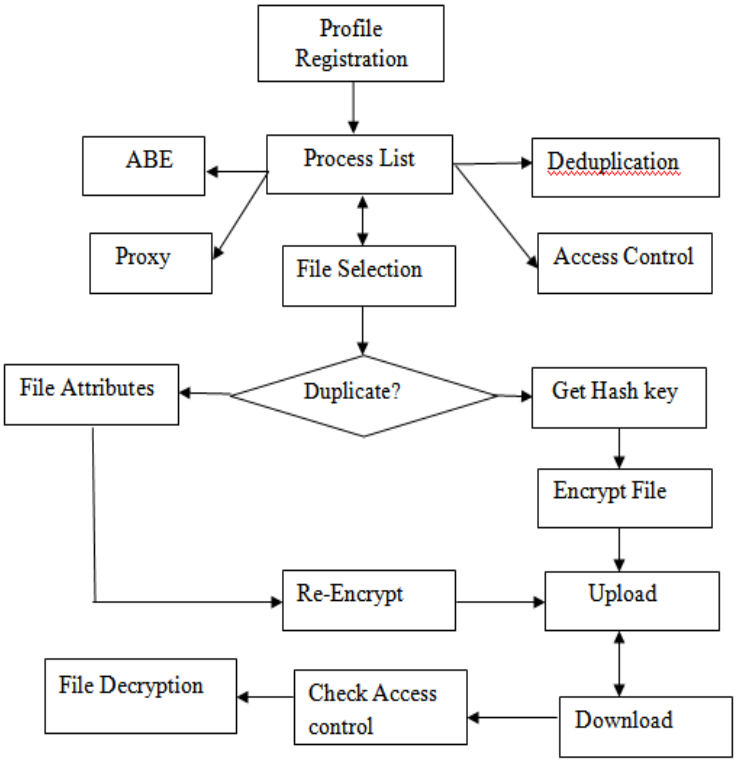

Fig.2 shows the flow diagram of proposed system design

The contributions of this paper are:

- Managing encrypted data storage using de-duplication in different conditions we can save cloud storage across multiple CSPs and data security and user privacy.
- We propose a scheme called Heterogeneous data management to support the de-duplication and also access control as per the demands of data owners.
  This scheme provides sharing of data between eligible users in a flexible way, and it can be controlled by the both data owners and other trusted parties or individual.
- We evaluate the performance of proposed scheme using security analysis and comparison with existing work. The results of performance evaluation are measured in terms of security, efficiency and potential applicability.

IV. RESULTS AND DISCUSSIONS

The proposed work implementation and execution is carried out using JAVA and NetBeans. The results are discussed in this section.The execution steps: First the user has to register using detail and then login. The process to carry out are shown in figure3.

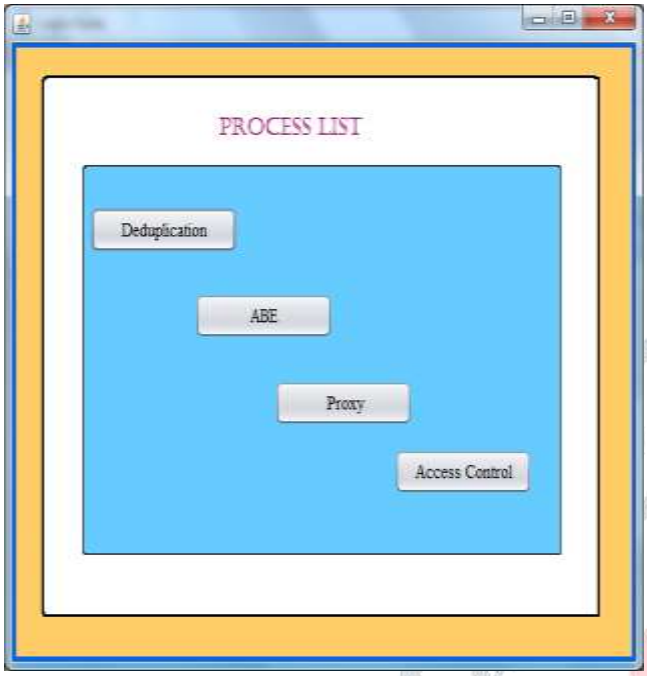

Figure 3: process list

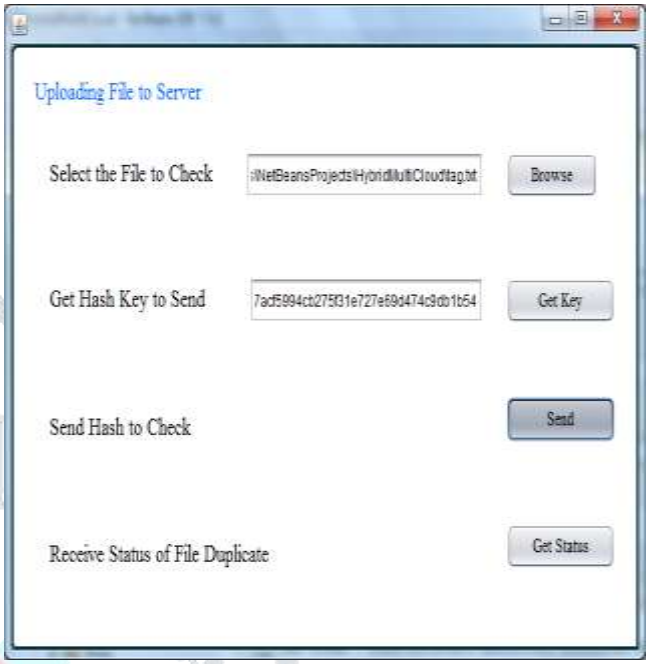

Figure 4: Uploading the File Server

i) Duplication process:

User has to register and then login. The uploading of file is done as it is shown in figure 4. Cheek for duplicate, and if there is no duplicate file present then it will display *No duplication occurred.* Also we can check block duplicate as shown in figure 5. Then it will display No duplcaition occurred. The file segmantation into blocks is done.

Figure 5: block duplicate check

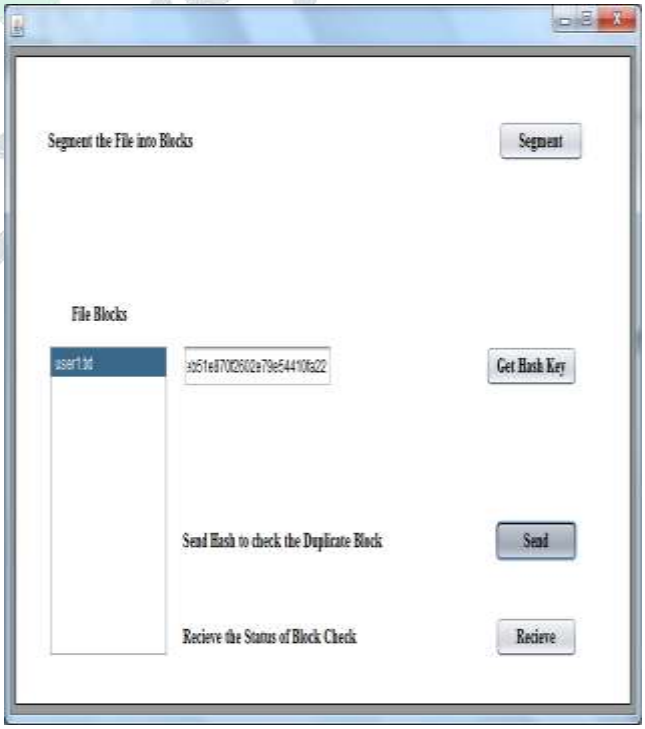

Figure 6: segmenting the file in blocks

The hash key is generated to perform file segmentation into blocks.this is shown in figure 6. The hash key is used to find duplicate block. If no duplicate block present then it will display *No duplication occurred.*

Figure 7 auditor checks for file and block duplicate

Figure 8 file encryption process and upload

The figure 8 shows the process of hash key and file encryption and uploading.

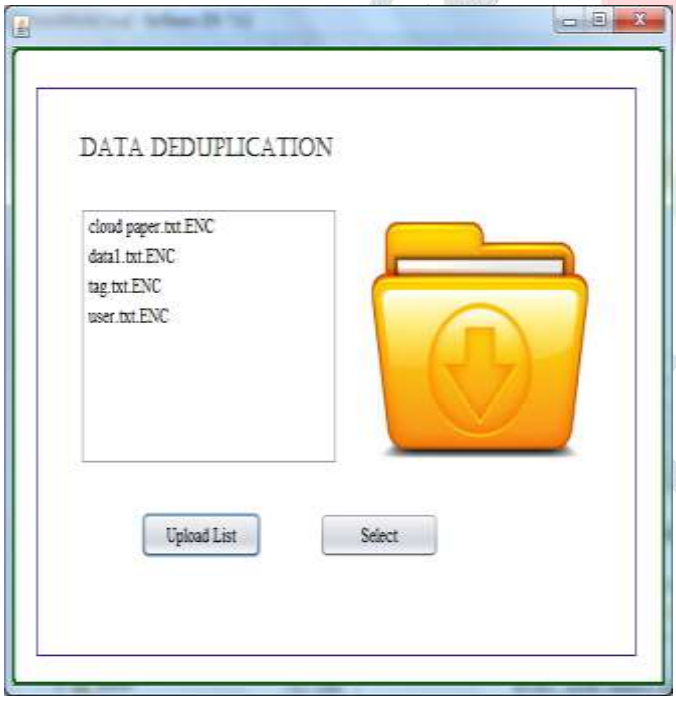

Figure 9 data de-duplication process and upload list

figure 10 cloud service provider

The figure 9 shows the uploaded list of data de-duplication and we can select the file using select option.

Figure 11 process carried out to download the file

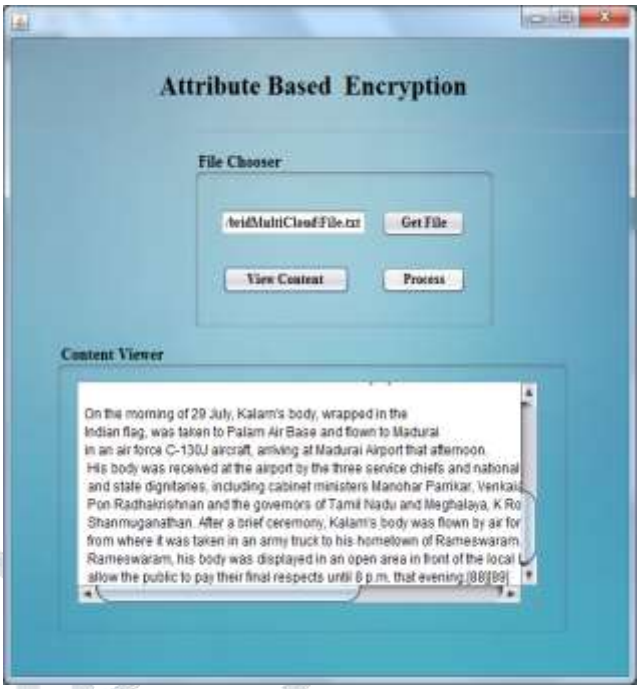

Figure 12 encrypted published data

ii. Attribute Based Encryption (ABE):

First the unique key is generated using publisher information and it display your ID is = 1534. The ID validation process is carried out using generated unique key and password. Then IP address validation is done using ID and password. Message is displayed as authenticate user and valid IP address.

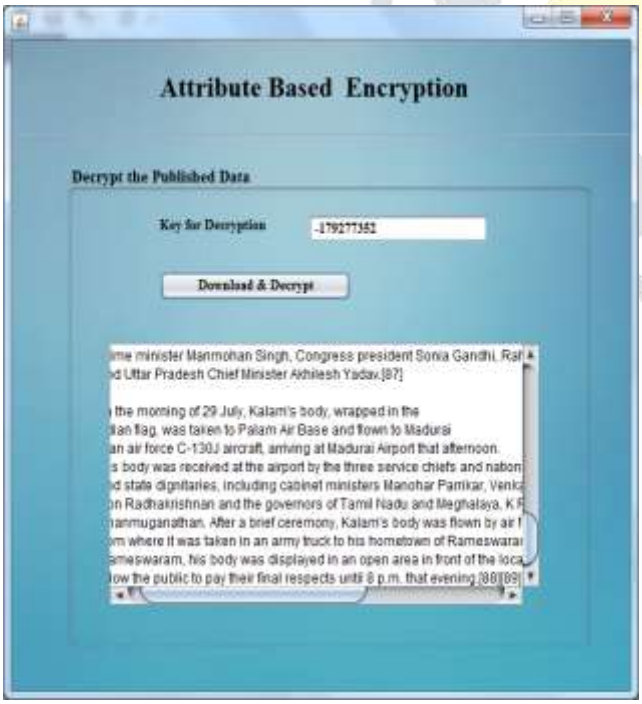

Figure 13 decrypted published data

Figure 14 proxy : key generation

The file encryption and decryption is carried out as shown in figure 12 and 13.The unique key is generated and assigned to corresponding user and then the process of file uploading is done.

iii. Proxy:

First the user has to login and it displays valid user Go ahead and then input as a file or message. Then Generate key is generated by sending the details to data owner and send to active user. User.null and send request with public key. The proxy re-encryption by data owner is performed and download the file.

iv. Access:
First choose your friends and accepts. File uploading process is done as shown in figure15 and figure 16 . The friend list file access permission: read and write. It displays permission granted.

Figure 15 file processing

Figure 16 file uploading

Friend list view with access permission is done, enter the friend name and access provider with file list. Generate file list.

CONCLUSION AND FUTURE WORK

Conclusion
In this paper, a heterogeneous big data storage management method, which presents flexible cloud data de-duplication and access control is proposed and implemented. The performance analysis based on security, comparison with existing work and implementation shows that our proposed scheme is efficient and secure. Also we study how to perform de-duplication process and proxy server set for access controls on multi-user and encryption techniques which is based on attributes.

Future Work
In future work, we consider the user privacy enhancement and performance of proposed scheme is improved towards practical deployment. .

REFERENCES

[1] C.-I. Fan, S.-Y. Huang, and W.-C. Hsu, "Hybrid data deduplication in cloud environment," in *2012 Int. Conf. Inf. Secur. Intell. Control (ISIC)*, pp. 174-177, 2012.
[2] J. W. Yuan, and S. C. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in *IEEE 2013 Conf. Commun. Netw. Secur. (CNS)*, pp. 145-153, 2013.
[3] N. Kaaniche, and M. Laurent, "A secure client side deduplication scheme in cloud storage environments," in *2014 6th Int. Conf. New Technol., Mobility Secur. (NTMS)*, pp. 1-7, 2014.
[4] Z. Yan, W. X. Ding, and H. Q. Zhu, "A scheme to manage encrypted data storage with deduplication in cloud," in *Proc. of ICA3PP2015*, pp. 547-561: Springer, 2015.
[5] Z. Yan, W. X. Ding, X. X. Yu, H. Q. Zhu, and R. H. Deng, "Deduplication on encrypted big data in cloud," *IEEE Trans. on Big Data*, vol. 2, no. 2, pp. 138-150, April-June 2016.
[6] Z. Yan, M. J. Wang, Y. X. Li, and A. V. Vasilakos, "Encrypted data management with deduplication in cloud computing," *IEEE Cloud Comput. Mag.*, vol. 3, no. 2, pp. 28-35, 2016.

[7] Z. Yan, X. Y. Li, M. J. Wang, A.V. Vasilakos, "Flexible data access control based on trust and reputation in cloud computing," *IEEE Trans. Cloud Comput.*, 2015. Doi: 10.1109/TCC.2015.2469662.
[8] J. Hur; D. Koo; Y. Shin; and K. Kang, "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 11, pp. 3113-3125, 2016
[9] R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, and J. Molina, "Controlling data in the cloud: outsourcing computation without outsourcing control," in *Proc. 2009 ACM Workshop Cloud Comput. Secur.*, pp. 85-90, 2009.
[10] S. Kamara, and K. Lauter, "Cryptographic cloud storage," *Financ. Crypto. Data Secur.*, pp. 136-149, Springer, 2010.
[11] Q. Liu, C. C. Tan, J. Wu, and G. Wang, "Efficient information retrieval for ranked queries in cost-effective cloud environments," in *Proc. 2012 IEEE INFOCOM*, pp. 2581-2585, 2012.
[12] 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. 29–42, 2003.
[13] 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.
[14] J. Bethencourt, A. Sahai, and B. Waters, "Ciphertext-policy attribute-based encryption," in *Proc. of IEEE Symp. Secur. Privacy (SP'07)*, pp. 321-334, 2007.
[15] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data", in *Proc. of 13th ACM Comput. Commun. Secur.*, pp. 89–98, 2006.
[16] S. Muller, S. Katzenbeisser, and C. Eckert, "Distributed attribute-based encryption," in *Proc. of 11th Annual Int. Conf. Inf. Secur. Crypto.*, pp. 20–36, 2008.
[17] A. Sahai, and B. Waters, "Fuzzy identity-based encryption," in *Proc. of 24th Int. Conf. Theory App. Cryptographic Tech.*, pp. 457–473, 2005
[18] S. C. Yu, C. Wang, K. Ren, and W. J. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in *Proc. of IEEE INFOCOM*, pp. 534–542, 2010.
[19] G. J. Wang, Q. Liu, J. Wu, and M. Y. Guo, "Hierarchical attribute-based encryption and scalable user revocation for sharing data in cloud servers," *Comput. Secur.*, vol. 30, no. 5, pp. 320–331, 2011.
[20] S. C. Yu, C. Wang, K. Ren, and W. J. Lou, "Attribute based data sharing with attribute revocation," in *Proc. ACM Asia Conf. Comput. Commun. Secur.*, pp. 261–270, 2010.
[21] G. J. Wang, Q. Liu, and J. Wu, "Hierarchical attribute-based encryption for fine-grained access control in cloud storage services," in *Proc. of 17th ACM Comput. Commun. Secur.*, pp. 735-737, 2010.
[22] M. Zhou, Y. Mu, W. Susilo, M. H. Au, and J. Yan, "Privacy-preserved access control for cloud computing," in *Proc. of IEEE 10th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, pp. 83-90, 2011.
[23] Z. G. Wan, J. E. Liu, and R. H. Deng, "HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 2, pp. 743-754, 2012.
[24] Z. Yan, Trust Management in Mobile Environments – Usable and Autonomic Models, IGI Global, Hershey, Pennsylvania, 2013.
[25] Y. Tang, P. P. Lee, J. C. Lui, and R. Perlman, "Secure overlay cloud storage with access control and assured deletion," *IEEE Trans. Dependable Secure Comput.*, vol. 9, no. 6, pp. 903-916, 2012.
[26] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: server aided encryption for deduplicated storage," in *Proc. of 22nd USENIX Conf. Secur.*, pp. 179-194, 2013.
[27] Dropbox, "A file-storage and sharing service," http://www.dropbox.com/.
[28] Google Drive. http://drive.google.com.
[29] Mozy, "Mozy: a file-storage and sharing Service," http://mozy.com/.
[30] J. R. Douceur, A. Adya, W. J. Bolosky, P. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in *Proc. of 22nd Int. Conf. Distributed Comput. Syst.*, pp. 617-624, 2002.
[31] G. Wallace, F. Douglis, H. W. Qian, P. Shilane, S. Smaldone, M. Chamness, and W. Hsu, "Characteristics of backup workloads in production systems," in *Proc. of USENIX Conf. File Storage Technol*., pp. 500, 2012.
[32] Z. O. Wilcox, "Convergent encryption reconsidered," 2011.
http://www.mail-archive.com/cryptography@metzdowd.com/msg08949.html.
[33] C. Yang, J. Ren, and J. F. Ma, "Provable ownership of file in de-duplication cloud storage," in *Proc. of IEEE Global Commun. Conf. (GLOBECOM)*, pp. 695-700, 2013.
[34] T.-Y. Wu, J.-S. Pan, and C.-F. Lin, "Improving accessing efficiency of cloud storage using de-duplication and feedback schemes," *IEEE Systems J.*, vol. 8, no. 1, pp. 208-218, 2014.
[27] C.-I. Fan, S.-Y. Huang, and W.-C. Hsu, "Hybrid data deduplication in cloud environment," in *2012 Int. Conf. Inf. Secur. Intell. Control (ISIC)*, pp. 174-177, 2012.
[35] J. Li, X. F. Chen, M. Q. Li, J. W. Li, P. P. C. Lee; and W. J. Lou, "Secure Deduplication with Efficient and Reliable Convergent Key Management," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 6, pp. 1615-1625, 2014.