ISSN 2347 - 3983
Volume 7, No. 11 November 2019
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter527112019.pdf
https://doi.org/10.30534/ijeter/2019/527112019

A Comprehensive review of Security Challenges for Data Deduplication and Integrity Auditing

K.Sreelatha[1], Dr. V. Krishna Reddy[2]
[1]KLEF, INDIA, gangireddysreelatha@gmail.com
[2]KLEF, INDIA, vkrishnareddy@kluniversity.in

ABSTRACT

Deduplication is a significant feature in the platform supporting the cloud platform to avoid storing redundant data on the cloud. On cloud storage servicers, Cross client data deduplication has become very popular. Data deduplication checking is carried out securely by the MLE scheme. Data-backup is an excellent solution to protect data. On making a small change in any file, a huge difference can be designed in the documentation of business. Data versions are developed to preserve the old data. In today's times, managing huge amount of growing data is a big challenge. The deduplication technique eliminates the redundancy of the data and supports the preservation of one copy of data and produces references of the copies of redundant data. To show the data deduplication as well as the integrity of data, several encryption techniques are discussed in the literature. These two problems increase within the growing amount of data in the cloud. This paper presents a review of the different works related to providing security to data deduplication and integrity privacy.

Key words: Data deduplication, integrity privacy, security, hash function, scheme, SecCloud

1. INTRODUCTION

Data deduplication minimizes the effect of repetitive data on the cost of storage. It develops an optimized space when enabled by assessing the volume data considering the portions involving duplicacy on the volume. Currently, the green-saving business has not yet cleared up more genuinely by people, particularly in the worldwide financial emergency when, how cost has always attached incredible importance to the issue of significant organizations. The green store data deduplication technology is against this framework to transform into a heated debated topic. A distance should be maintained from the different customers to set up their free extra room in the significant warehouse system condition by setting the power pool and offering the properties.

Deduplication of data can reduce the warehouse of big data, through warehouse limits, use of space and energy. Significant warehouse sellers are propelling related items or services; for example, IBM can request up to 1/25 in the IBM System Storage TS7650 G ProtecTIER items using data deduplication arrangement (Diligent). By reducing the value in efficiency facilities, the expenditures usually decrease and the use of energy decrease. NetApp V Series supports expelled excess data; at any level, 35 percent will lower data algorithm. The internet of things to come will be the evacuation of duplicate data in real-time in the necessary warehouse process. Replication is simple, witnessed rehashed replication of data where data is not spared reinforcement, but instead contains a reference to the single (one) list data. Duplication is not something else; in fact, it is only the branch of data tension. Data pressure inside a solitary record to erase copy data, supplanting the main data point to the index. Duplication is a method of data decrease, usually used in plate-based reinforcement frameworks, storage frameworks designed to decrease capacity limit utilization. This functions to find copy documents in different areas with variable-size data squares in an alternative timeframe. Copy the squares of data substituted by the predictor. Deduplication is the #1 highlight for which clients ask when they put resources into capacity arrangements.

Data deduplication distinguishes and wipes out redundancies in data, with the advantages applying to both warehouse limit investment funds ("data very still") and system transmission capacity reserve funds ("data on the wire"). Notwithstanding subduing the development away all out cost-of-proprietorship, the capacity limit reserve funds can make high IOPS gadgets like blaze based SSDs progressively possible as far as cost. The system data transfer capacity reserve funds can moderate WAN bottlenecks; in this way empowering client to-cloud and a hybrid private-open distributed storage situations.

Increased data deduplication has been around for about ten years, advocated by early new companies in the space, for example, Data Domain. Late improvements carry data deduplication to the more costly and quicker essential

warehouse level, where deduplication space investment funds are progressively significant, meaning decreases in the measure of data that should be imitated, geo-recreated, reserved, supported up, and moved over the network. Data duplication can likewise happen when you are attempting to syndicate data from different sources. Without much stretching, you can understand how this can turn into a challenge if you try to gather data about rival organizations. The Better Business Bureau, Yelp and other posting agencies will remove comments. Another daily explanation of data quality that endures is data arranging issues. It tends to be very tedious to process at a time when data organization is not uniform over your data pools. Yes, even the sincere Hadoop data mining tool will take longer to complete the undertakings exponentially. If the issues related to the organization of data are severe enough, it may be difficult to mine and process questions through any stretch of the imagination. Another daily problem that still hurts partnerships in 2019 is fragmented data. There are many potential reasons why data about an organization may not be correct. Outdated data reliability is another common problem that doesn't get as much attention as it deserves. This type of data is frequently overlooked because it is probably 100 per cent correct. This study discusses in detail the works carried out in data deduplication and the problems faced along with discussing data integrity.

Section 2 contains literature survey and in section 3 we discussed about some applications and the drawbacks of the same section 4 contains comparative table and last in section 5 we concluded the work.

2. LITERATURE REVIEW

This section discusses all the works carried out related to data deduplication and exploring the major security issues.

(El Ghazouani, El Kiram & Er-Rajy, 2019)[1] Developed a reliable trustworthy and safe known as Multi-Agent System for manipulating the technique of deduplication which provides permission to carry out the minimization technique involved in data mining, thereby minimizing the storage overhead. Due to the loss of physical control over data, issues related to data tampering, data loss and modification of data are introduced. To solve this kind of problems, Block chain has been used as the database for storing metadata of the files containing client information. This database is the logging database, which consists of auditing function supporting data integrity.

(Radhakrishnan& Jayakrishnan, 2018)[2] Concentrates on removing the duplicates and checking the integrity of the files. SecCloud+ is used to add the extra benefits of encrypting before carrying out uploading. The system is modified by inserting a procedure of compression following the procedure of encryption. Hence cloud space gets saved to a huge amount.

(Li et al., 2015)[3] Discusses the issue of auditing integrity and ensuring risk-free deduplication on cloud data. Two secure systems are proposed, including the SecCloud and SecCloud+. SecCloud discusses an auditing entity along with maintaining MapReduce cloud which supports the generation of data tags prior to the process of uploading along with auditing the data integrity present in cloud storage. In comparison to the existing work, during the stage of file uploading and auditing phase, there is a reduction in the computation in SecCloud. The constant demand of customers of encrypting their data before uploading motivates SecCloud+ and integrity auditing with deduplication of data is supported on encrypted data.

(Tian et al., 2019)[4] Discusses an updated study of the developments of cloud storage assessment (CSA) to motivate researchers to explore new research areas focusing on the efficient techniques of auditing to tackle issues of evolvement and the increasing demand for reliable and secure data storage.

(Hou, Yu & Hao, 2019)[5] Utilizes the secure encryption supporting semantic values for encrypting the private data for understanding the importance of semantic security and encrypting the big data by using encryption algorithm to show deduplication of cipher text. Due to the change in data popularity, a huge problem arises in the auditing of cloud storage. Due to the existing difference between the encryption algorithms, with the change in data popularity, the cipher text changes. Once the cipher text changes, the old authenticators become invalid for checking the integrity. This problem is tackled by exploring the numerical relationship between the old authenticators with the new ones. The designed scheme ensures that the users are free to show their presence online to do extra computation while changing the data popularity. The cloud carries out the task of transforming authentications to provide surety of smooth functioning of cloud storage auditing.

(Li & Hao, 2019)[6] Discussed the security issues found in Fan et al.'s public auditing scheme. They showed how any risky cloud could pass through the verification process of the auditor by providing valid proof even when the user's intact data is not maintained in the scheme.

(Reddy, Manjula & Venugopal, 2017)[7] Analyze the different issues, including the high computational complexity, improper method of using resources, low security and sharing of the public or private file. In several federated datacenters, different scheduling techniques are used. A more secure method is provided by the aggregate key to share file between sender and receiver. Techniques' supporting forward security encryption as well as re-encryption provides a high degree of security for the privately managed data. It provides the generation of new scale distributed computing which includes data center. The virtual machine and physical

machines consume more power. However, due to this, there is an increase in the cost involved in file transfer.

(Motghare & Satsungi, 2018)[8] Discusses the several issues faced in a cloud environment. Initially, explorations are carried out on implementing biometrics on the cloud environment. Then the existing schemes of public auditing are discussed identifying the best available option. Then the encryption and mechanism of key distribution are worked upon to find the data in the encrypted environment.

(Zhang et al., 2019)[9] Showed the risky factors involved in the auditing scheme, this supports checking the integrity and sharing of data publicly dynamically with modifications in multi-user characteristics and allows multiple users for altering data to ensure the integrity of cloud data. Moreover, a public proof of retrievability (POR) has been proposed in the cloud having fixed cost where POR obtains constant communication, public verifiability and costs of computation on users, hence proving the security of the proposed scheme.

(Li & Liu, 2019)[10] Developed a privacy-preserving authentication scheme which supports deduplication of data. The storage and the cost of the transmission of the tag can be minimized as there is a single element in the authentication tag of the message. The proposed scheme eliminates the private key of the user in the response against an auditor of the third party. In the case of deduplication of data, Bloom filter effectively checks who owns the data to match with the user claiming it. The proposed schemes prove to be flawless and anonymous under the assumption of harshness for BDH in any arbitrary oracle model. From the security analysis, it is found that during deduplication, the scheme is shaped and shows a higher degree of security and better functionality as compared to the existing schemes by evaluating the function and analyzing security.

(Masood & Muthusundar, 2018)[11] Discusses the overall view of the distributed, dependable Deduplication system. A high degree of reliability exists for the system and chunks of data are distributed among the several servers. An alternative type of protection is essential for utilizing encryption algorithms as observed in Deduplication systems.

(Sial et al., 2019)[12] Reviewed the different methods of handling issues of security at several layers supporting IoT. The present study brings out the various challenges in IoT, including the security challenges, limited hardware resources, network classification in the form of low, mid and high-level issues. The technologies involving Block chain are also reviewed.

(Verma, Gupta & Tyagi, 2019)[13] Analyzed the existing auditing schemes of data integrity by considering their distinctions. The proposed system employs Boneh–Lynn–Shacham-based signature technique and uses homomorphic linear authentication supporting the auditing supporting privacy-preserving integrity. This research helps the data owner to fasten the initial phase processing by using several cores of CPUs through a series of experiments as well as comparisons between models of single thread with multi thread.

(Aujla, Chaudhary & Kumar, 2018)[14] Develop secure storage, verification, and auditing (SecSVA) to support big data residing in IoT environment. This incorporates several models including a secure data deduplication frameworks supported by attribute to store data on the cloud, verification and authentication of Kerberos-based identity and third party auditing on the cloud which is supported by Merkle hash-tree-based trust.

(Meng, Ge & Jiang, 2019)[15] Designed a scheme where security depends on correct removal of keys in the wrapped key tree and periodic updation of the deduplication encryption keys. Further on including incremental data update the efficiency of the proposed scheme by encrypting or decrypting only the changed part along with carrying out uploading or downloading in data uploading.

(Harita & Suresh, 2018)[17] Proposed an auditing scheme supporting remote data integrity to observe the process of sensitive data hiding through data sharing. The data blocks are sanitized utilizing a sanitizer about the sensitive information from the files, and the data block signatures are transformed into valid signatures for a sanitized file. These files verify the sanitized file's integrity in the integrity auditing phase.

(Suguma & Raja, 2018)[18] Provides an unclear technique which intends to develop the numerical data in cloud data storage. This unclear technique supports encryption as well as decryption of data. A confidential system is proposed to provide improved security in the environment supporting cloud computing.

(Zhang, Huang & Wang, 2018)[19] Proposed a scheme supporting verification of data integrity considering the cloud storage of smart grid. The scheme represents an effective and secure deduplication method based on the characteristics of the bloom filter, which analyses the fast verification of the hash value of the user as well as the value of initialization. Additionally, the proposed method is combined with the mechanism of S-PDP integrity verification, random sampling strategy and data segmentation to show security deduplication as well as verification of integrity for clients of smart grid.

(Babu& Guruprakash, 2019)[20] Analyzed the several methods involved in cloud auditing considering the different parameters. Several techniques of the key update are used to improve the efficiency and privacy of the model needed in the technique of cloud auditing.

(Tian et al., 2019) [21] Discusses a public auditing scheme to store data in IoT scenarios supporting fog to cloud, helping to achieve remarkable performance as well as demands of security. A tag forming strategy is designed which is based on the technique of bilinear mapping for converting the mobile sink generated tags into the fog node generated tag in the proof generation phase which not only protects the privacy of the identity but also minimizes the cost involved in computational and communication.

(Periawamy & Latha, 2018)[22] Proposed a safe deduplication technique which develops reliable data. The study holds file level, as well as data deduplication which is fine-grained block. The consistency of the tag and integrity is provided with security.

(Jawale, 2018)[23] Discusses the implementation of deduplication. It shows data security in the environment of the cloud. The systems SecCloud and SecCloud+ are successfully implemented to show data deduplication and security auditing. SecCloud supports data integrity auditing kept in storage devices at data centers.

(Tang, Huang & Chang, 2019)[24] Proposes a real-time scheme of integrity auditing, which is effective to images of

It ensures that the cipher text present in the cloud users is accessible by valid users.

(Zhang et al., 2019)[28] Propose a scheme of integrity auditing by combining commitment of vectors and unknown decision to support signature of the group for the data present in cloud storage to support modification of data. This ensures that the server cannot compromise the privacy data of the user.

(Gan, Wang & Fang, 2018)[29] Propose an effective auditing scheme to support the outsourced big data by considering algebraic signatures and XOR-homomorphic functions, having several benefits with few drawbacks, preservation of data privacy and minimum cost of computation and communication. It allows a secure third party auditor for auditing the data in the cloud.

(Subramanian & Jeyaraj, 2018)[30] Explored the challenges associated with security for the entities of the cloud. These entities consist of Cloud Service Provider, Cloud Server and Data Owner.

(Park et al., 2018)[31] Proposed a deduplication proof of storage scheme supported by symmetric key to ensuring confidentiality along with maintaining resilience on dictionary attack and supporting integrity auditing through cryptography. Addition of the big level challenge minimizes the access of data.

cloud providing random support. The scheme uses the principle of a reversible watermarking algorithm providing images with the capacity of embedding to embed the data used for authentication.

(Tang, Zhou & Huang, 2018)[25] Proposes a deduplication scheme which is effective and supports the cross user to support the encrypted data, considering the efficiency observed during deduplication and on a lighter note obtaining the deduplication. The scheme is developed by generating a key which is non-interactive and convergent, removing the cost involved in users as well as verification of support batch data which is recovered.

(Ma et al., 2018)[26] Developed a deduplication scheme supporting random client-side that reduces attacks arising from several duplications and utilizes tags of random files to provide offline resistivity to the brute-force attacks by the outside attacker. Another method of data sharing proposed based on KEK tree.

(Yuan et al., 2018)[27] Propose a data deduplication scheme which is secure and scalable along with managing dynamic user safely updating the groups of users, limiting the unknown users from accessing the valid user's sensitive data.
(Jayapandian & Zubair Rahman, 2018)[32] Selects the method called interactive Message-Locked Encryption with Convergent Encryption (iMLEwCE) where initially after encrypting the data, the cipher text is again subjected to encryption. Block-level deduplication minimizes the space of the storage. The hacker can't deduce the critical configuration from the chunk of encrypted data.
(Subbalakshmi & Madhavi, 2018)[33] Discussed the problems related to the security of data storage as well as controlling access. It maximizes the issues related to data transfer, maintaining storage and issues related to access. Most research is performed on data security and clouds having private access.

(Batham et al., 2018)[34] Proposed a new method to support duplicate content observation and detection (DCOD). Most of the existing techniques use the tag of randomization and deduplication detection considering the size of the file, the name of the data and its annotation.

(Boyd et al., 2018) [35] Developed the natural notions of security for cloud storage and deduplication through genetic syntax for the storage of the cloud. The notions related to confidentiality and integrity is defined in existing encrypted cloud storage and identifying the relations between these notions.

(Taylor & Aboagye-Darko, 2019)[36] Discusses the various techniques of security and adopting algorithms for enhancing the security, integrity and reliability of data as well as information present in the cloud. The trending issues related to data security are also discussed.

(Space, 2018)[37] Proposed architecture, as well as a big data deduplication, is supporting privacy-preservation in the storage of the server-side. This helps to obtain privacy preservation and availability of data. The accountability is considered to get better assurances supporting privacy.

Data sharing the dynamic users to support distributed storage, CloudMe is proposed. More plans can be developed to protect the clients of the cloud.

(Zhao et al., 2018) Proposes the application of the convergent type of encryption for the application in the Medical cloud and utilizing the convey key's hash value in repeat detection and search of Bloom filter for obtaining the medical cloud's efficiency. Introducing fuzzy keywords improves the medical data's practicability that ensures secure storage through validation performed locally.

A comparative study was included in the form of table (1) in below sections.

3 APPLICATIONS INVOLVING DATA DEDUPLICATION AND HOW DRAWBACKS FACED BY THEM CAN BE IMPROVED

Data deduplication supports the reduction in the storage of repetitive data which provides an optimum solution for maintaining the bandwidth of the network and storage capacity. It has been observed that the data gets retained if stored on disks for long hours, thus maximizing the likelihood of rapid data recovery. Performance is improved by transferring fewer amounts of data. The reduction observed by a reduction in data transfer over a connection of WAN to permit the enterprises arranging their backup far away locations as well as areas with less protection. The data depute allows a higher amount of data recovery from the disk, which helps to save time and money of the enterprises and minimize

(Abdulghani et al., 2019)[38] Discusses the goals of IoT security and recognizes the stakeholders of IoT. Further potential threats and attacks related to IoT are presented. At last, a framework to support and guide privacy and security for the ideal IoT data.

(Alamelu & Akila, 2018)[39] Proposed safe management of the power, footprint and things essential for cooling in secondary storage. It enhances the protection of data.
It is necessary to understand the fine print while choosing the data depute product. There are several ways to optimize storage capacity. Few technologies involved in the process of data protection include delta differencing and compression to remove the repetitive data. Sadly in recent times, the enterprises have to analyze the advertisements thoroughly to understand what exactly is provided by the seller completely.

In applications dealing with fingerprint, a distinct data object is developed by the data dedupe integrating the hash algorithm with a fixed index for operation. The duplicated data is removed on using the file-level dedupe by performing a check on the attributes of the file and removing the repetitive files present on media, kept at backup.

The most hardware-based method in recent times solve the issue of data reduction observed in environments supporting disk-to-disk backup, but the issues arising due to the expansion and evolution of environment are hidden. The problem exists between hardware and software. In spite of providing a faster method of deployments, hardware faces limitations in terms of scalability and flexibility. The software shows a more flexible disk capacity.

Data deduplication is applicable within files, across different files, across different platforms and across clients over the globe. It supports to take backups of file systems, databases with minimum change rate, NAS, LAN, VMware environmentandSAN.

4. COMPARITIVE ANALYSIS

Reference No	Objective	Advantages	Disadvantages	Future Scope
3	To achieve the integrity of data and deduction in cloud through SecCloud and SecCloud+.	It helps to generate cloud data tags before carrying out uploading.	Computing all the exponentials for all the challenge block consumes a lot of time.	Future work needs to focus on improving the growth of auditor.
5	To understand the auditing of cloud storage along with deduplication along with providing support to various security methods.	It audits the cloud integrity even when there is a change in data popularity.	It consumes a lot of computation as well as resources of communication of the user.	Future work needs to solve the necessity of staying available constantly.
7	To study the various issues	It provides good	Due to the high power	Future works needs to

	related to security and minimize the energy consumed in virtual machines and physical machines.	demonstration of large scale distributed computing found in data centers.	consumption by virtual machines and physical machines there is an increase in the cost to transfer files.	upgrade the security algorithms and minimize the cost of file transfer by reducing consumption of power.
10	To develop a privacy-preserving public auditing scheme	It ensures privacy to user's identity minimizing the tag size to one element.	In times of damage to the storage devices, the data can be easily assessed.	Future work need stow ok on flawless against adaptive chosen-message attack
14	To develop SecSVA for allowing secure verification, auditing and storage of big data in an environment of cloud and designing a Kerberos.	It prevents data deduplication through multiple tags in cloud server.	It relies on timestamps to handle the replay attack.	Future work includes removal of replay attack without using system timestamps in the developed scheme.
16	To develop a remote data integrity auditing scheme to share sensitive data.	It ensures protection of the sensitive information shared through data sharing.	More computation overheads are spent on handling more challenged data blocks.	Future works focus on minimizing the overheads of computation.
19	To develop a data integrity verification method for supporting data deduplication and protection of smart grid in cloud storage.	It solves the issue of linear growth occurring due to the large number of keys and users sharing data.	No work is done on verifying third party.	Future work needs to work upon verifying smart grid data supporting verification of third party.
21	To develop a technique of tag-transformation using bilinear mapping.	It handles well block verification, protects identity-privacy and protects data identity.	There is an exponential growth in the amount of big-data in IoT.	Future works include designing a new model to support big data along with load balancing of several auditors.
23	To implement a deduplication technique to protect a file.	It achieves effective space for auditing.	Sharing OTP for large number of files will be tedious and time consuming.	Future work includes identifying an alternative way to protect a file.
25	To develop Encrypted Data Deduplication using Re-encryption.	It achieves light weight data deduplication.	High performance is achieved only when there is a control in the number of overheads.	Future work includes controlling overheads.
27	To develop a scalable scheme for data deduplication to provide protection from unknown attacks.	It prevents any unknown cloud user to access the sensitive data.	The time for generating key in encryption scheme is longer.	The key generation time needs to be reduced in future.
31	To develop Sec-DPoS scheme based using symmetric key cryptography.	It minimizes the access to data.	The cost of development increases with the growth in the size of the file.	Future work involves identifying alternative way to reduce the time cost on increasing the responses.

5. CONCLUSION

This study discusses the several security mechanisms to support data deduplication and integrity. The works related to SecCloud and SecCloud+ are discussed along with reviewing the encryption methodologies. The review the different kind of security challenges faced by deduplication and discussed how the growing size of data needs proper management in cloud server. The growing data are highly susceptible to unauthorized users and threats. The relationship between the notions of integrity is discussed with schemes of deduplication.

REFERENCES

1. El Ghazouani, M., El Kiram, M. A., & Er-Rajy, L. (2019). **Blockchain & Multi-Agent System: A New Promising Approach for Cloud Data Integrity Auditing with Deduplication**. International Journal of Communication Networks and Information Security, 11(1), 175-184.
2. Nair, A. S., Radhakrishnan, B., Jayakrishnan, R. P., & Kanthan, P. S. L. (2018, April). **Secure Data Deduplication and Efficient Storage Utilization in Cloud Servers Using Encryption, Compression and Integrity Auditing**. In International Conference on Soft Computing Systems (pp. 326-334). Springer, Singapore.
https://doi.org/10.1007/978-981-13-1936-5_35
3. Li, J., Li, J., Xie, D., & Cai, Z. (2015). **Secure auditing and deduplicating data in cloud**. IEEE Transactions on Computers, 65(8), 2386-2396.
4. Tian, H., Chen, Y., Jiang, H., Huang, Y., Nan, F., & Chen, Y. (2019). **Public Auditing for Trusted Cloud Storage Services.** IEEE Security & Privacy, 17(1), 10-22.
5. Hou, H., Yu, J., & Hao, R. (2019). **Cloud storage auditing with deduplication supporting different security levels according to data popularity**. Journal of Network and Computer Applications, 134, 26-39.
https://doi.org/10.1016/j.jnca.2019.02.015
6. Li, X., & Hao, R. (2019, August). **A Note on Enhancing cloud storage security against a new replay attack with an efficient public auditing scheme**. In Proceedings of the 2nd International Conference on Big Data Technologies (pp. 48-51). ACM.
7. Reddy, M., Manjula, S. H., & Venugopal, K. R. (2017). **Secure data sharing in cloud computing: a comprehensive review**. International Journal of Computer (IJC), 25(1), 80-115.
8. Motghare, S., & Satsangi, C. S. (2018). **A Review on Biometric based Privacy Preserving and Public Auditing Schemes in Cloud Computing Environment**.
https://doi.org/10.32628/CSEIT183869
9. Zhang, J., Wang, B., He, D., & Wang, X. A. (2019). **Improved secure fuzzy auditing protocol for cloud data storage**. Soft Computing, 23(10), 3411-3422.
10. Li, C., & Liu, Z. (2019). **A Secure Privacy-Preserving Cloud Auditing Scheme with Data Deduplication**. *IJ Network Security*, *21*(2), 199-210.
11. Masood, A. M. D., & Muthusundar, S. K. (2018). **Cryptographic Hashing Method using for Secure and SimilarityDetection in Distributed Cloud Data**. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(1), 107-110.
12. Sial, M. F. K. (2019). **Security Issues in Internet of Things: A Comprehensive Review**. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS), 53(1), 207-214.
13. Verma, D. K., Gupta, P., & Tyagi, R. K. (2019). **Cloud Storage–Optimization of Initial Phase for Privacy-Preserving Public Auditing**. In Ambient Communications and Computer Systems (pp. 353-365). Springer, Singapore.
https://doi.org/10.1007/978-981-13-5934-7_32
14. Aujla, G. S., Chaudhary, R., Kumar, N., Das, A. K., & Rodrigues, J. J. (2018). **SecSVA: secure storage, verification, and auditing of big data in the cloud environment**. IEEE Communications Magazine, 56(1), 78-85.
15. Meng, W., Ge, J., & Jiang, T. (2019). **Secure Data Deduplication with Reliable Data Deletion in Cloud**. International Journal of Foundations of Computer Science, 30(04), 551-570.
16. Shen, W., Qin, J., Yu, J., Hao, R., & Hu, J. (2018). **Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage**. IEEE Transactions on Information Forensics and Security, 14(2), 331-346.
17. Haritha, M. T. S. R., & Suresh, K. (2018). **Identity-based Data Auditing and Hiding for Secure Cloud Storage**.
18. Suguma, R., & Raja, K. (2018). **Data Security and Data Privacy in Cloud Computing Environment using Data Obfuscation Technique**. International Journal of Advanced Studies in Computers, Science and Engineering, 7(3), 24-29.
19. Zhang, S., Huang, K., & Wang, B. (2019). **A Data Integrity Verification Scheme with Secure Deduplication in Smart Grid Cloud Storage**. Advances in Computer, Signals and Systems, 3(1), 1-7.
20. Babu, T. K., & Guruprakash, C. D. (2019, March). **A Systematic Review of the Third Party Auditing in Cloud Security: Security Analysis, Computation Overhead and Performance Evaluation**. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 86-91). IEEE.
https://doi.org/10.1109/ICCMC.2019.8819848
21. Tian, H., Nan, F., Chang, C. C., Huang, Y., Lu, J., & Du, Y. (2019). **Privacy-preserving public auditing for secure**

data storage in fog-to-cloud computing. Journal of Network and Computer Applications, 127, 59-69.

22. Periasamy, J. K., & Latha, B. (2018). **Secure and duplication detection in cloud using cryptographic hashing method**. Int J Eng Technol, 7(1.7), 105-108.

23. Jawale, P. (2018). **Efficient Data Communication On Cloud by Secure Auditing and Deduplication**. International Journal Of Emerging Technology and Computer Science, 3(2), 49-53.

24. Tang, X., Huang, Y., Chang, C. C., & Zhou, L. (2019). **Efficient Real-Time Integrity Auditing With Privacy-Preserving Arbitration for Images in Cloud Storage System**. IEEE Access, 7, 33009-33023.

25. Tang, X., Zhou, L., Huang, Y., & Chang, C. C. (2018, August). **Efficient Cross-User Deduplication of Encrypted Data Through Re-Encryption**. In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 897-904). IEEE.

26. Ma, H., Tian, G., Liu, Z., & Zhang, L. (2018, November). **Secure Data Deduplication with Ownership Management and Sharing in Cloud Storage**. In International Conference on Frontiers in Cyber Security (pp. 168-176). Springer, Singapore.

27. Yuan, H., Chen, X., Jiang, T., Zhang, X., Yan, Z., & Xiang, Y. (2018). **DedupDUM: Secure and scalable data deduplication with dynamic user management**. Information Sciences, 456, 159-173. https://doi.org/10.1016/j.ins.2018.05.024

28. Zhang, Y., Chen, C., Zheng, D., Guo, R., & Xu, S. (2019). **Shared Dynamic Data Audit Supporting Anonymous User Revocation in Cloud Storage**. IEEE Access, 7, 113832-113843.

29. Gan, Q., Wang, X., & Fang, X. (2018). **Efficient and secure auditing scheme for outsourced big data with dynamicity in cloud**. Science China Information Sciences, 61(12), 122104.

30. Subramanian, N., & Jeyaraj, A. (2018). **Recent security challenges in cloud computing**. Computers & Electrical Engineering, 71, 28-42.

31. Park, C., Kim, H., Hong, D., & Seo, C. (2018). **A Symmetric Key Based Deduplicatable Proof of Storage for Encrypted Data in Cloud Storage Environments**. Security and Communication Networks, 2018.

32. Jayapandian, N., & Md Zubair Rahman, A. M. J. (2018). **Secure Deduplication for Cloud Storage Using Interactive Message-Locked Encryption with Convergent Encryption, To Reduce Storage Space**. *Brazilian Archives of Biology and Technology*, *61*.

33. Subbalakshmi, S., & Madhavi, K. (2018). **Security challenges of Big Data storage in Cloud environment: A Survey**. International Journal of Applied Engineering Research, 13(17), 13237-13244.

34. Batham, S., Prasad, R., Saurabh, P., & Verma, B. (2018, May). **A new approach for data security using deduplication over cloud data storage**. In Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT) (pp. 26-27).

35. Boyd, C., Davies, G. T., Gjøsteen, K., Raddum, H., & Toorani, M. (2018, October). **Security notions for cloud storage and deduplication**. In International Conference on Provable Security (pp. 347-365). Springer, Cham. https://doi.org/10.1007/978-3-030-01446-9_20

36. Taylor, M. E., & Aboagye-Darko, D. (2019, July). **Security Approaches and Crypto Algorithms in Mobile Cloud Storage Environment to Ensure Data Security**. In International Conference on Artificial Intelligence and Security (pp. 516-524). Springer, Cham.

37. Space, B. S. S. M. (2018). **Achieving Efficient, Privacy-Preserving and Data Replication for Balancing Server-side Memory Space**.

38. Abdulghani, H. A., Nijdam, N. A., Collen, A., & Konstantas, D. (2019). **A Study on Security and Privacy Guidelines, Countermeasures, Threats: IoT Data at Rest Perspective**. Symmetry, 11(6), 774.

39. Alamelu, S., & Akila, A. (2018). **A Secure Cloud Media Center Application with Secure Deduplication and Anticollusion Attack Using Saaa Model**. system, 1(11).

40. Zhao, H., Wang, L., Wang, Y., Shu, M., & Liu, J. (2018). **Feasibility study on security deduplication of medical cloud privacy data**. EURASIP Journal on Wireless Communications and Networking, 2018(1), 185. https://doi.org/10.1186/s13638-018-1192-4

41. Wankhade, Swapnil A. **Review Paper on Big Data**. International Journal of Emerging Technologies in Engineering Research (IJETER)Volume 5, Issue 5, May (2017).

42. Baghel, Sawan, P. R. M. I. T. R. Badnera, and Gaurav Saboo. **Efficient Cryptographic Algorithms for Cloud Storage Security**. International Journal of Emerging Technologies in Engineering Research (IJETER) Volume 3, Issue 2, November (2015).