Secure Cloud Storage with Dynamic Deduplication and Integrity Verification

D Asha[1], D Nagaraj[2]

[1]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: dharaasha001@gmail.com, ORC-ID: https://orcid.org/0009-0002-3382-8464

[2]Professor, Department of CSE, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: raj2dasari@gmail.com, ORC-ID: https://orcid.org/0000-0002-8511-1863

To Cite this Article

D Asha, D Nagaraj, "Secure Cloud Storage with Dynamic Deduplication and Integrity Verification", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 233-242, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp233-242

Submitted: 28-02-2026 Accepted: 01-04-2026 Published: 08-04-2026

Abstract: Today, cloud storage systems have huge issues with the process of securely erasing duplicate data and verifying the integrity of data fast and accurately. It is particularly so when active processes such as the addition or removal of blocks are taken into consideration. The traditional compression requires a lot of additional computing power since whole files must be re-encrypted and verified each time the data is modified. To bypass these issues, DIADD introduces a deduplication and data integrity check safe system, which is compatible with changing data. The scheme allows the owners of data to transmit encrypted data to other parties besides generating unique file tags through a multi-set hash facility. This will ensure that future users do not end up uploading the same information twice through proof of ownership methods. Using hash-based authenticators, integrity checks are given to a TPA. These authenticators are only used to verify the modified blocks rather than the entire file which reduces communication and computing expenses. It is also equipped with a safe and searchable trapdoor that allows you to search keywords in encrypted files stored in the cloud without revealing any information about you. The system also ensures that the storage process is efficient, the data operations are facilitated dynamically, data integrity is ensured and privacy ensured when searching. It offers a safe and scalable method of outsourcing data management in clouds.

"Index Terms: *Cloud storage, data integrity auditing, data deduplication, and data dynamics".*

1. INTRODUCTION

The massive growth in the volume of data being generated in the world has presented enormous opportunities and significant challenges to the storage and management of data. IDC predicts that by 2025, the amount of data created around the world will be more than 175 zettabytes [1]. There is an increasing requirement in people and businesses to store more digital data and it is occurring due to the increasing number of connected devices, social media, the Internet of Things (IoT) and digital transformation across most industries. This has increased the amount of digital information. Cloud storage advantages such as the ability to scale on demand, is always on, and economical to maintain infrastructure are some of the factors that make cloud storage an excellent option when dealing with a large volume of data [2]. The benefits of outsourcing data to the cloud, however, are associated with various risks associated with data security, accuracy, and access.

When individuals transfer information to third party cloud service providers, they do relinquish control of how it is handled and stored securely [3]. The inability to control it further increases the risk of data being deleted without authorization, hacked or even lost due to a system failure [4]. One issue remains a substantial one ensuring that individuals may get outsourced data safely and reliably, particularly in a shared and decentralized environment. Encryption technology is famous to ensure the security of confidential data. However, encryption adds another level of difficulty to handling redundant data and making storage more efficient.

A large portion of stored content on the cloud consists of copy files, which increase the cost of storage and reduce the efficiency of systems [5]. To prevent this, individuals have developed data deduplication techniques that identify and remove data blocks which are duplicated and retain a single copy and reference it when required [6]. Conventional deduplication algorithms, however, do not perform well in case data is encrypted. When you encrypt two different files using different keys, they will have different ciphertexts, i.e. deduplication is not possible. To avoid this problem CE was developed. It ensures that encryption of two plaintexts having the same value gives the same ciphertext [7]. Because the content is used to select the encryption keys, this technique simplifies the deduplication process without affecting the privacy of the data.

Nonetheless, CE also has certain security issues, including the vulnerability to dictionary attacks, as well as chosen-plaintext attacks. To assist in these issues, new cryptographic models and secure deduplication protocols have been developed that maintain data confidential and also simplify storage control [8]. As the amount of data continues to grow at an exponential rate, it is still important to make cloud storage solutions that are strong, safe, and good at removing duplicate data. This is in order to secure the security and privacy of outsourced data.

2. LITERATURE REVIEW

The paper by Wang et al. [9] proposed a data-management system (full of blockchain) that is GDPR-compliant to permit integrity and compliance audits in cloud systems that are IoT-powered. Their system will include a basic blockchain-based checking mechanism to ensure that information is processed in a transparent and immutable manner and it complies with legal demands such as the right to be forgotten. This method applies smart contracts and Merkle tree constructions to enhance performance in monitoring and maintain privacy and decentralization of data.

Tian et al. [10] developed a method of auditing data sharing by identities in multi-copy that has autonomous trust management. Their approach addresses the issues of altering the ownership of the data and ensuring its accuracy in the cloud storage environments where numerous copies of the data are distributed. The system ensures that a check of data integrity can be verified even with the change of ownership or copies of data because of the identity-based cryptography and blockchain integration, which allows building the trust without reliance on the central authorities.

Zhang [11] proposed the application of blockchain to verify the security of more than one copy of information in the decentralized storage systems. The primary objective of the work is to enhance storage efficiency and security with the help of smart contracts and the algorithm of verification that will allow data owners and third-party monitors to verify the integrity of the copies that are shared. The system enables the system to have dynamic operations and ensures that the existence of any malicious changes or deletion of any copy of data is reliably detected in real time.

Yang, Chen, and Chen [12] developed a safe and efficient compressive integrity auditing system of the cloud storage. Their plan reduces the level of communication and computing that requires to be carried out when conducting an audit and also prevents the risk associated with bad cloud service providers. The protocol verifies the integrity of the data where users do not need to download the data or re-encode the data completely. It achieves this by combining the compressive sensing and homomorphic message authentication codes.

John [13] in his study on IDC discussed the rate at which data is being generated and the need to prepare storage infrastructure with a zettabyte-scale digital growth. The report made us realize the size of the problems of big data and emphasized the importance of the scalable, safe, and efficient storage solutions that would be capable of keeping up with the massive expansion of the digital information.

Douceur et al. [14] thought of the methods of eliminating redundant files and solving the issue of duplicate data in serverless distributed file systems. Their initial contribution was the basis of deduplication in distributed settings since it demonstrated how to locate and eliminate duplicated material without impacting the integrity or availability of data.

In the case of the encrypted cloud storage, Song et al. [15] proposed a blockchain-based system, which would eliminate redundant data and verify its validity. The strategy is based on convergent encryption and ciphertext-policy attribute-based encryption to provide safe deduplication and retain fine-grained control of visibility to what is seen. Audit results and metadata are stored using blockchain so that they can never be modified. This ensures that there is transparency in all the storage activities and they can be monitored.

Liu et al. [16] introduced the One-Tag Checker scheme that allows to verify the security of the encrypted, deduplicated storage and leave the messages locked. The system operates by using a single verification tag on each unique piece of

data by using message-locked encryption. Quick integrity checks are then performed using this tag and hence reduces the cost of storage and connection and ensures that any unauthorized alterations on the data are detected immediately. In the case of decentralized storage systems, Tian [17] demonstrated how it is safe to discard duplicate information and distribute auditing information that is constructed on blockchain. The same data can be shared by multiple users and the model allows every user to verify the integrity of the data independently. Blockchain maintains a history of ownership of what and what has been audited and can never be altered. This creates trust in setups that are multi-tenant and which lack centralized control.

The original article by Yuan et al. [18] established a public auditing and secure deduplication system, which involves equitable arbitration based on blockchain. The system can resolve conflicts between individuals who possess data and cloud service providers through providing a reliable arbitration process. It combines the public-key encryption with the proving the ownership and maintains audit logs in blockchain. This ensures that data sharing and validation is equitable, transparent and responsible.

The new method of monitoring cloud storage was developed by Hou, Yu, and Hao [19] and considers various levels of security depending on the popularity of the data. In order to optimize the use of the resources, files that are regularly accessed are prioritized to be deduplicated and audited whereas files that are not regularly accessed receive less stringent security controls. The system dynamically varies its deduplication and tracking mechanisms depending on the manner of data access and this maintains a check on its performance and security.

Ma et al. [20] developed a method of eliminating the same data in the cloud which is founded on dynamic asset management. Their work ensures that the storage is secure and title changes are conducted in a hassle free manner in environments where there are numerous users where files are shared, copied or supplied to other individuals. The plan involves two steps where the ownership is verified and the deduplication is verified to ensure that the deletion of users, reassignment of files and their integrity is safely done. This is applicable to real life collaborative use cases of cloud.

3. MATERIALS AND METHODS

The proposed system is an efficient and secure method of data storage on the cloud. It integrates deduplication of data, dynamic data operations, integrity auditing and search of key words that uphold privacy. Encrypted data files are divided into blocks and a multi-set hash operation is applied to provide each block with a tag which is easy to find duplicates. A proof-of-ownership system prevents more than one transfer of a user by checking the access rights of the user. In the case of dynamic processes, performance is enhanced since changed blocks are only re-authenticated. A TPA verifies the integrity of stored data by ensuring that the data is right without viewing the actual data. Besides, safe search allows safe and encrypted trapdoor to access files with an encrypted query that keeps your data confidential [1518]. This holistic approach makes the current cloud storage systems more secure, effective, and convenient.

Fig.1 Proposed Architecture

The system design (Fig. 1) indicates a safe cloud storage system in which users may utilize block-wise deduplication and tagging in encrypting and uploading information. When uploading a file, it is scanned against duplicates, and it is instantly stored on the cloud. The system is dynamically updated and this implies that it is possible to add and delete

data bits without re-encryption. A TPA is a technique of integrity checking of a file in a manner of safe hash. There is encrypted search by key-word, which is safe to retrieve information. A graph on the overhead also demonstrates the effectiveness of the proposed encrypting procedures versus the existing ones.

a) Modules:

1) Signup with Cloud: This module allows new users of the cloud service to register the service with the passwords and the personal information required by the service. It retains user data in the system database through encrypted information to keep it secure. This will allow users to log-in and get secure cloud services in future.

2) Cloud User Login: This module allows the registered users to use their passwords safely to access the cloud server. It confirms the user using a password and allows him or her to access the functionality of the app, including data upload, dynamic updates, personal search, and audit of files.

3) Data Upload: Individuals who are logged in are able to send stuff to the cloud. The encryption keys, duplicate checks, file block breaking, and authentication codes and file tags are generated by this module. The validation data will be relayed to the TPA and the encrypted file will be relayed to the cloud.

4) Dynamic Updates: With the help of this tool, users can perform dynamic operations on the files they put in the cloud, such as adding new blocks of data or eliminating the old ones. This ensures the integrity of the data and ensures that updates do not disrupt the encrypted file hierarchy as well as authentication verification methods.

5) Auditing: The TPA provides users with the opportunity to request the verification of the security of the files they store in the cloud. The TPA relies on the saved authentication hashcodes to determine whether the file has changed or not. This makes sure that the data is still correct and safe without having to download the whole file.

6) Secured Search: Users can use encrypted trapdoors to do safe keyword-based searches over encrypted cloud data with this optional module. It prevents the cloud server to discover the actual keywords or contents of the files, which safeguards the privacy of the information as it is being scanned.

7) Overhead Graph: When this tool is applied, it demonstrates how much time and resources are lost on such aspects as the creation of file tags and authentications. It compares the proposed method to the already existing methods and presents the user with a graph of the system efficiency and how it can process data more efficiently.

b) Methods/Technologies:

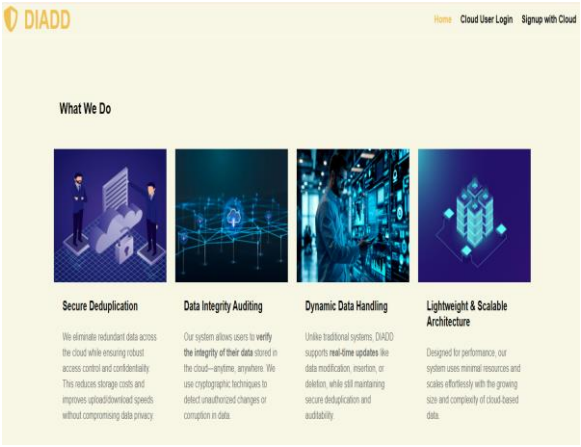
1. AES Encryption: AES is a symmetric key algorithm of encryption, which is employed to ensure that files are secure prior to their uploading to the cloud. It works without affecting privacy in that it encrypts plaintext data with a hidden key to encrypt it into ciphertext which is inaccessible. The reason why AES is very popular is that it is secure, fast and it is impossible to crack it through brute force.

2. SHA-based Authentication: The Secure Hash Algorithm (SHA) generates different hashcodes of every data. This ensures that the data is accurate and actual. SHA works in a similar manner as SHA-256, where the system uses the functions to compute the value of a set size of hash functions which act as digital fingerprints of the file. Whenever there are any changes on the file, it will generate a different number, and this will be used to identify tampering.

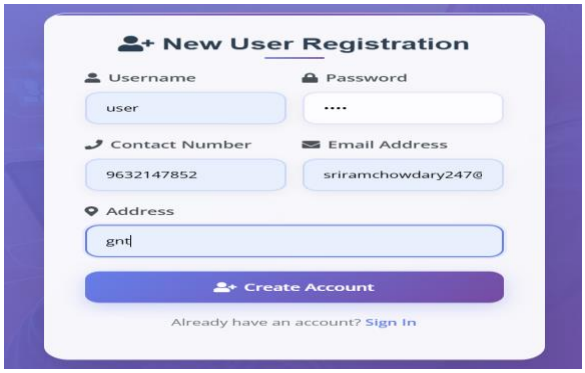
3. Trapdoor Searchable Encryption: Trapdoor Searchable Encryption allows to find what you need searching in encrypted files on clouds with safety. A trapdoor is a search token, which is encrypted and generated by users and compared with encrypted databases. This allows the users to have access to the files they require without allowing the cloud to have access to their search terms or files. This safeguards privacy and prevents leakage of data.

4. TPA-enabled Auditing: TPA allows an esteemed external party to verify the validity of information that was transmitted by another party without any knowledge of the data itself. The TPA verifies the presence of change or interference of files in the cloud by means of authentication information such as hashcodes. This way ensures that the information is dependable and simplifies the task of the individual that owns the information.

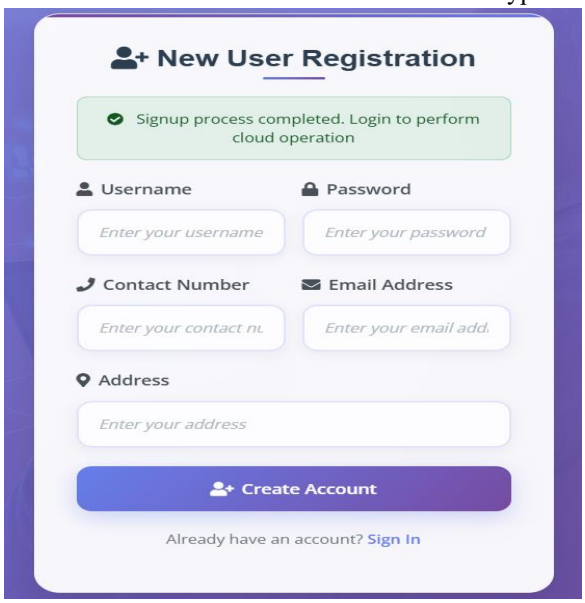
4. EXPERIMENTAL RESULTS



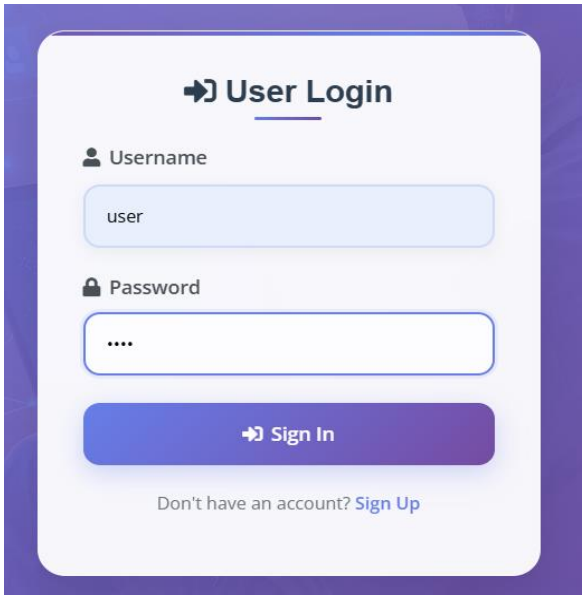
Then, the project window appears as in the figure above.



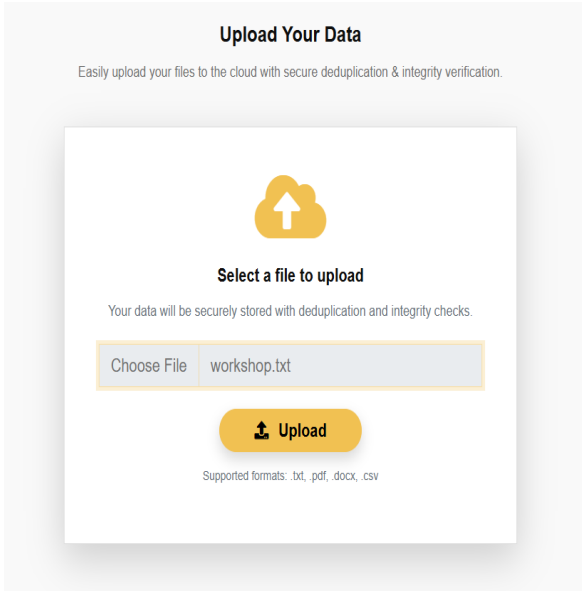
Fill in all the required fields in order to be a user and then select the type of user presented below.



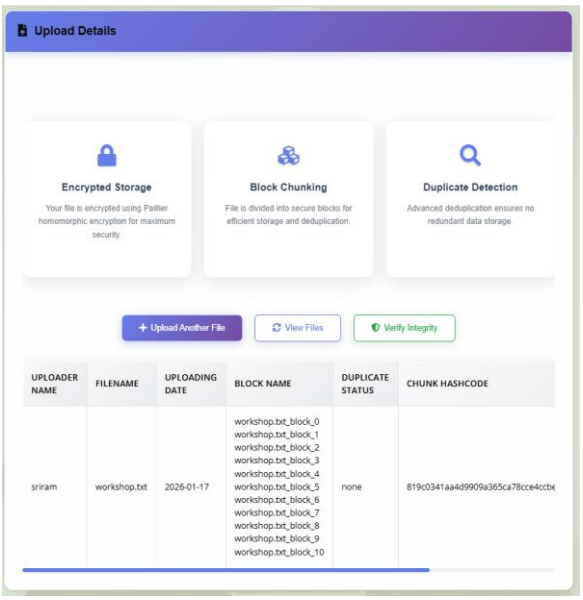
User registration is completed successfully and the data stored in the database is now safe.



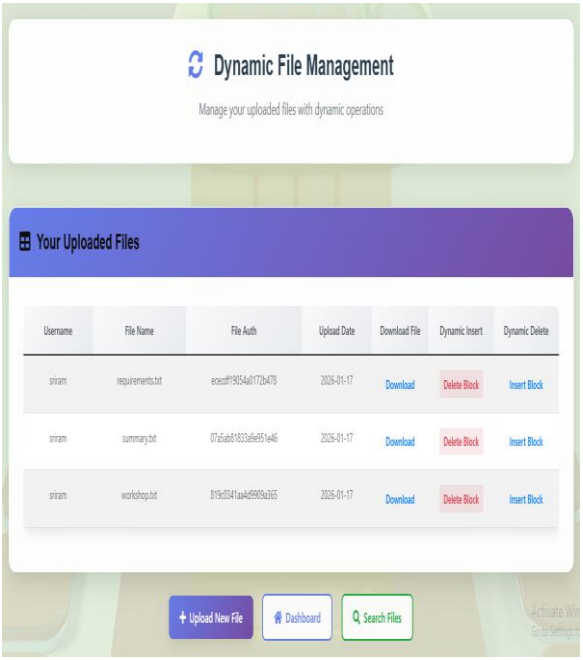
Once you have registered, fill and submit the appropriate password details that are displayed on the screen.



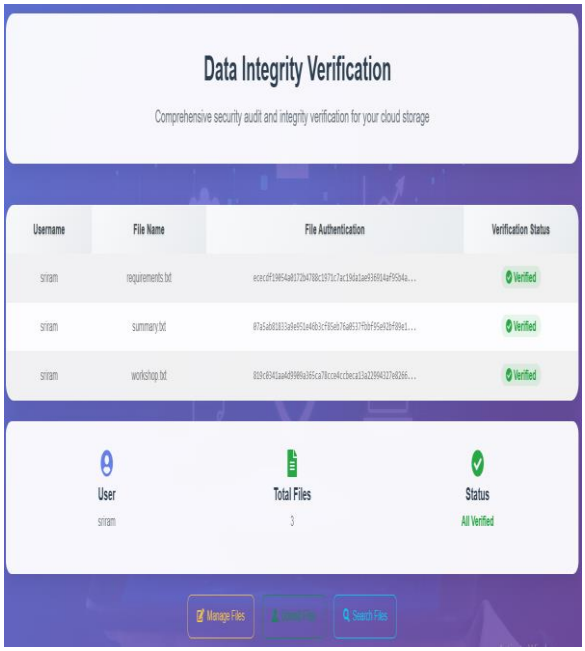
Take a file in your local storage and upload it to the cloud in order to create replicas of a file and ensure that they are right.



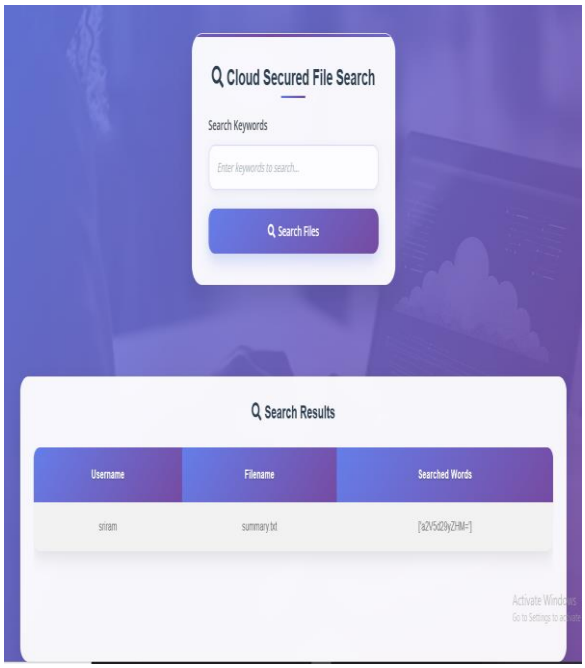
The individual that uploaded the file can get a glimpse of what the file is about on the cloud server.



One of the tasks to manage files dynamically is to select the desired options in the cloud server.



The system verifies that there is correct data in the file that is stored in the cloud server.



Depending on the keyword typed in, the system will search files stored in the cloud platform.

The x-axis in the screen below indicates the number of requests whereas the y-axis indicates the time of accuracy overhead.

5. CONCLUSION

Data deduce security, fast integrity checking and dynamic management of data are all what DIADD excels in cloud storage environment. The homomorphic encryption and a multi-set hash function are used in the system to allow data owners to send encrypted files to other individuals without causing unnecessary storage with the use of proving ownership methods. The real content and authenticators are uploaded by the first person to which the data belongs. Then, the users are verified without the need to re-enter the same information. This is of big difference in regards to the amount of data and bandwidth utilized.

The TPA helps to maintain the security of the outsourced data and ensures that files stored are correct with the help of block-level authentication. In this manner alterations do not need to reprocess the entire file. Adding, deleting and changing blocks are dynamic and not allowed to compromise integrity and security. This will make the system more practical in the actual sense.

There has also been a safe trapdoor based search facility which allows the users to complete a keyword based search on encrypted data without exposing their personal data to the cloud. This ensures privacy and that the info could be utilized. Take deduplication, integrity testing, data dynamics support, and safe search, and combine them and you have a robust, scalable, and confidential solution to the issues that arise with the new cloud storage.

REFERENCES

[1] Peng, X., Shen, W., Yang, Y., & Zhang, X. (2025). Secure Deduplication and Cloud Storage Auditing with Efficient Dynamic Ownership Management and Data Dynamics. IEEE Transactions on Network and Service Management.

[2] Shen, W., Su, Y., & Hao, R. (2020). Lightweight cloud storage auditing with deduplication supporting strong privacy protection. IEEE Access, 8, 44359-44372.

[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] Liu, X., Sun, W., Lou, W., Pei, Q., & Zhang, Y. (2017, May). One-tag checker: Message-locked integrity auditing on encrypted cloud deduplication storage. In IEEE INFOCOM 2017-IEEE conference on computer communications (pp. 1-9). IEEE.

[5] Viswanath G., Krishna Prasad K ., Dr. J Maha Lakshmi., Dr.G.Swapna (2024). Health Prediction Using Machine Learning with Drive HQ Cloud Security. Frontiers in HealthInformatics, 13(8), 2755-2761, https://doi.org/10.5281/zenodo.19128870.

[6] Ren, K., Wang, C., & Wang, Q. (2012). Security challenges for the public cloud. *IEEE Internet Computing, 16*(1), 69–73. https://doi.org/10.1109/MIC.2012.14

[7] Lakshmi, J. M., Prasad, K. K., & Viswanath, G. (2025). Proactive Security in Multi-Cloud Environments: A Blockchain Integrated Real-Time Anomaly Detection and Mitigation Framework. Cuestiones De Fisioterapia, 54(2), 392-417. [8] He, D., Zeadally, S., & Wu, L. (2018). Certificateless public auditing scheme for cloud-assisted wireless body area networks. *IEEE Systems Journal, 12*(1), 64–73. https://doi.org/10.1109/JSYST.2015.2434879

[9] Wang, L., Guan, Z., Chen, Z., & Hu, M. (2023). Enabling integrity and compliance auditing in blockchain-based GDPR-compliant data management. *IEEE Internet of Things Journal, 10*(23), 20955–20968. https://doi.org/10.1109/JIOT.2023.3285350

[10] Tian, Y., Tan, H., Shen, J., Pandi, V., Gupta, B. B., & Arya, V. (2023). Efficient identity-based multi-copy data sharing auditing scheme with decentralized trust management. *Information Sciences, 644*, Article 119255. https://doi.org/10.1016/j.ins.2023.119255

[11] Gudditti, V., & Krishna, P. V. (2021). Adaptive Light Weight Encryption Algorithm for Securing Multi-Cloud Storage. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(9), 545–554

[12] Yang, Y., Chen, Y., & Chen, F. (2021). A compressive integrity auditing protocol for secure cloud storage. *IEEE/ACM Transactions on Networking, 29*(3), 1197–1209. https://doi.org/10.1109/TNET.2021.3063930

[13] John, G. (2010). Digital universe decade—Are you ready? *IDC Report*. Retrieved from https://idcdocserv.com/925

[14] Douceur, J. R., Adya, A., Bolosky, W. J., Simon, P., & Theimer, M. (2002). Reclaiming space from duplicate files in a serverless distributed file system. In *Proceedings of the 22nd International Conference on Distributed Computing Systems* (pp. 617–624). https://doi.org/10.1109/ICDCS.2002.1022314

[15] Song, M., Hua, Z., Zheng, Y., Huang, H., & Jia, X. (2023). Blockchain-based deduplication and integrity auditing over encrypted cloud storage. *IEEE Transactions on Dependable and Secure Computing, 20*(6), 4928–4945. https://doi.org/10.1109/TDSC.2022.3207069

[16] Liu, X., Sun, W., Lou, W., Pei, Q., & Zhang, Y. (2017). One-tag checker: Message-locked integrity auditing on encrypted cloud deduplication storage. In *Proceedings of the IEEE Conference on Computer Communications* (pp. 1–9). https://doi.org/10.1109/INFOCOM.2017.8057069

[17] Tian, G. (2022). Blockchain-based secure deduplication and shared auditing in decentralized storage. *IEEE Transactions on Dependable and Secure Computing, 19*(6), 3941–3954. https://doi.org/10.1109/TDSC.2021.3053463

[18] Dr, K, Pushpa Latha., Mr, M, N, Mallikarjuna Reddy., Dr, B, Rajalingam., Malleswari Akurati., Dr, G, Swapna., Bakkala Santha Kumar., (2026). Blockchain-Enabled Trade Finance Framework for Secure Drug Supply Chain Transactions., International Journal of Drug Delivery Technology, 16(3s), 884-889.

[19] 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.2019.02.016

[20] Ma, X., Yang, W., Zhu, Y., & Bai, Z. (2022). A secure and efficient data deduplication scheme with dynamic ownership management in cloud computing. In *Proceedings of the IEEE International Performance, Computing, and Communications Conference (IPCCC)* (pp. 194–201). https://doi.org/10.1109/IPCCC55827.2022.00040