10.48047/jocaaa.2024.33.02.26

A Research based on Safeguarding Data Storage Security in Cloud Computing: Challenges and Solutions

Varun Kumar Singh
Ph.D. Research Scholar
Department of Computer Science
C. S. J. M. University, Kanpur
varunksingh07@gmail.com

Dr. Rashi Agarwal
Faculty of Computer Science
Department of Computer Science
C. S. J. M. University, Kanpur

Dr. Renu Jain
Head of the Department
Department of Computer Science, C. S. J. M. University, Kanpur

***Corresponding Author:** Varun Kumar Singh, Ph.D. Research Scholar**,** Department of Computer Science**,** C. S. J. M. University, Kanpur**,** varunksingh07@gmail.com

Abstract

Cloud computing has gained popularity, allowing customers to outsource processing and storage to public suppliers. However, this presents new challenges in data security and accuracy. To address these issues, numerous data integrity models and methodologies have been developed. Most research focuses on mitigating the dependency on Cloud Service Providers (CSPs) for data analysis. Proposed models focus on maintaining log data integrity without compromising system efficiency. They use a Minimized Deterministic Finite Automata (mDFA)-based encrypted search mechanism to protect data privacy and protect against Keyword Guessing Attacks. Additionally, a model for decentralized and transparent agri-transactions and a distributed power-aware secured scheme for cloud-oriented IoT are proposed. This research aims to maintain data integrity in cloud computing using Propagated Chain of Log Blocks and Hybrid Vector Committed BST. It proposes lightweight multikey hybrid storage structures, a systematic encrypted search model, efficient keystamp generation algorithm, and an efficient Agri-Transactions application. Future work will integrate SDN and IoT devices for secure cloud services. Limitations include space and time requirements.

Keywords: Cloud Computing, Security, IoT, Artificial Intelligence, System Architecture.

Introduction

Cloud computing is a model that enables ubiquitous, convenient, and on-demand network access to a shared pool of configurable computing resources, such as networks, servers, storage, applications, and services. It is made possible through a managed data center and server, such as the Cloud Service Provider (CSP), which makes the resource available through subscription or on-demand provisioning models [1].

10.48047/jocaaa.2024.33.02.26

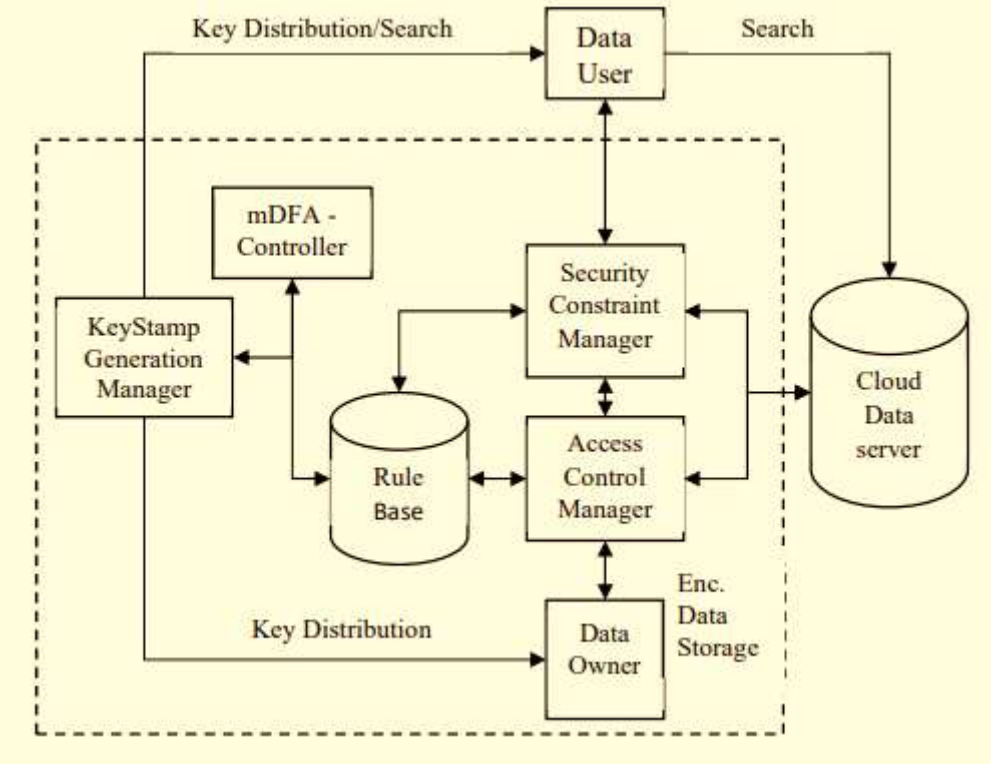

Fig.-1 Architecture model of search over encrypted cloud

some essential characteristics of cloud computing include wide access to networks, pooling of resources, elasticity, on-demand and self-serviced, accountable, and measured [2]. Cloud computing models can be categorized into service and deployment models, such as Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), Private Cloud, Community Cloud, Public Cloud, and Hybrid Cloud [3]. Cloud computing has various applications, including data storage, backup and recovery, big data analysis, testing and development, education, agriculture, and more [4]. Data storage allows businesses to store and retrieve data using various devices, while backup and recovery provide secure backups for data and resources. Big data analysis allows enterprises to store and analyze big data for priceless business insights due to its limitless storage capacity [5]. Testing and development applications offer a scalable and adaptable solution for product creation, testing, and deployment [6].

In the education sector, cloud computing applications have led to significant developments such as e-learning, online distance learning programs, and student information portals [7]. In agriculture, cloud computing helps facilitate the integration of IoT devices and capture sensor data from every device, making it easier to manage and optimize resources. In conclusion, cloud computing offers numerous benefits, including wide access to networks, pooling of resources, elasticity, on-demand and self-serviced, accountable, and measured service models [8]. By leveraging cloud computing, businesses can benefit from cost-effective, efficient, and secure solutions for their data storage, backup, and development needs [9].

Literature Survey

This chapter examines various literature on security and log data in the cloud, encrypted search techniques in the cloud, agricultural applications of the cloud, and security and feasibility of IoT in cloud [10]. The literature is categorized into four major categories: infrastructure as a service model for cloud management, encrypted search techniques in the

10.48047/jocaaa.2024.33.02.26

cloud, agricultural applications of the cloud, and security and feasibility of IoT in cloud [11].

Dykstra and Sherman (2011) justified using the infrastructure as a service model for cloud management, which was complex and extenuated users' reliance on the cloud service provider. Birk & Wegener (2011) suggested using the Application Programming Interface (API) to acquire log data, but this solution faced concerns about reliance on the cloud service provider. Marty (2011) proposed retrieving log data through a business-oriented logging framework through a SaaS architecture, but the outer delivery layer was a serious security concern [12].

Alex and Kishore (2017) highlighted challenges faced by data investigators in the cloud, implementing a centralized forensic layer and a data monitoring plane. Dykstra and Sherman (2011) simplified reliance on the Cloud Service Provider (CSP) through a 6-layer trust-enabled model.

Wang et al. (2021) studied and implemented a unified, distributed, and secure system with deep learning enabled frameworks, but their model was less efficient due to sample attack scenarios. Ray et al. (2013) used a reliable delivery mechanism for user log records, but their model lacks tight coupling with the operating system-based logging methodology.

Khan et al. (2017) introduced secure log as a service using reversible watermarking (SecLaaS-RW) scheme for data content authentication.

Cloud log management solutions face challenges in implementing hidden metadata within log files. Lemoudden & El (2015) proposed a new methodology for log retention and transfer, but it did not fit into an automated category. Pichan et al. (2018) proposed a cloud forensic logging framework, CFLog, but its speeding up was unrealistic due to the involvement of cloud service providers. Sandikkaya & Harmanci (2012) explored security factors in Platform as a Service clouds, but their plane architecture did not address the complexity of cloud log management [13-15].

Proposed System Architecture

This chapter discusses the architecture of a proposed system that consists of various layers for data processing. The data from the user interface goes through a cloud server and is processed in various layers, including the log integrity module, PCLB, encrypted search, and secured storage module. The data is stored in a cloud database, which can be verified and validated by a valid user or domain-specific forensic expert at times of need [16].

The system's architecture includes log data extraction and classification, which involves extracting log data from client applications hosted on the cloud and processing it through APIs. The nature of log data varies with the application being used, such as security constraints, ownerships, availability, modularity, dependency, and granularity [17].

Log data from agricultural devices and IoT-based sensor data are also extracted and processed. The Predictive System for Agri-Transactions and IoT device logs are inputs to the upper layer, with the latter being analyzed for its efficiency and effectiveness. The priority-based classifier classifies the log data based on the nature of the log, with the k-nearest neighbor algorithm being utilized for better efficacy [18].

10.48047/jocaaa.2024.33.02.26

The data is then forwarded to the Propagated Chain of Log Blocks to ensure data integrity throughout its life cycle. An efficient mDFA-based search over encrypted cloud is used to address search efficiency and probability of success. The system enables authorized users or applications to access the presence of the log data in the cloud without exactly reading the data [19].

Lastly, the system is validated and validated by end users and applications, who can be integrated into the existing log chain and granted legal and constrained access to the data system. Applications like digital and cyber forensics can leverage this research work to identify and process the log data [20].

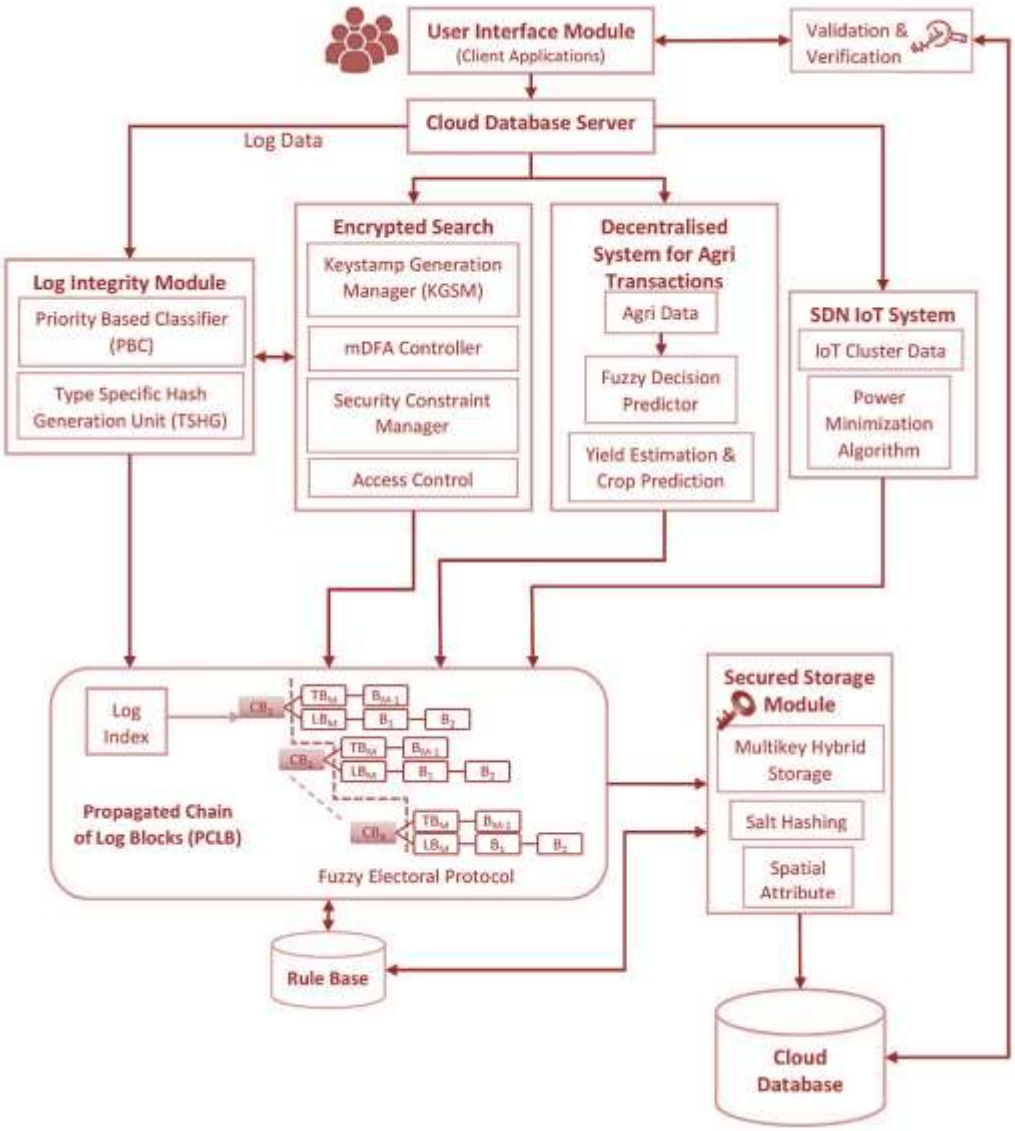

Fig.-2 System Architecture

10.48047/jocaaa.2024.33.02.26

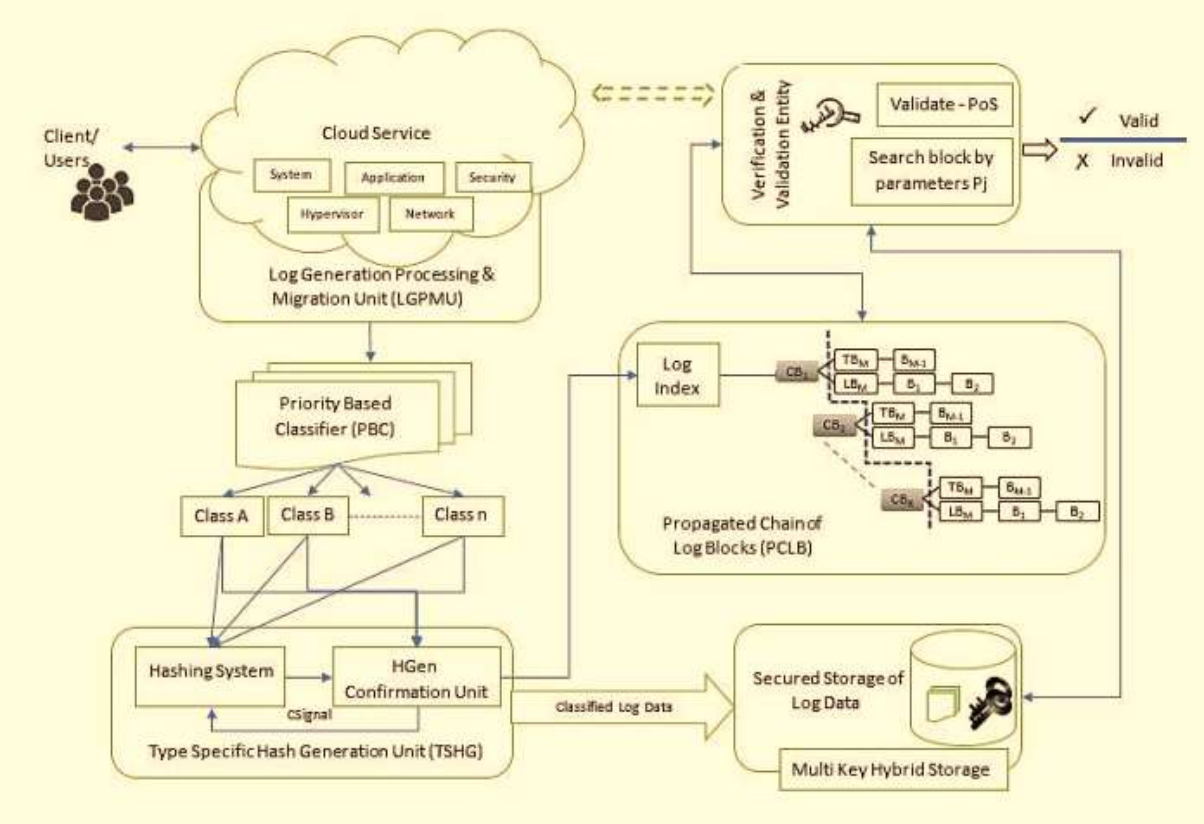

Fig.-3 System Architecture of Preserving Log Data Integrity

Design and Analysis

Table-1 Dataset Details

Dataset Detail	Category	Number of instances	Number of Attributes	Number of classes
TShark Archive of IoT	{1, 6}	477426	13	5
Judge Server Archive	{6, 8}	16007	3	10
CIDDS logs	{2, 4, 8}	172838	12	3
Apache Server	{5, 7}	1386	5	4
IEEE access logs	{2, 3, 4}	45156	9	3
Machine Access logs of Endpoint Security	{1, 2, 8}	18912	4	20
Incident server logs	{2, 5}	28856	17	2

10.48047/jocaaa.2024.33.02.26

Table-2 Success % of lazy and eager classification methodologies on our dataset using K-NN

Dataset Detail	Success % in Lazy Methodology Classification	Success % in Eager Methodology Classification
TShark Archive of IoT	97.30	98.91
Judge Server Archive	79.21	78.73
CIDDS logs	88.10	86.72
Apache Server	89.36	88.5
IEEE access logs	98.69	97.76
Machine Access logs of Endpoint Security	91.56	87.6
Incident server logs	95.25	94.49

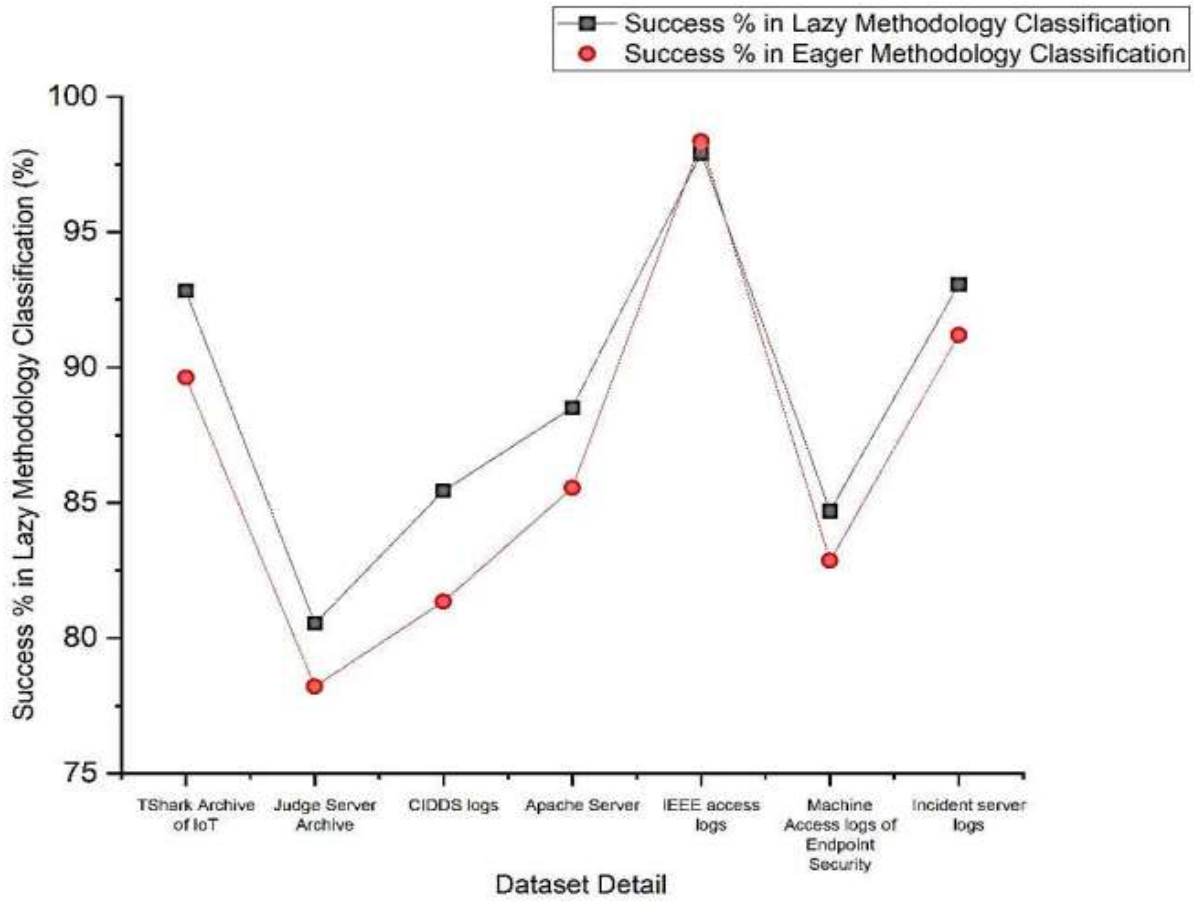

Fig.-4 Success % of lazy and eager classification methodologies using Naïve Bayes

10.48047/jocaaa.2024.33.02.26

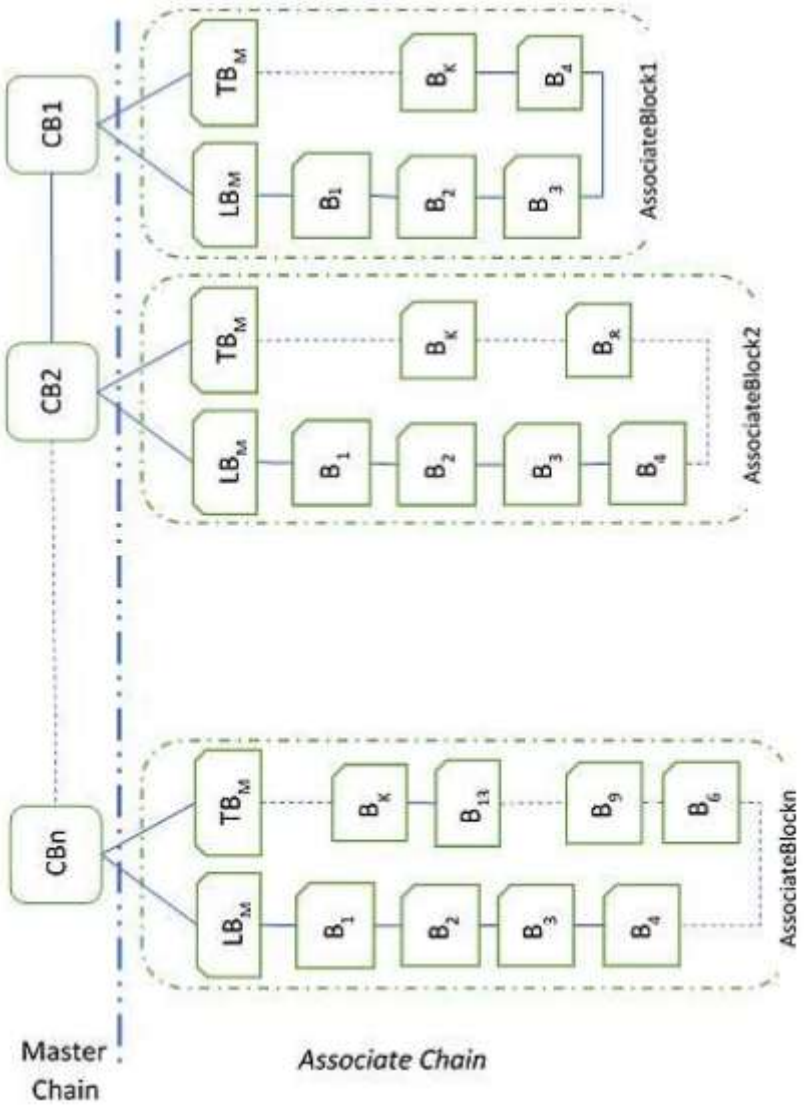

Fig.-5 PCLB architecture

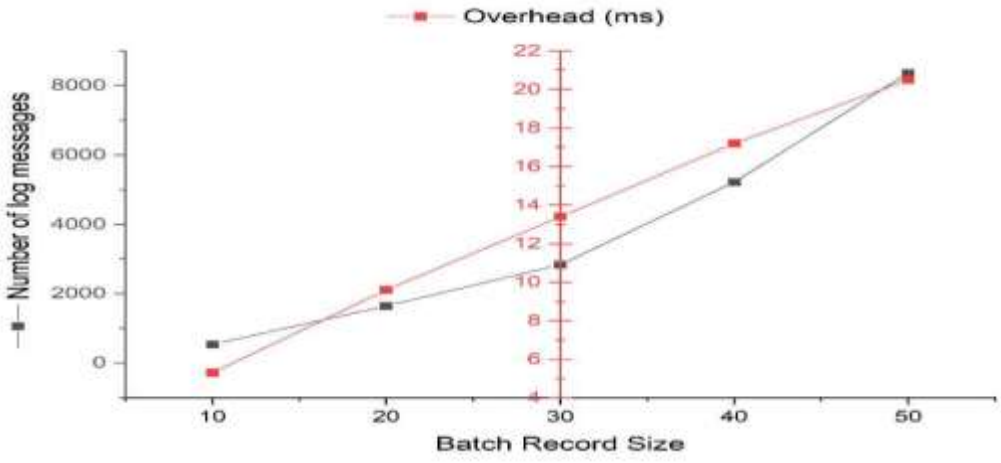

Fig.-6 Overhead with no enabled security constraints

10.48047/jocaaa.2024.33.02.26

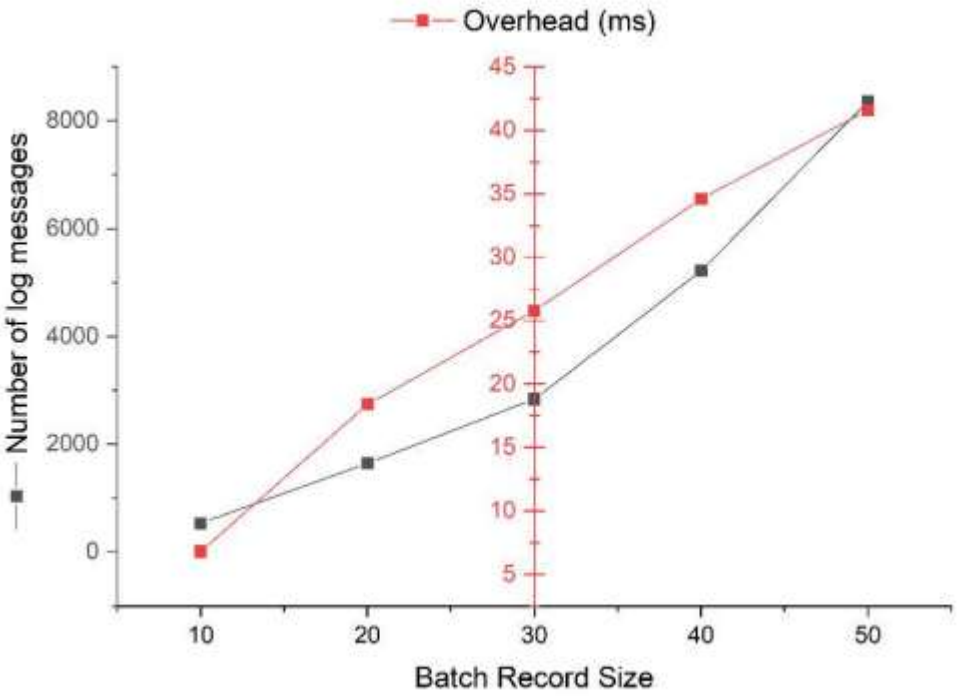

Fig.-7 Overhead with partially enabled security constraints

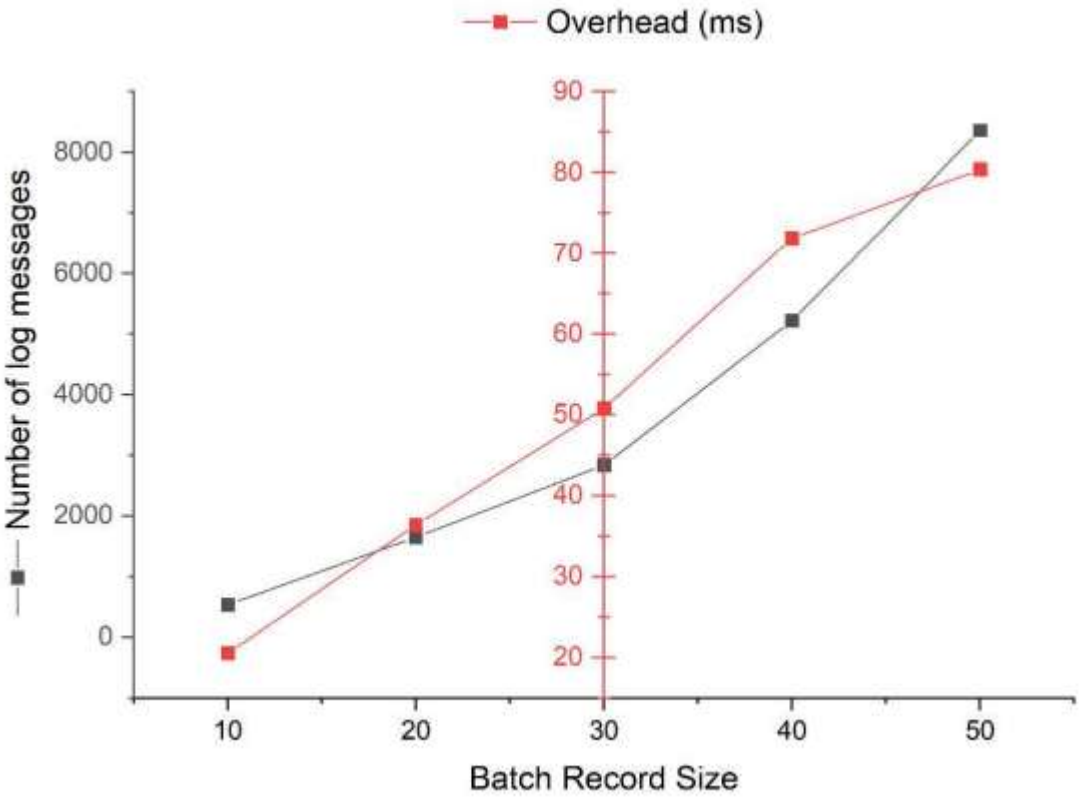

Fig.-8 Overhead with fully enabled security constraints

10.48047/jocaaa.2024.33.02.26

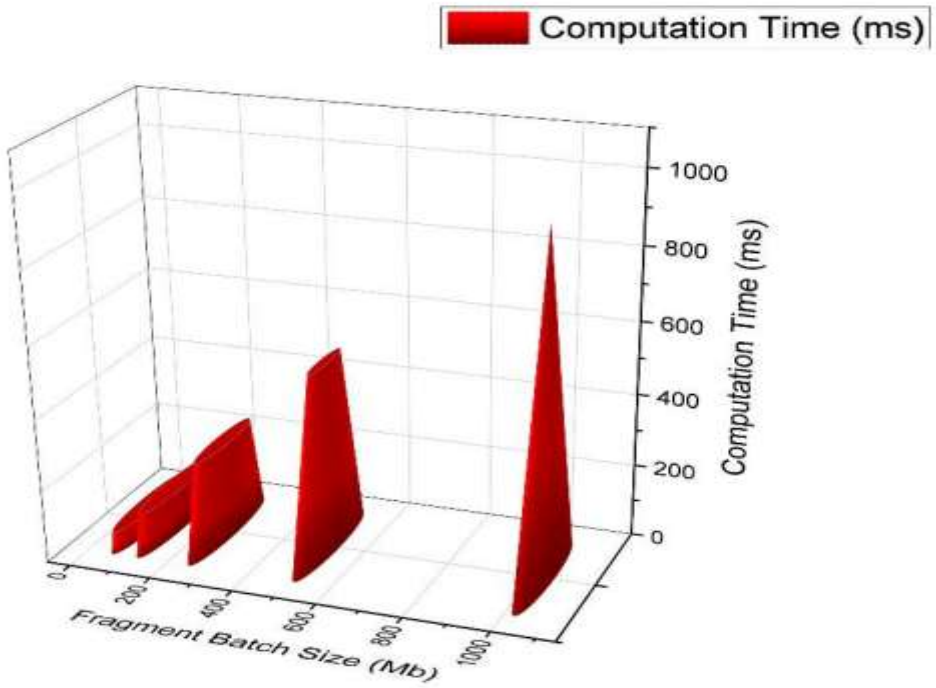

Fig.9 Computation Time for Batch Fragmentation

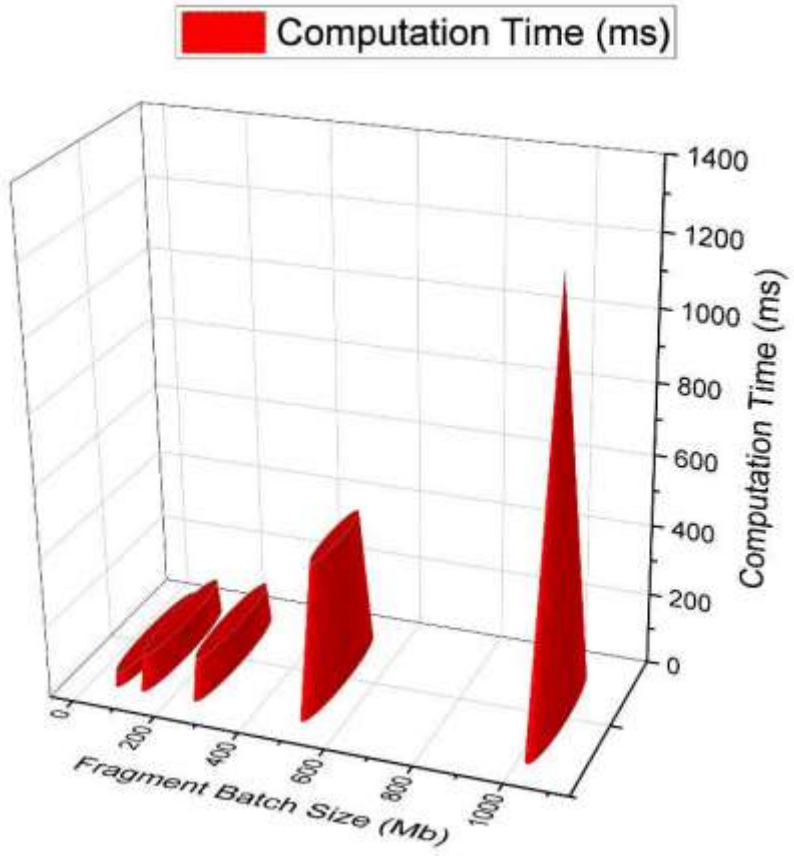

Fig.10 Computation Time for Batch Defragmentation

10.48047/jocaaa.2024.33.02.26

This study aids in the creation of a framework for preserving the accuracy of the log data. The log generation processing and migration unit first collects the data, which is then forwarded to the classification unit for additional user priority-based classification. The propagated chain of log blocks and the Hybrid Vector Committed BST (HVCBST) techniques preserve the integrity of log data. Additionally, lightweight multikey hybrid storage structures are used to secure the log data. Ethereum and self-coded platforms are used for the implementation, and the outcomes are verified and compared. When compared to the current frameworks and methodologies, the outcomes of the implemented systems have shown themselves to be significantly more effective and impenetrable.

Conclusion

Cloud computing has gained popularity due to its on-demand allocation and sharing among users. However, cybercriminals have been successful in using it for fraudulent operations due to its fundamental features and lack of appropriate safe techniques. This research focuses on maintaining data integrity in the cloud using Propagated Chain of Log Blocks and Hybrid Vector Committed BST. Lightweight multikey hybrid storage structures are used for secure storage, implemented through Ethereum. A systematic model for encrypted search is proposed to preserve cloud data privacy through mDFA search scheme. An efficient keystamp generation algorithm and regular encryption, decryption, and key exchange processes are also implemented. An efficient Agri-Transactions application is designed using blockchain and machine learning models. The next work focuses on integrating SDN and IoT enabled devices for secured cloud services. Limitations include increasing space and time required for hybrid vector committed BST, improving classification of log records, and testing IoT systems in private blockchains and fog edge platforms.

References

1. Alex, ME & K ishore, R 2017, ‘Forensics framework for cloud computing’, Computers & Electrical Engineering, vol. 60, pp. 193-205.
2. Bai, M, Jiang, S, Zhang, X & Wang, X 2022, ‘An efficient skyline query algorithm in the distributed environment’, Journal of Computational Science, vol. 58, p. 101524.
3. Balakumar, S & K avitha, AR 2021, ‘Quorum-based blockchain network with IPFS to improve data security in IoT network’, Studies in Informatics and Control, vol. 30, no. 3, pp. 85-98.
4. Boneh, D, Crescenzo, GD, Ostrovsky, R & Persiano, G 2004, ‘Public key encryption with keyword search’, In International conference on the theory and applications of cryptographic techniques, Springer, Berlin, Heidelberg, pp. 506-522.
5. Caro, MP, Ali, MS, V ecchio, M & Giaffreda, R 2018, ‘Blockchainbased traceability in Agri-Food supply chain management: A practical implementation’, In IoT V ertical and Topical Summit on AgricultureTuscany (IOT Tuscany), pp. 1-4.
6. Chen, R, Mu, Y , Y ang, G, Guo, F & Wang, X 2015, ‘Dual-server public-key encryption with keyword search for secure cloud storage’, IEEE transactions on information forensics and security, vol. 11, no. 4, pp. 789-798.

10.48047/jocaaa.2024.33.02.26

7. Dong, X , Cao, Z & Shen, J 2019, 'Revocable Public K ey Encryption with Authorized K eyword Search', In IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 857-860.
8. Galvez, JF, Mejuto, JC & Simal-Gandara, J 2018, 'Future challenges on the use of blockchain for food traceability analysis', TrAC Trends in Analytical Chemistry, vol. 107, pp. 222-232.
9. Geetha, K & K annan, A 2019, 'An efficient information system for providing location based services in network environments', Wireless Personal Communications, vol. 109, no. 4, pp. 2377-2398.
10. Islam, SJ, Chaudhury, ZH & Islam, S 2019, 'A simple and secured cryptography system of cloud computing', In IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pp. 1-3.
11. K han, A, Y aqoob, A, Sarwar, K , Tahir, M & Ahmed, M 2017, 'Secure logging as a service using reversible watermarking', Procedia Computer Science, vol. 110, pp. 336-343.
12. K im, M, Hilton, B, Burks, Z & Reyes, J 2018, 'Integrating blockchain, smart contract-tokens, and IoT to design a food traceability solution', In 2018 IEEE 9th annual information technology, electronics and mobile communication conference (IEMCON), pp. 335-340.
13. K shetri, N 2019, 'Blockchain and the economics of food safety', It Professional, vol. 21, no. 3, pp. 63-66.
14. Lei, K , Du, M, Huang, J & Jin, T 2020, 'Groupchain: Towards a scalable public blockchain in fog computing of IoT services computing', IEEE Transactions on Services Computing, vol. 13, no. 2, pp. 252-262.
15. Liu, Q, Tan, CC, Wu, J & Wang, G 2011, 'Reliable re-encryption in unreliable clouds', In IEEE Global Telecommunications ConferenceGLOBECOM, pp. 1-5.
16. Muthurajkumar, S, Ganapathy, S, V ijayalakshmi, M & K annan, A 2017, 'An intelligent secured and energy efficient routing algorithm for MANETs', Wireless Personal Communications, vol. 96, no. 2, pp. 1753-1769.
17. Osmanoglu, M, Tugrul, B, Dogantuna, T & Bostanci, E 2020, 'An effective yield estimation system based on blockchain technology', IEEE Transactions on Engineering Management, vol. 67, no. 4, pp. 1157-1168.
18. Pichan, A, Lazarescu, M & Soh, ST 2018, 'Towards a practical cloud forensics logging framework', Journal of information security and applications, vol. 42, pp. 18-28.
19. Ray, I, Belyaev, K , Strizhov, M, Mulamba, D & Rajaram, M 2013, 'Secure logging as a service— delegating log management to the cloud', IEEE systems journal, vol. 7, no. 2, pp. 323-334.
20. Selvi, M, Thangaramya, K , Ganapathy, S, K ulothungan, K, K hannah Nehemiah, H & K annan, A 2019, 'An energy aware trust based secure routing algorithm for effective communication in wireless sensor networks', Wireless Personal Communications, vol. 105, no. 4, pp. 1475-1490.