Enhancing Network Security In Hadoop-Based Secure Storage Solution For Multi-Cloud Authentication

Ms.T.Kalaiselvi[1], Mr.R. Pavin Rohith[2], Mr.T. C. Stalin[3], Mr.S.V. Nivethan[4]
[1] Associate Professor, Dept of Computer Science and Engineering
[2, 3, 4]Dept of Computer Science and Engineering
[1, 2, 3, 4] Erode Sengunthar Engineering College (Autonomous) , Thudupathi, Erode, Tamilnadu, India

Abstract- *Cloud users can verify data integrity without retrieving the entire file through Provable Data Possession (PDP) schemes that rely on Public Key Infrastructure (PKI). These schemes are efficient, flexible, and support private, delegated, and public verification. However, ID-DPDP is flawed and fails to achieve soundness. To fix this flaw, a generic construction is presented, resulting in a new ID-DPDP protocol that extends the basic ID-PDP to a multiple clouds environment. With data storage and sharing services in the cloud, users can easily modify and share data as a group. To ensure shared data integrity can be verified publicly, users in the group need to compute signatures on all the blocks in shared data. Different blocks in shared data are generally signed by different users due to data modifications performed by different users. When a user is revoked from the group, the blocks previously signed by this user must be re-signed by an existing user. However, the straightforward method of downloading the corresponding part of shared data and re-signing it during user revocation is inefficient due to the large size of shared data in the cloud.*

Keywords- Big Data Security, Data Encryption, Parallel Encrypted Storage

I. INTRODUCTION

The landscape of cloud computing is constantly evolving, and with the rise of big data, organizations are faced with both challenges and opportunities. To effectively process and analyse large-scale data, secure storage solutions based on Hadoop have become essential. These solutions address the critical need for scalable, reliable, and secure storage of vast data sets. Hadoop, an open-source distributed computing framework, provides the infrastructure for handling enormous data volumes, but its true potential can only be realized with a robust storage system in place. In the realm of cloud computing, where data security and privacy are of utmost importance, the integration of Hadoop with advanced security measures offers a comprehensive approach to managing big data while mitigating potential risks. A Hadoop-based secure storage solution for big data in the cloud encompasses a wide range of technologies and strategies. It involves the use of distributed file systems like Hadoop Distributed File System (HDFS) to efficiently store and manage data across a cluster of machines. Additionally, these solutions incorporate encryption techniques to protect data at rest and in transit, access control mechanisms to prevent unauthorized access, and monitoring and auditing capabilities for compliance and threat detection. As the demand for big data storage in the cloud continues to grow, organizations must prioritize the deployment of secure storage solutions to ensure the confidentiality, integrity, and availability of their data, while also harnessing the full potential of their big data analytics initiatives.

II. LITERATURE REVIEW

In this paper [1] r, Gang Li et.al. have put forth the objective of achieving a distributed, collaborative, and automated design and manufacturing workflow. To achieve this, they have leveraged the concept of Industry 4.0 and its associated technologies such as Cyber Physical System, Internet of Things, Cloud Computing, and Big Data Analytics. The cyber-physical system and internet of things enable the collection and transfer of industrial data through a fusion of peripherals such as software, sensors, and electronics. Cloud computing techniques aid in centralized data storage and offer a platform for collaboration to expedite and refine resource allocation and research for entire industry gains. Big data analytics has also been exploited to organize these digital assets and extract valuable insights from them. This has attracted significant attention from both industry and academia. In this special issue, the manifold relationship between Industry 4.0 and Big Data is investigated by bringing together active researchers from related fields.

Chunhui Wen [2] et.al. has proposed in their paper the successful utilization of the big data technology framework in the Internet of Things. The financial industry also aims to leverage the advanced technology of big data to integrate and enhance the internal and external data pertaining to credit risks. By relying on more efficient machine learning algorithms, a reasonable prediction of credit risk can be obtained, thereby reducing self-generated losses in the Internet

of Things finance and increasing profits. This article employs distributed search engine technology to customize web crawlers for acquiring the necessary bank card and transaction data from the diverse sources of data in the Internet of Things financial industry. It further designs a corresponding Spark parallel algorithm to preprocess the data and establishes an inverted table and two-level index file to serve as a data source for big data analysis platforms. Once the data source is determined, the Mutually Exclusive Collectively Exhaustive (MECE) analysis method is combined with the expertise of numerous financial business experts in the industry to derive a set of candidate indicators and quantification methods for evaluating the financial credit risk in the Internet of Things. Additionally, the correlation between these indicators and risk grading is analyzed.

The authors, Xiangfan Zhang [3] et.al., have proposed a system that aims to enhance the intelligence of the medical system. This paper presents the design and implementation of a secure medical big data ecosystem on top of the Hadoop big data platform. The system is developed in response to the growing concern over the security of medical big data ecosystems. The paper also introduces a personalized health information system that enables patients to access their treatment and rehabilitation status anytime and anywhere. The system ensures that all medical health data is stored independently, even if it is distributed across different independent medical institutions. The paper also explores the potential of blockchain as a distributed accounting technology for multi-party maintenance and backup information security. The system realizes the personal health datacenter on the Hadoop big data platform, and the original distributed data is stored and analyzed centrally through the data synchronization module and the independent data acquisition system.

Xin Huang [4] et.al. have proposed a system that addresses the challenges posed by the storage and frequent reading of massive electronic medical data. In this system, doctors utilize electronic images instead of traditional film for diagnosis in the context of electronic medical data. Additionally, patients have the convenience of accessing examination images through various electronic means at any time. To enhance storage performance, different merging strategies are suggested based on the characteristics of image files generated by different examination types. Furthermore, a two-level model combined with medical imaging information is proposed, taking into account the characteristics of medical data with examination as the fundamental unit. This model includes an indexing mechanism to enable random access to SEQ files. Considering the time characteristics of data access, an improved 2Q algorithm is introduced to cache optimized and read files in separate cache queues, thereby enhancing file reading efficiency. Experimental comparisons demonstrate that the proposed algorithm outperforms the baseline method in terms of storage and access performance. The advancement of smart medical practices is crucial for the development of smart cities.

The exponential growth of data being processed by computing systems has put a strain on existing technologies to provide scalable, fast, and efficient support. This has led to a shift from data-centric to knowledge-centric computing, but the challenge remains to optimally store and migrate large data sets across data centers. To address this challenge, the main objective is to find a better data storage location that improves overall data placement cost and application performance. In this survey paper, we provide an overview of Cloud-centric Big Data placement and data storage methodologies, focusing on non-functional properties. Our analysis of respective technologies related to Big Data management can guide readers towards selecting the best solutions for their non-functional application requirements. We also highlight current gaps and challenges in this field. Applications have undergone a significant transformation over time, transitioning from batch, compute, or memory-intensive applications to more complex and long-running streaming or interactive applications. This evolution has resulted in the need for frequent access to multiple distributed data sources during application deployment and provisioning.

This survey paper provides an overview of Cloud-centric[5] Big Data placement and data storage methodologies, focusing on non-functional properties. The authors analyze various technologies related to Big Data management and offer guidance for selecting appropriate solutions based on non-functional application requirements. Gaps and challenges in this field are also highlighted.

W. Rajeh explores Hadoop distributed file system security problems and investigates unauthorized access issues in this paper[6]. The study delves into security vulnerabilities and potential threats to Hadoop file systems, offering insights into enhancing security measures to mitigate unauthorized access risks.

G. S. Bhathal and A. Singh discuss weaknesses in the Hadoop framework, security challenges, and potential attacks in this paper[7]. The authors identify vulnerabilities within the Hadoop framework and analyze security challenges that organizations may face when implementing Hadoop-based solutions. Various attack vectors and their implications are also examined.

Kapil, A. Agrawal, and R. A. Khan address big data security challenges from a Hadoop perspective in this paper[8]. The study focuses on identifying security challenges associated with Hadoop-based environments and proposes strategies to mitigate these challenges. The authors offer insights into enhancing security measures to safeguard big data assets in Hadoop ecosystems.

B. H. Husain, S. R. Zeebaree, et al., present a review of improvised distributions framework of Hadoop in this paper[9]. The study examines enhancements and developments in Hadoop distributions to address evolving requirements and challenges. The authors provide insights into the evolution of Hadoop frameworks and their implications for big data management.

M. Naisuty, A. N. Hidayanto, N. C. Harahap, A. Rosyiq, and G. M. S. Hartono conduct a systematic literature review on data protection on Hadoop distributed file systems by applying encryption algorithms in this paper[10]. The study investigates various encryption algorithms and their effectiveness in protecting data stored in Hadoop distributed file systems. The authors analyze encryption techniques and their impact on data security in Hadoop environments.

2.1 TABLE

S.NO	PAPER	METHODOLOGY	DEMERITS	MERITS
1.	Industry 4.0 and big data innovations	Leveraging Industry 4.0 and associated technologies such as Cyber Physical System, Internet of Things, Cloud Computing, and Big Data Analytics. Utilizing cyber-physical systems and IoT for data collection and transfer, cloud computing for centralized storage, and big data analytics for insights extraction.	Increased complexity due to integration of multiple technologies. Potential maintenance overhead. Risk of compatibility issues with existing systems.	Enhanced security and scalability through distributed authentication and data possession. Improved resource allocation and research collaboration in the industry
2.	big data driven internet of things for credit evaluation and early warning in finance	Employing big data technology framework in the Internet of Things (IoT) for credit evaluation and early warning in finance. Utilizing distributed search engine technology, machine learning algorithms, and Spark parallel algorithm for data preprocessing. Exploring Mutually Exclusive Collectively Exhaustive (MECE) analysis method for evaluating financial credit risk in the IoT.	Potential performance issues due to computational demands of machine learning algorithms. Dependence on customized web crawlers may limit data collection.	mproved prediction of credit risk, reduced self-generated losses, and increased profits in IoT finance. Customized web crawlers for data acquisition. Spark parallel algorithm for data preprocessing.
3.	intelligent medical big data system based on Hadoop and blockchain	Developing a secure medical big data ecosystem on the Hadoop platform. Utilizing Hadoop big data platform, personalized health information system, and blockchain for data security.	Complexity in managing independent storage of medical health data. Reliance on blockchain introduces overhead and scalability concerns.	Enhanced security of medical data. Personalized health information system for patients. Potential use of blockchain for multi-party informatio

		Implementing a personalized health information system and exploring blockchain for multi-party maintenance and backup information security.		n security.
4.	Mathematical Problems in Engineering presented a Hadoop-based medical image storage and access method for examination series.	Addressing challenges of storage and frequent reading of massive electronic medical data. Introducing electronic images for diagnosis in electronic medical data. Proposing different merging strategies based on examination types for storage efficiency. Introducing an indexing mechanism and an improved caching algorithm for enhanced file reading efficiency.	Complexity in merging strategies and indexing mechanisms may increase implementation overhead. Reliance on improved caching algorithm may introduce additional computational overhead.	mproved storage and access efficiency for electronic medical data. Convenience for doctors and patients through electronic image usage. Introduction of optimized caching algorithm for enhanced file reading efficiency.
5.	data storage and placement methodologies for cloud big data ecosystem.	Providing an overview of Cloud-centric Big Data placement and data storage methodologies, focusing on non-functional properties. Analyzing various technologies related to Big Data management and offering guidance for selecting appropriate solutions based on non-functional application requirements. Highlighting gaps and challenges in this field.	Lack of specific case studies or experiments for validation of proposed methodologies. Potential biases in technology selection based on author perspectives.	Comprehensive overview of Cloud-centric Big Data placement and storage methodologies. Guidance for selecting appropriate solutions based on non-functional application requirements. Identification of current gaps and challenges in the field.
6.	Hadoop distributed file system security problems and investigation of unauthorized access issue	Investigating Hadoop distributed file system security problems and unauthorized access issues. Delving into security vulnerabilities and potential threats to Hadoop file systems. Offering insights into enhancing security measures to mitigate unauthorized access risks.	Focus primarily on identifying security issues rather than proposing solutions. Potential lack of practical implementation examples or case studies	Enhanced understanding of security vulnerabilities and risks associated with Hadoop file systems. Insights into improving security measures to mitigate unauthorized access risks.
7.	Hadoop framework weaknesses, security challenges, and potential	Analyzing weaknesses in the Hadoop framework, security challenges, and potential	Lack of detailed exploration of mitigation strategies for	Increased awareness of weaknesses and security challenges

	and attacks.	attacks. Identifying vulnerabilities within the Hadoop framework and examining security challenges faced by organizations. Exploring various attack vectors and their implications.	identified weaknesses and vulnerabilities. Potential bias towards highlighting vulnerabilities without offering corresponding solutions.	in the Hadoop framework. Insights into potential attack vectors and their implications for organizations utilizing Hadoop-based solutions.
8,	data security challenges: Hadoop perspective	Addressing big data security challenges from a Hadoop perspective. Identifying security challenges associated with Hadoop-based environments and proposing strategies to mitigate these challenges. Offering insights into enhancing security measures to safeguard big data assets in Hadoop ecosystems.	Potential lack of depth in exploration of specific security challenges and mitigation strategies. Reliance on theoretical approaches without practical implementation examples.	Enhanced understanding of security challenges in Hadoop ecosystems. Insights into strategies for mitigating security risks and safeguarding big data assets
9.	Improvised distributions framework of hadoop	Reviewing improvised distributions framework of Hadoop. Examining enhancements and developments in Hadoop distributions to address evolving requirements and challenges. Providing insights into the evolution of Hadoop frameworks and their implications for big data management.	Lack of in-depth analysis or case studies to illustrate practical applications of improvised distributions framework. Potential bias towards focusing on theoretical aspects rather than practical implications.	Comprehensive review of improvised distributions framework of Hadoop. Insights into enhancements and developments in Hadoop distributions and their implications for big data management
10.	Data protection on Hadoop distributed file system by applying encryption algorithms	Conducting a systematic literature review on data protection on Hadoop distributed file systems using encryption algorithms. Investigating various encryption techniques and their effectiveness in protecting data stored in Hadoop distributed file systems. Analyzing encryption techniques and their impact on data security in Hadoop environments	Limited exploration of encryption algorithms and their comparative analysis in terms of performance and security. Potential bias towards highlighting positive aspects of encryption without considering drawbacks.	Comprehensive review of encryption techniques for protecting data in Hadoop distributed file systems. Insights into the impact of encryption on data security in Hadoop environments

III. EXISTING SYSTEM

To overcome the challenges posed by a single encryption algorithm, including low encryption efficiency and unreliable metadata for static data storage in cloud computing

environments, we suggest a secure storage scheme for big data based on Hadoop. Our approach involves dispersing the Name Node service across multiple servers using HDFS federation and HDFS high-availability mechanisms, and coordinating each node using the Zookeeper distributed coordination mechanism to achieve dual-channel storage. We have also enhanced the ECC encryption algorithm for ordinary data encryption and adopted a homomorphic encryption algorithm for data that requires calculation. To speed up the encryption process, we have implemented a dual-thread encryption mode. Finally, we have designed the HDFS control module to integrate the encryption algorithm with the storage model. Our experimental results demonstrate that our proposed solution effectively addresses the single point of failure issue with metadata, improves metadata reliability, and enables server fault tolerance.

3.1 DISADVANTAGES

- Introducing a secure storage scheme built upon Hadoop, incorporating a distributed Name Node service, coordination facilitated by Zookeeper, and the incorporation of diverse encryption algorithms, introduces notable complexity to the system. This complexity amplifies maintenance demands and necessitates specialized expertise for effective deployment and management.
- Employing multiple encryption algorithms, particularly homomorphic encryption for data necessitating calculations, could result in substantial consumption of computational resources. This may elevate processing overhead and potentially affect the performance of the storage system, particularly in environments constrained by limited computational capabilities.
- The implementation of a dual-thread encryption mode and the integration of encryption algorithms with the storage model have the potential to introduce supplementary latency into the data storage and retrieval process. Consequently, this may lead to delayed response times for data access operations, posing challenges for applications that necessitate real-time or low-latency access to data.
- The introduction of bespoke encryption algorithms and alterations to the Hadoop storage model could give rise to compatibility and interoperability issues. This might restrict the system's capability to seamlessly integrate with established tools, frameworks, or third-party services that rely on conventional Hadoop implementations..

IV. PROPOSED SYSTEM

CP-HABE, also known as "Hierarchal Attribute-Based Distributed Provable Data Possession," is an innovative authentication protocol that aims to tackle the ever-evolving security challenges faced by large-scale network systems. This protocol introduces a unique dual scheme approach, where each subgroup is independently managed by a trusted group security intermediary. Essentially, it treats each subgroup as a separate multi-cloud group, alleviating the burden of concentrating the workload on a single entity. The significance of this lies in its ability to distribute the authentication and data possession processes across multiple entities, thereby enhancing both security and scalability. By leveraging hierarchical attribute-based access control, CP-HABE not only ensures secure data access but also simplifies the management of permissions and access policies within complex network systems. This ground-breaking protocol paves the way for secure, efficient, and distributed authentication and data possession in the era of multi-cloud and large-scale network environments.

4.1 ADVANTAGES

- CP-HABE employs a unique approach by splitting authentication and data possession tasks among several trusted entities. This helps minimize the vulnerability of relying on a single point and boosts the overall security of the system.
- CP-HABE treats every subgroup as its own multi-cloud group, spreading out the workload and easing the strain on any one entity. This makes the protocol highly adaptable, which is especially important in expansive network systems where user numbers and data amounts can grow rapidly.
- In CP-HABE, authentication tasks are spread out among several entities, which helps to avoid congestion and enhances the performance of the system. This means authentication runs smoothly even in networks that are widely spread out and constantly changing.
- Hierarchical attribute-based access control makes it simpler for administrators to handle permissions and access rules in intricate network setups. This ease of management allows administrators to establish and enforce access controls more effectively, resulting in an improved overall security stance.
- CP-HABE's hierarchical attribute-based method allows for tailored access policies based on user attributes. This ensures precise control over who can access specific data, guaranteeing that users only access information they're authorized to view.

4.2 HADOOP

Hadoop stands as a pioneering open-source framework that has revolutionized the realm of big data processing and storage. Renowned for its distributed computing architecture, Hadoop empowers organizations to effectively manage and analyze extensive datasets across clusters of commodity hardware. Through the utilization of Hadoop's scalability and fault tolerance features, businesses can extract invaluable insights from diverse data sources, thereby facilitating informed decision-making and fostering innovation. The widespread adoption of Hadoop underscores its indispensable role in addressing the formidable challenges presented by the exponential proliferation of data in contemporary digital landscapes.

4.3 CP-HABE

Cypher text Hierarchical Attribute-Based Encryption (CP-HABE) represents a cutting-edge cryptographic technique poised to revolutionize data security in modern computing environments. This innovative approach offers a robust solution for enforcing access control policies based on user attributes within hierarchical structures. By leveraging CP-HABE, organizations can ensure secure and efficient data sharing while maintaining granular control over access permissions. This technology empowers administrators to define complex access policies that encompass multiple levels of hierarchical attributes, thereby safeguarding sensitive information from unauthorized access. CP-HABE's ability to seamlessly integrate with existing systems makes it a versatile solution for protecting data in diverse applications, ranging from cloud computing to IoT ecosystems. As the digital landscape continues to evolve, CP-HABE emerges as a cornerstone in the quest for robust and flexible data security solutions.

4.4 BIG DATA

Big data represents a transformative force reshaping industries and revolutionizing how organizations harness and analyze vast volumes of information. Characterized by the 'three Vs'—volume, velocity, and variety—big data encompasses datasets of unprecedented size, generated at high speeds and spanning diverse sources and formats. This deluge of data presents both challenges and opportunities for enterprises seeking to extract actionable insights and drive informed decision-making. Leveraging advanced analytics tools and technologies, organizations can unlock the potential of big data to uncover hidden patterns, trends, and correlations that traditional methods might overlook. From predictive analytics and machine learning to real-time data processing and sentiment analysis, big data empowers businesses to gain a competitive edge, optimize operations, enhance customer experiences, and fuel innovation. However, harnessing the full potential of big data requires robust data management strategies, scalable infrastructure, and skilled data professionals capable of navigating the complexities of large-scale data ecosystems. As big data continues to evolve and proliferate, organizations must adapt and embrace data-driven approaches to thrive in an increasingly interconnected and data-rich world.

V. MODULE DESCRIPTION

5.1 Secure Multi-Cloud User Registration and Authentication Protocols

The initial user provided their username, password, and selected a group ID to register with the Data Cloud Server. This user was subsequently added to the specified group. Afterwards, they entered their username, password, and selected their group ID to log in.

5.2 Efficient Data Encryption and Key Management Strategies for Multi-Cloud Environments

Within the Key Generation module, each member of the group is responsible for generating their own public and private keys. By generating a random key, the user is able to output both their public and private keys. For the sake of simplicity, let us assume that user u1 is the original creator of the shared data and is therefore the first user in the group. Additionally, the original user is responsible for creating a public user list (UL) that includes the IDs of all users in the group. This list is then signed by the original user and made public.

5.3 Distributed File Storage Architecture for Multi-Cloud Servers

In order to upload a file, the user has divided it into multiple blocks. Each block has been encrypted using the public key and a signature has been generated for authentication purposes. The user has then proceeded to upload each block's cipher text along with its signature, block ID, and signer ID. All of this metadata and key details have been stored in the Public Verifier for public auditing.

5.4 Streamlined File Retrieval Mechanisms from Multi-Cloud Storage Systems

The subsequent user or member of the group intends to retrieve a file. Therefore, the user provides the filename and

receives the confidential key. Subsequently, the user inputs this confidential key. If the confidential key is valid, the user can successfully decipher the downloaded file. However, if the next user enters an incorrect confidential key, user1 will be blocked by the Public Verifier. In the event that the confidential key is valid, each block is decrypted and the signature is verified. If both signatures are identical, all blocks are merged to obtain the original file.

5.5 Privacy-Preserving Public Auditing Techniques with User Collision Prevention

The Public Verifier method involves blocking a User who enters the wrong secret key. The blocked User is then added to the Public Verifier collision user list. When the User attempts to download a file, the Data Cloud Server responds with their blocked information. To resolve the collision, the User requests assistance from the Public Verifier. Once the Public Verifier unrevoked the User, they are able to download files using their corresponding secret key. This approach utilizes proxy re-signatures to allow the Data Cloud Server to re-sign blocks previously signed by the collision User with a resigning key.

```
function CP-HABE(plaintext_data, attributes, security_parameters):
Key Generation
master_key, public_parametersKeyGeneration(security_parameters)
Encryption
 Ciphertext Encrypt(plaintext_data, attributes, public_parameters)
Data Distribution
   distribute(ciphertext)
Authentication
authentication_result Authenticate(ciphertext, public_parameters)
  return ciphertext, authentication_result
```

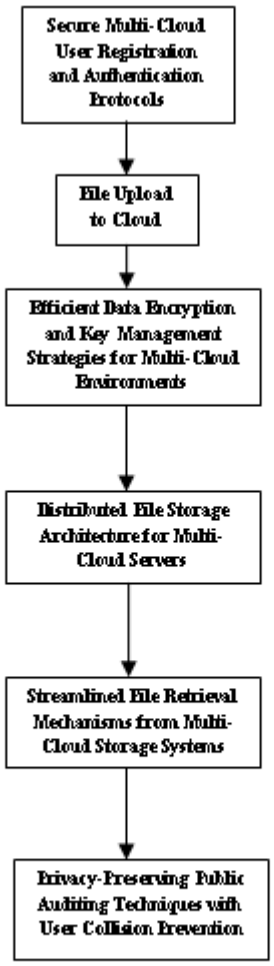

Figure 1. Block diagram

VI. RESULT ANALYSIS

In large-scale network systems, the CP-HABE protocol presents a dual scheme approach to authentication and data custody, dividing the responsibility across several trusted organizations. Through better resource usage and a decrease in the possibility of a single point of failure, this decentralized structure improves both security and scalability. CP-HABE facilitates secure data access while streamlining permission administration in intricate networks by utilizing hierarchical attribute-based access control.

Table 1. Comparison table

algorithm	**accuracy**
Existing system	75
Proposed system	81

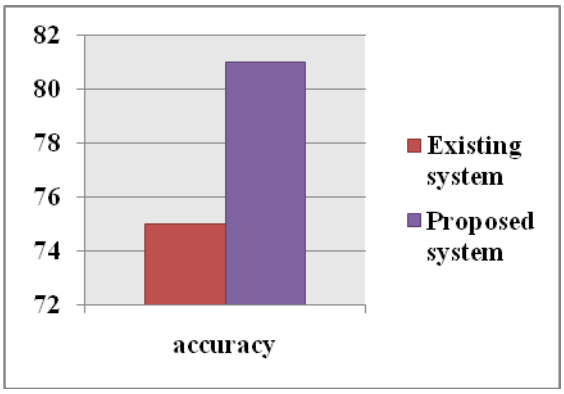

Figure 1. Comparison graph

VII. CONCLUSION

Multiple cloud storage can be secured with a provable data possession (PDP) scheme that enables users to verify data integrity without downloading the entire data. This is particularly useful for sensitive data storage and shared data verification. The PDP scheme for multiple clouds comprises a user client, a cloud server, and an identity management system (IMS). The user client generates and verifies proofs of possession (POPs) and public proofs of possession (PPOPs), while the cloud server stores user data and generates POPs and PPOPs upon request. The IMS manages user identities and public keys, and maintains a revocation list (RL) of revoked users.

VIII. FUTURE WORK

PDP schemes can be computationally demanding, particularly when dealing with large volumes of data. There is a need for future research to concentrate on the development of more efficient PDP schemes, as well as techniques to minimize the overhead associated with generating and verifying POPs and PPOPs. At present, PDP schemes primarily cater to simple data structures like files and databases. However, there is scope for future work to focus on the creation of PDP schemes that can accommodate more intricate data structures, such as graphs and networks. Additionally, current PDP schemes assume that the data stored in the cloud remains static. To address this limitation, future research could concentrate on the development of PDP schemes that can support dynamic data, including data that undergoes frequent updates or deletions.

REFERENCES

[1] Enterprise Information Systems published an article on Industry 4.0 and big data innovations authored by G. Li, J. Tan, and S. S. Chaudhry in 2019.

[2] Future Generation Computer Systems featured a paper on big data driven internet of things for credit evaluation and early warning in finance by C. Wen, J. Yang, L. Gan, and Y. Pan in 2021.

[3] EURASIP Journal on Wireless Communications and Networking published a research paper on intelligent medical big data system based on Hadoop and blockchain by X. Zhang and Y. Wang in 2021.

[4] Mathematical Problems in Engineering presented a Hadoop-based medical image storage and access method for examination series.

[5] Journal of Big Data conducted a survey on data storage and placement methodologies for cloud big data ecosystem.

[6] W. Rajeh, Journal of Information Security, 13 (2) (2022) 23–42, Hadoop distributed file system security problems and investigation of unauthorized access issue

[7] Array. 1-2 (4) (2019) 1–8, G. S. Bhathal, A. Singh, Big data: Hadoop framework weaknesses, security challenges and attacks.

[8] International Journal of Pure and Applied Mathematics, 120 (6) (2020), 11767–11784; Kapil, A. Agrawal, R. A. Khan, Big data security challenges: Hadoop perspective.

[9] "Improvised distributions framework of hadoop: A review," by B. H. Husain, S. R. Zeebaree, et al., International Journal of Science and Business, 5 (2) (2021), 31–41.

[10] Data protection on Hadoop distributed file system by applying encryption algorithms: a systematic literature review by M. Naisuty, A. N. Hidayanto, N. C. Harahap, A. Rosyiq, and G. M. S. Hartono, Journal of Physics Conference Series, 1444 (4) (2020) 1–8.