

Review: Big Data Storage and Privacy in Cloud Environment by using Triple Data Encryption Standard

Prachi N. Deshmukh¹ and Dr. Vaishali H. Deshmukh²

¹⁻²PRMIT&R, Department of Computer Science and Engineering, Badnera, SGBAU, Amravati, Maharashtra, India.

Email: dprachi437@gmail.com, vmdeshmukh@mitra.ac.in

Abstract— With the creation of big data and the increasing adoption of cloud computing, secure and efficient storage solutions for big data in a cloud environment are of paramount importance. This paper presents a novel approach for achieving both efficiency and security in the storage of big data by utilizing Triple Data Encryption Standard (3DES). The proposed solution leverages the strength of 3DES, which applies the Data Encryption Standard (DES) algorithm three times sequentially, to provide a higher level of security. Additionally, to ensure efficiency in the cloud environment. The existing privacy preserving approaches showed several disadvantages such as a lack of data privacy and accurate data analysis, a lack of efficiency of performance and reliance on a third party, which considers the susceptibility of security and reveals the secret data of the individual. The outcomes demonstrate storage and privacy achieved by the proposed solution, making it a promising system for the storage of big data in cloud environments as compare to existing approaches.

Index Terms— Big Data; Cloud; Security; Triple Data Encryption Standard, Data Compression; Cloud Data Storage, etc.

I. INTRODUCTION

Cloud Computing is a fast-developing business computing model and has many advantages, such as large storage, low cost, and scalability. More and more enterprise users are apt to outsource their Big Data to cloud servers for storage and processing [3]. The storage of Internet of Things (IoT) Big Data in Cloud Computing is a research focus in the field of IoT. Since cloud storage systems provide on-demand storage services, storing IoT Big Data in cloud storage systems is both economic and environmental [9]. However, because cloud service interruption occurs frequently [22-25], IoT Big Data stored in single-cloud storage systems are at risk of being unreachable. To ensure that data can be accessed when several cloud services are interrupted, many institutions attempt to store data in modified cloud storage systems. However, existing cross-cloud storage systems are difficult to implement as it is hard to integrate cloud storage services from different cloud service providers [1]. The size of generated data per day on the Internet has already exceeded two exabytes [26]. Outsourced storage service has become the new trend for data owners to upload their Big Data to remote storage servers, in which the cloud storage service, e.g., Amazon Simple Storage Service (S3), Microsoft Azure Storage, has become a popular option for data owners [27]. However, as a centralized storage paradigm, cloud storage suffers from single point of failures [28] due to system failures, power outages, or even malicious attacks [29-30], which results in many cases of data breaches or data loss in the cloud [2].

In addition, the creation and maintenance of data centers is expensive, requires specialized workforce, and can create hurdles to sharing. Conversely, attributes like cost-effectiveness, ease of use, and (almost) infinite scalability make public cloud services natural candidates to address data storage problem [5]. The efficient data storage for storing the Big Data is enabled by data deduplication and to ensure privacy in existing data deduplication is a complex task. It's also suggested that about 75% of digital data are identical (or duplicate) [31], and data redundancy in backup and archival storage system is significantly more than 90% [32]. While cost of storage is relatively cheap and advances in cloud storage solutions allow us to store increasing amount of data, there are associated costs for the management, maintenance, processing and handling of such Big Data [33-34]. It is, therefore, unsurprising that efforts have been made to reduce overheads due to data duplication. The technique of data deduplication is to eliminate duplicate data, by storing only a single copy of redundant data [4]. In other words, data deduplication technique can significantly reduce storage and bandwidth requirements [35]. Also in many situations, the cloud is not considered fully trusted. Thus, the client may employ encryption schemes to keep data confidential before uploading it to the cloud. However, in MSC-based applications, data always be related to certain information, such as location information in location based services. In this situation, which item of data is being accessed leaks additional information to the cloud server. By utilizing this leaked information of access pattern, the cloud may infer the operation of the client and even the content of the encrypted data. Generally a cloud can identify approximately 80% of the search queries by applying a general inference attack with access pattern leakage and minimal background knowledge [6].

Also it is seen that the cloud server is rich in storage and computing resources, edge servers are near to IoT devices, and IoT devices usually have limited resources. The collaborative storage enables the edge servers rapidly respond to requests from IoT devices, e.g., analyzing data collected from IoT devices in real-time and forwarding processed data to the cloud server to save the cost of IoT devices [8]. Among all security issues of cloud storage, data integrity is the most basic one, as it convinces data users that the data they store on the service provider is complete. To deal with the data integrity issue, the concept of data integrity auditing has been proposed, which adopts a third party auditor to audit whether the service provider has stored the users' data intact. Up to now, many public data integrity auditing protocols have been proposed by researchers to achieve different security and functional features, such as privacy preserving, storage correctness, batch verification, and dynamic support., most of solutions are designed based on the public key infrastructure (PKI), which brings a heavy key management burden to the data users. As a result, identity- based cryptography has been exploited by researchers in the public auditing protocols to avoid the potential key management problem [7].

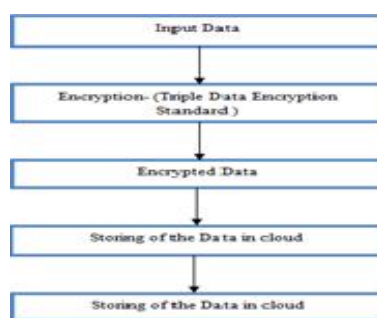


Fig. 1 TDES block diagram for encryption

II. RELATES WORK OR LITERATURE STUDIES

Cloud Computing, plays a major role in information and knowledge extraction. Efficient handling of Big Data often poses innovative security challenges for usual data encryption standards, methodologies and algorithms. Existing studies of data encryption focused on small-to-medium-size information, which does not work well for Big Data due to issues in the presentation and scalability. Thus, important policies for data access control and safety running need to be formulated for dealing with Big Data and this requires incorporating new data management systems. [21].

• An Overview of Big Data And Cloud Computing

Big Data refers to extremely big and complex datasets that cannot be efficiently managed, processed, or analyzed using conventional data processing techniques. It encompasses huge amounts of structured, semi-structured, and

unstructured data that is generated from a variety of sources, including business transactions, social media, sensor data, scientific research, and more [6].

A new popular term known as Big Data technology, used in order to describe the astonishingly accelerated growth of the amount of data (data volume) of structured and unstructured form.. Infrequently, it also relates to a certain size of data set. Precision in Big Data could result in more secure decision making, and better decisions may lead to an increased operational efficiency, reduced costs, and minimized risk [13].

• Methodology

Cloud Computing technology offers utilities which consist of Platform as a Service (PaaS), Infrastructure as a service (IaaS), and Software as a service (SaaS). It is most continent way to offer computation services and can be accessed remotely [4].

Encryption techniques are being used to ensure isolation of data in Internet of Things (IoT) cloud and ensure privacy. Machine Learning (ML) Algorithms are in use including Decision Tree (DT), Support Vector Machine (SVM), Naïve Bayes (NB), K-nearest Neighbors (KNN), and Random Forest (RF) [6] These are the five different Machine Learning Algorithms are useful to classify the dataset. Each of the layout strategies displays a characteristic appropriateness and specificity that depends on the type of dataset [10].

Clustering it is fundamentally a type of unsupervised learning technique. An unsupervised learning technique is a technique in which draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find important structure, illustrative underlying processes, generative features, and groupings inbuilt in a set of examples.

K-means clustering algorithm it is the simplest unsupervised learning algorithm that solves clustering problem. K-means algorithm is partitions n observations into k clusters where each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster [15-19].

In this methodology, also provide the role of the optimization will be to design the keys of large length for ensuring the data privacy without being intruded by any attacks. This methodology provide an optimal 3DES based on the nature-inspired characters, such as coyote and deers, and the role of the optimization will be to render the best keys for data privacy protection. The 3-DES algorithm can reduced the influence of weak and semi weak keys. An optimal 3DES is more flexible and faster [36].

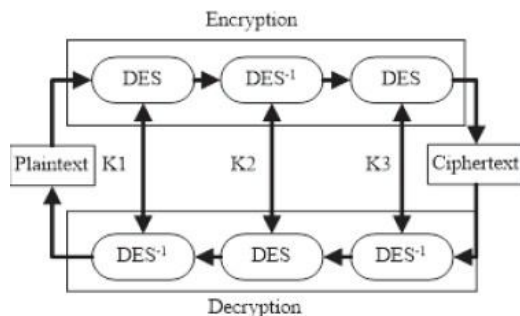


Figure 2: Architecture of 3DES

• Related Work

Efficient handling of big data is a process in cloud storage to exclude duplicate data and achieves data privacy. Many researchers have investigated different aspects of securing big data in cloud environments, including encryption techniques, access control mechanisms, and data integrity verification [11 - 12].

TABLE I. RELATED WORK IN BIG DATA AND CLOUD COMPUTING

Sr. No	Author	Methodology	Result	Limitation
1	Haiyang Yu et al. (2021) [2]	Continuous Big Data integrity Checking approach.	The method has reduced computation and transmission cost.	The method observed the challenges of inconsistency between files and verification tags.
2	Stergiou Christos et al. (2017) [13]	3DES Algorithm	Security and privacy issues of both technologies	security and privacy issues

3	Amanpreet Kaur Sandhu et al. (2022) [14]	Classification Algorithm	Classification and characteristics of Big Data are discussed, along with various Cloud services, such as Microsoft Azure, Google Cloud, Amazon Web Services, International Business Machine cloud, Hortonworks, and MapR	Distributed database storage, data privacy/security, and heterogeneity data formats
4	Kamalpreet Kaure et al. (2017) [16]	MapReduce Algorithm	Describe and identify most vulnerable aspects of security threats in cloud environment	Protecting the system from attacks by utilizing the resources and mitigate risks is only one important problem
5	Yong-Hyun Kim et al. (2017) [20]	BigPros system, Dynamic Workflow model, Directed Acyclic Graph; MapReduce	Provide an efficient data processing method for a medical Big Data analytics in the hybrid cloud	Data Ownership and Control, Compliance Challenges

TABLE II: THE RELATED WORK ON CLOUD STORAGE METHOD

Sr. No	Author	Methodology	Result	Limitation
1	Han Bao et al. (2019) [1]	Adaptive erasure code for Joint Cloud storage Method	The method has reduced Network Resource consumption (NRC) and low repair time.	The system has increased complexity due to more number of storage nodes.
2	Han Yu et al. (2020) [3]	The public auditing scheme for dynamic data storage.	The system is resilient to key exposure and verifies the integrity of data on the demand of users.	The system have the less storage and integrity verification speed.
3	Ricardo Mendes et al. (2019) [5]	Cloud backed storage system with Byzantine resilient data centric algorithm.	The system can store and share a large critical datasets in a controlled manner.	The system observed the potential security issues.
4	Jia-Nan Liu et al. (2020) [6]	Cloud storage using Oblivious Selection and Update protocol.	The system provides improved resistance to malicious clouds.	The system observed the computation overhead.
5	Xiong Li et al. (2020) [7]	Data storage using identity-based dynamic data auditing protocol.	The system has minimized the data integrity issues.	The system observed the challenge of high computational cost and security issues.

TABLE III : THE RELATED WORK ON CLOUD SECURITY

Sr. No	Author	Methodology	Result	Limitation
1	Xue Yang et al. (2020) [4]	Efficient Data Deduplication scheme	The system provided high data availability, accountability and improved resiliency to brute-force attacks.	The system observed the challenge of data loss and corruption.
2	Harsha S. Gardiyawasam Pussewalag et al. (2017) [9]	Attribute based encryption (ABE) mechanisms	Distributed multi-authority ABE scheme and thereby they show how a patient-centric and attribute based PHR sharing scheme	Central TA to manage all attributes in the public domain
3	E. Shanmugapriya et al. (2019) [10]	Triple DES	Big database reduce process Map-Reduce framework with Optimal Fuzzy C means (OFCM) to Clustered data are accumulated in the cloud and furthermore using classification approach to classify sensitive and non-sensitive data in the cloud to encryption.	The system observed the time and security issues.
4	Hui Cui et al. (2016) [11]	Attribute-based encryption	Attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage.	To modify a cipher text over one access policy into cipher texts of the same plaintext but under other access policies without revealing the underlying Plaintext.
5	A. Alabdulatif et al. (2019) [12]	Fully Homomorphic Encryption (FHE)	A privacy-preserving distributed analytics framework for Big Data in cloud. Fully Homomorphic Encryption (FHE) is used as an emerging and powerful cryptosystem that can carry out analysis tasks on encrypted data.	Secure distributed data processing model with possible enhancements to its capabilities for more advanced data mining models.

III. MOTIVATION

Efficient and secure Big Data storage in a cloud environment is a crucial concern in today's data-driven world. The increasing volume, velocity, and variety of data generated by various sources necessitate robust storage solutions that can handle large-scale datasets while ensuring data confidentiality and integrity. One approach to achieving this is by using Triple Data Encryption Standard (3DES) for securing Big Data in the cloud [7].

Triple Data Encryption Standard (3DES) is an encryption algorithm that applies the Data Encryption Standard (DES) algorithm three times to each data block. It provides a higher level of security compared to the original DES by using a longer key length and multiple encryption rounds. The use of 3DES can help protect sensitive data stored in the cloud from unauthorized access and mitigate the risk of data breaches [9].

IV. SECURITY AND CHALLENGES

The cloud-based data systems require the more storage space for the data abundance and maintenance that increases the expensive cost [37].

- The cloud storage faces the data governance and revocation problem due to the users request for the data modification and deletion operations.
- Most of the big data storage it only tolerates crash failures and is not improved to handle big data effectively which should be minimized.
- The big data storage, in which choosing the right data Visualization Tools is important that affects the efficiency.
- The big data storage in which choosing the self-service analytics or business intelligence solution is considered for enhancing the efficiency.

V. APPLICATION OVERVIEW

1. Data Encryption with Triple DES

Encryption Process: Employ Triple DES encryption algorithm to encrypt the data before storing it in the cloud. This involves dividing the data into blocks and applying the encryption algorithm multiple times to enhance security.

Key Management: Implement secure key management practices, including key generation, storage, rotation, and distribution to authorized users, ensuring the confidentiality of keys used in the encryption process.

2. Cloud Storage Integration

Secure Cloud Storage: Utilize a reputable and secure cloud storage service (like AWS S3, Google Cloud Storage, etc.) with robust security measures to store the encrypted data.

Access Control: Implement access controls and authentication mechanisms to restrict unauthorized access to the stored data.

3. Data Retrieval and Decryption

Authorized Access: Implement mechanisms to ensure that only authenticated and authorized users can retrieve and decrypt the stored data.

Decryption Process: Use the appropriate keys and decryption algorithms to decrypt the data retrieved from the cloud before presenting it to authorized users.

4. Privacy and Compliance

Compliance Measures: Ensure compliance with relevant data protection regulations (e.g., GDPR, HIPAA, etc.) and industry standards concerning data privacy and security.

VI. CONCLUSION

The privacy and security of big data is a problematic issue that restricts organizations in utilizing Cloud services. In order to solve such an issue, TDES methodology is used to provide security for the big data in Cloud environments. The proposed research intent to leverages the strength of 3DES, which applies the Data Encryption Standard (DES) algorithm three times sequentially, to provide a higher level of security. Additionally, to ensure efficiency in the Cloud environment, the research work aims towards optimizations through parallel processing and data compression techniques. The combination of 3DES encryption and the use of optimization strategies aims

to address the challenges of securing Big Data while minimizing the impact on storage and retrieval performance. The effectiveness and performance of the proposed approaches will be evaluated through experiments and compared with the existing encryption schemes. The proposed research outcome is to demonstrate the issues related efficiency and security of storing Big Data in Cloud Environment.

REFERENCE

- [1] Bao, H., Wang, Y. and Xu, F., 2019. An adaptive erasure code for joint cloud storage of internet of things Big Data. *IEEE Internet of Things Journal*, 7(3), pp.1613-1624.
- [2] Yu, H., Hu, Q., Yang, Z. and Liu, H., 2021. Efficient continuous Big Data integrity checking for decentralized storage. *IEEE Transactions on Network Science and Engineering*, 8(2), pp.1658-1673.
- [3] Yu, H., Lu, X. and Pan, Z., 2020. An authorized public auditing scheme for dynamic Big Data storage in Cloud Computing. *IEEE Access*, 8, pp.151465-151473.
- [4] Yang, X., Lu, R., Shao, J., Tang, X. and Ghorbani, A.A., 2020. Achieving efficient and privacy-preserving multi-domain Big Data deduplication in cloud. *IEEE Transactions on Services Computing*, 14(5), pp.1292-1305.
- [5] Mendes, R., Oliveira, T., Cogo, V., Neves, N. and Bessani, A., 2019. Charon: A secure cloud-of-clouds system for storing and sharing Big Data. *IEEE Transactions on Cloud Computing*, 9(4), pp.1349-1361.
- [6] Liu, J.N., Luo, X., Weng, J., Yang, A., Wang, X.A., Li, M. and Lin, X., 2020. Enabling efficient, secure and privacy-preserving mobile cloud storage. *IEEE Transactions on Dependable and Secure Computing*, 19(3), pp.1518-1531.
- [7] Li, X., Liu, S., Lu, R. and Zhang, X., 2020. On security of an identity-based dynamic data auditing protocol for Big Data storage. *IEEE Transactions on Big Data*, 7(6), pp.975-977.
- [8] Tao, Y., Xu, P. and Jin, H., 2019. Secure data sharing and search for cloud-edge-collaborative storage. *IEEE Access*, 8, pp.15963-15972.
- [9] Harsha S. Gardiyawasam Pussewalage, Vladimir A. Oleshchuk "A Distributed Multi-Authority Attribute Based Encryption Scheme for Secure Sharing of Personal Health Records SACMAT'17, June 21–23, 2017, Indianapolis, IN, USA
- [10] E. Shanmugapriya & R. Kavitha, "Efficient and Secure Privacy Analysis for Medical Big Data Using TDES and MKSVM with Access Control in Cloud" Springer Science+Business Media, LLC, part of Springer Nature 2019.
- [11] Hui Cui, Robert H. Deng, Yingjiu Li, and Guowei Wu "Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud" *JOURNAL OF LATEX CLASS FILES*, VOL. , NO. , MONTH 2016
- [12] A. Alabdulatif, I. Khalil and X. Yi, Towards secure Big Data analytic for cloud-enabled applications with fully homomorphic encryption, *Journal of Parallel and Distributed Computing* (2019)
- [13] Stergiou Christos, Kostas E. Psannis, "Efficient and secure Big Data delivery in Cloud Computing" 2017.
- [14] Amanpreet Kaur Sandhu "Big Data with Cloud Computing: Discussions and Challenges" *BIG DATA MINING AND ANALYTICS* ISSN 2096-0654 03/06 pp 32 – 40 Volume 5, Number 1, March 2022.
- [15] N. Jayapandian "Cloud Dynamic Scheduling for Multimedia Data Encryption Using Tabu Search Algorithm" The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021
- [16] Kamalpreet Kaur, Ali Syed, Azeem Mohammad, Malka N. Halgamuge "Review: An Evaluation of Major Threats in Cloud Computing Associated with Big Data" 2017 IEEE 2nd International Conference on Big Data Analysis.
- [17] Yibin Li , Keke Gai, Longfei Qiu , Meikang Qiu , Hui Zhao "Intelligent cryptography approach for secure distributed big data storage in Cloud Computing" 2017 Elsevier.
- [18] YANG MING , (Member, IEEE), BAO KANG HE, AND CHEN HAO WANG "Efficient Revocable Multi-Authority Attribute-Based Encryption for Cloud Storage" 2021, IEEE access
- [19] Farzana Rahmana, Md Zakirul Alam Bhuiyan, Sheikh Iqbal Ahamed "A privacy preserving framework for RFID based healthcare systems", 2016 Elsevier journal.
- [20] Yong-Hyun Kim, Eui-Nam Huh "Towards the Design of a System and a Workflow Model for Medical Big Data Processing in the Hybrid Cloud" 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress.
- [21] Liu, S., Zhang, C. and Chen, Y., 2018. HASG: Security and efficient frame for accessing cloud storage. *China Communications*, 15(1), pp.86-94.
- [22] Available: <http://www.kejixun.com/article/190603/464156.shtml>
- [23] <https://www.datacenterknowledge.com/archives/2015/08/21/explosion-downtown-los-angeles-disrupts-data-center-operations>
- [24] <http://it.sohu.com/20110311/n279778961.shtm>
- [25] <http://tech.sina.com.cn/i/2015-05-27/docivxaeafs8200893.shtml>
- [26] Gandomi, A. and Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), pp.137-144.
- [27] Hogan, M., Liu, F., Sokol, A. and Tong, J., 2011. Nist Cloud Computing standards roadmap. NIST Special Publication, 35, pp.6-11.

- [28] Al Nuaimi, K., Mohamed, N., Al Nuaimi, M. and Al-Jaroodi, J., 2012, December. A survey of load balancing in Cloud Computing: Challenges and algorithms. In 2012 second symposium on network Cloud Computing and applications (pp. 137-142). IEEE.
- [29] Subramanian, N. and Jeyaraj, A., 2018. Recent security challenges in Cloud Computing. *Computers & Electrical Engineering*, 71, pp.28-42.
- [30] <https://technode.com/2018/08/06/tencent-cloud-user-claims -1-6-million-compensation-for-data-loss/>
- [31] Harnik, D., Pinkas, B. and Shulman-Peleg, A., 2010. Side channels in cloud services: Deduplication in cloud storage. *IEEE Security & Privacy*, 8(6), pp.40-47.
- [32] Douceur, J.R., Adya, A., Bolosky, W.J., Simon, P. and Theimer, M., 2002, July. Reclaiming space from duplicate files in a serverless distributed file system. In *Proceedings 22nd international conference on distributed computing systems* (pp. 617-624). IEEE.
- [33] Bellare, M., Keelveedhi, S. and Ristenpart, T., 2013. Message-locked encryption and secure deduplication. In *Advances in Cryptology–EUROCRYPT 2013: 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Athens, Greece, May 26-30, 2013. *Proceedings 32* (pp. 296-312). Springer Berlin Heidelberg.
- [34] Bellare, M., Keelveedhi, S. and Ristenpart, T., 2013. DupLESS: Server-Aided Encryption for Deduplicated Storage. *IACR Cryptol. ePrint Arch.*, 2013, p.429.
- [35] Miao, M., Wang, J., Li, H. and Chen, X., 2015. Secure multi-server-aided data deduplication in Cloud Computing. *Pervasive and Mobile Computing*, 24, pp.129-137.
- [36] Wang, 2019 " The Design of 3-DES Encryption System Using Optimizing Keys " IEEE Access.
- [37] G. P. Bherde and M. A. Pund, "Recent attack prevention techniques in web service applications," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 2016, pp. 1174-1180, doi: 10.1109/ICACDOT.2016.7877771.