Archives available at journals.mriindia.com

International Journal of Recent Advances in Engineering and Technology

ISSN: 2347 - 2812

Volume 14 Issue 02s, 2025

Review paper on Secure and Efficient Framework for Big Data Storage and Access Control in Cloud Environments Using Optimized Cryptographic Algorithms

[1]Mrs. Anuja R.Tungar, [2]Dr.D.S.Deshpane

[1]CSE Department, MGM University, Ch.Sambahjinagar, Maharashtra, India

[2]CSE Department, MGM University, Ch.Sambahjinagar, Maharashtra, India

Email: [1]atungar@mgmu.ac.in , [2]ddeshpande@mgmu.ac.in

Peer Review Information

Submission: 21 Oct 2025

Revision: 18 Nov 2025

Acceptance: 05 Dec 2025

Keywords

Data security, cloud computing, Access control mechanism, secure storage framework, optimal Elliptic Curve Cryptography

Abstract

A transition of industries from analogue to digital in last few decades has given rise to the requirement for reliable and strong big data handling systems. These systems must be able to handle complex and large databases at lowest possible cost but with high speed and reliability. Finance and retail industry, healthcare industry are the examples using big data and they rely heavily on big data analytics for decision making. Their fraud detection and compliance activities heavily depend in big data analytics. Cloud platform can satisfy these requirements. Cloud is cost effective, scalable as well as real time. Big data and Cloud technologies combined, forms Big Data Cloud, that provides necessary support for structured and unstructured data. But these systems, being shared among multiple users are prone to security and reliability attacks. This work proposes a secure storage framework and access control mechanism utilizing optimized cryptographic computations. Specifically, an Optimal Elliptic Curve Cryptography (OECC) approach is proposed, enhanced by the Modified Seagull Optimization Algorithm (MSOA) for key generation. The goal is to ensure secure data storage in the cloud while maintaining efficient data access and addressing critical privacy and computational efficiency needs.

Introduction

The revolution in Big Data demands the development of infrastructure and architecture capable of handling large volumes of varied data. As the data size is scaled, classical algorithms' robustness, stability and reliability are compromised and are called the curse of dimensionality. The requirement is to develop high dimensionality reduction algorithms for heterogeneous data. Data of big data is sourced from various sources, so there is always a security issue when this data is stored on the cloud; it is always prone to various attacks like collusion attacks, Denial of Serve attacks and insider attacks, to name a few. Existing big data systems do not show the expected level of excellence in data protection systems.

Hence, to solve these problems and to protect the data from attacks, it is necessary to develop a more secure and reliable framework for big data storage The developed system must be capable of providing a mechanism that is secure as well as fast and less costly. It should have minimum memory requirement and minimum number of operations too so that the processor requirement is minimized. Optimized cryptographic algorithms seems to a feasible solution to address this problem. So a system can be developed using this algorithm for secure, reliable as well as fast storage system

that provides the necessary access control mechanism for data stored on cloud. This will be a three-step implementation, data preprocessing to reduce size of data, Set-up and perform encryption followed by data storage construction phase.

Literature survey

A literature review is divided into three sections based on the three phases of implementation.

1. Data preprocessing: This step consists of converting raw data, which is received from various types of sources, to usable format. Basically, Data preprocessing is to clean data, or in some application data reduction, transformation. Data cleaning is performed to remove possible errors and some missing values is the received data. Data reduction is to reduce size of data. Dimensionality reduction, data compression are the techniques that can be used. Many papers that were studied highlight the problem of "Curse of dimensionality".

Raina Mkhinini Gahar et al. [1] introduced a distributed statistical approach that used the MapReduce algorithm for high dimensionality reduction of heterogeneous data. This paper studies the challenge of combining quantitive and qualitative variables for analysis. They also used MCA (Multiple Correspondence Analysis), PCA with indicators and FAMD (Factor Analysis of Mixed Data). Two methods are discussed in depth: The statistical methods for descriptive analysis and the distributed approach for dimensionality reduction. The distributive algorithm includes data preparation, separation, and the application MapReduce algorithm. The method's performance is evaluated based on reduction rate runtime and accuracy. This method also ensures that there is no data loss while minimising the size of the data. Ireneusz Czarnowski and Piotr Jędrzejowicz et al. [2] discuss a method of combining stacking, rotation and then data reduction, which is done to improve machine classification performance. They discuss many data reduction techniques suitable for big data sets in machine learning. The authors found that solving large-scale machine- learning problems through innovative methods is necessary. G. Thippa Reddy et al. [3] have studied Linear Discriminant Analysis(LDA) and Principal Component Analysis(PCA) methods on four popular machine learning algorithms, namely Decision Tree Induction, Support Vector Machine (SVM), Naïve Bayes Classifier and Random Forest Classifier using Cardiography (CTG) dataset. This data set was available from the University of California and Irvine Machine Learning Repository. These experiments concluded that PCA performs better than LDA in all measures. Tonglin Zhang et al. [4] justify how traditional methods are ineffective in performing big data analysis and the reasons for PCA's limitations towards big data analysis.

2. Cryptography: Encryption is applying a mathematical function to plain text to convert it into cypher text [5]. A Cipher is an algorithm for performing conversion. Key size plays an important role to decide the security level. The capability of the selected algorithm for encryption should be such that, it should be very hard, theoretically impossible to interpret the encrypted text to a readable one. But as the key size increases, the computational operations and memory requirement of the processing also increased. Encryption must fulfil the requirements: Integrity, privacy, authentication, and non-repudiation. Symmetric and asymmetric key algorithms exist. Key distribution is a main problem in symmetric key cryptography. Both parties need to have a predefined set-up in terms of key sharing, and the key must be kept secret by both.

3. ECC and Bigdata: Resource sharing for data storage or application development is now possible due to cloud computing. Cloud services are modelled as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Algorithms like RSA and AES were popular when cloud computing was introduced. So, they were used initially for cloud services for data security. As the utilisation of cloud services increased, RSA faced major challenges for large databases. It required a large amount of time for encryption. Eg. To encrypt 1TB AES with a 128-bit key size takes seven hours. Hashing 1TB, SHA-1 takes 45 minutes, and RSA is too slow to apply big data directly. Lei Xu [6] mentioned in 2014 that privacy-preserving data mining (PPDM) modifies data to allow mining without exposing sensitive information. PPDM then became Do Not Track (DNT), and controlled access to online activities was then possible partly, as, in their opinion, data privacy cannot be granted once it is shared on any common platform.

Rajashsree. V. and Dr J. Gnanajayanthi [7]performed a survey on the "evolution of cryptographic systems and the cryptanalysis that resulted in failures". This paper discussed societal acceptance, system drawbacks, and emerging research opportunities. They emphasised that no digital system is entirely secure; it is always likely to be hacked or eavesdropped. A. Malik, M. Aggarwal, B. Sharma, A. Singh, and K.

K. Singh [8] proposed two options to store data on the cloud: (1) to split and store or(2) to store on a single virtual machine. They analysed files for sensitivity, and the decision was made. The authors used the Cuckoo Search Algorithm. RSA and OECC were used for encryption. E. K. Subramanian and Latha Tamilselvan proposed the Elliptic Curve Diffie-Hellman (ECDH) algorithm to enhance cloud data security. ECDH reduces computational complexity and improves encryption efficiency. They obtained 70% better results measured against the parameters: Encryption time, computational overhead, and key generation time. The comparison was with RSA, MRSA and MRSAC. The data set used was the Enron corpus database, which contained 200,399 messages from 158 users. The problems identified were: (i) The existing data encryption system is still complicated and can be reduced if ECC with ECDH is used. (ii) Existing encryption techniques do not secure the pattern access information, which affects the cloud's privacy. They suggested that the pattern access information should also be encrypted to increase security level. Though ECDH gives 70% better performance, a secure, relevant data retrieval mechanism can be incorporated into the cloud storage. In a multi-tenancy architecture, Pawan Kumar and Ashutosh Kumar Bhatt [9] developed a secure data encryption approach using ECC and DNA encryption. They recognised that SaaS services provided by cloud computing service providers (CSPs) often use shared software instances. This multi- tenancy model suffers a major threat from unauthorised sources. To overcome this problem, secured authentication is needed. Rakesh Kumar and Rinkaj Goyal's survey on cloud security requirements [10] emphasises the need for self-adaptive security measures to address the dynamic nature of cloud computing threats. The trust- based adaptive security they proposed opens many research areas for improving cloud security. This approach can support secure and innovative business services, leveraging technologies like IoT, Big Data, 5G, SDN, and NFV. The paper highlights the importance of adapting security strategies to changing environments.

The summary of the literature survey highlights the following points:

a. Do Not Trace cannot guarantee the safety of users' privacy
b. Two computationally heavy algorithms, RSA and ECC, were used to improve security levels, combined with the Cuckoo search algorithm. However, this is not a complete success.
c. As per [11], The execution time of the ECDH algorithm is around 70% better performance than the existing cryptographic methods. In future, a secure, relevant data retrieval mechanism can be incorporated into the cloud storage.
d. [12] have mentioned that Distributed DoS (DDoS) is a very dangerous attack for multi-tenant cloud storage devices. If sufficient resources are available, one can avoid DDoS attacks. So, designing the best dynamic resource allocation strategy and load balancing for multi- tenant cloud storage is necessary to counter DDoS attacks.
e. After surveying many research articles on data preprocessing, it was observed that the dimensionality reduction technique is mainly useful for feature extraction, followed by data analysis. We had thought of using this technique for data preprocessing to reduce the data size. However, we now conclude that dimensionality reduction will not be used, and alternate approaches must be searched, including data compression or equivalent.

Methodology

The volume of big data comes from various sources such as business records, share market, healthcare and many more. There is a possibility of many security issues in the big data stored in the cloud. Existing access control techniques are weak and hence lead to different attacks like collusion attacks, denial of service attacks and insider attacks on the data stored in the cloud.

Advanced encryption techniques and optimization algorithms have been introduced to enhance security and efficiency. Recent research explores AI-driven encryption assists in optimizing encryption processes, making them faster and more secure. Additionally, metaheuristic optimization techniques, such as Binary Banyan Tree Optimization (BBTO), are being integrated with cryptographic algorithms to enhance key generation and reduce computational complexity. These advancements aim to improve encryption speed, reduce storage overhead, and strengthen cloud security, ensuring that big data remains protected from unauthorized access and cyber threats. As cloud computing continues to evolve, securing big data storage through optimized encryption and decryption will remain a key research focus. The development of intelligent cryptographic solutions will not only enhance

data protection but also ensure seamless and secure cloud computing experiences for users worldwide.

1. Lossless compression: The lossless compression is a crucial technique for big data storage in the cloud, enabling efficient data management without sacrificing information integrity. By reducing file sizes while preserving original content, lossless compression optimizes storage utilization, data transfer speeds, and processing efficiency. Burrows-Wheeler Transform (BWT) model is utilized minimize redundancy and enhance cloud-based data accessibility. In cloud environments, where massive datasets require frequent access and processing, lossless compression ensures cost-effective storage, bandwidth optimization, and faster retrieval, making it essential for data-intensive applications.

2. Secure Data Storage Using Enhanced RSA (ERSA) Algorithm

- The big data are initially preprocessed and stored in cloud storage space.
- The Enhanced RSA (ERSA) algorithm is introduced to improve data encryption and decryption performance. ERSA consists of three key phases:
- Key Generation – Generates asymmetric key pairs (private and public).
- Encryption – Encrypts data using two diverse public keys (P1 and P2) with a modular exponential technique.
- Decryption – Recovers plaintext using two private keys.
- Optimization of RSA with the Modified Binary Banyan Tree Optimization (BBTO) Algorithm

3. The standard BBTO algorithm is modified to reduce encryption and decryption processing time.

- This modification optimizes key selection and computational efficiency, making the encryption process faster while maintaining strong security.

Expected Outcomes

The proposed model will improve big data integrity and speed, enabling effective big data management in cloud storage. The ERSA algorithm with modified BBTO will provide higher security with faster encryption and decryption compared to conventional RSA. Reduced Processing Time – The optimized BBTO algorithm will significantly enhance encryption speed, reducing cloud storage overhead.

References

R. M. Gahar, O. Arfaoui, M. S. Hidri, and N. B. Hadj-Alouane, "A Distributed Approach for High-Dimensionality Heterogeneous Data Reduction," IEEE Access, vol. 7, pp. 151006–151022, 2019, doi: 10.1109/ACCESS.2019.2945889.

I. Czarnowski and P. Jędrzejowicz, "An Approach to Data Reduction for Learning from Big Datasets: Integrating Stacking, Rotation, and Agent Population Learning Techniques," *Complexity*, vol. 2018, no. 1, p. 7404627, Jan. 2018, doi: 10.1155/2018/7404627.

G. T. Reddy *et al.*, "Analysis of Dimensionality Reduction Techniques on Big Data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020, doi:10.1109/ACCESS.2020.2980942.

T. Zhang and B. Yang, "Big Data Dimension Reduction Using PCA," in *2016 IEEE International Conference on Smart Cloud (SmartCloud)*, New York, NY, USA: IEEE, Nov. 2016, pp. 152–157. doi:10.1109/SmartCloud.2016.33. "Elliptic_Curve_AnnopMS.pdf."

Z. Wang, Y. Wang, L. Zhang, C. Zhang, and X. Zhang, "A Dimensionality Reduction Algorithm for Unstructured Campus Big Data Fusion," *Symmetry*, vol. 13, no. 2, p. 345, Feb. 2021, doi: 10.3390/sym13020345.

V. Rojasree and J. Gnanajayanthi, "Cryptographic Algorithms to Secure Networks - A Technical Survey on Research Perspectives," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India: IEEE, Aug. 2020, pp. 159–165. doi: 10.1109/ICSSIT48917.2020.9214259.

A. Malik, M. Aggarwal, B. Sharma, A. Singh, and K. K. Singh, "Optimal Elliptic Curve Cryptography-Based Effective Approach for Secure Data Storage in Clouds:," *International Journal of Knowledge and Systems Science*, vol. 11, no. 4, pp. 65–81, Oct. 2020, doi: 10.4018/IJKSS.2020100105.

P. Kumar and A. Kumar Bhatt, "Enhancing multi-tenancy security in the cloud computing using hybrid ECC- based data encryption approach," *IET Communications*, vol. 14, no. 18, pp. 3212–3222, Nov. 2020, doi: 10.1049/iet-com.2020.0255.

R. Kumar and R. Goyal, "On cloud security requirements, threats, vulnerabilities and countermeasures: A survey," *Computer Science Review*, vol. 33, pp. 1–48, Aug. 2019,

doi:10.1016/j.cosrev.2019.05.002.

D. R. Hankerson, S. A. Vanstone, and A. J. Menezes, Guide to elliptic curve cryptography. New York: Springer, 2003.

W. Stallings, “Cryptography and Network Security: Principles and Practice,” p. 900.

Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren, “Information Security in Big Data: Privacy and Data Mining,” IEEE Access, vol. 2, pp. 1149–1176, 2014, doi: 10.1109/ACCESS.2014.2362522.