JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 26 Issue 04, 2023 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

AN OPTIMIZED NOVEL ALGORITHM FOR BIG DATA SECURITY: XPP DOUBLE ELLIPTIC CURVE CRYPTOGRAPHY

Hriday Kumar Gupta [1,2], Rafat Parveen[1]

[1]Department of Computer Science, Jamia Millia Islamia, New Delhi, India
[2]KIET Group of Institutions, Ghaziabad, Delhi NCR

Abstract

The research presents a new encryption system that combines Big Data-Map Reduce (BD-MR) for efficient data processing, Access Pattern Matching (APM) for analyzing user interactions, and the XOR public and Private Key Double Elliptic Curve Cryptography (XPP-DECC) algorithm for strong cryptographic protection. The primary goal of this research is to address the urgent need to protect the easy availability, confidentiality, and reliability of sensitive data in cloud services. It specifically covers the issues of securing huge databases given limited resources.

Keywords: Big Data, Cloud environment, Cryptography, Elliptic Curve Cryptography, XOR Public Private Key

INTRODUCTION

Cryptographic techniques are essential for improving the security of cloud computing. Various cryptographic techniques are utilized to secure data while it is being stored, shared, and processed. Public-key cryptography, symmetric-key cryptography, and homomorphic encryption are cryptographic approaches used to accomplish various security goals. Public-key cryptography guarantees the security of communication and authentication, while symmetric-key cryptography facilitates the encryption and decryption of data. Homomorphic encryption enables the execution of operations on encrypted data, ensuring the confidentiality of the information while carrying out computations . These strategies enhance the strength of data security in cloud computing environments. [1]. Cloud computing has grown into a very potential and rapidly growing field within the information technology industry. In Q1 2022, worldwide spending on cloud infrastructure services surged by 34%, reaching a total of $55.9 billion. This increase was driven by companies prioritizing digitalization initiatives to address market obstacles [2]. The Elliptic Curve Cryptography (ECC) is a notable cryptographic scheme with power imbalance in key distribution. It was independently developed by Miller in the late 1980s. [3] and Koblitz [4]. Cloud computing is an innovation in the rapidly changing field of IT; it has altered commercial practices and people's relationships with online tools. The introduction of the cloud, which provides a variety of services including infrastructure, platforms, and software, has completely altered the data lifecycle.

LITERATURE REVIEW

Submitted: 22/11/2023 **Accepted**: 30/12/2023

This section reviews cloud computing security paradigm research. Subramanian et al. [5] established the Elliptic Curve with Diffie–Hellman (ECDH) approach for data encryption and decryption for enhancing data security in the cloud. Here, the input data was divided into various blocks. After that, a secret key was created for each block and positioned in the cloud. Lastly, all the data blocks were decrypted and merged to access the data. Thus, data authentication was guaranteed with reduced information delay. But, the consumption time was increased by the adaption of the challenge and proof generation process.For cloud-based health data security and privacy, ArakapuVineela et al. [6] created an authentication method. Encrypting and storing patient sensor node data in the cloud allowed clinicians and patients to analyse it. The approach simplified computing and communication error detection but lacked privacy and attack resistance. SADS-Cloud was created by Uma Narayanan et al. [7]. Data Owners (DO) were registered with a trusted centre, input files were split into blocks using MapReduce, and compression, clustering, and indexing were used for effective data management. However, encrypting after compression raised data loss worries. The Holistic Big Data Integrated Artificial Intelligent Modelling (HBDIAIM) technique was utilized by Jie Chen and colleagues [8] in order to ensure the efficient and secure administration of data stored in the cloud. Enhanced scalability and accessibility were achieved by the use of a differential evolutionary algorithm that was accompanied by a decision privacy system that was supported by data analytics. In spite of the fact that the model exhibited security, precision, and dependability, it was vulnerable to passive attacks.

Abdulatif Alabdulatif [9],has built a secure framework for analysing big datasets in the cloud. In order to protect the confidentiality of the data, they have utilised a revolutionary method known as Fully Homomorphic Encryption (FHE). By distributing computations among independent cloud nodes, the framework is able to accomplish both speed and accuracy in its analysis. As a result, it is an invaluable instrument for the development of cloud-based analytics applications that are both secure and efficient. [10] For the purpose of processing large amounts of data, the author suggested a security model that would be constructed on top of the Hadoop architecture that was already in place. In addition to the typical Hadoop Distributed File System (HDFS) and MapReduce (MR) layers, this approach contains an additional layer known as the ”Secured Map Reduce (SMR) Layer.” The initial step is to collect the data and then store it in the HDFS. After that, the SMR model encrypts the data in order to protect the confidentiality of the information while it is being processed in the MapReduce layer [11]. The data is encrypted in the private cloud using this method, which makes use of vertical partitioning within the HybrEx paradigm. After that, the data is transferred to the public cloud for additional processing, where it is decrypted when it arrives there. Trials have demonstrated that this approach strikes a healthy balance between privacy and security, which is a significant advantage when it comes to managing large amounts of data [12].

MOTIVATION

As the difficulties of protecting sensitive big data in cloud systems continue to increase, this research is being pushed by those difficulties. As the volume of data and the usage of cloud

computing continue to grow, so do issues surrounding security, particularly with regard to the confidentiality of sensitive information as well as the overall integrity and availability of the system. These vulnerabilities are the targets of attacks from both inside and outside the organisation. The purpose of this research is to keep one step ahead of ever-evolving threats and to ensure that big data processing in the cloud is carried out in a secure manner. This research investigates security on several levels, ranging from sophisticated encryption methods to strong system architectures.

A. Security Level

When saving and processing large amounts of data in the cloud, data security is very important. The reason for looking at the amount of security in this study is to come up with methods, frameworks, and techniques that make big data in the cloud safer overall [13]. This includes steps to keep data safe from people who shouldn't have access to it, make sure the data is correct, stop data breaches or leaks, and make sure that only authorized users can access the data. The study is trying to find ways to store and process sensitive data in the cloud that are as safe as possible by focusing on the level of security.

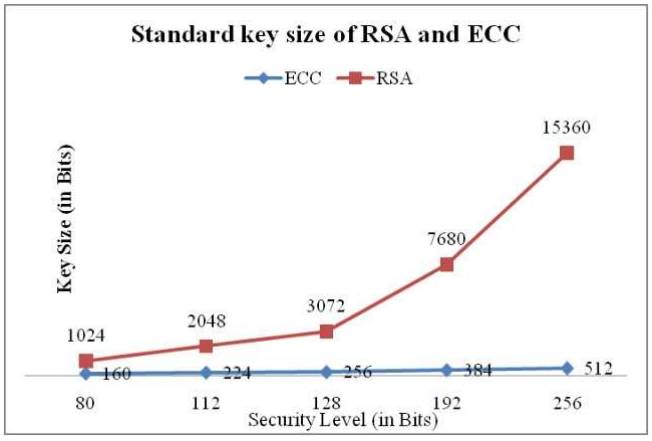

Figure 1: Key size and security level: NIST recommended [14]

B. Attack Level

The reason for the attack level is that cyber threats and attacks that target big data in cloud computer environments are getting smarter and more varied. Attackers may use unauthorised access, data interception, data manipulation, denial of service, and malware injection, among other methods, to get into big data and risk its security and privacy [15]. By looking at the attack level in this study, we hope to come up with countermeasures, detection systems, and defensive methods that can find and stop attacks that might be aimed at big data in the cloud. This includes making strong encryption methods, breach detection systems, access control systems, and threat intelligence systems so that attacks can be found and stopped quickly.

PROPOSED METHODOLOGY

The implementation of a proposed architecture that is comprised of four stages is the research methodology for the study that is titled ”XPP-DECC Algorithm for Enhancing Cloud Storage Security.” These stages are as follows: the data owner stage, the encryption process stage, the decryption process stage, and the text matching process stage. In order to produce and distribute keys for cloud storage in a secure manner, the primary goal of this security model is stated. For the purpose of key generation, the XPP-DECC algorithm is utilised. This algorithm makes it possible to utilise encryption keys that are less in size and utilises a secure method of key generation.

In the Data Owner phase, large data files are to be uploaded to the cloud server at the Data Owner stage. After that, these files are partitioned into blocks, which are then mapped onto the Mapper phase of the Hadoop Distributed File System (HDFS) framework.For the purpose of encrypting the data in an effective manner, the encryption procedure step makes use of the XPPDECC algorithm. The security of cloud data is ensured by the implementation of symmetric encryption, as well as the utilization of an elliptic curve function that is founded on XOR Public and Private Keys. In order to guarantee the safety of the generation of shared secret keys, the method is utilized.

decryption phase entails carrying out a series of operations on the private keys in order to decrypt the messages that have been encrypted. A particular equation is utilised in order to carry out the mathematical operations that are carried out over an elliptic curve. The public keys that are generated by generator G are accessible to all of the persons involved, while the private keys are kept secret.Following the transformation of ASCII data into affine points on an elliptic curve, the next stage of the text-matching process involves conducting cryptographic operations on the grouped ASCII values of the characters. In order to determine the size of the group, the size of the list that is generated by the Integer Digits function in Mathematica is taken into consideration. After that, the values that have been aggregated are transformed into big integers, and a pairing is made in order to transfer them to elliptic coordinates. This eliminates the requirement for a jointly maintained lookup database.

The approach of the research places an emphasis on the implementation and assessment of the XPP-DECC algorithm for the purpose of improving the security of cloud storage applications. The implementation of the proposed design, the execution of tests and experiments to evaluate the efficiency and efficacy of the algorithm, and the analysis of the findings are all things that are involved in this process. In order to guarantee the safe generation and distribution of keys for cloud storage, the technique takes a methodical approach. Additionally, it takes into account the efficacy and efficiency of the encryption and decryption procedures. The security framework for cloud-based big data that uses the XPP-DECC algorithm is shown in Figure 1.

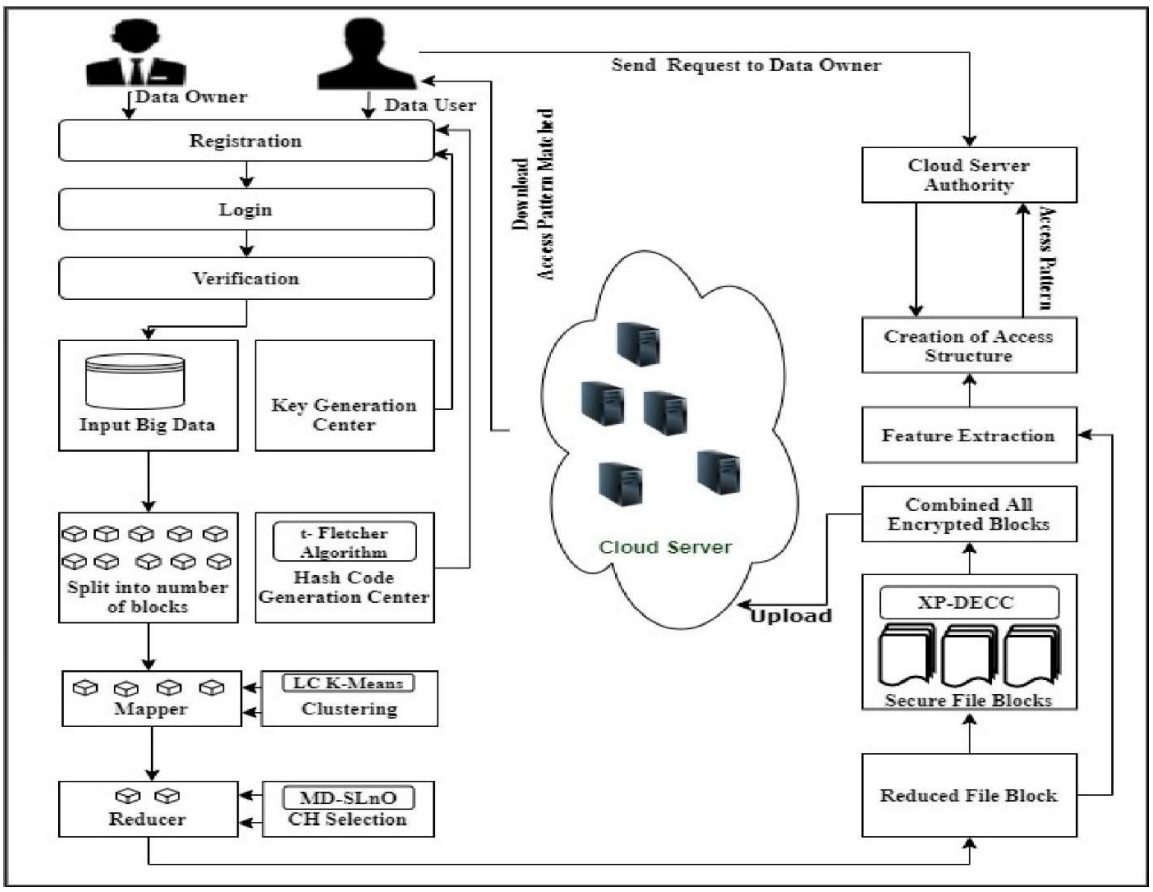

Figure 2: Proposed framework for Big data Security

Elliptic Curve Cryptography (ECC) provides a high level of security by employing keys that are substantially smaller than those that are utilised in conventional approaches. ECC is being considered an innovative successor to RSA due to the fact that it provides a number of benefits. In the first place, the ECC keys and signatures are substantially smaller than the ones that are used in RSA, yet they nevertheless offer the same level of security. Additionally, it has a speedier process for the development of keys, the establishment of agreements, and the fabrication of signatures. The sophisticated mathematical processes that involve points on elliptic curves, which are at the core of ECC operations, are the source of this advantage. As a result of these intricate computations, it is incredibly difficult to break the encryption, which ensures that your communications remain private.

Encoding a plain text message into a point on the curve is the first step in the process involved. The public key and the private key are both created at the second step of the process. As we move on to the third step, the message is encrypted by making use of the public key. In the final step, which is Step 4, the decryption process is carried out with the private key.

The public keys Pub and PUa, which are generated by the generator G, are known to all of the parties that participate in the exchange. However, the PRb and PRa private keys have not been

disclosed to anyone. In this particular scenario, the conventional method of encoding ASCII data into affine points on an elliptic curve is not taken into consideration.

The equation describes an essential mathematical operation over the elliptic curve $y^2 = (x^3 +ax+b) \bmod p$ in such a way that $4a^3 +27b^2 \bmod p \neq 0$, where a and b are the coefficients that generate distinct points on the elliptic curve (x,y) and p is a

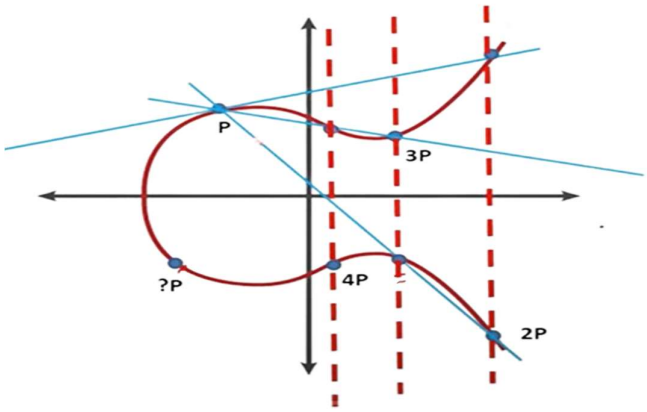

Figure 3: Private key with elliptic curve

substantial prime integer. The provided equation denotes the Weierstrass normal form applied to elliptic curves [16].

Before the Key generation, Encryption and Decryption a message (M) is represented as an elliptic curve point (P) within the context of a constructed finite field, the point P is a member of the elliptic curve $E_p(a, b)$. where, $E_p(a, b)$ denotes an elliptic curve formed over the finite field with the constants 'a' and 'b'.

Key Generation

Assign a generator point 'G' to the elliptic curve E_p, which is defined as (a, b). Following this, select 'n' as the private key, guaranteeing that it is a prime number from 1 to (p-1) inclusive, with 'p' denoting the order of the elliptic curve.

Represented as Public Key $(P_u) = nG$,

(1)

Where nG, represented by the coordinates (ψ_3, Φ_3). The result of aggregating two points on the elliptic curve, $P_1(\psi_1, \Phi_1)$ and $P_2(\psi_2, \Phi_2)$, is denoted by this point. The point denoted by the coordinates (ψ_3, Φ_3) is the result of the addition operation defined below.

$$\psi_3=(\lambda^2- \psi_1- \Phi_2) \bmod p \quad \text{and} \quad \Phi_3=(\lambda(\psi_1- \psi_3)- \Phi_1) \bmod p \tag{2}$$

For the purpose of determining the lambda value, Equation 3 applies in cases where P1 and P2 are identical to one another, but Equation 4 is applied in cases where P1 and P2 represent different points on the elliptic curve. It implies that the related equation is utilized in the process

of determining the lambda value in elliptic curve computations, and this calculation is carried out in accordance with whether the two points are equal or different.

$\lambda=(3\ \psi^2+a)/2\Phi \bmod p$

(3) $\lambda=(\ \Phi_2-\ \Phi_1)/(\ \psi_2-\ \psi_1)\bmod p$

(4)

Encryption

Choose a random key 'k' such that $1\leq k\leq p-1$ where 'p' denotes the order of the elliptic curve
The initial portion of the ciphertext C_1 is computed using the outcome of the scalar multiplication (kG).
The second portion of the ciphertext C_2 should be computed using the sum of two elements.(M+kPu).
Following this, the ciphertext C is denoted by the ordered pair (c_1,c_2).

$C = [(kP_u), (M+kG)]$

(5)

This procedure employs the stochastic nature of 'k' in order to bolster the encryption's security. By utilising the inverse operations, the recipient, who possesses the corresponding private key, can decrypt the message.
Double encryption can provide an additional layer of security,
First Encryption: ECC_Encrypt(Public Key1, plaintext) -> ciphertext1
Second Encryption: ECC_Encrypt(Public Key2, ciphertext1) -> final_ciphertext

Decryption

Compute $M+kP_u$ as the sum of the ciphertext components that were received.
Determine n(kG) by performing a scalar multiplication between the private key 'n' and the random key 'kG'.
n(kG) is subtracted from M+kPu.
The formula for the decrypted message

$M=(M+kP_u)-[n(kG)]$.

Double encryption can provide an additional layer of security,
First Encryption: ECC_Encrypt(Public Key1, plaintext) -> ciphertext1
Second Encryption: ECC_Encrypt(Public Key2, ciphertext1) -> final_ciphertext

(6)

By executing this procedure, the encryption is successfully reversed, guaranteeing that solely the individual in possession of the private key can decipher the initial message.

EXPERIMENTAL RESULT AND DISCUSSION

Table 1 illustrate the time required for encryption and decryption, as well as the level of attack and the level of security. also presents the comparative evaluation of the suggested XOR-ECC algorithm in comparison to the traditional ECC and RSA methods. Here, the RSA demonstrates an inadequate performance. Comparatively, the ECC algorithm has relatively higher efficiency

than the RSA algorithm. For example, the time required for encryption and decryption with ECC is 4578 milliseconds and 4495 milliseconds, respectively, which is a less amount of time than the RSA system. The suggested XOR-ECC technique, on the other hand, achieves encryption and decryption times of 3125 milliseconds and 3157 milliseconds, respectively, which are considerably longer than the ECC approach. In a similar vein, the proposed methodology offers superior performance with relation to other criteria as well. Because of this, the results demonstrate that the XOR-ECC that was presented can be utilised for the secure uploading of BD files to the cloud.

Table I: Proposed XOR-ECC algorithm performance analysis

Performance Metrics	Proposed XOR-ECC	ECC	RSA
Encryption time	3245	4578	5124
Decryption time	3157	4495	5348
Attack Level	6	12	18
Security Level	96	93	91

The performance evaluation of the proposed XOR Elliptic Curve Cryptography (XOR-ECC) is carried out in comparison with the existing ECC and RSA algorithms. This evaluation takes into consideration the amount of time required for encryption and decryption, the level of attack, the level of security, and the amount of memory that is utilised for encryption and decoding.

CONCLUSION AND FUTURE WORK

XOR-Private Public elliptic curve cryptography-based architecture improves cloud-based large data security and privacy. Elliptical Curve Cryptography has a smaller key size than other cryptographic methods and provides more security. ECC is used for authentication on low-computational devices because to its strong security and small key length. To prevent unauthorised decryption, the output feedback approach divides encrypted data into chunks. Thus, ECC with output feedback mode provides better security with less hardware. The XOR-ECC-based secure data uploading performance is compared to ECC and RSA methods. The proposed XOR-ECC has a 6% attack threshold, lower than ECC (12%) and RSA (18%). This investigation reveals that cloud-based BD is secure and secret. Many improvements are possible in the future. Distributing computational load and storage with a distributed cloud server can improve performance and dependability. Signature-based authentication verifies data integrity and authenticity, strengthening data protection.These improvements should make the suggested methods and architecture more adaptive, scalable, and secure. By tackling these areas, the research can improve cloud algorithms, data security, and privacy. Several improvements are worth considering. First, investigating a distributed cloud server

infrastructure could distribute computational load and storage, boosting system performance and dependability. Signature-based authentication can also improve data security by validating integrity and authenticity. These future directions aim to improve the architecture and methodologies' adaptability, scalability, and security. These domains can increase algorithms, data security, and privacy in cloud-based systems, enabling additional advancements.

REFERENCES

[1] The rise of "big data" on cloud computing: Review and open research issues, Information Systems, vol. 47, pp. 98–115, 2015.

[2] Canalys 2022. https://canalys.com/newsroom/global-cloud-services-q1-2022. accessed 12 aug 2022.

[3] V. S. Miller, "Use of elliptic curves in cryptography," in Conference on the theory and application of cryptographic techniques. Springer, 1985, pp. 417–426. [4] N. Koblitz, "Elliptic curve cryptosystems," Mathematics of computation, vol. 48, no. 177, pp. 203–209, 1987.

[5] E. Subramanian and L. Tamilselvan, "Elliptic curve diffie–hellman cryptosystem in big data cloud security," Cluster Computing, vol. 23, pp. 3057–3067, 2020.

[6] A. Vineela, N. Kasiviswanath, and C. ShobaBindu, "Theoretical analysis on applications aspects of smart materials preserving the security and privacy in medical big data and cloud," Materials Today: Proceedings, vol. 81, pp. 977–982, 2023.

[7] U. Narayanan, V. Paul, and S. Joseph, "A novel system architecture for secure authentication and data sharing in cloud enabled big data environment," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 6, pp. 3121–3135, 2022.

[8] J. Chen, L. Ramanathan, and M. Alazab, "Holistic big data integrated artificial intelligent modeling to improve privacy and security in data management of smart cities," Microprocessors and Microsystems, vol. 81, p. 103722, 2021.

[9] 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, vol. 137, pp. 192–204, 2020.

[10] P. Jain, M. Gyanchandani, and N. Khare, "Enhanced secured map reduce layer for big data privacy and security," Journal of Big Data, vol. 6, no. 1, pp. 1–17, 2019.

[11] H. K. Gupta and R. Parveen, "Comparative study of big data frameworks," in 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), vol. 1. IEEE, 2019, pp. 1–4.

[12] An efficient cluster by cluster head selection approach in big data, in 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2022, pp. 1–6.

[13] J. Zhang, J. Ma, X. Li, and W. Wang, "A secure and efficient remote user authentication scheme for multi-server environments using ecc," KSII Transactions on Internet and Information Systems (TIIS), vol. 8, no. 8, pp. 2930–2947, 2014.

[14] E. Barker, W. Barker, W. Burr, W. Polk, and M. Smid, “Nist special publication 800-57,” NIST Special publication, vol. 800, no. 57, pp. 1–142, 2007.

[15] M. S. Srinath and V. Chandrasekaran, “Isogeny-based quantum-resistant undeniable blind signature scheme,” Cryptology ePrint Archive, 2016.

[16] L. D. Singh and K. M. Singh, “Implementation of text encryption using elliptic curve cryptography,” Procedia Computer Science, vol. 54, pp. 73–82, 2015.