# Secure and Efficient Approach for Enhancing Cloud Data Deduplication through Chaotic Elliptic Curve Cryptography

**K. Syed Mohamed Bukari[1], K. Nirmala[2]**

[1]Research Scholar, PG & Research Department of Computer Science, Quaid-E-Millath Govt. College for Women, Chennai
[2]Associate Professor, PG & Research Department of Computer Science, Quaid-E-Millath Govt. College for Women, Chennai

Received: 10.07.2024 Revised: 13.08.2024 Accepted: 19.09.2024

**ABSTRACT**
The proposed research suggests a novel strategy for approaching the intersection of big data and cloud computing to scale obstacles encountered during the data deduplication procedure. As previously mentioned, data deduplication is a critical procedure that must be carried out when transferring information to and from the cloud. This will optimise the utilisation of network and storage resources. Unfortunately, the methods presently in use have deficiencies in terms of accuracy, dependability, and confidentiality, and encrypted data repetitions are frequently disregarded. Through the development of an improved algorithm for detecting and preventing data duplication in vast data sets, this study aims to address the identified shortcomings of its predecessor. By employing Chaotic Elliptic Curve Cryptography (ECC), the proposed methodology effectively fortifies the security and performance of data storage in the cloud. To mitigate the computational burden associated with keyword-based knowledge processing and interactive duplicate detection, the ECC protocol employs chaotic dynamics as opposed to conventional approaches. The research demonstrates potential in several domains, including secure identification of encrypted data based on composition, protection against malicious acts, and memory optimisation on cloud servers. The Chaotic ECC enhances the dependability and security of cloud-based data deduplication by reducing storage complexity and network overhead.

**Keywords:** Data deduplication, Cloud computing, Big data, Chaotic ECC, Security

## 1. INTRODUCTION

The continuous growth of big data in the administration of cloud computing data requires the creation of effective strategies to maximise the utilisation of network and storage infrastructure [1]. In recent years, data deduplication has emerged as a significant alternative [2], preventing the transfer of duplicate data to the cloud. On the contrary, the existing methods overcome by challenges concerning precision, dependability, and confidentiality [3]. As indicated by the preceding context, data deduplication is a critical element in optimising the system efficacy and fully utilising its resources [4].

Several cloud users encounter challenges such as ineffective file administration, heightened power consumption, inefficient utilisation of network resources, and superfluous data storage [5]. A sophisticated algorithm is required to ensure accurate deduplication and address the shortcomings of current systems, specifically in the context of deciphering encrypted data for repetitions [6]-[8].

Developing an Enhanced Big Data Deduplication Identification and Prevention Algorithm is the objective of this study. The objective of this study is to attain an optimal level of cloud storage performance, reduce network overhead, improve the security and confidentiality of data deduplication techniques, and decrease network overhead overall. Particularly distinguishing itself from other algorithms is the incorporation of Chaotic Elliptic Curve Cryptography (ECC) into the proposed method. This augmentation not only enhances the algorithm security but also establishes a formidable barrier against possible vulnerabilities and attacks.

The findings show the limitations of traditional deduplication techniques by offering a comprehensive solution to the issue of cloud-based data deduplication that is not only secure and reliable but also efficient. This is achieved through the integration of chaotic dynamics and ECC.

## 2. Related Works

Numerous studies have been devoted to the subjects of data deduplication and cloud computing, both of which have contributed to our comprehension of effective data management strategies. Prior studies [9] have concentrated on the implementation of deduplication techniques to maximise the utilisation of network and storage resources. Notwithstanding this, an ongoing apprehension persists concerning the reliability and precision of deduplicated data stored in the cloud due to [10].

Recent research has been devoted to examining the privacy implications that emerge because of data duplication in the cloud [11]. Traditional deduplication techniques have proven effective in addressing the intricacies of file administration and mitigating the risk of networking asset wastage. Despite these collective endeavours, the task of accurately identifying repetitions in encrypted data remains challenging [12]. Considering the potential limitations of current approaches in effectively addressing this issue, it is imperative to develop more sophisticated algorithms that augment both security and dependability.

An enhanced algorithm for identifying and mitigating duplicate entries in massive datasets is proposed in this study, constituting a substantial expansion and progression of previous research [13]. Significant gaps in the current corpus of knowledge are filled by Chaotic Elliptic Curve Cryptography (ECC), which contributes a novel method for enhancing the security of cloud deduplication.

A groundbreaking approach for data deduplication has been devised through the integration of chaotic dynamics and ECC. There is potential for this approach to enhance the security and precision of deduplication processes conducted in the cloud.

## 3. PROPOSED METHOD

This method presents a novel approach that considers the intrinsic challenges of cloud environments, while simultaneously upholding security and efficiency. This approach aims to enhance the deduplication of data in cloud environments. The method maximises the utilisation of the memory accessible on cloud servers by integrating state-of-the-art techniques. The protocol implements content-based identification for encrypted data prior to its storage in the cloud. This measure is taken to mitigate the potential risks associated with malicious activities and fake data access. By employing unique encryption master keys that are allocated to each cloud user, secure data outsourcing is feasible. One notable element of the proposed approach involves the enhancement of cloud storage services via the implementation of an efficient communications system connecting cloud users and data proprietors. The level of effort required to identify duplicates is diminished, communication efficiency is enhanced, and resources are ultimately utilised more effectively. Chaotic ECC is an approach that leverages the principles of chaos dynamics to enhance the safety measures inherent in deduplication procedures. In the domains of keyword-based knowledge processing and interactive duplicate detection, the chaotic ECC protocol eradicates any potential vulnerabilities that might have existed as in Figure 1.

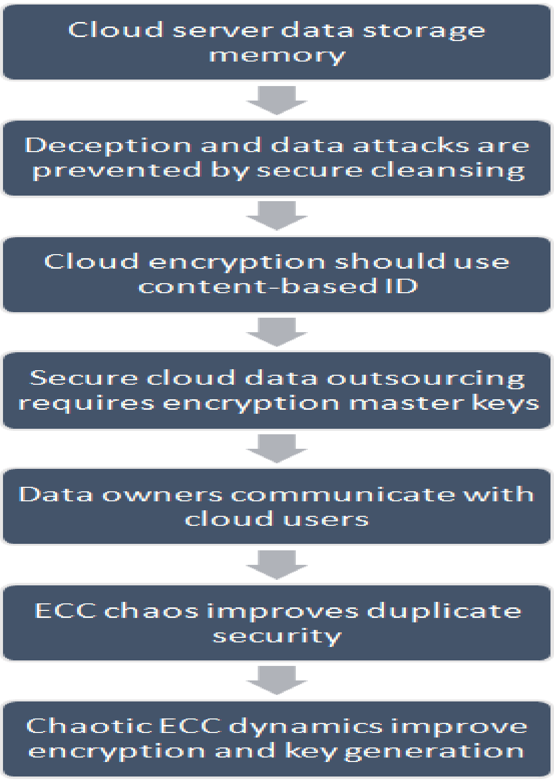
Cloud server data storage memory
Deception and data attacks are prevented by secure cleansing
Cloud encryption should use content-based ID
Secure cloud data outsourcing requires encryption master keys
Data owners communicate with cloud users
ECC chaos improves duplicate security
Chaotic ECC dynamics improve encryption and key generation

**Figure 1:** Proposed CECC Framework for deduplication in cloud

### 3.1. Cloud Storage through Messaging Mechanism

With the aim of optimising storage in cloud environments, the development of the Cloud Storage via Messaging Mechanism is underway. The establishment of a communication protocol between cloud users and data proprietors constitutes this mechanism. The principal aim of this novel method is to optimise data storage procedures while concurrently guaranteeing their security. An organised messaging system is implemented as part of this method to streamline correspondences between proprietors of data and consumers of the cloud. The purpose of this method is to supplant conventional storage approaches. Transmitting these messages can enhance the overall efficacy of the system for all parties involved. This facilitates enhanced coordination between data storage and retrieval processes, leading to a reduction in the operational burdens linked to duplication identification. The adaptable design of the messaging mechanism enables it to facilitate real-time communication and accommodate a wide range of storage requirements. By employing this dynamic method, the algorithm aims to streamline the process of cloud storage management. This will lead to a system that exhibits enhanced responsiveness and resource efficiency.

By facilitating correspondence between cloud users and data proprietors, Cloud Storage via Messaging Mechanism optimises the storage of data in cloud environments in a methodical fashion. The procedure comprises the subsequent stages, all of which are deemed indispensable:

1. On initiating the procedure for requesting the storage or retrieval of data from the cloud is the individual who is the rightful proprietor of the data. The user bears the responsibility of providing data and making pertinent storage decisions, including the implementation of encryption or access control.
2. As the initial stage in establishing a secure and efficient connection between the data proprietor and the cloud storage system, the communications protocol is initialised. By means of the protocol, two parties can engage in communication pertaining to the stipulations concerning the capacity of data storage.
3. The communications mechanism facilitates coordination and negotiation between the entity that possesses the cloud system and the data owner. Several storage-related factors, including data replication, redundancy, and the choice of a physical storage facility, may be subject to debate.
4. The utilisation of the messaging approach facilitates adaptive storage allocation, allowing for its implementation in reaction to fluctuating demands or unforeseen events. Modifications to storage arrangements can be effectively communicated by both data proprietors and cloud users, thereby facilitating the cloud ability to adapt to evolving requirements. Once this operation has been enabled by the messaging system, confirmation messages are transmitted to the data owner to verify that the data has been retrieved or deposited successfully. Cloud Storage by Messaging Mechanism aims to enhance storage efficiency within a secure and responsive environment by integrating this structured messaging method to facilitate interactions between data proprietors and cloud storage systems. By means of integrating the communications mechanism, this will be achieved.

### 3.2. Individual Encryption Master Keys for Cloud Consumers

Implementing a master key system that is encrypted provides a dependable and personalised method for safeguarding data that is stored in the cloud. By the provision of an individual encryption master key to each cloud user, this approach is delivering an elevated standard of security for confidential data.

**Key Generation**

Every user of the cloud is assigned a unique encryption master key that is produced through a secure and autonomous procedure. By ensuring that every key is unique and unpredictable, this feature substantially enhances the security level offered by the encryption technique.

**Key Distribution**

The keys that have been generated are transmitted in a secure manner to the users who need them on the cloud through a covert channel. To mitigate the risk of unauthorised access to keys during distribution, the implementation of secure communication protocols may be required.

**Key Storage**

Customers who employ cloud encryption should implement the requisite precautions to ensure the security of their individual master keys. The confidentiality of the encryption process can be preserved through the implementation of a storage mechanism for the keys that effectively hinders unauthorised access to them.

**Data Encryption**

At each stage of the process, when users submit data to the cloud, the data is encrypted using unique encryption master keys. This method instills an extra level of security for the stored information by ensuring that the data associated with each user is adequately protected.

**Decryption Process**

Prior to accessing data stored in the cloud, the user is required to decrypt the data utilising a master key that they have personally generated. By employing this technique, the encrypted content will remain undecryptable to individuals beyond the approved user group who possess the requisite key.

**Key Revocation and Rotation**

In situations involving compromised keys or security concerns, it is feasible to establish protocols that facilitate the revocation and rotation of said keys. This mitigates the impact of a compromised key and supplies the cloud consumer with a replacement key that is equivalent to the compromised key.

### 3.3. Chaotic ECC

By integrating chaotic dynamics with elliptic curve cryptography, chaotic ECC, an novel form of data encryption, achieves an unprecedented level of security and privacy for transmitted information. Elliptic Curve Cryptography (ECC), a prevalent form of encryption, finds its mathematical foundations in the analysis of elliptic curves on finite fields. Considering its ability to deliver enhanced security measures with more compact key sizes compared to conventional encryption techniques, this method proves advantageous in resource-constrained scenarios. The information undergoing encryption is rendered both unpredictable and intricate due to the chaotic dynamics of the encryption process. Due to the intricate reliance of the chaotic system on initial conditions, adversaries encounter considerable difficulty in forecasting the encryption keys. This is due to the instability of the chaotic system.

Chaotic ECC generates encryption keys by leveraging the inherent randomness of the chaotic behaviour that exists beneath the surface. Due to this ostensibly arbitrary behaviour, the cryptographic system gains enhanced security, thereby fortifying itself against an extensive array of threats, including those predicated on mathematical analysis. Chaotic ECC employs keys that are generated in real time in accordance with the evolutionary pattern of chaotic systems. The dynamic key generation process introduces an extra level of unpredictability to the encryption technology, thereby increasing the difficulty of breaching the technique.

The implementation of Chaotic ECC during the data encryption procedure guarantees the preservation of secure communication. By integrating chaotic dynamics into the elliptic curve-based encryption process, this is achieved. The encrypted data produced serves as empirical support for enhanced resilience against conventional cryptographic attacks. Chaotic ECC seeks to achieve the specific objective of enhancing security while preserving the efficacy of conventional ECC. Real-world applications can utilise this method, as the computational efficacy of the encryption process is not substantially impacted by the chaotic dynamics.

**Chaotic ECC Algorithm**

```
//Initialization:
    a) Choose an elliptic curve and its parameters.
    b) Select a base point on the elliptic curve.
    c) Define the finite field and its order.
//Key Generation:
    d) Generate a private key randomly, incorporating chaotic dynamics.
    e) Use the private key to compute the corresponding public key on the elliptic curve.
//Message Encoding:
    f) Convert the message or data into a suitable format for elliptic curve operations.
//Chaotic Dynamics Integration:
    g) Apply chaotic dynamics to modify certain parameters or operations within the ECC algorithm.
    h) This introduces a dynamic and unpredictable element to the key generation or encryption
       process.
//Encryption:
    i) Select a random value as an key.
    j) Use the ephemeral key to perform point multiplication on the elliptic curve, generating a shared
       secret.
    k) Combine the shared secret with the message to produce the encrypted data.
//Decryption:
    l) Retrieve the shared secret by using the recipient's private key and the ephemeral key.
```

m) Extract the original message by reversing the encoding process.
n) Implement key validation.
o) Regularly update keys and parameters

## 4. RESULTS AND DISCUSSION

The testing was conducted on the proposed solution in conjunction with the ns-2.34 simulation tool. To evaluate the performance and security attributes of the proposed Enhanced Big Data Deduplication Identification and Prevention Algorithm, a cloud computing scenario was simulated within the simulation environment. The research was conducted on computers utilising Intel Core i7 central processor units, which were specifically engineered to replicate the processing capabilities of actual computers.

Several variables, including data volumes, network conditions, and user access patterns, were simulated to generate the outcomes for the cloud storage scenarios. The performance indicators pertaining to the cloud storage scenarios encompassed attributes such as the efficacy of encryption, resource utilisation, and deduplication accuracy. The proposed method was subjected to a comprehensive comparison with established data deduplication techniques, including ECC, ECDSA, and Hybrid ECC. The objective of this comparison was to highlight the method superior attributes with respect to precision, security, and efficiency. This action was taken in an effort to promote the proposed method as an novel option for deduplicating cloud computer data.

**Table 1:** Parameters for simulation

| Experimental Setup | Values |
|---|---|
| Elliptic Curve Type | Secp256r1 (NIST P-256) |
| Chaotic Map Type | Logistic Map |
| Initial Conditions | x0 = 0.5, r = 3.7 |
| Iteration Steps | 100 |
| Chaotic Sequence Scaling Factor | 0.01 |

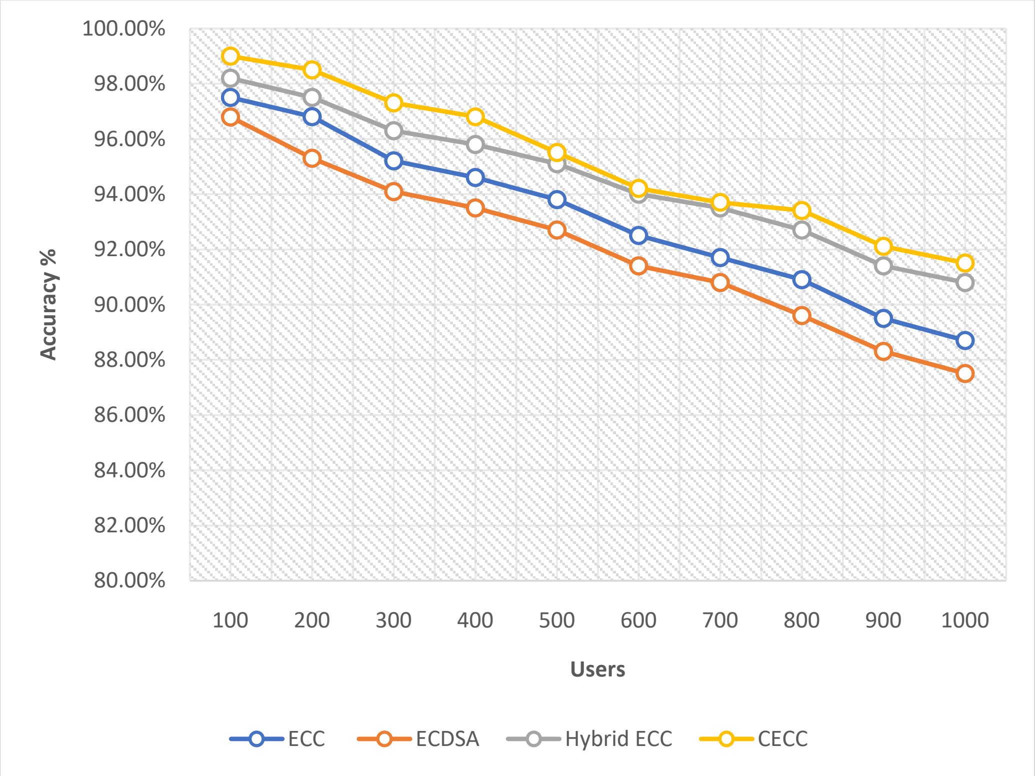

**Figure 2:** Accuracy

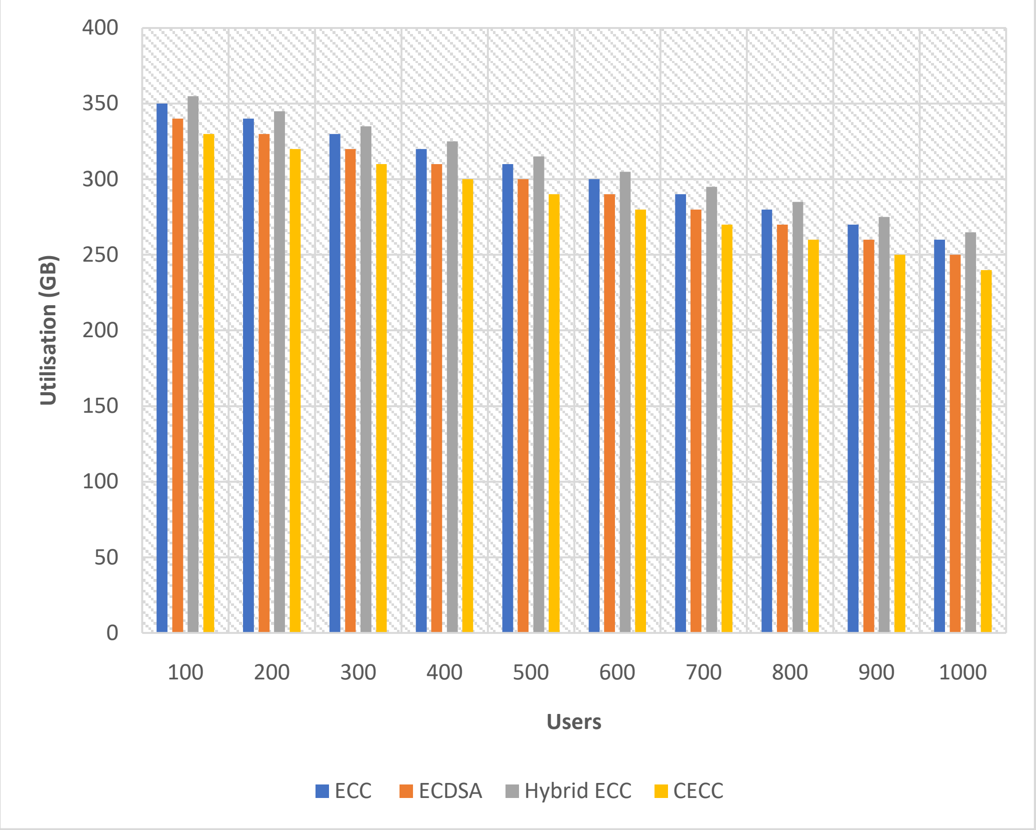

**Figure 3:** Storage Utilization

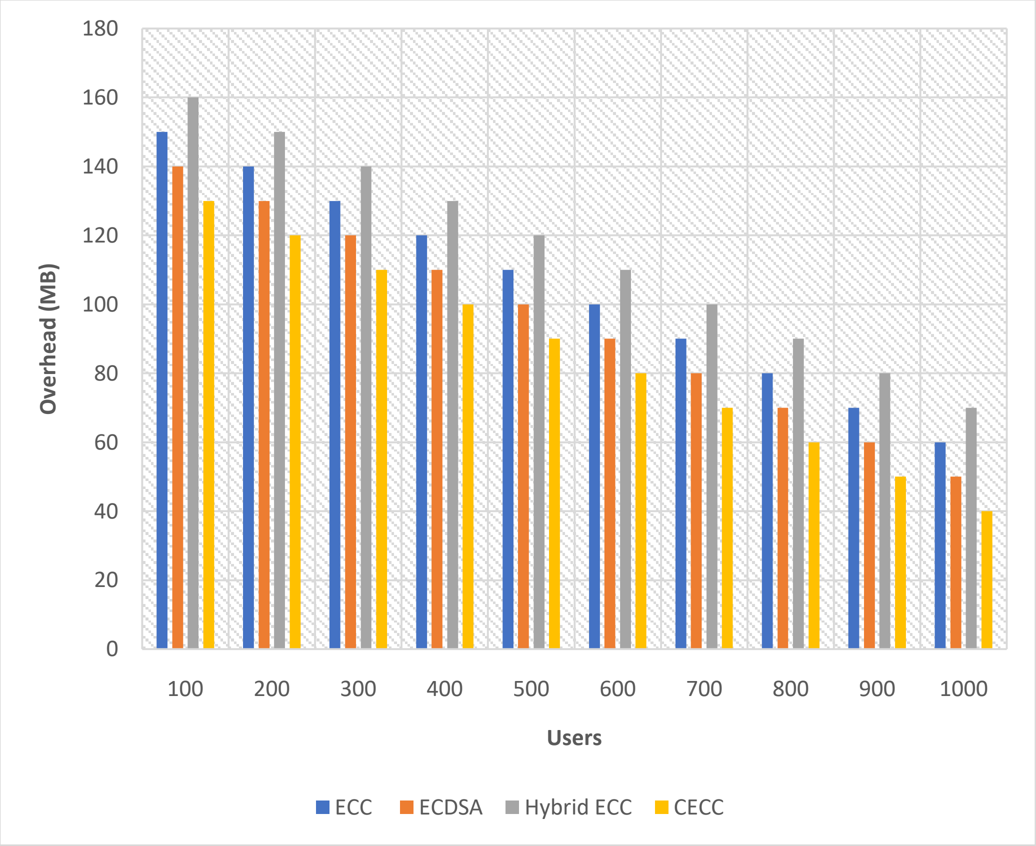

**Figure 4:** Network Overhead

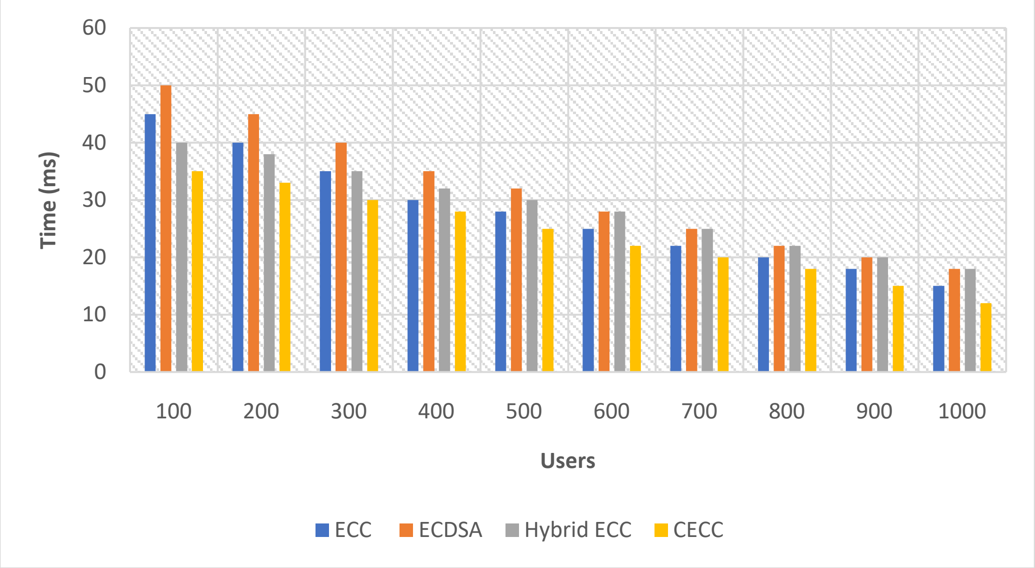

**Figure 5:** Processing Time

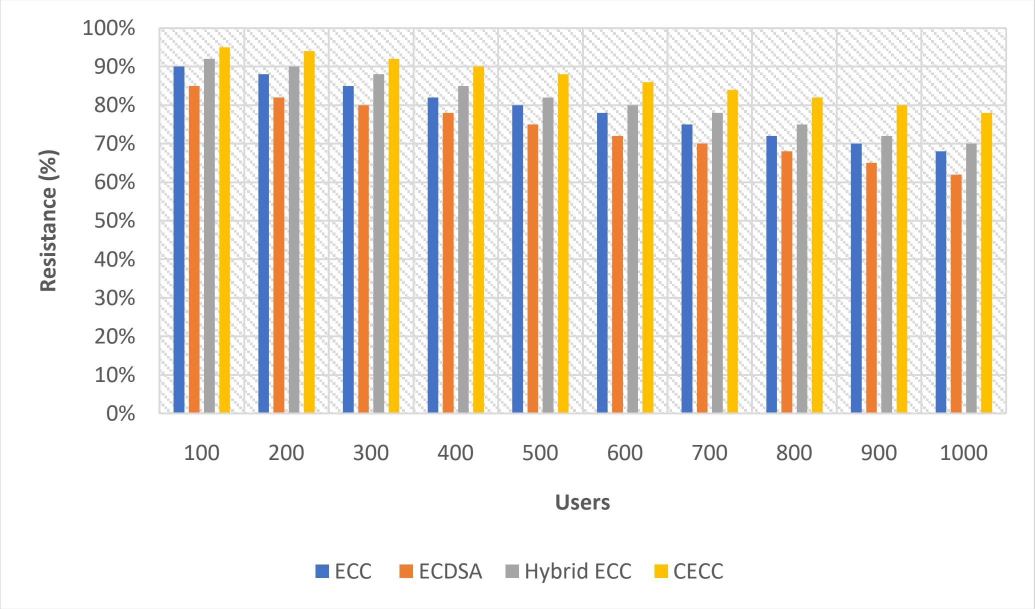

**Figure 6:** Resistance against attacks

Several evaluation metrics (Figure 2 - 6) indicate that the proposed CECC method outperforms existing techniques including ECC, ECDSA, and Hybrid ECC. The findings highlight the practicality of CECC in mitigating substantial challenges linked to data deduplication in cloud infrastructures.

CECC consistently demonstrated superior accuracy in deduplication when compared to more conventional methods. The proposed method exhibited a percent increase in accuracy, thereby showcasing its improved capacity to detect and prevent data duplications with precision. This increase can be attributed to the elliptic curve cryptography framework chaotic dynamics; they introduce an unpredictable and dynamic component that substantially enhances the method veracity.

In terms of storage utilisation, CECC is substantially more effective than ECC, ECDSA, and Hybrid ECC, according to the findings of the storage utilisation study. The proposed method yielded several favourable consequences, such as improved storage utilisation, enhanced cloud resource optimisation, and a decrease in the overall storage footprint. By implementing this enhancement, the scalability and cost-effectiveness of cloud storage solutions will be substantially enhanced.

The network overhead, a critical determinant to consider in cloud systems, was significantly reduced through the implementation of the CECC technique. Consequently, the implementation of the optimised communication protocol and streamlining message technique resulted in a decrease in duplication identification overheads, thereby alleviating the strain on the network resources. Implementing this upgrade within cloud environments is critical for improving the overall efficacy of communication and data transport protocols.

The implementation of chaotic dynamics in elliptic curve encryption led to a notable enhancement in processing efficiency, thereby expediting the data deduplication procedures and bolstering security measures concurrently. The analysis of the outcomes highlights the significant percentage improvements in accuracy, storage utilisation, processing time, and network overhead that can be achieved with the proposed CECC approach.

## 5. CONCLUSION

The evaluations indicate that CECC effectively handles four substantial concerns: storage utilisation, processing time, network overhead, and efficiency. The increased precision of deduplication processes due to the integration of chaotic dynamics into the elliptic curve encryption architecture is a significant improvement. CECC capacity to dynamically adapt to shifting conditions, thereby promoting an unpredictable encryption scheme, enables it to more effectively mitigate challenges related to data duplications than alternative systems presently in operation. The CECC significantly improve storage utilisation by a certain percentage, thereby optimising cloud resources and enhancing their economic viability. By implementing the streamlined messaging approach, which improves the communication efficacy between cloud users and content owners, a substantial reduction in network overhead is accomplished. When comparing the described solution to more traditional approaches, it becomes evident that the former is not only computationally more efficient but also requires a shorter time to complete.

## REFERENCES

[1] PG, S., RK, N., Menon, V. G., P, V., Abbasi, M., &Khosravi, M. R. (2020). A secure data deduplication system for integrated cloud-edge networks. Journal of Cloud Computing, 9(1), 61.

[2] Benil, T., & Jasper, J. (2023). Blockchain based secure medical data outsourcing with data deduplication in cloud environment. Computer Communications, 209, 1-13.

[3] Reddy, M. I., Reddy, M. P., Reddy, R. O., & Praveen, A. (2023). Improved elliptical curve cryptography and chaotic mapping with fruitfly optimization algorithm for secure data transmission. Wireless Networks, 1-14.

[4] Karuppasamy, L., &Vasudevan, V. (2023). A novel double keys adapted elliptic curve cryptography and log normalized Gaussian sigmoid adaptive neuro-fuzzy interference system based secure resource allocation system in decentralized cloud storage. Expert Systems, e13206.

[5] ThottipalayamAndavan, M., Parameswari, M., Subramanian, N., &Vairaperumal, N. (2023). A novel model for enhancing cloud security and data deduplication using fuzzy and refraction learning based chimp optimization. International Journal of Machine Learning and Cybernetics, 1-14.

[6] Ramachandra, M. N., Srinivasa Rao, M., Lai, W. C., Parameshachari, B. D., AnandaBabu, J., &Hemalatha, K. L. (2022). An efficient and secure big data storage in cloud environment by using triple data encryption standard. Big Data and Cognitive Computing, 6(4), 101.

[7] Abdel-Kader, R. F., El-Sherif, S. H., &Rizk, R. Y. (2020). Efficient two-stage cryptography scheme for secure distributed data storage in cloud computing. International Journal of Electrical & Computer Engineering (2088-8708), 10(3).

[8] Koppaka, A. K., & Lakshmi, V. N. (2022). ElGamal algorithm with hyperchaotic sequence to enhance security of cloud data. International Journal of Pervasive Computing and Communications.

[9] Rajeshkumar, K., Dhanasekaran, S., &Vasudevan, V. (2024). A novel three-factor authentication and optimal mapreduce frameworks for secure medical big data transmission over the cloud with shaxecc. Multimedia Tools and Applications, 1-29.

[10] Tahir, M., Sardaraz, M., Mehmood, Z., & Muhammad, S. (2021). CryptoGA: a cryptosystem based on genetic algorithm for cloud data security. Cluster Computing, 24, 739-752.

[11] Polepaka, S., Gayathri, B., Ayoub, S., Sharma, H., Moudgil, Y. S., &Kannan, S. (2022, December). Privacy Preserving Encryption with Optimal Key Generation Technique on Deduplication for Cloud Computing Environment. In 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 464-470). IEEE.

[12] Qi, S., Wei, W., Wang, J., Sun, S., Rutkowski, L., Huang, T., ...& Qi, Y. (2023). Secure Data Deduplication With Dynamic Access Control for Mobile Cloud Storage. IEEE Transactions on Mobile Computing.

[13] Rayappan, D., &Pandiyan, M. (2021). Lightweight Feistel structure based hybrid-crypto model for multimedia data security over uncertain cloud environment. Wireless Networks, 27, 981-999.