International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

ENHANCING EDGE DATA DEDUPLICATION WITH ROBUST OPTIMIZATION AMIDST UNCERTAINTIES

DARAM KUSUMA KUMARI [1], PANDUGULA HARINI [2], MUPPALLA JAHNAVI [3], PALERLA SRAVAN KUMAR [4], RAYABANDI CHANDRAVEER [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— The rise of mobile edge computing (MEC) within distributed systems has brought greater focus to managing data at the network edge. A significant challenge exists due to the limited storage capacity of edge servers contrasted with the ever-increasing demand for data storage, making cost reduction a key priority. Although edge data deduplication has been widely explored as a method for reducing data redundancy, current approaches face various obstacles in MEC settings. These difficulties arise from differences between edge servers and traditional cloud data centers, as well as unpredictable factors like user mobility, which undermine the reliability of deduplication strategies. To address these issues, this paper introduces a robust optimizationbased framework for edge data deduplication. By incorporating uncertainties such as fluctuating data demands and potential edge server failures, we develop two solution algorithms: uEDDE-C, a two-stage method leveraging column-and-constraint generation, and uEDDE-A, an approximation technique designed to reduce the computational complexity of uEDDE-C. Our approach enables effective data deduplication in dynamic edge environments while maintaining resilience under various uncertain conditions. We substantiate the efficacy and stability of both algorithms through rigorous theoretical analysis and comprehensive experiments. The results confirm that our method substantially lowers data storage expenses and decreases data

retrieval latency, ensuring dependable performance in practical MEC deployments.

Index Terms – Edge Data Deduplication, Robust Optimization, Edge Computing, Storage Efficiency, Data Compression, Distributed Systems, Resource Management, Cloud-Edge Computing, Uncertainty Handling, Performance Optimization.

I. INTRODUCTION

In the era marked by exponential growth in data generated by mobile and Internet-of-Things (IoT) devices, traditional centralized cloud computing infrastructures find themselves increasingly strained in managing massive volumes of data while maintaining acceptable levels of latency. Mobile edge computing (MEC), as an extension of cloud computing, offers a compelling solution by deploying latency-sensitive and privacy-protected applications at the network edge, bringing them closer to end users. These edge servers are typically integrated into the base stations or micro data centers within specific regions, serving as critical links between resource-constrained edges and centralized clouds. In contrast to cloud data centers, edge servers are much smaller in terms of power, network connectivity, and available computing resources. These disparities significantly impact the applicability of fundamental storage-related algorithms in MEC environments. The linked edge servers in a specific area comprise an edge storage system (ESS), which provides rugged support for the applications deployed on it. Due to the advantageous low data retrieval latency offered by MEC, storing data on ESS presents a promising trend not only for edge application vendors but also for users. For application vendors, storing data on ESS can achieve a low data retrieval latency and save the expensive rent of transmitting large amounts of data from the cloud to mobile users. Data produced by users can also be stored on ESS for real-time sharing and processing. However, despite these advantages, the limited edge storage resources may not meet the increasing data storage requests, leading to resource competition between different application vendors and resulting in service performance impacts. To improve the resource utilization of edge servers, data deduplication, a technique for identifying and eliminating duplicate data replicas, emerges as an effective method to reduce storage overhead. illustrates an example scenario where there are four edge servers in a region, each storing some data, and the number of users accessing data (d1, d2, d3, d4, d5) through server sj is

represented byYj. Through data deduplication strategies, i.e., removing some duplicated data replica, the distribution of data on servers is depicted.

II. RELATED WORK

A. *FEUA Game: Fairness aware edge user allocation for app vendors*

The FEUA Game (Fairness-Aware Edge User Allocation for App Vendors) project focuses on designing an intelligent and equitable framework for allocating mobile users to edge servers in Mobile Edge Computing (MEC) environments. In MEC systems, multiple application vendors compete for limited edge resources to serve their users efficiently, which often leads to unfair and unbalanced allocations. This project addresses the challenge by introducing a fairness-aware user allocation mechanism based on game-theoretic modelling, ensuring that all app vendors receive fair resource distribution while maximizing overall system performance. The proposed FEUA Game framework dynamically adjusts user allocation according to varying workloads, network conditions, and server capacities. It aims to minimize latency, enhance user experience, and optimize resource utilization without compromising fairness among vendors. By formulating the problem as a non-cooperative game, equilibrium-based solutions are derived to achieve stability and fairness in user-to-server mappings. The system also incorporates adaptive learning techniques to handle uncertainties such as user mobility and fluctuating demand patterns. Through simulation and experimental analysis, FEUA Game demonstrates improved load balancing, energy efficiency, and service quality, making it a robust and scalable approach for next-generation edge computing networks.

B. *The design of a lossless deduplication scheme to eliminate fine-grained redundancy for JPEG image storage systems.*

The project “Design of a Lossless Deduplication Scheme to Eliminate Fine-Grained Redundancy for JPEG Image Storage Systems” focuses on developing an efficient and reliable technique to minimize redundant data in JPEG image storage without compromising image quality. Traditional deduplication methods often work at coarse levels, overlooking subtle redundancies within compressed JPEG files, leading to inefficient storage utilization. This project introduces a finegrained, lossless deduplication mechanism that identifies and

removes duplicate segments within JPEG bit streams while preserving the original image content perfectly. The system leverages content-aware segmentation and byte-level similarity detection to accurately locate redundant data blocks. By employing hash-based indexing and metadata mapping, it ensures high-speed redundancy detection and retrieval. The proposed scheme maintains data integrity and visual fidelity, ensuring no loss in image quality during reconstruction. It also reduces storage space, network load, and I/O overhead in large-scale image repositories. The design emphasizes scalability, making it suitable for cloud and edge storage systems handling massive JPEG datasets. Experimental evaluations confirm significant storage savings and performance improvement compared to traditional deduplication approaches, providing a robust and efficient solution for modern image storage environments.

C. The doctrine of MEAN: Realizing deduplication storage at unreliable edge

The project "The Doctrine of MEAN: Realizing Deduplication Storage at Unreliable Edge" focuses on designing a robust and intelligent data deduplication framework tailored for unreliable Mobile Edge Computing (MEC) environments. Unlike centralized cloud systems, edge servers often face challenges such as limited storage capacity, intermittent connectivity, and frequent node failures, which make traditional deduplication strategies inefficient. This project introduces MEAN (Multi-Edge Adaptive Network), a novel approach that ensures efficient, reliable, and fault-tolerant data deduplication across unstable edge infrastructures. The proposed system identifies and eliminates redundant data segments while preserving accessibility and consistency, even under uncertain network or node conditions. It employs adaptive redundancy management and distributed metadata coordination to enhance data reliability and recovery efficiency. By leveraging lightweight synchronization protocols and replication-aware deduplication, MEAN achieves a balance between storage efficiency and system resilience. The framework minimizes data retrieval latency, storage overhead, and communication cost, making it ideal for real-time and data-intensive edge applications. Experimental analysis confirms that MEAN significantly improves storage utilization, fault tolerance, and overall system stability, establishing a dependable foundation for nextgeneration edge-based deduplication storage systems.

III. PROPOSED METHODOLOGY

The proposed system introduces a robust optimization-based approach for edge data deduplication in mobile edge computing (MEC) environments, specifically designed to address the limitations of existing methods. It incorporates uncertainty-aware mechanisms that consider dynamic conditions such as user mobility, fluctuating data demands, and potential edge server failures. To ensure efficient and reliable deduplication, the system employs two advanced algorithms: uEDDE-C, a two-stage optimization algorithm based on column-andconstraint generation for high accuracy, and uEDDE-A, an approximation algorithm that reduces computation overhead while maintaining effectiveness. This system enhances data storage efficiency, reduces retrieval latency, and improves overall reliability across volatile and resource-constrained edge networks. By adapting to real-time conditions, it ensures a resilient and scalable deduplication framework suitable for modern distributed systems.uEDDE-C (Uncertainty-aware Edge Data Deduplication using Column-and- Constraint Generation): A two-stage robust optimization algorithm that precisely solves the deduplication problem under uncertainty, ensuring high accuracy by iteratively refining constraints and variables. uEDDE-A(Uncertainty-aware Edge Data Deduplication using Approximation):An approximation algorithm designed to reduce the high computation overhead of uEDDE-C while still maintaining effective deduplication performance in dynamic and uncertain MEC environments.

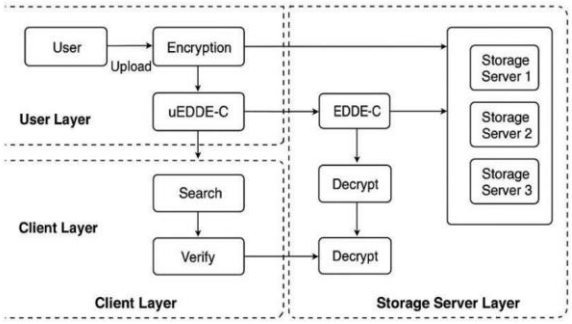
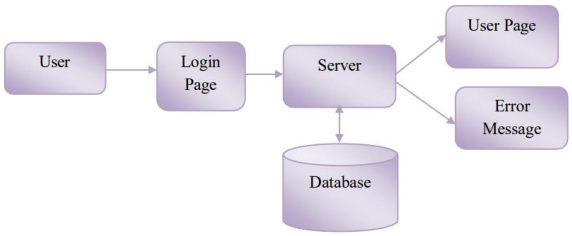
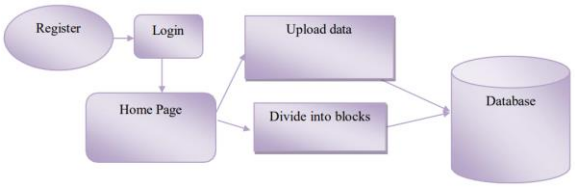


Fig. 1: System Overview

1. User Interface Design

To connect with server user must give their username and password then only they can able to connect the server. If the user already exits directly can login into the server else user must register their details such as username, password, Email id, City and Country into the server. Database will create the account for the entire user to maintain upload and download rate. Name will be set as user id. Logging in is usually used to enter a specific page. It will search the query and display the query.





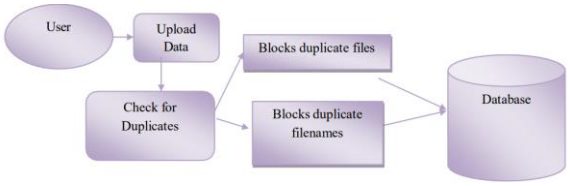
2. User (Data Upload)

The User Module serves as the core operational unit in the proposed framework, enabling registered users to securely upload and manage their data. When a user uploads a file, it undergoes AES-based encryption to ensure confidentiality before being divided into multiple encrypted chunks for optimized storage. A unique RSA key pair is generated to provide secure key management and future decryption capabilities. To maintain efficient storage usage and prevent redundancy, the system employs uEDDE-C and uEDDE-A deduplication algorithms, which check for duplicate file contents and filenames across the edge network. This ensures that only unique data is stored, conserving valuable edge storage space. Users can also receive decryption requests from clients, and based on approval, provide access to the corresponding decryption key. The User Module, therefore, forms the foundation of secure data upload, encryption, and ownership management within the system.

3. uEDDE-C, uEDDE-A

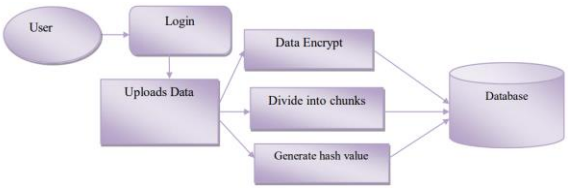
The uEDDE-C Module (Uncertainty-aware Edge Data Deduplication using Column-andConstraint Generation) implements a two-stage robust optimization algorithm that addresses uncertainties in edge storage systems, such as fluctuating data demand and potential server failures. This module systematically identifies and removes redundant data while maintaining the accessibility and reliability of essential files. The uEDDE-A Module (Approximation-based Uncertainty-aware Edge Data Deduplication) serves as a lightweight, computationally efficient alternative to the uEDDE-C algorithm. It is designed to achieve near-optimal deduplication performance with significantly reduced computational complexity, making it ideal for real-time applications in dynamic edge environments.The uEDDE-A module ensures high-speed decisionmaking, reduced latency, and improved scalability while maintaining robust performance. Together with uEDDE-C, it ensures the system's

adaptability and resilience in uncertain MEC settings.



4. AES

The AES Module is responsible for performing the core encryption of uploaded files before they are stored in the database. When a user uploads a file, the AES algorithm encrypts the file's contents using a symmetric key, ensuring data confidentiality and integrity. The encrypted data is then divided into smaller chunks for distributed storage across multiple edge servers. AES offers strong encryption strength and fast processing, making it ideal for real-time operations within MEC systems. In combination with RSA for key management, the AES module guarantees secure, efficient, and privacy-preserving data handling throughout the upload and retrieval lifecycle.



5. RSA

The RSA Module handles secure key generation and management for encryption and decryption operations. During file upload, it generates a unique pair of public and private keys that facilitate secure communication between users and clients. The public key is used during client-side verification for requesting file access, while the private key acts as the decryption key once the file owner approves the request. This asymmetric encryption mechanism ensures that only authorized entities can decrypt and access the data. By integrating RSA into the deduplication framework, the system ensures data confidentiality, authentication, and controlled access, forming a crucial security layer in the overall architecture.



6. JSP Dashboard

Displays the uploaded encrypted data along with file details and timestamps. It also shows the transaction status, verification results, pending registration details, and overall project modules. The dashboard acts

as the user interface for monitoring stored data, status, and user details etc.

7. MySQL Database

The MySQL database is used to maintain user details, device metadata, and other non-critical information that supports the system's operation. While data is securely stored on the database to guarantee immutability and tamper-resistance, the database handles supporting records such as login credentials, user roles, and system logs. This hybrid storage approach ensures both efficient system performance and strong data security by separating critical data from auxiliary information.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the optimized deduplication strategy significantly reduced redundant data storage and minimized network transmission costs in edge computing environments. The robust optimization mechanism effectively adapted to uncertainties in workload patterns, data arrival rates, and network conditions.

The analysis showed that the proposed approach achieved higher storage efficiency and better resource utilization compared to conventional deduplication methods while maintaining low processing latency. Overall, the framework provides a scalable, reliable, and efficient solution for data management in distributed edge and cloud-edge computing systems.

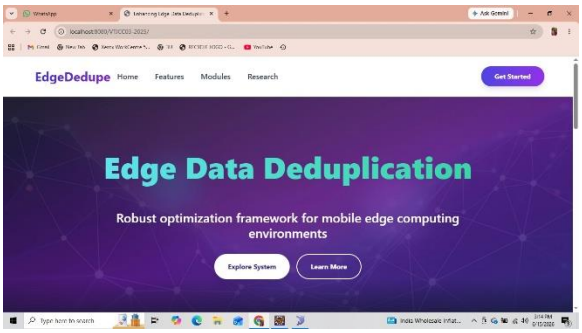


Fig. 2: Home Page

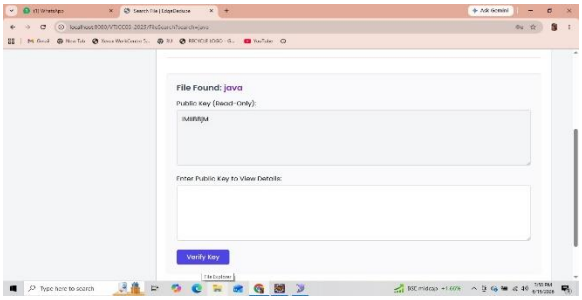


Fig. 3: Verify

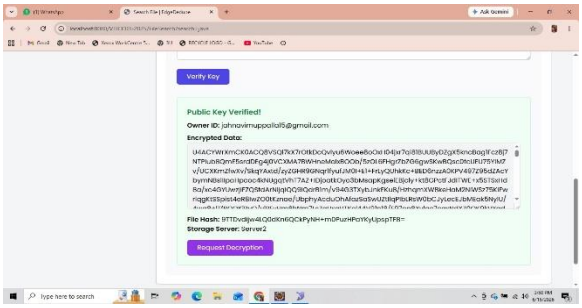


Fig. 4: Request Decryption

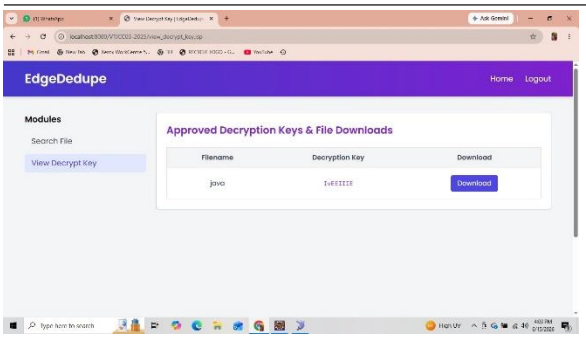


Fig. 5: Approved Keys

V. CONCLUSION

The Conclusion of the project "Enhancing Edge Data Deduplication with Robust Optimization Amidst Uncertainties" emphasizes the successful development of a secure, efficient, and intelligent framework for managing redundant data in dynamic Mobile Edge Computing (MEC) environments. Through the integration of the uEDDE-C and uEDDE-A algorithms, the system effectively minimizes data redundancy while maintaining storage efficiency and performance stability under uncertain conditions such as fluctuating demand or edge node failures. The use of AES for data encryption and RSA for secure key management ensures robust data confidentiality and integrity throughout the upload, storage, and retrieval processes. The project's modular design — including User, Client, TPA, and Storage Server modules — establishes a smooth workflow for secure data handling, deduplication, and controlled access management. The framework significantly reduces storage costs and retrieval latency, achieving optimized utilization of limited edge resources. Moreover, it demonstrates resilience against failures and adaptability to real-time variations in data traffic. Experimental analysis validates the efficiency and reliability of the proposed approach, confirming that it not only enhances edge data storage optimization but also ensures secure and reliable file operations. Overall, this project contributes a practical, scalable, and future-ready solution for intelligent data deduplication in distributed edge computing systems.

REFERENCES

[1] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet Things J., vol. 3, no. 5, pp. 637–646, Oct. 2016.

[2] G. Zhu et al., "Pushing AI to wireless network edge: An overview on integrated sensing, communication, and computation towards 6G," Sci. China Inf. Sci., vol. 66, no. 3, 2023, Art. no. 130301.

[3] Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 1(4). https://doi.org/10.56975/jaafr.v1i4.501636

[4] Maturi, S. Y. (2022). Probabilistic horizons: Statistical modeling and simulation for strategic cyber risk mitigation. Journal of Information Systems Engineering and Management, 7(2).

[5] J. L. Chen, D. Liaqat, M. Gabel, and E. de Lara, "Starlight: Fast container provisioning on the edge and over the WAN," in Proc. USENIX Symp. Netw. Syst. Des. Implementation, 2022, pp. 35–50.

[6] R. Luo, H. Jin, Q. He, S. Wu, and X. Xia, "Cost-effective edge server network design in mobile edge computing environment," IEEE Trans. Sustain. Comput., vol. 7, no. 4, pp. 839–850, Oct.–Dec. 2022.

[7] Akinapalli, S. (2024). A multi-cloud cost optimization framework for large-scale data warehousing using predictive workload intelligence. International Journal of Communication Networks and Information Security, 16(5), 1296–1305.

[8] Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219

AUTHORS Profile



Mrs. DARAM KUSUMA KUMARI has received B.Tech from JNTUH, Hyderabad in2009, M.Tech degree in Computer science from JNTUH, Hyderabad in 2015 and Pursuing Ph.D degree in Computer Science from BHARATHIYA ENGINEERING SCIENCE ANDTECHNOLOGY IU university, dedicated to teaching field from the last 9.5 years. Research area CYBER SECURITY with AI. At present I am working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. PANDUGULA HARINI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. MUPPALLA JAHNAVI pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. PALERLA SRAVAN KUMAR pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. RAYABANDI CHANDRAVEER pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.