ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

A SYSTEMATIC RESEARCH ON SECURE DEDUPLICATION SUPPORT SYSTEM FOR ATTRIBUTE-BASED STORAGE IN CLOUD

Vaishnavi Sunil Chaudhari[1], Prof. Dr. S. A. Vyawahare[2], Prof. Pallavi P. Rane[3]

[1]*Student, Rajarshi Shahu College of Engineering, Buldhana, Maharashtra*

[2]*Assistant Professor, Sanmati engineering College, Washim, Maharashtra*

[3]*Assistant Professor, Rajarshi Shahu College of Engineering, Buldhana, Maharashtra*

Abstract: *The contemporary challenge in cloud computing lies in reconciling two often conflicting mandates: enforcing fine-grained data confidentiality through Attribute-Based Encryption (ABE) and achieving storage efficiency via data deduplication. Standard ABE systems fail to support secure deduplication because identical plaintext files, when encrypted under different access policies, yield cryptographically distinct ciphertexts, thereby eliminating the possibility of recognizing duplicate copies. This research presents the design and architecture of an attribute-based storage system that successfully integrates secure deduplication within a hybrid cloud environment. The proposed system employs Ciphertext-Policy ABE (CP-ABE) and achieves the standard notion of semantic security for data confidentiality, a critical advantage over systems utilizing weaker security models. The core innovation is a methodology that allows for the modification of a ciphertext, transforming it from one access policy to another without revealing the underlying plaintext, thus enabling secure deduplication and substantial savings in storage space and network bandwidth.*

Keywords: Cloud Computing, Attribute, Deduplication, Security Models.

I. INTRODUCTION

Cloud computing has revolutionized data management, offering data providers the ability to outsource their information to external servers and share it securely with users possessing specific credentials. This architecture reduces data management costs and enhances productivity by allowing data access regardless of location. However, the offloading of data to third-party cloud servers brings two major, concurrent challenges: maintaining the confidentiality of sensitive data and managing the explosive growth of data volume.

Attribute-Based Encryption (ABE) has become a widely adopted solution for secure data sharing. ABE allows a data owner to encrypt a message under an access policy defined over a set of attributes, ensuring that only users whose private keys satisfy this policy can decrypt the data. This approach offers fine-grained access control by specifying policies rather than distributing decryption keys, granting data owners direct control over their information. Conversely, data deduplication is a critical technique employed by cloud service providers to save storage space and network bandwidth by eliminating redundant copies of identical data. This feature is vital for mitigating storage overhead and saving upload bandwidth.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

The fundamental design conflict arises because standard ABE is incompatible with secure deduplication.3 Since two identical files encrypted under different access policies result in two unique ciphertexts, the cloud server cannot detect the duplication. Therefore, designing a storage system that possesses both the fine-grained access control of ABE and the efficiency of secure deduplication is highly desirable.

1.1 Cryptographic Foundation and Security Requirements

The proposed system is founded on the robust security guarantees of Attribute-Based Encryption and explicitly addresses the shortcomings of conventional deduplication methods.

- ***Ciphertext-Policy Attribute-Based Encryption (CP-ABE)***
  The system utilizes Ciphertext-Policy Attribute-Based Encryption (CP-ABE), which is the preferred choice for access control in public cloud storage. In the CP-ABE model, the data owner encrypts a message M using an access structure over the universe of attributes. Only a user possessing a set of attributes that satisfies the access structure embedded in the ciphertext will be able to decrypt the message.
- ***Semantic Security Standard***
  A critical requirement for this system is achieving the standard notion of semantic security for data confidentiality. Unlike certain existing systems that define a weaker security notion, this design ensures a strong level of security against adversaries. This is typically formalized as selective IND-CPA security, where any polynomial-time adversary is proven to have at most a negligible advantage in the security game. This rigorous security framework is fundamental to protecting sensitive data outsourced to the cloud.

1.2 Objectives

- To design a secure deduplication framework that integrates Data Provider, Attribute Authority, Public Cloud, and Private Cloud for confidentiality-preserving storage.
- To develop a cryptographically strong mechanism for generating and verifying unique tags, labels, and proofs to support secure duplicate detection.
- To enhance storage efficiency by enabling the Private Cloud to perform secure, privacy-preserving tag checking without accessing plaintext data.
- To implement an attribute-based access control scheme through the Attribute Authority (AA) for fine-grained, policy-driven key issuance.

II. LITERATURE REVIEW

Periasamy et. al. states that Cloud computing enables data storage and application deployment over the internet, offering benefits such as mobility, resource pooling, and scalability. However, it also presents major challenges, particularly in managing shared resources, ensuring data security, and controlling distributed applications in the absence of centralized oversight. One key issue is data duplication, which leads to inefficient storage, increased costs, and potential privacy and security risks. To address these challenges, this study proposes a post-quantum mechanism that enhances both cloud security and deduplication efficiency.

 DOI: 10.5281/zenodo.17846700

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

The proposed SALIGP method leverages Genetic Programming and a Geometric Approach, integrating Bloom Filters for efficient duplication detection. The Cryptographic Deduplication Authentication Scheme (CDAS) is introduced, which utilizes blockchain technology to securely store and retrieve files, while ensuring that encrypted access is limited to authorized users. This dual-layered approach effectively resolves the issue of redundant data in dynamic, distributed cloud environments. Experimental results demonstrate that the proposed method significantly reduces computation and communication times at various network nodes, particularly in key generation and group operations. Encrypting user data prior to outsourcing ensures enhanced privacy protection during the deduplication process. Overall, the proposed system leads to substantial improvements in cloud data security, reliability, and storage efficiency, offering a scalable and secure framework for modern cloud computing environments [1]

Pavithra et. al. states that Cloud computing technology offers flexible and expedient services that carry a variety of profits for both societies as well as individuals. De-duplication techniques were developed to minimize redundant data in the cloud storage. But, one of the main challenges of cloud storage is data deduplication with secure data storage.To overcome the issue, we propose Boneh Goh Nissim Bilinear Attribute-based Optimal Cache Oblivious (BGNBA-OCO) access control and secure de-duplication for data storage in cloud computing in this paper. The proposed method achieves fne-grained access control with low computation consumption. We design Boneh Goh Nissim Privacy Preserving Revocable Attribute-based Encryption that reinforces attribute revocation and averts the discharge of sensitive information. Furthermore, we utilize Optimal Cache Oblivious algorithm to prevent disclosure of access patterns to hide the access patterns in cloud storage via rand pattern matching. We support updating both encrypted data and access control policies to minimize communication and computation overhead of data duplication and encryption processes concurrently. We perform secure data sharing to achieve higher data confidentiality and integrity. Finally, we conducted the extensive experiments in cloud and the results illustrated that our proposed BGNBA-OCO method is more efficient than related works. [2]

Prajapati et. al. states that the cloud storage service providers caters to the need of organizations and individuals by allowing them to store, transfer and backup their ever-increasing amount of data at low cost along with providing access to the other resources of cloud. For providing efficient data storage, cloud service providers utilize most widely employed deduplication technique as it allows storage of single instance of data and removes duplicate copies of data, thus mitigating storage overhead and saving upload bandwidth.

Clients uploading their data on cloud are most concerned about the security, integrity, privacy and confidentiality of their data. Conventional encryption usually employed to encrypt data while outsourcing it, is not recommended as it conflicts with data deduplication technique and so in most cases, Convergent Encryption (CE) and Proof of Ownership (PoW) are used to protect confidentiality and integrity of data. Several other approaches such as Provable Data Possession (PDP), Proof of Retrievability (POR), secure keyword search, DupLESS, Proof of Storage with Deduplication (PoSD),

 DOI: 10.5281/zenodo.17846700

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

Dekey, Message-Locked Encryption, Attribute Based Encryption (ABE) and Identity Based Encryption (IBE) have been researched to address client's security concerns and this paper does a literature review on such various proposed approaches for secure deduplication techniques in cloud storage. [3]

Kim et. al. states that data deduplication technology improves data storage efficiency while storing and managing large amounts of data. It reduces storage requirements by determining whether replicated data is being added to storage and omitting these uploads. Data deduplication technologies require data confidentiality and integrity when applied to cloud storage environments, and they require a variety of security measures, such as encryption. However, because the source data cannot be transformed, common encryption techniques generally cannot be applied at the same time as data deduplication. Various studies have been conducted to solve this problem. This white paper describes the basic environment for data deduplication technology. It also analyzes and compares multiple proposed technologies to address security threats.

The rapid development of information and communication technology (ICT) has induced various changes to data storage environments. In the early computing environment, punch cards, magnetic tapes, etc., were used as auxiliary memory devices. Since then, auxiliary storage has made considerable progress in terms of speed and data integration, from hard disk drives to flash memory. Consequently, it was predicted that the lack of storage space would be solved by the increase in data density. However, improved data processing technology developed have improving the productivity and quality of the data, thereby increasing its volume. Therefore, it is still important to secure the storage space available in a computing environment. Today, the primarily used auxiliary devices include hard disk drives, solid-state drives, and flash memory.

The advantages of such storage mediums are that they can store a large amount of data and are portable. However, they are always exposed to loss and failure, and the stored data can only be accessed by carrying the storage medium around. Therefore, to solve this problem, storage mediums have become lighter, more compact, and recoverable even in case of failure. However, carrying storage mediums around is still inconvenient. Cloud storage has emerged as a good alternative to this problem. [4]

Shynu et. al. states that Data redundancy is a significant issue that wastes plenty of storage space in the cloud-fog storage integrated environments. Most of the current techniques, which mainly center around the static scenes, for example, the backup and archive systems, are not appropriate because of the dynamic nature of data in the cloud or integrated cloud environments. This problem can be effectively reduced and successfully managed by data deduplication techniques, eliminating duplicate data in cloud storage systems.

Implementation of data deduplication (DD) over encrypted data is always a significant challenge in an integrated cloud-fog storage and computing environment to optimize the storage efficiently in a highly secured manner. This paper develops a new method using Convergent and Modified Elliptic Curve Cryptography (MECC) algorithms over the cloud and fog environment to construct secure deduplication systems. The proposed method focuses on the two most important goals of such

 DOI: 10.5281/zenodo.17846700

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

systems. On one side, the redundancy of data needs to be reduced to its minimum, and on the other hand, a robust encryption approach must be developed to ensure the security of the data.

The proposed technique is well suited for operations such as uploading new files by a user to the fog or cloud storage. The file is first encrypted using the Convergent Encryption (CE) technique and then re-encrypted using the Modified Elliptic Curve Cryptography (MECC) algorithm. The proposed method can recognize data redundancy at the block level, reducing the redundancy of data more effectively. Testing results show that the proposed approach can outperform a few stateof-the-art methods of computational efficiency and security levels. [5]

III. SYSTEM ARCHITECTURE

To overcome the inherent security-efficiency conflict, the secure deduplication system is built upon a hybrid cloud architecture involving four distinct, interacting entities.

The system entities and their roles are defined as follows:

- ***Data Provider (Do):*** The entity that initiates the storage request. The Do encrypts the file (F) under a chosen access structure ($\mathbb{A}$) and creates a unique tag (T) and a label (L) associated with the data.3 The Do also generates a cryptographic proof (pf) on the relationship between T, L, and the encrypted message.
- ***Attribute Authority (AA):*** A trusted entity responsible for system setup and the issuance of attribute-based private keys (SK) to users based on their assigned attribute sets.
- ***Public Cloud (CS):*** The entity responsible for large-scale data storage. The CS is generally assumed to be untrusted.
- ***Private Cloud (PC):*** A semi-trusted entity that performs limited, sensitive computations, primarily tag checking and duplicate detection. The PC manages a tag-label list to facilitate the deduplication process.

This hybrid environment strategically places the trusted components (AA and Do) and the semi-trusted computational engine (PC) away from the high-volume, untrusted storage component (CS), allowing for secure processing without compromising confidentiality.

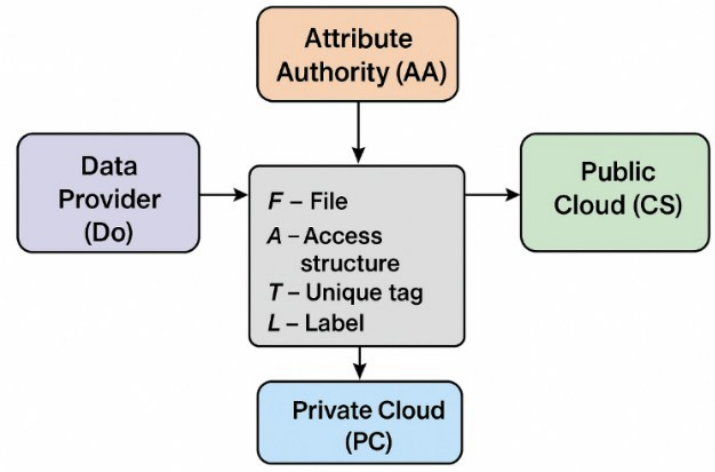

Fig. 1 System Architecture of the Proposed System

 DOI: 10.5281/zenodo.17846700

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

IV. SECURE DEDUPLICATION PROTOCOL AND POLICY TRANSFORMATION

The crucial challenge in this system is enabling deduplication when data is encrypted under variable access policies. The proposed protocol achieves this through a mechanism that modifies the ciphertext structure.

4.1. Data Ingestion and Duplicate Detection

When a Data Provider submits a file storage request, the following steps occur at the Private Cloud (PC):

- Proof Verification: The PC first checks the validity of the proof (pf) provided by the Data Provider. This proof ensures the integrity of the tag-label relationship.
- Tag Checking: The PC then tests the equality of the newly created tag (T) with the existing tags in its tag-label list.
- Storage Decision:
  - If there is no match for the new tag T, the PC adds the new tag T and the label L to its list and forwards the encrypted data and label to the Public Cloud for storage.
  - If a match is found (a duplicate exists), the PC proceeds to the Ciphertext Transformation Protocol to enable data sharing under the new policy without re-uploading the data.

4.2. Ciphertext Transformation Protocol

When a duplicate file is detected but the new request specifies a different access policy ($\mathbb{A}'$) than the stored ciphertext (CT under policy $\mathbb{A}$), the Private Cloud executes a methodology to modify the ciphertext.

This core functionality is a put forth methodology to modify a ciphertext over one access policy ($\mathbb{A}$) into a ciphertext of the same plaintext but under other access policies ($\mathbb{A}'$) without revealing the underlying plaintext. The Private Cloud, when provided with a trapdoor key, is able to convert the ciphertext C with access policy $\mathbb{A}$ to a new ciphertext C' with another access policy $\mathbb{A}'$.

This process ensures that the single stored instance of the encrypted data can be securely shared with users based on the new access policy, fully enabling secure deduplication.3 If multiple access policies are involved, the Private Cloud can regenerate a ciphertext that satisfies a combined access policy ($\mathbb{A} \cup \mathbb{A}'$)..

V. SECURITY ANALYSIS AND OPERATIONAL RESILIENCE

The architectural design must guarantee security and resilience against specific threats prevalent in cloud environments.

5.1. Defense Against Duplicate Faking Attacks

A critical security concern is the duplicate faking attack, where an adversary attempts to replace an honest message with a fake message that shares the same cryptographic tag. This system provides security against such attacks through the tag/label consistency verification process.

The Private Cloud's check of the cryptographic proof (pf) upon receiving a storage request ensures the tag and the label are correctly tied to the encrypted message. If an adversary attempts to tamper with the label or forge the tag, the Private Cloud detects the tampering

 DOI: 10.5281/zenodo.17846700

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

immediately. Consequently, a user having decryption privilege to the ciphertext can always check the correctness of the plaintext via the label.

5.2. Enhancing Integrity and Decentralization

To further enhance trust and data integrity, advanced technologies can be incorporated:

- *Blockchain-Assisted Verification:* Blockchain technology can be adopted to assist users in verifying file integrity and detecting deduplicated data based on its tamper-evident nature. This provides an additional layer of integrity assurance, allowing users to verify the legitimacy of the ciphertext.
- *Decentralization:* Traditional ABE systems rely on a single, centralized Attribute Authority (AA). Investigating Decentralized Attribute-Based Encryption (DABE) expands upon the conventional framework by dispersing the responsibility and control of cryptographic keys among several groups or authorities. This approach enhances resilience and mitigates collusion threats.

5.3. User Revocation

The system supports the ability to efficiently conduct both deduplication and revocation. Revocation is essential to ensure that an expired user cannot decrypt the data, even if their attributes technically still satisfy the access criteria. While efficient user revocation complicates ABE key management, it is a necessary feature for secure data sharing.

VI. PERFORMANCE AND EFFICIENCY GAINS

Although achieving high-level semantic security and flexibility (CP-ABE with Ciphertext Transformation) introduces computational overhead, the overall system design yields significant efficiency gains in the long term.

Performance evaluation metrics include assessing latency, throughput, and, crucially, storage capacity. The primary advantage of the SADD system is the efficiency gained from eliminating redundant copies of data. The security analysis and experiments on schemes that support batch checking and semantically secure storage have demonstrated that the system can save data storage by up to 89.84%. By mitigating duplicate storage, the system also saves network bandwidth during the upload phase.

It is recognized that highly secure methodologies, such as those employing advanced encryption and deduplication techniques, may result in a slight increase in execution time compared to less secure existing systems. However, this minimal trade-off is a strategic investment that ensures the enhanced security level required for data confidentiality in sensitive cloud environments. The ability to efficiently handle fine-grained access control while substantially reducing storage overhead validates the design's overall performance and efficiency.

VII. CONCLUSION

The secure deduplication support system successfully resolves the incompatibility between Attribute-Based Encryption and cloud data deduplication through a meticulously designed hybrid cloud architecture. By leveraging CP-ABE to ensure fine-grained, policy-based access control and achieving

 DOI: 10.5281/zenodo.17846700

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

the standard notion of semantic security, the system provides a robust security posture. The unique methodology for ciphertext transformation within the semi-trusted Private Cloud enables the elimination of duplicate files, even if they are encrypted under different policies, thereby achieving massive storage efficiency. Future work should focus on implementing decentralized authority structures (DABE) and leveraging blockchain for enhanced integrity verification to further eliminate trust bottlenecks and secure the entire architecture.

REFERENCES

[1] J. K. Periasamy, S. Prabhakar, A.Vanathi, LiuYu4, "Enhancing cloud security and deduplication efficiency with SALIGP and cryptographic authentication", Scientific Reports, (2025) 15:30112, https://doi.org/10.1038/s41598-025-14972-3

[2] M. Pavithra, M. Prakash, V. Vennila, "BGNBA OCO based privacy preserving attribute based access control with data duplication for secure storage in cloud", Journal of Cloud Computing (2023) 13:8 https://doi.org/10.1186/s13677-023-00544-1

[3] Priteshkumar Prajapati, Parth Shah, "A Review on Secure Data Deduplication: Cloud Storage Security Issue", Journal of King Saud University – Computer and Information Sciences 34 (2022) 3996–4007

[4] Won-Bin Kim, Im-Yeong Lee, "Survey on Data Deduplication in Cloud Storage Environments", J Inf Process Syst, Vol.17, No.3, pp.658~673, June 2021 ISSN 1976-913X https://doi.org/10.3745/JIPS.03.0160

[5] Shynu P. G., Nadesh R. K., Varun G. Menon, Venu P., Mahdi Abbasi, Mohammad R. Khosravi, "A secure data deduplication system for integrated cloud-edge networks", Journal of Cloud Computing: Advances, Systems and Applications (2020) 9:61 https://doi.org/10.1186/s13677-020-00214-6

[6] Thottipalayam Andavan, M., Parameswari, M., Subramanian, N. & Vairaperumal, N. A novel model for enhancing cloud security and data deduplication using fuzzy and refraction learning based chimp optimization. Int. J. Mach. Learn. Cybern. 15 (3), 1025– 1038. https://doi.org/10.1007/s13042-023-01953-z (2024).

[7] Yang, Z., Zhu, H., Li, Z., Wang, G. & Su, M. A malware detection method based on genetic algorithm optimized CNN-Senet network. IEEE Access https://doi.org/10.1109/ACCESS.2024.3485917 (2024).

[8] Pampattiwar, K. N. & Chavan, P. V. Blockchain-based composite access control and secret sharing-based data distribution for security-aware deployments. Int. J. Inf. Comput. Secur. 25 (3–4), 292–332. https://doi.org/10.1504/IJICS.2024.143920 (2024).

[9] Brahmam, M. G. & Anand, R. V. VMMISD: An efficient load balancing model for virtual machine migrations via fused metaheuristics with iterative security measures and deep learning optimizations. IEEE Access https://doi.org/10.1007/s41060-02 5-00718-x (2024).

[10] Alsadie, D. Artificial intelligence techniques for Securing fog computing environments: Trends, challenges, and future directions. IEEE Access https://doi.org/10.1109/ACCESS.2024.3463791 (2024).

[11] Ha, G. et al. Scalable and popularity-based secure deduplication schemes with fully random tags. IEEE Trans. Dependable Secur. Comput. 14, 1–17. https://doi.org/10.1109/TDSC.2023.3285173 (2023).

[12] Reddy, M. I., Rao, P. V., Kumar, T. S. & K, S. R Encryption with access policy and cloud data selection for secure and energy-efficient cloud computing. Multimed. Tools Appl. 83 (6), 15649–15675. https://doi.org/10.1007/s11042-023-16082-6 (2024).

[13] Vijayakumar, D., Srinivasagan, K. G. & Vivekrabinson, K. Enhancing cloud storage security through blockchain-enabled data deduplication and auditing with a fair payment. Peer-to-Peer Netw. Appl. 18 (3), 147. https://doi.org/10.1007/s12083-025-01970-5 (2025).

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

[14] Fu, Y. et al. Distributed data deduplication for big data: A survey. ACM Comput. Surveys. https://doi.org/10.1145/3735508 (2025).

[15] Zeydan, E., Arslan, S. S. & Liyanage, M. Managing distributed machine learning lifecycle for healthcare data in the cloud. IEEE Access https://doi.org/10.1109/ACCESS.2024.3443520 (2024).

[16] Liu, B. et al. Blockchain-assisted fine-grained deduplication and integrity auditing for outsourced Large-Scale data in cloud storage. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2025.3548681 (2025).

[17] Zhang, Q. et al. Blockchain-based Privacy-preserving deduplication and integrity auditing in cloud storage. IEEE Trans. Comput. https://doi.org/10.1109/TC.2025.3540670 (2025).

[18] Zheng, X., Shen, W., Su, Y. & Gao, Y. DIADD: Secure deduplication and efficient data integrity auditing with data dynamics for cloud storage. IEEE Trans. Netw. Serv. Manag. https://doi.org/10.1109/TNSM.2025.3535708 (2025).

[19] Lapmoon, J. & Fugkeaw, S. A verifiable and secure industrial IoT data deduplication scheme with real-time data integrity checking in fog-assisted cloud environments. IEEE Access https://doi.org/10.1109/ACCESS.2025.3529765 (2025).

[20] Abdeddine, A., Mekouar, L. & Iraqi, Y. A generic framework for mobile crowdsensing: A comprehensive survey. IEEE Access https://doi.org/10.1109/ACCESS.2025.3526739 (2025).

