



Privacy Preservation in Cloud Computing Using Attribute-Based Access Control, Spectral Clustering and Hybrid Anonymization

¹Anoop Srivastava and ²Dr. Shakeeb Khan

¹Research Scholar, Department of Computer Applications, Maharaja Agrasen Himalayan Garhwal University, Uttarakhand, India

²Assistant Professor, Department of Computer Applications, Maharaja Agrasen Himalayan Garhwal University, Uttarakhand, India

DOI: <https://doi.org/10.5281/zenodo.23068514>

Corresponding Author: Anoop Srivastava

Abstract

Cloud computing facilitates extensive data sharing, storage, and analytical processing, creating significant benefits for organizations while simultaneously increasing the potential exposure of sensitive personal, healthcare, financial, and institutional information. This study presents a focused research-paper adaptation of the privacy-preservation aspects discussed in the supplied doctoral paper. The research investigates the application of Attribute-Based Encryption (ABE) for protecting and securely sharing medical records in cloud environments, along with Walrus-based cloud storage, spectral clustering for privacy-oriented user segmentation, the Graph-Based Anonymization Technique with Attribute Probability Matrix (GAT-APM), and Integrated Geometric and Ordinal Incremental Data-Based Privacy Preservation (IGOID-PP).

The work is grounded in an algorithm-based research design. Privacy mechanisms are framed as computational procedures rather than solely policy recommendations. The proposed approaches attempt to preserve confidentiality while retaining data utility, supporting controlled sharing and identifying potentially risky access behavior. The paper also discusses k-anonymity, differential privacy, federated learning, homomorphic encryption and blockchain as complementary privacy-enhancing approaches.

The reported results include qualitative findings and correlation-based security analysis from the paper, together with a comparative time graph for anonymization methods. The proposed hybrid model combines geometric perturbation and probability-based attribute sanitization. Because the paper does not provide a complete experimental dataset, statistical protocol or full benchmark specification for every technique, this paper distinguishes reported paper observations from interpretations and avoids introducing unsupported numerical claims.

Keywords: Cloud privacy, Attribute-Based Encryption, Walrus, spectral clustering, GAT-APM, IGOID-PP, anonymization, healthcare data, data utility

Introduction

Cloud computing allows information to be stored, managed, and processed beyond the user's immediate physical control, creating significant challenges for maintaining data privacy. The paper identifies this transfer of control as a key concern, as sensitive information may be vulnerable to unauthorized access, insider threats, insecure APIs, incorrect system configurations, inadequate authentication, third-party involvement, and limited transparency regarding data-handling practices.

Privacy is a broader concept than confidentiality or anonymity, although the three are closely related. Confidentiality focuses on preventing unauthorized

disclosure of information, whereas anonymity seeks to conceal an individual's identity. Privacy, in contrast, encompasses broader control over the collection, storage, sharing, processing, and utilization of personal information. Accordingly, the paper emphasizes that effective cloud privacy requires a combination of suitable confidentiality and anonymity techniques, together with meaningful user control and transparent governance mechanisms. The research is particularly relevant to healthcare. Medical records contain highly sensitive information but must be shared among authorized doctors, nurses, pharmacists and other stakeholders. A privacy mechanism must therefore protect data without making legitimate access impossible.

ABE is used in the paper as a way to attach access policies to encrypted information so that decryption depends on user attributes.

Healthcare and Cloud Data Sharing

The healthcare use case illustrates why privacy-aware sharing is difficult. A hospital may need to share a patient record with a cardiologist while limiting unrelated staff, or allow emergency personnel access under an emergency attribute. Traditional single-key encryption does not naturally express these fine-grained policies. ABE addresses this by associating access with attributes and policies.

The paper describes CP-ABE as particularly suitable for cloud healthcare. The data owner defines an access policy and attaches it to the ciphertext. Users receive attribute-bearing decryption keys. If the user's attributes satisfy the policy, the record can be decrypted. If they do not, access is denied even when the encrypted object is accessible.

Walrus is described in the paper as an open-source storage system associated with the Eucalyptus cloud platform. It is used as the secure storage backend in the healthcare scenario. The combination of Walrus and ABE separates scalable storage from fine-grained cryptographic authorization, providing a model for secure cloud medical-record management.

Review of Literature

The literature surveyed in the paper shows rapid development in cloud privacy protection. Recent healthcare research addresses patient trust, IoT-generated medical data, EHR security, medical-image privacy, digital twins and cloud-based healthcare architectures. These studies commonly emphasize confidentiality, integrity, authentication, encryption and continuous monitoring.

The literature also covers cryptographic and privacy-preserving methods. Traditional AES and RSA remain important for practical protection, while homomorphic encryption, searchable encryption and secure multi-party computation support more advanced privacy-preserving processing. Differential privacy and federated learning address statistical and machine-learning privacy problems by limiting the exposure of individual records.

The paper identifies a research need for systematic comparison because individual techniques often optimize different properties. Encryption protects content but does not itself solve re-identification in released datasets. K-anonymity reduces uniqueness but may suffer from utility loss and linkage attacks. Graph-based and geometric methods can preserve structural information but require careful parameterization. The present paper organizes these methods around access control, behavioral analysis and high-dimensional anonymization.

Attribute-Based Encryption

ABE is a public-key cryptographic approach in which access is determined by attributes rather than a single identity-specific key. The paper distinguishes Key-Policy ABE and Ciphertext-Policy ABE. In KP-ABE, the access policy is embedded in the user's key; in CP-ABE, the data owner defines the policy associated with the ciphertext while user keys contain attributes.

For healthcare cloud sharing, CP-ABE allows a patient or

hospital to define a policy such as access for users having a doctor and cardiology attribute, or emergency department status. The cloud stores encrypted records without requiring the cloud operator to understand the plaintext. The security model consequently moves authorization logic into cryptographic policy enforcement.

The paper identifies benefits including fine-grained access control, enhanced privacy even when storage is compromised, flexible role-based sharing and scalability across multiple users or institutions. It also connects ABE to regulatory concerns such as HIPAA and GDPR. However, practical deployment still requires careful key management, attribute revocation and performance evaluation.

Behavioral Privacy and AI

The paper proposes spectral clustering as an unsupervised method for privacy-aware user segmentation. User access behavior, data-transfer patterns and security events can be represented as a graph. Similarity relationships are captured in an affinity matrix, followed by graph-Laplacian construction and eigenvalue decomposition. K-means can then be applied in the transformed feature space.

This approach can distinguish groups of normal and suspicious behavior. The paper describes applications including detection of unusual login behavior, privacy-aware segmentation according to security needs, detection of abnormal data movement and adaptive encryption optimization. Users with higher apparent risk can be assigned stronger protection or monitoring.

The broader literature summarized in the paper also discusses AI-driven anomaly detection, blockchain-based identity, federated learning and privacy-preserving AI. These methods can be combined, but each introduces implementation considerations such as computational overhead, model complexity, transparency and governance. The research therefore treats AI as a complementary security layer rather than a replacement for cryptographic controls.

Objectives and Scope – Proposed Privacy Architecture

The proposed privacy architecture can be viewed as four coordinated layers. The storage layer uses encrypted cloud objects. The access layer uses ABE policies and authenticated user attributes. The behavioral layer uses clustering and monitoring to identify risky patterns. The data-sharing layer uses anonymization techniques when datasets must be released or analyzed by parties that should not receive raw identifiers.

In healthcare, the data owner encrypts a medical record under a policy before uploading it to Walrus. A doctor, nurse or pharmacist receives attributes associated with their role or specialization. Decryption succeeds only when the attributes satisfy the policy. This approach allows multiple access categories without creating a separate plaintext copy for every user.

For analytical data, GAT-APM and IGoid-PP provide different forms of transformation. GAT-APM uses graph relationships and attribute probabilities. IGoid-PP uses geometric perturbation and ordinal/incremental anonymization. Their hybrid formulation seeks to balance both mechanisms.

Research Methodology / Experimental

The study adopts an algorithm-driven design in which privacy mechanisms are described, implemented or simulated, and evaluated. The methodology starts by identifying privacy threats such as unauthorized access, data leakage and re-identification. It then selects cryptographic, access-control, clustering and anonymization mechanisms appropriate to each threat.

Python is identified in the paper as the main programming environment. Cryptographic libraries such as PyCryptodome and OpenSSL are discussed for encryption-related operations, while Scikit-Learn is used for machine-learning and clustering tasks. Cloud simulation can be supported using AWS SDK and Google Cloud APIs. The methodological aim is reproducibility rather than dependence on one commercial cloud.

Evaluation dimensions include privacy protection, security strength, computational overhead, scalability and data utility. The paper also stresses ethical considerations, including protection of user data, transparent reporting and independent verification. Where a numerical benchmark is not fully specified in the paper, this paper reports the qualitative or graph-based result rather than inventing a new value.

Walrus and CP-ABE Workflow

The CP-ABE workflow begins with a data owner, such as a patient or hospital, defining an access policy. The medical record is encrypted according to that policy before it is uploaded to cloud storage. The cloud therefore receives ciphertext rather than readable medical information.

A user requests the record and presents a decryption key containing attributes. The policy may require a single attribute or a logical combination of attributes. If the attributes satisfy the policy, the record is decrypted; otherwise, the request fails. The paper also describes dynamic revocation or policy changes as a way to prevent users who no longer require access from decrypting future data.

The approach is intended to provide fine-grained authorization while retaining scalable cloud storage. Its security benefit depends on correct cryptographic implementation and secure attribute/key management. The architecture is therefore complemented by authentication, audit trails and monitoring rather than treating ABE as a complete security system.

Spectral Clustering

The paper implements spectral clustering on synthetic cloud user-behavior data using 300 samples with two clusters. The conceptual inputs are user access data, the number of clusters and an affinity function. Each user or behavior is represented as a node, and similarity relationships form the graph.

The method computes an affinity matrix A , a degree matrix D and a graph Laplacian $L = D - A$. Eigenvalue decomposition transforms the observations into a feature space in which clustering is applied. K-means is then used to produce clusters. The paper interprets one cluster as normal behavior and another as suspicious or privacy-risk behavior, after which stronger monitoring or encryption can be assigned.

This experiment illustrates how unsupervised learning can be integrated with privacy management. It does not by itself prove that all suspicious behavior will be detected. Real deployments would require representative access logs, validated labels or expert review, threshold analysis and measurement of false positives and false negatives.

IGOID-PP

IGOID-PP combines geometric perturbation with ordinal and incremental anonymization. Its purpose is to modify sensitive values while preserving useful relationships. The paper gives the transformation $T(X) = X + r \cdot \cos(\theta) + s \cdot \sin(\theta)$, where r and s are random perturbation factors and θ is a rotation factor.

The geometric transformation changes the original values so that direct reconstruction is more difficult while attempting to preserve relationships among observations. Ordinal anonymization maintains ranking order, which is important for datasets in which relative position carries analytical meaning. Incremental anonymization adjusts protection according to exposure risk so that higher-risk records can receive stronger transformation.

The paper identifies healthcare, financial transactions and large-scale enterprise data as application areas. The claimed benefits are enhanced privacy, preserved utility, scalability and adaptive protection. As with GAT-APM, practical validation should include re-identification testing, utility metrics and computational benchmarking.

Hybrid Privacy Model

The paper combines IGOID-PP and GAT-APM into a hybrid model. IGOID-PP contributes geometric perturbation and ordinal/incremental anonymization, while GAT-APM contributes graph-based structural protection and probabilistic attribute sanitization. The combined formulation is $S_{cloud} = \alpha \cdot T(X) + \beta \cdot A_i$, where α and β balance the two components.

The conceptual advantage is multi-layered privacy. Geometric perturbation reduces direct exposure of values, while probability-based sanitization reduces attribute-level linkage. The paper describes the hybrid model as adaptable to dynamic cloud environments and suitable for large datasets.

The key methodological issue is parameter selection. If perturbation is too weak, privacy protection may be insufficient; if it is too strong, data utility can deteriorate. The weights α and β therefore represent a research parameter that should be calibrated according to sensitivity, utility requirements and threat models.

5. Hyprpaper

H1: A cloud architecture integrating fine-grained access control and encryption provides stronger protection against unauthorized access than relying on unrestricted cloud storage.

H2: CP-ABE can support data-owner-defined access policies for secure sharing of sensitive medical records in cloud environments.

H3: Spectral clustering can provide a computational mechanism for grouping cloud user behaviors into privacy-relevant segments and identifying potentially anomalous patterns.

- H4:** GAT-APM can reduce direct attribute exposure in high-dimensional data while retaining relationships needed for legitimate analysis.
- H5:** IGOID-PP can improve privacy protection by combining geometric perturbation with ordinal and incremental anonymization while attempting to preserve data utility.
- H6:** A hybrid IGOID-PP and GAT-APM model can provide complementary privacy protection against direct exposure and attribute-linkage risks.

These hypotheses are framed as research propositions derived from the paper architecture. They should be tested empirically with clearly defined datasets, attack models, privacy metrics and utility measures in future validation studies.

Limitation of Research Work (Research Gap)

The paper identifies a research gap around the absence of a comprehensive evaluation of privacy-preserving solutions across different cloud environments. Existing methods often address specific risks, but their effectiveness, scalability, usability and privacy-utility trade-offs are not always evaluated under a common framework.

A further limitation is that the supplied paper combines several methods with different purposes. ABE addresses access authorization, spectral clustering addresses behavioral segmentation, and GAT-APM/IGOID-PP address anonymization. Their outputs are not directly comparable using one universal metric. A research-paper evaluation therefore needs clearly separated outcome measures.

The paper also provides algorithms and illustrative results but does not fully specify a single large-scale benchmark dataset, complete experimental configuration, statistical significance analysis or standardized re-identification attack suite for all methods. This paper therefore treats the reported methods and findings as paper-derived evidence rather than claiming a new independent benchmark.

Privacy-Utility Trade-off

Anonymization creates an unavoidable design problem: increasing privacy protection can reduce data utility. The paper explicitly notes this problem for k-anonymity and perturbation methods. High K values may reduce uniqueness but can also generalize data so strongly that analytical precision declines. Likewise, excessive perturbation may protect values but distort relationships.

Graph-based methods introduce another issue: relationships themselves can become identifying information. A graph can preserve structure while still exposing rare patterns. Probability-based sanitization therefore needs testing against linkage and background-knowledge attacks. Geometric methods similarly require evaluation against reconstruction attacks and distributional inference.

The research gap is therefore not simply the creation of another anonymization algorithm. It is the development of an evaluation framework that simultaneously measures disclosure risk, data utility, computational cost, scalability and policy compliance. Such an evaluation would make it possible to understand which mechanism is appropriate for a particular cloud data-sharing scenario.

Results and Discussion

The paper reports that the Walrus and ABE combination provides a model for secure storage and fine-grained sharing of medical records. Sensitive data is encrypted before cloud upload, and only users whose attributes satisfy the defined policy can decrypt it. The architecture is described as supporting strong patient privacy, flexible access levels and scalable multi-user healthcare environments.

The spectral clustering implementation uses synthetic two-cluster data generated with 300 samples. The resulting segmentation is intended to separate standard user behavior from suspicious behavior. The paper proposes using the cluster assignment to adjust monitoring or encryption requirements. This provides an adaptive security concept in which privacy controls respond to observed behavior rather than remaining identical for every user.

The anonymization results are represented through GAT-APM and IGOID-PP formulations and a time-comparison graph. The paper states that GAT-APM reduces temporal and spatial complexity in high-dimensional anonymization and that IGOID-PP is designed to preserve relationships while altering sensitive values. These findings are interpreted as evidence of complementary rather than identical privacy mechanisms.

Comparative Anonymization Findings

The paper's compares time across GA07B, LZD, WCW, IBPRE and the proposed method at different data sizes represented by 1, 100, 500 and 1000. The plotted lines show that processing time increases as data size increases, with the largest workload producing the greatest time values across the methods. The proposed method is plotted at a comparatively low time level in the displayed figure.

Because the figure is graphical and the paper does not provide the underlying numerical dataset in the supplied text, the exact values should not be treated as precise benchmark measurements in this paper. The useful result is the observed trend: the proposed method remains comparatively efficient as workload size increases in the illustrated experiment.

The paper attributes the hybrid model's privacy benefits to the combination of geometric and probability-based transformations. It identifies prevention of attribute linkage attacks, scalability and adaptability as key benefits. These claims are best understood as design findings supported by the paper formulation and figure rather than as a universal guarantee against all attack models.

K-Anonymity and Data Sharing

The paper explains that removing direct identifiers does not necessarily eliminate privacy leakage. An attacker may combine quasi-identifiers with external information and re-identify individuals. K-anonymity addresses this by making records indistinguishable from at least K-1 other records with respect to selected identifying attributes.

The implementation process described in the paper includes identifying private attributes and quasi-identifiers, applying generalization or suppression, translating privacy policy into measurable parameters, selecting an appropriate K value and producing the anonymized dataset. The paper emphasizes that choosing K requires balancing protection and utility.

The principal limitations are linkage attacks, background knowledge, computational cost and utility loss. L-diversity and t-closeness are identified as extensions that can strengthen protection. The result is therefore that k-anonymity remains useful but should be integrated with more advanced privacy-preserving mechanisms rather than used as the only defense.

Conclusions

The study concludes that cloud privacy requires coordinated cryptographic, access-control, anonymization and monitoring mechanisms. In healthcare, Walrus and CP-ABE provide a structured approach to encrypted storage and data-owner-defined access policies. The model supports different access privileges for different professional attributes and reduces the impact of a storage compromise because encrypted records cannot be read without satisfying the access policy.

For behavioral security, spectral clustering provides an unsupervised graph-based method for segmenting users or activities. For data publication, GAT-APM and IGOID-PP offer complementary mechanisms: graph/probability-based sanitization and geometric/ordinal/incremental perturbation. The hybrid model combines these properties in a unified formulation.

The conclusions should be interpreted within the evidence available in the paper. The methods are promising architecture and algorithmic approaches, but large-scale independent experiments, attack simulations and standardized privacy/utility metrics would strengthen validation. The research nevertheless provides a coherent framework for privacy-aware cloud computing.

Social Relevance

Privacy-preserving cloud computing has direct relevance to citizens because medical, financial, educational and personal records are increasingly processed by remote systems. The paper emphasizes that privacy protection contributes to trust in digital services. Effective access control can reduce unauthorized disclosure, while anonymization can allow useful analysis without exposing individual identities unnecessarily.

In healthcare, privacy mechanisms can support controlled sharing among professionals while reducing the risk that patients' records become openly readable in cloud storage. In research and analytics, anonymization can enable secondary use of data while attempting to preserve useful relationships. The social benefit depends on responsible governance, consent, security awareness and appropriate legal compliance.

The paper also identifies social risks such as digital inequality, limited user awareness and differences among regional privacy laws. It recommends educational initiatives, standardized privacy protection and privacy-focused services for underserved populations. Thus, the social relevance of the research extends beyond cryptography: it concerns digital rights, ethical data use, organizational accountability and public confidence.

References

1. Vempali V, Rao K, Prathap TMR, Jhansi Rani M, Mahendher S, Leonard, et al. Strengthening customer data protection in healthcare: an empirical analysis of data privacy and security measures for ensuring customer information security. *European Economics Letters*. 2024;14:386-394. doi:10.52783/eel.v14i2.1337.
2. Manukondakrupa AC. Fortifying patient privacy: a cloud-based IoT data security architecture in healthcare. 2024;[publication details not provided].
3. Zarichuk O. Security in cloud computing: methods for ensuring privacy and integration in modern applications. *Development Management*. 2024;23:37-45. doi:10.57111/devt/1.2024.37.
4. Yi H. Improving cloud storage and privacy security for digital twin based medical records. *Journal of Cloud Computing*. 2023;12. doi:10.1186/s13677-023-00523-6.
5. Patel RK, Gidwani P, Patel NR. Privacy preservation and cloud computing. In: *Privacy Preservation and Secured Data Storage in Cloud Computing*. IGI Global; 2023. p.88-107. doi:10.4018/979-8-3693-0593-5.ch004.
6. Gayathri S, Gowri S. Securing medical image privacy in cloud using deep learning network. *Journal of Cloud Computing*. 2023;12. doi:10.1186/s13677-023-00422-w.
7. Geetha Rani E, Chetana DT. A survey of recent cloud computing data security and privacy disputes and defending strategies. Springer; 2023. doi:10.1007/978-981-99-2468-4_31.
8. Abdulsalam Y, Hedabou M. Security and privacy in cloud computing: technical review. *Future Internet*. 2022;14(1):11. doi:10.3390/fi14010011.
9. Verma G. Blockchain-based privacy preservation framework for healthcare data in cloud environment. *Journal of Experimental & Theoretical Artificial Intelligence*. 2022;36. doi:10.1080/0952813X.2022.2135611.
10. Gayathri S, Gowri S. CUNA: a privacy preserving medical records storage in cloud environment using deep encryption. *Measurement: Sensors*. 2022;24:100528. doi:10.1016/j.measen.2022.100528.
11. Ukeje N, Gutierrez J, Petrova K. Information security and privacy challenges of cloud computing for government adoption: a systematic review. *International Journal of Information Security*. 2024. doi:10.1007/s10207-023-00797-6.
12. Sivan R, Zukarnain Z. Security and privacy in cloud-based e-health system. *Symmetry*. 2021;13:742. doi:10.3390/sym13050742.
13. Uma Maheswari J, Vijayalakshmi S, Rajiv Gandhi N, Alzubaidi LH, Khonimkulov A, Elangovan R. Data privacy and security in cloud computing environments. *E3S Web of Conferences*. 2023;399:04040. doi:10.1051/e3sconf/202339904040.
14. Chen CL, Huang PT, Deng YY, Chen HC, Wang YC. A secure electronic medical record authorization system for smart device application in cloud computing environments. *Human-centric Computing and*

Information Sciences. 2020;10:21. doi:10.1186/s13673-020-00221-1.

15. Ganiga R, Pai R, MM, Manohara, Sinha R. Security framework for cloud based electronic health record (EHR) system. International Journal of Electrical and Computer Engineering. 2020;10:455-466. doi:10.11591/ijece.v10i1.pp455-466.

Creative Commons (CC) License

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.