A Survey of Privacy-Preserving Techniques in Distributed Systems

Munib Fahmid Rafeed
ID: 24341128
BRAC University
Dhaka, Bangladesh
munib.fahmid.rafeed@g.bracu.ac.bd

Abstract

Privacy-preserving computing is critical for safeguarding sensitive information in distributed systems. This survey examines key techniques, including encryption-based methods, differential privacy, federated learning, secure multi-party computation, and blockchain, categorizing them based on their strengths, challenges, and applications. Emerging trends such as quantum-resistant cryptography and decentralized identity systems are also explored. The report highlights current challenges like scalability, adversarial threats, and regulatory compliance, proposing directions for future research to achieve scalable and interoperable solutions.

Keywords

Privacy-preserving computing, distributed systems, encryption, differential privacy, federated learning, blockchain, secure multi-party computation.

ACM Reference Format:
Munib Fahmid Rafeed, ID: 24341128. . A Survey of Privacy-Preserving Techniques in Distributed Systems. In *Proceedings of* . ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

In an increasingly interconnected world, distributed systems have become the backbone of modern computing. From cloud services and Internet of Things (IoT) devices to federated learning frameworks, these systems enable seamless data exchange and processing across geographically dispersed networks. However, this growth has been accompanied by heightened concerns about data privacy and security.

Privacy-preserving computing aims to protect sensitive data while still enabling valuable computations and insights. The importance of this field has grown with the rise of stringent regulations like the General Data Protection Regulation (GDPR) and growing awareness of privacy risks. These challenges necessitate innovative approaches that balance security, efficiency, and functionality in distributed systems.

This survey explores the landscape of privacy-preserving techniques in distributed systems. It categorizes the methods into encryption-based solutions, differential privacy, federated learning, secure multi-party computation, and blockchain-based approaches. By analyzing their effectiveness, scalability, and use cases, this report provides a comprehensive overview of the current state and future potential of privacy-preserving computing.

,
©
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/XXXXXXX.XXXXXXX

2 Background

2.1 Distributed Computing and Systems

Distributed computing refers to a network of interconnected systems collaborating to perform computational tasks. These systems distribute workloads among multiple nodes, providing advantages such as scalability, fault tolerance, and high availability. Applications of distributed computing range from cloud platforms and IoT devices to high-performance computing clusters and peer-to-peer networks.

2.2 Privacy Challenges in Distributed Systems

Despite the efficiencies and scalability improvements offered by distributed systems, they also present unique privacy challenges:

- **Data Breaches:** Sensitive information transmitted across multiple nodes is at risk of unauthorized access or cyberattacks.
- **Regulatory Compliance:** Legislation such as GDPR, HIPAA, and CCPA impose rigorous privacy requirements, necessitating organizations to protect user data diligently.
- **Interdisciplinary Concerns:** Privacy considerations must coexist with distributed system characteristics like real-time processing, reliability, and minimal latency, often resulting in trade-offs.

3 Literature Review

3.1 Encryption-Based Techniques

Encryption methodologies serve as the foundational elements of privacy-preserving strategies within distributed systems. Notably, homomorphic encryption facilitates operations on encrypted data, thereby maintaining confidentiality during the entire process. Reddy and Nallapa Reddy [10] illustrated its application in secure content management systems to support personalization while safeguarding data. Similarly, Yu et al. [13] investigated the implementation of Paillier encryption in sampled-data consensus for multi-agent systems, ensuring secure distributed computations in nonlinear settings.

3.2 Differential Privacy

Differential privacy incorporates statistical noise into datasets to thwart individual re-identification in analytical processes. Mao and Zhuang [7] presented a privacy-aware cooperative edge inference framework that utilizes differential privacy to improve the efficacy of distributed systems while ensuring data integrity. Their methodology successfully reconciles privacy preservation with utility, addressing a common obstacle in large-scale implementations.

3.3 Federated Learning (FL)

Federated learning permits collaborative model training across distributed nodes without the necessity of sharing raw data, thus safeguarding privacy. Forootani and Iervolino [3] examined asynchronous federated learning techniques to overcome scalability and heterogeneity challenges in distributed machine learning systems. Furthermore, Rauch [9] investigated federated learning structures optimized for privacy-preserving distributed asset transfer, underscoring their potential applicability in financial services and IoT environments.

3.4 Secure Multi-Party Computation (SMPC)

SMPC enables secure collaborative computations without exposing raw data among participants. Nguyen et al. [8] realized context-aware SMPC methods for federated anomaly detection, delivering strong privacy safeguards in IoT contexts. Their research highlights SMPC's significance in essential systems such as healthcare and financial networks.

3.5 Blockchain Integration

The application of blockchain technology has been progressively employed to bolster privacy within distributed systems. Imteaj et al. [4] proposed blockchain-enabled federated learning paradigms for edge computing, introducing robust trust frameworks while diminishing cyber threats. In a similar context, Sharma and Khullar [11] designed a blockchain-centric federated learning architecture for precision agriculture, ensuring secure and privacy-preserving communications among IoT devices.

3.6 Emerging Trends

Recent trends in privacy-preserving computing emphasize quantum-resistant cryptography and individualized federated learning. Zhu and Fan [14] developed a decentralized federated learning framework utilizing quantum-resistant cryptographic methods to counter the increasing risks posed by quantum computing to conventional encryption protocols. Additionally, Wang et al. [12] employed digital twin-powered federated incremental learning, demonstrating its effectiveness in addressing non-independent and identically distributed (non-IID) data in distributed networks.

3.7 Privacy Challenges

Despite notable advancements, challenges such as scalability, adversarial resilience, and adherence to regulatory requirements remain. Di and Shi [2] underscored the shortcomings of current cross-domain recommendation systems in sustaining privacy while ensuring high accuracy across distributed environments. Similarly, Kong and Ye [5] tackled privacy issues in multi-virtual power plant scheduling by proposing innovative privacy-preserving optimization techniques for energy distribution systems.

3.8 Real-World Applications

The application spectrum of privacy-preserving techniques continues to widen. Liu et al. [6] illustrated utility in nonlinear multi-agent systems for industrial automation, highlighting adaptive control and encrypted communications. Furthermore, Marmol Campos [1] focused on intrusion detection systems (IDS) within IoT contexts, employing federated learning to improve privacy while identifying cyber threats.

4 Privacy-Preserving Techniques

4.1 Overview of Privacy-Preserving Computing

Privacy-preserving computing tackles these challenges by facilitating secure data processing and sharing without exposing raw information. Its primary objectives include:

- **Confidentiality:** Safeguarding sensitive data from unauthorized access.
- **Integrity:** Ensuring data accuracy and protection against alterations.
- **Usability:** Preserving system performance while enforcing privacy measures.

Technologies such as encryption, differential privacy, federated learning, and blockchain serve as foundational components within this domain. These methodologies not only protect individual data but also facilitate advanced analytics and machine learning securely.

4.2 Privacy-Preserving Techniques

4.2.1 1. Encryption-Based Methods. Encryption forms the bedrock of privacy-preserving computing, ensuring that data remains unintelligible to unauthorized users, even if intercepted. Key approaches include:

- **Homomorphic Encryption:** This technique allows computations on encrypted data without requiring decryption. The results of such computations are also encrypted and can be decrypted by the data owner to reveal the correct output.
  - **Advantages:** Maintains data confidentiality throughout processing.
  - **Challenges:** Computationally intensive, particularly for complex operations.
  - **Applications:** Secure cloud computing, encrypted search.
- **End-to-End Encryption (E2EE):** Guarantees that data is encrypted from sender to receiver, preventing intermediaries from accessing it.
  - **Applications:** Messaging platforms such as WhatsApp and Signal.

4.2.2 2. Differential Privacy. Differential privacy introduces carefully calibrated noise to data or query responses to prevent the identification of individual data points, enabling aggregate analysis without exposing sensitive information.

- **Advantages:** Provides robust privacy guarantees while facilitating statistical analysis.

- **Challenges:** Balancing the level of noise with data utility.
- **Applications:** User data collection, recommendation systems, and public health research.

4.2.3 3. Federated Learning. Federated learning promotes decentralized machine learning by training models locally on user devices, eliminating the need to transfer raw data to a central server.

- **Advantages:** Preserves data locality and enhances privacy.
- **Challenges:** Communication overhead, model synchronization, and ensuring robustness against adversarial attacks.
- **Applications:** Healthcare (collaborative diagnosis), mobile applications (keyboard suggestions).

4.2.4 4. Secure Multi-Party Computation (SMPC). Secure Multi-Party Computation allows multiple parties to collaboratively compute a function based on their inputs while protecting the privacy of those inputs.

- **Advantages:** Provides strong security guarantees without necessitating a trusted third party.
- **Challenges:** Computationally demanding for large-scale systems.
- **Applications:** Joint market analysis, secure voting systems.

4.2.5 5. Blockchain for Privacy. Blockchain technology ensures tamper-proof data storage through cryptographic mechanisms and decentralized consensus methods. Privacy-enhancing features, including zero-knowledge proofs (ZKPs), can be integrated.

- **Advantages:** Offers transparency, immutability, and decentralized trust.
- **Challenges:** Scalability and energy consumption concerns.
- **Applications:** Supply chain management, secure data sharing.

4.2.6 6. Secure Hardware. Trusted Execution Environments (TEEs), such as Intel SGX, provide hardware-based isolation to safeguard sensitive computations.

- **Advantages:** High security with minimal software overhead.
- **Challenges:** Limited scalability and dependence on hardware vendors.
- **Applications:** Financial computations, secure cloud services.

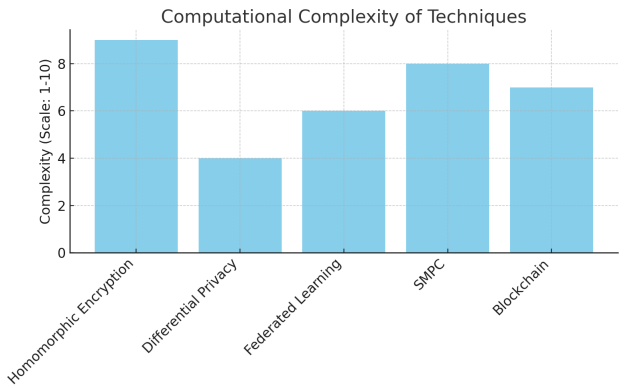

Figure 1: Computational complexity of different techniques

5 Analysis

Table 1: Key Metrics for Technique Evaluation

Technique	Complexity	Scalability	Privacy Guarantees
Homomorphic Encryption	High	Medium	Strong
Differential Privacy	Low to Medium	High	Medium
Federated Learning	Medium	High	Strong
Secure Multi-Party Computation	High	Low to Medium	Strong
Blockchain	Medium	Low to Medium	High

Based on the techniques discussed previously:

- **Homomorphic Encryption** and **Secure Multi-Party Computation (SMPC)** are best suited for high-security applications where confidentiality is critical, albeit with a potential impact on performance.
- **Differential Privacy** and **Federated Learning** strike a balance between privacy and usability, making them well-suited for real-world applications like healthcare and user analytics.
- **Blockchain** provides a solid foundation for privacy and transparency but faces challenges regarding scalability.
- **Secure Hardware** presents a pragmatic approach for isolated sensitive tasks, although it is restricted by physical infrastructure.

Table 2: Comparison of Privacy-Preserving Techniques

Technique	Features	Advantages	Challenges
Homomorphic Encryption	Computation on encrypted data	High confidentiality	High computational cost
Differential Privacy	Noise addition to data	Strong privacy guarantees	Balancing noise and utility
Federated Learning	Decentralized training	Preserves data locality	Communication overhead
Secure Multi-Party Computation	Joint computation, no data sharing	Strong privacy without trust	High computational complexity
Blockchain	Decentralized storage	Immutable and transparent	Scalability, energy consumption

6 Emerging Trends in Privacy-Preserving Computing

Table 3: Emerging Trends in Privacy-Preserving Computing

Trend	Description	Example Use Cases
Quantum-Resistant Cryptography	Protects against quantum computing threats	Financial systems
Federated Blockchain	Combines FL with blockchain for secure learning	IoT, healthcare
Privacy-Preserving AI	AI models trained on encrypted or noisy data	Personalized healthcare
Decentralized Identity	Self-sovereign identity systems based on blockchain	Secure digital identity management
Trusted Execution Environments	Hardware-secured environments for sensitive data	Cloud computing, banking

6.1 Quantum-Resistant Cryptography

As quantum computing advances, traditional encryption techniques like RSA and ECC may become vulnerable. Researchers are investigating quantum-resistant cryptographic algorithms, such as lattice-based cryptography, to protect future systems.

- **Impact:** Ensures long-term security for distributed systems in a post-quantum landscape.
- **Challenges:** Involves computational overhead and complexities in adoption.

6.2 Privacy-Preserving AI

Recent developments in machine learning are embedding privacy-preserving mechanisms within models. Techniques include:

- **Encrypted Model Training:** Training AI models on encrypted data using homomorphic encryption.
- **Privacy-Preserving Transfer Learning:** Protecting sensitive data while adapting pre-trained models for specific tasks.
- **Impact:** Revolutionizes applications in healthcare, finance, and personalized services.
- **Challenges:** Balancing model accuracy with privacy requirements.

6.3 Decentralized Identity (DID) Systems

Blockchain-based identity solutions, such as self-sovereign identity (SSI), empower users to control their data and share minimal necessary information.

- **Impact:** Reduces reliance on centralized identity providers while enhancing user privacy.
- **Challenges:** Issues with scalability and interoperability across platforms.

6.4 Secure Federated Learning Platforms

Federated learning is evolving with enhanced privacy mechanisms, including:

- Deployment of secure aggregation protocols to maintain data anonymity.
- Cross-silo and cross-device federated learning applications for diverse scenarios.
- **Impact:** Promotes privacy-focused collaboration among institutions without compromising data security.
- **Challenges:** High communication and computation costs in large-scale networks.

6.5 Trusted Execution Environments (TEEs)

Advancements in secure hardware technology are enhancing TEEs with innovations such as:

- Cryptographic operations accelerated by hardware.
- Integration of TEEs with cloud platforms for scalable, secure computing.
- **Impact:** Supports high-security applications in cloud and edge environments.
- **Challenges:** Dependence on vendor-specific solutions and limited availability.

6.6 Federated Blockchain

The integration of federated learning with blockchain results in decentralized learning frameworks that offer strong privacy and trust guarantees.

- **Impact:** Strengthens transparency and data ownership in cooperative learning settings.
- **Challenges:** Performance trade-offs due to the latency inherent in blockchain technology.

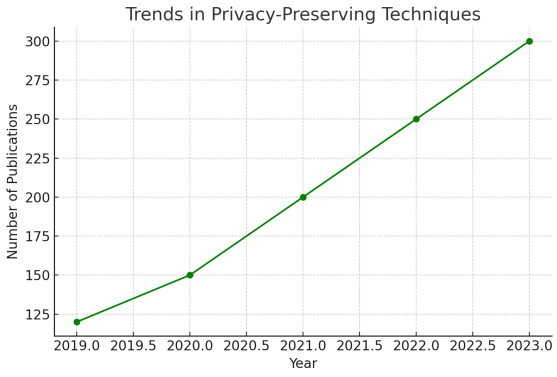

Figure 2: trends in privacy-preserving techniques by looking at the number of publications

7 Challenges and Open Issues

Table 4: Challenges and Open Issues

Category	Challenge	Description
Scalability	Communication overhead	Growth limits in federated learning
Adversarial Resistance	Model inversion attacks	Security risks in federated systems
Regulatory Compliance	Cross-border data laws	Limits to international data sharing
Computational Efficiency	High computational costs	Limits for real-time applications
Interoperability	Lack of standards	Integration challenges

7.1 Trade-Offs Between Privacy and Utility

Many privacy-preserving techniques require a compromise between privacy assurances and data utility. For example:

- Differential Privacy: Introducing noise can reduce the accuracy of statistical results or machine learning outputs.
- Homomorphic Encryption: Performing computations on encrypted data is slower than on plain text.

Open Issue: How can these methods be optimized to balance privacy and utility without major compromises?

7.2 Scalability

As distributed systems grow, scalability becomes increasingly critical:

- Federated Learning: The communication burden increases with more participant devices.

- Blockchain: Throughput diminishes as the number of network nodes grows.

Open Issue: How can these methodologies scale effectively to accommodate large networks or datasets?

7.3 Interoperability

Privacy-preserving solutions often need to integrate with existing systems and standards. A lack of interoperability poses challenges to widespread adoption. **Open Issue:** What strategies can ensure seamless integration of privacy-enhancing solutions with diverse platforms, frameworks, and protocols?

7.4 Adversarial Threats

Emerging adversarial techniques present challenges to the robustness of privacy-preserving methods:

- In federated learning, model inversion attacks can reconstruct private data.
- Side-channel attacks may exploit vulnerabilities in secure hardware.

Open Issue: How can systems designed to preserve privacy be made resilient against such adversarial threats?

7.5 Computational Overhead

Many privacy-preserving methods introduce significant computational costs:

- Homomorphic Encryption and SMPC: Notable latency is involved.
- Secure Hardware: Requires specialized infrastructure.

Open Issue: Can emerging algorithms or hardware advancements reduce computational expenses without compromising strong privacy standards?

7.6 Ethical and Regulatory Concerns

Global regulations like GDPR, HIPAA, and CCPA vary significantly, creating obstacles for international data sharing and collaboration.

Open Issue: How can privacy-preserving strategies align with varying regulatory requirements while enabling global systems?

8 Conclusion

Privacy-preserving computing is vital for maintaining data confidentiality, integrity, and usability in distributed systems. As reliance on cloud services, IoT, and machine learning grows, the demand for effective methods to address privacy concerns intensifies.

This survey analyzed significant approaches, including encryption-based methods, differential privacy, federated learning, secure multi-party computation, and blockchain-based strategies. Each method provides distinct advantages while confronting unique challenges.

Despite progress, unresolved issues persist, such as scalability, computational overhead, and resilience against adversarial threats. Emerging trends like quantum-resistant cryptography, decentralized identity frameworks, and federated blockchain systems show promise in addressing these challenges while unlocking new opportunities.

Future research should aim to develop scalable, efficient, and interoperable solutions that integrate smoothly with existing systems. Aligning these techniques with evolving regulatory frameworks is crucial for successful implementation.

Privacy-preserving computing has the potential to revolutionize sectors like healthcare, finance, and IoT by facilitating secure data sharing and collaboration. Continued innovation will be essential for protecting sensitive information and fostering trust in distributed systems.

Acknowledgments

A special thanks to Annajiat Alim Rasel sir for the guidance and dedication towards the CSE449 course.

References

[1] E. M. Campos. 2024. Intrusion Detection Based on Federated Learning for Internet of Things Scenarios. *Portal Investigacion UM* (2024). https://portalinvestigacion.um.es/documentos/6740d5ded21bb866449f5411

[2] Y. Di and H. Shi. 2024. Federated Cross-Domain Recommendation System Based on Bias Eliminator and Personalized Extractor. *Springer* (2024). https://link.springer.com/article/10.1007/s10115-024-02316-y

[3] A. Forootani and R. Iervolino. 2024. Asynchronous Federated Learning: A Scalable Approach for Decentralized Machine Learning. *arXiv* (2024). https://arxiv.org/abs/2412.17723

[4] A. Imteaj and S. Rezapour. 2024. Blockchain-Empowered Cyber-Secure Federated Learning for Trustworthy Edge Computing. *arXiv* (2024). https://arxiv.org/abs/2412.20674

[5] W. Kong and H. Ye. 2024. Privacy-Preserving Multi-VPPs Scheduling for Peak Ramp Minimization. *Elsevier* (2024). https://www.sciencedirect.com/science/article/pii/S0378779624012616

[6] J. Liu. 2024. Distributed Event-Triggered-Based Encrypted Control for Nonlinear Multi-Agent Systems via Privacy-Preserving Approach. *Springer* (2024). https://link.springer.com/article/10.1007/s11071-024-10779-5

[7] Y. Mao and W. Zhuang. 2024. Privacy-Aware Multi-Device Cooperative Edge Inference with Distributed Resource Bidding. *arXiv* (2024). https://arxiv.org/abs/2412.21069

[8] D. T. Nguyen. 2024. IoT Security: From Context-Based Authentication to Secure Federated Learning Anomaly Detection. *TU Darmstadt* (2024). https://tuprints.ulb.tu-darmstadt.de/28827/

[9] A. Rauch. 2024. Towards more scalable and privacy-preserving distributed asset transfer systems. *HAL* (2024). https://inria.hal.science/tel-04857796/

[10] V. S. S. N. Reddy. 2024. Leveraging Machine Learning for Personalization and Security in Content Management Systems. *ResearchGate* (2024). https://www.researchgate.net/publication/387156536_Leveraging_Machine_Learning_for_Personalization_and_Security_in_Content_Management_Systems

[11] I. Sharma and V. Khullar. 2024. Blockchain-Enabled Federated Learning-Based Privacy Preservation Framework for Secure IoT in Precision Agriculture. *Elsevier* (2024). https://www.sciencedirect.com/science/article/pii/S2452414X24002085

[12] Q. Wang, S. Chen, M. Wu, and X. Li. 2024. Digital Twin-Empowered Federated Incremental Learning for Non-IID Privacy Data. *IEEE* (2024). https://ieeexplore.ieee.org/abstract/document/10803097/

[13] L. Yu, Z. Wang, Y. Liu, and Z. Yang. 2024. Sampled-Data-Based Privacy-Preserving Scaled Consensus for Nonlinear Multiagent Systems: A Paillier Encryption Approach. *IEEE Transactions on Systems* (2024). https://ieeexplore.ieee.org/abstract/document/10816094/

[14] H. Zhu and Y. Fan. 2024. UA-PDFL: A Personalized Approach for Decentralized Federated Learning. *arXiv* (2024). https://arxiv.org/abs/2412.11674