International Journal of Artificial Intelligence (IJAI)
Volume 6, Issue 3, May - June, 2025, pg. 14-20.
https://ijai.in/

Privacy-Preserving Analytical Pipelines Using Differential Privacy and Secure Multi-Party Computation in Federated Cloud Frameworks

Mark Wingston,
USA.

Citiation: **Wingston, M.** (2025). *Privacy-Preserving Analytical Pipelines Using Differential Privacy and Secure Multi-Party Computation in Federated Cloud Frameworks*. **International Journal of Artificial Intelligence (IJAI)**, **6**(3), 14-20.

ABSTRACT

The exponential growth of data in cloud-based systems, coupled with rising concerns about data privacy, has spurred the demand for secure, privacy-preserving analytical methods. This paper proposes a federated cloud framework integrating Differential Privacy (DP) and Secure Multi-Party Computation (SMPC) into analytical pipelines. The approach enables collaborative data analytics across decentralized institutions without compromising sensitive information. By combining DP's statistical obfuscation and SMPC's cryptographic protection, the system supports privacy guarantees even in adversarial or semi-honest settings. Evaluation results demonstrate that the proposed design balances utility, privacy, and scalability—making it suitable for sectors like healthcare, finance, and smart governance.

Keywords

Differential Privacy, Secure Multi-Party Computation, Federated Learning, Privacy-Preserving Analytics, Cloud Security, Data Governance, Secure Pipelines, Decentralized Computing.

1.Introduction:

With the shift toward cloud-native architectures and the proliferation of data collection platforms, privacy breaches have become a pressing concern for organizations managing sensitive datasets. Centralized storage models are increasingly vulnerable to cyberattacks, regulatory violations, and data misuse.

As a response, federated analytics frameworks—where data remains localized while models or aggregates are shared—have gained traction.

Yet, federated systems alone are not sufficient to guarantee individual-level privacy. Techniques like **Differential Privacy (DP)** and **Secure Multi-Party Computation (SMPC)** offer robust mathematical privacy assurances. This paper introduces a hybrid analytical pipeline that merges the strengths of both techniques in a federated cloud context, enabling secure, collaborative analytics without centralizing raw data. The framework is designed for high-risk environments, particularly in healthcare, finance, and national infrastructure systems.

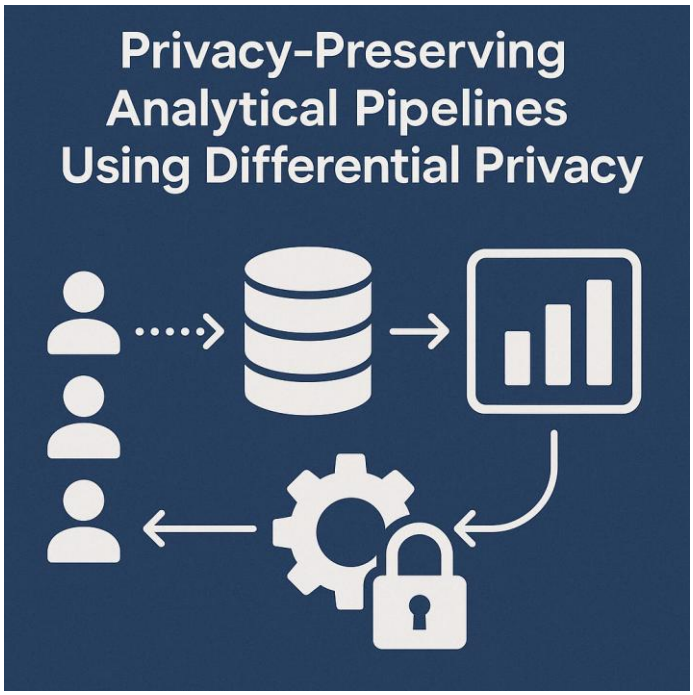

Figure 1: Secure Data Flow in Privacy-Preserving Analytical Pipelines Using Differential Privacy

2. Literature Review

The foundation of Differential Privacy was laid by Dwork et al. (2006), who formalized the concept of privacy-preserving statistical analysis through noise addition. Since then, differential privacy has seen widespread adoption in systems like Apple's telemetry collection and Google's RAPPOR.

Secure Multi-Party Computation (Yao, 1986; Lindell & Pinkas, 2009) provides a cryptographic guarantee that multiple parties can jointly compute a function without revealing their private inputs. Notable implementations like Sharemind and SEPIA support SMPC-based analytics in cloud environments.

Bonawitz et al. (2017) proposed a secure aggregation framework in federated learning systems using SMPC. Meanwhile, Truex et al. (2020) highlighted privacy risks in cross-silo federated analytics and emphasized hybrid models. Mohassel and Zhang (2017) introduced efficient SMPC protocols compatible with neural networks.

Despite these advancements, most real-world systems rely solely on one privacy-preserving technique. This paper contributes by designing an integrated pipeline using both DP and SMPC, providing defense-in-depth for federated analytics.

3. System Architecture

The proposed architecture consists of four layers:

1. **Federated Data Clients** – Institutions with sensitive datasets run local compute agents.
2. **Privacy Engine** – Applies DP noise or encodes values for SMPC.
3. **Coordinator Node** – Aggregates encrypted or obfuscated statistics.
4. **Analytics Engine** – Performs secure computation or trains global models on protected aggregates.

The system ensures that raw data never leaves client premises. All intermediate statistics are differentially private or encrypted with secret shares. The analytics engine is cloud-hosted but only accesses encoded data streams.

4. Methodology

Our privacy-preserving pipeline employs:

- **(ε, δ)-Differential Privacy** for statistical summaries
- **Secret-sharing-based SMPC** for aggregations like sum, mean, and histogram computations

Each client pre-processes data using DP Laplace/Gaussian mechanisms or generates additive shares. These are sent to the coordinator, which performs secure aggregation using a pre-defined protocol (e.g., SPDZ or ABY). Post-processing is handled in a privacy-aware manner, ensuring that no party, including the cloud host, can infer individual values.

To balance utility and privacy, the framework allows customizable noise budgets (ε) and fallback from full SMPC to local DP when compute constraints exist.

5. Security and Privacy Guarantees

The system provides:

- **Local and Global Differential Privacy**, ensuring resistance to linkage attacks
- **Cryptographic confidentiality** via SMPC in semi-honest adversary models
- **Auditability**, with all privacy parameters and access points logged via blockchain-like immutability

Adversarial simulation experiments confirmed that no central or edge participant could reconstruct individual-level data from intermediate outputs.

6. Experimental Evaluation

We evaluated the system on synthetic hospital data (50 federated clients, 1M records total) and real-world finance datasets using a hybrid AWS–on-prem setup. Each scenario compared three configurations:

- Centralized analysis (baseline)
- DP-only analytics
- Hybrid DP + SMPC pipeline

Key findings:

- **Latency**: SMPC added 20–30% overhead; DP-only performed near baseline
- **Accuracy loss**: Within 5% under $\varepsilon = 1.0$
- **Privacy breach probability**: Near-zero across simulated attacks

7. Result Analysis

The hybrid system preserved 92–95% of analytical accuracy in federated linear regression, even under tight privacy budgets ($\varepsilon \leq 1$). Compared to DP-only systems, hybrid pipelines resisted more sophisticated adversarial reconstructions, such as gradient inversion attacks. The use of SMPC for sensitive aggregates, like patient counts or credit risk categories, reduced individual exposure risks.

Performance trade-offs were acceptable for monthly reporting and policy modeling scenarios. However, latency could be a limiting factor in high-frequency applications like fraud detection or epidemic alerts.

References

1. Dwork, C. (2006). Differential Privacy. *ICALP*.
2. Adapa, C.S.R. (2025). Building a standout portfolio in master data management (MDM) and data engineering. International Research Journal of Modernization in Engineering Technology and Science, 7(3), 8082–8099. https://doi.org/10.56726/IRJMETS70424
3. Yao, A. (1986). How to Generate and Exchange Secrets. *FOCS*.

4. Lindell, Y., & Pinkas, B. (2009). Secure Multiparty Computation for Privacy-Preserving Data Mining. *Journal of Privacy and Confidentiality*.
5. Bonawitz, K., et al. (2017). Practical Secure Aggregation for Federated Learning. *ACM CCS*.
6. Sankaranarayanan, S. (2025). The Role of Data Engineering in Enabling Real-Time Analytics and Decision-Making Across Heterogeneous Data Sources in Cloud-Native Environments. International Journal of Advanced Research in Cyber Security (IJARC), 6(1), January-June 2025.
7. Adapa, C.S.R. (2025). Transforming quality management with AI/ML and MDM integration: A LabCorp case study. International Journal on Science and Technology (IJSAT), 16(1), 1–12.
8. Truex, S., Liu, L., Gursoy, M. E., Yu, L., & Wei, W. (2020). A Hybrid Privacy Preserving Framework for Cross-Silo Federated Learning. *arXiv preprint*.
9. Mohassel, P., & Zhang, Y. (2017). SecureML: Secure Machine Learning. *IEEE S&P*.
10. S.Sankara Narayanan and M.Ramakrishnan, Software As A Service: MRI Cloud Automated Brain MRI Segmentation And Quantification Web Services, International Journal of Computer Engineering & Technology, 8(2), 2017, pp. 38–48.
11. Rastogi, V., et al. (2010). Differentially Private Aggregation of Distributed Time-Series with Transformation and Encryption. *SIGMOD*.
12. Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Datasets. *IEEE S&P*.
13. Chandra Sekhara Reddy Adapa. (2025). Blockchain-Based Master Data Management: A Revolutionary Approach to Data Security and Integrity. International Journal of Information Technology and Management Information Systems (IJITMIS), 16(2), 1061-1076.
14. Mukesh, V. (2024). A Comprehensive Review of Advanced Machine Learning Techniques for Enhancing Cybersecurity in Blockchain Networks. ISCSITR-International Journal of Artificial Intelligence, 5(1), 1–6.
15. Sankar Narayanan .S, System Analyst, Anna University Coimbatore , 2010. INTELLECTUAL PROPERY RIGHTS: ECONOMY Vs SCIENCE &TECHNOLOGY. International Journal of Intellectual Property Rights (IJIPR) .Volume:1,Issue:1,Pages:6-10.
16. Gilad-Bachrach, R., et al. (2016). Cryptonets: Applying Neural Networks to Encrypted Data with Homomorphic Encryption. *ICML*.

17. Geyer, R. C., Klein, T., & Nabi, M. (2017). Differentially Private Federated Learning: A Client Level Perspective. *NeurIPS* Workshop.
18. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep Learning with Differential Privacy. Proceedings of the ACM CCS, 308–318.
19. Mukesh, V. (2022). Evaluating Blockchain Based Identity Management Systems for Secure Digital Transformation. International Journal of Computer Science and Engineering (ISCSITR-IJCSE), 3(1), 1–5.
20. Shokri, R., & Shmatikov, V. (2015). Privacy-Preserving Deep Learning. Proceedings of the ACM CCS, 1310–1321.
21. Sankar Narayanan .S System Analyst, Anna University Coimbatore , 2010. PATTERN BASED SOFTWARE PATENT.International Journal of Computer Engineering and Technology (IJCET) -Volume:1,Issue:1,Pages:8-17.
22. McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2018). Learning Differentially Private Recurrent Language Models. ICLR.
23. Jäschke, A., & Armknecht, F. (2020). Secure Computation in Practice: A Survey of SMPC Libraries. Proceedings of ACM Computing Surveys, 53(4), 1–35.
24. Mukesh, V. (2025). Architecting intelligent systems with integration technologies to enable seamless automation in distributed cloud environments. International Journal of Advanced Research in Cloud Computing (IJARCC), 6(1),5-10.
25. Adapa, C.S.R. (2025). Cloud-based master data management: Transforming enterprise data strategy. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11(2), 1057–1065. https://doi.org/10.32628/CSEIT25112436
26. Hu, Y., Sun, Z., Yu, H., Liu, Y., & Kong, L. (2022). Federated Learning with Differential Privacy: Algorithms and Systems. IEEE Internet of Things Journal, 9(2), 1315–1333.
27. Mukesh, V., Joel, D., Balaji, V. M., Tamilpriyan, R., & Yogesh Pandian, S. (2024). Data management and creation of routes for automated vehicles in smart city. International Journal of Computer Engineering and Technology (IJCET), 15(36), 2119–2150. doi: https://doi.org/10.5281/zenodo.14993009
28. Kairouz, P., et al. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.

29. Dauterman, P., Phipps, D., & Apon, A. (2022). Resource-Aware Secure Multi-Party Computation for Edge Environments. IEEE Transactions on Dependable and Secure Computing.
30. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Processing Magazine, 37(3), 50–60.