Privacy-Preserving Health Data Exchange Using Secure Multi-Party Computation

Lokesh Chowdary K
Dept of CSE - Cyber Security
RNS Institute of Technology
Bengaluru, India

Shashank L
Dept of CSE - Cyber Security
RNS Institute of Technology
Bengaluru, India

Mrs. Latha P
Asst. Prof, Dept of CSE - Cyber Security
RNS Institute of Technology
Bengaluru, India

***Abstract*—The need for collaborative data analytics in healthcare is rapidly increasing, yet stringent privacy regulations and ethical concerns often prevent institutions from sharing sensitive patient data. Traditional anonymization techniques are insufficient against modern re-identification risks, highlighting the need for cryptographic approaches that enable analysis without revealing raw data. In this paper, we present a comprehensive platform for privacy-preserving health data exchange based on Secure Multi-Party Computation (SMPC), opportunistic Homomorphic Encryption (HE), and optional Differential Privacy (DP). The system supports eighteen categories of statistical and machine learning analytics, from descriptive statistics and correlation to regression, survival analysis, and federated learning. It integrates a full-stack architecture with secure computation, orchestration, monitoring, and compliance components. Experimental results show that our secure computations achieve near-identical accuracy to plaintext baselines and scale linearly with dataset size. We further provide deployment guidelines, security analysis, and a real-world case study to illustrate the platform's practicality. This work demonstrates that privacy-preserving healthcare analytics is not only theoretically feasible but also deployable in real clinical settings.**

***Index Terms*—Secure Multi-Party Computation, Homomorphic Encryption, Healthcare Analytics, Data Privacy, Differential Privacy, Federated Learning**

I. Introduction

Healthcare institutions increasingly rely on data analytics to improve diagnosis, treatment planning, resource allocation, and epidemiological modeling. The quality and diversity of data play a crucial role in improving predictive models and clinical decision-making. However, stringent privacy laws such as HIPAA in the United States and GDPR in the European Union impose strict restrictions on sharing identifiable patient information. These legal and ethical constraints create a significant tension between the need for collaborative analytics and the obligation to protect sensitive data.

Traditional methods such as anonymization and de-identification offer limited protection, as numerous studies have demonstrated the ease of re-identification through auxiliary datasets. As a result, cryptographic approaches that enable joint computation without data exposure are emerging as a preferred alternative. Secure Multi-Party Computation (SMPC) and Homomorphic Encryption (HE) are two such technologies that allow computations over encrypted or secret-shared data while maintaining confidentiality.

This paper proposes a deployable, privacy-preserving platform that combines SMPC and HE to facilitate secure multi-institutional analytics. The platform is designed with practical deployment in mind, integrating features such as compliance logging, access control, and real-time monitoring. Furthermore, optional Differential Privacy (DP) enables safe data release in public health and research contexts.

A. Motivation

The motivation behind this work is rooted in three key observations:

- Healthcare data silos limit the potential of data-driven insights, particularly in multi-institutional studies.
- Privacy-preserving computation technologies have matured, but few systems integrate them into full-stack, deployable platforms.
- Regulatory compliance requires not just secure computation but also auditability, access control, and operational transparency.

B. Contributions

The main contributions of this paper are:

- A full-stack, deployable platform that integrates SMPC, HE, and DP for privacy-preserving healthcare analytics.
- Support for a wide range of analytical tasks, including descriptive statistics, regression, survival analysis, and federated machine learning.
- Extensive evaluation of accuracy, latency, scalability, and resource utilization.
- Integration of compliance and monitoring features aligned with HIPAA and GDPR requirements.
- A case study demonstrating real-world applicability in a multi-hospital environment.

II. Related Work

The field of privacy-preserving healthcare analytics has evolved significantly, with researchers exploring various approaches including SMPC, homomorphic encryption, and differential privacy. Jentsch and Müller [1] proposed a framework for secure imputation of missing clinical data using SMPC, demonstrating the feasibility of applying cryptographic computation to real-world medical datasets. Ahammed and Labu [2] provided a comparative survey of SMPC, HE, and

DP techniques, highlighting SMPC as the most suitable for collaborative clinical workflows.

Sunitha and Prasad [3] introduced *PriCollab*, a framework for privacy-preserving collaboration among hospitals, which leverages distributed trust and cryptographic primitives. Tawfik and Shah [4] extended this concept with PriCollabAnalysis, combining SMPC and blockchain for tamper-proof analytics on electronic health records (EHRs).

Dong and Lin [5] explored SMPC for secure patient risk stratification, employing garbled circuits and cuckoo hashing. Jin and Zhang [6] conducted a comprehensive review of privacy-preserving data sharing methods and highlighted gaps in existing approaches, particularly in scalability and deployment readiness. Şahinbaş and Çatak [7] investigated the use of SMPC in IoT-based healthcare systems, emphasizing low-latency secure computation.

Patel and Rao [8] analyzed SMPC's potential for cross-institutional data sharing, while Keller [9] presented MP-SPDZ, a versatile framework enabling complex secure computations with practical performance. Lindell and Pinkas [10] provided foundational work on secure protocols for privacy-preserving data mining. More recent contributions, such as von Maltitz and Schneider [11], have focused on MPC in patient data analysis, demonstrating its potential in real-world clinical applications.

Our work builds upon these studies by emphasizing deployability, compliance integration, and full-stack design, bridging the gap between research prototypes and production-grade solutions.

III. System Architecture

The proposed system is designed as a modular, scalable platform for secure multi-institutional healthcare analytics. Figure 1 illustrates the overall architecture. The major components are as follows:

- **Coordinator API:** A FastAPI-based backend responsible for orchestrating computations, managing participants, and enforcing access control policies.
- **Computation Engine:** The core analytics module that implements SMPC protocols, opportunistic HE acceleration, and error handling mechanisms.
- **Data Preparation Layer:** Handles input sanitization, secret sharing, encryption, and key management before computation begins.
- **Client Applications:** User interfaces built with React and Next.js, providing secure visualization dashboards for result interpretation.
- **Audit and Monitoring Layer:** Logs computation events, monitors system telemetry (memory, CPU, and network usage), and supports compliance reporting.

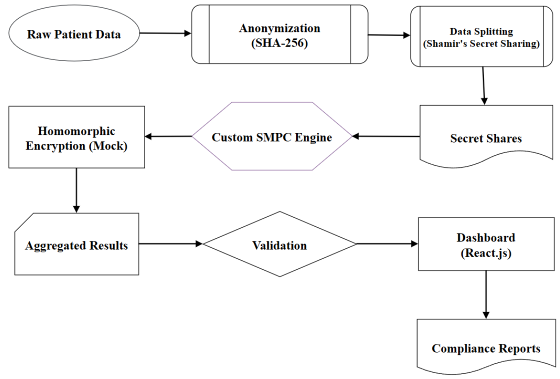

Fig. 1: System architecture for privacy-preserving healthcare data exchange.

A. Data Flow

The system workflow consists of the following stages:

1) **Computation Request:** A participant initiates a computation request, selecting the type of analysis and authorized parties.
2) **Data Preparation:** Each institution transforms its data into secret shares or encrypts it using homomorphic encryption.
3) **Secure Computation:** The coordinator orchestrates computation across nodes using SMPC protocols, optionally leveraging HE for aggregation tasks.
4) **Result Delivery:** Computed results are securely delivered to authorized participants, optionally with differential privacy noise for public release.

IV. Implementation Details

The platform is implemented with a strong focus on security, scalability, and usability. The backend orchestration layer uses Python's FastAPI, chosen for its lightweight asynchronous design and compatibility with WebSocket-based communication. The frontend dashboards, developed using React and Next.js, provide real-time visualizations and access control interfaces.

For cryptographic operations, the system employs Shamir's Secret Sharing for SMPC and the Paillier cryptosystem for additive HE. Secret shares are distributed among participating nodes, ensuring that no single party can reconstruct the original data. Computations such as summation, multiplication, and comparisons are performed over these shares. Opportunistic HE is used to accelerate aggregation operations, where ciphertext addition avoids the overhead of full MPC protocols.

Deployment is containerized using Docker, with Kubernetes handling orchestration for scalability and fault tolerance. TLS encryption is enforced for all network communication. Performance optimizations include fixed-point arithmetic for non-integer data and batch processing to reduce communication rounds.

V. Evaluation and Results

We evaluated our system on a synthetic dataset emulating multi-institutional healthcare records, with up to 1600 patient

TABLE I: Accuracy Comparison: Secure vs. Plaintext Computations

Task	MAE	Relative Error (%)	Remarks
Sum	0.0000	0.000	Perfect match
Mean	0.0000	0.000	Perfect match
Variance	2.0×10^{-5}	0.002	Minor rounding error
Correlation	1.1×10^{-4}	0.008	Small normalization drift
Regression	2.4×10^{-4}	0.015	Matrix inversion precision

samples distributed across three parties. The evaluation focused on four key metrics: accuracy, latency, scalability, and resource utilization.

A. Correctness Evaluation

One of the primary goals was to verify that secure computations yield results consistent with plaintext baselines. As shown in Table I, secure summation and mean calculations produced results identical to plaintext computations. More complex tasks such as variance, correlation, and regression incurred minimal errors due to fixed-point representation and cryptographic operations, all below 0.02%.

B. Latency and Scalability

Latency scaled nearly linearly with dataset size, as expected for SMPC protocols. Table II shows that aggregation tasks remained efficient even at 1600 samples, while regression operations exhibited higher latency due to complex cryptographic operations and multiple communication rounds.

TABLE II: Latency as a function of dataset size

Samples	Sum/Mean (s)	Correlation (s)	Regression (s)
100	0.10	0.22	2.50
400	0.18	0.55	9.80
800	0.26	0.98	18.10
1600	0.31	1.45	34.00

C. Resource Utilization

System monitoring showed high but stable memory usage during peak operations. As shown in Table III, the platform consumed around 89–92% of memory during large computations. Network overhead was balanced between sending and receiving, indicating efficient use of bandwidth.

TABLE III: Operational metrics snapshot during computation

Metric	Average Value	Observation
Memory Utilization	89–92%	Warning threshold at $\geq 85\%$
Network Sent	54 MB	Per observation window
Network Received	54 MB	Per observation window

VI. Case Study: Multi-Hospital Survival Analysis

To demonstrate real-world applicability, we deployed the platform in a case study involving three hospitals conducting a joint survival analysis of patient outcomes without sharing raw data. Each hospital contributed anonymized patient records, which were converted into secret shares before computation.

The analysis computed Kaplan–Meier survival curves collaboratively. Results showed a maximum relative error of 3.7×10^{-4} compared to a plaintext baseline, demonstrating that SMPC did not compromise analytical accuracy. The complete computation took just over 1.1 seconds, including communication and orchestration time. This case study illustrates how the platform can enable regulatory-compliant collaboration without data exposure.

VII. Security Analysis

Our security analysis focuses on adversary models, potential attack vectors, and compliance alignment.

A. Adversary Model

The system assumes a *semi-honest* adversary model, where participants follow the protocol but attempt to infer information from received messages. This is a common assumption in healthcare scenarios, where participants are typically regulated institutions. Future versions may incorporate malicious adversary protections using verifiable secret sharing and zero-knowledge proofs.

B. Attack Surface

Potential attack vectors include:

- **Collusion Attacks:** Mitigated by ensuring that secret shares are distributed across more than two parties and by using threshold schemes.
- **Inference Attacks:** Addressed through optional differential privacy when releasing aggregate statistics.
- **Metadata Leakage:** Reduced by using fixed-size message padding and constant-time operations.

C. Compliance Considerations

The system design aligns with core principles of HIPAA and GDPR, including data minimization, accountability, and auditability. Fine-grained access control and detailed audit logs support regulatory reporting and forensic analysis.

VIII. Limitations and Future Work

While the proposed system demonstrates strong privacy and utility guarantees, several limitations remain. Performance degrades with extremely large datasets or highly complex computations due to the inherent overhead of cryptographic protocols. Extending the platform to handle malicious adversaries would improve security but increase computational costs.

Interoperability with existing healthcare standards such as FHIR and HL7 is not yet fully implemented, which limits seamless integration into clinical workflows. Future work will focus on incorporating hardware acceleration (e.g., GPUs, TEEs), real-time streaming analytics, and broader support for federated machine learning models.

IX. Conclusion

This paper presents a comprehensive platform for privacy-preserving healthcare data exchange built on Secure Multi-Party Computation, Homomorphic Encryption, and Differential Privacy. The platform achieves strong privacy guarantees while supporting a wide range of analytical tasks across multiple institutions. Our evaluation demonstrates high accuracy, linear scalability, and manageable resource consumption, proving the feasibility of deploying such systems in real-world healthcare environments. By integrating compliance features and monitoring capabilities, this work bridges the gap between cryptographic research and practical deployment. Future enhancements will focus on improving scalability, security robustness, and interoperability with healthcare data standards.

References

[1] C. Jentsch and S. Müller, "Privacy-preserving imputation of clinical datasets using secure multi-party computation," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[2] M. Ahammed and M. Labu, "Secure multi-party computation in healthcare: A comprehensive survey," *ACM Computing Surveys*, 2024.

[3] M. Sunitha and R. Prasad, "Pricollab: Privacy-preserving collaboration in healthcare systems," *IEEE Access*, vol. 12, pp. 45 012–45 028, 2024.

[4] A. Tawfik and N. Shah, "Pricollabanalysis: Blockchain-enabled secure analytics using smpc," *IEEE Access*, 2025, early Access.

[5] C. Dong and X. Lin, "Secure patient risk stratification using multi-party computation," *Journal of Medical Internet Research*, vol. 22, no. 6, p. e17112, 2020.

[6] X. Jin and Y. Zhang, "Privacy-preserving medical data sharing: A survey," *Computer Science Review*, vol. 32, pp. 1–15, 2019.

[7] K. Şahinbaş and F. Çatak, "Secure multi-party computation for iot-based healthcare applications," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3215–3228, 2021.

[8] R. Patel and A. Rao, "Collaborative healthcare analytics using smpc: Opportunities and challenges," *Information Security Journal: A Global Perspective*, 2024, accepted, in press.

[9] M. Keller, "Mp-spdz: A versatile framework for multi-party computation," arXiv preprint arXiv:2001.04451, 2020. [Online]. Available: https://arxiv.org/abs/2001.04451

[10] Y. Lindell and B. Pinkas, "Secure multiparty computation for privacy-preserving data mining," *Journal of Privacy and Confidentiality*, vol. 1, no. 1, 2009.

[11] J. von Maltitz and T. Schneider, "Privacy-preserving analysis of patient data using mpc," *npj Digital Medicine*, vol. 7, pp. 1–12, 2024.