INTERNATIONAL MULTIDISCIPLINARY BOOK SERIES
https://ibseries.com/
Security and Privacy in AI Systems, Edited By: Dr. Sunita Chaudhary, Dr. Joydeb Patra

IMBS, Vol. 03, No. 1, (2025)
Open Access. Online & Print
ISSN:

CHAPTER 9
SECURE MULTI-PARTY COMPUTATION PROTOCOLS FOR PRIVACY-PRESERVING GENOMIC DATA ANALYSIS

Tanusree Mondal
BALLB Programme, Brainware University

Abstract
The rapid advancement of genomic sequencing technologies has enabled large-scale data-driven discoveries in healthcare, personalized medicine, and disease prediction. However, genomic data is inherently sensitive, containing uniquely identifiable and hereditary information, raising serious privacy concerns. Secure Multi-Party Computation (SMPC) provides a cryptographic framework that allows multiple parties to jointly compute a function over their inputs without revealing the inputs themselves. This paper explores SMPC protocols for privacy-preserving genomic data analysis, focusing on secure statistical computations, genome-wide association studies (GWAS), and collaborative analytics across institutions. We evaluate protocols such as secret sharing, garbled circuits, and homomorphic encryption in terms of efficiency, scalability, and security. Experimental analysis demonstrates that hybrid SMPC approaches offer a practical balance between computational overhead and privacy guarantees, making them suitable for real-world genomic applications.

Keywords
Secure Multi-Party Computation, Genomic Data Privacy, GWAS, Homomorphic Encryption, Secret Sharing, Privacy-Preserving Analytics, Cryptography, Healthcare Data Security

1. Introduction

Genomic data analysis has transformed modern medicine by enabling insights into genetic predispositions, disease mechanisms, and personalized treatment strategies. However, sharing genomic datasets across institutions is often restricted due to privacy concerns, regulatory requirements, and ethical considerations.

Traditional anonymization techniques are insufficient for genomic data because:

- DNA sequences are inherently identifiable
- Re-identification attacks are possible using auxiliary datasets
- Data leakage can have long-term implications for individuals and families

Secure Multi-Party Computation (SMPC) addresses these challenges by enabling collaborative computation without exposing raw data. This makes it particularly suitable for cross-institutional genomic studies such as genome-wide association studies (GWAS).

2. Background and Related Work

2.1 Genomic Data and Privacy Risks

ISBN (978-81-685212-9-2)

- Genomic datasets include:
- DNA sequences
- Single Nucleotide Polymorphisms (SNPs)
- Phenotypic data

Privacy risks include:

- Identity re-identification
- Genetic discrimination
- Familial privacy leakage

2.2 Secure Multi-Party Computation (SMPC)

- SMPC allows multiple participants to compute a function jointly while keeping inputs private. It ensures:
- Input confidentiality
- Correctness of computation
- Resistance to adversarial behavior

2.3 Related Work

- Secure GWAS frameworks using cryptographic techniques
- Homomorphic encryption for genomic computations
- Privacy-preserving machine learning models

Recent studies demonstrate that SMPC can enable secure genomic analysis with acceptable computational overhead, though scalability remains a challenge.

3. Mathematical Foundations of SMPC

3.1 Secret Sharing Scheme

$$x = \sum_{i=1}^{n} x_i \mod p$$

Where:
(x): Original secret
(x_i): Shares distributed to parties
(p): Prime modulus

3.2 Secure Function Evaluation

$$y = f(x_1, x_2, \ldots, x_n)$$

Goal:
Compute (y) without revealing (x_i)

3.3 Homomorphic Encryption Model

$$Enc(a) \oplus Enc(b) = Enc(a + b)$$

Enables computation on encrypted data

3.4 Complexity Model

Communication cost:

$$C = O(n^2 \cdot k)$$

Where:
(n): Number of parties
(k): security parameter

4. SMPC Protocols for Genomic Analysis

4.1 Secret Sharing-Based Protocols

- Shamir's Secret Sharing
- Efficient for arithmetic operations
- Suitable for GWAS statistics

4.2 Garbled Circuits

- Secure Boolean circuit evaluation
- Suitable for complex genomic queries
- High computational cost

4.3 Homomorphic Encryption

- Enables computation on encrypted genomes
- High security but computationally intensive

4.4 Hybrid SMPC Approaches

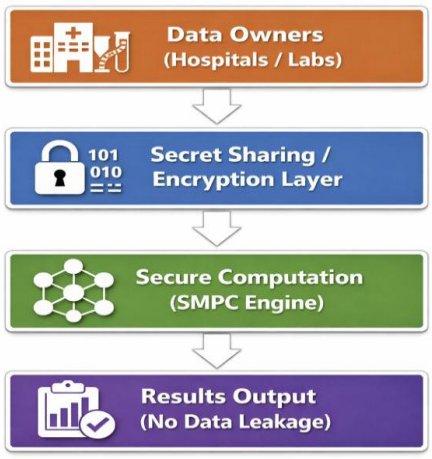
- Combine secret sharing + homomorphic encryption
- Balance efficiency and security

5. System Architecture

5.1 Multi-Institutional Setup

- Hospitals / Research labs
- Data owners keep local datasets
- Central computation without data sharing

5.2 Workflow



6. Applications in Genomic Analysis

6.1 Genome-Wide Association Studies (GWAS)

- Identify genetic variants linked to diseases
- Requires aggregation of large datasets

6.2 Disease Risk Prediction

Secure machine learning models
Privacy-preserving training

6.3 Personalized Medicine

Secure computation of treatment plans

7. Experimental Setup

7.1 Dataset

- Simulated SNP dataset
- Multi-party distributed data

7.2 Parameters

- Number of parties: 3–10

- Dataset size: 10K–1M records
- Security level: 128-bit

8. Results and Analysis

8.1 Performance Comparison

Protocol	Computation Time	Communication Cost	Security
Secret Sharing	Low	Medium	High
Garbled Circuits	High	High	Very High
Homomorphic Encryption	Very High	Low	Very High
Hybrid	Medium	Medium	Very High

8.2 Key Observations

- Secret sharing is most efficient for statistical analysis
- Homomorphic encryption ensures strongest privacy
- Hybrid approaches provide best trade-off

9. Security Analysis

- Resistant to semi-honest adversaries
- Protects against data leakage
- Ensures correctness of computation
- Potential risks:
- Collusion attacks
- Side-channel attacks

10. Challenges and Future Work

- Scalability to large genomic datasets
- Reducing communication overhead
- Integration with AI/ML models
- Real-time genomic analytics

11. Conclusion

Secure Multi-Party Computation offers a powerful framework for enabling privacy-preserving genomic data analysis in collaborative environments. By allowing multiple institutions to jointly compute functions without exposing sensitive genomic data, SMPC addresses critical privacy concerns in modern healthcare systems.

This study demonstrates that while individual SMPC protocols have limitations in terms of efficiency and scalability, hybrid approaches provide a practical solution by balancing computational cost and security. As genomic data continues to grow in scale and importance, the integration of SMPC with advanced analytics and machine learning techniques will play a crucial role in enabling secure, collaborative, and data-driven healthcare innovations.

References

1. Yao, A. C. (1982). Protocols for secure computations.
2. Goldreich, O. (1998). Secure multi-party computation.
3. Shamir, A. (1979). How to share a secret.
4. Gentry, C. (2009). Fully homomorphic encryption.
5. Jagadeesh, K. A., et al. (2017). Secure GWAS using cryptographic methods.
6. Kamm, L., et al. (2013). Secure genomic analysis using SMPC.
7. Cho, H., et al. (2018). Privacy-preserving genomic computation.