Secure Multi-Party Computation for Collaborative Data Analysis

Oliv J Patel[1], Dhruvil R Patel[2], Riddhi A Mehta[3]

1,2 Student, 3 Assistant Professor
Computer Science & Engineering
Parul Institute of Technology, Parul University
Vadodara, Gujarat, India

Abstract - *A powerful encryption mechanism called Secure Multi-Party Computation (SMPC) was developed to allow many participants to collaborate and perform data analysis tasks while maintaining the privacy and secrecy of their personal information. In many fields, such as health care, finance, and social sciences, where multiple stakeholders need to exchange and evaluate sensitive information without disclosing it to others, collaborative data analysis is becoming more common. This study provides an in-depth study of SMPC for group data analysis. The main goal is to provide a comprehensive understanding of the leading ideas, protocols, and applications of SMPC, while highlighting the benefits and challenges it brings to promote secure collaboration between different data owners. In summary, this study provides a comprehensive and up-to-date study on secure multiparty computing in collaborative data review. It provides a comprehensive overview of SMPC implementation issues and the underlying ideas, protocols, and applications. The article is intended to be a useful resource for researchers, practitioners, and policy makers interested in using SMPC to facilitate group data analysis while protecting confidentiality and privacy.*

1.INTRODUCTION

The paper first introduces the basic ideas of SMPC, such as security function evaluation, secret sharing, and cryptographic primitives. It looks at how these ideas could be used to facilitate collaborative analysis without revealing private information. Yao's Adversary Chains, Secure Multiparty Computation via Boolean Chains (SMC-BC), and Fully Homomorphic Encryption (FHE) are just a few of the SMPC protocols that are explored in depth and their advantages and disadvantages in various situations. In addition, the research examines the precise uses of SMPC in the analysis of group data. It looks at scenarios where multiple hospitals can collaborate to examine patient data for medical research while protecting patient privacy. In addition, it explores financial scenarios where multiple agencies could work together to identify money laundering trends without revealing specific client activities. The threat model, assumptions and level of protection offered by various protocols are all discussed in detail in relation to SMPC security features. The study also discusses the trade-offs between privacy and efficiency and highlights the computational and communication costs associated with SMPC. Difficulties and unresolved research issues in analyzing group data at SMPC are also noted. Scalability issues, performance optimization, dynamic party management and fighting malicious adversaries are some of them. The report suggests potential research opportunities to address these issues and increase the use of SMPC in the real world. The study analyses case studies and real-world applications to ensure the effectiveness of SMPC. It tells about the effective use of SMPC in many fields, demonstrates its applicability and highlights the implications of these experiences. The need for secure and privacy-friendly data analysis methods has grown in the era of big data and collaborative research. Traditional data analysis techniques fail to provide adequate protection against privacy breaches and unauthorized access due to the collection and exchange of sensitive data between multiple parties. Secure Multi-Party Computation (MPC), developed to address these problems, has proven to be an effective solution because many parties can collaboratively compute functions based on their private data without revealing background information.

Secure MPC, also known as Secure Multiparty Computing, is a cryptographic system that allows multiple parties to perform calculations on their shared data while protecting the privacy and confidentiality of individual inputs. MPC enables decentralized computing where each partner retains ownership of its data, unlike traditional systems that require data sharing or outsourcing of computations to a central server. The primary purpose of Secure MPC is to enable collaborative data analysis while protecting confidentiality and privacy. MPC guarantees that each partner's inputs and intermediate calculations remain private during the analysis process using various encryption methods, including homomorphic encryption, secret sharing, and security protocols. It allows companies, researchers, and individuals to collaborate and analyses integrated datasets without having to reveal or disclose their sensitive data. The concept of "privacy by design" is one of the core ideas of Secure MPC. This demonstrates that security and privacy considerations are considered during the development and use of computing protocols. This ensures that privacy is preserved by default and removes the need for additional layers of security that can be vulnerable to errors or mistakes. Secure MPC implements privacy through carefully designed encryption algorithm selection, secure key management, and thorough protocol testing and auditing. Secure MPC has many uses in many industries, including machine learning, healthcare, banking, and social sciences. For example, medical institutions and researchers in the health sector often collaborate to examine sensitive patient data for disease surveillance, clinical trials, and demographic studies. Secure MPC allows them to collaborate on calculations using encrypted data without revealing specific patient information or violating privacy laws. Similarly, secure collaborative data analytics can be used in banking to analyses consumer behavior, assess risk, and detect fraud. Although banks and other financial institutions protect the confidentiality of their customers' financial information, they may share information about questionable activities or practices. Researchers can use secure MPC to bring together data sets from multiple sources for social science statistical analysis, surveys, and social network analysis. This allows for a better understanding of social dynamics and trends without compromising authors' right to privacy. Furthermore, protected MPC has important implications for artificial intelligence and machine learning. Organizations often need to train models on large, heterogeneous data sets as the use of artificial intelligence increases. Using secure MPC, multiple parties can train models collaboratively without revealing their information, protecting their privacy, and ensuring the security of sensitive information. Protected MPC has several advantages, but it also has disadvantages and trade-offs. Encryption protocols can cause computation and data transfer overhead, which increases latency and resource requirements. Secure MPC aims to create more efficient cryptographic primitives and improve protocols to achieve a balance between security and efficiency.

2. LITERATURE REVIEW

[1] This study describes secure multipart computation (MPC) approaches to protect privacy in the analysis of shared data. It examines many MPC procedures and their application in various situations and examines their advantages and disadvantages.

[2] The proprietary MPC methods discussed in this study are specifically designed for array genome analysis. It reviews current methods, examines their computational efficiency, and considers possible improvements to increase speed while preserving data privacy.

[3] This work provides a comprehensive analysis of secure MPC methods for collaborative machine learning environments. It examines the problems of illicit financing related to privacy-preserving cooperation and summarizes the remedies suggested by the study.

[4] This review article explores the use of secure MPC for data analytics that protects user privacy. It evaluates the performance and functionality of several MPC protocols for collaborative analysis tasks including clustering, classification, and outlier detection.

[5] This study systematically investigates secure MPC algorithms for group financial analysis. It examines the difficulties and needs of financial data analysis and evaluates the safety, accuracy, and efficiency of current MPC methods.

[6] This study investigates the use of secure MPC in collaborative data mining that protects privacy. It explores several data mining methods and how they can be integrated into secure MPC protocols to enable collaborative analysis without compromising data security.

[7] This study investigates the use of secure MPC for internet of data (IoT) analytics. It covers current MPC protocols designed for IoT applications and addresses the challenges and needs of privacy-preserving IoT computing.

[8] The main topic of this research is secure MPC methods for collaborative recommender systems. It examines current strategies and improvements to protect user privacy and provides accurate and effective recommendations for collaboration.

[9] This review article examines the application of protected MPC in cluster analysis of health data. It explores privacy and security issues in healthcare and summarizes MPC protocols already in use for this purpose.

[10] This paper explores the use of protected MPC in group analysis of social networks. To promote collaborative analysis without revealing sensitive information, it explores various social network analysis tasks and analyses the creation of privacy-preserving MPC protocols.

[11] This methodological review study evaluates secure MPC methods for batch fraud detection. It explores privacy and security criteria unique to fraud detection situations and provides an overview of the accuracy, scalability, and computational power of available MPC methods.

[12] This work focuses on privacy-preserving collaborative MPC techniques for natural language processing (NLP). It reviews current methods, examines the difficulties associated with NLP tasks, and explores the creation of efficient MPC protocols for ensemble NLP analysis.

[13] This study investigates the use of secure MPC in joint traffic analysis to protect passenger privacy. It discusses the difficulties associated with the study of traffic data and evaluates the MPC methods currently used in group analysis of traffic data.

[14] This paper reviews secure MPC protocols for joint energy consumption analysis. It addresses privacy and security issues related to energy data and presents powerful MPC solutions that enable joint analysis while maintaining the confidentiality of sensitive data.

[15] This paper focuses on the implementation of secure MPC in collaborative video surveillance analysis. It addresses

the challenges and privacy requirements of video surveillance scenarios and explores existing MPC protocols designed for collaborative analysis of video data while preserving privacy.

3. PROPOSED METHODOLOGY

Collaborative data analytics involves multiple parties with sensitive data working together to analyses and gain insights without disclosing their personal data. The proposed system aims to solve data privacy issues related to collaborative data analysis using secure multi-party computing techniques. This section presents the importance of privacy-preserving data analytics and the need for secure collaboration frameworks.

System Architecture

The proposed system architecture consists of several components that work together to facilitate secure collaborative data analysis. These parts include: Data preprocessing in this phase, participating parties preprocess their data locally to ensure data compatibility and remove all personally identifiable information (PII). Data anonymization techniques such as k-anonymity or differential privacy can be used to protect privacy.

Secure Multi-Party Computing Protocol (MPC)

The secure MPC protocol, which allows parties to jointly evaluate their data while protecting anonymity, is the brain of the proposed system. The protocol allows parties to calculate desired statistical metrics, such as means, variances, or correlations, without revealing individual data inputs. Depending on the individual needs of the study, other MPC protocols can be used, including secret sharing, homomorphic encryption, and Yao's scrambled chains.

Secure Communication

Secure communication channels between parties must be developed to ensure data confidentiality and integrity during computing. This can be achieved using digital signatures of authentication and encryption technologies such as Secure Socket Layer (SSL) and Transport Layer Security (TLS).

Combination of results

After a secure calculation phase, the calculated results are summed without revealing the contributions of the individual parties. Privacy-preserving aggregation methods such as safe sum or safe mean can be used to derive final analysis results. Security measures the proposed system includes several security measures to protect data protection and data integrity during joint analysis. These measures include: Maintaining privacy The secure MPC protocol ensures that no party can learn anything other than what is revealed in the final analysis results. The system ensures that the privacy of participants is preserved even from malicious or conspiratorial parties. Secure Computing Choosing appropriate MPC protocols and encryption techniques ensures secure computing of analytical tasks. Technologies such as zero-data certificates, secure activity evaluation and discreet transmission help prevent data leakage and unauthorized access.

Access Control

The system implements strict access control mechanisms to prevent unauthorized use. Parties must authenticate themselves before participating in collaborative analysis. Access rights and permissions are assigned based on predefined policies.

Finance

Financial institutions can collectively analyse transaction data to identify patterns, detect fraud or assess risk while maintaining the confidentiality of their customers' financial information.

Research Academic institutions and researchers can conduct joint data analyses without sharing raw data, enabling collaboration between organizations and industries.

3. DESIGN AND IMPLEMENTATION

Algorithm:

1. step: install step: a. Protocol Initialization: Each party generates a pair of public and private keys for encryption and decryption. b) Secure communication channels: The parties establish secure communication channels to exchange encrypted messages between themselves.

2. step: input step: a. Each party privately stores a portion of the data for analysis. b) Each party encrypts its data with its own public key.

3. step: calculation step: a. Each party performs local calculations with its encrypted data without revealing the plaintext. b) The parties safely calculate jointly agreed operations such as addition, multiplication, or more complex functions. c. Secure protocols such as Yao's Garbled Circuits or Secret Sharing can be used to perform calculations while maintaining privacy.

4. step: result Step: a. The parties decrypt the calculated results using their private keys. b) The extracted results are tightly combined to produce the result. Table Output: If we have three parties (Party A, Party B and Party C) collaborating on a secure data analysis task, let us look at an example where they want to calculate the average performance of their combined data sets.

Table 4.1 - Party A: Dataset A (in Rupees)

Name	Income
Dhaval	50000
Bharat	60000
Jay	45000
Jenish	55000

Table 4.2 - Party B: Dataset B (in Rupees)

Name	Income
Het	40000
Shrey	65000
Nirav	55000
Neel	48000

Table 4.3- Party C: Dataset C (in Rupees)

Name	Income
Yash	55000
Harsh	52000
Aditya	50000
Raviraj	60000

Table 4.4- Result:

Party	Income
Average	52000

In this table, each party stores an encrypted subset of the data (data sets A, B, and C) and uses it for local computations. The halves then combine the encrypted observations to safely calculate the average income. The calculated average income is distributed to obtain a result of 52000 Rupees.

The proposed scheme creates a comprehensive basis for secure multiparty computing in shared data analysis. The solution allows multiple parties to collaboratively analyses their data without compromising privacy, using privacy-preserving mechanisms and strong security measures. The system is a useful tool for many industries that require unified data analysis while maintaining security and privacy through flexibility and possible applications.

4. CONCLUSION

As an effective framework, secure multiparty computing enables the analysis of group data while protecting the secrecy and privacy of individual participants. Secure MPC uses cryptographic methods to allow many people to work together to compute their personal information without revealing sensitive information. It has applications in many fields, including social sciences, finance, healthcare, and machine learning. Despite the obstacles, ongoing research and development efforts aim to improve the efficiency and usability of Secure MPC, making it a key data analysis tool that protects privacy in the digital age.

References

[1] Smith, J., & Johnson, A. (2019). Secure Multi-Party Computation for Privacy Preserving Collaborative Data Analysis. Journal of Privacy and Security, 15(2), 123145.

[2] Brown, M., & Davis, R. (2020). Efficient Secure Multi-Party Computation for Collaborative Genomic Analysis. Journal of Bioinformatics and Computational Biology, 18(3), 235-257.

[3] Lee, H., & Wang, S. (2021). Secure Multi-Party Computation for Collaborative Machine Learning: Challenges and Solutions. IEEE Transactions on Knowledge and Data Engineering, 33(8), 1234-1256.

[4] Chen, L., et al. (2018). Privacy-Preserving Data Analytics using Secure Multi-Party Computation: A Survey. ACM Computing Surveys, 51(3), 1-35.

[5] Liu, X., et al. (2022). Secure Multi-Party Computation for Collaborative Financial Analysis: A Systematic Review. Journal of Financial Data Science, 2(1), 45-68.

[6] Wang, Y., & Li, Q. (2019). Privacy-Preserving Collaborative Data Mining using Secure Multi-Party Computation. Data Mining and Knowledge Discovery, 33(4), 789-813.

[7] Zhang, W., & Zhang, L. (2020). Secure Multi-Party Computation for Collaborative Internet of Things Data Analysis. IEEE Internet of Things Journal, 7(5), 3789-3807.

[8] Li, X., et al. (2021). Efficient Secure Multi-Party Computation for Collaborative Recommender Systems. ACM Transactions on Information Systems, 39(4), 1-28.

[9] Wang, L., et al. (2019). Secure Multi-Party Computation for Collaborative Healthcare Data Analysis: A Review. Journal of Biomedical Informatics, 92, 103148.

Yang, C., et al. (2020). Privacy-Preserving Collaborative Social Network Analysis using Secure Multi-Party Computation. Social Network Analysis and Mining, 10(1), 122.