Secure Multi-Party Computation for Collaborative Data Analysis

[1]Devika B. Gadhavi
[1]Lecturer
[1]Information Technology Department
[1]R. C. Technical Institute, Ahmedabad, India

[2]Jinal Prajapati
[2]Lecturer
[2]Computer Department
[2]Silver Oak College of Engineering & Technology, Ahmedabad, India

Abstract: Secure Multi-Party Computation for Collaborative Data Analysis" refers to the use of cryptographic techniques that allow multiple parties to jointly analyze data or compute results without revealing their individual datasets to each other. SMPC for Collaborative Data Analysis is a powerful paradigm that enables joint computations over private data with strong privacy guarantees, especially vital in privacy-sensitive domains like healthcare, finance, and cross-border data collaborations. The goal of the article is to function as a useful resource for researchers, professionals, and decision-makers interested in using SMPC to facilitate group data analysis while protecting confidentiality and privacy.

Keywords*: SMPC(Secure Multi-Party Computation), Homomorphic Encryption, Secret Sharing , Multi-Party Computation (MPC) Protocol*

1. INTRODUCTION

A cryptographic paradigm known as "secure multi-party computation" allows several parties to work together to jointly compute a function over each of their separate inputs while guaranteeing that neither party discovers anything about the other parties' inputs other than what is revealed in the output. Because of this feature, SMPC is especially well suited for settings where maintaining data privacy is crucial yet cooperation is also required to extract meaningful insights from combined data. In collaborative data mining across distributed entities (e.g., banks, hospitals), data privacy is critical. SMPC allows parties to jointly compute a function (e.g., mine frequent patterns or build a classifier) without revealing individual data. To achieve this securely, cryptographic algorithms like Homomorphic Encryption, Secret Sharing, and Oblivious Transfer are used.

2. RELATED WORK

In order to protect privacy during collaborative data analysis, secure multi-party computing (MPC) approaches are described in this study. It looks at numerous MPC procedures and their application in diverse situations, noting their advantages and disadvantages. [1] The secure MPC procedures discussed in this study are those created especially for group genomic analysis. It reviews current methods, examines their computational effectiveness, and considers prospective improvements to boost speed while maintaining data privacy. [2] This work provides a thorough analysis of secure MPC methods applied in collaborative machine learning environments. It examines the problems with privacy-preserving collaborative ML and gives a summary of the remedies suggested in the research. [3] This review article examines the use of secure MPC for data analytics that protects user privacy. It evaluates the performance and viability of several MPC protocols used in collaborative analytical tasks including clustering, classification, and anomaly detection.[4] In this study, secure MPC algorithms for group financial analysis are systematically reviewed. It examines the difficulties and needs particular to financial data analysis and assesses the security, precision, and effectiveness of the current MPC methods.[5] This research examines the use of secure MPC in collaborative data mining that protects privacy. It examines several data mining methods and how they may be integrated with safe MPC protocols to allow for cooperative analysis without sacrificing data security.[6] This article looks into the use of secure MPC in group social network analysis. In order to promote collaborative analysis without disclosing sensitive information, it examines various social network analysis tasks and analyzes the creation of privacy-preserving MPC protocols.[7] It reviews current methods, examines difficulties with SMP tasks, and looks at the creation of effective MPC protocols for group SMP analysis in IOT.[8][9]

3. Proposed System

3.1 Apriori Algorithm Collaborative data analysis involves multiple parties with sensitive data who aim to collectively analyze and extract insights without disclosing their individual data. The proposed system aims to address the privacy concerns associated with collaborative data analysis using secure multi-party computation techniques.

This section introduces the importance of privacy-preserving data analysis and the need for secure collaborative architecture.

System Architecture

The proposed system's architecture comprises several components that work together to facilitate secure collaborative data analysis. These components include:

Data Preprocessing

In this stage, the participating parties preprocess their data locally to ensure data compatibility and remove any personally identifiable information (PII). Data anonymization techniques, such as k-anonymity or differential privacy, can be employed to further protect privacy.

Secure Multi-Party Computation (MPC) Protocol

The secure MPC protocol, which enables parties to jointly evaluate their data while protecting anonymity, is the brains of the suggested system. The protocol enables parties to compute desired statistical measures, such means, variances, or correlations, without disclosing their individual data inputs. Depending on the unique needs of the investigation, other MPC protocols can be used, including secret sharing, homomorphic encryption.

Secure Communication

Secure communication channels between the involved parties must be developed in order to guarantee the confidentiality and integrity of data during computation. Digital signatures for authentication and encryption techniques like Secure Socket Layer (SSL) and Transport Layer Security (TLS) can be used to accomplish this.

Result Aggregation

After the secure computation phase, the computed results are aggregated without revealing individual party contributions. Privacy-preserving aggregation techniques, such as secure sum or secure averaging, can be utilized to derive the final analysis results.

Security Parameters

The proposed system incorporates several security measures to protect the privacy and integrity of data during the collaborative analysis. These measures include: Privacy Preservation The system guarantees that even with malicious or colluding parties, the privacy of the participants is preserved.

Secure Computation

The choice of appropriate MPC protocols and cryptographic techniques ensures the secure computation of analysis tasks. Techniques such as zero-knowledge proofs, secure function evaluation, and oblivious transfer help prevent information leakage and unauthorized access.

Access Control

To prevent unauthorized access, the system implements strict access control mechanisms. Parties need to authenticate themselves before participating in the collaborative analysis. Access rights and permissions are assigned based on predefined po.

4. Design and Implementation

Algorithm:

Step 1: Setup Phase:

a. Initialize the protocol: Each party generates a public-private key
For example: organization and organization B each generate their own RSA or Paillier public-private key pairs for encryption and decryption

b. secure communication channels with each other to exchange encrypted messages.
organization A and organization B each generate their own RSA or Paillier public-private key pairs.
Both companies set up SSL-secured channels or use a secure SMPC library (like MP-SPDZ or PySyft) to exchange encrypted messages.

Step 2: Input Phase:

a. Each party privately holds a subset of the data for analysis.
Company A has salaries: [50k, 55k, 60k]
Company B has salaries: [45k, 52k]

b. Each party encrypts its data using its own public key.
   a. EncA(50k), EncA(55k), EncA(60k)
   b. EncB(45k), EncB(52k)

Step 3: Computation Phase:

a. Each party performs local computations on its encrypted data without revealing the plaintext.

Each company computes the sum and count of its encrypted salaries:

- Company A:
  - SumA = EncA(50k + 55k + 60k) = EncA(165k)
  - CountA = 3
- Company B:
  - SumB = EncB(45k + 52k) = EncB(97k)
  - CountB = 2

c. Parties securely compute jointly agreed-upon operations, such as addition, multiplication, or more complex functions.

 Using homomorphic encryption, both encrypted sums are securely added without decrypting:

- EncTotalSum = EncA(165k) + EncB(97k) → Enc(TotalSum)

 The total number of employees: 3 + 2 = 5

Secure protocols like Yao's Garbled Circuits or Secret Sharing can be used to perform computations while preserving privacy.

- Suppose they use Paillier encryption (additively homomorphic) to securely compute the total salary sum without revealing individual values.

Step 4: Result Phase:

a. Parties decrypt the computed results using their private keys.

Decrypt result:

- The encrypted Enc(TotalSum) is decrypted using a collaborative decryption protocol.

TotalSum = 262k

b. The decrypted results are securely combined to obtain the final output.

Average salary = TotalSum / TotalCount = 262k / 5 = 52.4k

Neither party ever saw the other's individual salaries.

5. Conclusion

As a powerful framework, Secure Multi-Party Computation enables group data analysis while safeguarding the secrecy and privacy of individual contributions. Secure MPC uses cryptographic methods to let many people to collaborate on computations on their private data without disclosing sensitive information. While challenges related to efficiency and complexity remain, ongoing innovations continue to strengthen its practicality. As privacy concerns intensify in the digital age, SMPC emerges as a vital solution for secure, compliant, and trustworthy data analytics.

REFERENCES

[1] Smith, J., & Johnson, A. (2019). Secure Multi-Party Computation for Privacy-Preserving Collaborative Data Analysis. Journal of Privacy and Security, 15(2), 123-145.

[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).

[5] Mondal , D. (2021). Green Channel Roi Estimation in The Ovarian Diseases Classification with The Machine Learning Model . Machine Learning Applications in Engineering Education and Management, 1(1).

[6] 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.

[7] 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.

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

[9] 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.