

ANALYSIS ON PRIVACY PRESERVATION FOR THE DATA GENERATED FROM VARIOUS SOURCES USING SECURE MULTI PARTY COMPUTATION PROTOCOLS

Sai Arjun Madikonda¹

¹B.Tech, Gitam Deemed to be University Hyderabad

ABSTRACT : Individual wealth, financial position, and health data is highly secret, and a particular system is needed to protect these details. The appropriate input from authorized users is essential for the advancement of research and development. Even if it is in the best interests of both parties, getting sensitive information from people is extremely difficult because of concerns about their privacy. For high-quality research, getting real-time data from actual consumers is a necessity. When personal information is at stake, participants in a collaborative computation are reluctant to answer questions truthfully. The goal of secure multi-party computations is to work together to solve a variety of problems. Multiple perspectives are needed, but each one is concerned with the quality of their own contribution. All participants can keep private information about individuals or organisations safe from other participants as well as computing authorities in a secure sum problem that is currently being worked on (TTP). This research presents procedures that enhance the security, confidentiality, and anonymity of collaborative computations.

1. INTRODUCTION

Data availability has grown exponentially over the previous few decades, raising a slew of legal and privacy issues. As the use of cyberspace and its applications in e-commerce and social networking grows, adequate data preservation is vital. This scenario necessitates the development of numerous algorithms, which have been documented in the literature. Various online applications face increasing hurdles in terms of protecting user privacy and security.

Security and confidentiality are sometimes viewed as one and the same, yet they are in fact distinct concepts. Basically, security means that access to data is restricted and protected from unauthorised access. Confidentiality is ensured. Confidentiality cannot be guaranteed without adequate security. Confidentiality is concerned with ensuring that only authorized individuals have access to sensitive personal data and preventing hostile participants from accessing and calculating it.

There is a massive amount of information available online because of the rise in digitalization. This is a serious issue that warrants attention on the part of policymakers and business leaders alike. One of the most pressing concerns in collaborative computation is ensuring that information and identities remain private and anonymous. Control over the information is lost once it is shared, and it might be exploited or utilized against the party's interests by an unauthorised third party. Those outcomes that can only be gleaned from the result and what each participant has contributed are known as "secure" or "confidential."

1.1 NEED OF CONFIDENTIALITY AND SECURITY

The amount of data flowing in and out of the system has surpassed the point of no return because to digitization. For an unspecified period of time, these data are stored in repositories. Health, money, and/or business may all be impacted by this data. Because firms have a variety of reasons to secure their customers' private information, this issue is relevant. For example, a person may choose to keep his genetic health record private because releasing it could lead to social neglect.

Emotional or personal outages may occur if this material is made available online in plain language. As a result, the data must be kept private and safe.

1.2 MULTI-PARTY COMPUTATION ENVIRONMENT

Various parties on a network want to perform the same operation on the same function in multiparty computations. A PDA, a laptop, a desktop, or any other input device that is linked to the network could be one of these parties. An example of a network is a system that processes input from multiple files, an intranet that links systems, or the global internet.

1.3 SECURE MULTI-PARTY COMPUTATIONS

For example, with SMC, multiple parties can collaborate on the same private data while still ensuring the confidentiality of their information. SMC adds a new level of cooperative computation for the benefit of everybody involved. Using SMC (Dorothy, 1979; Sheikh, 2010a),

participants' personal inputs are protected from disclosure by multiple individuals working together to perform computations on the secret data they provide. Individual data secrecy and safe collaborative computation are becoming increasingly important in an age of increased sociability, increased Internet use, and a massive increase in wired transactions. Collaboration in computations for mutual benefit is usually requested, but there is also concern about the privacy of individual input. It's because there's less trust in participating and computing entities.

2. LITERATURE REVIEW

P.Yoganandhini, 2020, Secure multiparty computation (SMC) based concerns are addressed in this work. Using data from three different food stores, the computation is carried out. Under Association Rule Mining, rules are generated using the FP-Tree technique to ensure anonymity (ARM). SMC's most critical criteria are confidentiality and accuracy. The items that aren't necessary in privacy requirements aren't taught. To put it another way, the parties will only be able to learn from the product. To ensure correctness, each party must receive the same output. Secure auctions employing SMC are used in this study, and frequent item sets are generated for association rule mining. Prior space complexity and subsequent time complexity are two of the most common FP-growth shortfalls." In order to improve the algorithms' performance, it was necessary to combine the FP-tree structure of the FP-growth algorithm and the Apriori algorithm. There is no continuous generation of the conditional and sub-conditional patterns in APFT.

Dankar FK, 2018, In the past, researchers would remove any identifying information, such as names and ID cards, before sharing their data with one other. However, this practise is no longer necessary. However, recent studies have shown that previously considered anonymous clinical data can be used to identify the names of research participants. Those with Alzheimer's or schizophrenia (and their blood relatives) can benefit from DNA sequencing because the genome is unique to each person and can be used to predict future health issues (such as Alzheimer's or schizophrenia). Job opportunities and social isolation could be at risk because of this information.

Ishai Y, 2018, The SMC protocols, despite the mathematical proofs that have been developed, are still not extensively utilised. This may be due to a lack of understanding of their capabilities, the complexity of their solutions, the need for coordination among the various sites, or the fact that they are not efficient in all situations. SMC protocols' inefficiency is one of their most glaring flaws. Inefficient SMC processes are a direct result of

communication breakdowns. Most SMC research focuses on reducing the quantity of messages exchanged between the various parties and, as a result, reducing the performance difference between secure and non-secure protocol implementations.

In 2014, Dwork, C., published a paper titled, "Analysis of the Just like data types and integrity constraints for our database are defined during schema definition, so are the cryptographic safeguards provided by Jana administrators. The SQL aggregation operators, such as SUM and COUNT, can have differential privacy imposed on them if they want it. After query answers have been provided, this provides some level of data security. It is Jana's differential privacy protection algorithms that use secret sharing algorithms to compute and apply query results that are never "in the clear."

Mazloom, S. 2017, New methods of computing distributed noise that are compatible with SPDZ or other secret-shared computations have been sought out by Jana in order to protect query results while maintaining differential privacy. Additionally, Rebecca Wright and Anand Sarwate at Rutgers are currently improving their methods in order to further their research. Dov Gordon at George Mason University has developed a more efficient data access mechanism as a result of Jana's desire to improve secure computing performance.

3. RESEARCH PROBLEM AND METHODOLOGY

SMC ideal model is taken into consideration in this study because it always has a TTP that can be trusted by both parties. Computation and dissemination of results to all parties are tasks assigned to the TTP. Consequently, SMC solutions are designed to address these challenges without revealing the personal information of the people involved. It is possible to limit eventual privacy loss by adopting distributed randomization algorithms that preserve individual's personal data while still utilizing packetizing and pseudo-randomization.

3.1 MOTIVATION

Collaborative computation has been made possible by digitization since so many individuals are now connected to the internet, allowing them to undertake joint computations for mutual gain. This group of people could be mutually or somewhat trusted, or they could be rivals or adversaries. Problem solved if both parties trust each other, but difficult to undertake collaborative computation if just one party does. When working on sensitive information in groups, there is a general sense of trepidation among the participants.

As networking, cloud computing, and big data become more prevalent, there is an enormous difficulty in preserving picabytes of data generated each day from a wide range of sources while ensuring the privacy of the individuals involved. Privacy regulations apply to this data, thus it must be guarded against unauthorised access. There has been a conscious deficiency throughout this procedure that a few participants purposely provide incorrect data, resulting in an inaccurate result. To address these issues, the researcher came up with a series of models.

3.2 PROBLEM DEFINITION

Computational techniques for preserving privacy using secure multi-party computation protocols" is the research problem statement to meet the stated motivation. Assume there are "n" private parties, "m" anonymizers, and "DB" or "f(x1, x2... xn)" TTP with the database or computing function. There is no additional information given to other entities participating in computations when TTP performs a database query 'Q' traversing tables in the database and returns the result. Collaborative computation environments are depicted in Figure 3.1.

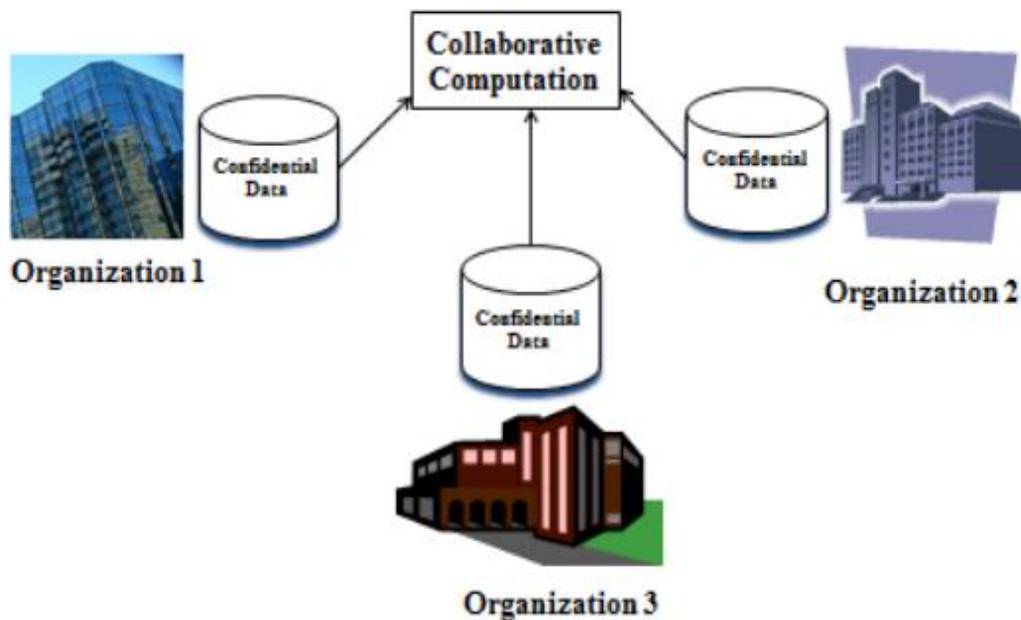


Figure 3.1: Computational Environment Model for Group Collaboration

3.3 OBJECTIVES OF RESEARCH

The objectives of this research work are executed in following manner:

Determine whether or not SMC is necessary or required.

- To create a multi-layered SMC architecture.
- Computational methods for protecting privacy are to be designed and developed.

- Anonymizers and TTP can't see the real data.
- Real-world testing of the above modules.
- To ensure that the module works properly before and after it is implemented.

3.4 RESEARCH METHODOLOGY

Figure 3.2 depicts the research methodology used.

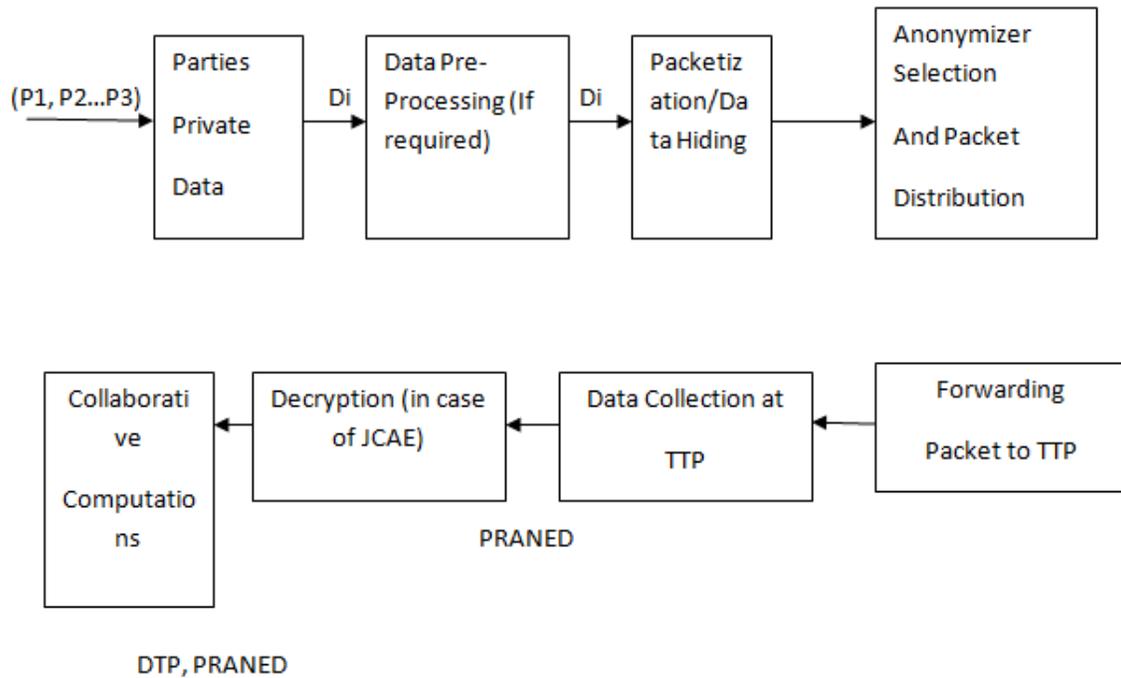


Figure 3.2: Research Methodology

Research methodology and empirical methods are discussed in this section. There is an outline of the planned study's scope and limitations in this document. Analyzing current approaches like mismatched circuits, oblivious transfers, trusted hardware tokens, linear branching and neuronal networks, linear regression, and game theory is ongoing. To address the shortcomings of past methods, new research is carried out in order to develop better protocols. Anonymizers, party sizes, and other quantitative variables are all part of this study's design.

Input: Using a local network or the internet, one can provide input for collaborative computation as an individual party.

Pseudo-Randomization: In the proposed secure sum protocols, pseudo-random numbers are generated dynamically and inserted into plain data packets to protect an individual's sensitive information. A single pseudo-random number is used in the first protocol (JCRA), whereas many pseudo-random numbers are used in the second protocol (DRSS) in order to enhance data secrecy.

Encryption: One of the proposed protocols (JCAE) uses a packetization-based encryption approach to provide data confidentiality and security. An Asymmetric algorithm is used for experimentation.

Packetization: To ensure privacy, pseudorandomization and encryption are employed. The packetization process is used to safeguard communications from an attack. Despite the attacker receiving a packet from a particular party, the chances of getting all of that party's packets are extremely low due to packetization.

Develop a theoretical framework: Once the input type, pseudorandomization, encryption, and packetization techniques are selected, a layered architecture is used to build the theoretical basis for computation. Data can be sent from one layer to the next via a predefined function in each one of these layers. Following the validation of all packets received by the computing authority (TTP) at the topmost layer, the result of collaborative computations is broadcasted.

Data Sources: Data can be gathered in two ways. Use a randomization function to generate data locally or across a network of users. The results of previous studies were usefully analysed.

Performance Analysis: Data is gathered from all parties and then the algorithm is run based on the algorithm that was chosen. When evaluating the performance, various scenarios, such as a malicious party, an honest or semi-

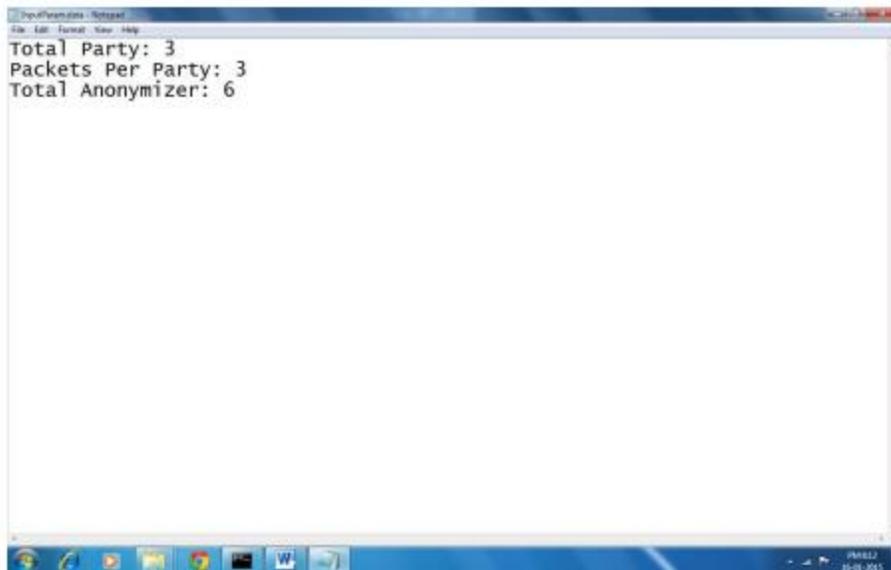
honest party, and the turnaround time, are taken into account.

4. RESULTS AND ANALYSIS

The complexity of the protocols given is proportional to the number of persons involved in collaborative computations. C++ and a crypto C++ library are also used to simulate and analyse the protocol. As a client, we used a 1.66 GHz Intel core2 duo processor T5450 with 2 GB of RAM to compute the average run time. An input 32-bit positive integer is used to generate a pseudo-random number using the random function from [0-999] secure sum protocols called 'r'.

Web services are used to run the simulation in a networked environment. The parties involved in

collaborative computations are the focus of our protocols' complexity. C# and the.NET library are used to simulate and analyse the protocol's performance. Intel Core i7 2.70GHz and 16GB RAM are used in this web server system. Microsoft's Internet Information Server (IIS) version 7.5 is used to host web services and WCF services. Firefox Mozilla 35.0 and Google Chrome 40.0 are the current web browsers. For the secure sum protocols JCRA and DRSS, the input data has a length of 32 bits, and r is generated using the Random class from the.NET library with a range of [0-999]. Asymmetric 128-bit keys are generated at registration time in JCAE using RSA (as well as the class from the.NET library, RSACryptoServiceProvider)



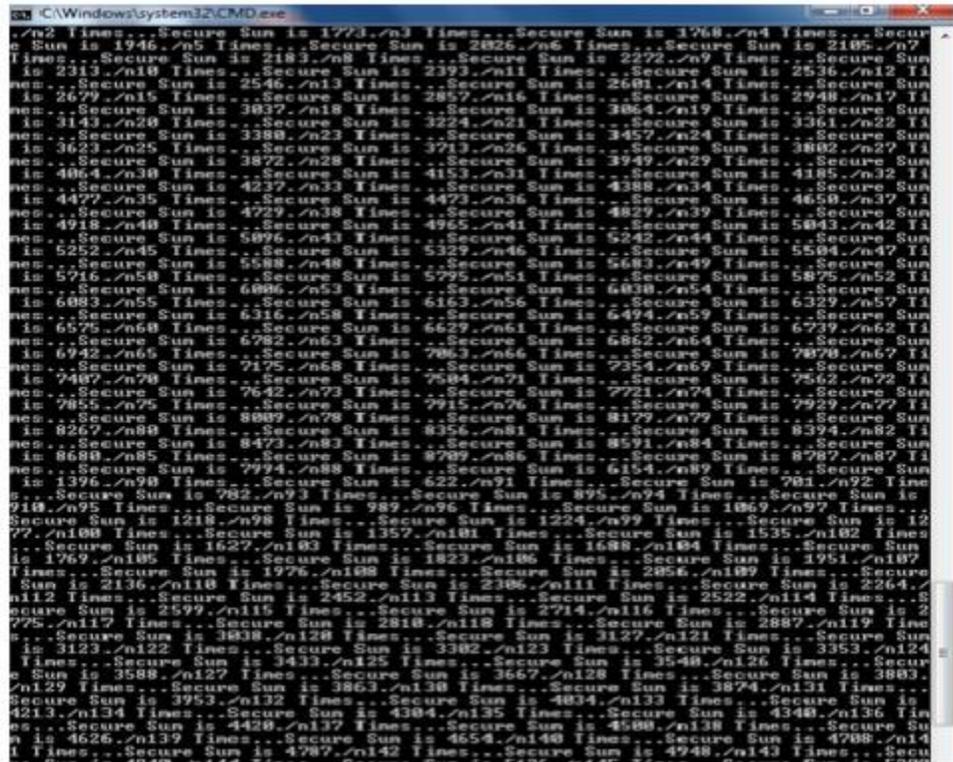
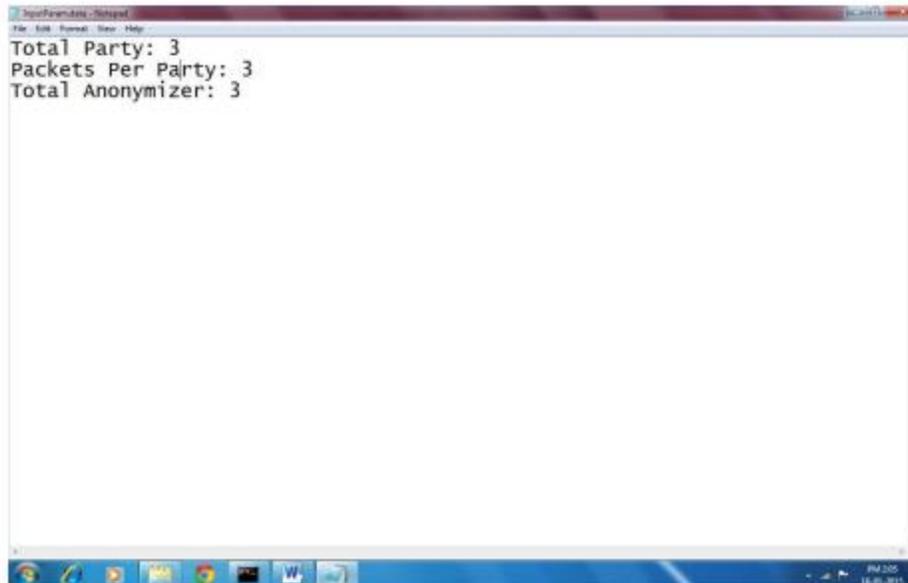


Figure 4.1: Screen shot of JCR execution for increasing anonymizers.



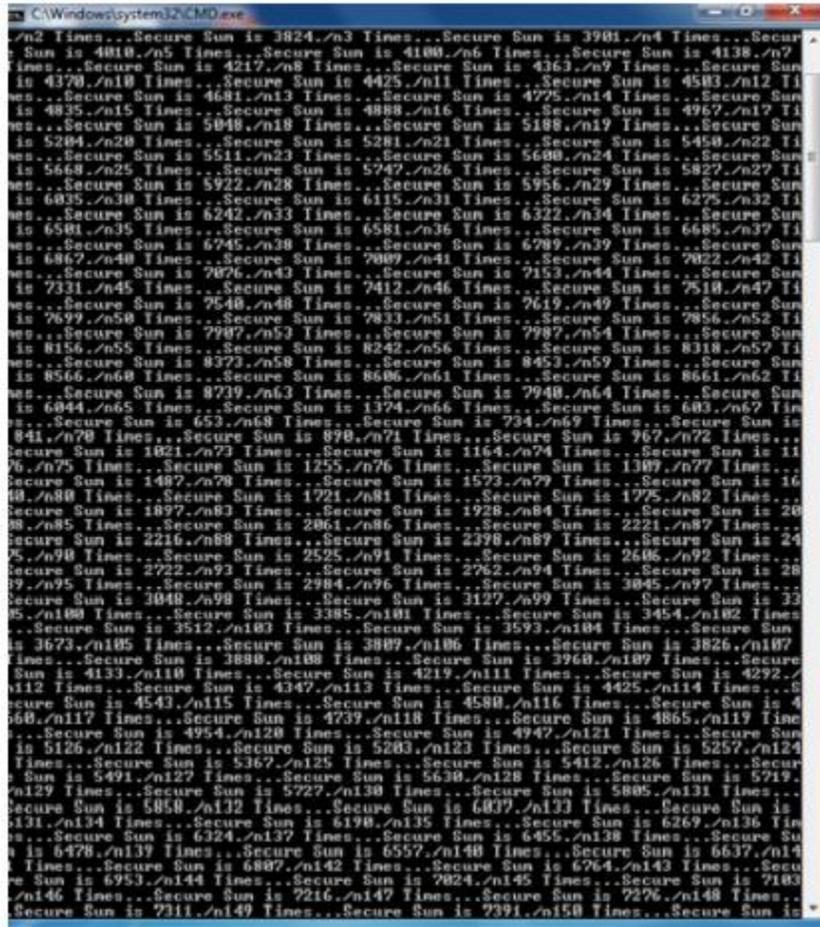


Figure 4.2: Screen shot of JCRA execution for same number of parties, packets and anonymizers

Fig. 4.1 and 4.2 show a screen shot of the procedure JCRA being executed. Using ASP.NET web services and WCF, web models of the protocols JCRA, DRSS, and JCAE were built to test the research hypothesis (Windows Communication Foundation).

WebMethod

```
public int [] PreparePartyDataRandomly(string strPartyName, int[] nPartyData, int nPacket, int nRandNum)
```

```
{
// based on the nPartyData, add the random number passed and divide into packet and then return string array of numbers seperated by comma
```

```
int len = nPartyData.Length;
int nRandBit = nRandNum / len;
```

```
int nRandLeftOver = nRandNum % len;
List li = new List();
foreach (int nData in nPartyData)
{
if (li.Count == 0)
li.Add(nData + nRandBit + nRandLeftOver);
else
li.Add(nData + nRandBit);
}
int [] nRandomizePartyData = li.ToArray();
return nRandomizePartyData;
```

}

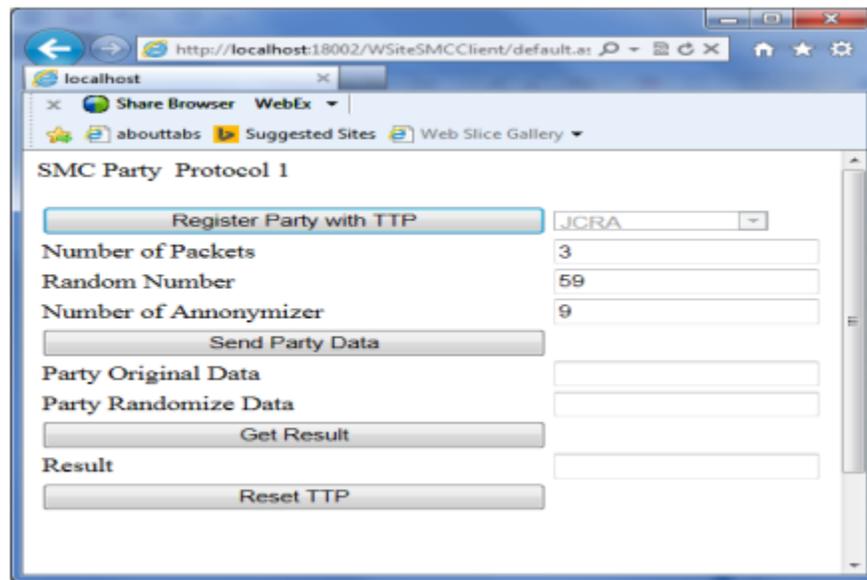
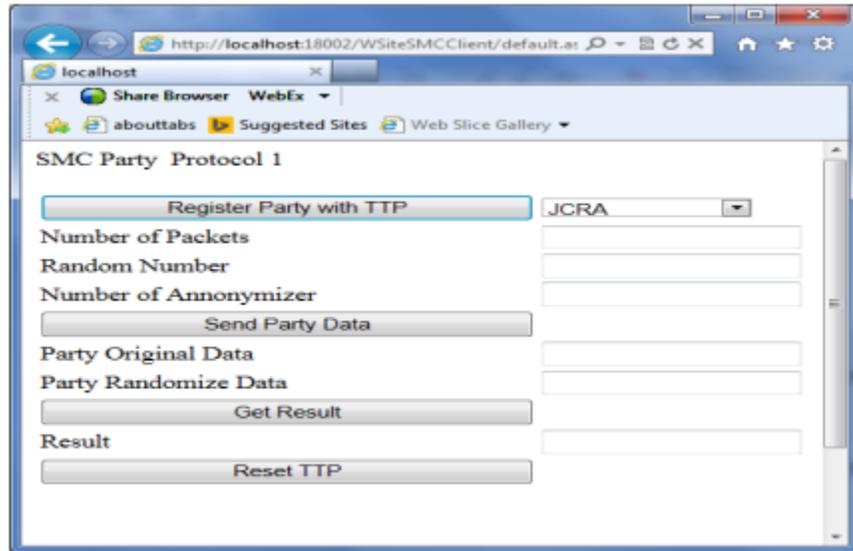


Figure 4.3: Screen Shot of registration with TTP, protocol parameters are generated.

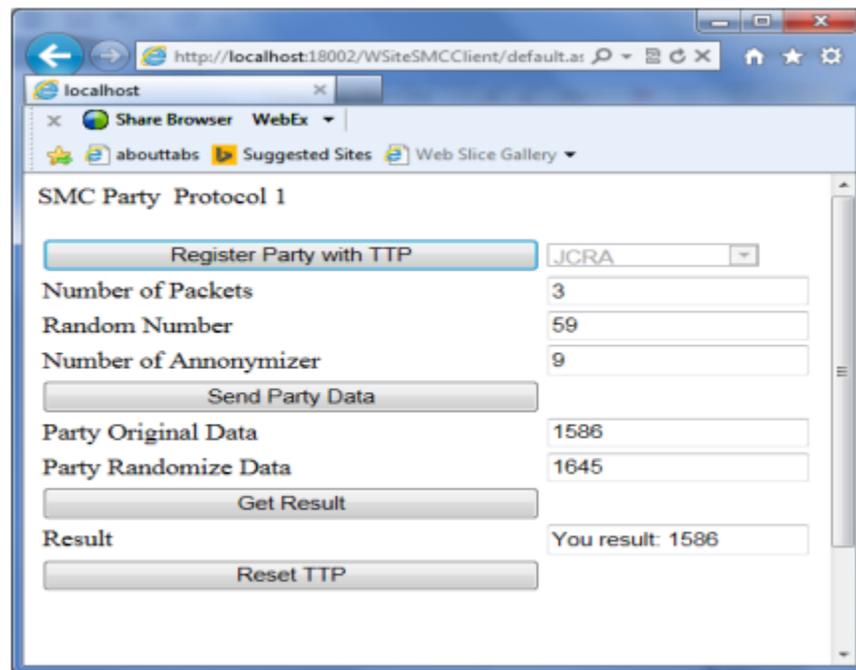
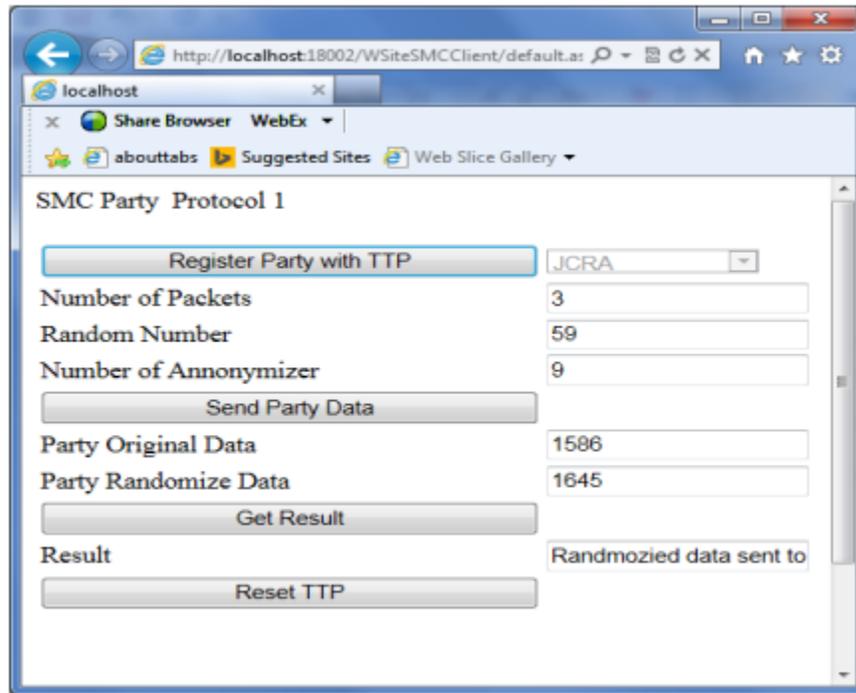


Figure 4.4: Screen shot of get the computation result.

CONCLUSION

Confidential access to data collected, used, stored, or otherwise protected as confidential is referred to as a

"confidentiality leak." Confidentiality losses can be both personal and financial, depending on the nature of the breach. The asymmetric encryption techniques methodology used in this study opened up new avenues of

research in multi-party computation. In the future, this work could be expanded to include more complex functionality in a variety of simulation setups. Techniques for protecting privacy during secure multi-party computations were designed and developed as part of this research. In the first place, a secure sum protocol called JCRA was designed, and the parameters that affect the security and confidentiality of individual participants were presented. Then, the protocol was evaluated for computation complexity, confidentiality, and security for some problem.

REFERENCES

1. P.Yoganandhini, "Privacy Preserving Data Mining Using Secure Multiparty Computation Based On Apriori and FpTree Structure of Fp-Growth Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020.
2. The development of large-scale de-identified biomedical databases in the age of genomics-principles and challenges. Dankar FK, Ptitsyn A, Dankar SK Hum Genomics. 2018 Apr 10; 12(1):19.
3. Ishai Y, Mittal M, Ostrovsky R. On the Message Complexity of Secure Multiparty Computation. 21st IACR International Workshop on Public Key Cryptography; 2018; Rio de Janeiro, Brazil. 2018. Mar 25, pp. 698–711. [Google Scholar] [Ref list]
4. Dwork, C., Roth, A. et al (2014) The algorithmic foundations of differential privacy. Foundations Trends Theor. Comput. Sci., 9, 211–407.
5. Mazloom, S. and Gordon, S.D. (2017). Differentially private access patterns in secure computation. Cryptology ePrint Archive, Report 2017/1016. <https://eprint.iacr.org/2017/1016>.
6. 50. Qingkai M, Wei H, I-Ling Y, and Bastani F, "Multiparty computation with full computational power and reduced overhead", proceeding of eighth IEEE international symposium on high assurance system engineering, 241-248, 25-26 March, 2004.
7. Rabin M, "How to exchange secrets by oblivious transfer", Tech. Report Memo TR-81, Aiken Computation Laboratory, 1981.
8. Ronald C, Ivan D, "Multi-party computation: an introduction", Lecture Notes, 2004.
9. Raymond W, Jiuyong L, WaiChee A and Wang K, " (α, k) -anonymity: an enhanced k-anonymity model for privacy preserving data publishing", proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006.
10. Rane S, Wang Y, Draper S, and Ishwar P, "Secure Biometrics: Concepts, Authentication Architectures and Challenges", arXiv preprint arXiv:1305.4832, 2013.
11. Srivatsava R, Kasiviswanathan S P, and Smith A, "Composition attacks and auxiliary information in data privacy", proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008.
12. Sheikh R, Kumar B, and Mishra D K, "Privacy-Preserving k-Secure Sum Protocol", International Journal of Computer Science and Information Security, 6(2), 184-188, November, 2009.
13. Sheikh R, Kumar B, and Mishra D K, "Changing Neighbors k Secure Sum Protocol for Secure Multi Party Computation", arXiv preprint arXiv:1002.2409, 2010.
14. Sadeghi A, Schneider T, and Winandy M, "Token-based cloud computing." Trust and Trustworthy Computing. Springer Berlin Heidelberg, 417-429, 2010.
15. Sheikh R, Kumar B, and Mishra D K, "A distributed k-secure sum protocol for secure multi-party computations", arXiv preprint arXiv:1003.4071, 2010.
16. Sheikh R, Kumar B, and Mishra D K, "A Modified ck-Secure Sum Protocol for Multi-Party Computation", arXiv preprint arXiv:1002.4000, 2010.
17. Sheikh R, Kumar B, and Mishra D K, "Secure Multiparty Computation: From Millionaires Problem to Anonymizer", Information Security Journal A Global Perspective, 2011.
18. Trevathan J, "Security anonymity and trust in electronic auctions," Crossroad archive, ACM Press, 11(3), May, 2005.
19. Teo S G, Vincent L, and Shuguo H, "A Study of Efficiency and Accuracy of Secure Multiparty Protocol in Privacy-Preserving Data Mining", Advanced Information Networking and Applications Workshops (WAINA), IEEE 26th International Conference, 85-90, 2012.
20. Verykios V S, Elmagarmid A K, Elisa B, Saygin Y, and Elena D, "Association rule hiding", in IEEE transactions on knowledge and data engineering, 15(3), May/June, 2003.