CHAPTER 3

RESEARCH PROBLEM AND METHODOLOGY

This chapter discusses motivation behind the research, problem definition, research methodology used, system domain, and the outcome of the research.

The security concerns can be achieved using real model as well as ideal model [*See Section 2.2.1*] of SMC. In this research Ideal model of SMC is taken into account, because in ideal model there is always a TTP, which can be trusted by all the parties. The TTP is responsible for collaborative computation and broadcast of results to all the parties.

Thus, SMC solutions attempt to solve these issues without revealing parties personal data. For example, distributed randomization algorithm work out the issue by preserving party's' personal data, by using packetization and pseudo-randomization so that the subsequent privacy loss can be minimized.

3.1 MOTIVATION

Digitization triggered the opportunity for collaborative computation as huge number of people are using internet, they can cooperate with each-other to perform joint computation for mutual benefit. This participants could be mutually trusted, partially trusted or competitors. If they are mutually trusted problem is solved but if they are partially trusted or not trusted then it is difficult to perform collaborative computation. All the participants have underlying fear in collaborative computation on sensitive information.

In the new age of socialization, cloud computing and big data there is a big challenge for preserving pica bytes of data generated daily from various sources and use it for medical, financial and government related

research while maintaining the confidentiality of people involved. This data is subject to privacy laws and need to be protected. All through this procedure a thoughtful inadequacy has been brought into picture that few of the participants intentionally furnish their data inaccurately, affecting the accuracy of the result. These circumstances motivated the researcher to propose models which can handle this type of problems.

Due to rise in cloud computing people are moving from in-premises infrastructure to cloud storage, which attracted a lot of researchers from industries and academics. Outsourcing data to distributed cloud environment avail data and resources easily. This relieves participating parties from the load of data management and gives this task to service providers with committed resources and high-end systems techniques. There is a requirement of mechanism in which data is not directly accessed, for outsourcing these data to a third party data need to be deployed and data should be shared in encrypted form so that confidentiality can be improved.

The basic motivation of SMC studies is to design protocols that permit maximum information utilization without compromising individual confidentiality.

Secure multi-party computation is a very broad term with great potential for real life application, but unfortunately very few practical applications of SMC have been implemented so far. For example: e-voting, in this each party holds a 0/1(D_i) vote, and the function to be computed is $(x) = \sum_{i=1}^{n} D_i$.

3.2 PROBLEM DEFINITION

To satisfy above motivation, following is the research problem statement "Development of Computational Techniques for Preserving Privacy Using Secure Multi-Party Computation Protocols"

Let there are *'n'* parties with private inputs $(x_1, x_2 \ldots x_n)$, *m* anonymizers and a TTP with database '*DB*' or with a computation function $f(x_1, x_2 \ldots x_n)$. Given the inputs TTP compute joint function or on database query '*Q*' traversing the table in *DB*, TTP compute the query and return the result without revealing any additional information to another entities involved in computations. *Figure 3.1* shows the model for collaborative computation environment.

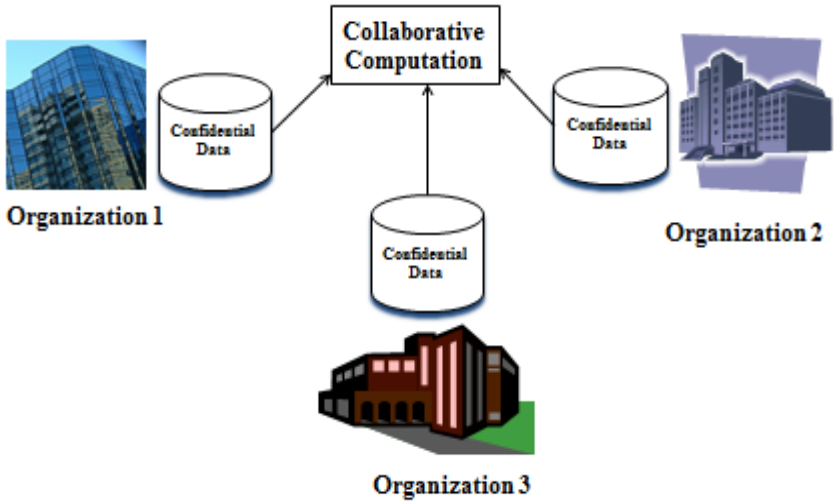

Figure 3.1: Model for Collaborative Computation Environment

3.3 OBJECTIVES OF RESEARCH

The main objective of this research work is to develop protocols to preserve confidentiality and security of parties' private input while computing an arbitrary function over the set of individual input and reducing the computation complexity. The protocols should not be depended on

particular application so that it could be practically used for various real life scenarios.

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

- To analyse necessity and requirement of SMC.
- To design layered architecture for SMC.
- To design and develop computational techniques for preserving privacy.
- To hide actual data from anonymizers and TTP.
- Testing of above modules for real life data.
- To validate the module before and after implementation.

Existing anonymization-based, cryptography-based generalized outcomes are quite complex to be used practically. These results are more competent than the generalized solutions; the proposed algorithms are applicable to some specific problems like EHR, Biometric Identification, and Outsourcing etc. There is the requirement of generalized solution which is not application specific, even though there are many SMC algorithms but complete confidentiality is not assured. This research work will identify the gap in existing algorithms for solving the multi-party computations problem. Thus, this thesis will present cryptography and pseudo-randomization based protocol to retain the confidentiality for SMC problems, through which gap in present conventional algorithms could be removed and novel idea to overcome gap will be presented.

3.4 DESIGN ISSUES

Dealing with adaptive adversaries is one of the fundamental problems in the designing of secure multi-party protocols. To make constructions adaptively secure, some property can be added in standard encryption scheme, to implement this idea on standard encryption algorithms like RSA. Goldreich *et. al.* (1987) suggested the way to play the mental games in case of honest majority. Designer has to consider all the cases where confidentiality and security between the players should not be compromised. It is better to design a protocol for the party who are semi-honest. Consider uncorrupted parties as semi-honest rather than honest. Semi-honest behaviour can be distinguished in three types: (i) Non-erasing, (ii) Honest-looking and (iii) Weakly-honest. Security can be defined by ideal-model for SMC and real-model for SMC.

Whenever system design is weak there is probability of security breaches, which degrades the practicality of the research work. The basic design issue while designing a SMC protocol could be fairness in parties, efficient communication, trust and anonymity.

Fairness of Parties: In SMC environment, it is assumed that participating parties are honest; it means the parties are providing correct input as in joint computation all the parties are going to get benefit from the computation result. In case if any party is not honest and provide wrong input for computation protocol may not work as per the expectation.

Trust Issues: Third party corruption is major problem in multi-party computation environment. Third party can influence the final result by providing fake input or corrupted input. For example a third party might insert fake input before computations, may reveal parties identity by

collaborating with anonymizers or collaborate with some parties to perform maliciously. In the proposed work following trust model is used:

Selection of TTP: This model selects a non-beneficiary trusted third party, who will perform computation on behalf of parties is major challenge. However this arrangement needs all the participants to have confidence in trusted third party. Even though it is assumed the party have trust on a TTP, this research has incorporated packetization, encryption and anonymization techniques to achieve better confidentiality and security.

Anonymity Issue: In SMC environment party-input relationship must be hidden in such a way that no party can be linked with the input they have furnished. To achieve this objective, the work presented in the thesis uses packetization and anonymization. Here, the input is distributed among multiple anonymizers; a malicious anonymizer cannot break the protocol until a threshold $1 \leq t \leq m$ of anonymizers collude. The value of t is generally around m/k. (Here m is total number of anonymizers and k is total number of packet for each party). Such mechanism is fairly better than a real model of SMC where malicious behaviour of a party acting as a protocol initiator can break the protocol. However this requires more communication as data is divided into packets to provide more secure environment.

3.5 RESEARCH METHODOLOGY

The *figure 3.2* shows the methodology used for research work.

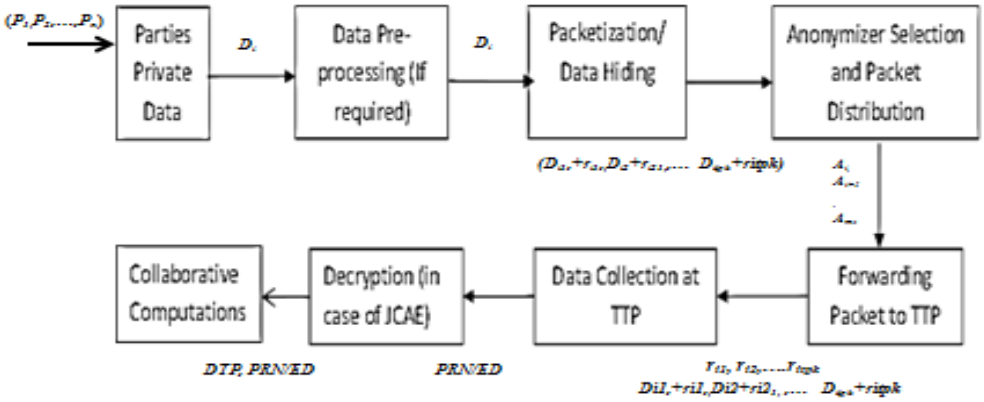

Figure 3.2: Research Methodology

This section discusses the research strategy and empirical techniques used. It defines scope and limitation of the proposed research design. The research proceeds with analysing already existing approaches such as garbled circuit, 1-out-N oblivious transfer, trusted hardware token, linear branching and Neural Networks, linear regression and game theory. After identification of gaps in previous approaches, further research is performed to provide improved protocols. This research is quantifiable as it involves variables such as parties, input parameters, and anonymizers. Once the variables are decided protocols are designed considering the gap identified in literature.

Input: Input for collaborative computation could be given as individual parties input, if connected in local network or internet.

Pseudo-Randomization: To preserve privacy of individual sensitive data, pseudo-randomization techniques are used, in the proposed secure sum

protocols pseudo-random numbers are generated dynamically and added to the plain data packets. First protocol (*JCRA*) uses single pseudo-random number; second protocol (*DRSS*) uses multiple pseudo-random numbers, in order to increase data confidentiality.

Encryption: To preserve confidentiality and security encryption mechanism could be used; in one of the proposed protocol (*JCAE*) encryption technique with packetization is used to maintain data confidentiality and security. For experiment purpose, Asymmetric algorithm is used.

Packetization: The confidentiality concerns are achieved using pseudo-randomization and encryption mechanism. To ensure security from the attack during communication the packetization mechanism is used. Because of packetization, even though attacker gets a packet of a party, probability of getting all the packets of same party is insignificant.

Develop a theoretical framework: Once the input type, pseudo-randomization, encryption and packetization techniques are finalized, the theoretical framework for computation is developed using layered architecture. In this each layer has a predefined functioning and a mechanism to transfer data to next layer. The topmost layer is computation layer, after receiving input of all the parties, computation authority (TTP) at the topmost layer validate all the packets received then performs collaborative computations and broadcast the result.

Data Sources: Data is collected in two ways. Locally generating data using randomization function, or in network from different users. Valuable insight was gained by analysis of previous research studies.

Performance Analysis: After data collection from the parties the algorithm is executed based on the selected algorithm. The performance is analysed considering different scenario such as malicious party, honest or semi-honest party, and turnaround time.

3.6 OUTLOOK OF SECURE MULTI-PARTY COMPUTATIONS SYSTEM

In SMC data confidentiality is essential feature. Our protocols for SMC are designed considering confidentiality and security concerns using different techniques to make it efficient. This leads to various system domain of SMC as follows:

3.6.1 DATA SOURCE

Data is the key element in every SMC application. It is applicable where multiple parties are involved for joint computation. And the requirement is to perform collaborative computation in secure manner. It can be done using TTP as independent entity. It can accept data coming from homogeneous or heterogeneous environment. This data source could be real time data entered by party at run time, conventional database or a data warehouse. Here the main use of SMC is to maintain the confidentiality, security and anonymity of individual participants. This data source cannot get affected by geographical location of the database or input providers. As it needs data from different parties and parties may be situated at different location working on different environment. It does not get affected by the operating system platform of the file system used.

3.6.2 NETWORK

The fundamental need of SMC is the parties and TTP, where parties provide the data and TTP performs collaborative computation over the data. For data transfer a network environment is required. SMC can work with different network environment. The result of SMC can’t get affected

by the topology, encryption algorithm, encryption-mode, operating system or protocols. As parties can be at different locations, or systems so to communicate among them networking system is required.

3.6.3 APPLICATION

SMC is applicable to all the real life applications where collaborative computation is required in secure manner. It is implemented using packetization, encryption and anonymization mechanism. These SMC applications can collect data from any web browser (in network setting) or GUI (in desktop setting), for communication any secure data transfer protocol such as https, http & SOAP etc. or secure file sharing. (for details about SMC applications *cf.* Chapter 6, section 6.8)

3.7 APPLICATION DOMAIN

In this time, huge amount of electronic data is available. To enhance the performance and the development of the organization, data mining and joint computation is required. In globalization similar organizations mine/compute their data jointly. In this process, the participant requires to preserve the privacy of data which is very essential for the organization. When joint computation is performed, the correctness of results needed. Wrong results can adversely affect the growth of the organization. The application domain SMC is suitable for the above mentioned situation. Below sections discusses the application domain of SMC.

3.7.1 HEALTHCARE

SMC has wide application in healthcare domain. The electronic health record makes health provider more equipped and quick to identify disease. But at the same time raises concerns about the confidentiality and security of patients' sensitive information. Now a day's patients are more aware of the information usage, they are selecting the healthcare provider who is preserving privacy of their sensitive information during and after the

treatment. Privacy and security of patients' sensitive clinical and genomic data should be primary objective in a healthcare sector for faster and better treatment and medical research.

Health care system must cultivate trust such that no one can question the health care institution for security of sensitive information. Various methods have been proposed to preserve privacy during genomic and clinical data processing. Fundamentally SMC problems are solved using trusted third party and broad cast the result. In general main issue with SMC approach is finding a third party who is trusted by all the participating parties.

For research and manufacturing of drugs, the research centres require real life data from different hospitals. For security and legal point of view and as per their commitment to the patients, hospitals do not want to disclose the patients' sensitive information. SMC helps here to collect the hospital data for analysis and research, guaranteeing hospitals for the security of the data. This analysis results helps pharmaceutical industry to manufacture better drug for individual or group health. Therefore the special attention is to be given for the correctness of the results.

3.7.2 ELECTRONIC VOTING

In electronic voting systems the voters cast their votes using some electronic media like cell phone, Internet, telephone, etc. For the unbiased election each voter should be restricted to one vote only and their casted votes must be counted correctly. SMC can be applied to meet this objective.

For human centred computing "Prime III" voting system was developed by William *et al.* in this method authors considered users first then system was designed which suits the users. Hence, Prime III became user friendly electronic voting system which incorporates the essential

security, integrity and user satisfaction security measures which must be satisfied by every electronic voting system. It could be easily integrated in present system and improved upon. It is designed considering human computer interaction rules to provide familiarity with the system to make voter comfortable and confident. With user friendly environment Prime III can increase voter involvement in the voting process by allowing individuals with different impairments to vote, i.e. visual, auditory, and/or physical.

3.7.3 FINANCIAL ANALYSIS

In the age of globalization all the organizations are making their data available in electronic format. On the basis of this they perform collaborative computation but they are not willing to disclose their private data to any individual participating in computations. Thus, some privacy concerns need to be assured for the organizations to willingly participate.

For example, two business organizations choose to work collaboratively for an assignment for their joint benefit. Each group would like its own desires to be fulfilled (Agrawal, 2000). Though, their desires are patented data which includes the customer's plans for likely future growth of certain product or service prices, economic information, inflation rates and interest rates. Hence, nobody wish to reveal its desires to the other party, or to a TTP.

3.7.4 CLOUD COMPUTING

Cloud Computing is one of the game changer this day’s. Organization and Individuals are moving from in-premises infrastructures to clouds. But there are certain issue which needs to be considered when moving to cloud. One of the most important thing to be take care in cloud computing is

security of individual participating in the resource, software or platform as a service mechanism of cloud .

3.7.5 ELECTRONIC AUCTION

Now a day's auctions are a widely used electronic commerce technology. Bids are never disclosed to any party, even after the auction completion of auction. In the electronic auction first-price and second-price auctions are provided, and the overall computational costs of electronic auctions are kept sufficiently low so that it can be used for various real-world auction application.

Electronic auction is an application of SMC. (Chowdhry, 2007; Mishra, 2007a) shows there are many companies which conducts online auction. In this type of auction the buyer and seller are not physically present. One of the reasons could be the distance between the auctioneers being geographically large, they can cheat each other. In this a bidder is desired who has the trust of the parties that he will keep their identity and the biding information secure. If the winning party is corrupt, then he can award auction to someone else. Hence, security of bidder is required. This problem is a basically belongs to SMC. (Trevathan, 2005) presents some fundamental examples of electronic auction schemes and illustrates the complexity involved in designing of a secure and anonymous auction system.

3.7.6 GOVERNMENT ORGANIZATION

SMC could be one of the solutions for the government organization who want to share information without disclosing the actual details of the information due to security or personal concerns. For example two government department like public water supply services and public telephone services. Let telephone service provider want to dig telephone lines in an area where water line already exists so they want the information

at which height they should dig or on which side of road they can easily dig. Due to security/personal concerns they cannot tell the exact details with the other department, here the SMC protocol can be applied so that the computation of both the input can give the proper result whether the region specified by the departments collude if so they need to select other region. (Example of secure set intersection)

3.7.7 BIG DATA

It is one of the emerging fields in computer science. As data is growing exponentially it is becoming difficult to process the data. SMC can be useful on big data to get the meaningful information for various applications such as voting, medical, personal and legal without disclosing individual input. The advancement of processor speed, cryptographic tools and cheap and huge storage availability has motivated the new development in secure multi-party computations. The algorithm using secure multi-party computation in big data can help to improve various applications such as medical, government for city planning, voting or legal.

3.7.8 SOCIAL NETWORKING

Social networking has reached even, the remote areas. People are well connected through social networking site. The communication frequency is increased in multiples. There are 1.23 billion users on one of the social networking site, out of which 950 million users are active at a time. This is one of the best platforms for target marketing, survey to promote awareness, or to perform joint computation on the users input. This leads to privacy concerns as users may not be interested in sharing their personal information with others. Here, SMC mechanism can be adopted so that users' information and identity will be secured during joint computation for mutual benefit.

3.8 SUMMARY

In this chapter section one discusses the motivation behind research. On the basis of this problem has been defined. The problem definition gives clear idea about the research work to be conducted. Once the problem is defined we can list the objective of research. Then the system design has been formed. Section four discusses the various issues faced during the protocol design. Section five shows various performance parameters considered for the proposed research work. In section six the research methodology is formulated on the basis of above findings. At last section seven presents the system domain such as: i) what could be the data source and how it works? ii) Which network environment is used? iii) What are probable application and how it works?