CHAPTER 2

LITERATURE REVIEW

Data is an asset; it can be utilised to create a decision support system by capturing various transactional databases, extracting, transporting and integrating together in a data warehouse. The pools of data in data warehouse are usually being explored by data mining tools to acquire significant data that might be unknown (Mishra, 2008; Ronald, 2004). So Data mining (DM) can be stated as the process of mining information from large amount of data.

As data is an asset; it can be personal or corporate data which can be helpful in extracting pattern for the databases. It can be from different sources so the need is that without disclosing the individual data, how it can reach to the final computation that can reveal results for decision making. This problem was considered as secure computation from the database without revealing sensitive information.

2.1 MAJOR CONCEPTS

In the proposed research the major concepts are represented in *figure 2.1* are as follows:

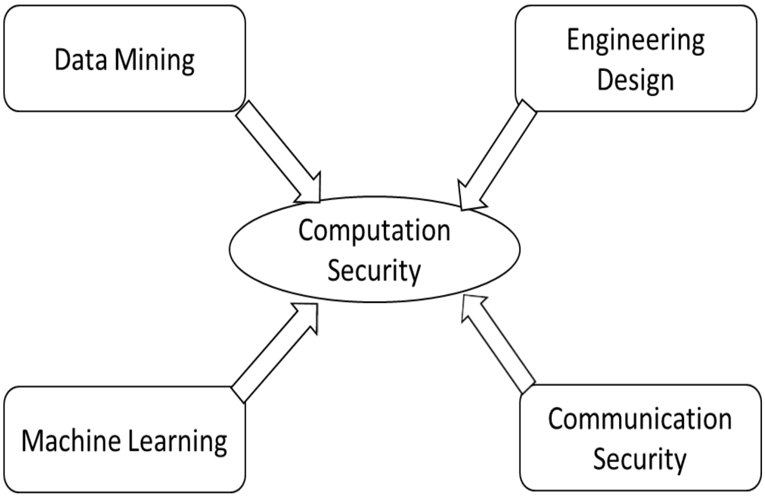

Figure 2.1: Area of Relevance and Contribution diagram (ARC diagram)

- **Computation Security:** Computation Security is desired when multiple party collaborate to compute over same function. To enable confidentiality conservation with the help of TTP mechanism. It enables confidentiality preservation with or without TTP [*See Section 2.2.1*]. There can be significant communication between participants to get the outcome but participants do not discover anything from this communication (Yao, 1986).
- **Data Mining:** Data mining is a well-known process by which useful knowledge can be extracted from the raw dataset. Data mining is defined as "Knowledge Discovery from Data" or in other words, "Data Mining is a practice or a process which is used to draw useful understanding from the huge amount of dataset collected by human". Dataset can be collected from different organizations, which is stored by them for different purposes. For example, the business organization such as amazon, flipkart etc. maintain customer data and buying pattern record for future marketing; the federal agencies for countrywide security keep records of people with felonious intent and criminal's for security reasons; and the health organization collect medical records of the patients for better treatment and medical research. Some of the major domains are Business, Education, Government and Medical.
- **Machine Learning:** It is about the building and training of systems that can learn from data.
- **Communication Security:** It is the area which prevents unauthorized attacker and eavesdropper from retrieving communications in an intelligible way, while still providing message to the designated receivers.

- **Engineering Design:** It is the origination of a plan to help an engineer form a product with a stated performance objective. This procedure comprises of a number of phases, and each phase of the procedure may require to be repeated many times before fabrication of a final product.

2.2 LITERATURE REVIEW AND RELATED WORKS

The initial concept of SMC has been brought in notice by *Yao* in (Yao, 1986; Yao, 1982) and then *Maurer* extended it by defining different type of security and the applications of SMC (Maurer, 2004). The first solution for this problem has been proposed by "*Yao"* in the form of two party computations for semi-honest party. Its extensions to malicious party were given by *Lindell* (Lindell, 2007). *Lindell et al.* (2009) presented the SMC issue as; there exist two participants with their personal databases and wish to perform data mining operation on joint database. Similar problem was addressed by *agarwal et al*. (2000). Both researches came out with different solutions *agarwal* solved it using data perturbation while *Lindell* considered SMC techniques.

The background of SMC was initially identified as two millionaire problem; in this two parties wish to know who is wealthier without revealing their wealth (Goldreich, 1987).

SMC problem comes, when there are multiple participants who wish to perform a collaborative computation to produce outcome without revealing their personal data. This joint computation can be performed in different ways: firstly between the competitors, secondly between the trusted participants and computation authority and lastly between the semi-honest parties (Agarwal, 2003). This is applicable to various applications such as preserving privacy in databases, healthcare and financial analysis.

To solve the single-input computation model same transformation cannot be applied. One possible solution is given, where they follow some assumptions; if this computation is assumed to be *C*, and there exist single input as a set *D* of multiple tuples. *D* can be divided into two or more distinct dataset $D_1, D_2, \ldots D_n$ and so it could be changed to multi-input computation model. By such partition different types of transformations are focused: heterogeneous transformation and homogeneous transformation. Heterogeneous transformation states that, every single data item is divided into parts, each part goes to isolated dataset and homogeneous transformation means *D's* data items are divided into sets of same type, but every single attribute does not split into parts (Du, 2001b; Agrawal, 2000).

Current problems, under investigation are confidentiality conservation in geometric computation, statistical analysis, database query, intrusion detection and cooperative scientific computation, etc. There are two basic models to provide input in SMC mechanism, first model is, multi-input computation model and second model is, single input model. First computation model has two or more distinct inputs but usually it has only two. For the *Yao's* protocol a proof has been given in (Lindell, 2008) for the two party computations. On the other hand another research provides a solution which is given as the practical approach for secure two party computations. For example, client-server computation is a multiple-input computation model. Data mining is single-input computation model; all the inputs generally come from one dataset which contains various data items. Here, if we assume that each input comes from different parties then a new problem comes up as "how to perform the same computation while assuring the confidentiality for all the data" (Pinkas, 2009).

With SMC, a number of parties can cooperatively perform some global function on their private data without any loss of data privacy. It provides support for end-to-end secure multiparty protocol development.

Dorothy et al. considered security as control system. According to them in order to standardize reading, deletion and updation of data and program, needs to have access control mechanism. It avoids the malicious or inadvertent disclosure, deletion or modification of records, and program segments. They looked at two important classes of access control mechanisms: transaction programming system and general programming system. In transaction processing system there are two types of limitations. They are data dependent limitations and history dependent limitations. Data dependent limitations lock the records which are currently accessed by the transaction and history dependent limitations are those functions of the records previously accessed. The general purpose system provides access control mechanism as a part of run time environment. In this access to the object is controlled regardless of the value stored in the object (Dorothy, 1979).

An oblivious transfer is a protocol in which a source sends specific data to the receiver, but remains unaware of what is sent. *Rabin* (1981) presented the first form of oblivious transfer. Suppose party *A* has two messages m_0 and m_1 and wants to interact with party *B*. In this process party '*A*' sends its messages to party '*B*'. Party '*B*' has to select only one message out of two. In 1-out-of-2 oblivious transfer party *A* never knows that which message was selected by '*B*'. The concept can be extended to *1-N* obvious transfer (Even, 1985).

Yao observed simple two millionaires problem. This problem is about two millionaires who wish to identify that, who is wealthier amongst them, without revealing their wealth. The protocol states that Alice and Bob have a public one way function and the inverse is known to them only. They exchange string with each other one after another. When the string derived from Alice, Bob compare it with his sequence. The main advantage of this technique is that no one could understand and change the data as it is in encrypted form, but still could decide that who is richer (Yao, 1982).

Goldreich et al. (1987) gave polynomial-time algorithm in which inputs are the description of the games with partial information of number of players. This protocol is able to secure all information for playing game, provided the majority of players are honest. They consider two types of faulty machine in a game network one passive and other malicious. It uses the TTP to preserve confidentiality and maintain correctness of distinct player. The approach for computation function *F* follows a combinatorial circuit and then the participants run a protocol for all the gates in the circuit. The protocol size depends on the input size. This is all related to the circuitry theory for playing any mental game.

Du et al. (2000) discuss secure remote database access problem and presents solutions for achieving privacy for the different secure remote database access models. *Omote et al.* (2000) present a scheme for electronic English auction which satisfies all the characteristics of an English auction.

Wahlstrom et al. (2000) present the social issues arising from the data mining. Overview of technology is presented with reference to cultural context of each issue. The authors have done feasibility analysis of existing solutions and an efficient solution is proposed and outlined.

Du et al. (2001a) present various computation geometry problem which could be solved using SMC. And protocols to perform statistical analysis in cooperative environment, based on cryptography and data perturbation techniques (Atallah, 2001).

Lindell et al. (2002) presents an improved implementation of the two party cases, using Yao's garbled circuits (GCs). *Du and Atallah* gives statement of SMC problem and various applications (Du, 2001b). This paper gives guidelines for SMC research with different applications where SMC can be applied efficiently.

Verykios et al. presents various approaches to protect sensitive rules during transaction processing. (Clifton, 2002) *Clifton et al.* gives tools for privacy preserving data mining; in this random number mechanism is used to preserve privacy of individuals. In this, if two parties collaborate they can get the data of third party (Verykios, 2003).

Agrawal et al. presented new protocols for different functions intersections, size and equi-join. And demonstrated that these protocols revealed insignificant information apart from what can be revealed from the query result. They presented a method to compute equi-join size but this methodology outflows some information about tuples which are combined, on the basis of duplicates distribution (Agarwal, 2003).

Maurer presented the role of cryptography to achieve security in databases and addresses the issue of specifying and accomplishing confidentiality in a framework where the database is not fully trusted (Maurer, 2004). *Verykios et al.* gives an overview of privacy preserving data mining techniques. A detailed review and classification hierarchy of previous published work has been given (Verykios, 2004).

Zhan et al. (2004) present the randomized response techniques to perform privacy preserving data mining operations. In this paper authors considered multi-group i.e. attributes are partitioned in specific number of groups. *Brickell et al.* (2005) present a SMC based algorithm to compute shortest distance and secure union in the environment where parties are “honest but curious”.

Trevathan(2005) present a model to conduct secure and anonymous online auctions. Methods are proposed to detect fraudulent in e-commerce. The proposed models have been implemented on online auction server. It can be used for various real life online applications.

Karr et al. (2005) presented the case, when data is stored in distributed databases and regulated by various statistical organizations then what is the way of accomplishing, secure linear regression for “horizontally partitioned data”. They also proposed the methods for the records that use the secure sum protocols, MPC protocol, to find the least squares estimators for disjoint sets of data.

Liu et al. (2006) explore probability of using multiplicative random projection matrix for protecting distributed data privacy during data mining. *Raymond et al.* (2006) present (α, k) anonymity prototype to protect identification and associations of critical information in data. In this paper, quasi-identifier and equivalence class concepts are used for global and local recoding. This work, experiments different variables set and comparative study of the result.

Mishra et al. (2007b) provide an outline in the form of protocol for SMC, during data mining for Indian Business process outsourcing. It is proposed that solution will help to secure BPO processing to enhance business. The work presented a conceptual framework, which can be implemented to find out efficiency of proposed protocol. It doesn’t deal with individual privacy.

Srivatsava et al. (2008) give partitioning based anonymization based scheme and randomization based scheme for composition attacks. Authors considered two different data sets, from UCI machine learning repository, and considered subset of attributes which are common for experiment of intersection attack.

Mishra et al. (2007a) proposed the protocol closely related to the proposed protocol. They considered a setting which includes multiple parties with private input and a trusted third party without any contributory input. In these setting parties direct data to randomly chosen anonymizer and then anonymizers forward the data to the TTP. In this protocol there is a possibility of misconduct as data is sent as it is to the anonymizer and if anonymizer becomes malicious and collide with any one party then protocol may break.

Kargupta et al. (2007) provides a game-theoretic framework for development and analysis of privacy preserving data mining games. It delivers a game theoretic result built on the idea of "cheap-talk".

Mishra et al. (2008) presents a zero knowledge protocol for computation security. In this authors used multiple TTP to ensure trust in third party. This protocol needs information to send result to respective parties. It could be a security breach as the address of parties is stored to deliver the outcome of computations.

The PPDM problem is an explicit SMC problem that has been extended in the literature review. (Lindell, 2009) the problem is presented as: two parties, with their personal record, willing to cooperatively perform DM operation on the intersection of their private databases. How can these parties conduct this computation without revealing their personal data to each other or even to another TTP? Other than this problem SMC problem occur in various other computation domains, such as EHR, statistical study, arithmetical computations (Liu, 2006; Du, 2001b; Barni, 2009).

Bryant et al. (2008) present “Future of data security and privacy: controlling big data” identified that security authorities apply most of the conditions at network level. In this case if the attacker breaks the circumference of the network, they can get complete unrestricted access to the big data. So it is beneficial to keep these conditions, closures to the data. As the participants' first concern is their data security so the data packets must be highly secure against attacks.

Bogetoft et al. (2009) present the practical application of multi-party computations for secure auction. *Sheikh et al.* (2009) present a secure sum technique using unidirectional ring. In this data is divided into segments and shared among the other participating parties to add subsequent segment, this process is repeated until all the participants add their segments. At last protocol initiator announces the result. It does not consider semi-honest and malicious parties, and communication threat.

Mishra et al. (2009) present extension of their previous work. In this authors are considering multiple TTP’s to overcome untrusted TTP. TTP is selected at runtime to hide the identity of TTP. Here, the function pool is accessible to all the TTP’s and parties.

Barni et al. (2009) give a privacy preserving system, where a server can classify an electrocardiogram (ECG) signals without getting any knowledge from the signals. In this client doesn’t learn anything about classification algorithm used by the server (*i.e.* parameters of classification are secret). This protocol experimentally compares different implementation of the proposed system, one based on garbled circuit, and other based on paillier crypto system and garbled circuits (hybrid).

Sheikh et al. (2010a) present a ck-secure sum protocol, in the proposed work authors divided data into fixed segments. The authors claim that, here probability of data leakage is zero. As this secure sum protocol uses changing neighbour mechanism, where neighbours are changed in each round with fixed protocol initiator. For this protocol to work accurately, minimum four parties are required. It doesn’t deal with malicious behaviour of party and attacker during communications.

Sadeghi et al. present how to combine a trusted hardware token (a cryptographic coprocessor like IBM CCP 4758) with secure function evaluation (SFE). It computes function on encrypted data where no information is leaked and outcome is verifiable. In this paper authors’ achieved low latency (i.e. time from submitting the query until the outcome of computations is received). The token is integrated in setup phase. In time critical online phase cloud performs computations on encrypted function on encrypted data using symmetric encryption mechanism. In this the hardware token ‘*T*’ is integrated in infrastructure of the server ‘*S*’, which is capable of computations on behalf of client ‘*C*’. Here, cryptographic coprocessor generates secret key and internally & securely transport them to ’*C*’. The cryptographic coprocessor is provided by ‘*S*’, if the client doesn’t trust manufacturer of cryptographic coprocessor, client can choose his own manufacturer for hardware token and ship it to ‘*S*’, to integrate in its infrastructure. The limitation of the proposed design is that, if client want to ensure trust, client has to tolerate all the expense of coprocessor of his choice (Sadeghi, 2010).

Sheikh et al. present dk-secure sum protocol using ring arrangement. In this paper, parties exchange any 1 out of *k* segments, with any 1 party out of *k*, so all the parties have *k-1* segment, plus 1 received from other party. In this protocol if two parties collaborate to get third parties data, then it may break the protocol (in case of three parties). This protocol works efficiently for four or more parties. In this paper authors assumed that communication channels are secure, it doesn't deal with insecure communication channels (Sheikh, 2010b).

Sheikh et al. explain the importance of modified ck-secure sum protocol over ck-secure sum protocol. In this protocol initiator changes its position in unidirectional ring, so that no neighbour remains together for more than one round. Here, data is divided into *n* segments (where '*n*' is the number of party). On n^{th} round initiator announces the sum. For this protocol to work effectively minimum four parties are required. It doesn't deal with malicious parties and security threat (Sheikh, 2010c).

Barni et al. (2011) present development of automatic diagnosis system where a remote server classifies biomedical signals, provided by the clients. During this classification server doesn't get any get any information about signal and the final result of classification. In this authors' uses linear branching program and neural networks for classification of electrocardiogram (ECG) signals.

Kamara et al. (2011) present a protocol design which reduces computation of the parties at the expense of server. In this general and special purpose server aided multi-party computation protocols are proposed. Authors' states that these protocols are more efficient than a standard SMC protocol without server. In this general purpose protocol provides security when at least one party is honest. This protocol is extension of FKN protocol. The limitation of the protocol is that, it doesn't guarantee output delivery.

Bogdanov et al. (2012) describe a secure system, for joint analysis of financial data of companies. To ensure privacy secret sharing and SMC are used. In this author doesn’t deal with confidentiality concerns of participants.

Miyaji et al. (2012) presented an efficient and private set operation without using homomorphic encryption, oblivious transfer and zero knowledge proof. It is constructed using game-theoretic model. This protocol works for rational parties. It is assumed that party will never deviate from the protocol.

Teo et al. (2012) give experimental analysis of accuracy and efficiency of fundamental secure multi-party computations protocols. It introduces the new dimensions for improvement of computation efficiency in multi-party online real data privacy preserving data mining in cloud computing. Nergiz *et al.* (2012) present a look-ahead approach, to achieve distributed k-anonymity to help parties to decide, whether the selected protocol will meet the expectation before initiating it.

Kui et al. (2012) discuss security demands for the public cloud. It is most challenging computing paradigm shift in information technology. In this article authors presented various confidentiality and security aspects which are the main complication for extensive acceptance of the cloud.

Bogdanov (2013) present the sharemind a tool for programmable secure computation and its practical applications. *Jurczyk et al.* (2012) deals with challenges emerged from cloud computing. The framework presented here, enables privacy in data federation services. In this paper, distributed anonymization service for multiple parties is provided using virtual anonymized database. It doesn’t deal with adversarial behaviour of parties, communication and computations security threats.

Dunning et al. (2013) proposed an algorithm, to share private data among multiple parties using anonymity. The sharing algorithm is used in iteration for anonymous ID assignment (AIDA) to parties. In this communication authors assumed that the communication channel is secure. It doesn't deal with insecure network.

Fournet et al. (2013) present a query language for computation on private data along with verification of correctness of result. They proposed a zero knowledge protocol which assures integrity of result and data privacy. For performance evaluation queries are tested on smart-meter, location based services and pay-as-you-drive insurance policy.

Rane et al.(2013) consider biometric as an important and widely used methods for identity verification and access control. The usage of biometric raises various security and privacy challenges. There are various different mechanism to deal with biometric authentication such as secure sketches, fuzzy commitment, SMC and cancellable biometric. As it is related to individual's sensitive data, it needs to be preserved against various adversaries. In this authors presented various methods which can be used for secure biometric identification and future directions for research.

Erman *et al.* (2013) present a confidentiality conservation mechanism for storing and processing clinical, genomic and environmental data by using privacy preserving integer comparison and homomorphic encryption. In this DNA sequence of patient is created by certified institution using the sample provided by the patient. The environmental and clinical data of the patients are collected directly from the doctor, patient visits to laboratory, or could be directly provided by the patient. (For ex. age weight, family history by patient whereas cholesterol level blood sugar level by his/her doctor's visits). All these information is considered as sensitive and need to be protected. Here, SMC is applied to preserve privacy of patients, against curious parties at storage and processing unit (SPU) and malicious parties at medical unit (MU). Genome is next big thing in medical science to identify disease risks. It could be possible by ensuring privacy of patients' sensitive data during the tests (Erman, 2013).

Kolesnikov et al. (2013) gives generic secure function evaluation framework for secure two-party computations. It represents functions as, Boolean circuits (BC), Arithmetic circuits (AC), and ordered binary decision diagram (OBDD) for secure function evaluations. Various security concepts such as semi-honest, malicious and covert adversaries are described.

Recently, *Padwalkar et al.* (2014) presents a hybrid technique of secure multi-party computation, in this author uses random number for data privacy, in the hybrid protocol participating parties and third party contribute for computation so it will be faster. This paper does not deal with communication security threat, when some parties are targeted purposefully and the case when third party becomes malicious.

2.2.1 BASIC MODELS OF SECURE MULTI-PARTY COMPUTATIONS

Let parties P_1, P_2... P_n be n parties (organizations or individuals) want to conduct cooperative computation C_i on their personal sensitive data. Since, computation is to be carried out on private data, it is key constraint that this private data should not be available to any other party, i.e. if $D_1, \ldots, D_n$ be the data corresponding to n parties and let D_i be data corresponding to ith party, then it is required for computation that, D_i should not be accessible to any D_j where $i \neq j$ and $j=1, 2\dots n$. Therefore, each party gets the final results of cooperative computation without being aware of inputs involved and the computations made. Various models have been presented in the literature for the SMC problems. Generally, two model prototypes of SMC are used:

- Ideal Model Prototype of SMC.
- Real Model Prototype of SMC.

In the Ideal model (Sheikh, 2011) an uncorrupted TTP among participating parties or out-side the parties is considered. Participants direct their personal inputs to the computation authority that will perform collaborative computation for all the participants. The TTP is supposed to be trusted by all the participants. It means that, TTP will never disclose the personal data of one participant to others. The TTP, after computation of function on personal input projects the outcome result to all the participants. Here, only computations result is known to all the parties. In this way, the individual security and confidentiality is conserved. In Ideal model, if a few parties behave maliciously then the outcome of the collaborative computation may be incorrect because the party may provide wrong input to the TTP but the participants' confidentiality will be protected. If the computation authority becomes dishonest, the confidentiality could be compromised. The ideal model of SMC for two parties is shown in *figure 2.2.* It can be carried for multiple parties. Participating parties offer their personal data inputs $x_1, x_2\dots x_n$ to the TTP. The TTP then calculates common function $f(x_1, x_2\dots x_n)$ and directs the computation result to the participants. In real life the role of computation authority is performed by government or private association which works as a service provider for collaborative computation. This model is costly as getting trusted computation authority needs to be paid more. Here, main requirement is the reliability of the computation authority. When computation authority becomes corrupt complete concept of the secure joint computation becomes insignificant. This day's Ideal model is widely used because of availability of intelligent tools.

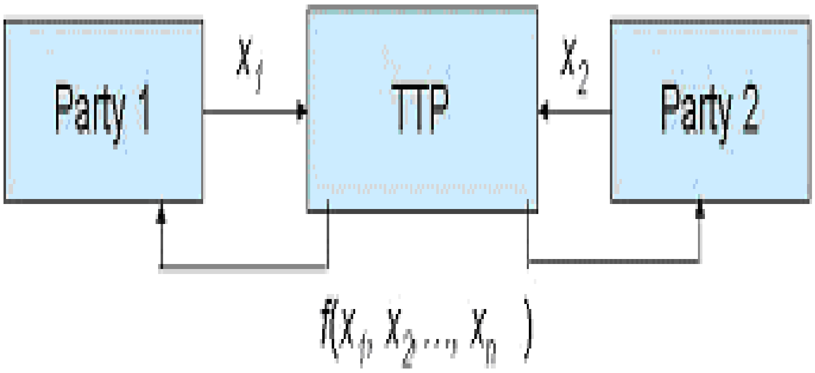

Figure 2.2: Ideal Model of SMC

In real model (Sheikh, 2011) no outside participants are assumed who can be trusted. In this model participating parties agree on a protocol which is to be executed among them in order to maintain confidentiality and correctness of the result. The prototype for two participants is shown in *figure 2.3*. Here, the computation participants do not contribute definite inputs. The inputs provided by participants are function of their personal data. What exists between the participants is a hypothetic computation mechanism. An adversary is a participant with malicious purpose.

An adversary (Sheikh, 2011) could be static or adaptive. A static opponent is malicious before the execution of the protocol. A semi-honest opponent monitors the protocol but tries to learn something more than the computation outcome. An adaptive opponent becomes malicious during the protocol execution. A malicious opponent is one who does not monitor the phases of the protocol instead tries to learn some information other than the outcome.

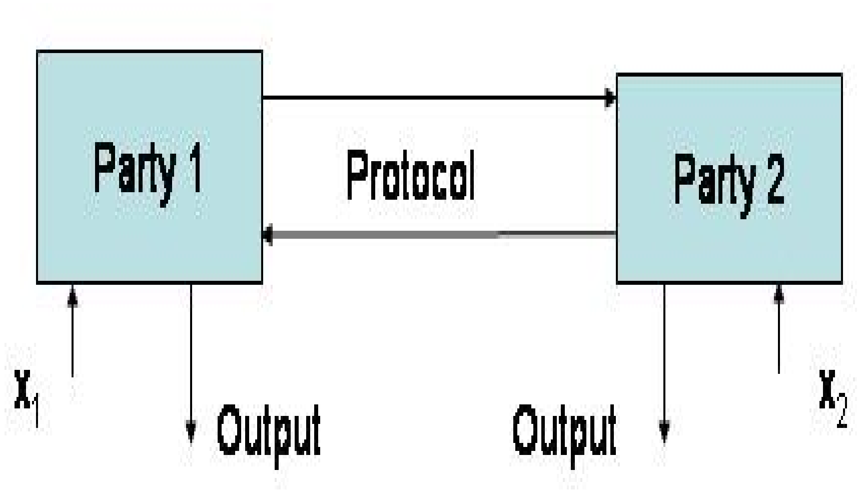

Figure 2.3: Real Model of SMC

2.3 PROBLEM AREA

This day’s electronic data is increasing at an unpredictable rate. And as the data is available electronically we can access it and can be used for joint computation so the confidentiality is required in this case. This problem has been accosted by organizations many times.

In SMC, research work is based on some probabilistic function ‘*f*’ on ‘*n*’ inputs in a distributed system where each participant holds an input ‘D_i’, all the inputs are independent of each other and correctness of computation should be ensured without revealing anything other than what can be revealed from the output. Main objective here is to provide confidentiality and security but it increases the complexity of protocol. *Kalle et al.* (1987) have discussed the different perspectives of the information system. To design the PPMC architecture following issues should be considered.

2.3.1 PROTECTION OF SENSITIVE INFORMATION

People are cautious about their confidentiality. While performing collaborative computation, we need to ensure some kind of confidentiality, than only joint effort can be made successful. In order to ensure that some confidentiality issues need to be taken into consideration that is beneficial for overall organizational growth.

The data is the lifeline of any organization. Millions of dollars are spent to keep the data private. When multiple organizations collectively want to get some results on similar data, multi-party computation is the answer. During this practise the confidentiality of data is an important issue. SMC uses the data of several parties to compute the result while maintaining the confidentiality of individuals. Data independence is an important issue for the system. It must be taken care that the data of one party must not leak to others during the process. The user wants to keep his data confidential but at the same time want to perform collaborative computation for mutual growth. Collaborative computation is helpful for city planning, education, and to improve public health. For this purpose the system should be such that, all data should be taken at one place, keeping individual confidentiality in mind. To maintain the confidentiality of individuals the pseudo-randomization and encryption techniques are used in the protocols presented in the thesis.

In collaborative computation multiple information providers collide and share their input this could be there personal, health, wealth or business etc. related information. This information is private to the individual or organization of concern. This information sharing is essential for better decision making, research and development. In collaborative computation parties with same interest join together, so there are chances of information leakage, due to this reason computation participants are unwilling to share their personal information. If there is some assurance of data confidentiality then parties can share their private data without any fear.

2.3.2 SECURE DATA TRANSFER FOR COLLABORATIVE COMPUTATION

Although communication security is a separate area of research, it is required in SMC environment also. There are few protocols available which can be used for secure data transfer over network. This research work is using https in the presented protocols for secure data transfer. As this research work is using packetization, anonymization and encryption, there are three level of security provided. Here, in-order to get intended parties data all the packets along with decryption key should be known then only it will be meaningful, it is insignificant in this case as decryption key known to TTP only and identity of packet is not known to TTP.

For collaborative computation everybody wants security. For this there should be some reliable participants to whom data can be given without any doubt of data being misused. In this research work, TTP is involved in computation as a reliable party to whom every participating party trust.

2.3.3 CONFIDENTIALITY AND ANONYMITY ISSUES

Anonymity and Confidentiality are two mandatory factors for collaborative computation in case of ideal model of SMC. Confidentiality feature assures parties that, they can share data without any worry about data being seen by competitors. Anonymity is incorporated in protocol to cover the identity of participants from the TTP. Both the features are required to achieve the basic objective of the research.

Mostly, the contributors in SMC mechanism get joint profit, so they are expected to be cooperative and honest. Although they wish to solve common problem with cooperation but they want to protect their private input from each other. In various conventional algorithm encryption techniques cannot be employed. So there is a requirement of plain data (non-encrypted) for computation using conventional algorithm. For this parties need to share their private data for computation which may lead to security challenges. The confidentiality concern in SMC is to prevent disclosure of data from unauthorised access, while sharing data among parties for the collaborative computations.

For example, in the meeting scheduling problem manager wants to know whether the scheduled meeting overlap with any other team members' pre-decided meeting. However no one is interested in disclosing his/her personal schedule so in this case without revealing any other details about the individual meeting and identity of the parties, manger has to choose the appropriate time for the meeting. It can cause substantial harm to the party if it is revealed to other participants, as it is related to personal calendar information. Therefore there is requirement of a protocol to solve such problems while protecting the confidentiality of their personal information.

In the Electronic Health Record (EHR) problem, John has disease database, and Marry want to find out whether she has the disease without disclosing her identity, to John. In millionaire problem, two participants want to know who is wealthier without revealing anything about her/his wealth. For this situation SMC is the best suited method for maintaining confidentiality concerns.

2.3.4 IMPLEMENTATION AND TESTING OF PROTOCOLS FOR SECURE MULTI-PARTY COMPUTATIONS

The protocols are implemented and tested in network as well as standalone environment. There is few existing protocol, based on real model of SMC, which work for at least 4 parties, security can be breached if '*n-1*' parties combine to get *n*th parties data. The proposed protocols are designed and developed such that it can work in the above mentioned situation and maintain the basic objective of the research.

Secure sum is a real world problem in which all the organizations deals with their authoritative data. Here, the system is designed to implement in real world. To implement a new system current system must be analysed. The existing secure sum protocol proposed by *Clifton et al. and Sheikh et al.* have been analysed to identify the gap and research work is improvement over this protocol in case of minimum number of participants, anonymity and confidentiality.

To implement secure sum protocol in the research work confidentiality and security are considered. To provide security and confidentiality different protocols are designed. To communicate between different layers data must be in cipher text form, to ensure that there are no security threats.

2.3.5 COMPLEXITY ISSUE

As amount of the data is increasing at tremendous rate so it increases the complexity of data being fetched as for getting information one has to go through the complete database. Here, data confidentiality and security is taken into consideration so, it becomes more complex to come up with the desired results.

The system should not be too complex to implement practically. In general, when there is increase confidentiality and security of a system the complexity increases. Because confidentiality and security are the main issues of SMC, the system is bounded to become complex. So while designing the architecture of SMC applications, the complexity issues should be considered otherwise the system will remain theoretical.

The protocol to solve to this problem is defined by Ueli Maurer and Dalhli (Maurer, 2004; Dalhli, 2002). In this protocol it has been observed that if all students follow the protocol, no student get the information about the marks of other student. On the other hand at the end of the protocol all students have an average of their marks. In this protocol, what happens if one of the students becomes malicious? If he follows the protocol and doesn't get information from others, he still can learn nothing about the data of other students. On the other hand, nothing prevents him from breaking the protocol or leaking information to other student. He easily can halt the computation just by not sending his share to the next student. A number of protocols are available to secure the data confidentiality. In research it has been seen that, the protocols which provide more security, increases the computation and communication complexity.

2.4 ADVERSARIES IN SECURE MULTI-PARTY COMPUTATIONS

An adversary is a malicious entity whose aim is to prevent the users of cryptographic computations system from attaining their objective. An adversary may try to find out secret data of other user's, falsify the identity of a sender or receiver, modify some data provided by the users.

Adversaries are the parties who will either try to get the data of other parties during the computation or corrupt the entire or part of the data which will be provided for computation. It is common, when malicious behaviour by an adversary may corrupt partial data of parties. An adversary may be thought as hackers who try to get parties data during communication. When a party is corrupted, the adversary gets all the data of the party including all the private information (Ronald, 2004). Various types of adversaries exist in secure multi-party computation (Chowdhry, 2007; Goldwasser, 1996).

2.4.1 HONEST ADVERSARY

Honest adversaries are group of parties who follow the protocol and never deviate from the protocol. This is very exceptional in real time environment.

2.4.2 SEMI-HONEST ADVERSARIES

It follows the protocol but tries to acquire more information from received message. It is a group of participants in the system who contribute to the computation without diverging from it, but try to get other's information from the output of computation.

2.4.3 MALICIOUS ADVERSARIES

This type of adversary diverges from the protocol in arbitrary ways, alters their inputs, and may leave the protocol at any point from the system. In this correctness of result will not be maintained.

2.4.4 BYZANTINE ADVERSARIES

Byzantine adversaries are the collection of users in the network who can diverge from the protocol in an arbitrary way based on their input as well as message received in the protocol.

2.5 SCOPE OF RESEARCH WORK

SMC can be used wherever a collaborative computation is required, where individuals or organizations are interested in preserving their personal data and identity. The proposed protocols are based on SMC. Scope of SMC is not limited to secure sum but it can be applied to following areas:

(a) Privacy Preserving Biometric Identification: Now a day biometric identification is widely adopted and proved to be practical and efficient for personal identification, identity verification and access control. Biometric are unique and can be utilized to recognize someone among a large group. If this biometric data is stolen, it can be used for unethical activities like tracing individuals or identity stealing. These need new approaches to develop secure biometric systems to reduce leakage of biometric data.

(b) Electronic Health Record: Health care industries are adopting technology at faster rate, it include personalize online services, remote diagnosis services. Biomedical signal are very crucial as it contains individuals health information which he/she may not be willing to reveal. Confidentiality prevention is very essential in various biomedical signals processing application such as electrocardiogram (ECG).

(c) Outsourcing: Outsourcing work to a trusted third party is adopted by many organizations as they need to parallelize the work at the same time they want to reduce the expenses of maintaining own resources. It needs assurance of security and confidentiality from the service provider as it may lead to security breach and party may not be interested in such service providers.

(d) Collaborative Query on Distributed Database: All the organizations have their database server. The numbers of database servers are based on the geographical positioning of the office of the establishment and their applications. All these databases are independent. Sometimes organization desires information which is the collaborative results of a query over the distributed data. As all the databases are autonomous so they do not want to disclose their information.

(e) Database Query: *A* has a database string '*q*' and *B* has a string's database $S = S_1,\ldots,S_n;$ instantly *A* wish to find out whether there exist a string S_i in *B's* string database which matches or almost match with '*q*'. For ex. John has the database to identify diseases and Marry want to know whether she has the disease by hiding her identity. So that John will not come to know about her disease.

(f) Market Analysis: In an organization, for each manufacturing unit there is a Product life cycle (PLC) and in order to maintain the inventory system market analysis is required, let a manufacturing company wants to know its relative place in the market with regard to competitors. To do this competitors personal information is required which they may not be interested to share. This valuation and assessment will help the organizations to improve their productivity and maintain inventory. SMC can be useful for such type of problem to review the performances and styles followed in the market, promising the confidentiality of entities involved.

(g) Collaborative Audit: A collaborative audit is an audit of a legitimate individual by two or more examiners to create an inspection report, there by dividing duty for the audit. In the collaborative audit, plan is made taking concerns from all the auditors together and the field work is divided among the auditors. Typically auditors are group of people, rather auditing firms. The audit work is rotated among collaborating auditors and work performed by one auditor is re-examined by other. The decisive concerns at group level and merging of individual audit report are re-examined collaboratively and there is collaborative reporting to the legitimate individual's management or its concerned authorities.

(h) Privacy Preserving Geometric Computation: Marry and John has their private shapes *a, b* respectively. They want to know whether they overlap with each other or not. Both the party does not want to disclose their private shape to each other. Moreover no-one should acquire anything information about relative position between the shapes *a*, and *b.*

For example: Two organizations want to expand their market share in some regions and both the organizations do not want to invest in same region. They only want to know that whether the selected region has some overlapping points.

(i) Electronic-Commerce: E-Commerce is the word used for electronic business transactions, ordering and retailing of commodities and services online. E-Commerce refers to the trade transactions between organizations and their consumers (B2C) or between organizations (B2B). E-Commerce can be defined as the use of digital information processing through electronic media to do business online. Through this customers can speed up ordering and payment systems. The experiences of people using E-commerce is not very good as for purchasing products on internet they have to uncover their credit/debit card number which could be misused by the vendors. SMC could be used to get over this problem.

(j) Privacy-Preserving Statistical Analysis: The problem such as, for completion of periodical banking transactions, a bank has to share its client's information to other bank. Sometimes the malicious banks may misuse the information for their profits. SMC can be applied in such scenario to secure the information of clients of one bank from another. It can be applied for collaborative computation of banking/financial data.

2.6 SUMMARY

In this chapter, literature has been reviewed and presented the research perspective of SMC. The related literature was studied and the prospects of SMC were explored. The different terminologies related to SMC were presented in chapter one. The special methods used for computations were explored and used to develop the new protocols. Various types of adversaries and their effects were identified during computation. Appropriate changes were advised during the designing of different protocols. In the some research papers and literature, the solution for the SMC is mainly based on packetization and anonymization techniques and complete belief on TTPs. In the proposed work encryption and pseudo-randomization are used between parties and TTP. This concept improves the confidentiality and security of the system.

During the research work various aspects of SMC have been studied. The main issues identified in the literature survey are confidentiality, security and complexity. Complete system is designed looking at the problem in different perspectives like data sources view, network view and application view. Some SMC application areas have been identified.