SECURE MULTIPARTY COMPUTATION VIA OBLIVIOUS POLYNOMIAL EVALUATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

MERT ÖZARAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

SEPTEMBER 2012

Approval of the thesis:

SECURE MULTIPARTY COMPUTATION VIA OBLIVIOUS POLYNOMIAL EVALUATION

submitted by **MERT ÖZARAR** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Adnan Yazıcı _____
Head of Department, **Computer Engineering**

Prof. Dr. F. Payidar Genç _____
Supervisor, **Computer Engineering Dept., METU**

Dr. Attila Özgit _____
Co-Supervisor, **Computer Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Ersan Akyıldız _____
Mathematics Dept., METU

Prof. Dr. F. Payidar Genç _____
Computer Engineering Dept., METU

Prof. Dr. M. Ufuk Çağlayan _____
Computer Engineering Dept., Boğaziçi University

Prof. Dr. Hakkı Toroslu _____
Computer Engineering Dept., METU

Prof. Dr. G. Wilhelm Weber _____
Institute of Applied Mathematics, METU

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Mert ÖZARAR

Signature:

ABSTRACT

SECURE MULTIPARTY COMPUTATION VIA OBLIVIOUS POLYNOMIAL EVALUATION

Özarar, Mert

Ph.D., Department of Computer Engineering

Supervisor: Prof. Dr. Fethi Payidar Genç

Co-Supervisor: Dr. Attila Özgit

September 2012, 95 pages

The number of opportunities for cooperative computation has exponentially been increasing with growing interaction via Internet technologies. These computations could occur between trusted partners, between partially trusted partners, or even between competitors. Most of the time, the communicating parties may not want to disclose their private data to the other principal while taking the advantage of collaboration, hence concentrating on the results rather than private and perhaps useless data values. For performing such computations, one party must know inputs from all the participants; however if none of the parties can be trusted enough to know all the inputs, privacy will become a primary concern. Hence the techniques for Secure Multiparty Computation (SMC) are quite relevant and practical to overcome such kind of privacy gaps. The subject of SMC has evolved from earlier solutions of combinational logic circuits to the recent proposals of anonymity-enabled computation. In this thesis, we put together the significant research that has been carried out on SMC. We demonstrate the concept by concentrating on a

specific technique called Oblivious Polynomial Evaluation (OPE) together with concrete examples. We put critical issues, challenges and the level of adaptation achieved before the researchers. We also provide some future research opportunities based on the literature survey.

Keywords: Privacy Preservation, Secure Multiparty Computation, Oblivious Polynomial Evaluation

ÖZ

İLGİSİZ POLİNOM DEĞERLEMESİ ÜZERİNDEN
GÜVENLİ ÇOK PARTİLİ HESAPLAMA

Özarar, Mert

Doktora, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Fethi Payidar Genç

Tez Ortak Yöneticisi: Dr. Attila Özgit

Eylül 2012, 95 sayfa

Gelişen Internet teknolojileriyle beraber birlikte hesaplama yapma fırsatları üstel olarak artmaktadır. Bu tür hesaplamalar güvenilir, kısmen güvenilir ya da rekabetçi taraflar arasında cereyan edebilir. Çoğu zaman iletişim halindeki taraflar mahrem verilerini açığa çıkarmak istemezler lakin beraber çalışmanın getirdiği avantajlardan faydalanarak özel ve belki gereksiz veri değerlerinden ziyade sonuçlara odaklanırlar. Bu tür hesaplamaları gerçekleştirmek için bir taraf katılımcılardan gelen tüm girdileri bilmelidir. Bununla birlikte eğer hiçbir taraf yeterince güvenilir değilse mahremiyet birincil öncelik haline gelecektir. Bundan mütevellit Güvenilir Çok Partili Hesaplama (GÇH) adını verdiğimiz teknikler bu konu ile alakalı olup bahsedilen tipteki mahremiyet açıklarının üstesinden gelmede pratik yollar açarlar. GÇH konusu ilkel çözümleri olan kombinatorik mantık devrelerinden başlayarak günümüzdeki anonimi sağlayan hesaplama yöntemlerine kadar evrilmiştir. Bu tezde GÇH hakkında derin ve anlamlı çalışmaları biraraya getireceğiz. Özel bir teknik olan İlgisiz Polinom Değerlemesi'ne konsantre olup konuyu somut örneklerle göstereceğiz. Daha evvelki çalışmalara kıyasla kritik hususları, meydan okumaları

ve adaptasyon seviyelerini ortaya koyacağız. Literatür taramasına binaen bazı gelecekteki araştırma fırsatlarına da değiniriz.

Anahtar Sözcükler: Mahremiyet Koruma, Güvenli Çok Partili Hesaplama, İlgisiz Polinom Değerlemesi

To all “Özarar” family members; both in the past and in the future, with love

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to my supervisor Prof. Dr. F. Payidar Genç and my co-supervisor Dr. Attila Özgit for their guidance and insight throughout this thesis study. The invaluable comments of my thesis committee members, namely Prof. Dr. Ersan Akyıldız and Prof. Dr. İsmail Hakkı Toroslu, surely helped me while preparing the final version. I would like to express my debt of regards to final jury members Prof. Dr. M. Ufuk Çağlayan and Prof. Dr. Gerhard-Wilhelm Weber.

I sincerely acknowledge the support and understanding of my family members, namely Serpil and Tunçer Özarar. They have supported my education period for twenty six years. I believe that I have honoured them with five consecutive higher education degrees obtained from simply the best technical university in Turkey.

My special thanks to everyone I met in my department. Each of them motivates me during this long journey.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

SMC	Secure Multiparty Computation
OPE	Oblivious Polynomial Evaluation
TTP	Trusted Third Party
DDH	Decisional Diffie Hellman
OMP	Overall Mean Problem
DPP	Dot Product Problem
BTS	Bilinear Terms Summation
HAC	Hierarchical Agglomerative Clustering
HDP	Hamming Distance Problem
PKI	Public Key Infrastructure

CHAPTER 1

INTRODUCTION

The serenity and tranquility of qualified information flow on the Web has expanded concerns of private protection. Network perusing, emails, and different utilities unvaryingly hole informative data about who we are and what we mind about. It's all in all plain that some protection may be lost in trade for the benefits of such informative data innovation based aids. However, in different spaces security is so significant that its insurance is commanded.

Engineerings for acquiring protection are improving in response to the aforementioned developing concerns. More accentuation has been set on protecting the security of user information accumulations, in well known eld. Access to these parties is, however, absolutely supportive. It is from this balance between privacy and utility that privacy-preserving data mining emerged.

Occasionally, the communicating parties might not desire to unveil their protected data to the other principal while taking the advantage of coaction, thus concentrating on the consequences rather than private and occasionally, the put acrossing political parties might not desire to unveil their protected data to the other principal while taking the advantage of coaction, thus concentrating on the consequences rather than private. The straightforward solution is one party must cognize input signals from all the participants however if none of the parties can be relied enough to cognize all the input signals, the straightforward solution is one political party must cognize inputs from all the participants however if none of the parties can be relied enough to cognize all the input signals.. Hence the techniques for Secure Multiparty Computation are quite crucial and practical to overcome the privacy gaps.

1.1. Secure Multiparty Computation

Depending on if numerous parties prefer to perform a calculation dependent upon their private inputs, in any case none, of these party is ready to uncover its particular include to anyone else, then the fundamental situation is the manner by which to lead quite an impressive reckoning while protecting the protection of the inputs. This is pointed to as Secure Multiparty Processing situation (SMC) in the writing.

For example, consider the following real life scenarios where Secure Multiparty Computation can directly be applicable;

1. Some hospitals situated in various different countries having their medical databases and patient's history stored on some remote database sites. If an insurance company wishes to verify the med claim of a particular person, he can get that patient's information from hospital's database, but the hospital's database is not completely provided, instead only the requested information is allowed to access.
2. In a given written examination, the results are privately shared with the students. No student wants to disclose its exam grade yet all of them want to calculate the average of the exam.
3. Let us assume that an international airline company that has a reservation database for each country exists. If a person wishes to make a reservation from city A located in country X to a city B located in country Y, then we need to consult each intermediate countries databases. These databases provide only the queried details without disclosing their whole reservation database.

In general, a SMC issue bargains with processing any probabilistic capacity on any data, in a dispersed system where every member keeps one of the inputs,

guaranteeing freedom of the inputs, effectiveness of the calculation, and that no more informative data is uncovered to a member in the calculation different than that might be gathered from the member's enter and yield (Du, 2002).

At present, to deal with the above scenarios, a consistently framework is to gather the trustworthiness of the utility suppliers, or to set the being of a TTP, which is dangerous in nowadays' powerful and malicious cyber world. Acknowledge a trusted party who gathers all members' information then afterward performs the processing and sends the outcomes to the members. Without having a relied party, some imparting near the members is absolutely needed; yet we do not cognize how to guarantee that this imparting does not unveil any of the above. Subsequently, methodologies that can once again down joint estimations while ensuring the members' protection are of developing criticalness.

In speculation, the general SMC is resolvable (Yao, 1986; Goldreich, 1987; Goldreich, 2004) yet utilizing the fixes inferred by these general effects for exceptional instances of multiparty reckoning could be unfeasible; uncommon fixes may as well be improved for proficiency explanations. A thought is to permit non-determinism in the accurate qualities sent for the middle correspondence and show that a party with simply its particular data and the consequence can create an anticipated transitional reckoning that is as imaginable as the true qualities (Özarar, 2007).

1.2. Privacy

The burgeoning of qualified data and conveyance mechanics in the final few decades has raised the concern over security as every expanding degree common issue. The growth of building has carried major updates to the plans of data airing and the infrastructure of engineering has carried major updates to the plans of informative content airing. In the meantime, the expanding limit for data openness determined by innovative advancements has accumulated huge chances terms of misuse of security.

When contending about protection, researchers and theorists routinely recognize that security exemplifies several interweaved significances, in particular physical protection (spatial disconnection and singularity), educational security (confidentiality, mystery, information insurance, and control over private informative content), and decisional protection (constrained interruption into determination making about sex, families, religion, and health awareness), as proclaimed in the expositive expression. Segregation, singularity, mystery, confidentiality, and secrecy are recognized necessities for a liberal being.

The advancement of the proposed several sorts of security contrast. The level of physical protection – as a byproduct of expanded fortune – in current improved public orders is remarkably towering by authentic models. Besides, current human rights and protection enactment help to concede the common emancipations on the rights to settle on determinations. Yet, the scenario observing enlightening security is worrisome. In the rest of the thesis, the term privacy and informational privacy are used interchangeably.

1.3. Model Paradigms

A significant number of models have been suggested in the written works for the research project and dissection of Secure Multiparty Computation situations. Near them, two model standards are notorious:

1. Ideal Model Paradigm
2. Real Model Paradigm

In a straightforward demonstration, a third party, whom we posit that it is trusted, performs processings. The parties send their information in secure mode to such Trusted Third Party (TTP). There are various orders suggested via analysts for this model. In this model, in the event that some party carries on noxiously, the consequence of the processing may be mistaken in light of the fact that the party may supply invalid enter to the TTP but the security might be safeguarded.

The perfect model of SMC is not leaned toward because of the expense of going with the TTP. Just another inconvenience of this design is that the adherence of the TTP is noteworthy. When Trusted Third Party turns into an undermined one, the entire thought of the SMC comes to be worthless. In any case, today this model is often utilized because of straightforward implementation and utilize of instruments that avoid the TTP from ending up being malignant.

Moreover, in actual model there is no Trusted Third Party for reckoning. Collaborating parties in this model coincide on some order which is to be run near them for security preserving and processing of right consequence. Parties do not share exact inputs to one another. The values sent by parties are some limit of their private informative data. What exists between parties is a speculative handling

machine. The real model of SMC is stated to be secure if a foe can satisfy some strike, which is moreover conceivable in the perfect structure of SMC.

We fixate on the legit model ideal model. In this model, the sort and action of antagonists may as well be painstakingly researched.

1.4. Adversarial Behavior

Privacy protecting algorithms are composed keeping in mind the end goal to protect security even in the presence of rivaling members that try to assemble informative data regarding the inputs of their companions. There are, nonetheless, better levels of ill-disposed conduct. Cryptographic research normally thinks about two sorts of foes: A semi-honest antagonist (in addition regarded as an inactive, or honest but inquisitive antagonist) is a party that accurately takes after the order particular, yet endeavors to memorize extra informative data by investigating the memos appropriated in the midst of the order execution.

Furthermore, a malevolent foe (engaged) may self-assertively change from the methodology detail. Case in point, recognize a step in the order where one of the parties is needed to decide on an erratic number and telecast it. In the event that the party is semi-honest then we can posit that this number is for sure irregular. Further, if the party is vindictive, then he may pick the number in a complex way that prepares him to increase more informative content.

It is obviously simpler to plan a result that is secured in opposition to semi-legit rivals, than for noxious contestants. A regular way is hence, first to plan a secure methodology for the semi-honest case, then after that transform it into an order that is secure in opposition to malevolent opponents. This transformation might be finished by needing every party to utilize zero-learning confirmations to advocate that every

step that it is taking takes after the determination of the order. More effective transformations are regularly needed, after this nonexclusive methodology may be fairly wasteful and add significant overhead to every step of the methodology. It's exceptional that the semi-legitimate antagonistic model is regularly a reasonable one. This is resulting from the fact that going astray from a specified order which may be buried in a perplexing requisition is a non-insignificant work, and resulting from the fact that a semi-trustworthy rivaling conduct can model a situation in which the parties that cooperate in the methodology are legit, anyway accompanying the order execution an antagonist may acquire a transcript of the order execution by breaking into a machine utilized by one of the members (Özarar, 2007).

Right away, it gives the idea that any request that is secure in the presence of malevolent foes is furthermore secure in the presence of semi-honest antagonists. This is on account of a semi-honest adversary is simply an exceptional instance of a malicious adversary who steadfastly accompanies the order specification. Granted that this is what we could anticipate, it creates be false. This peculiarity is on account of the way that although a certifiable semi-honest opponent is likely an outstanding case of a certified noxious foe, this is not revised of the unattached rivals in the ideal model. Specifically, the foe in the ideal model for malignant contestants is allowed to redesign its drop in, however the antagonist in the perfect model for semi-legit rivals is not. Along these lines, the adversary/simulator for the occasion of vindictive rivals has more power than the adversary/simulator for the occasion of semi-honest enemies.

In that capacity, it might be plausible to recreate a methodology in the vindictive model, however not in the semi-trustworthy case. We consider the semi-honest case throughout this chapter.

1.5. Scope of the Study

As demonstrated by the past scenarios, severe security threats in joined network procurements exist. This scenario could not be understood by passing on secure channels or keeping security-delicate data of the users encrypted on the server side. While the proposed undertakings to build security shed diverse security threats from outside assaulters, they are not insufficient to secure the unstable qualified information in restriction to misuse by the support supplier which makes the most immense potential risk.

In this dissertation, we center on principled explanations to secure the protection of users in some SMC requisitions utilizing the method called Careless Polynomial Assessment (OPE). For this reason we suggest to keep the Privacy-delicate information sheltered by method of encryption around the same time as handling. This methodology takes out the danger of conceivable protection misuses as the delicate information is just good to go to the possessor but not to the different parties. In any case, once encrypted, the structure in information is obliterated as a result of the encryption methodology. To handle encrypted information, we thoroughly research the cryptographic instruments grouped OPE en masse with some multi party computation systems.

To delineate the thought of the combination of SMC and cryptography, we have chosen prototypical requisitions. Specifically, we center on generally speaking mean calculation, network arithmetical situations and Hamming Distance estimation. The proposed requisitions are chosen as they comprise of normal indicator transforming operations for example separating, duplicating, separation reckoning, thresholding and finding most conservative levels which might be perceived in different provisions simultaneously. To apprehend privacy-protecting form of provisions, such operations might as well be apprehended in the encrypted space powerfully. To actualize this object, we have delivered a speech to the emulating challenges:

• Data representation,

• Apprehending linear and non-linear operations in the encrypted domain,

• Information hiding because of encryption,

• Correspondence and reckoning expenses of utilizing cryptographic methodologies.

1.6. Organization of the Thesis

This thesis is arranged to blanket all angles of the chose prototypical requisitions. To have a clear perspective on the ready cryptographic devices and existing explanations that location comparable issue, we begin with a review part. The diagram is accompanied by various sections each of which fixes all available attention on one specific requisition and shows a complete fix.

In Chapter 2, the background information on the SMC and OPE are discussed. In Chapter 3, the proposed methodology for the definition of SMC and its properties is presented. The building blocks of the methodology, relations of these methods and the supporting structures form the content of this chapter. The heart of the thesis is chapter 4 where SMC via OPE is explained equipped with concrete examples. In Chapter 5, privacy preserving Hierarchical Agglomerative Clustering is explained to demonstrate a concrete application of the underlying concept. In chapter 6, we present a real life computer security problematic case scenario and offer a solution to overcome it via OPE. We finalize the thesis with a conclusion section that summarizes what has been obtained and what type of difficulties need for further analysis.

In the appendix section, the concrete protocol for OPE together with its evaluation is presented.

CHAPTER 2

BACKGROUND

In this chapter, the background information for the study is presented, mainly concentrated on secure multiparty computation together with its application to target domains. Obviously, the foundations of Oblivious Polynomial Evaluation are discussed thoroughly.

2.1. Background on SMC

There has been an unyielding work in helpful calculation between substances that commonly doubt each other. This processing may be of any sort: scientific, information transforming or even mystery imparting. The first paper on SMC belongs to Yao who has stated the famous problem called “Millionaire’s Problem” (Yao, 1986). By what means can two tycoons know who is richer without revealing their single fortune to one another? Yao has utilized cryptographic methods to take care of the situation. Goldreich et. al. enlarged the thought from two party case to multiparty case (Goldreich, 1987). These works are straightforwardly identified with the field called “Secret Sharing” which is close to SMC (Shamir, 1979).

The first solutions to SMC problems used combinatorial circuit in which each party cooperatively runs a dedicated protocol for every gate in the circuit. Each member get (irregularly picked) imparts of the qualities of the info and yield for every entryway; the elite or of the portions is the exact esteem. One give separated from a

lot of people else moves no data concerning the qualified information or system regard, as which party gets which confer is determined heedlessly. At the completion, the parties exchange their portions, approving each to figure the irreversible outcome. This methodology has been exhibited to give the desired impact without revealing whatever than the result. This strategy, in any case participating in its smoothness and accord, recommends that the compass of the methodology relies on the compass of the circuit, which relies on the size of the data. Different general methods have been recommended by Chaum et. al. (Chaum, 1988) and Naor et. al. (Naor, 2001). While this way is engaging in its consensus and straightforwardness, the produced methodologies rely on the size of the circuit. This size hinges on the measure of the information realm, and on the many-sided quality of communicating quite an impressive processing.

Du and Atallah complete a magnificent study of secure multiparty reckoning situations and their provisions (Du, 2001). They demarcated different SMC situations for their particular processings for example Privacy-PreservingInformation mining, Privacy-Preserving Interruption Location and Privacy-Preserving Geometric Calculation. Also, Du and Zhan recommend a useful methodology to applying secure multiparty reckoning by receiving a trade off on security (Du, 2002). Verykios et. al. put forth a diagram of the newfangled and quickly developing research zone of security Preservinginformation mining, in addition group the procedures, audit and assessment of protection preserving algorithms (Verykios, 2004). Clifton et. al. put forth certain devices and demonstrate how they could be utilized to illuminate some security preserving information mining situations (Clifton, 2004). Agrawal and Srikant suggest a novel recreation system to faultlessly assess the dispersion of initial information esteem by the proposed reproduced conveyances (Agrawal, 2000). Pinkas centers to show essential thoughts from a substantial assortment of cryptographic research on secure circulated calculation and their requisitions to information mining (Pinkas, 2003). Vaidya et. al. study ways to secure multiparty

processing and give a system where by a powerful order incorporating two parties (Vaidya, 2003).

2.2. Specific Notable SMC Problems

A significant number of particular SMC situations and theirs results are devised by the scientists. The notable SMC problems can be presented together with their solutions:

2.2.1. Privacy-Preserving Decision Trees

Lindell et. al. centered on the issue of "Decision Tree picking up with the ID3 functional process" and methodology is all in all effective (Lindell, 2000). Vaidya et. al. likewise tackle the same situation of arrangement and they present a summed up “privacy preserving variant of the ID3” functional process for vertically divided information circulated over two or more parties (Vaidya, 2003). Tooth et. al. put forth a novel “Privacy-Preserving Decision Tree” studying system (Tooth, 2008). Emekci et. al. in addition centers on the same situation and advance a “privacy-preserving ID3” equation simultaneously (Emekci, 2007). Agrawal and Srikant define the situation as how one party could be permitted to perform information mining operation on the private database of a different party without the first party knowing any portions of the database of the second party (Agrawal, 2000). Distinctive from past ones, they utilize information bother system to take care of the situation. Dowd et. al. display an information bother procedure dependent upon erratic substitutions and demonstrated that the coming about security-protecting determination tree mining system is safe to assaults that are evidently applicable (Dowd, 2006). Efficient examinations indicate that it is additionally adequate.

2.2.2. Privacy-Preserving Naïve Bayes Classification

Vaidya and Clifton solve the same problem on vertically partitioned data (Vaidya, 2003). Kantarcioglu et. al. exhibit orders to improve a Naive Bayes classifier on a level plane divided information (Kantarcioglu, 2004). Vaidya et. al. compare the results of the both horizontal and vertical case and combined a solution for hybrid model as well (Vaidya, 2004).

2.2.3. Private Information Retrieval

"Private Information Retrieval (PIR)" is an order that permits a customer to recover a component of a database without the manager of that database having the capacity to confirm which component was chosen. While this situation confirms a trifling result-sending the whole database to the client permits the customer to question with flawless Privacy-there are procedures to diminish the correspondence many-sided quality of this situation, which might be basic for great databases. Moreover, "Strong Private Informative content Recovery (SPIR)" is private informative content recovery with the extra prerequisite that the customer just pick up concerning the components he is questioning for, and nothing else. This prerequisite catches the average security necessities of a database manager. This situation was presented by Chor et. al. (Chor, 1995). The situation was augmented by Gertner et. al. (Gertner, 1998).Many keys are suggested for the PIR situation centering on lessening the conveyance cost (Chor, 1997; Kushilevitz 1997; Di-Crescenzo, 1998; Cachin 1999; Ishai 1999).

2.2.4. Privacy-Preserving Clustering

In the recent past privacy-preserving information mining has been a truly dynamic territory of examination. Beginning center around there was on project of choice trees from conveyed information sets (Agrawal, 2000; Lindell, 2000). There is moreover a strong collection of exploration on Privacy-protecting mining of companionship leads (Evfimievski 2002; Rizvi 2002). As a rule, there are two ways for planning "privacy-preserving" machine learning ordered systems. The first way is

to utilize transformations to bother the information set soon after the equation is connected. This methodology for outlining "privacy-preserving" bunching algorithms is taken by some scientists (Klusch, 2003; Merugu, 2003; Oliviera, 2003). The second methodology to objective security protecting contrivances is to utilize ordered systems from the secure-multiparty reckoning literary works. The preference of this methodology over the bother way is that formal assurances of security might be given for the proposed ordered systems. Vaidya and Clifton introduce a "privacy-preserving" k-means contrivance for vertically-apportioned information sets (Vaidya, 2003). There are dispersed grouping equations where the object is to decrease correspondence expenses (Dhillon, 1999; Kargupta, 2001; Jha, 2005).

2.2.5. Privacy-Preserving Statistical Analysis

Franconi and Merola give an earlier review on the subject, with a center on collected information discharged by means of net access (Franconi, 2003). Evfimievski et. al. give an absolutely fantastic idea exchange of work in randomization of information, in which information benefactors autonomously add clamor to their particular reactions (Evfimievski, 2003). A large number of studies in the statistics written works bargain with creating reproduced information while upholding certain amounts, for example marginals. Different extensively-pondered systems incorporate cell suppression, including re-enacted information, discharging just a subset of recognitions, discharging just a subset of characteristics, discharging synthetic or incompletely synthetic information-swapping, and post-randomization. Agrawal and Srikant start to location protection in information mining (Agrawal, 2000). That work tries to formalize security in terms of certainty interims and additionally demonstrates to reproduce an initial circulation from uproarious tests.

2.3. Background on OPE

The usage of OPE in privacy preserving applications is presented by Naor by the introduction of "oblivious transfer" (Naor, 1999). "Oblivious transfer" is a

fundamental methodology that is the principle fabricating square of secure calculation. It may appear abnormal in the first place; anyway its part in secure processing might as well end up being clear later. It was indicated by again Naor that absent transfer is sufficient for secure calculation in the sense that given an implementation of oblivious transfer, and no different cryptographic primitive, one may build any secure calculation methodology (Naor, 2005).

OPE is applied to SMC problems especially under privacy preserving data mining concept which is introduced by designing privacy preserving ID3 decision tree algorithm (Lindell, 2000). Jha et. al. solve weighted average problem by two techniques (Jha, 2005). Former is by Oblivious Polynomial Evaluation and latter is by encryption techniques based on homomorphism yet both of them is used as a tool for k-means clustering for two parties. Neural grid requisitions are concentrated on utilizing OPE techniques also within the setting of security-protecting information mining (Chang, 2001), Goethals et. al. display a private scalar item order dependent upon standard cryptographic procedures and authenticated that it's secure (Goethals, 2004).

Özarar and Özgit pay special attention to the subject and do extensive research on it. They solve the secure multiparty mean computation problem using OPE (Özarar, 2007). Moreover, they develop algorithms for matrix algebraic concepts like eigenvalue, eigenvector and determinant computations using Oblivious Polynomial Evaluation in horizontally partitioned data (Özarar, 2008). The presentation of a novel protocol for secure multiparty Hamming Distance algorithm exists which is designed to be used as a building block for Hierarchical Agglomerative Clustering (HAC) of documents in Özarar et. al (Özarar, 2011).

CHAPTER 3

METHODS TO SOLVE SMC PROBLEMS

In this chapter, we present the known methods which are designed to solve SMC problems from different viewpoints. In the past, there exist three types of methods for SMC problems:

1. Randomization Methods
2. Cryptographic Techniques
3. Anonymization Methods.

3.1. Randomization Methods

The randomization procedure furnishes a viable yet effortless way of staying away from the single from thinking about sensitive informative content, which could be effectively connected at informative data determination stage for solace securing qualified data investigation, being as how the unsettling influence incorporated to a given history is partitioned of the movements of different informative data records. In the randomization strategy, occasions utilize interesting figures for disguising their informative data and execute computations over undetectable informative content. Systems are made such that the conclusions of the counts over undetectable informative data are the same as the conclusions of estimations over honest qualified data.

Information randomization system acts for one normal way to tackle some sort of SMC situations where the initial (private) dataset is irritated and the effect is discharged for information dissection. Information bother incorporates a broad assortment of strategies incorporating: added substance, multiplicative (Kim, 2003), framework multiplicative, k-anonymization (Sweeney, 2002) and micro-accumulation (Li, 2006).

The added substance bother is a system for privacy-preserving information mining in which tumult is added to the information keeping in mind the end goal to cover the quality qualities of records (Agrawal, 2000). The commotion included is sufficiently vast so single record qualities can't be recuperated. Thusly, procedures are outlined to determine total disseminations from the annoyed information.

The randomization technique has been augmented to an assortment of information mining situations. Agrawal and Srikant exchange ideas about how to utilize the way for taking care of the protection protecting characterization situation (Agrawal, 2000). Furthermore, Evfimievski et. al. suggest a key to the security preserving appropriated acquaintanceship mining situation (Evfimievski, 2002). The situation of companionship leads is particularly challenging in view of the discrete nature of the ascribes relating to presence or unlucky deficiency of articles. Keeping in mind the end goal to bargain with this issue, the randomization strategy ought to be altered to a limited extent. Rather than including quantitative commotion, haphazard articles are dropped or incorporated with a certain possibility. The bothered transactions are then utilized for total affiliation lead mining. The randomization methodology has additionally been expanded to different provisions; case in point, peculiar disintegration based community oriented sifting (Polat, 2005).

3.2. Cryptographic Techniques

The cryptographic techniques solutions to SMC problems include some basic building blocks which are used as components while handling secure computations. Some of the building blocks are as follows (Oleschchuk, 2007):

3.2.1 The Millionaires Problem

It is a SMC situation which was presented by (Yao, 1986). The situation examines two moguls, who are intrigued by knowing which of them is richer without disclosing their true fortune. This situation is similar to a more general situation where there are two numbers x and y and the objective is to settle the bias "y > x" without disclosing the real qualities of x and y.

Countless answers have been presented for the situation, near which the first explanation improved by utilizing symmetric cryptography, put forth by Yao himself, was exponential in time and space (Yao, 1986).

This part for SMC is of service in provisions for example connected offering and barters. A significant number of scientists suggest keys to this situation (Cachin, 1999; Ioannidis, 2003; Amirbekyan, 2009). The same situation could be augmented to multiparty case and is handy for the SMC explanation. Subsequently result to this situation can fill in as the manufacturing square for a considerable number of SMC situations

3.2.2. Homomorphic Encryption

The SMC problem becomes more complex when asking for the possibility to compute (publicly) with encrypted data or to modify functions in such a way that

they are still executable while privacy is ensured. That is where homomorphic cryptosystems can be used.

A homomorphic encryption plan is an encryption plan which permits certain logarithmic operations to be completed on the encrypted plaintext, by applying an effective operation to the comparing figure message. For a cement occurrence, an additively "homomorphic encryption" plan (Paillier ,1999) that is similar with the encryption prepare of RSA in terms of the processing cost, while the decryption transform of the added substance homomorphism is faster than the decryption handle of RSA might be exhibited.

An additively homomorphic cryptosystem has the superb property that for two plain quick message m_1 and m_2:

$$E(m_1) + E(m_2) = E(m_1 + m_2)$$

This basically denotes that we can have the whole of two numbers without recognizing what the aforementioned numbers are. Besides, due to the property of associability,

$$E(m_1) + E(m_2) + \ldots + E(m_n) = E(m_1 + m_2 + \ldots + m_n)$$

and we may quickly deduce:

$$E(m_1)^{m}{}_2 = E(m_2)^{m}{}_1 = E(m_1 \,.\, m_2) \text{ where } E(m_i) \neq 0.$$

Many homomorphic systems with semantic security are given in the past (Benaloh, 1994; Naccache, 1998; Paillier, 1999). There are a lot of important works related to usage of homomorphic encryption as a tool for SMC (Lindell, 2002; Du, 2002; Vaidya, 2002).

3.2.3. Oblivious Transfer and Oblivious Polynomial Evaluation

The oblivious transfer order includes two parties, the sender and the receiver. The sender's include is a couple (x_0, x_1) and the recipient's include is a bit $\sigma \in \{0, 1\}$. At the close of methodology, the collector gathers x_σ (and nothing else) and the sender memorizes nothing. Oblivious transfer is frequently the most computationally concentrated operation of secure methodologies and is rehashed a significant number of times. Every conjuring of neglectful transfer commonly needs an unvarying number of summonses of trapdoor changes. It is plausible to decrease the amortized overhead of unaware transfer to one exponentiations for each a logarithmic number of neglectful transfers, all the more for the instance of malignant foes (Pinkas, 2003).

Oblivious polynomial assessment is a procedure dependent upon oblivious transfer. To plan a secure order for registering a method **f(x,y)** permits two parties, a receiver who knows x and a sender who knows y, to together register the quality of **f(x,y)** in a protection protecting way. The way that for each processable method **f(x,y)** in polynomial time, there exists a (polynomially-computable) methodology that is recently accomplished in the cryptographic research (Naor, 2005). In the OPE, the data of the sender is a polynomial **P** of degree **k** over some field **F**. The receiver can get **P(x)** for any component **x Є F** without memorizing all else regarding the polynomial P and without disclosing to the sender any qualified data about **x**. The information and yield for the usefulness of OPE as a two party order run between a receiver and a sender over a field **F** as takes after:

- Input
  - Receiver: a field element **x Є F**.
  - Sender: A polynomial **P** defined over **F**.

- Output
  - Receiver: **P(x)**.
  - Sender: nothing.

There are various protocols to solve the OPE yet the protocol given by Naor et. al. is preferred for the target algorithms (Naor, 2005).

The literature survey on OPE is given in the previous section.

3.2.4. Private Matching

Agrawal et. al. display a paper that investigates the accompanying "private matching" situation: two parties each have a database and they wish to figure normal passages without uncovering any qualified information about passages just discovered in one database (Agrawal, 2003). This paper has created significant investment in the exploration neighborhood and systems cal press. While the Agrawal/Evfimievski/Srikant (AgES) methodology portrayed in the paper is right within in its suspicions, it is not strong in a mixed bag of contrasting situations. Actually, in numerous feasible situations, the AgES methodology can effectively be misused to acquire a vital bargain of data about an additional database. The "private matching" situation has truly better results hinging on suppositions regarding the diverse parties, the way they cooperate, and cryptographic mechanisms good to go.

The methodologies utilized for private matching utilize the lands of homomorphic encryption. Contrasting methodologies are good to go for semi legit parties and that for malignant parties (Freedman, 2004).

3.3. Anonymization Methods

Keeping in mind the end goal to protect security, Sweeney et. al. recommend the "k-anonymity" model which accomplishes anonymity utilizing generalization and suppression, so that, any distinct is undefined from at slightest k-1 different ones as for the semi-identifier traits in the "anonymized dataset" (Sweeney, 2002).

In well known years, various algorithms have been recommended for executing k-anonymity by means of generalization and suppression. Bayardo and Agrawal put forth an optimal functional process that begins from a completely summed up table and practices the dataset in a negligible k-unnamed table (Bayardo, 2005). Lefevre et. al. portray a functional process that utilizes a base-up system and a priori calculation (Lefevre, 2005). Fung et al. display a top-down heuristic to make a table to be discharged k-unnamed (Fung, 2005). As to the speculative effects, Sweeney show the optimal k-anonymity is NP-hard and furnished estimate equations for optimal k-anonymity (Sweeney, 2002). Nonetheless, Machanavajjhala et. al. sharp out that the user may figure the touchy qualities with elevated trust when the touchy information is absence of assorted qualities, and presented the "l-diversity" qualities strategy (Machanavajjhala, 2007).

The k-anonymity methods chiefly fix all available attention on a global procedure that puts the same product of upkeep for all folks, without accommodating their substantial necessities. The effect may be furnishing insufficient security to a part of individuals, while enabling utmost solace administration to an additional part. "k-Anonymous" data investigation is on the other hand a last examination region and a

significant number of situations are still to be examined, for example, the mixture of k-anonymity with different conceivable data investigation procedures; the examination of unique methods for identifying and counteracting k-anonymity offenses. The anonymization strategy can verify that the altered informative content is honest; anyhow it in addition conclusions in qualified information misfortune in some level.

CHAPTER 4

SECURE MULTIPARTY COMPUTATION VIA OBLIVIOUS POLYNOMIAL EVALUATION

In this section, the application of OPE (as a building block) is discussed to handle SMC problems. Two concrete examples together with solution algorithms are presented to demonstrate the usage. Moreover, under given security assumptions like passive adversaries, the privacy validities of the algorithms are justified using statistical indistinguishability and semantic security. The related definitions used throughout this chapter together with intriguing theorems and their proofs can be accessible through Appendix section of the thesis.

Another important concept while developing a secure algorithm is the model of computation. We present a transformation skeleton that efficiently transforms standard processings to secure multiparty processings.

4.1. Statistical Indistinguishability

Before all, the privacy proof concept in SMC and semantic security should be defined. The definition of protection is dependent upon the instinct that parties might as well memorize nothing more from the notes utilized within “privacy-preserving” order, i.e., the memos appropriated by a party around the same time as an execution of a Privacy-protecting methodology could be “conclusively processed” by just knowing its enter and yield. This thought is formalized beneath:

Definition 1 Let x and y be inputs of the two parties and $f_1(x,y)$, $f_2(x,y)$ be the coveted functionalities, i.e., the first party prefers to figure $f_1(x, y)$, and the second prefers to figure $f_2(x,y)$. Let P be a two-party methodology to register f. The view of the first party ultimately having cooperated in methodology P (signified by $VIEW_1(x,y)$) is $(x, r, m_1. . . m_k)$, where r are the irregular bits produced by party 1 and $m_1. . . m_k$ is the arrangement of memos accepted by party 1, while cooperating in methodology P. $VIEW_2(x,y)$, for the second party might be demarcated in a comparative way.

We declare that P privately processes f if there are probabilistic polynomial-time contrivances (PPTA), meant by S_1 and S_2 such that,

$$\{S_1(x, f_1(x,y))\}_{x,y} \equiv s \{VIEW_1(x,y)\}_{x,y}$$

$$\{S_2(x, f_2(x,y))\}_{x,y} \equiv s \{VIEW_2(x,y)\}_{x,y}$$

In the equation given above, ≡s denotes statistically indistinguishable.

Goldreich states that "two probability ensembles $X = \{X_w\}_{w \in S}$ and $Y = \{Y_w\}_{w \in S}$ indexed by S are statistically indistinguishable if for some negligible function μ: ℵ→ [0, 1] and all w Є S,

$$\sum_{\alpha} | Prob(X_w = \alpha) - Prob(Y_w = \alpha) | < \mu(|w|)$$

A function μ: ℵ→ [0, 1] is called negligible if for every positive polynomial q, and all sufficiently large n's, $\mu(n) < q(n)^{-1}$. There is a weaker notion of indistinguishability called computationally indistinguishable."

We utilize statistical lack of definition all through the section, yet every last trace of the effects keep regardless of the possibility that the weaker thought of indistinctness is utilized. Goldreich has given a formal definition of protection and has furnished a strong speculative grounding that explanations to a specific secure multiparty calculation situation might as well base on (Goldreich, 2004).

4.2. Models of Computation

In this section, the flow of events from the initial definition of properties till the end of attestation process is described step by step.

Assume that the data to be ready is an arranged D of information parts. Gave that we can parcel D into two disjoint information sets D_1 and D_2, we will have a distinctive-incorporate processing model. There are a significant number of courses to partition D into two information sets, and each way may development to an odd SMC scenario. We are fixating on two sorts of transformations: homogeneous transformation and heterogeneous transformation.

In the homogeneous transformation, D's information articles are secluded to two sets, meanwhile each and every information article is not cut into two parts. Case in point, if D is a database of patient records, the homogeneous transformation will put a subset of the records into one information set, and the final leftover of the records into an extra information set; then again, each patient's record is not cut into two parts. In unexpected comments, the two made datasets keep up the same set of main events.

In the heterogeneous transformation, every last information article is cut into two parts, with each part taking off to a specific dataset. Taking the same case used above, if each patient record keeps a patient's ID record and medicinal record, the heterogeneous transformation may put all patients' ID records into one information set, and all patients' remedial records into a more information set. In diverse articulations, the two generated information sets oversee different set of main events.

In the next three subsections, we demonstrate the usage of OPE to concrete problems like secure overall mean, matrix algebra and Hamming distance computations, respectively. We believe that the notion and role of OPE can be explained best by those examples.

4.3. Application of OPE to Mean Computation

The beginning validation of a client stage can consequence with its embracing by the host aid, in any case constant design updates are constantly an issue. In this section, an enlargement of the structural engineering for the taking care of this situation is clarified.

In this subsection, the problem definition for privacy preserving two party and multiparty overall mean computation problems are defined mathematically whose algorithms are presented.

4.3.1. Privacy Preserving Two-Party Overall Mean Computation Problem (OMP)

Suppose that Alice (party 1) has n examples and Bob (party 2) has m-n specimens of legit numbers. Every party prefers to get the mean of their examples without disclosing any private qualified data. We are party that recognizing the mean of the union of examples from the two parties is more charming than computing the two examples independently.

Let μ_A represents the mean of Alice's samples and μ_B represents the mean of Bob's, respectively. The means are weighted with respect to their cardinalities and joined together with multiplication and then divided by the total size of samples. Hence the result to be computed is,

$$\mu = (\mu_A \cdot n + \mu_B \cdot (m-n)) / m \qquad \textbf{(1)}$$

Remark that the terms in the first product (μ_A , n) are only known by Alice and (μ_B , m-n) are only known by Bob.

4.3.2. Privacy Preserving Two-Party OMP Algorithm (via OPE)

To develop such an algorithm, a functional should be taken as a target to place the terms in the OPE. Let $\boldsymbol{f}$ be the functional for such a computation, its domain set must be two-dimensional vectors (i.e. mean and cardinality) for both parties and range set must be the same value (overall mean) as a two dimensional vector. Since the cardinalities are multiplied with individual means in the numerator and the total sum of samples exist in the denominator of (1), $\boldsymbol{f}$ is constructed as

$$\boldsymbol{f}\text{:}\ \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$\boldsymbol{f}\,((x,m), (y,n)) = ((x\cdot m+y\cdot n)/ (m+n), (x\cdot m+y\cdot n)/ (m+n)) \qquad \textbf{(2)}$$

The straight forward solution is approximating $\boldsymbol{f}$ by a circuit. However, it is well known that the cost for implementing such a circuit is so inefficient that a new solution should be developed for the specific case (via OPE). We describe the protocol in a top-down fashion. The steps are as follows:

i. Define the private rational polynomial evaluation problem (RPE)
ii. Develop a protocol for RPE using OPE.
iii. Find a suitable case for RPE by placing polynomials and field elements for OMP. (Reduction from private-RPE)

i. RPE Problem

For any finite field F, construct ***f*** as;

$$\boldsymbol{f}\text{:}\ \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$\boldsymbol{f}\ ((P,Q), (\alpha, \beta)) = (P(\alpha)/Q(\beta)\ ,\ P(\alpha)/Q(\beta)) \qquad \textbf{(3)}$$

where P, Q ϵ F[x] (polynomials for party 1) and $\alpha, \beta \epsilon$ F (field elements of party 2).

ii. RPE Protocol

The protocol can be developed via OPE in the following scheme:

(1) Party 1 computes the polynomials P and Q by multiplying with a predetermined field element $\gamma \epsilon$ F.

(2) Party 2 calculates $\gamma P(\alpha)$ and $\gamma Q(\beta)$ by applying OPE twice.

(3) Party 2 calculates $P(\alpha)/Q(\beta)$, by obtaining $\gamma P(\alpha) / \gamma Q(\beta)$, and shares it with party 1.

The first two steps are adapted from Jha et. al. where the aim is to solve Weighted Average Problem which is relatively more trivial than OMP (Jha, 2005). In the below, the reduction from private-RPE to OMP is stated by placing suitable polynomials and field elements.

iii. Reduction from private-RPE

Recall that party 1 has inputs (x,m) and party 2 has inputs (y,n). Since the reduction is from RPE, in the format of (3), party 1 needs to construct polynomials and party 2 needs to choose field elements. The polynomials for party 1 are:

$$P(w) = x+w$$

$$Q(w) = m+w$$

$$R(w) = x.w$$

$$S(w) = m.w$$

Note that all polynomials are linear and coefficients are known by party 1. On the other hand, the field elements of party 2 are:

$$\alpha = y$$

$$\beta = n$$

$$\gamma = -n$$

$$\delta = -y$$

The field elements γ and δ are well-defined since every field is an algebraic group and inverse with respect to the addition exists for y and n.

Let us define the $T(w_1,w_2,w_3,w_4)$ as the linear combination of P, Q, R and S polynomials in the following way:

$$T(w_1, w_2,w_3,w_4) = P(w_1) \cdot Q(w_2) + R(w_3) + S(w_4)$$

T is nothing but a dummy polynomial to handle the bilinear terms. Since multiplication and closed operations in a field hence T is also well-defined. If $(w_1, w_2,w_3,w_4) = (\alpha, \beta, \gamma, \delta)$ variable replacement is done then it yields:

$$T(w_1, w_2,w_3,w_4) = x \cdot m + y \cdot n$$

The numerator of the desired functional is constructed and the denominator is nothing but P(α). Using OPE, the reduction is complete by choosing suitable field elements together with constructing the polynomial, T as a combination of party 1's polynomials. Since the functional T is formed as a linear combination of polynomials, the computational complexity of reduction is in polynomial-time.

4.3.3. Proof of Privacy of the Algorithm

Two lemmas are critical for the proof that ***f*** computes the overall mean privately. The former belongs to Canetti (Canetti, 2000) and the latter belongs to Jha et. al. (Jha, 2005). Their proofs are not given here as they can be reached from original sources.

Lemma 1 *(Composition Theorem for passive adversary)* In the event that **g** is privately reducible to **f** and there exists an order for figuring **f** privately then there is a methodology for registering **g** privately.

Lemma 2 *(Private RPE)* The protocol given as RPE protocol privately computes RPE problem.

Theorem 1 *(Two-party Private OMP)* The protocol formed by ***f*** yields a privacy preserving algorithm for two-party Overall Mean Computation problem.

Proof It is clear that OMP is privately reducible to RPE by choosing the numerator as T and the denominator as Q and there exists a protocol for private-RPE (Lemma 2) then by Lemma 1, given protocol is privately computes two-party Overall Mean Computation problem. It is trivial that reduction is polynomially-computable.

The multiparty case is nothing but extension of two parties to multiple.

4.3.4. Multiparty OMP Protocol

Assume that there are N parties in the computation. The protocol for multiparty OMP can be designed from two-party OMP in the following way:

(1) Parties are ordered from 1 to N in a manner that consecutive parties are involved to two-party OMP computation. This can be done with a common share or coin-tossing into well protocol (Özarar, 2007).

(2) Between party j and j+1, $0<j<N$, the two-party overall mean computation problem protocol works and

$$((\mu_j,\ c_j), (\mu_{j+1},\ c_{j+1})) \rightarrow ((\mu_j \cdot c_j + \mu_{j+1} \cdot c_{j+1})/ (c_j + c_{j+1}), (\mu_j \cdot c_j + \mu_{j+1} \cdot c_{j+1})/ (c_j + c_{j+1}))$$

is computed.

(3) Furthermore, party j and j+1 privately compute their cardinality sum via the functional given in lemma 3. In other words, consecutive parties compute the partial mean and partial size of their samples:

$$((\mu_j,\ c_j), (\mu_{j+1},\ c_{j+1})) \rightarrow (c_j + c_{j+1}, c_j + c_{j+1})$$

(4) The mean and cardinality values are updated for party j+1 with the new values calculated at the closure of the order involved with the previously ordered party. (i.e. party j+1 gets the partial mean and partial sum of samples up to her)

(5) Apply the previous two steps for all consecutive parties. Total computation is linear in size and k-1 times for k parties.

(6) At the end of the computation, the last party gets the overall mean together with total sample size and shares them with remaining parties.

The unimportant gap of the protocol is; party j learns the size of the total previous samples yet it is not give the size of the each individual party. The only exception is for the first party, party 2 gets the size of its sample. This can be overcome by choosing the order in a circular round-robin fashion so the order of consecutive

parties are preserved but only the first party changes. The probability to be the first party is 1/N which is Pareto-optimal for such a scheme.

After demonstrating the usage of OPE in overall mean case, let us analyze one more concrete application.

4.3.5. Two-Party OMP Protocol Numerical Example

Assume that there are two parties Alice and Bob having two data sets and their cardinalities, related Galois Field and blinding factor are as follows:

Alice has (5, 60)

Bob has (15, 20)

F = GF(1024577)

Blinding Factor = 3

Step 1: Alice presents her polynomials in the given form and multiplies them:

$$P(w) = w+5$$

$$Q(w) = w+60$$

$$R(w) = w \cdot 5$$

$$S(w) = w \cdot 60$$

Step 2: Bob chooses his field elements:

$$\alpha = 15$$

$$\beta = 20$$

$$\gamma = -20$$

$$\delta = -15$$

Step 3: Alice forms her functional by putting respective values and gets the numerator:

$$T(w_1, w_2, w_3, w_4) = P(w_1) \cdot Q(w_2) + R(w_3) + S(w_4)$$

$$T(w_1, w_2, w_3, w_4) = (w_1+5)(w_2+60) + 5w_3 + 60w_4$$

$$T(w_1, w_2, w_3, w_4) = w_1w_2 + 60w_1 + 5w_2 + 5w_3 + 60w_4 + 300$$

Step 4: Alice forms her functional by putting respective values and gets the denominator:

$$P(w) = w+5$$

Step 5: Alice blinds her polynomials multiplying with the blinding factor and sends them to Bob:

$$3T(w_1, w_2, w_3, w_4) = 3w_1w_2 + 180w_1 + 15w_2 + 15w_3 + 180w_4 + 900$$

$$3P(w) = 3w+15$$

Step 6: Bob computes $3T(\alpha, \beta, \gamma, \delta)$ and $3P(\alpha)$ by applying OPE twice:

If $(w_1, w_2, w_3, w_4) = (\alpha, \beta, \gamma, \delta)$ variable replacement is done then it yields:

$$3T(15, 20, -20, -15) = 3(300 + 900 + 100 – 100 – 900 + 300)$$

$$3T(15, 20, -20, -15) = 1800$$

If $w = \alpha$ variable replacement is done then it yields:

$$3P(15) = 3(15+5) = 60$$

Step 7: Bob gets the result and shares with Alice:

$$3T(15, 20, -20, -15) / 3P(15) = 1800 / 60 = 30$$

4.4. Application of OPE to Secure Matrix Algebra

In this subsection, we examine how diverse matrix algebra situations might be explained in an agreeable nature, where parties should tackle a computational situation dependent upon their joint information, anyhow not, one or the other prefers to reveal its private information to the different party. Some of the target situations in this schema are as takes after:

Problem 1 (*Dot Product*) Alice has a vector (x_1, ..., x_n), and Bob has also another vector (y_1, ..., y_n). They prefer to calculate the dot product $z = x_1.y_1 + \ldots + x_n.y_n$, without revealing each other's vector.

For the remaining problems, we have the following assumption:

Assumption (N-Party *Homogenous Cooperation*) There exist *N* parties each having a matrix M_i. The size of M_i is $m_i x k$ and the total sum of m_i*'s* is equal to *k.* Let *K* be a square matrix with dimension *kxk.* Construct *K* as a horizontal concatenation:

$$K_{k\times k} = \begin{pmatrix} [M_1] \\ [0] \end{pmatrix} + \ldots \begin{pmatrix} [0] \\ [M_i] \\ [0] \end{pmatrix} + \ldots \begin{pmatrix} [0] \\ [M_N] \end{pmatrix}$$

Problem 2 (*Determinant*) All joint parties want to compute the determinant of the common matrix K. Notice that K is square and the determinant is well-defined.

Problem 3 (*Trace*) All joint parties want to compute the trace of the common matrix K to the corresponding diagonal elements found in the previous problem.

Problem 4 (*Eigenvalue*) All joint parties want to compute the eigenvalue of the common matrix K without sharing their secret values to the other principals. Since characteristic polynomial for calculating the eigenvalues is a special kind of determinant computation, the algorithm for problem 2 can be suitably applicable.

4.4.1. Privacy Preserving Dot Product Problem (DPP)

In this subsection, a protocol for privacy preserving two-party Dot Product Problem (DPP) is designed. A function should be taken as a target to place the terms in the OPE for developing such an algorithm.

First we analyze the two-party case then extend it to the multi-party case. Let f be the function for such a computation, its domain set must be n-dimensional vectors for both parties and range set must be a singleton numeric value (dot product). f can be constructed as:

$$f: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$f((x_1, ..., x_n), (y_1, ..., y_n)) = x_1.y_1 + \ldots + x_n.y_n \qquad \textbf{(2)}$$

The straight forward solution is approximating f by a circuit for the privacy preserving case. Yet it is well known that the cost for implementing such a circuit is so inefficient that a new solution should be developed for the specific case. We describe the protocol in a top-down fashion.

Steps are as follows:

i. Define the private Bilinear Terms Summation problem (BTS)

ii. Produce a protocol for BTS using OPE.

iii. Find a suitable case for BTS by placing polynomials and field elements for DPP. (Reduction from private-BTS → DPP)

BTS: For any finite field *F*, construct *f* as

$$f:(F^n[x] \times F) \rightarrow F \times F$$

$$f(P, \alpha) = (P(\alpha), P(\alpha))$$

where $P \in F^n[x]$ (polynomial for party 1) and $\alpha \in F$ (field element of party 2).

Protocol BTS: The construction of BTS via OPE is as follows:

(1) Party 1 "blinds" the polynomials P by multiplying with a pre-determined field element $\gamma_1 \in F$.

(2) Party 2 computes $\gamma_1 P(\alpha)$ by OPE.

(3) Party 2 blinds by multiplying with $\gamma_2 \in F$ and sends the product $\gamma_1 P(\alpha)\gamma_2$ to party 1.

(4) Party 1 cancels out γ_1 by dividing from the product and sends $\gamma_2 P(\alpha)$ to party 2.

(5) Party 2 computes $P(\alpha)$, by dividing $(\gamma_2 P(\alpha))/\gamma_2$, and sends it to party

The construction of private-BTS to DPP is stated by placing suitable polynomials and field elements.

Private-BTS to DPP reduction: Recall that party 1 has a vector (x_1, ..., x_n)**,** and party 2 has (y_1, ..., y_n), Since the reduction is from BTS, the party 1 needs to construct polynomials and party 2 needs to choose field elements. The polynomials for party 1 are:

$$P(w) = \sum_{i=1}^{n} R_i(w)$$
$$R_i(w) = w x_i$$

Note that all polynomials are linear and coefficients are known by party 1. *P* is formed as a summation of sub-polynomials R_i's for each bilinear terms in it. On the other hand, the field elements for party 2 are:

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$$
$$\alpha_i = y_i$$

After the desired function *f* is constructed, the lemmas and the theorem in the work of [ours] can be manipulated to prove that reduction can be done in polynomial time (due to finite summation of polynomials) and *f* privately computes the Dot Product Problem.

Lemma 3 *(Private-BTS)* The protocol $P_{PBTS}((R_1, \ldots, R_n), (\alpha_1, \ldots, \alpha_n))$ is a "privacy preserving protocol" for private-BTS problem.

Proof The views of the two parties are

$$VIEW_1 (R_1, \ldots, R_n) = (R_1, \ldots, R_{n,}, R_1(\alpha_1), \ldots, R_n(\alpha_n))$$

$$VIEW_2 (\alpha_1, \ldots, \alpha_n) = (\alpha_1, \ldots, \alpha_{n,}, \gamma R_1(\alpha_1), \ldots, \gamma R_n(\alpha_n))$$

The view of party 1 consists of its input ($R_1, \ldots, R_{n,}$) and output ($R_1(\alpha_1), \ldots, R_n(\alpha_n)$) so there is nothing to prove (S_1 can be used as the identity function). The input and

output of party 2 are (α_1, …, α_n) and ($\gamma R_1(\alpha_1)$, … , $\gamma R_n(\alpha_n)$), respectively. It is time to determine probabilistic polynomial-time algorithm S_2 such that $S_2(\alpha_1, \ldots, \alpha_n, \gamma R_1(\alpha_1), \ldots, \gamma R_n(\alpha_n))$ and $VIEW_2$ (α_1, …, α_n) are "statistically indistinguishable".

Let δ be a random element of F and $S_2(\alpha_1, \ldots, \alpha_n, \delta R_1(\alpha_1), \ldots, \delta R_n(\alpha_n))$ be defined as follows:

$$(\alpha_1, \ldots, \alpha_n, \delta R_1(\alpha_1), \ldots, \delta R_n(\alpha_n), \delta')$$

It is inconsequential to see that the taking after two groups is statistically vague:

$$(\alpha_1, \ldots, \alpha_n, \delta R_1(\alpha_1), \ldots, \delta R_n(\alpha_n), \delta')$$

$$(\alpha_1, \ldots, \alpha_n, \gamma R_1(\alpha_1), \ldots, \gamma R_n(\alpha_n))$$

This is because if δ is a random element of F then $\gamma R_i(\alpha_i)$ is a random element of F as well, for i=1, …, n.

Theorem 2 *(Two-party DPP)* The protocol formed by *f* yields a privacy-preserving protocol for two-party DPP.

Proof It is clear that DPP is privately reducible to BTS by choosing the given numerator and the denominator and there exists a protocol for private-RPE (Lemma 2) then by Lemma 1, given protocol is privately computes two-party Dot Product Problem.

4.4.2. Two-Party DPP Protocol Numerical Example

Assume that there are two parties Alice and Bob having two dimensional vectors, related Galois Field and blinding factor are as follows:

Alice has (5, 60)

Bob has (15, 20)

F = GF(1024577)

Blinding Factor of Alice = 3

Blinding Factor of Bob = 7

Step 1: Alice presents her polynomials in the given form and multiplies them:

$$R_1(w) = 5w$$

$$R_2(w) = 60w$$

Step 2: Bob chooses his field elements:

$$\alpha = 15$$

$$\beta = 20$$

Step 3: Alice forms her functional by putting respective values:

$$R(w_1, w_2) = R_1(w_1) + R_2(w_2)$$

$$R(w_1, w_2) = 5w_1 + 60w_2$$

Step 4: Alice blinds her polynomial multiplying with the blinding factor and sends it to Bob:

$$3R(w_1, w_2) = 3(5w_1 + 60w_2)$$

$$3R(w_1, w_2) = 15w_1 + 180w_2$$

Step 5: Bob computes $3T(\alpha, \beta, \gamma, \delta)$ by applying OPE.

If $(w_1, w_2) = (\alpha, \beta)$ variable replacement is done then it yields:

$$3R(15, 20) = 15.15 + 180.20 = 3825$$

Step 6: Bob blinds by multiplying with 7 Є F and sends the product 7(3R(α, β)) to Alice.

$$7(3R(15, 20)) = 7(15.15 + 180.20) = 7.3825 = 26775$$

Step 7: Alice cancels out her blinding factor by dividing from the product and sends the result to Bob.

$$7(3R(15, 20)) / 3 = 8925$$

Step 8: Bob cancels out his blinding factor and obtains the dot product.

$$7(R(15, 20)) / 7 = 1275$$

4.4.3. Multi-Party DPP

In the multi-party case, the BTS protocol should be applied (m-1) times if there are m parties are joined in. The following construction summarizes the reduction of multi-party DPP to two-party DPP. The key point about the calculation is all the coefficients of the target polynomial can be gathered from traversing all the parties except the last one. Hence rather than a single blinding factor for the step(2) of BTS protocol, there exist a multiplication of blinding factors, a new one is added at each iteration traversing from one to other.

Multi-party DPP Protocol: The protocol for multi-party DPP can be designed from two-party DPP in the following way:

(1) Parties are ordered from 1 to k in a manner that consecutive parties are involved in two-party DPP computation. This can be done with a common share or coin-

tossing into well protocol [9]. Hence the last party can be differed from previous attempts.

(2) Between party i and i+1, 0<i<k, the two-party DPP protocol works and Ri (polynomial regarding to i^{th} feature of the input vectors) can be iterated as:

$$R_t(w) = \left(\prod_{j=1}^{t} \alpha_{ji}\, x_{ji} \right) w \rightarrow R_{t+1}(w) = R_t(w) \left(\alpha_{(t+1)i}\, x_{(t+1)i} \right) w$$

Note that α_{ji}'s are the blinding factors.

(3) Hence for the i^{th} feature of the input vectors, the product is computed by multiplying all i^{th} coefficients except the last:

$$R_i(w) = \left(\prod_{j=1}^{k-1} \alpha_{ji}\, x_{ji} \right) w$$

(4) Summing up into all dimensions (features);

$$P(w) = \left(\sum_{j=1}^{n} \left(\prod_{j=1}^{k-1} \alpha_{ji}\, x_{ji} \right) \right) w$$

(5) The field elements are determined by the last party;

$$\beta = (\beta_1, \beta_2, \ldots, \beta_n)$$
$$\beta_i = x_{ki}$$

Total computation is quadratic in size and at the end of the computation; the last party (i.e. party k) gets the blinded dot product together with total sample size with remaining parties.

(6) In accordance with step (3) of BTS, the blinded dot product is traversed from all party k to party 2 backwards by dividing with corresponding blinding factor. Thus, the first party gets the dot product and send to other parties.

The unimportant gap of the protocol is the first party is more privileged than others yet this can be overcome by choosing the order in a circular round-robin fashion so the order of consecutive parties are preserved but only the first party changes. The probability to be the first party is 1/k which is Pareto-optimal for such a scheme (Özarar, 2007).

4.4.3. Homogeneous Matrix Algebra Protocols

In this section, multi-party determinant, trace, eigenvalue problems are discussed, respectively. Multi-party DPP protocol is used with *specific* input vectors in each of them.

4.4.3.1. Multi-Party Determinant Protocol

In this subsection, we demonstrate how the determinant of a nxn matrix could be registered utilizing the changes of the succession {1 ... n} by the extremely popular Leibnitz's recipe.

Theorem 3 (Determinant of a matrix with given permutations) Let A be a *n* x *n* matrix and V_i be a *1* x *n* matrix where for all $j \neq i$, $v_{1,j} = 0$ and $v_{1,i}=1$. Let S(n) be the set of all permutations on the sequence {1, ...,n}with n! elements and $\sigma_1(i), \ldots, \sigma_{n!}(i)$ represent each of the possible permutations of the sequence {1, ..., n} so that for all $\sigma(i) \in S(i)$, there exists k such that $\sigma(i) = \sigma_k(i)$."

Then;

$$\det(A) = \sum_{k=1}^{n!} \left\{ sgn(\sigma_k) \left[a_{1,\sigma_k(1)}, \ldots, a_{n,\sigma_k(n)} \right] \right\}$$

Suppose that we have a kxk matrix, *A*, then an elementary product from this matrix will be a product of *k* entries from *A* and none of the entries in the product can be from the same row or column. Thus, each permutation corresponds to a signed elementary product. Some remarks have to be done since there is an analogy between dot product and determinant computations. The sign is always positive and total number of n-tuple products is n in dot product, on the contrary in determinant; the sign is alternating with the parity of permutation and n! copies of elementary products are involved to the computation. Yet the multi-party dot protocol might be changed in the taking after route to handle multi-party determinant processing situation in the light of our first surmise. So there are N parties which form the rectangular kxk matrix zero-padded matrix.

Multi-party Determinant Protocol: The protocol for multi-party determinant protocol where all that needs to be done is to calculate the product of every possible permutation and then sum them up can be designed from multi-party DPP. We have two define a number of algorithms to get the notion.

```
1 Calculate_Determinant(Input Matrix, Output Determinant)
2 {
3           Check for invalid conditions (not square etc)
4           if number of rows = 1 then return only element
5           Initialize determinant value to zero
6           while not used all permutations
7           {
8                 Get next permutation vector
9                 Calculate the elementary product of the current permutation
```

10 **if** (the current permutation is even) **then**

11 Add the product to the determinant value

12 **else**

13 Subtract the product from determinant value

14 }

15 **return** determinant value

17 }

Algorithm I. Determinant Calculation

The Johnson-Trotter algorithm proposes an astute course to straightforwardly create stages of the needed length without figuring more limited changes. The ordered system needs the definition of a guided whole number is declared to be portable depending on if it is more terrific than its instantaneous neighbor in the bearing its taking a gander at.

The Johnson-Trotter algorithm can be given with the pseudo-code:

1 Initialize the first *permutation* with {1, 2, ..., n}

2 **while** there exists iki kelime bir arada olmamış a mobile integer

3 {

4 Find the heftiest versatile whole number k

5 Swap k and the contiguous whole number its looking

6 Reverse the course of all numbers larger than k

7 }

Algorithm II. Johnson-Trotter Algorithm for Permutation Calculation

The Algorithm I should be modified in order to handle multi-party case:

(1) Parties are ordered from 1 to k in a manner that consecutive parties are involved to two-party DPP computation. This can be done with a common share or coin-tossing into well protocol [9]. Hence the last party can be differed from previous attempts.

(2) Lines 2, 3 and 4 are handled by the first party.

(3) The **while** loop should be managed in the coordination of the first party. All the permutations are calculated by the result of Johnson-Trotter Algorithm for all parties in parallel at line 8.

(4) Calculation of the current permutation should be handled by using multi-party DPP protocol. *For each calculation while the parties are being traversed, each party includes his/her term(s) exist(s) in the formation of permutation for the target elementary product in the direction of the DPP protocol. The vectors given as input to the protocol are partially created and should have all-zero values in all the dimensions except the first one which has the target term.*

(5) Lines 10-13 are handled by the first party and the determinant result stored in the return value is distributed to the parties.

Example 1:

Let Alice, Bob and Cindy are involved in a determinant computation where Alice and Cindy have single vectors yet Bob has two vectors. Each vector has 4 dimensions suitable with our Assumption 1. There are totally 4! permutations that form the determinant computation. An elementary product corresponding to the permutation $\sigma_{3412}=(A_3B_4B_5C_2)$ is illustrated which is the eleventh element in the Johnson-Trotter order. The vectors from x_1 to x_4 are partially created which are given as input to multi-party DPP protocol. The D matrix is also used in the following examples.

$$D = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ b_5 & b_6 & b_7 & b_8 \\ c_1 & c_2 & c_3 & c_4 \end{pmatrix}$$

$$|D| = sgn(\sigma_{a_1 b_2 b_7 c_4}) a_1 b_2 b_7 c_4 + \ldots + sgn(\sigma_{a_3 b_4 b_5 c_2}) a_3 b_4 b_5 c_2 + \ldots$$
$$\ldots + sgn(\sigma_{a_2 b_1 b_7 c_4}) a_2 b_1 b_7 c_4$$

$$\Rightarrow \{(a_3, 000), (0 b_4, 00), (00 b_5, 0), (0, 0, 0, c_2)\}$$

$$\Rightarrow a_3 b_4 b_5 c_2$$

4.4.3.2. Multi-Party Trace Protocol

In linear algebra, the trace of an n-by-n square matrix A is outlined to be the summation of the elements on the main diagonal (the diagonal from the upper left to the lower right) of A, i.e.,

$$\operatorname{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$$

where a_{ij} denotes the value on the i^{th} row and i^{th} column of *A*.

The idea for trace computation is based on selecting the target vectors given as an input to DPP in the form of (1, 1, ..., d_{ii}, …, 1, 1) where d_{ii} is the diagonal element which corresponds to the identity permutation.

1 *Calculate_Trace*(**Input** Matrix, **Output** Trace)

2 {

3 Check for invalid conditions (not square etc.)

4 **If** number of rows = 1 **Then** return only element

5 Initialize trace value to zero

6 Get identity permutation vector, σ_{id}

7 Calculate the elementary product of σ_{id} via DPP

8 **Return** trace value

9 }

Algorithm III. Multi-party Trace Calculation

The algorithm is the reduced version of determinant protocol where parties are traversed just once with the identity permutation. Hence the input vectors are formed of all 1's except the diagonal element.

Example 2:

Regarding the previous example while taking another permutation into account σ_{1234} = ($A_1B_2B_7C_4$) which yields the trace, x_1 to x_4 are formed and the trace value is computed.

$$\left\{x_1=(A_1,1,1,1),\ x_2=(1,B_2,1,1),\ x_3=(1,1,B_7,1),\ x_4=(1,1,1,C_4)\right\}$$
$$\Rightarrow tr(D)=A_1+B_2+B_7+C_4$$

4.4.3.3. Multi-Party Trace Protocol

Consider the square matrix A It is said that "λ is an eigenvalue of A" if there is a non-zero vector x such that "**Ax = λx".** In this case, x is called an "eigenvector (corresponding to λ)", and the pair *(λ,x)* is called an "eigenpair for A". A nontrivial key of this comparison is conceivable if and just if the coefficient matrix A−λI is non-invertible. Such a condition can be expressed as the vanishing of the determinant

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

When this determinant is calculated, the "characteristic equation of A" can be obtained:

$$P(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \cdot \quad \cdot \quad \cdot + \alpha_n = 0$$

The algorithm for eigenvalue computation is quite similar to determinant computation yet the terms of elementary products include indeterminate λ values as well. Hence, the first party can compute the characteristic equation of the matrix instead of exact determinant. The eigenvalues can be obtained solving the characteristic equation for λ.

The vectors given as input to the protocol are partially created and should have all-zero values in all the dimensions except the first one which has the target term decremented by indeterminate. The following example summarizes the point.

Example 3:

For the matrix D, in example 1, the partially created vectors for the permutation $\sigma_{3412}=(A_3B_4B_5C_2)$ to compute eigenvalues are as follows:

$$\{(a_3-\lambda,0,0,0),(0,b_4-\lambda,0,0),(0,0,b_5-\lambda,0),(0,0,0,c_2-\lambda)\}$$
$$\Rightarrow \quad (a_3-\lambda)(b_4-\lambda)(b_5-\lambda)(c_2-\lambda)$$

After demonstrating the usage of OPE in matrix algebra protocols, let us analyze final application which is on Hamming Distance Computation. Different from previous ones, its application to Hierarchical Document Clustering is also presented.

4.5. Application of OPE to Secure Hamming Distance Computation

4.5.1. Privacy Preserving Two-Party Hamming Distance Computation Problem (HDP)

Suppose Alice has a binary vector $X = (x_1, \cdots , x_k)$ and Bob has also another binary vector $Y = (y_1, \cdots , y_k)$. They want to determine the Hamming Distance of X and Y without revealing each other's vector. In informative data speculation, the "Hamming Distance" between useless successions of equivalent length is the number of positions at which the relating images are contrasting. In an additional way, it measures the least number of substitutions needed to update one string into the different, or the number of slips that transformed one string into the different.

For the remaining scope, the previous assumption on homogeneous cooperation should be assumed as well.

4.5.2. Privacy Preserving Two-Party HDP Algorithm (via OPE)

In this subsection, an algorithm (in a more programmer-friendly way) for privacy preserving two-party "Hamming Distance" problem (HDP) is designed. A function should be taken as a target to place the terms in the OPE for developing such an algorithm.

Let $\boldsymbol{f}$ be the function for such a computation, its domain set must be *n*-dimensional vectors for both parties and range set must be a singleton numeric value ("Hamming Distance"). $\boldsymbol{f}$ can be constructed as:

$$\boldsymbol{f}: \mathbb{Z}_2^{\,k} \times \mathbb{Z}_2^{\,k} \rightarrow \mathbb{Z}$$

The straight forward solution is approximating f by a circuit. Yet it is well known that the cost for implementing such a circuit is so inefficient that a new solution should be developed for the specific case. We now develop a protocol using OPE by placing meaningful polynomials and field elements for Hamming Distance Problem.

Party 1 and Party 2 have binary vectors $X = (x_1, \dots, x_k)$ and $Y = (y_1, \dots, y_k)$, respectively. Moreover, Party 1 generates a random vector $R = (r_1, \dots, r_k)$ where each $r_i \in \mathbb{Z}$ and Party 2 generates a singleton random value S from the set of integers as well. Party 1 is the sender who determines the polynomials and Party 2 is the receiver who chooses the field elements. The result should be known by Party 2. Let Pi(x) denotes the ith polynomial of Party 1 during the protocol hence $P = (P_1(x), \dots, P_k(x))$. The return value d denotes the Hamming Distance of X and Y.

The construction of the algorithm via OPE is as follows:

Inputs; k, X, Y, R, P, S

Output: d

Algorithm: $P_{HDP}(P,X,Y)$:

```
1 d := 0;                                   /* Party 2 */
2 for i = 1 … k
3 begin                                     /* OPE starts*/
4     if (x_i = 0) then P_i(x) = r_i+x;     /* Party 1 */
5     else P_i(x) = r_i+1-x;                /* Party 1 */
6     d := d + P_i(y_i);                    /* Party 2  */
7 end                                       /* OPE ends*/
```

8 d := d + S; /* Party 2 */

9 d := d - $\sum_i r_i$; /* Party 1 */

10 d := d - S; /* Party 2 */

Algorithm IV. HDP via OPE Algorithm

At the beginning, Party 2 initializes the output (Hamming distance) to 0. In the main loop, both parties apply OPE protocol. Party 1 determines two types of linear polynomials according to the value of the vector element. Party 2 chooses nothing but the corresponding value at each step.

At line 4, Party 1 blinds the polynomials P by adding with a pre-determined random integer. This is because at line 6, Party 2 cannot get any intermediate result about the distance. Between lines 3-7, OPE is applied between parties. The cardinality of the loop is equal to the cardinality of input vectors. Party 1 prepares the polynomials and sends them to Party 2. Party 2 evaluates the polynomials and keeps the summation.

At line 8, Party 2 has calculated the distance yet the result has been blinded. If he directly sends the result to Party 1 then Party 1 will have a chance to get the exact distance. It is time for Party 2 to blind the result by adding a singleton value and sends it to party 1. Party 1 sums up the total value of random array, extracts the summation from the value sent by Party 1. Party 1 still cannot determine the distance because it was blinded with S. After that, it sends the value to party 2 again at line 9. At line 10, party 2 obtains the final distance value by subtracting S from d.

Privacy Proof of Private HDP

It is time to prove that Private Hamming Distance Problem is privacy preserving. We need to use Composition Theorem stated in Lemma 1 for Passive Adversary as a lemma for main proof.

A method to solve OPE was given by Naor and Pinkas (Naor, 2005). Let $P_{OPE}(P,\alpha)$ denote the "privacy-preserving protocol" for OPE. We present a protocol $P_{HDP}(P,X,Y)$ for HDP, which uses $P_{OPE}(P,\alpha)$ as an oracle.

Theorem 4 *(Two-party Private HDP)* The algorithm PHDP(P,X,Y) yields a privacy preserving algorithm for two-party HDP.

Proof The views of the two parties are;

$$VIEW_1\ (P) = (P,\ d+(\Sigma_i\ r_i)+S)$$

$$VIEW_2\ (Y) = (Y,\ d+P_1(x),\ d+P_2(x),\ \cdots\ d+P_k(x),\ d+\ (\Sigma_i\ r_i),\ d)$$

We have to show two PPTAs, $S_1(P,d)$ and $S_2(Y,d)$ are statistically indistinguishable with respective views, $VIEW_1(P)$ and $VIEW_2(Y)$.

Let $z, z', z_1, \cdots z_k$ are random elements from $\mathbb{Z}$.

Define;

$$S_1(P,d) = (P,\ d+z)$$

$$S_2(Y,d) = (Y,\ d+z_1,\ \cdots,d+z_k,\ d+z',\ d)$$

It is simple to observe that the following two ensembles are statistically indistinguishable:

$$(P,\ d+z)$$

$$(P,\ d+(\Sigma_i\ r_i)+S)$$

The reason is that if z is a random integer then d+z is a random element of $\mathbb{Z}$ as well. P(d+z) = P(d+z|z) since the cardinality of $\mathbb{Z}$ is equal to the cardinality of the integers greater than or equal to z.

Similarly,

$$(Y, d+P_1(x), d+P_2(x), \cdots, d+P_k(x), d+ (\Sigma_i r_i), d)$$

$$(Y, d+z_1, \cdots, d+z_k, d+z', d)$$

are also statistically indistinguishable in accordance with the previous approach.

Recall that P_{HDP} uses the protocol P_{OPE}. Using the Canetti's composition theorem as a lemma, we conclude that P_{HDP} is privacy preserving.

CHAPTER 5

PRIVACY PRESERVING HIERARCHICAL CLUSTERING

5.1. Document Clustering

Document clustering has been examined for usage in different differentiating areas of expressions mining and data recuperation. At first, document clustering was examined for upgrading the accuracy or review in informative content recovery frameworks and as an effective way of recognizing the closest neighbors of a document. As of late, clustering has been recommended for utilization in scanning an accumulation of documents or in arranging the effects reverted by a web based info search according to a user's question.

Document clustering has likewise been utilized to mechanically create hierarchical groups of documents. A to a degree special methodology discovers the indigenous groups in a presently existing document taxonomy, then after that utilizes the aforementioned groups to process a viable document classifier for unique documents.

“Agglomerative hierarchical clustering” and “k-means” are two clustering systems that are ordinarily utilized for document clustering. “Agglomerative hierarchical clustering” is regularly depicted as “preferred” than K-means, in spite of the fact that slower.

5.2. Hierarchical Agglomerative Clustering

Progressive routines produce a settled party of segments, with a singular, complete aggregation at the top and singleton packs of one of a kind demonstrates at the base. Each transitional level may be viewed as combining two parties (or part an aggregation from the thereafter higher level). The result of a progressive grouping equation may be graphically indicated as tree, called a dendogram. This tree graphically showcases the uniting get ready and the partly group. The dendogram at the right exhibits how four demonstrates may be combined into a particular aggregation. For record bunching, this dendogram outfits taxonomy, or various leveled record.

There are two basic approaches for hierarchical clustering:

Agglomerative: Start with the focuses as unique clusters and, at every step, consolidation the most comparative or closest match of clusters. This needs a definition of cluster likeness or separation.

Divisive: Start with one, comprehensive cluster and, at every step, part a cluster until just singleton clusters of distinctive indicates remains. In this case, we ought to choose, at every step, which cluster to part and how to perform the part.

Agglomerative methods are more regular, and we condense the conventional agglomerative hierarchical clustering strategy as takes after:

Simple Agglomerative Clustering Algorithm:

Given a set of N items to be clustered, and an NxN distance (or similarity) matrix, the basic algorithm of Johnson's (Johnson, 1967) hierarchical clustering is this:

1. Start by relegating every part to its particular bunch, so that in the event that you have N parts, you now have N groups, every holding simply one article. Let the separations (likenesses) between the groups meet the separations (likenesses) between the articles they hold.

2. Find the closest (most comparable) match of bunches and consolidation them into a specific bunch, for the purpose that now you have one less group.

3. Compute separations (similitudes) between the newfangled group and each of the old groups.

4. Repeat steps 2 and 3 until all parts are grouped into a solitary group of size N.

Step 3 could be finished specially, which is what differentiates single-connection from complete-connect and normal-channel grouping. Without abatement in all inclusive statement, we decide on single-connection grouping (likewise reputed to be the connectedness or least technique), we recognize the reach between one assembly and an additional aggregation to be comparative to the most modest run from any member of one assembly to any member of the different party.

5.3. Private Hierarchical Agglomerative Clustering

Assumptions. Our aim is to design a privacy-preserving HAC that does not use a TTP. When we show the embodiment of an equation, we state surmises made in the outline of our privacy-preserving protocol.

Number of parties. Any number of parties can be involved in protocol.

The adversary model. We collect a semi-honest adversary (likewise called legitimate but inquisitive antagonist model). There are standard projects that transform a methodology that is secure in the semi-honest model and transform a methodology that is secure in a more general malevolent model.

Information disclosure. Our privacy-preserving functional process uncovers the Hamming Distance at the diverse steps between two parties. Along these lines, the reckoning of grouping specimens consistent with the HAC could be performed generally. Hence, the complexity of our privacy-preserving algorithm depends only on the number of steps of Hamming Distance computation which is $O(n^2)$.

Inputs. For document clustering, each document can be represented as a binary vector where each element indicates where each element indicates whether a given/word term present or not.

5.4. Multiparty PP-HAC Algorithm

Inputs*; M parties, each party has N_m binary encoded document vectors denoted by v_i*

Oracle: P_{HDP} (x,y)

Output*: Dendogram*

Algorithm:

/* Each Party gets the total number of vectors */

Party 1 produces a random number *R*

sum := *R* + N_1;// Total number of document vectors

Party 1 sends *sum* to Party 2

for i = 2 … M

begin

sum := sum + N_i;

if (i != M)

Party i sends Sum to Party i+1

else

Party k sends Sum to Party 1

end

sum := sum – R;

Party 1 broadcasts sum to other parties

/* Each Party calculates the Hamming Distance of vectors in his own data set */

for i = 1 … M

for k = 1 … N_i

for t = 1 … N_i

begin

if (k <t)

HD (v_k, v_t) {v_k, ϵ Party i and v_t ϵ Party i}

end

/* Parties calculate the Hamming Distance of vectors between their data sets */

for i = 1 … M-**1**

for j = 2 … M

Apply P_{HDP} (v_k, v_t) {v_k ϵ Party i and v_t ϵ Party j}

Party j shares the result with Party i

end

end

Algorithm V. HAC via OPE Algorithm

Each party can form the dissimilarity matrix of size ***sum*** **x** ***sum*** by combining values from Hamming Distances in their data sets and between parties data sets

Each party applies simple Hierarchical Agglomerative Clustering according to dissimilarity matrix and gets the dendogram.

5.5. A Concrete Example

Let there be 3 parties, denoted by A, B and C.

Party A has $a_1 = \{0, 1, 1, 0\}$ and $a_2 = \{1, 0, 0, 1\}$

Party B has $b_1 = \{0, 1, 1, 1\}$, $b_2 = \{1, 1, 1, 1\}$ and $b_3 = \{0, 0, 0, 0\}$

Party C has $c_1 = \{1, 1, 0, 0\}$

Let us trace P_{HDP} ({0, 1, 1, 0}, {0, 1, 1, 1})

Party A determines R = [12, 34, 56, 78], $\sum r_i = 180$

Party B determines S = 67

$P_1(x) = r_1+x = x+12$

$P_2(x) = r_2+1-x = 34+1-x = 35-x$

$P_3(x) = r_3+1-x = 56+1-x = 57-x$

$P_4(x) = r_4+x = x+78$

$P_1(0) = 12$

$P_2(1) = 34$

$P_3(1) = 56$

$P_4(1) = 79$

d = 181+S = 181+67 = 248 */* *Party 2 */*

d = 248–$\sum r_i$ = 248-180 = 68 */* *Party 1 */*

d = 68-S = 68-67 = 1 */* *Party 2 */*

The Hamming Distance matrix formed by P_{HDP}:

	a_1	**a_2**	**b_1**	**b_2**	**b_3**	**c_1**
a_1	**X**	**4**	**1**	**2**	**2**	**2**
a_2		**X**	**3**	**2**	**2**	**2**
b_1			**X**	**1**	**3**	**3**
b_2				**X**	**4**	**2**
b_3					**X**	**2**
c_1						**X**

The first cluster is between a_1 and b_1. Each party locally applies HAC and for the final dendogram depicted in the list as:

{(((((a1,b1) , b2) , a2) , b3), c1)}

CHAPTER 6

PRIVACY PRESERVING COMPARISON OF INFORMATION

In this chapter, a real life scenario from computer security concept, namely remote server authentication problem, is discussed and a solution with Oblivious Polynomial Evaluation is developed to overcome it. This can be achieved with privacy-preserving comparison of information. We present an algorithm that solves this case.

6.1. Comparing Information Without Leakage

Assume that two parties Alice and Bob keep a specific name. The two get-togethers may like to check positing that they both have the same consolidate, under the condition that if the inputs are divergent they do not prefer to uncover any advantageous qualified data about them (forbidding for the way that they are uncommon). The major impediment in organizing a procedure for this situation is that the territory of inputs, e.g. names, is clearly modest enough to distinguish an animal drive pursue over all conceivable inputs.

This scenario was deeply examined by Fagin et. al. (Fagin, 1996), with subsequent studies by Crepeau et.al. (Crepeau, 1995). To differentiate the system adequately, if the parties inputs are (α, β), freely, then their yields are 1) if α = β, and 0) usually. By ideals of noxious parties we unwind the crucial and announce that while for α = β the yields may be noncompulsory (following to a poisonous party can perseveringly upgrade its fuse), by temperance of inputs α = β, the yield of the respectable get-

together may besides be 0. Careless estimation of linear polynomials may be utilized to create an actually fundamental award a demonstration to this situation.

6.2. PP-Comparison of Information Algorithm

Inputs Alice has α and Bob has β

Oracle: Oblivious evaluation of linear polynomials

Algorithm:

• Alice creates a random $P_A(\cdot)$.

• Bob creates a random $P_B(\cdot)$.

• The parties execute the OPE twice, switching roles;

In the first summon, Alice obliviously calculates Bob's polynomial (she may as well decide on to memorize $P_B(\alpha)$.)

In the second summon, Bob neglectfully computes Alice's polynomial (he might as well pick to memorize $P_A(\beta)$.)

The parties figure and think about the two qualities, $P_A(\alpha) + P_B(\alpha)$ (processed by Alice), and $P_A(\beta) + P_B(\beta)$ (figured by Bob).

Provided that $\alpha = \beta$ then the two qualities are the same, elsewise they are unexpected with chance 1/|F|

For semi-honest parties, security is protected following the parties investigate qualities of the method $P_A(x) + P_B(x)$ that has the emulating lands (which are unimportant to substantiate):

i. The capacity is match-wise free.

ii. Every party just registers $P_A(x) + P_B(x)$ once.

iii. Every party registers $P_A(x) + P_B(x)$ without disclosing x to the different party.

By virtue of vindictive parties, the verification is smooth if the methodology uses an OPE methodology which plans the extraction of the receiver's incorporate. To be particular, given a TTP which figures the capacity in the perfect model, we can recreate the joint movement of the malevolent party and the yield of the offbeat party:

Suppose that Alice is malicious. We disconnect her information α from her summons of the OPE order and outfit α to the TTP. Relying on if the reaction is 1 we bear on the methodology by surveying her polynomial at $\beta = \alpha$, sending the worth $P_A(\alpha) + P_B(\alpha)$ to the examination, and fixing Bob's yield needy upon the conclusion of the outline. Note that because of a malevolent Weave, who registers a value of Alice's polynomial in the wake of letting her survey his, we absence the limit to focus Bob's drop in before surveying $P_B()$. We can, in any case, execute the OPE of Bob's polynomial twice in the re-establishment, remember $P_B()$ extensively, and following that have the limit to figure any worth of $P_B()$ and use it to give the right regard to the association

6.3. Application to Password Security

This methodology could serve as a backing for a usually validated key exchange subordinate upon conceivably frail passwords. Consider a user who wishes to login to a remote server over a unstable arrangement. She attempts not to like to send her secret key unhindered, and is not even beyond any doubt that the remote party is the target server. An extraneous scenario is that the secret key should not have enough entropy and may in these lines be exposed to vocabulary attacks. Suppose that there is no public key infrastructure (PKI), and that the user does not pass on with her an accessible key of the remote server.

The most instinctive definition of the case is as a “looking at information without emitting it” situation. Hinging on if the right user contacts the right server they both may simultaneously be considering the same secret word, and they can verify if this is the scenario using the given contrivance. Thus, relying on assuming that its placed that there is no dynamic adversary, and that the unequivocally operation of the opponent is to listen to the transport between the two parties then a while later endeavor to copy the user, then the functional process may be used for secret word verification. Additionally, it may be used to process a session key for the two parties, whose entropy does not rely on the entropy of the passwords.

Unequivocally, each of the parties picks a spasmodic direct polynomial and (neglectfully) courses of action the total of the two polynomials at x = “secret word”. They use the initial part of the yield for affirmation and the second part as a “session key”.

Given that the hopeful can additionally be occupied, i.e., redesign the movement sent between the two parties, and then the above order is insufficient. Regardless of the way that the foe would not be able to decrypt the notes sent between the parties (e.g. in the summons of the neglectful transfer methodology), it can overhaul them and make the yield of the neglectful transfer methodology to be different, be that as it may related to its legitimate yield. The adversary can use this article to ambush different summons of the methodology that are once again being executed in parallel.

CHAPTER 7

SUMMARY, RESULTS AND CONCLUSIONS

7.1. Summary

In this thesis study, secure multiparty computation concept is deeply investigated while focusing on a specific cryptographic technique called Oblivious Polynomial Evaluation.

Privacy is beyond all doubt the most paramount lands of an informative content framework should fulfill, in which frameworks the requirement to impart informative data right around better, untrusted elements, the security of sensible data has a significant part. Accordingly secure dispersed computation, which was finished as a major aspect of a more substantial assembly of examination in the hypothesis of cryptography, has attained exceptional consequences. The suggested conclusions were exhibited using non particular works in advancement that might be had a cooperation with any capacity that has a successful representation as a circuit. Privacy preserving algorithms have been in the recent past presented with the point of averting the disclosure of sensible informative data.

Cryptographic orders for secure processing accomplished grand results: it was shown that non particular advancements may be used to enroll any role securely and it was besides demonstrated that certain capacities might be figured outstandingly profitably using particular works in advancement. Still, a secure request for figuring a certain capacity might persistently be more exorbitant than a guileless request that does not outfit any security. By making use of cryptographic procedures to recovery

sensitive data and outfitting access to the filed informative data needy upon a specific's part, we ensure that the informative content is protected from protection breaks.

In this thesis, the broad concept of Secure Multiparty Computation is analyzed especially concentrating on a cryptographic building block called Oblivious Polynomial Evaluation. The model paradigms in SMC are given and two types of adversarial behavior are explained. The literature survey and notable specific problems on SMC are briefly discussed. Methods to solve SMC problems are given and mainly focused on a cryptographic one, namely OPE. The usage of OPE to concrete problems like secure overall mean and Hamming Distance computations are demonstrated including privacy proofs. The applications and complexity of OPE protocols are presented.

The most common drawback of SMC protocols with OPE is their inefficiency. They require considerable computation and communication costs. We think that encourage research around there is pivotal for the improvement of secure and effective methodologies in this field with the assistance of tamper-invulnerable devices to give ease viable keys besides.

7.2. Implementation of OPE

Remark that any method from k bits to k bits could be acted for as a polynomial over a limited field GF(2^k), in any case its degree might go as towering as $2^k - 1$. Gilboa et al. states that “Hence one could jump at the chance to center on these works that might be stood for by level degree polynomials. This makes have a few fascinating provisions”(Gilboa, 1999; Naor, 1999). The plan recommended in Naor et. al. is a great deal more effective than the tried and true way of enduring oblivious circuit assessment orders, in any case its security is dependent upon two suspicions (Naor,

1999). One suspicion is the being of a secure oblivious transfer order while the diverse, a patched up one, is the relentlessness of a “Noisy Polynomial Interpolation scenario”. It was later shown in Bleichenbacher et. al. that this unique suspicion may be much weaker than envisioned and suggested the use of a possibly stronger stubbornness supposition on a “Polynomial Reconstruction Problem”. (Bleichenbacher, 2000). The methodology exhibited in Gilboa et. al. is dependent upon a surmise that the Decisional Diffie-Hellman (DDH) suspicion moreover keeps over the aggregation Z_{nxn} (Gilboa, 1999), where n is the result of two hefty primes. Despite the for the most part contemplated DDH over Zn, the hardness of this scenario in this redid setting is yet to be examined.

7.3. Complexity of OPE in the context of Cryptography

The overhead of an algorithm is unequivocally polynomial assuming that it is limited by a polynomial function of the number of information articles in the information, as opposed to the span of the input values. The issue of fifinding protocols for Oblivious Polynomial Evaluation whose overhead does not hinge on the field size, may be viewed as being a contender with scanning for decidedly polynomial algorithms in combinatorial improvement (e.g. for straight customizing). In the setting of cryptographic protocols, we measure the overhead in terms of the number of public key operations (i.e. operations dependent upon trapdoor functions, or comparative operations) with a specific security parameter, where the extent of the inputs to the public key operations is straight in the security parameter. Tallying only public key operations is justified following the overhead of public key operations relies on the length of their inputs, and is more fabulous by requests of greatness than the overhead of symmetric key operations (i.e. operations dependent upon hash functions) (Naor, 2000).

Hence, we declare that a cryptographic protocol is strongly polynomial if the accompanying two lands keep: (1) the number of public key operations performed by

the protocol is limited by a polynomial function of a security parameter and of the number of inputs (but not their size), and (2) the length of the inputs to the public key operations is straight in the security parameter. Remark that the number of symmetric key operations that the protocol performs could be polynomial in the size of its inputs (Naor, 2000).

Oblivious polynomial assessment may be actualized using general methodologies for secure two-get-together reckoning. Be that as it may, as was specified above, the aforementioned orders deal with a twofold circuit that registers the capacity and is not determinedly polynomial, as the number of oblivious transfers they use is at slightest straight as a part of log |F|, where F is the field over which the polynomial is defined. A different work in advancement of OPE may be subordinate upon using homomorphic encryption. That task, too, is not emphatically polynomial, as the extent of the data to the homomorphic encryption method may as well be with the anticipation that the authority's enter in the OPE order. Interestingly, the number of oblivious transfers used by the methodologies displayed as a part of this paper does not rely on the measure of the underlying field: The length of the parts transferred in the oblivious transfer methodologies is of size log |F|, in any case they require actually O(1) popular key operations for each transfer. Specifically, if log |F| is longer than the length of the data of the OT methodology, then the parts to be transferred in the OT methodology are encrypted using sporadic keys, and the examining keys are transferred in the certified OT methodology.

7.4. Applications of OPE

There are two major mandates of an OPE request. One is whenever k-wise self-rule can swap full self-legislation or pseudo-discontinuity. Such property is needed, case in point, for the interest of collecting unacknowledged coupons that arrangement unacknowledged regulation of obliged stakes (e.g., for raising a unacknowledged dissent box). The differentiating sorts of obtainments utilizes OPE for distinguishing

qualified data without spilling it, or preserving obscurity when Receiver may besides select the worth of a polynomial at a certain exhibit. Acquirements of this nature solidify a system that permits reliable and security shielding metering.

7.5. Future Research Directions

A new method for solving SMC problems will be tamper-resistant devices. This can be added as a fourth alternative to already existing methods or techniques.

The answers weighed in on so far meet some of the targets of the preferable structural planning, anyhow need an assembly of commonly suspicious suppliers running associate-to-associate SMC protocols. Setting up and managing such a consortium is quite demanding. Likewise, the cryptographic routines utilized as a part of these protocols give restricted, rather underprivileged underpin for legit essence requisitions, yet for the latent antagonist model.

We can develop the techniques and devices utilized within SMC by incorporating tamper-resistant cryptographic co-processors. These gadgets combine cryptographic techniques and physical insurance to furnish respectability of executable code, information mystery, and information trustworthiness. A few results utilizing tamper-resistant gadgets have as of late been suggested; anyhow this line of exploration has not accepted sufficient regard (Benenson, 2006; Katz, 2007). In a server-based key, we recommend the utilization of such mechanisms to apprehend preferable functionalities for secure computation undertakings, as trusted implementations of the undertakings, running in a shut, secure execution earth.

An influential cryptographic co-processor has sufficient assets to run inside a secure computation, for effortless requisitions and a humble number of parties. The unit can cooperate with the parties utilizing the perfect case, amplified to prepare the parties

to verify that the gadget is suitable and runs trusted code. These mechanisms are costly and their assets are innately restrained. A more general explanation utilizes hardware security modules just for a humble set of undertakings that cannot be (prudently) done by cryptographic protocols, for example cycle and dispersion of edge decryption keys and a comprehensive set of arithmetic operations and numerical functions required good to go provisions.

Utilizing this methodology, straight-line systems for scientific computations could be done all in all effortlessly in a specific server setting. Taking care of control course and decryption of the outcomes without compromising privacy appears more demanding. Edge decryption could at present help, and if the server may memorize the outcome, regulate face to face time between parties is not vital for decryption.

APPENDIX A

PROTOCOL FOR OBLIVIOUS POLYNOMIAL EVALUATION AND ITS EVALUATION

Assume that there are two parties, Alice who has a function f and Bob who has an input x. They want to collaborate in a way for Alice to learn nothing and for Bob to learn f(x) and nothing more. A protocol achieving this task for any function f and any input x is called an Oblivious Function Evaluation protocol. The remarkable results of Yao (Yao, 1986) and Goldreich, Micali, and Wigderson (Goldreich, 1987) showed that such protocols exist, under some standard cryptographic assumptions. Their protocols utilize a Boolean circuit to stand for the function f and afterward re-enact the computation of this circuit in certain oblivious way. The computational or communicational overhead of their protocols depends just linearly on the circuit size of the function f, which is the best one can anticipate from a unpredictability-speculative outlook. However, their protocols are far from being down to earth when all is said in done, and this situation still requirements a ton of work to be finished. One line of exploration is to think about different representations of functions and see if more effective reproduction might be realized through such representations.

Note that any function from m bits to m bits can be represented as a polynomial over a finite field $GF(2^m)$, but its degree could go as high as $2^m - 1$. Thus one would like to focus on those functions that can be represented by low degree polynomials. This turns out to have several interesting applications (Gilboa, 1999; Naor, 1999). The scheme proposed in Naor et. al. is much more efficient than the conventional way of

going through oblivious circuit evaluation protocols, but its security is based on two assumptions (Naor, 1999). One assumption is the existence of a secure OT protocol while the other, a new one, is the intractability of a Noisy Polynomial Interpolation problem. It was later shown in Bleichenbacher et. al. that this new assumption may be much weaker than expected and suggested the use of a possibly stronger intractability assumption on a Polynomial Reconstruction Problem (Bleichenbacher, 2000). The protocol presented in Gilboa et. al. is based on an assumption that the Decisional Diffie-Hellman (DDH) assumption also holds over the group $\mathbb{Z}_{nxn}$ (Gilboa, 1999), where n is the product of two large primes. Contrary to the well studied DDH over $\mathbb{Z}_n$, the hardness of this problem in this new setting is yet to be studied. A novel OPE protocol is proposed in 2009 and we strongly recommend to use this novel one for the implementation of OPE (Vanishree, 2009).

Breakdowns indicate that the protocol furnishes unconditional security as in opposition to the computational security gave by the awhile ago existing protocols. The essential computational bottleneck of the existing developments is the OT protocol, the computational cost of which is basically exponentiations in limited fields. As an additional major stake of the protocol, this overhead is deterred and subsequently the protocol is authenticated to be more proficient.

One magnetic emphasize of our protocol is that they might be modified truly effortlessly to handle floating-indicate numbers. This is not the case for existing OPE protocols which rely on some specific lands of finite fields. Numerous vital provisions in legit essence include numerical computation over floating-indicate numbers, as a substitute for over numbers or discretionary finite fields. There is no powerful mapping known that inserts floating-indicate numbers into finite fields where mathematics might be fulfilled effortlessly.

The approach of Naor is to scale floating-point numbers up to integers with some book-keeping, apply some existing OPE protocol over integers, and then do a normalization to get back floating-point numbers (Naor, 1999). The extra work of scaling up, scaling down, and book-keeping makes their algorithm less appealing.

A.1. Preliminaries

We fix a security parameter τ, so that any number within a small factor of $2^{-\tau}$ is considered negligible.

For a distribution D over a set S, let D(i), for $i \in S$, denote the probability of i according to D, and define D(A), for A as a subset of S, to be $\Sigma_{i \in A} D(i)$.

Definition 2 Let D and D' be two distributions over a set S. Let $d_A(D,D') = |D(A) - D'(A)|$. The distance of D and D' is defined as $d(D,D') = \max_{A \subseteq S} d_A(D,D')$.

Note that $d(D,D') = \Sigma_{i \in S} |D(i) - D'(i)|$, which is a useful way for calculating d(D,D').

Definition 3 Let D and D' be two distributions. They are statistically indistinguishable, if d(D,D') is negligible. They are computationally indistinguishable, if $d_A(D,D')$ is negligible for any subset A decided by a polynomial-size circuit.

An important cryptographic primitive is the 1-out-of-2 oblivious transfer, denoted as 1-OT-2. There are several variants which are all equivalent, and the one most suited for us is the following string version of 1-OT-2. Let F be a set.

Definition 4 An 1-OT-2 protocol has two parties, Sender who has input $(x_0, x_1) \in F^2$ and Chooser who has a choice $c \in \{0, 1\}$. The protocol is correct if the Sender learns x_c for any (x_0, x_1) and c. The protocol is secure if both conditions below are satisfied for any (x_0, x_1) and c:

Chooser cannot distinguish the distribution of Sender's messages from that induced by Sender having a different value of $x_{1-c.}$

Sender cannot distinguish the distributions of Chooser's messages induced by c and $1 - c$.

Definition 5 A protocol for Oblivious Polynomial Evaluation has two parties, Alice who has a polynomial P over some finite field F and Bob who has an input $x_* \in F$. An OPE protocol is correct if Bob learns $P(x^*)$ for any x_* and P. It is secure if both conditions below are satisfied for any x_* and P:

Alice cannot distinguish the distribution of Bob's messages from that induced by Bob having a different x_*'.

Bob cannot distinguish the distribution of Alice's messages from that induced by Alice having a different $P'(x_*) = P(x_*)$.

A.2. Requirements of a Private OPE Protocol

OPE requires privacy for both receiver and sender. Namely, in an OPE protocol neither party learns anything more than is defined by the OPE functionality. The strongest way of formalizing this notion and ensuring simple composition of the protocols is through the definition of secure two-party computation (Goldreich, 2004) and studies on universal composition. Yet, this definition is somewhat complex, while there are a significant number of requisitions that do not need the full power of the general definition and might utilize non-ideal protocols. We, along these lines, like to utilize a loose definition for OPE, which guarantees security for both parties but does not need the sender to confer to its enter (i.e., to dedicate to the

polynomial P). We call this definition private computation. The definition of private computation is important moreover to the instance of malicious parties (and is in this way stronger than a definition for the semi-honest case just). It safeguards the security of the customers, however does not need one to mimic the joint dissemination of the view of a malicious sender and the yield of a honest collector, as is needed by the general definition of secure computation.

The necessities of a private OPE protocol could be separated into effectiveness, recipient protection, and server security. Let us first characterize the proposed lands freely then after that describe a private OPE protocol as a protocol fulfilling the aforementioned definitions. In the definitions, the running time of polynomial time algorithms is polynomial in the span of their inputs, and additionally in the log |F|, where F is the field in which the polynomial P is described, and in a security parameter k. The length of representations of components in F should be polynomial in the security parameter forasmuch as generally the cryptographic operations may be insecure given antagonists with poly-log |F| running time.) We do not need in the definitions themselves that the number of public-key operations is free of F. To disentangle the documentation we in addition discard any reference to assistant inputs.

We claim that this relaxation is justified by efficiency considerations, in particular when constructing specific OPE protocols rather than black-box reductions of OPE to other primitives.

Moreover, the definition of private computation is standard for identified primitives for example oblivious transfer or private informative content recovery. Note additionally that the definition of private computation is proportionate to the definition of secure computation on account of semi-honest parties. Besides, we make arrangements for a collector-sender (i.e. client-server) situation, where one

exclusive party, the recipient, has a yield in the protocol. Along these lines, the two definitions are proportional regarding a malicious client, as there is no issue of mimicking the joint circulation of the client's see and the server's yield.

Definition 6 (Correctness, or Functionality) At the end of the protocol the receiver obtains the output of the OPE functionality, namely P(x).

The definition of the receiver's privacy is simplified by the fact that the sender gets no output. It is as follows:

Definition 7 (Receiver's privacy – indistinguishability) For any probabilistic polynomial time device performing the sender's aspect, for any x and x' in F, the opinions that it recognizes in situation the receiver's feedback is x and in situation the receiver's feedback is x' are computationally (statistically) indistinguishable.

The definition of sender's protection is a spot trickier, subsequent to the collector gets some qualified data, and we have a desire to state that the recipient does not get more or special data than she might as well. We contrast the protocol with the ideal implementation. In the ideal implementation there is a Trusted Third Party, which gets the sender's polynomial P and the recipient's solicit x and gives P(x) to the recipient. The security prerequisite is that the protocol does not hole to the receiver more qualified data than in the ideal implementation.

Definition 8 (Sender's security – comparison with the ideal model) For every probabilistic polynomial-time device A replacing the device, there is a probabilistic polynomial-time device A' that performs the receiver's part in the perfect execution, such that the perspective of A and the outcome of A' are computationally indistinguishable.

Definition 9 (Privacy Preserving Protocol) A two-party protocol satisfying Definitions 6, 7 and 8.

Observe that the definition of recipient protection does not block the sender from cheating by utilizing a polynomial of degree higher than the degree of P (and consequently it may not be plausible to concentrate from the sender a degree k polynomial.) We do not need that the sender be submitted to single polynomial, and that the receiver might verify that the worth she appropriates relates to this polynomial. Our project permits such cheating; yet, in a significant number of requisitions this is insignificant.

A.3. Protocol for OPE

We will present an OPE protocol in this section. Assume that both parties have agreed that polynomials are over a finite field F and have degrees at most d. The set of such polynomials can be identified with the set $T = F^{d+1}$ in a natural way. Suppose now Alice has a polynomial $P(x) = \sum_{i=0}^{d} a_i x^i \in T$ and Bob has $x_* \in F$.

To make the picture clear, we only discuss the case F = GF(p) for some prime p. The generalization of $GF(p^k)$ with k>1 is straightforward. Each coefficient a_i in the polynomial can be represented as $a_{ij} = \Sigma_{j\in[\log 2|F|]} a_{ij} 2^{j-1}$ with $a_{ij} \in \{0, 1\}$. For $i \in [d]$ and $j \in [\log 2|F|]$, let $v_{ij} = 2^{j-1} x_*^i$. Note that for each $i \in [d]$, $\Sigma_{j\in[\log 2|F|]} a_{ij} v_{ij} = a_i x_*^i$. The idea is to have Bob prepare $(v_{ij})_{j \in [\log 2|F|]}$ and have Alice get those v_{ij} with $a_{ij} = 1$, in some secret way. This is achieved by having Bob prepare the pair $(r_{ij}, v_{ij}+r_{ij})$ for a random noise r_{ij}, and having Alice get what she wants via 1-OT-2. Note that what Alice obtains is $a_{ij}v_{ij}+r_{ij}$.

Protocol 1

1. Bob prepares d[log2|F|] pairs $(r_{ij}, v_{ij}+r_{ij})$, $i \in [d], j \in [\log 2|F|]$, with each r_{ij} chosen randomly from F.
2. For each pair $(r_{ij}, v_{ij}+r_{ij})$, Alice runs an independent 1-OT-2 with Bob to get r_{ij} if $a_{ij} = 0$ and $v_{ij}+r_{ij}$ otherwise.
3. Alice sends to Bob the sum of a_0 and those d[log2|F|] values she got. Bob subtracts $\Sigma_{i,j} r_{ij}$ from it to obtain $P(x_*)$.

Lemma 4 Protocol 1 is correct when parties are semi-honest.

Proof The sum Bob obtains in Step 3 is $a_0 + \Sigma_i \Sigma_j (a_{ij} v_{ij} + r_{ij}) = P(x_*) + \Sigma_{i,j} r_{ij}$.

Lemma 5 Protocol 1 is secure when parties are semi-honest.

Proof First, we prove Alice's security. Suppose P and P' are two distinct polynomials with $P(x_*) = P'(x_*) = y_*$. According to Lemma 2.1, it suffices to show that for any fixed r_{ij} , Alice's respective message distributions D and D' induced by P and P' are indistinguishable. Note that the last message from Alice is $y_* + \Sigma_{i,j} r_{ij}$ for both P and P' can be ignored. So we focus on Alice's d[log2|F|] messages from the d[log2|F|] independent executions of OT's. For $0 \leq k \leq d[\log 2|F|]$, let D_k denote the distribution with the first k messages from D and the remaining messages from D'.

Posit that there exists a distinguisher C for D and D'. A standard contention show that C can also distinguish D_{s-1} and D_s for some s. Note that Alice must select different elements from that pair in the s'th OT, as otherwise the two circulations are indistinguishable.

Then one can break Chooser's security in 1-OT-2 when Sender has this input, because with Chooser's messages for different choices replacing the s'th message of D_{s-1}, we get exactly D_{s-1} and D_s, which can be distinguished by C. As 1-OT-2 is assumed to be secure, D and D' are indistinguishable, and Alice is secure.

Next, we prove Bob's security. Note that Bob sends dm messages to Alice for the d[log2|F|] independent executions of OT's. Let $x_* \neq x_*$'. Let E and E' be Bob's respective message distributions, and let E_k denote the distribution with the first k messages from E and the remaining messages from E'.

Suppose a distinguisher for E and E' exists. Then it can also distinguish E_{s-1} and E_s for some s. The pairs in that s'th OT have the forms (r, v+r) and (r', v'+r'), for some fixed v and v' and for random r and r'.

Alice's polynomial is fixed, so which element to choose in that s'th OT is also fixed. Suppose Alice chooses the first one in that pair. Then according to Lemma 1, there is a fixed r_0 such that E_{s-1} conditioned on Bob having (r_0, v+r_0) and E_s conditioned on Bob having (r_0, v'+r_0) are distinguishable. Similarly as before, one can distinguish Sender's messages when Sender has (r_0, v+r_0) and (r_0, v'+r_0) respectively and Chooser selects the first element, which violates Sender's security in 1-OT-2.

The case when Alice chooses the second one in that pair can be argued similarly, by noticing that the distribution (r, v+r) and the distribution (−v+r, r) are identical. As 1-OT-2 is assumed to be secure, so is Bob**.**

Theorem 4 Protocol 1 is correct and secure when parties are semi-honest.

Proof Note that only dm invocations of 1-OT-2 are required and they can be done concurrently. If 1-OT-2 can be carried out in one round, Protocol 1 runs in one round. Also observe that if 1-OT-2 can achieve perfect security for Chooser, then Protocol 1 is perfectly secure for Alice, in the information-theoretical sense.

REFERENCES

Agrawal, R., Srikant, R. (2000). Privacy-preserving Data Mining. Proceedings of the 2000 ACM SIGMOD on Management of Data, 439–450.

Agrawal, R., Evfimievski, A., Srikant, R. (2003). SIGMOD '03 Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, 86-97.

Amirbekyan, A., Estivill-Castro, V. (2009). Practical protocol for Yao's millionaires problem enables secure multiparty computation of metrics and efficient privacy-preserving k-NN for large data sets. Knowledge and Information Systems, volume 21, 327-363.

Atallah M. J., Du W. (2001). Secure Multiparty Computational Geometry. In Proceedings of 7th International Workshop on Algorithms and Data Structures, 165-179.

Bayardo R., Agrawal R. (2005). Data Privacy Through Optimal k-Anonymization. In Proceedings the 21st International Conference on Data Engineering, 217-228.

Benaloh, J. (1994). Dense probabilistic encryption. Proceedings of the Workshop on Selected Areas of Cryptography, 120-128.

Benenson Z. (2006). TrustedPals: Secure Multiparty Computation Implemented with Smart Cards. In Proc. of the 11th European Symposium

on Research in Computer Security (ESORICS 2006), 306-314.

Bleichenbacher D., Nguyen P. (2000). Noisy polynomial interpolation and noisy chinese remaindering. In EURO-CRYPT 2000, 53−69.

Cachin, C., Micali, S., and Stadler, M. (1999). Computationally private information retrieval with polylogarithmic communication. Advances in Cryptology: EUROCRYPT '99, Lecture Notes in Computer Science, 1592, 402–414.

Canetti. R. (2000). Security and Composition of Multiparty Cryptographic Protocols. Journal of Cryptology, 13(1), 143–202.

Chang Y., Lu C. (2001). Oblivious polynomial evaluation and oblivious neural learning, Theoretical Computer Science, 369-384.

Chaum D., Crepeau C., Damgard I.. (1988) Multiparty unconditionally secure protocols. In Proceedings of the 20th ACM Symposium on the Theory of Computing.ACM Press, 11–19.

Chor, B., Kushilevitz, E., Goldreich, O., Sudan, M. (1995). Private information retrieval. Proceedings of the 36th Annual IEEE Symposium on Foundations of Computer Science, 41–50.

Chor, B. and Gilboa, N. (1997). Computationally private information retrieval. Proceedings of 29th Annual ACM Symposium on Theory of Computing, 304–313.

Clifton C., Kantarcioglu M., Vaidya J. (2004). Tools for privacy preserving distributed data mining. ACM SIGKDD Explorations Newsletter, 4(2), 28-34.

Crepeau C., Salvail L. (1995) Oblivious Verification of Common String. CWI Quarterly special issue for the Crypto Course 10th Anniversary, Vol. 8, N 2, 97-109.

Dhillon I.S., Modha D.S. (1999). A data-clustering algorithm on distributed memory multiprocessors. In Proceedings of Large-scale Parallel KDD Systems Workshop (ACM SIGKDD), 245-260.

Di-Crescenzo G., Ishai, Y., and Ostrovsky, R. (1998). Universal service-providers for database private information retrieval. Proceedings of the 17th Annual ACM Symposium on Principles of Distributed Computing, 91–100.

Dowd J., Xu S., Zhang W. (2006). Privacy-Preserving Decision Tree Mining Based on Random Substitutions. ETRICS 2006, 145-159.

Du W., Atallah M. J. (2001). Secure Multiparty Computation Problems and their Applications: Review and Open Problems. In New Security Paradigms Workshop, 11-20.

Du W., Atallah M. J. (2002). A Practical Approach to Solve Secure Multiparty Computation Problems. In New Security Paradigms Workshop, 127-135.

Du W., Zhan Z., (2002). Building Decision Tree Classifier on Private Data. CRPIT '14 Proceedings of the IEEE international conference on Privacy, security and data mining - Volume 14, 1-8.

Duda R.O., Hart P.E. and Stork D.G. (2001) Pattern Classification. New York: John Wiley & Sons, 2001, ISBN: 0-471-05669-3

Emekci F., Sahin O.D., Agrawal D., El Abbadi A. (2007). Privacy preserving decision tree learning over multiple parties. Data & Knowledge Engineering 63, 348-361.

Evfimievski A., Srikant R., Agrawal R., Gehrke J. (2002). Privacy preserving mining of association rules. In Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 217–228.

Evfimievski A., Gehrke J., Srikant R. (2003). Limiting privacy breaches in privacy preserving data mining, Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 211-222.

Fagin R., Naor M. and Winkler P. (1996). Comparing Information Without Leaking It. Communications of the ACM 39, 77-85.

Fang W., Yang B. (2008). Privacy Preserving Decision Tree Learning Over Vertically Partitioned Data. ACM Transactions on Knowledge Discovery from Data (TKDD) Volume 2 Issue 3, October 2008 Article No. 14.

Franconi L., Merola G. (2003). Implementing Statistical Disclosure Control for Aggregated Data Released Via Remote Access, Working Paper No. 30, United Nations Statistical Commission and European Commission.

Freedman M., Nissim, K., Pinkas, B. (2004). Efficient private matching and set intersection. Advances in Cryptology Eurocrypt ‘2004 Proceedings, 1–19.

Fung B., Wang K., Yu P. (2005). Top-down Specialization for Information

and Privacy Preservation, In Proceedings of the 21st IEEE International Conference on Data Engineering, 205-216.

Gertner, Y., Ishai, Y., Kushilevitz, E., and Malkin, T. (1998). Protecting data privacy in information retrieval schemes. Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98, 151–160.

Gilboa N. (1999). Two party RSA Key Generation. In CRYPTO 1999, 116−129.

Goethals B., Laur S., Lipmaa H. and Mielikäinen T. (2004). On Private Scalar Product Computation for Privacy-Preserving Data Mining. The 7th Annual International Conference in Information Security and Cryptology (ICISC 2004), volume 3506 of Lecture Notes in Computer Science, 104-120.

Goldreich O., Micali S. and Wigderson A. (1987). How to play any mental game - a completeness theorem for protocols with honest majority. In 19th Symposium on Theory of Computer Science, 218–229.

Goldreich O. (2004). Foundations of Cryptography: Volume 2, Basic Applications. Cambridge University Press, ISBN 0-521-83084-2.

Ioannidis, I., Grama, A. (2003). An efficient protocol for Yao's millionaires problem. Proceedings of the 36th Hawaii International Conference on System Sciences, 6–9.

Ishai, Y. and Kushilevitz, E. (1999). Improved upper bounds on information-theoretic private information retrieval. Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing, 79–88.

Johnson S. C. (1967). Hierarchical Clustering Schemes. *Psychometrika*, 2:241-254

Jha S., Kruger L. and McDaniel P. (2005). Privacy Preserving Clustering. 10th European Symposium on Research in Computer Security (ESORICS), 397-417.

Kargupta H., Huang W., Sivakumar K., Johnson E. (2001). Distributed clustering using collective principal component analysis. Knowledge and Information Systems, 3(4), 405–421.

Kantarcioglu M. and Clifton C. (2004). Privacy Preserving Data Mining of Association Rules on Horizontally Partitioned Data, Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 639-644.

Kantarcioglu M., Vaidya J. (2003). Privacy Preserving Naive Bayes Classifier for Horizontally Partitioned Data, In Proceedings of the Workshop on Privacy Preserving Data Mining held in association with The Third IEEE International Conference on Data Mining.

Katz J. (2007). Universally Composable Multiparty Computation Using Tamper-Proof Hardware. Advances in Cryptology - EUROCRYPT '07, LNCS 4515, Springer-Verlag, 115-128.

Kim J. (1986). A method for limiting disclosure in microdata based on random noise and transformation. In Proceedings of the American Statistical Association on Survey Research Methods, 370–374.

Klusch M., Lodi S., Moro G. (2003). Distributed clustering based on sampling local density estimates. In Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI 2003), 485–

490.

Kushilevitz, E., Ostrovsky, R. (1997). Replication is not needed: Single database, computationally private information retrieval. Proceedings of the 38th Annual IEEE Computer Society Conference on Foundation of Computer Science, 20–22.

Lefevre K., Dewittd J., Ramakrishnan R. (2005). Incognito: Efficient Full-Domain k-Anonymity, In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, 49-60.

Li B., Sarkar S. (2006). A tree-based data perturbation approach for privacy-preserving data mining. IEEE Transactions on Knowledge and Data Engineering (TKDE), 18(9), 1278–1283.

Lindell Y., Pinkas B. (2000). Privacy preserving data mining. In Advances in Cryptology (Crypto 2000), 36–54.

Machanavajjhala A., Gehrke J., Kifer D. (2007). l-Diversity: Privacy Beyond k-Anonymity. ACM Transactions on Knowledge Discovery from Data, 24-35.

Merugu S., Ghosh J. (2003). Privacy-preserving distributed clustering using generative models. In Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003), 211–218, 2003.

Naccache, D., Stern, J. (1998). A new public key cryptosystem based on higher residues. Proceedings of the 5th ACM Conference on Computer and Communications Security, CCS ‘98, 59–66.

Naor M., Pinkas B. (1999). Oblivious transfer and polynomial evaluation. In

31st Symposium on Theory of Computer Science, 245–254.

Naor M., Nissim K. (2001). Communication preserving protocols for secure function evaluation. STOC '01 Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, 590 - 599.

Naor M., Pinkas B. (2005). Computationally Secure Oblivious Transfer. Journal of Cryptology, Vol. 18, No. 1, 1-35.

Oleshchuk,V., Zadorozhny,V. (2007). Secure Multiparty Computations and Privacy Preservation: Results and Open Problems. Telektronikk: Telenor's Journal of Technology, v. 103, N2, 2007.

Oliveira S., Zaiane O.R. (2003). Privacy preserving clustering by data transformation. In XVIII Simposio Brasileiro de Bancos de Dados, 304–318.

Özarar M., Özgit A. (2007). Secure Multiparty Overall Mean Computation via Oblivious Polynomial Evaluation. Proceedings of First International Conference on Security of Information and Networks (SIN 2007), 84-95.

Özarar M., Özgit A. (2008). Secure Homogeneous Matrix Algebra with Oblivious Polynomial Evaluation. Proceedings of Third Information Security and Cryptology Conference (ISCTURKEY), 157-163.

Özarar M., Özgit A. (2011). Privacy Preserving Hierarchical Agglomerative Document Clustering. Technical Report, METU.

Paillier P. (1999). Public-key cryptosystems based on composite degree residuosity classes. In Proceedings of Advances in Cryptology (EUROCRYPT'99).

Pinkas B. (2003). Cryptographic Techniques for Privacy-Preserving Data Mining. SIGKDD Explorations, the newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining, 4 (2), 12-19.

Polat H., Du W. (2005). Privacy-Preserving Collaborative Filtering. International Journal of Electronic Commerce, 9 (4), 9-35.

Rizvi S. J., Harista J. R. (2002). Maintaining data privacy in association rule mining. In Proceedings of 28th International Conference on Very Large Data Bases (VLDB), 682-693.

Shamir A. (1979). How to share a secret. Communications of the ACM 22 (11), 612–613.

Sweeney L. (2002). k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10(5), 557–570.

Vaidya J., Clifton C. (2003). Leveraging the "Multi" in Secure Multiparty Computation. In Proceedings of the Workshop on Privacy in the Electronic Society, 53-59.

Vaidya J., Clifton C. (2003). Privacy-preserving k-means clustering over vertically partitioned data. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 206–215.

Vaidya J., Clifton C. (2004). Privacy Preserving Naive Bayes Classifier for Vertically Partitioned Data. In Proceedings of the 2004 SIAM International Conference on Data Mining, 522-526.

Vanishree H., George K. (2009). A Novel Unconditionally Secure Oblivious Polynomial Evaluation Protocol. Proceedings of 2009 International Workshop on Information Security and Application (IWISA 2009).

Verykios, V. S., Bertino, E., Parasiliti, L., Favino, I. N., Saygin, Y., Theodoridis, Y. (2004). State-of-the-art in Privacy preserving Data Mining. SIGMOD, 33(1), 50-57.

Yao A.C. (1986). How to generate and exchange secrets. SFCS '86 Proceedings of the 27th Annual Symposium on Foundations of Computer Science, 162-167.

CURRICULUM VITAE

Personal Information

Surname, Name: Özarar, Mert
Nationality: Turkish
Date and Place of Birth: 07 October 1979, Ankara
Marital Status: Single
Phone: +90 533 776 46 56
e-mail: mert.ozarar@gmail.com

Education

Degree	Institution	Year of Graduation
MS	METU Cryptography	2004
MS	METU Computer Engineering	2003
BS	METU Computer Engineering	2002
BS	METU Mathematics	2001
High School	TED Ankara College	1996

Work Experience

Year	Place	Enrollment
2004 – …	TÜRKTRUST Inc.	Project Manager
2001 – 2004	METU Computer Engineering	Research Assistant

Foreign Languages

Advanced English

Intermediate German

Beginner Russian

Developed Projects in Brief

AVEA Mobile Signature: Mobile Signature Services Platform, Software Eng., DB Admin (http://www.aveamobilimza.com)

PLATAN: PKI Certificate Authority Software (Java and web-based modules), Software Eng.

TSSYY: PKI Certificate Authority Software (Java application with GUI), Software Eng.

ARNICA: Software Library for digital signatures (Java, C++), Software Eng. (http://www.turktrust.com.tr/en/yazilim-arnika.html)

PALMA: Smart Card Management Software for certificate owners (C++, DLL) (http://www.turktrust.com.tr/en/yazilim-palma.html)

TILIA: Secure Electronic Signature Creation and Verification Software (C#. NET) (http://www.turktrust.com.tr/en/yazilim-tilia.html)

Central Bank of Qatar PKI System: Establishment of Certificate Management Software, Hardware Security Modules and Timestamp Server

e-Invoice Framework Development Project (together with Innova): Developed models for a technological framework to conduct e-invoice and e-billing transformation projects (72 manXmonths)

Various products of digital signature, time-stamping, web-based PKI authentication etc., Software Eng.

Integration and Data Transfer of DB systems, performing required adjustments to the structure

Publications

1. ”Enhancement of Multi-Level Grid File Structure for Fuzzy-Spatial Querying”, Technical Report, Computer Eng. Dept., METU, February 2002

2. ”Prediction of Protein Subcellular Localization Based on Primary Sequence Data”, Lecture Notes on Computer Science 2869, pp 611-619, November 2003

3. ”Birincil Dizi Veri Temelli Protein Hücre İçi Yer Belirleme Tahmini”, Sinyal İşleme ve İletişim Uygulamaları Konferansı, Kuşadası, Turkey, April 2004

4. ”BooleCrypt: Boole Tipi Fonksiyon Özelliklerini Değerlendirme Amaçlı bir Yazılım Kütüphanesi”, Abant Kriptografi Günleri, Bolu, Turkey, July 2004

5. ”Güvenlik Mekanizmalarında Kriptografik Akıllı Kartlar Üzerine”, Ulusal Elektronik İmza Sempozyumu, Ankara, Turkey, December 2006

6. ”Secure Multiparty Overall Mean Computation via Oblivious Polynomial Evaluation”, International Conference on Security of Information and Networks, Magosa, Cyprus, May 2007

7. ”Güvenli Elektronik Arşivleme”, International Security Conference, ISC Turkey, Ankara, Turkey, November 2007

8. ”Secure Homogeneous Matrix Algebra via OPE”, Proc. of 3rd Information Security and Cryptology Conference", (2008), s.157-163. Ankara, Turkey, November 2007

9. ”Secure Multiparty Computation via Oblivious Polynomial Evaluation”, Book Chapter, Theory and Practice of Cryptography Solutions for Secure Information Systems, 2012, IGI Global

Personal Interests

European Soccer, NBA, Smartphone Applications, Disc jockeying, Politics