MINISTRY OF
EDUCATION AND TRAINING

VIETNAM ACADEMY OF
SIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SIENCE AND TECHNOLOGY

Vu Duy Hien

DEVELOPING EFFICIENT AND SECURE MULTI-PARTY SUM COMPUTATION PROTOCOLS AND THEIR APPLICATIONS

SUMMARY OF DISSERTATION ON INFORMATION SYSTEM
Code: 9 48 01 04

Hanoi – 2024

The dissertation is completed at: Graduate University of Science and Technology, Vietnam Academy of Science and Technology

Supervisors:
1. Supervisor 1: Prof. Dr. Ho Tu Bao, Vietnam Institute for Advanced Study in Mathematics
2. Supervisor 2: Assoc. Prof. Dr. Luong The Dung, Academy of Cryptography Technologies

Referee 1:

Referee 2:

Referee 3:

The dissertation will be examined by Examination Board of Graduate University of Science and Technology, Vietnam Academy of Science and Technology at…………….. (date: , year:)

This dissertation can be found at:
1) Graduate University of Science and Technology Library
2) National Library of Vietnam

CONTENTS

INTRODUCTION 3

1 OVERVIEW OF SECURE MULTI-PARTY SUM COMPUTATION 4

1.1 Background of secure multi-party computation 4

1.1.1 Introduction 4

1.1.2 Definition of security 5

1.1.3 Cryptographic preliminaries 5

1.2 Secure multi-party sum computation problem 5

1.2.1 Problem formulation 5

1.2.2 Related work 5

1.3 Conclusion 5

2 PROPOSING EFFICIENT SECURE MULTI-PARTY SUM COMPUTATION PROTOCOLS 6

2.1 Analysis of typical secure multi-party sum computation protocols 6

2.1.1 Secure multi-party sum computation protocol of Urabe et al. 6

2.1.2 Secure multi-party sum computation protocol of Hao et al., 2010 in an electronic voting system 6

2.1.3 Privacy-preserving frequency computation protocol in fully distributed setting of Yang et al. 7

2.1.4 Further discussion 7

2.2 Proposed secure multi-party sum computation protocols . . . 7

2.2.1 Privacy-preserving frequency computation protocol based on elliptic curve ElGamal cryptosystem 7

2.2.2 An efficient approach for secure multi-party sum computation without pre-establishing secure/authenticated channels 10

2.2.3 Secure multi-sum computation protocol 12

2.3 Conclusion 16

3 DEVELOPING NEW SOLUTIONS BASED ON SECURE MULTI-PARTY SUM COMPUTATION PROTOCOLS FOR PRACTICAL PROBLEMS 18
3.1 An efficient solution for the secure electronic voting scheme without pre-establishing authenticated channel 18
3.1.1 Introduction 18
3.1.2 A secure end-to-end electronic voting scheme 18
3.1.3 Security analysis 18
3.1.4 Experimental evaluation 19
3.2 An efficient and practical solution for privacy-preserving Naive Bayes classification in the horizontally distributed data setting 20
3.2.1 Introduction 20
3.2.2 New privacy-preserving Naive Bayes classifier for the horizontally distributed data setting 21
3.2.3 Privacy analysis 21
3.2.4 Accuracy analysis 21
3.2.5 Experimental evaluation 21
3.3 Conclusion 22
CONCLUSION 24

INTRODUCTION

A. Motivation

In essence, the birth of SECURE MULTI-PARTY COMPUTATION area (SMC) is based on the distributed computing field and privacy constraints of data owners. This thesis has investigated the secure multi-party sum computation problem (SMS) where it is assumed that where there are some parties, in which each party owns a private value as his/her input, and the parties wish to obtain the sum of all inputs but they reveal nothing about their inputs beyond the sum value.

Currently, the existing solutions for the SMS problem have low level of security, poor performance or trade-off between the security and performance properties. Thus, developing SMS protocols that have both high security level and good performance is perfectly necessary.

B. Research objectives

The research objectives of this thesis include:

- Designing efficient and secure multi-party sum computation protocols that are both secure and efficient.
- Developing solutions for practical problems based on new protocols.

C. Main contributions

The contributions of this thesis are summarized as follows:

- The first contribution is to propose three novel SMS protocols.
- The second contribution is to develop new solutions based on the proposed protocols for practical applications.

D. Thesis organization

The rest of this thesis is organized as follows:

- Chapter 1 provides a general background. After that, this chapter of the thesis comprehensively reviews related work to identify research gap and new directions.
- Chapter 2 analyzes typical SMS protocols in detail. Based on the analysis result, this chapter proposes three new protocols.
- Chapter 3 develops new solutions based on the proposed SMS protocols for two practical applications.

CHAPTER 1. OVERVIEW OF SECURE MULTI-PARTY SUM COMPUTATION

In this chapter, the thesis first provides a background of secure multi-party computation field and the secure multi-party sum computation problem. Next, the previous work is meticulously analyzed.

1.1. Background of secure multi-party computation

1.1.1. Introduction

- *Input*: there are n parties where each participant i owns a private input v_i.
- *Output*: the participants obtain the result $f(v_1,...,v_n)$ of the specific function f over the inputs $(v_1,...,v_n)$, and each party reveals nothing about his/her input but the output result.

Here, it needs to be expressed that the ”secure” concept means the two following constraints:

- The correctness of the function’s output is guaranteed.
- Each party’s input is privately kept by himself/herself.

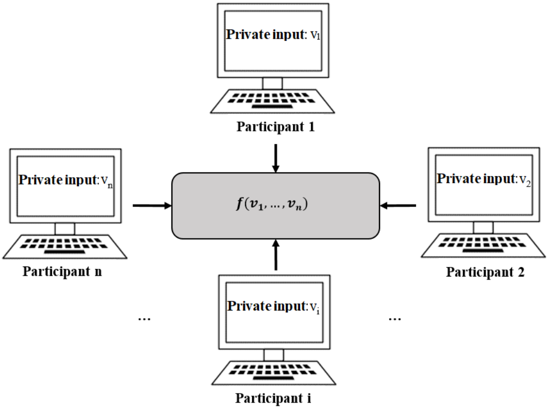


Figure 1.1: The distributed computing model in a secure manner

Generally, the security property of a SMC protocol depends on the adversary model including type of adversary, type of communication channels, and capabilities of adversary.

It can be seen that there are many practical problems related to SMC.

1.1.2. Definition of security

The standard security definition of privacy with respect to the semi-honest model using public channels in the book[1] is used in this thesis. Besides, there is a composition theorem often used to construct SMC protocols in the semi-honest model.

1.1.3. Cryptographic preliminaries

This thesis is based on the following cryptographic preliminaries: hard discrete logarithm problems over standard cyclic groups, ElGamal public-key cryptosystem - a homomorphic encryption.

1.2. Secure multi-party sum computation problem

1.2.1. Problem formulation

- *Input*: there are n parties, where each party i owns a private value v_i.
- *Output*: the parties obtain the sum $f = v_1 + ... + v_n$, and each party reveals nothing about his/her input.

1.2.2. Related work

Up to now, SMS protocols have been based on two approaches: *non-cryptographic* and *cryptographic* ones. For convenience, it is assumed that there are n parties joining a SMS protocol execution, in which the i^{th} party and his/her private input value are correspondingly denoted as U_i and v_i.

The thesis has reviewed typical SMS, privacy-preserving frequency computation, and secure multi-sum computation protocols. It is necessary and significant to design SMS protocols that should be not only secure against malicious adversary but also efficient in real-life applications.

1.3. Conclusion

In this chapter, the thesis has represented background and preliminaries in the SMC field. The thesis then formulated the SMS problem and pointed its importance in practice. To find out potential research issues for the SMS problem, the related work has been fully analyzed.

[1] Oded Goldreich. Basic Applications. In Foundations of Cryptography, volume II. Cambridge University Press, 2004

CHAPTER 2. PROPOSING EFFICIENT SECURE MULTI-PARTY SUM COMPUTATION PROTOCOLS

This chapter first meticulously analyzes typical secure multi-party sum computation protocols. Based on the analysis result, this chapter proposed three new secure multi-party sum computation protocols having both high security level and good performance.

2.1. Analysis of typical secure multi-party sum computation protocols

This section fully analyzes the most popular SMS protocols related to the thesis. It is also recalled that there are n parties $\{U_1, U_2, ..., U_n\}$ where the i^{th} party U_i owns a private value v_i $(i = 1, 2, ..., n)$. The aim of SMS protocols is to correctly compute the sum value $V = \sum_{i=1}^{n} v_i$ while all parties do not reveal their private values.

2.1.1. Secure multi-party sum computation protocol of Urabe et al.

Security analysis: this protocol is secure and it can protect the privacy of each honest party against $(n-2)$ parties colluding together.

Performance analysis: the communication costs of this protocol are high, and it is impractical to require each tuple of parties to setup a communication channel together

2.1.2. Secure multi-party sum computation protocol of Hao et al., 2010 in an electronic voting system

Security analysis: this protocol is secure and it can protect the privacy of each honest party against $(n-2)$ parties colluding together.

Performance analysis: the computational complexity required for the server is low, but each voter must spend high cost performing his/her tasks. The total messages transferred in this protocol is quite large (i.e. $(n^2 + n)$).

2.1.2.4. Variants of the protocol of Hao et al., 2010

To decrease each voter's computational complexity, the protocol of Hao et al.[1], 2014 was proposed for the DRE-based electronic voting sys-

[1] Feng Hao, Matthew N. Kreeger, Brian Randell, Dylan Clarke, Siamak F. Shahandashti, and Peter Hyun-Jeen Lee. Every Vote Counts: Ensuring Integrity in Large-Scale Electronic Voting. USENIX Journal of Election Technology and Systems, 2:1–25, 2014

tem. Moreover, there exists another variant in Hao et al.[2], 2018 where the computations are performed over an elliptic curve.

2.1.3. Privacy-preserving frequency computation protocol in fully distributed setting of Yang et al.

Security analysis: this protocol is secure and it can protect the privacy of each honest party against $(n-2)$ parties colluding together.

Performance analysis: if the number of parties n is large, the computational cost of the miner will be expensive. The total of communication cost of the Yang et al.'s protocol is $6n\,|p|$ bits.

2.1.4. Further discussion

It can be seen that the protocols of Yang et al. and Hao et al. are the most remarkable ones among the typical `SMS` protocols.

2.2. Proposed secure multi-party sum computation protocols

This section proposes three novel `SMS` protocols having unique features. It is also recalled that all of the new proposals are based on the semi-honest model.

2.2.1. Privacy-preserving frequency computation protocol based on elliptic curve ElGamal cryptosystem

2.2.1.1. Introduction

In this section, the thesis propounds a privacy-preserving frequency computation protocol based on elliptic curve ElGamal cryptosystem. This proposal is related to **Publication** 1.

2.2.1.2. Privacy-preserving frequency computation protocol

Before the `PPFC` protocol starts, each user chooses two private keys $p_i, q_i \in [1, d-1]$ and computes the corresponding public keys $P_i = p_iG$, $Q_i = q_iG$. Then these public keys sent to the miner.

The proposed `PPFC` protocol consists of three main phases described in follows.

[2] Feng Hao, Dylan Clarke, Brian Randell, and Siamak F. Shahandashti. Verifable Classroom Voting in Practice. IEEE Security & Privacy, 16:72–81, 2018

Protocol 2.1: A privacy-preserving frequency computation protocol for fully distributed setting

Phase 1: Pre-computing

- Miner pre-computes the public values: $P = \sum_{i=1}^{n} P_i$, $Q = \sum_{i=1}^{n} Q_i$
- Miner → U_i: P, Q

Phase 2: Computing the messages

- U_i computes: $M_i = v_i G + q_i P - p_i Q$
- U_i → Miner: M_i

Phase 3: Secure frequency computation

- Miner computes: $M = \sum_{i=1}^{n} M_i$
- Miner runs Shanks' algorithm to find out v that satisfies $vG = M$

2.2.1.3. Security analysis

i. Proof of Correctness: if the miner finds out a value v, then v is the secure sum of all parties' private values.

ii. Privacy Analysis: it is proven that the proposed PPFC protocol in fully distributed setting protects each honest user's privacy in the case of $(n-2)$ parties colluding with the miner.

2.2.1.4. Performance evaluation

This section compares the proposed protocol with Hao et al.[3], 2018 and a variant of Yang et al.'s[4] protocol.

- *Experimental setting*

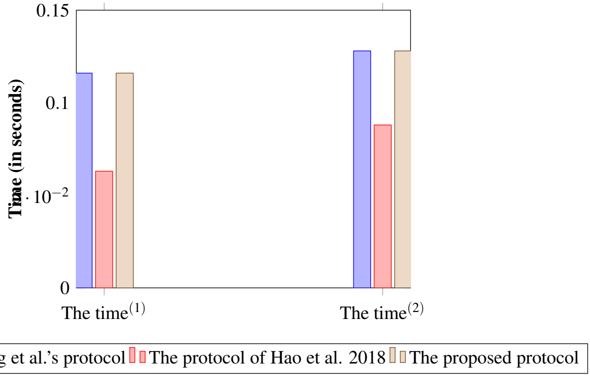
The experiments are implemented on the Lenovo Thinkpad X280 laptop with different numbers of users, from 10000 to 50000 and use the safe elliptic curve 25519.

- *Experimental results*

The running time of each user comparisons among three protocols are presented in Figure 2.1. It can be seen that that time in the new proposal and the variant of Yang et al.'s protocol is larger than in the protocol of Hao et al., 2018. However, the amount of differences are negligible.

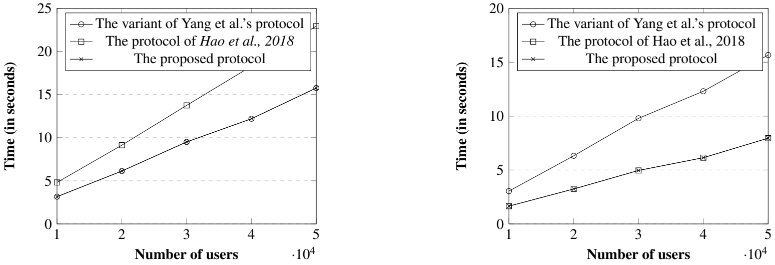
[3] Feng Hao, Dylan Clarke, Brian Randell, and Siamak F. Shahandashti. Verifable Classroom Voting in Practice. IEEE Security & Privacy, 16:72–81, 2018

[4] Zhiqiang Yang, Sheng Zhong, and Rebecca N. Wright. Privacy-Preserving Classification of Customer Data without Loss of Accuracy. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 92–102. SIAM, 2005



$^{(1)}$ The time for preparing the public keys, $^{(2)}$ The time for computing the messages

Figure 2.1: The running time of each user comparisons among the typical PPFC protocols.



(a) *The time for computing the public keys* (b) *The time for computing frequency*

Figure 2.2: *The time for the miner/the server comparisons among the typical PPFC protocols.*

According to the above comparisons in Figures 2.1, 2.2, and Table 2.1, it can be seen that the new solution has more advantages than the others. If the cost for establishing authenticated communication between the miner/the server and each user is considered, the communication cost and computational complexity of the protocol of Hao et al., 2018 would have significantly increased.

Table 2.1: The stored data volume of the miner comparisons among the typical PPFC protocols (in megabytes).

Number of parties Protocols	10000	20000	30000	40000	50000
The variant of Yang et al.'s protocol	2.4	4.9	7.3	9.8	12.2
The protocol of *Hao et al., 2018*	1.2	2.4	3.7	4.9	6.1
The new solution	1.8	3.7	5.5	7.3	9.2

2.2.2. An efficient approach for secure multi-party sum computation without pre-establishing secure/authenticated channels

2.2.2.1. Introduction

In this section, an efficient secure multi-party sum protocol without pre-establishing secure/authenticated channels is proposed. This proposal relates to **Publication** 3.

2.2.2.2. An efficient secure multi-party sum computation protocol without pre-establishing secure/authenticated channels

The new protocol requires the parameters as follows. Let $(\mathbb{G}, p, q, g)$ be standard public parameters. Each user U_i has already owned a private key x_i, the corresponding public key is X_i, and chooses a secret number y_i (the public value is Y_i). H is a secure hash function.

Before starting the protocol, each user U_i directly sends a message Y_i to the miner. Continuously, the miner pre-computes $X = \prod_{i=1}^{n} X_i$; $Y = \prod_{i=1}^{n} Y_i$. The miner then sends $M = (X \parallel Y)$ and the Schnorr signature of M to all users.

iv. The proposed protocol

Three main phases of the new protocol are summarized as follows.

Protocol 2.2: A secure n-parties sum protocol without pre-establishing secure/authenticated channels.

Phase 1: Data submission

- Each user verifies the miner's Schnorr signature on $M = (X \parallel Y)$
- Each user U_i computes: $P_i = \frac{g^{v_i} X^{y_i}}{Y^{x_i}}, r_i = Y_i, s_i \equiv y_i - x_i H(r_i \parallel P_i) \pmod q$
- Each user $U_i \to$ Miner: P_i, s_i

Phase 2: User authentication

- Miner computes: $\gamma_i = H(r_i \parallel P_i), r_i^{'} = g^{s_i} X_i^{\gamma_i}$
- Miner authenticates each user U_i by verifying the equation: $r_i \overset{?}{=} r_i^{'}$

Phase 3: Secure n-parties sum computation

- Miner computes: $K = \prod_{i=1}^{n} P_i$
- Miner executes **Shanks' algorithm** to find out V that satisfies $g^V = K$

2.2.2.3. Security analysis

As mentioned before, the new protocol is composed from two secure sub-protocols. Hence, the proposed protocol are secure.

2.2.2.4. Performance evaluation

This section compares the performance of the new solution with the typical protocols of Yang et al.[5] and Hao et al.[6], 2014, briefly named Yang's protocol and Hao2014's protocol, respectively.

- *Experimental setting*

The experiments are run on the Lenovo Thinkpad X280 laptop with an Intel core i5 with different numbers of users from 200000 to 1000000., the prime number p is 2048 bits length, and the length of prime number q is 256 bits.

- *Experimental results*

In summary, considering the above experimental results, it can be seen that the proposed protocol has more advantages than the others.

[5] Zhiqiang Yang, Sheng Zhong, and Rebecca N. Wright. Privacy-Preserving Classification of Customer Data without Loss of Accuracy. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 92–102. SIAM, 2005

[6] Feng Hao, Matthew N. Kreeger, Brian Randell, Dylan Clarke, Siamak F. Shahandashti, and Peter Hyun-Jeen Lee. Every Vote Counts: Ensuring Integrity in Large-Scale Electronic Voting. USENIX Journal of Election Technology and Systems, 2:1–25, 2014

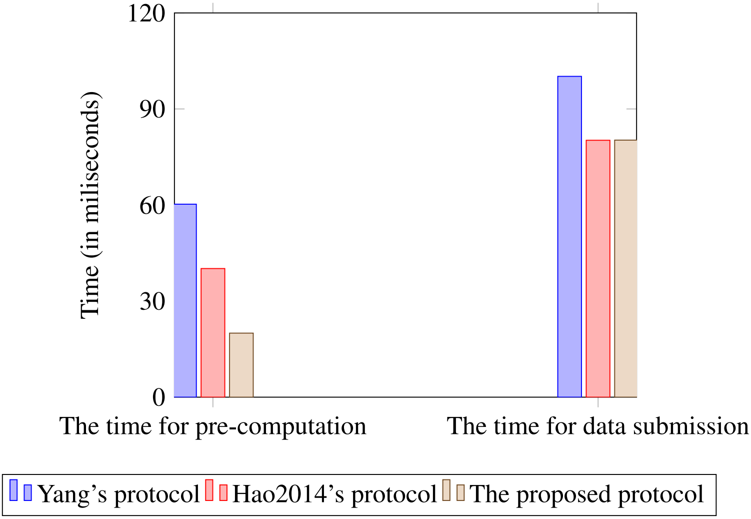
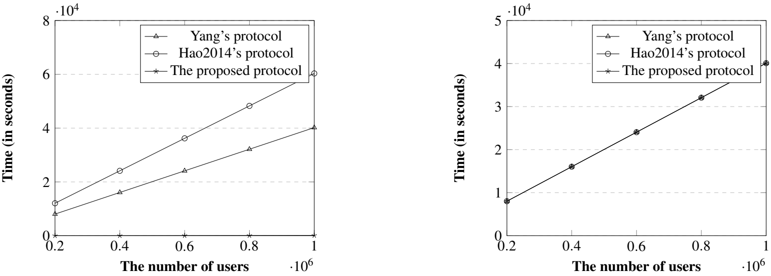


Figure 2.3: The running time of each user comparisons among the proposed protocol and the typical protocols.



(a) The time for pre-computating.

(b) *The time for authenticating users*

Figure 2.4: *The time for the miner pre-computing and authenticating users comparisons among the proposed protocol and the typical protocols.*

2.2.3. Secure multi-sum computation protocol

2.2.3.1. Introduction

This section presents an efficient secure multi-sum computation protocol that can securely compute multiple sum values only in one round of

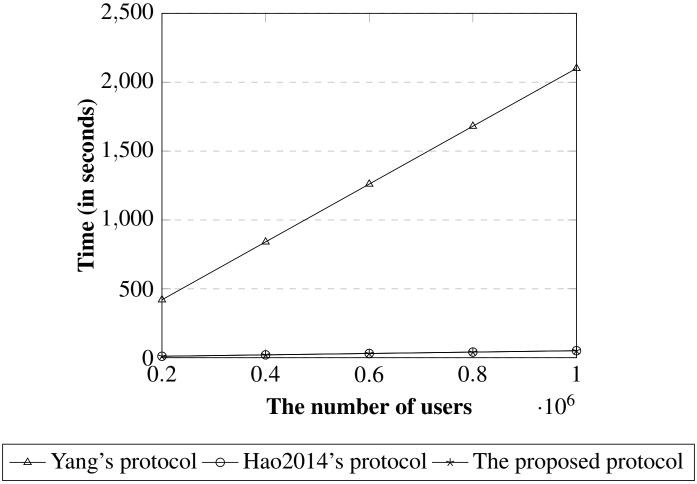


Figure 2.5: The time of the secure n-parties sum phase comparisons among the proposed protocol and the typical protocols.

Table 2.2: The stored data volume of the miner comparisons among the proposed protocol and the typical protocols (in megabytes).

Number of parties **Protocols**	**200000**	**400000**	**600000**	**800000**	**1000000**
Yang's protocol	215.1	430.3	645.4	860.6	1075.7
Hao2014's protocol	215.1	430.3	645.4	860.6	1075.7
The proposed protocol	107.6	215.1	322.7	430.3	537.9

computation. This proposal of the thesis is related to **Publication** 4.

2.2.3.2. The secure protocol for multi-sum computation

Let $(\mathbb{G}, p, q, g)$ be standard cryptographic parameters. The main stages of the new secure multi-sum computation protocol are presented as follows.

Protocol 2.3: A secure protocol for computing multi-sum in one round of computation

```
1. Phase 1: The users $U_i$ do
forall $i$ where $1 \le i \le np$ do
    forall $j$ where $1 \le j \le nk$ do
        $Prv_i^j = Random(1, q-1)$
        $Pub_i^j = g^{Prv_i^j}$
        Sends to Miner: $Pub_i^j$
    end
end

2. Phase 2: Miner does
forall $j$ where $1 \le j \le nk$ do
    $Pub^j = 1$
    forall $i$ where $1 \le i \le np$ do
        $Pub^j = Pub^j * Pub_i^j$
    end
    Sends to all $U_i$: $Pub^j$
end

3. Phase 3: The users $U_i$ do
forall $i$ where $1 \le i \le np$ do
    $j = 1$
    forall $t$ where $1 \le t \le nk - 1$ do
        forall $k$ where $t + 1 \le k \le nk$ do
            $p_i^j = g^{v_i^j}(Pub^t)^{Prv_i^k}(Pub^k)^{q-Prv_i^t}$
            if $j == ns$ then
                break
            else
                $j{+}{+}$
            end
        end
    end
    Sends to Miner: $p_i^j$
end

4. Phase 4: Miner does
forall $j$ where $1 \le j \le ns$ do
    $K^j = 1$
    forall $i$ where $1 \le i \le np$ do
        $K^j = K^j * p_i^j$
    end
end
Solves the problems $g^{Sum_j} = K^j$ ($j \in \{1, ..., ns\}$) by performing the brute-force algorithm once
```

Each party only needs to prepare nk tuples of keys with $nk = \left\lceil \frac{1}{2} + \sqrt{2ns + \frac{1}{4}} \right\rceil$.

2.2.3.3. Security analysis

i. Proof of Correctness

This section has showed that the proposed secure multi-sum protocol's correctness.

ii. Privacy analysis

It is proved that the proposed protocol securely protects honest parties' privacy in the semi-honest model under several necessary assumptions.

2.2.3.4. Performance evaluation

This section evaluates the new proposal and the solutions created by executing multiple times the protocol of Yang et al.[7], the protocol of Hao et al.[8], and the protocol in the third publication (denoted as Yang's-based solution, Hao2014's-based solution, and Pub3-based solution, respectively).

- *Experimental setting*

All compared solutions are implemented on the laptop with an Intel core i5 8250U with different tuples (number of users, number of sums), i.e. (1000, 500), (2000, 1000), (3000, 1500), (4000, 2000), (5000, 2500) denoted as $C1, C2, C3, C4, C5$, respectively. The prime number p is 2048 bits length, and the length of prime number q is 256 bits.

- *Experimental results*

Table 2.3: The running time for the miner to compute the sum values comparisons among the compared solutions (in seconds).

Number of users-Number of sums **Solutions**	**1000-500**	**2000-1000**	**3000-1500**	**4000-2000**	**5000-2500**
Yang's-based solution	823.142	3217.324	7299.532	13320.492	19310.021
Hao2014's-based solution	10.000	35.556	74.314	128.160	195.650
Pub3-based solution	10.005	35.004	74.102	128.533	195.647
The new solution	**7.104**	**27.588**	**62.085**	**110.262**	**172.094**

In summary, based on the experimental results, the new proposal is the most suitable solution for applications in practice when compared to the typical ones.

[7] Zhiqiang Yang, Sheng Zhong, and Rebecca N. Wright. Privacy-Preserving Classification of Customer Data without Loss of Accuracy. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 92–102. SIAM, 2005

[8] Feng Hao, Matthew N. Kreeger, Brian Randell, Dylan Clarke, Siamak F. Shahandashti, and Peter Hyun-Jeen Lee. Every Vote Counts: Ensuring Integrity in Large-Scale Electronic Voting. USENIX Journal of Election Technology and Systems, 2:1–25, 2014

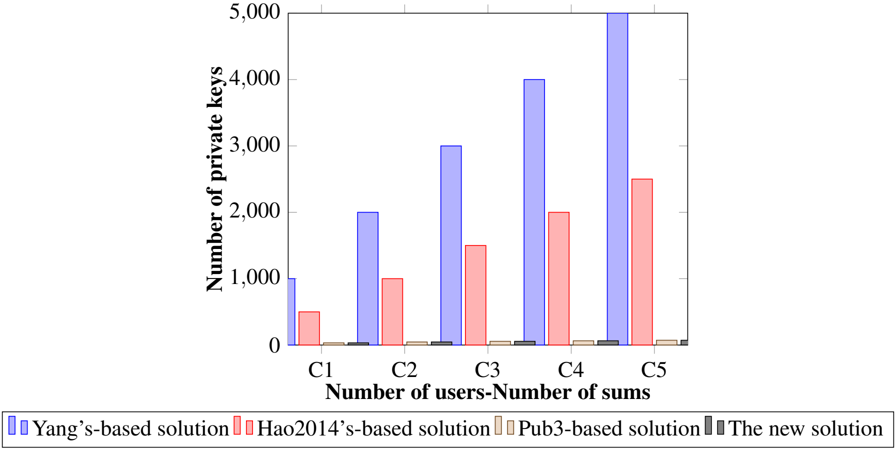


Figure 2.6: The number of private keys comparisons among the compared solutions.

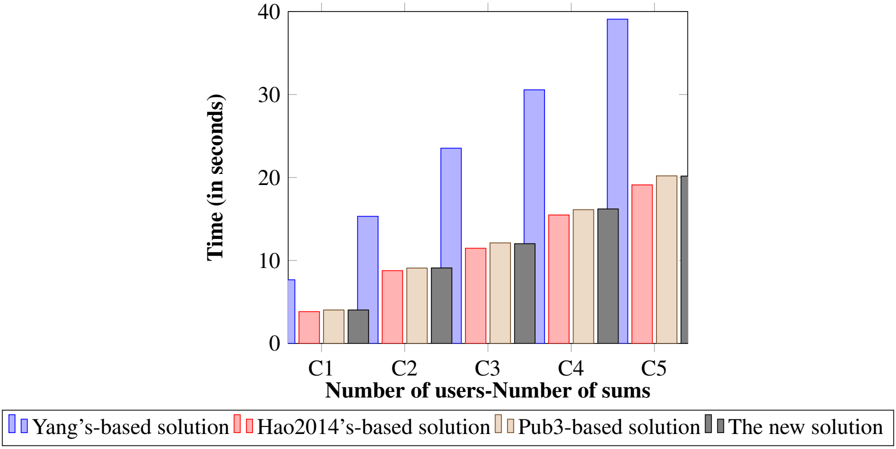


Figure 2.7: The total running time of each user comparisons among the compared solutions.

2.3. Conclusion

This chapter has provided a background of secure multi-party computation field and comprehensively analyzed the most typical previous work related to this study. Based on the analysis results, three new protocols have

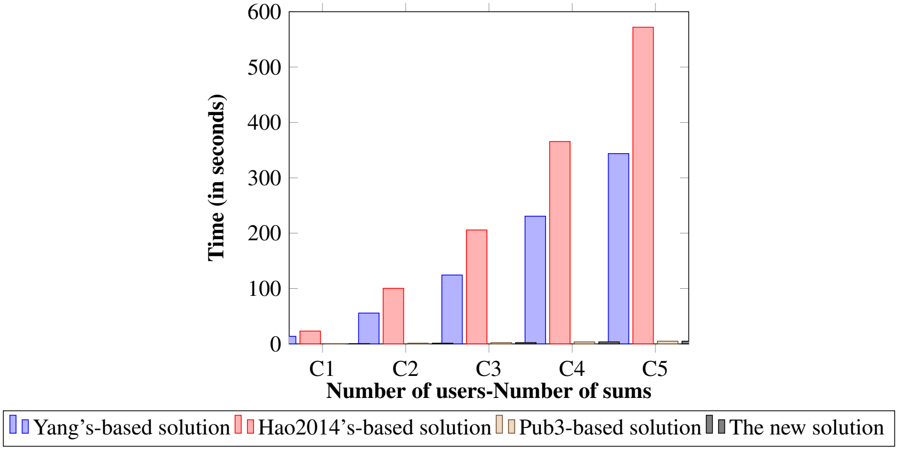


Figure 2.8: The running time for the miner to compute the public keys comparisons among the compared solutions.

Table 2.4: The stored data volume of the miner comparisons among the compared solutions (in megabytes).

Number of users-Number of sums / Solutions	1000-500	2000-1000	3000-1500	4000-2000	5000-2500
Yang's-based solution	488.8	1954.1	4396.0	7814.5	12209.5
Hao2014's-based solution	366.2	1464.8	3295.9	5859.4	9155.3
Hien's-based solution	130.1	510.8	1139.7	2015.6	3139.7
The new solution	130.1	510.8	1139.7	2015.6	3139.7

been propounded. The evaluations also show their efficiency. The proposed protocols have the capability to be applied in practical problems requiring to securely compute frequency or sum values.

CHAPTER 3. DEVELOPING NEW SOLUTIONS BASED ON SECURE MULTI-PARTY SUM COMPUTATION PROTOCOLS FOR PRACTICAL PROBLEMS

Based on the proposed protocols presented in Chapter 2, this chapter constructs solutions for two very practical problems that are the secure end-to-end decentralized voting scheme and the privacy-preserving Naive Bayes classifier in the horizontally distributed data setting. These proposals of the thesis are related to **Publications** 2, 4, and 5.

3.1. An efficient solution for the secure electronic voting scheme without pre-establishing authenticated channel

3.1.1. Introduction

In this section, an efficient solution for electronic voting scheme without pre-establishing authenticated channel is proposed. This proposal has related to **Publication** 2.

3.1.2. A secure end-to-end electronic voting scheme

It is assumed that the cryptographic parameters $q, E(\mathbb{Z}_q), \mathbb{O}, G$ are known by all participants.

The main phases of the single-candidate decentralized e-voting scheme based on privacy-preserving frequency computation protocol are detailed in Protocol 3.1.

3.1.3. Security analysis

It has been proven that the new solution meets the most important security properties of the proposed electronic voting scheme (i.e. accuracy, privacy, and verifiability).

Protocol 3.1: A single-candidate decentralized electronic voting scheme based on privacy-preserving frequency computation protocol

Phase 0: Preparation

- Each voter U_i chooses $p_i, q_i \in \{1,...,q-1\}$ and computes: $P_i = p_i G$, $Q_i = q_i G$
- $U_i \rightarrow$ Voting server: P_i, Q_i
- Voting server pre-computes the shared public values: $P = \sum_{i=1}^{n} P_i$, $Q = \sum_{i=1}^{n} Q_i$
- Voting server $\rightarrow$ All voters: P, Q

Phase 1: Ballot submission

- Each voter U_i computes: $M_i = v_i G + q_i P - p_i Q$, $r_i = x_{Q_i}$, $s_i = q_i - p_i H(x_{M_i} \parallel r_i)$
- $U_i \rightarrow$ Voting server: M_i, s_i

Phase 2: Voter authentication

- Voting server publishes M_i, s_i on its bulletin board
- Voting server authenticates U_i by computing $\gamma_i = H(x_{M_i} \parallel r_i)$, $R_i^{'} = s_i G + \gamma_i P_i$

and verifying the equation $r_i \overset{?}{=} x_{R_i^{'}}$

Phase 3: Vote counting

- Voting server computes: $M = \sum_{i=1}^{n} M_i$
- Executes Shanks' algorithm to obtain the output v satisfying $vG = M$

3.1.4. Experimental evaluation

This section considers the running time between the proposed solution and one of the most typical e-voting schemes in Hao et al.[1], 2018 (Hao2018's scheme, for short). It is assumed that all voters interact with the voting server at the same time, and the network latency is not considered.

3.1.4.1. Experimental setting

The experiments are implemented on the Lenovo Thinkpad X280 laptop with an Intel core i5 with different numbers of voters, from 2000 to 10000. The curve 25519, and the secure hash function $RIPEMD-160$ are used.

3.1.4.2. Experimental results

The experimental results are detailed in Figures 3.1 and 3.2.

It can be stated that the proposed electronic voting scheme is efficient and practical.

[1] Feng Hao, Dylan Clarke, Brian Randell, and Siamak F. Shahandashti. Verifable Classroom Voting in Practice. IEEE Security & Privacy, 16:72–81, 2018

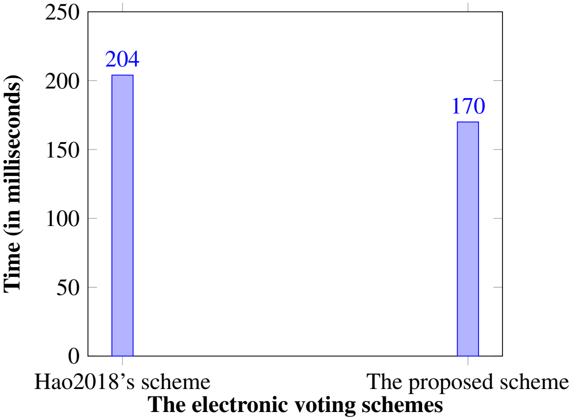


Figure 3.1: The total running time of each voter comparisons between the new solution and Hao2018's scheme.

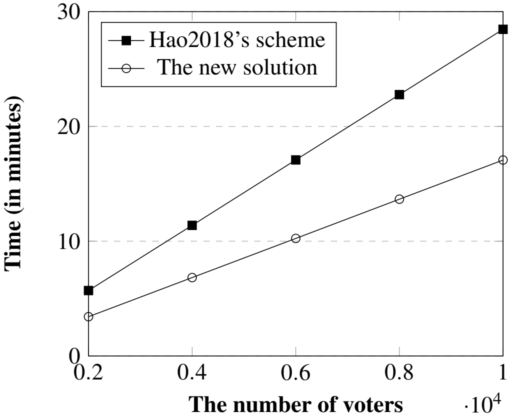


Figure 3.2: The voting server's total running time comparisons between the new solution and Hao2018's scheme.

3.2. An efficient and practical solution for privacy-preserving Naive Bayes classification in the horizontally distributed data setting

3.2.1. Introduction

This section presents an efficient and practical solution for privacy-preserving Naive Bayes classification in the horizontally distributed data setting. This proposal of the thesis has related to **Publications** 4 **and** 5.

3.2.2. New privacy-preserving Naive Bayes classifier for the horizontally distributed data setting

Protocol 3.2 describes the proposed PPNBC solution based on the new secure multi-party computation protocol, in which *np* is the number of participants, *nk* is the number of tuples of keys, *ns* is the number of sum values that are used for constructing the Naive Bayessian model.

3.2.3. Privacy analysis

As a result, according to Theorem 2.1, the proposed PPNBC solution can securely protect each data owner's privacy, even when there are up to $(np-2)$ parties colluding with the miner.

3.2.4. Accuracy analysis

The proposed PPNBC solution accurately ensures the classification model result because of the secure multi-sum protocol's correctness.

3.2.5. Experimental evaluation

This section compares the experimental results of this solution with the ones of three typical privacy-preserving Naive Bayes classifiers: the solution of Yang et al.[2], the solution using Hao et al.'s protocol[3] to compute frequency values, and the private Naive Bayes classifier in the fifth publication (denoted as Yang's solution, Hao2014's-based solution, and Pub5 solution, respectively).

3.2.7.1. Experimental setting

This section considers a practical PPML problem, i.e. spam messages detection using Naive Bayes classifier with privacy concerns.

[2] Zhiqiang Yang, Sheng Zhong, and Rebecca N. Wright. Privacy-Preserving Classification of Customer Data without Loss of Accuracy. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 92–102. SIAM, 2005

[3] Feng Hao, Matthew N. Kreeger, Brian Randell, Dylan Clarke, Siamak F. Shahandashti, and Peter Hyun-Jeen Lee. Every Vote Counts: Ensuring Integrity in Large-Scale Electronic Voting. USENIX Journal of Election Technology and Systems, 2:1–25, 2014

3.2.7.3. Experimental results

i. Running time

The experimental results are summarized in Table 3.1.

Table 3.1: The running time comparisons among the new proposal and the typical PPNBC solutions on the real dataset (in seconds).

Solutions	**The time**$^{(1)}$	**The time**$^{(2)}$	**The time**$^{(3)}$	**The time**$^{(4)}$
Yang's solution	46.196	40.506	46.602	2808.035
Hao2014's-based solution	23.036	76.287	**23.263**	81.390
Pub5 solution	**0.429**	**0.382**	46.734	81.390
The new solution	**0.429**	**0.382**	46.734	**20.394**

Note: The time$^{(1)}$ for the first phase, The time$^{(2)}$ for the second phase.
The time$^{(3)}$ for third phase, The time$^{(4)}$ for the fourth phase.

The experimental results show that the new PPNBC solution has more advantages than the others. Therefore, the new proposal is the most suitable solution for real-life applications.

ii. Classification accuracy discussion

The accuracy rates of the proposed PPNBC model are equal to the ones of the traditional Naive Bayes classification model tested on the same pre-processed spam short-messages detection dataset.

iii. Deployment issues of the proposed PPNBC solution

In practice, the proposed PPNBC solution is quite simple to be developed on smart devices.

3.3. Conclusion

In this chapter, based on the new protocols presented in Chapter 2, the thesis has developed new solutions for two practical problems, the secure voting scheme in the end-to-end decentralized environment and the privacy-preserving Naive Bayes classifier in the horizontally distributed scenario. The necessary experiments have been run, and the experimental results proved that the new proposals not only satisfy the requirements of practical problems, but also outperform the previous solutions.

Protocol 3.2: An efficient solution based on secure multi-party computation protocol for privacy-preserving Naive Bayes classification

```
1. Keys preparation: The data owners $U_i$ do
forall $i$ where $1 \le i \le np$ do
    forall $j$ where $1 \le j \le nk$ do
        $Prv_i^j = Random(1, q-1)$
        $Pub_i^j = g^{Prv_i^j}$
        Sends to Miner: $Pub_i^j$
    end
end
3. Data submission: The data owners $U_i$ do
forall $i$ where $1 \le i \le np$ do
    $j = 1$
    forall $t$ where $1 \le t \le nk-1$ do
        forall $k$ where $t+1 \le k \le nk$ do
            $p_i^j = g^{v_i^j}(Pub^t)^{Prv_i^k}(Pub^k)^{q-Prv_i^t}$
            if $j == ns$ then
                break
            else
                $j{+}{+}$
            end
        end
    end
    Sends to Miner: $p_i^j$
end
2. Shared public keys computation: Miner does
forall $j$ where $1 \le j \le nk$ do
    $Pub^j = 1$
    forall $i$ where $1 \le i \le np$ do
        $Pub^j = Pub^j * Pub_i^j$
    end
    Sends to all data owners: $Pub^j$
end
4. Results Extraction: Miner does
forall $j$ where $1 \le j \le ns$ do
    $K^j = 1$
    forall $i$ where $1 \le i \le np$ do
        $K^j = K^j * p_i^j$
    end
end
Solves the problems $g^{Sum_j} = K^j$ ($Sum_j \in \{0, 1, ..., np\}$, $j \in \{1, ..., ns\}$) by performing the brute-force algorithm once
```

CONCLUSION AND FUTURE WORK

The thesis has proposed three secure multi-party sum protocols and developed two practical applications based on these protocols. For each proposal, the thesis analyzed the security aspect as well as evaluated the performance. The results of the thesis can be summarized as follows:

- In the first work, the thesis has propounded three novel secure multi-party sum protocols that are the elliptic curve analog of the ElGamal cryptosystem-based protocol for privacy-preserving frequency computation in fully distributed setting, the secure multi-party sum protocol without pre-establishing any authenticated/ or secure channel, and the secure multi-sum computation protocol in one round of computation. The proposed protocols are secure and efficient enough to be implemented in real-life applications.
- In the second work, the thesis has developed new solutions based on the proposed protocols for two practical applications that are the secure E2E e-voting scheme without pre-establishing secure/ or authenticated channels, and an efficient and practical method for privacy-preserving Naive Bayes classification in the horizontally distributed data setting. The security analysis results have proved that the new solutions satisfy the requirements of applications. Besides, the experimental evaluations has showed the new proposals' efficacy.

Next, some potential issues of the general SMC field are considered.

- Firstly, based on new distributed computing scenarios or the requirements of practical problems, new secure multi-party computation protocols are necessary to be investigated by research community.
- Secondly, secure multi-party computation protocols should be based on modern cryptography (e.g. post-quantum cryptography) with the aim of preventing potential attacks on the next generation of computing that is coming closer to us.
- Thirdly, because of SMC protocols' applicability in various domains, such protocols should be considered for the implementation in real-life applications. This helps not only distributed computing tasks to be performed more easily, but also the parties' safety to be protected.

LIST OF THE PUBLICATIONS
RELATED TO THE DISSERTATION

1. Duy-Hien Vu, The-Dung Luong, Tu-Bao Ho, and Chung-Tien Nguyen. Privacy-preserving frequency mining protocol based on elliptic curve ElGamal cryptosystem. *HNUE Journal of Science*, 63:87-96, 2018.

2. Duy-Hien Vu, The-Dung Luong, Tu-Bao Ho, and Chung-Tien Nguyen. An Efficient Approach for Electronic Voting Scheme without An Authenticated Channel. In *Proceedings of the 10th International Conference on Knowledge and Systems Engineering*, pages 376-381. IEEE, 2018.

3. Duy-Hien Vu, The-Dung Luong, and Tu-Bao Ho. An efficient approach for secure multi-party computation without authenticated channel. *Information Sciences*, 527:356-368, 2020. (SCI/ISI indexed, Scopus Q1).

4. Duy-Hien Vu, Trong-Sinh Vu, and The-Dung Luong. An efficient and practical approach for privacypreserving Naive Bayes classifcation. *Journal of Information Security and Applications*, 68, 2022. (SCIE/ISI indexed, Scopus Q1).

5. Vu Duy Hien, Luong The Dung, and Hoang Duc Tho. An Efficient Solution for Privacy-preserving Naive Bayes Classifcation in Fully Distributed Data Model. *Journal of Science and Technology on Information Security*, 15:56-62, 2022.