BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Vũ Duy Hiến

NGHIÊN CỨU PHÁT TRIỂN MỘT SỐ GIAO THỨC TÍNH TỔNG BẢO MẬT HIỆU QUẢ TRONG MÔ HÌNH DỮ LIỆU PHÂN TÁN ĐẦY ĐỦ VÀ ỨNG DỤNG

TÓM TẮT LUẬN ÁN TIẾN SĨ
HỆ THỐNG THÔNG TIN
Mã số: 9 48 01 04

Hà Nội – 2024

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ,

Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học:
1. Người hướng dẫn 1: GS. TSKH. Hồ Tú Bảo, Viện Nghiên cứu cao cấp về Toán
2. Người hướng dẫn 2: PGS. TS. Lương Thế Dũng, Học viện Kỹ thuật Mật mã

Phản biện 1:

Phản biện 2:

Phản biện 3:

Luận án được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi giờ , ngày tháng năm

Có thể tìm hiểu luận án tại:
1. Thư viện Học viện Khoa học và Công nghệ
2. Thư viện Quốc gia Việt Nam

MỤC LỤC

GIỚI THIỆU

A. Động lực nghiên cứu

Luận án này tập trung nghiên cứu bài toán tính tổng bảo mật nhiều thành viên (viết tắt là SMS). Trong mô hình của bài toán SMS, giả sử rằng có các thành viên tham gia trong đó mỗi thành viên sở hữu một giá trị đầu vào riêng tư và tất cả các thành viên mong muốn có được tổng của các giá trị này nhưng họ không tiết lộ gì về đầu vào của họ.

Đến nay, các giải pháp hiện nay cho bài toán tính tổng bảo mật thường có mức độ an toàn thấp, hiệu năng nghèo nàn hoặc phải đánh đổi giữa sự an toàn và hiệu quả. Vì vậy, việc phát triển các giao thức tổng bảo mật vừa an toàn, vừa hiệu quả là một thách thức lớn đối với cộng đồng nghiên cứu.

B. Mục tiêu nghiên cứu

Mục tiêu nghiên cứu của luận án này bao gồm:

- Thiết kế một số giao thức tính tổng bảo mật nhiều thành viên an toàn và hiệu quả.
- Phát triển các giải pháp mới dựa trên giao thức tính tổng bảo mật mới cho một số bài toán thực tế.

C. Những đóng góp chính của luận án

Các đóng góp chính của luận án được tổng kết như sau:

- Đề xuất ba giao thức SMS mới an toàn và hiệu quả.
- Phát triển các giải pháp dựa trên các giao thức mới đề xuất cho hai bài toán trong thực tế.

D. Cấu trúc của luận án

Những nội dung chính của luận án được tổ chức như sau:

- Chương 1 cung cấp kiến thức tổng quan về lĩnh vực tính toán bảo mật nhiều thành viên và khảo sát các nghiên cứu liên quan.
- Chương 2 2 phân tích chi tiết các giao thức SMS điển hình sau đó đề xuất ba giao thức mới.
- Chương 3 phát triển các giải pháp mới dựa trên các giao thức SMS đề xuất cho hai ứng dụng thực tế.

CHƯƠNG 1. TỔNG QUAN VỀ TÍNH TỔNG BẢO MẬT NHIỀU THÀNH VIÊN

Trong chương này, đầu tiên luận án cung cấp khái quát về lĩnh vực tính toán bảo mật nhiều thành viên và bài toán tính tổng bảo mật nhiều thành viên, sau đó những nghiên cứu liên quan được phân tích.

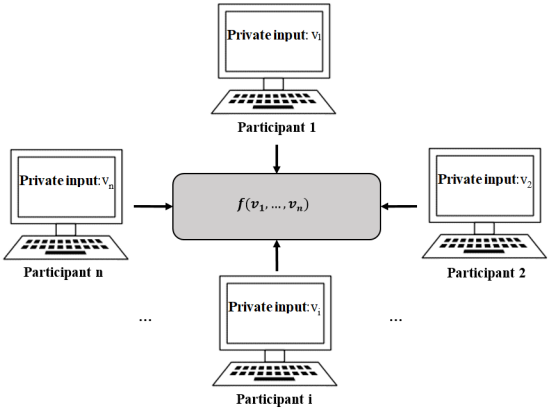
1.1. Khái quát về tính toán bảo mật nhiều thành viên

1.1.1. Giới thiệu

- *Đầu vào*: có n thành viên trong đó thành viên thứ i sở hữu giá trị đầu vào riêng tư v_i.
- *Đầu ra*: các thành viên đạt được kết quả của hàm $f(v_1, ..., v_n)$ và mỗi thành viên không tiết lộ điều gì về giá trị đầu vào của họ.

Ở đây, cần nhấn mạnh rằng khái niệm "`bảo mật/an toàn`" nghĩa là hai ràng buộc sau được thỏa mãn:

- Sự đúng đắn của kết quả đầu ra được đảm bảo.
- Đầu vào của mỗi thành viên được giữ riêng tư bởi chính họ mà thôi.



Hình 1.1: Mô hình tính toán phân tán trong ngữ cảnh an toàn

Sự an toàn của một giao thức `SMC` phụ thuộc vào mô hình tấn công, loại kênh kết nối, và khả năng của địch thủ.

Có thể thấy rằng nhiều bài toán thực tế liên quan đến `SMC`.

1.1.2. Định nghĩa an toàn

Luận án sử dụng một định nghĩa chuẩn an toàn đối với các giao thức tính toán bảo mật nhiều thành viên trong mô hình bán trung thực sử dụng kênh kết nối công khai của O.Goldreich[1].

1.1.3. Cơ sở mật mã học

Luận án này được dựa trên hai cơ sở mật mã học quan trọng là bài toán logarit rời rạc khó trong các nhóm cyclic tiêu chuẩn, và hệ mã hóa đồng cấu ElGamal

1.2. Bài toán tính tổng bảo mật nhiều thành viên

1.2.1. Phát biểu bài toán

- *Đầu vào*: n thành viên, mỗi thành viên thứ i sở hữu giá trị đầu vào riêng tư v_i.
- *Đầu ra*: các thành viên đạt được giá trị tổng $f(v_1,...,v_n) = v_1 + ... + v_n$, và mỗi thành viên không tiết lộ gì về giá trị đầu vào của họ, ngoại trừ giá trị tổng.

1.2.2. Các nghiên cứu liên quan

Có thể thấy rằng những giao thức SMS phi mật mã thường có chi phí tính toán thấp nhưng chúng phải đánh đổi giữa tính an toàn và chi phí truyền thông Ngược lại, các giao thức SMS dựa trên mật mã có thể dễ dàng đạt được mức an toàn cao nhưng cũng thường có chi phí tính toán cao. Vì vậy, việc thiết kế các giao thức SMS vừa an toàn chống lại địch thủ động hại, vừa hiệu quả trong ứng dụng thực tế là rất cần thiết.

1.3. Kết luận

Trong chương này, luận án đã trình bày những nội dung tổng quan về lĩnh vực tính toán bảo mật nhiều thành viên. Luận án sau đó đã mô hình hóa bài toán tính tổng bảo mật nhiều thành viên và chỉ ra vai trò quan trọng của nó trong thực tế.

---

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

CHƯƠNG 2. ĐỀ XUẤT MỘT SỐ GIAO THỨC TÍNH TỔNG BẢO MẬT NHIỀU THÀNH VIÊN HIỆU QUẢ

Đầu tiên, chương này phân tích kỹ lưỡng những giao thức tính tổng bảo mật điển hình. Dựa trên kết quả phân tích, chương này đề xuất ba giao thức tính tổng bảo mật mới vừa đạt được mức độ an toàn cao và có hiệu năng tốt.

2.1. Phân tích những giao thức tính tổng bảo mật điển hình

Phần này phân tích chi tiết những giao thức SMS phổ biến nhất liên quan đến luận án. Nhắc lại rằng, có n thành viên $\{U_1, U_2, ..., U_n\}$ trong đó thành viên thứ i là U_i sở hữu giá trị riêng tư v_i $(i = 1, 2, ..., n)$. Mục đích của các giao thức SMS là tính toán chính xác giá trị tổng $V = \sum_{i=1}^{n} v_i$ trong khi tất cả thành viên không tiết lộ các giá trị riêng tư của họ với ai.

2.1.1. Giao thức tính tổng bảo mật nhiều thành viên của Urabe và cộng sự

Phân tích độ an toàn: giao thức an toàn và có thể chống lại $(n-2)$ thành viên thông đồng.

Phân tích hiệu năng: chi phí truyền thông cao và không thực tế vì yêu cầu các thành viên kết nối với nhau.

2.1.2. Giao thức tính tổng bảo mật nhiều thành viên của Hao và cộng sự, 2010 trong hệ thống bỏ phiếu an toàn

Phân tích độ an toàn: giao thức an toàn và có thể chống lại $(n-2)$ thành viên thông đồng.

Phân tích hiệu năng: độ phức tạp tính toán của máy chủ thấp nhưng của mỗi thành viên lại cao. Tổng số thông điệp truyền thông lớn, cụ thể là (n^2+n).

2.1.2.4. Các biến thể giao thức của Hao và cộng sự, 2010

Để giảm chi phí tính toán của mỗi người bỏ phiếu, giao thức của Hao và cộng sự[1] năm 2014 đã được đề xuất cho hệ thống bỏ phiếu điện tử an toàn sử dụng thiết bị chuyên dụng DRE. Ngoài ra, tồn tại một biến thể khác trong

[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

công trình của Hao và cộng sự[2] năm 2018, trong đó các tính toán được thực hiện trên một đường cong elliptic.

2.1.3. Giao thức tính tần suất đảm bảo tính riêng tư của Yang và cộng sự

Phân tích độ an toàn: giao thức an toàn và có thể chống lại $(n-2)$ thành viên thông đồng.

Phân tích hiệu năng: nếu số thành viên tham gia lớn thì chi phí tính toán của bên khai phá sẽ trở nên tốn kém. Tổng chi phí truyền thông của giao thức này là $6n|p|$ bits.

2.1.4. Thảo luận nâng cao

Có thể thấy rằng, giao thức của Yang và cộng sự là tiềm năng nhất trong số các giao thức SMS điển hình. Những ý tưởng phát triển giao thức SMS mới hiện thực hóa trong các mục 2.3.1, 2.3.2, 2.3.3 của luận án.

2.2. Các giao thức tính tổng bảo mật nhiều thành viên được đề xuất

Luận án đề xuất ba giao thức SMS mới đặc trưng độc đáo và an toàn trong mô hình bán trung thực.

2.2.1. Giao thức tính tần suất đảm bảo tính riêng tư dựa trên mật mã đường cong elliptic

2.2.1.1. Giới thiệu

Trong phần này, luận án trình bày giao thức tính tần suất đảm bảo tính riêng tư dựa trên hệ mã hóa ElGamal trên đường cong elliptic. Đề xuất này liên quan đến công trình thứ nhất của luận án.

2.2.1.2. Giao thức tính tần suất đảm bảo tính riêng tư

Giao thức đề xuất bao gồm ba bước chính như được mô tả dưới đây.

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

Giao thức 2.1: Một giao thức tính tần suất đảm bảo tính riêng tư cho mô hình phân tán đầy đủ

Bước 1: Chuẩn bị

- Bên khai phá tính trước các khóa công khai dùng chung: $P = \sum_{i=1}^{n} P_i$, $Q = \sum_{i=1}^{n} Q_i$
- Bên khai phá $\rightarrow U_i$: P, Q

Bước 2: Tính toán các thông điệp

- U_i tính: $M_i = v_i G + q_i P - p_i Q$
- $U_i \rightarrow$ Bên khai phá: M_i

Bước 3: Tính tần suất an toàn

- Bên khai phá tính: $M = \sum_{i=1}^{n} M_i$
- Bên khai phá thực thi thuật toán Shanks để tìm ra v thỏa mãn $vG = M$

2.2.1.3. Phân tích độ an toàn

i. Chứng minh tính đúng đắn: nếu bên khai phá tìm ra giá trị v thỏa mãn thì v chính là giá trị tần suất.

ii. Phân tích tính riêng tư: giao thức trình bày trong Hình 2.2 được chứng minh bảo vệ sự riêng tư của mỗi thành viên trong mô hình bán trung thực. Sau đó, nó cũng được chỉ ra là vẫn có thể bảo vệ sự riêng tư của mỗi thành viên ngay cả trong trường hợp có tới $(n-2)$ thành viên cấu kết với Bên khai phá.

2.2.1.4. Đánh giá hiệu năng

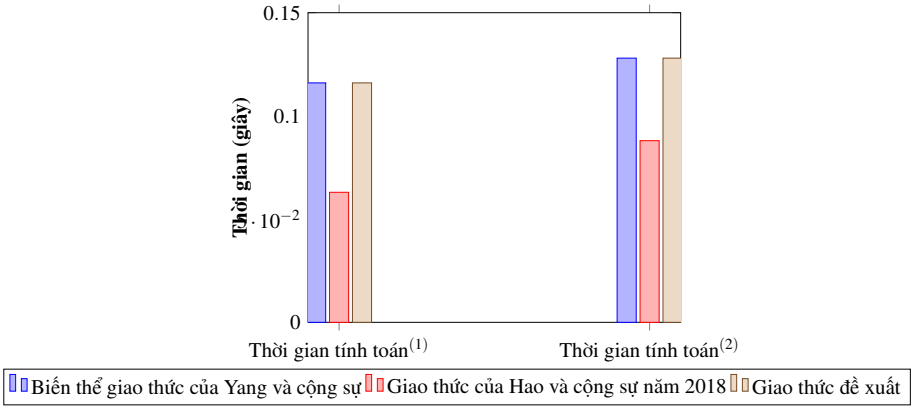
Để đánh giá hiệu năng của giao thức `PPFC` đề xuất, đầu tiên luận án so sánh độ phức tạp tính toán và chi phí truyền thông của giao thức này với giao thức của Hao và cộng sự năm 2018[3] biến thể trên đường cong elliptic của giao thức của Yang và cộng sự[4]. Cả ba giao thức trên được giả sử rằng thực thi thuật toán Shanks để tìm ra giá trị tổng ở bước cuối cùng và các kênh xác thực luôn sẵn sàng cho máy chủ/bên khai phá giao tiếp với mỗi người dùng.

- *Thiết kế thí nghiệm*

Các thí nghiệm được thực hiện trên máy tính xác tay Lenovo Thinkpad X280 với các số người dùng khác nhau từ 10000 đến 50000 và đường cong

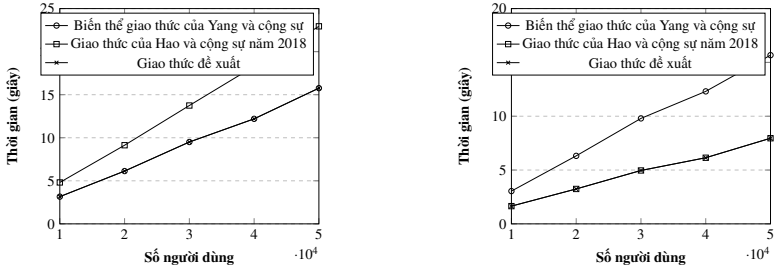
[3]Feng Hao, Dylan Clarke, Brian Randell, and Siamak F. Shahandashti. Verifable Classroom Voting in Practice. IEEE Security & Privacy, 16:72–81, 2018

[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) Thời gian mỗi người dùng chuẩn bị các khóa công khai, (2) Thời gian mỗi người dùng tính các thông điệp truyền thông

Hình 2.1: So sánh thời gian thực thi của mỗi người dùng trong giao thức đề xuất với các giao thức PPFC điển hình.



(a) Thời gian tính khóa công khai của bên khai phá/máy chủ

(b) Thời gian tính giá trị tần suất của bên khai phá/máy chủ

Hình 2.2: So sánh thời gian tính toán của bên khai phá/máy chủ của các giao thức điển hình

elliptic an toàn 25519.

- *Kết quả thực nghiệm*

Theo các kết quả so sánh kể trên, có thể thấy rằng giải pháp mới đề xuất có nhiều ưu điểm hơn so với biến thể giao thức của Yang và cộng sự và giao thức của Hao và cộng sự năm 2018. Ngoài ra, nếu thiết lập kết nối xác thực giữa máy chủ/bên khai phá với từng người dùng, chi phí tính toán và

Bảng 2.1: So sánh lượng dữ liệu lưu trữ của máy chủ giữa các giao thức (theo megabytes).

Số thành viên / Các giao thức	10000	20000	30000	40000	50000
Biến thể giao thức của Yang và cộng sự	2.4	4.9	7.3	9.8	12.2
Giao thức của Hao và cộng sự năm 2018	1.2	2.4	3.7	4.9	6.1
Giao thức đề xuất	1.8	3.7	5.5	7.3	9.2

truyền thông trong giao thức của Hao và cộng sự năm 2018 còn tăng lên đáng kể nữa.

2.2.2. Giao thức tính tổng bảo mật nhiều thành viên hiệu quả không cần thiết lập trước kênh kết nối xác thực

2.2.2.1. Giới thiệu

Trong phần này, một giao thức SMS hiệu quả mà không cần thiết lập trước kênh kết nối xác thực được đề xuất. Đề xuất này liên quan đến công trình thứ ba.

2.2.2.2. Giao thức tính tổng bảo mật nhiều thành viên hiệu quả đề xuất

Giao thức đề xuất yêu cầu các tham số sau: $(\mathbb{G}, p, q, g)$ là các tham số mật mã công khai cần thiết. Mỗi người dùng U_i sở hữu sẵn một khóa riêng tư x_i và khóa công khai tương ứng X_i. Người dùng này chọn thêm một số ngẫu nhiên bí mật y_i (giá trị công khai tương ứng là Y_i). H là một hàm băm an toàn.

Trước khi bắt đầu giao thức, mỗi người dùng U_i gửi Y_i đến bên khai phá. Tiếp theo, bên khai phá tính toán trước: $X = \prod_{i=1}^{n} X_i$; $Y = \prod_{i=1}^{n} Y_i$. Sau đó, bên khai phá gửi thông điệp $M = (X \parallel Y)$ và chữ ký số Schnorr trên M đến tất cả người dùng.

iv. Giao thức đề xuất

Ba bước chính của giao thức đề xuất được trình bày dưới đây.

Giao thức 2.2: Giao thức tính tổng bảo mật n thành viên không yêu cầu kênh xác thực

Bước 1: Gửi đi dữ liệu

- Mỗi người dùng xác thực chữ ký số Schnorr của bên khai phá trên $M = (X \parallel Y)$
- Mỗi người dùng U_i tính: $P_i = \frac{g^{v_i} X^{y_i}}{Y^{x_i}}, r_i = Y_i, s_i \equiv y_i - x_i H(r_i \parallel P_i) \pmod{q}$
- Mỗi người dùng $U_i \rightarrow$ Miner: P_i, s_i

Bước 2: Xác thực người dùng

- Bên khai phá tính: $\gamma_i = H(r_i \parallel P_i), r_i^{'} = g^{s_i} X_i^{\gamma_i}$
- Bên khai phá xác thực U_i bằng cách kiểm tra: $r_i \stackrel{?}{=} r_i^{'}$

Bước 3: Tính toán tổng bảo mật

- Bên khai phá tính: $K = \prod_{i=1}^{n} P_i$
- Bên khai phá thực thi thuật toán cuả Shanks để tìm ra V thỏa mãn $g^V = K$

2.2.2.3. Phân tích độ an toàn

i. Giao thức con tính toán tổng bảo mật

- *Chứng minh tính đúng đắn*

Phần này đã chứng minh rằng kết quả đầu ra của giao thức con này chính là tổng của các giá trị đầu vào riêng tư.

- *Phân tích tính riêng tư*

Tiếp theo, giao thức con này cũng được chứng minh khả năng bảo vệ sự riêng tư của mỗi người dùng trong mô hình bán trung thực dưới các giả thiết quan trọng.

ii. Giao thức con xác thực người dùng

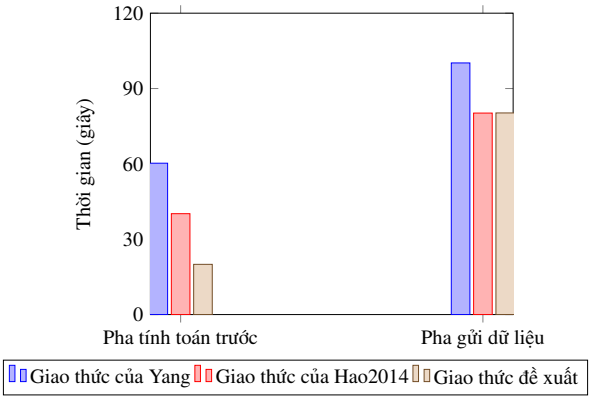
Phần này đã chỉ ra rằng (i) mỗi người dùng U_i với bộ $\{P_i, r_i, s_i\}$ được xác thực đúng đắn, và (ii) giao thức con này an toàn chống lại các tấn công có thể có trong mô hình tiên tri ngẫu nhiên.

2.2.2.4. Đánh giá hiệu năng

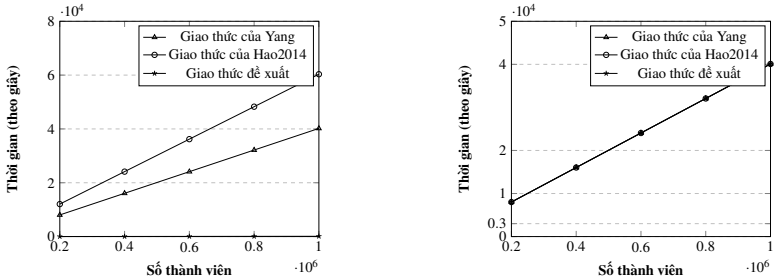
Phần này so sánh hiệu năng của giải pháp đề suất với các giao thức điển hình của Yang và cộng sự[5] và của Hao và cộng sự[6] năm 2014 (gọi tắt

[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



Hình 2.3: So sánh thời gian thực thi của mỗi người dùng trong giao thức đề xuất và các giao thức SMS điển hình.



(a) Thời gian pha tính toán trước (b) Thời gian xác thực người dùng.

Hình 2.4: *So sánh thời gian tính toàn trước và xác thực người dùng giữa giao thức đề xuất và các giao thức điển hình.*

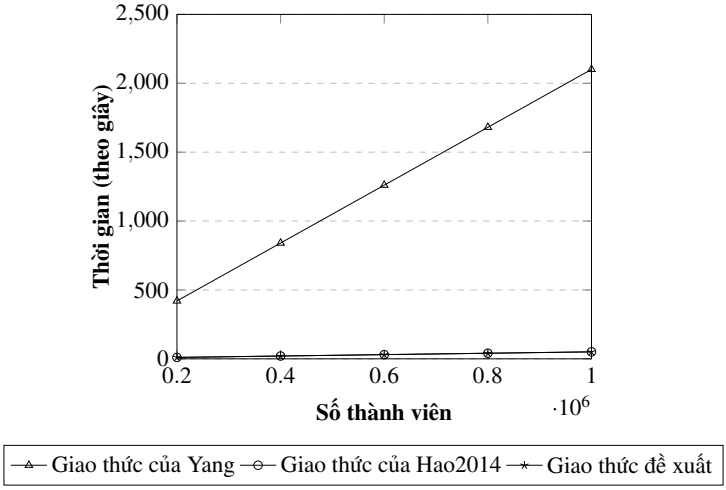
tương ứng là giao thức của Yang, giao thức của Hao2014).

- *Thiết kế thí nghiệm*

Các thí nghiệm được thực thi trên máy tính xách tay Lenovo Thinkpad X280 với các số người dùng khác nhau từ 200000 đến 1000000 và số nguyên tố p dài 2048 bits, còn độ dài của q là 256 bits.

- *Kết quả thực nghiệm*

Xem xét tất cả các kết quả thực nghiệm ở trên, có thể thấy rằng giao



Hình 2.5: So sánh thời gian thực thi pha tính tổng của bên khai phá trong giao thức đề xuất với các giao thức điển hình.

Bảng 2.2: So sánh lượng dữ liệu được lưu trữ của máy chủ giữa giao thức đề xuất và các giao thức điển hình (theo megabytes).

Số thành viên / Các giao thức	200000	400000	600000	800000	1000000
Giao thức của Yang	215.1	430.3	645.4	860.6	1075.7
Giao thức của Hao2014	215.1	430.3	645.4	860.6	1075.7
Giao thức đề xuất	107.6	215.1	322.7	430.3	537.9

thức đề xuất có nhiều ưu điểm hơn so với các giao thức SMS đặc trưng khác. Do đó, trong số các giao thức được so sánh thì giao thức đề xuất là giải pháp phù hợp nhất cho các ứng dụng thực tế.

2.2.3. Giao thức tính đa tổng bảo mật nhiều thành viên

2.2.3.1. Giới thiệu

Phần này trình bày một giao thức hiệu quả để tính đa tổng bảo mật, nghĩa là có thể tính toán an toàn nhiều giá trị tổng chỉ trong một vòng tính toán. Đề xuất này liên quan đến Công trình thứ tư của luận án.

2.2.3.2. Giao thức an toàn tính toán đa tổng

Cho $(\mathbb{G}, p, q, g)$ là các tham số chuẩn mật mã công khai. Trong giao thức mới, mỗi thành viên chỉ cần chuẩn bị nk bộ khóa với $nk = \left\lceil \frac{1}{2} + \sqrt{2ns + \frac{1}{4}} \right\rceil$. Các bước chính của giao thức mới được trình bày như dưới đây.

Giao thức 2.3: Giao thức an toàn tính toán đa tổng trong một vòng tính toán.

1. Bước 1: Thành viên U_i thực hiện
forall *i where* $1 \leq i \leq np$ **do**
  forall *j where* $1 \leq j \leq nk$ **do**
    $Prv_i^j = Random(1, q-1)$
    $Pub_i^j = g^{Prv_i^j}$
    Gửi tới bên khai phá: Pub_i^j
  end
end

2. Bước 2: Bên khai phá thực hiện
forall *j where* $1 \leq j \leq nk$ **do**
  $Pub^j = 1$
  forall *i where* $1 \leq i \leq np$ **do**
    $Pub^j = Pub^j * Pub_i^j$
  end
  Gửi tới mọi thành viên: Pub^j
end

3. Bước 3: Thành viên U_i thực hiện
forall *i where* $1 \leq i \leq np$ **do**
  $j = 1$
  forall *t where* $1 \leq t \leq nk-1$ **do**
    forall *k where* $t+1 \leq k \leq nk$ **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
  Gửi tới bên khai phá: p_i^j
end

4. Bước 4: Bên khai phá thực hiện
forall *j where* $1 \leq j \leq ns$ **do**
  $K^j = 1$
  forall *i where* $1 \leq i \leq np$ **do**
    $K^j = K^j * p_i^j$
  end
end
Giải các bài toán $g^{Sum_j} = K^j$ ($j \in \{1, \ldots, ns\}$) bằng cách thực thi thuật toán vét cạn một lần

2.2.3.3. Phân tích độ an toàn

i. Chứng minh tính đúng đắn

Phần này đã chỉ ra tính đúng đắn của giao thức tính bảo mật đa tổng mới đề xuất dựa vào tính chất đồng cấu của hệ mã ElGamal.

ii. Phân tích tính riêng tư

Giao thức mới đề xuất được chứng minh là có thể bảo vệ sự riêng tư của các thành viên trong mô hình bán trung thực dưới các giả sử cần thiết.

2.2.3.4. Đánh giá hiệu năng

Phần này đánh giá giao thức đề xuất với các giải pháp được tạo ra bằng cách thực thi nhiều lần giao thức của Yang và cộng sự[7], giao thức của Hao và cộng sự[8], và giao thức trong công trình thứ ba (ký hiệu tương ứng là giải pháp dựa trên Yang, giải pháp dựa trên Hao2014, và giải pháp dựa trên CT3).

• *Thiết kế thí nghiệm*

Trong các thí nghiệm, các giải pháp được triển khai bằng ngôn ngữ lập trình Python trong môi trường Anaconda trên máy tính xác tay Lenovo Thinkpad X280 với các cặp (số người dùng-số giá trị tổng) khác nhau là (1000, 500), (2000, 1000), (3000, 1500), (4000, 2000), (5000, 2500) được ký hiệu lần lượt là $C1, C2, C3, C4, C5$. Số nguyên lớn hơn p dài 2048 bits, còn độ dài của số nguyên nhỏ hơn q là 256 bits.

• *Kết quả thực nghiệm*

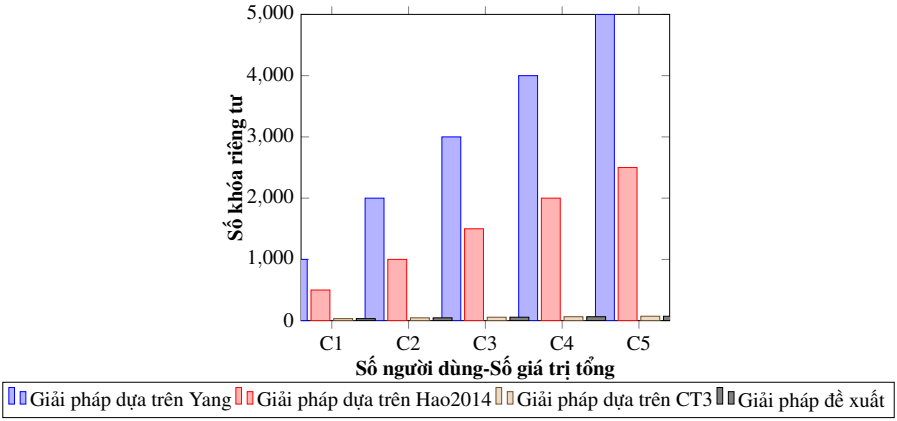
Bảng 2.3: So sánh thời gian thực thi của bên khai phá tính toán các giá trị tổng trong giải pháp đề xuất với các giải pháp điển hình.

Số người dùng-Số giá trị tổng / Các giải pháp	1000-500	2000-1000	3000-1500	4000-2000	5000-2500
Giải pháp dựa trên Yang	823.142	3217.324	7299.532	13320.492	19310.021
Giải pháp dựa trên Hao2014	10.000	35.556	74.314	128.160	195.650
Giải pháp dựa trên CT3	10.005	35.004	74.102	128.533	195.647
Giải pháp đề xuất	**7.104**	**27.588**	**62.085**	**110.262**	**172.094**

Xem xét tất cả kết quả thực nghiệm ở Hình 2.6, 2.7, 2.8 và Bảng 2.3, 2.4, có thể khẳng định rằng giải pháp đề xuất có nhiều ưu điểm hơn các giải pháp khác.

[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



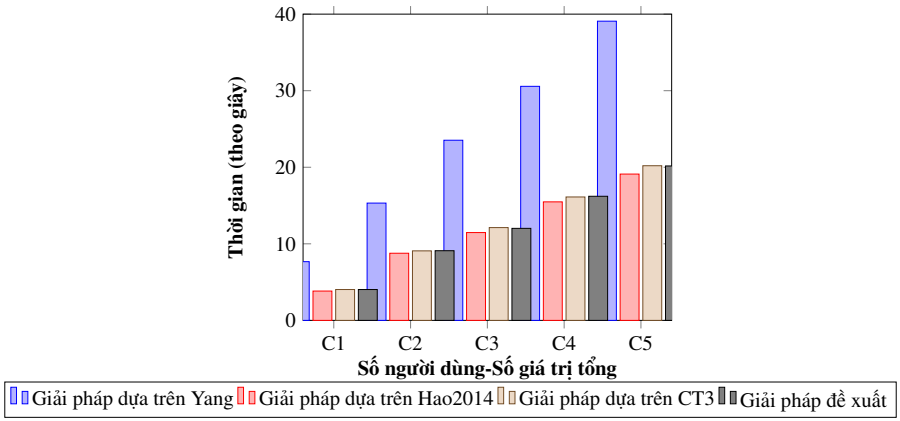
Hình 2.6: So sánh số khóa riêng tư trong giải pháp đề xuất với các giải pháp điển hình.

Bảng 2.4: So sánh lượng dữ liệu được lưu trữ bởi máy chủ giữa các giải pháp điển hình (theo megabytes).

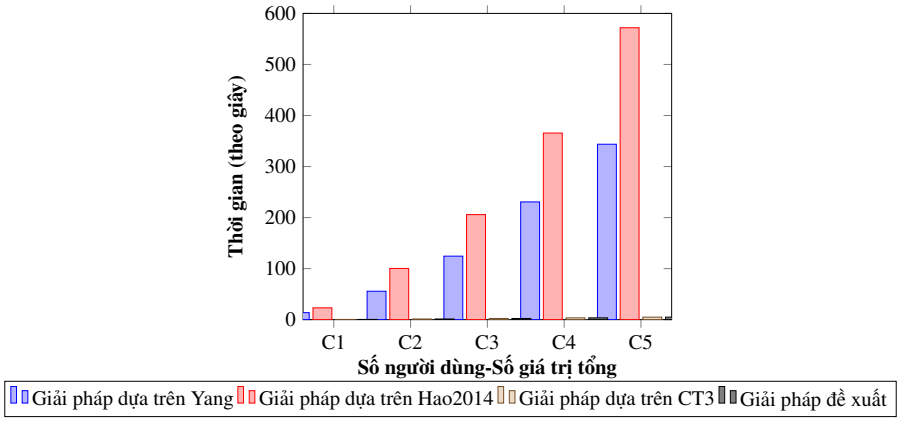
Số người dùng-Số giá trị tổng / Các giải pháp	1000-500	2000-1000	3000-1500	4000-2000	5000-2500
Giải pháp dựa trên Yang	488.8	1954.1	4396.0	7814.5	12209.5
Giải pháp dựa trên Hao2014	366.2	1464.8	3295.9	5859.4	9155.3
Giải pháp dựa trên CT3	130.1	510.8	1139.7	2015.6	3139.7
Giải pháp đề xuất	130.1	510.8	1139.7	2015.6	3139.7

2.3. Kết luận

Chương này đã phân tích và đề xuất ba giao thức mới. Các giao thức đề xuất đã được chứng minh an toàn trong mô hình bán trung thực. Các đánh giá cũng đã chỉ ra sự hiệu quả của chúng. Vì thế, các giao thức đề xuất có khả năng ứng dụng vào các bài toán trong thực tế đòi hỏi tính toán bảo mật các giá trị tổng hoặc tần suất.



Hình 2.7: So sánh tổng thời gian thực thi của mỗi người dùng trong giải pháp đề xuất với các giải pháp điển hình.



Hình 2.8: So sánh thời gian thực thi của bên khai phá tính toán các khóa công khai trong giải pháp đề xuất với các giải pháp điển hình.

CHƯƠNG 3. PHÁT TRIỂN GIẢI PHÁP MỚI CHO MỘT SỐ ỨNG DỤNG THỰC TẾ DỰA TRÊN CÁC GIAO THỨC TÍNH TỔNG BẢO MẬT NHIỀU THÀNH VIÊN ĐƯỢC ĐỀ XUẤT

Chương này phát triển giải pháp cho hai ứng dụng thực tế là hệ thống bỏ phiếu điện tử đầu-cuối an toàn và bộ phân lớp Naive Bayes đảm bảo tính riêng tư trong mô hình dữ liệu phân tán ngang. Các đề xuất này của luận án liên quan tới các công trình thứ hai, thứ tư và thứ năm.

3.1. Một giải pháp hiệu quả cho hệ thống bỏ phiếu điện tử an toàn không cần thiết lập trước các kênh kết nối xác thực

3.1.1. Giới thiệu

Trong nội dung này, một giải pháp hiệu quả cho hệ thống bỏ phiếu điện tử an toàn không cần thiết lập trước các kênh kết nối xác thực được trình bày. Đề xuất này liên quan đến công trình thứ hai.

3.1.2. Một hệ thống bỏ phiếu điện tử đầu-cuối an toàn

Trong giải pháp đề xuất, mọi bên tham gia đồng ý sử dụng các tham số chung bao gồm: q, $E(\mathbb{Z}_q)$, $\mathbb{O}$, G. Dưới đây là giải pháp đề xuất.

3.1.3. Phân tích độ an toàn

Các thuộc tính an toàn của giải pháp đề xuất (tính chính xác, tính riêng tư, và khả năng xác minh kết quả) được chứng minh.

Giao thức 3.1: Giải pháp bỏ phiếu điện tử đầu-cuối một ứng viên dựa trên giao thức tính tần suất đảm bảo tính riêng tư

Bước 0: Chuẩn bị

- Mỗi bên bỏ phiếu U_i chọn $p_i, q_i \in \{1, ..., q-1\}$
- Mỗi bên bỏ phiếu U_i tính: $P_i = p_i G$, $Q_i = q_i G$
- $U_i \rightarrow$ Máy chủ: P_i, Q_i
- Máy chủ tính toán các khóa công khai dùng chung: $P = \sum_{i=1}^{n} P_i$, $Q = \sum_{i=1}^{n} Q_i$
- Máy chủ $\rightarrow$ các bên bỏ phiếu: P, Q

Bước 1: Bỏ phiếu

- Mỗi bên bỏ phiếu U_i tính: $M_i = v_i G + q_i P - p_i Q$, $r_i = x_{Q_i}$, $s_i = q_i - p_i H(x_{M_i} \parallel r_i)$
- $U_i \rightarrow$ Máy chủ: M_i, s_i

Bước 2: Xác thực bên bỏ phiếu

- Máy chủ công khai M_i, s_i lên bảng tin
- Máy chủ xác thực U_i bằng cách:
  - Tính $\gamma_i = H(x_{M_i} \parallel r_i)$, $R_i^{'} = s_i G + \gamma_i P_i$
  - Kiểm tra phương trình $r_i \stackrel{?}{=} x_{R_i^{'}}$

Bước 3: Kiểm phiếu

- Máy chủ tính: $M = \sum_{i=1}^{n} M_i$
- Máy chủ thực thi biến thể của thuật toán Shank để đạt được kết quả v thỏa mãn $vG = M$

3.1.4. Đánh giá thực nghiệm

Phần này xem xét thời gian thực thi giữa giải pháp đề xuất và một trong các hệ thống bỏ phiếu điển hình là của Hao và cộng sự[1] năm 2018 (gọi tắt là giải pháp của Hao2018).

3.1.4.1. Thiết kế thí nghiệm

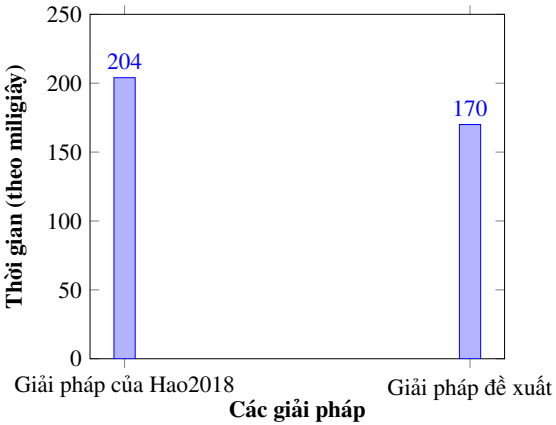
Các thí nghiệm được triển khai trên máy tính xách tay Lenovo Thinkpad X280 với các số bên bỏ phiếu từ 2000 đến 10000, đường cong 25519, và hàm băm được chọn là $RIPEMD-160$.

3.1.4.2. Kết quả thực nghiệm

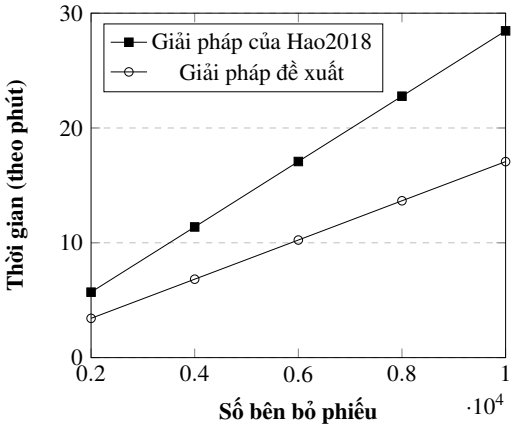
Các kết quả thực nghiệm được chi tiết trong Hình 3.1 và 3.2.

Có thể phát biểu rằng giải pháp bỏ phiếu điện tử đề xuất hiệu quả và

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



Hình 3.1: So sánh tổng thời gian thực thi của mỗi bên bỏ phiếu giữa giải pháp đề xuất và giải pháp của Hao.



Hình 3.2: So sánh tổng thời gian thực thi của máy chủ bỏ phiếu giữa giải pháp đề xuất và giải pháp của Hao.

mang tính thực tế.

3.2. Một giải pháp hiệu quả và thực tế cho kỹ thuật phân lớp Naive Bayes đảm bảo tính riêng tư trong mô hình dữ liệu phân tán ngang

3.2.1. Giới thiệu

Phần này trình bày một giải pháp hiệu quả và thực tế cho kỹ thuật phân lớp Naive Bayes đảm bảo tính riêng tư trong mô hình dữ liệu phân tán ngang. Đề xuất này liên quan đến công trình thứ tư và thứ năm.

3.2.2. Bộ phân lớp Naive Bayes đảm bảo tính riêng tư cho mô hình dữ liệu phân tán ngang

Dưới đây là giải pháp phân lớp Naive Bayes đảm bảo tính riêng tư dựa trên giao thức tính toán bảo mật nhiều thành viên, trong đó *np* là số thành viên, *kn* là số cặp khóa riêng tư & công khai, *ns* là số giá trị tổng được dùng cho việc xây dựng mô hình Naive Bayes.

3.2.3. Phân tích tính riêng tư

Do được kết hợp nhiều giao thức tính đa tổng bảo mật nên giao thức cho mô hình phân lớp Naive Bayes trình bày trong giao thức 3.2 bảo vệ an toàn sự riêng tư của mỗi bên sở hữu dữ liệu chống lại các thành viên thông đồng.

3.2.4. Phân tích độ chính xác

Do được dựa trên giao thức tính toán đa tổng chính xác nên giao thức cho mô hình phân lớp Naive Bayes trình bày trong giao thức 3.2 bảo toàn được độ chính xác của mô hình phân lớp so với phương thức truyền thống.

3.2.5. Đánh giá thực nghiệm

Phần này so sánh các kết quả thực nghiệm của giải pháp đề xuất với ba bộ phân lớp đảm bảo tính riêng tư điển hình là: giải pháp của Yang và cộng sự[2], giải pháp dựa trên giao thức tính toán của Hao và cộng sự[3] năm 2014, và giải pháp trong công trình 5 (ký hiệu tương ứng là giải pháp của Yang, giải pháp dựa trên Hao2014, và giải pháp của CT5). Với mỗi giải pháp, thời gian

[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

thực thi của các bước sẽ được đo lường.

3.2.7.1. Thiết kế thí nghiệm

Phần này xem xét một bài toán PPML thực tế, đó là bài toán phát hiện tin nhắn rác sử dụng bộ phân lớp Naive Bayes với ràng buộc tính riêng tư.

3.2.7.3. Các kết quả thực nghiệm

i. Thời gian thực thi

Các kết quả thực nghiệm được tổng hợp trong Bảng 3.1.

Bảng 3.1: So sánh thời gian thực thi giữa giải pháp đề xuất với các giải pháp PPNBC điển hình chạy trên bộ dữ liệu thực tế.

Các giải pháp	**Thời gian thực thi**$^{(1)}$	**Thời gian thực thi**$^{(2)}$	**Thời gian thực thi**$^{(3)}$	**Thời gian thực thi**$^{(4)}$
Giải pháp của Yang	46.196	40.506	46.602	2808.035
Giải pháp dựa trên Hao2014	23.036	76.287	**23.263**	81.390
Giải pháp của CT5	**0.429**	**0.382**	46.734	81.390
Giải pháp đề xuất	**0.429**	**0.382**	46.734	**20.394**

Ghi chú: Thời gian thực thi$^{(1)}$ ở pha đầu tiên, Thời gian thực thi$^{(2)}$ ở pha thứ hai. Thời gian thực thi$^{(3)}$ ở pha thứ ba, Thời gian thực thi$^{(4)}$ ở pha thứ tư.

Các kết quả thí nghiệm chỉ ra rằng giải pháp PPNBC mới đề xuất có nhiều ưu điểm vượt trội so với các giải pháp điển hình khác. Do đó, giải pháp này được xem là phù hợp nhất với các ứng dụng thực tế.

ii. Thảo luận về độ chính xác trong phân lớp

Trong các thí nghiệm, giải pháp đề xuất đạt được chỉ số "accuracy": 0.97, "balanced_accuracy": 0.92, và "F1-score": 0.93.

iii. Vấn đề triển khai ứng dụng giải pháp đề xuất

Trên thực tế, việc triển khai giải pháp đề xuất trên các thiết bị di động thông minh tương đối đơn giản.

Giao thức 3.2: Một giải pháp hiệu quả dựa trên giao thức tính đa tổng bảo mật cho bài toán phân lớp Naive Bayes đảm bảo tính riêng tư.

1. Chuẩn bị khóa: Các bên sở hữu dữ liệu U_i thực hiện
forall i *where* $1 \leq i \leq np$ **do**
  forall j *where* $1 \leq j \leq nk$ **do**
    $Prv_i^j = Random(1, q-1)$
    $Pub_i^j = g^{Prv_i^j}$
    Gửi tới bên khai phá: Pub_i^j
  end
end
3. Gửi đi dữ liệu: Các bên sở hữu dữ liệu U_i thực hiện
forall i *where* $1 \leq i \leq np$ **do**
  $j = 1$
  forall t *where* $1 \leq t \leq nk-1$ **do**
    forall k *where* $t+1 \leq k \leq nk$ **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
  Gửi tới bên khai phá: p_i^j
end

2. Tính toán các khóa công khai dùng chung: Bên khai phá thực hiện
forall j *where* $1 \leq j \leq nk$ **do**
  $Pub^j = 1$
  forall i *where* $1 \leq i \leq np$ **do**
    $Pub^j = Pub^j * Pub_i^j$
  end
  Gửi tới tất cả bên sở hữu dữ liệu: Pub^j
end
4. Trích xuất kết quả: Bên khai phá thực hiện
forall j *where* $1 \leq j \leq ns$ **do**
  $K^j = 1$
  forall i *where* $1 \leq i \leq np$ **do**
    $K^j = K^j * p_i^j$
  end
end
Giải các bài toán $g^{Sum_j} = K^j$ ($Sum_j \in \{0, 1, ..., np\}$, $j \in \{1, ..., ns\}$) bằng cách thực thi thuật toán vét cạn một lần

3.3. Kết luận

Trong chương này, luận án đã phát triển giải pháp cho bài toán bỏ phiếu điện tử đầu cuối an toàn và cây phân lớp Naive Bayes có đảm bảo tính riêng tư cho mô hình dữ liệu phân tán ngang dựa trên các giao thức SMS mới ở chương 2. Các kết quả thực nghiệm đã chứng minh các đề xuất này không chỉ thỏa mãn những đòi hỏi từ ứng dụng thực tế mà còn mang những ưu điểm vượt trội so với những giải pháp hiện có.

KẾT LUẬN

Luận án đã đề xuất ba giao thức tính tổng bảo mật nhiều thành viên cũng như phát triển hai ứng dụng thực tế dựa trên các giao thức này. Với mỗi đề xuất, luận án đã phân tích khía cạnh an toàn và đánh giá hiệu năng. Các kết quả chính của luận án được tổng kết lại như sau:

- Trong kết quả đầu tiên, luận án đã đề xuất ba giao thức tính tổng bảo mật nhiều thành viên mới, đó là giao thức tính tần suất đảm bảo tính riêng tư dựa trên biến thể trên đường cong elliptic của hệ mã hóa ElGamal, giao thức tính tổng bảo mật nhiều thành viên không cần thiết lập trước các kênh kết nối an toàn/xác thực, và giao thức tính đa tổng trong một vòng tính toán. Các giao thức đề xuất trên an toàn và hiệu quả để được triển khai trong các ứng dụng thực tế.
- Trong kết quả thứ hai, luận án đã phát triển giải pháp mới cho hai ứng dụng thực tế dựa trên các giao thức đề xuất, đó là hệ thống bỏ phiếu đầu cuối an toàn không cần thiết lập trước các kênh kết nối an toàn/xác thực, và một phương thức hiệu quả và thực tế cho kỹ thuật phân lớp Naive Bayes đảm bảo tính riêng tư. Các kết quả phân tích về khía cạnh an toàn đã chứng minh rằng những giải pháp mới này thỏa mãn được các yêu cầu của những ứng dụng và hiệu quả.

Tiếp theo, luận án thảo luận một số bài toán nghiên cứu tiềm năng trong lĩnh vực SMC nói chung trong tương lai.

- Vấn đề đầu tiên là dựa trên các kịch bản tính toán phân tán mới hoặc các yêu cầu từ các bài toán thực tế, những giao thức SMC mới cần được đầu tư nghiên cứu.
- Vấn đề thứ hai là các giao thức SMC nên được dựa trên nền tảng mật mã hiện đại (ví dụ các kỹ thuật mật mã hậu lượng tử) nhằm mục tiêu chống lại các tấn công tiềm ẩn ở thế hệ tính toán tiếp theo đang đến gần.
- Vấn đề thứ ba là, bởi vì khả năng ứng dụng của các giao thức SMC trong nhiều lĩnh vực khác nhau, nên những giao thức này nên được xem xét trong việc triển khai ứng dụng thực tế. Điều này không chỉ giúp cho các tác vụ tính toán phân tán được thực thi thuận lợi hơn và còn bảo vệ sự an toàn của các bên tham gia.

DANH MỤC CÁC BÀI BÁO ĐÃ XUẤT BẢN LIÊN QUAN ĐẾN LUẬN ÁN

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.