MỤC LỤC

PHẦN MỞ ĐẦU

1. Lí do chọn đề tài

Hệ tư vấn người dùng (Recommender Systems –RS) nhằm giải quyết vấn đề quá tải thông tin đó đồng thời cá nhân hóa trải nghiệm người dùng bằng cách cung cấp các đề xuất chính xác, được cá nhân hóa. Tuy nhiên, các hệ thống này yêu cầu người dùng chia sẻ thông tin của mình với máy chủ tư vấn trung tâm, điều này dẫn đến các lo ngại về tính riêng tư dữ liệu của người dùng. Vì vậy cần có giải pháp đảm bảo tính riêng tư cho RS. Các phương pháp được sử dụng để bảo vệ tính riêng tư của RS có thể được chia thành hai loại [29]: dựa trên thống kê, dựa trên mật mã. Tuy nhiên phương pháp dựa trên thống kê gây ra thêm nhiễu, vì thế phải đánh đổi giữa tính riêng tư và độ chính xác. Phương pháp dựa trên mật mã là một phương pháp an toàn hơn nhưng tiêu tốn tài nguyên tính toán. Mà chất lượng tư vấn là một yếu tố vô cùng quan trọng đối với một hệ tư vấn người dùng. Vì thế Luận án này thực hiện việc xây dựng giải pháp phân tích dữ liệu đảm bảo tính riêng tư dựa trên tính toán bảo mật (SMC).

2. Tính mới trong khoa học của Luận án

Những đóng góp khoa học mới của Luận án bao gồm:

- Phát triển 02 giao thức tính toán bảo mật hiệu quả.
- Đề xuất 02 giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư sử dụng các giao thức đã được đề xuất.

3. Cấu trúc của Luận án

Ngoài phần mở đầu, kết luận và phụ lục, nội dung của Luận án được bố cục thành 3 chương như sau:

CHƯƠNG 1. Tổng quan về hệ tư vấn người dùng có đảm bảo tính riêng tư và tính toán bảo mật.

CHƯƠNG 2. Phát triển một số giao thức tính toán bảo mật hiệu quả.

CHƯƠNG 3. Giải pháp đảm bảo tính riêng tư cho hệ tư vấn người dùng dựa trên các giao thức tính toán bảo mật.

Chương 1. TỔNG QUAN VỀ HỆ TƯ VẤN NGƯỜI DÙNG CÓ ĐẢM BẢO TÍNH RIÊNG TƯ VÀ TÍNH TOÁN BẢO MẬT

1.1. Mở đầu chương

Chương này trình bày tóm lược về hệ tư vấn người dùng, bài toán đảm bảo tính riêng tư cho hệ tư vấn người dùng và từ đó xác định những vấn đề cần được giải quyết.

1.2. Hệ tư vấn người dùng

1.3. Những rủi ro về vi phạm tính riêng tư trong hệ tư vấn người

Các rủi ro về vi phạm tính riêng tư của người dùng có thể xảy ra trong cả ba giai đoạn của RS [29]: Giai đoạn mô hình hóa người dùng, Giai đoạn tính toán và Giai đoạn tạo ra kết quả tư vấn.

Việc lưu trữ tập trung dữ liệu của nhưng người dùng hệ thống dù ở dạng này hay dạng khác cũng đem đến các mối lo ngại về quyền riêng tư của dữ liệu. Vì vậy, luận án sẽ tập trung giải quyết bài toán hệ tư vấn cho mô hình dữ liệu phân tán..

1.4. Các kỹ thuật đảm bảo tính riêng tư cho hệ tư vấn

Các kỹ thuật đảm bảo tính riêng tư gồm có [5]: kỹ thuật dựa trên mật mã, kỹ thuật dựa trên nhiễu, Mã thông báo và kỹ thuật dựa trên mặt nạ . Trong phần này, các kỹ thuật đảm bảo tính riêng tư cho hệ tư vấn được trình bày và so sánh các ưu, nhược điểm của chúng.

Đối với bài toán RS thì chất lượng của hệ tư vấn (tính đúng đắn) là quan trọng nhất, nó có vai trò quyết định trong việc có đưa hệ tư vấn này vào sử dụng hay không. Do đó, nghiên cứu sinh lựa chọn kỹ thuật dựa trên mật mã để đảm bảo tính riêng tư cho bài toán cần giải quyết của luận án vì kỹ thuật này có thể đảm bảo được tính đúng đắn và dễ dàng đáp ứng các ràng buộc an toàn cao.

1.5. Tính toán bảo mật

Trong phần này, luận án giới thiệu một số khái niệm về tính toán bảo mật và giao thức tính toán bảo mật.

1.5.1. Các thực thể trong giao thức SMC

Trong mô hình tính toán của SMC gồm ba loại thực thể [44]: (1) các bên trung thực (honest parties), (2) các bên bán trung thực (corrupted parties) và (3) đối thủ bên ngoài (external adversary).

1.5.2. Định nghĩa an toàn

1.5.2.1. Mô hình định nghĩa

Có hai cách tiếp cận để định nghĩa mô hình an toàn trong SMC: dựa trên lý thuyết trò chơi (game-based) và dựa trên mô phỏng (simulation). Trong đó cách tiếp cận dựa trên mô phỏng có lợi thế và đầy đủ hơn [44].

1.5.2.2. Mô hình kẻ tấn công

Luận án sẽ tập trung vào hai tham số chính để xác định mô hình kẻ tấn công cho một giao thức SMC [44]: hành vi đối nghịch được cho phép, chiến lược mua chuộc của đối thủ.

1.5.2.2.1. Hành vi của đối thủ

Đây là tham số quan trọng nhất để xác định mô hình đối thủ cho một giao thức SMC [44]. Có hai loại đối thủ chính: đối thủ bán trung thực và đối thủ độc hại. Luận án tập trung vào mô hình bán trung thực.

1.5.2.2.2. Chiến lược mua chuộc

Có ba mô hình chính [44]: Mô hình mua chuộc tĩnh, Mô hình mua chuộc thích ứng và Mô hình bảo mật chủ động. Trong đó, Luận án sẽ tập trung giải quyết các bài toán trong mô hình mua chuộc thích ứng với số lượng của các bên không trung thực sẽ bị giới hạn tuỳ thuộc vào từng bài toán cụ thể.

1.5.2.2.3. Một số tham số khác

Bên cạnh hai tham số chính xác định mô hình đối thủ, còn có một số tham số khác như: Kênh liên lạc, Sức mạnh tính toán của đối thủ, Giới hạn trên đối với số lượng các bên không trung thực. Trong mô hình đối thủ được xem xét trong luận án như sau: các bên liên lạc với nhau qua kênh kênh an toàn, đối thủ bị giới hạn về mặt tính toán, nghĩa là nó chạy trong thời gian đa thức và giới hạn trên đối số lượng các bên thông đồng thông qua thiết lập kích thước cụ thể của một tập con các bên thông đồng .

1.5.2.3. Khái niệm không thể phân biệt về mặt tính toán

Phần này trình bày về một khái niệm quạn trọng giúp định nghĩa về an toàn của giao thức SMC: không thể phân biệt về mặt tính toán.

1.5.2.4. Định nghĩa về an toàn của giao thức SMC

Trong phần này, luận án trình bày định nghĩa về an toàn của giao thức SMC đối với mô hình bán trung thực (Định nghĩa 1.5 [24]) và định lý quan trọng liên quan đến quá trình chứng minh tính riêng tư của giao thức đề xuất (Định lí 1.1 [24]).

Định nghĩa 1.5 [24]. *(Tính riêng tư đối với mô hình bán trung thực) Cho f là một hàm tính toán nhiều thành viên như được định nghĩa trong Định nghĩa 1.1.*

• Trong trường hợp f là hàm tất định: giao thứcπ tính toán an toàn hàm f với t bên tham gia không trung thực nếu $\forall I \subseteq \{1,2,\dots,n\}$ sao cho $|I| = t$, tồn tại thuật toán xác suất thời gian đa thức M sao cho:

$$\{M(I, \bar{x}_I,\ f_I(\bar{x}))\}_{\bar{x}\in(\{0,1\}^*)^n} \overset{c}{\equiv} \{VIEW_{A,I}^{\pi}(\bar{x})\}_{\bar{x}\in(\{0,1\}^*)^n} \qquad (1.5.5)$$

• Trong trường hợp tổng quát: giao thứcπ tính toán an toàn hàm f với t bên tham gia không trung thực nếu $\forall I \subseteq \{1,2,\dots,n\}$ sao cho $|I| = t$, tồn tại thuật toán xác suất thời gian đa thức M sao cho:

$$\{M(I, \bar{x}_I,\ f_I(\bar{x}), f(\bar{x}))\}_{\bar{x}\in(\{0,1\}^*)^n} \overset{c}{\equiv} \{VIEW_{A,I}^{\pi}(\bar{x}), OUTPUT^{\pi}(\bar{x})\}_{\bar{x}\in(\{0,1\}^*)^n} \qquad (1.5.6)$$

Trong đó:

• $VIEW_{A,I}^{\pi}(\bar{x})$: những gì mà t bên thông đồng quan sát được và tất cả các thông điệp (được truyền giữa các bên trung thực) mà kẻ tấn công A nghe lén trong quá trình thực thi giao thức π với đầu vào $\bar{x} = (x_1, x_2, \dots, x_n)$.

• $\bar{x}_I$: đầu vào của tất cả các bên tham gia bị mua chuộc trong giao thức π.

• $f_I(\bar{x})$: chuỗi đầu ra của tất cả các bên tham gia bị mua chuộc trong giao thức π.

• $OUTPUT^{\pi}(\bar{x})$: chuỗi đầu ra của tất cả các thành viên tham gia giao thức π. Trong trường hợp đầu tiên, $OUTPUT^{\pi}(\bar{x}) \equiv f(\bar{x})$.

• $\overset{c}{\equiv}$ là không thể phân biệt được về mặt tính toán (mô hình tính toán (computational model)).

1.5.2.5. Một số nguyên thuỷ mật mã

1.6. Các nghiên cứu liên quan

Phần này, luận án tiến hành phân tích, đánh giá các giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư (PPRS) dựa trên mật mã hiện có về tính đúng đắn, tính riêng tư và hiệu suất.

1.7. Những vấn đề cần được nghiên cứu của Luận án

1.7.1. Những vấn đề cần được nghiên cứu trong hệ tư vấn người dùng có đảm bảo tính riêng tư với mô hình dữ liệu phân tán đầy đủ

1.7.1.1. Mô hình dữ liệu phân tán đầy đủ

1.7.1.2. Vấn đề cần được nghiên cứu

Từ những đánh giá, luận án tập trung vào vấn đề nâng cao hiệu suất của PPRS cho mô hình dữ liệu phân tán đầy đủ thông qua việc cải tiến hiệu suất của giai đoạn tính trung bình xếp hạng của các mục tin, độ tương tự giữa các cặp mục tin và giai đoạn sinh tư vấn cho người dùng.

Trong hệ thống sẽ sử dụng các xếp hạng của người dùng đối với các sản phẩm để đánh giá mức độ tương tự giữa những sản phẩm cho mục tạo ra các khuyến nghị cho người dùng theo hai cách: dựa trên nội dung (CBF) và dựa trên lọc cộng tác (CF) theo hai công thức (1.7.11) và (1.7.12) tương ứng.

$$P_{i,k} = \frac{\sum_{j=1}^{m} r_{i,j}.S(i_j, i_k)}{\sum_{j=1}^{m} S(i_j, i_k)} \quad (1.7.11)$$

$$P_{i,k} = \frac{R_k.\sum_{j=1}^{m}.S(i_k, i_j) + \sum_{j=1}^{m}(r_{i,j} - R_j).S(i_k, i_j)}{\sum_{j=1}^{m} S(i_k, i_j)} \quad (1.7.12)$$

Trong đó:

$$R_j = \frac{\sum_{i=1}^{n} r_{i,j}}{\sum_{i=1}^{n} e_{i,j}} \quad (1.7.13)$$

$$S(i_j, i_k) = \frac{\sum_{i=1}^{n} r_{i,j}.r_{i,k}}{\sqrt{\sum_{i=1}^{n} {r_{i,j}}^2}.\sqrt{\sum_{i=1}^{n} {r_{i,k}}^2}} \quad (1.7.14)$$

1.7.2. Những vấn đề cần được nghiên cứu trong hệ tư vấn người dùng có đảm bảo tính riêng tư với mô hình dữ liệu phân tán đầy đủ hai bên

1.7.2.1. Mô hình dữ liệu phân tán đầy đủ hai bên

1.7.2.2. Vấn đề cần được nghiên cứu

Một vấn đề nữa có thể thấy được qua phần đánh giá về các giải pháp hiện có là chưa hề có giải pháp PPRS nào cho mô hình dữ liệu 2PFD. *Vì thế, luận án sẽ nghiên cứu để đề xuất giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư cho mô hình 2PFD sử dụng SMC.*

Kết luận chương

Chương 2. PHÁT TRIỂN MỘT SỐ GIAO THỨC TÍNH TOÁN BẢO MẬT HIỆU QUẢ

2.1. Mở đầu chương

Chương này sẽ trình bày các giao thức được phát triển phục vụ cho việc tính các giá trị trung bình và các giá trị độ tương tự có bảo mật trong mô hình dữ liệu phân tán đầy đủ và trong mô hình dữ liệu 2PFD. Những đề xuất này có liên quan đến Công trình [CT1, CT3, CT4, CT5, CT6]. Các giao thức đề xuất tập trung đảm bảo tính riêng tư trong mô hình bán trung thực, trong đó các bên tham gia sẽ tuân thủ giao thức nhưng các bên cũng có thể thông đồng với nhau để tìm hiểu các thông tin nhạy cảm của các bên tham gia khác. Tất cả người dùng có một kênh được xác thực riêng tư với trung tâm tính toán. Các tham số của các hệ mật được sử dụng trong luận án được giả định là thoả mãn yêu cầu sao cho bài toán logarit rời rạc trên trường hữu hạn và trên đường cong Elliptic là bài toán khó.

2.2. Giao thức tính đồng thời nhiều giá trị trung bình và độ tương tự có bảo mật

2.2.1. Ý tưởng

Mô hình bài toán gồm n bên tham gia $U = \{U_1, U_2, \dots, U_n\}$, mỗi bên U_i nắm giữ m giá trị bí mật là số nguyên không âm nhỏ hoặc vừa $r_{i,j}$ $(1 \le j \le m)$, nếu $r_{i,j} \neq 0$ thì $f_{i,j} = 1$, ngược lại $f_{i,j} = 0$. Một trung tâm tính toán (CC) cần tính các giá trị trung bình của các giá trị bí mật $r_i \neq 0$ mà các bên này nắm giữ theo công thức (1.7.13) mà không tiết lộ $r_{i,j}$, $f_{i,j}$ và tính các giá trị độ tương tự của cặp mục tin dựa trên độ tương tự cosin như công thức (1.7.14) mà không tiết lộ $r_{i,j}, r_{i,k}$.

2.2.1.1. Tính trung bình có bảo mật

2.2.1.2. Tính độ tương tự có bảo mật

Mục 2.2.1.1 và 2.2.1.2 trình bày phân tích đầy đủ các giao thức SMAC, SMSC phổ biến nhất liên quan đến Luận án và nhận thấy có thể đề xuất giao thức nhiều giá trị trung bình và độ tương tự hiệu quả được dựa trên ý tưởng gồm hai bước chính như sau:

Bước 1: Thiết kế lại giao thức trong [7] bằng cách:

- Ở pha khởi tạo: chỉ yêu cầu mỗi U_i sử dụng $n_k = \left\lceil \frac{1}{2} + \sqrt{m(m+5) + \frac{1}{4}} \right\rceil$ cặp khoá, tương ứng CC chỉ tính n_k khoá công khai chung dùng cho quá trình mã hoá các giá trị riêng tư của người dùng thay vì $\frac{m(m+5)}{2} + 1$ cặp khoá.

- Ở pha 1: yêu cầu mỗi U_i thực hiện mã hoá các giá trị riêng tư của mình bằng khoá công khai chung và giải mã thành phần sử dụng khoá bí mật tương ứng của mình.

- Ở pha 2: CC chỉ tính các bản mã tổng hợp của các giá trị trung bình, các giá trị độ tương tự và thực hiện tính logarit rời rạc một lần tất cả các giá trị này.

Bước 2: Chuyển đổi giao thức được thiết kế lại ở bước 1 sang một biến thể sử dụng hệ mật mã trên đường cong Elliptic.

Đề xuất này được công bố trong các công trình [CT3, CT4, CT6].

2.2.2. Giao thức đề xuất

Luận án thực hiện thiết kế lại giao thức trong công trình [7] theo ý tưởng cải tiến số 1 đã được đề xuất trong mục 2.2.1 và được giao thức tính đồng thời nhiều giá trị trung bình và độ tương tự bảo mật cải tiến được mô tả như trong Giao thức 2.3. Giao thức sử dụng hệ mật mã Elgamal với các tham số như được mô tả trong Mục 1.5.2.5.1.

2.2.3. Phân tích tính đúng đắn

Luận án đã chứng minh được tính đúng đắn của giao thức đề xuất.

2.2.4. Phân tích tính riêng tư

Luận án đã chứng minh giao thức cải tiến có mức độ an toàn cao hơn so với giao thức gốc [7] và tương đương với giao của nhóm tác giả Verma và các cộng sự [75] theo mệnh đề sau:

Mệnh đề 2.1. *Giao thức tính đồng thời nhiều giá trị trung bình và độ tương tự có bảo mật đảm bảo tính riêng tư của mỗi người dùng trong mô hình bán trung thực trong trường hợp không có sự thông đồng giữa các bên. Trong trường hợp có sự thông đồng lên đến tối đa n-2 người dùng, giao thức vẫn bảo vệ tính riêng tư của những người dùng trung thực.*

Đầu vào:
n: Số lượng người dùng.
U_i: Người dùng thứ i, với $i \in [1,n]$.
m: Số lượng các giá trị trung bình cần tính.
RM: Ma trận xếp hạng của những người dùng đối với các mục tin, trong đó:$r_{i,j}$: Các số nguyên không âm nhỏ hoặc vừa, là các giá trị bí mật mà mỗi U_i nắm giữ ($0 \le j < m$).
max: Giá trị lớn nhất trong các giá trị bí mật mà các U_i nắm giữ.
Đầu ra: Tại trung tâm tính toán (CC):
$R_j = \frac{\sum_{i=1}^{n} r_{i,j}}{\sum_{i=1}^{n} f_{i,j}}$, $S_{j,k} = \frac{\sum_{i=1}^{n} r_{i,j} \cdot r_{i,k}}{\sqrt{\sum_{i=1}^{n} r_{i,j}^2} \cdot \sqrt{\sum_{i=1}^{n} r_{i,k}^2}}$, với $0 \le j < k < m$.

*** Pha khởi tạo:**
- Mỗi U_i thực hiện:
1. for $(0 \le j < n_k)$
2. $ksu_{i,j} \in_R \mathbb{Z}_q^*; KPU_{i,j} = g^{ksu_{i,j}};$
3. Gửi $\{KPU_{i,j}\}_{j=0}^{n_k-1}$ cho CC.
- CC thực hiện:
4. for $(0 \le j < n_k)$
5. $KP_j = \prod_{i=1}^{n} KPU_{i,j};$
6. Gửi $\{KP_j\}_{j=0}^{n_k-1}$ cho $\{U_i\}_{i=1}^{n};$

*** Pha 1: Mỗi U_i thực hiện**
7. for $(0 \le j < m)$
8. $a_{i,j} = r_{i,j}; f_{i,j} = 0;$
9. if $(r_{i,j} \ne 0)$ $f_{i,j} = 1;$
10. for $(m \le j < 2m)$
11. $a_{i,j} = f_{i,j-m};$
12. for $(2m \le j < 3m)$
13. $a_{i,j} = r_{i,j-2m} \cdot r_{i,j-2m};$
14. for $(0 \le t < m-1)$
15. for $(t+1 \le k < m)$
16. $a_{i,j} = r_{i,t} \cdot r_{i,k};$
17. j++;
18. j = 0;
19. for $(0 \le t < n_k - 1)$
20. for $(t+1 \le k < n_k)$
21. $AU_{i,j} = g^{a_{i,j}} \cdot KP_k^{-ksu_{i,t}} \cdot KP_t^{ksu_{i,k}};$
22. j++;
23. if $(j == n_s - 1)$ break;
24. if $(j == n_s - 1)$ break;
25. Gửi $\{AU_{i,j}\}_{j=0}^{n_s-1}$ cho CC.

*** Pha 2: CC thực hiện**
26. for $(0 \le j < n_s)$
27. $A_j = \prod_{i=1}^{n} AU_{i,j};$
28. $sa = \mathrm{Dlog}(p, g, \max^2, A);$
/*Sử dụng thuật toán vét cạn một lần với $A = \{A_j, j \in [1,n_s]\}$, $sa = \{sa_j, j \in [1,n_s]\}$*/.
29. for $(0 \le j < m)$
30. $R_j = \frac{sa_j}{sa_{j+m}};$
31. $l = 0;$
32. for $(0 \le j < m-1)$
33. for $(j+1 \le k < m)$
34. $S_{j,k} = \frac{sa_{3m+l}}{\sqrt{sa_{2m+j}} \cdot \sqrt{sa_{2m+k}}};$
35. l++;

Giao thức 2.3. *Giao thức tính đồng thời nhiều giá trị trung bình và độ tương tự có bảo mật*

2.2.5. Phân tích tính hiệu quả

Luận án thực hiện đánh giá, so sánh về chi phí truyền thông, chi phí tính toán giữa giao thức ban đầu [7], giao thức cải tiến [75] và giao thức được đề xuất. Kết quả cho thấy giao thức đề xuất có truyền thông và chi phí tính toán thấp hơn các giao thức còn lại.

2.2.5.1. Đánh giá về chi phí truyền thông

2.2.5.2. Đánh giá về chi phí tính toán

2.2.6. Chuyển đổi giao thức đề xuất sử dụng hệ mật đường cong Elliptic

Để tiếp tục nâng cao hiệu suất của giao thức đề xuất, luận án chuyển đổi giao thức đề xuất sang một biến thể sử dụng hệ mật mã trên đường cong Elliptic theo ý tưởng cải tiến số 2 đã được đề xuất trong mục 2.2.1. Luận án cũng so sánh chi tiết về chi phí tính toán giữa giao thức đề xuất trước và sau khi chuyển đổi sử dụng hệ mật đường cong Elliptic. Từ kết quả so sánh trên cho thấy việc chuyển đổi giao thức đề xuất sang sử dụng hệ mật mã Elgamal trên đường cong Elliptic giúp cải thiện đáng kể về cả chi phí truyền thông và chi phí tính toán của giao thức.

2.2.7. Đánh giá về mặt thực nghiệm

Luận án tiến hành thực thi giao thức đề xuất, giao thức gốc được đánh giá có hiệu suất tốt nhất [7] và giao thức được đánh giá có hiệu suất khá tốt nhưng có mức độ an toàn cao [75] để so sánh về thời gian thực thi và nhận thấy giao thức đề xuất có thời gian thực thi thấp hơn nhiều so với các giao thức còn lại.

2.2.8. Ví dụ minh hoạ

Phần này trình bày một ví dụ minh hoạ dưới dạng số cụ thể đối với giao thức được đề xuất.

2.3. Giao thức tính nhiều giá trị tần suất có bảo mật nhiều thành viên trong mô hình 2PFD

2.3.1. Ý tưởng

Mô hình của bài toán gồm hai nhóm người dùng $U = \{U_1, U_2, \dots, U_{n_1}\}$ và $V = \{V_1, V_2, \dots, V_{n_1}\}$, $n = 2n_1$, trong đó mỗi người dùng U_i, V_i giữ các giá trị nhị phân riêng của mình$\{u_{i,0}, u_{i,1}, \dots, u_{i,m_u-1}\}$, $\{v_{i,0}, v_{i,1}, \dots, v_{i,m_v-1}\} \in \{0,1\}$. Mỗi cặp người dùng (U_i, V_i) sở hữu một bản ghi trong đó người dùng U_i biết các giá trị của một tập con các thuộc tính thích hợp, và người dùng V_i biết các giá trị của các thuộc tính còn lại. CC cần tính tập hợp $m_u.m_v$ giá trị tần suất $SU = \{su_j = \sum_{i=1}^{n_1} u_{i,\lfloor j/m_v \rfloor}.v_{i,j\%m_v}\}_{j\in[0,m_u.m_v-1)}$.

Luận án đánh giá các giao thức tính tần suất bảo mật nhiều thành viên trong mô hình 2PFD hiện có và nhận thấy có thể đề xuất một giao thức tính nhiều giá trị tần suất có đảm bảo tính riêng tư mới trong mô

hình 2PFD thông qua việc cải tiến hiệu suất của giao thức ban đầu [27] dựa trên ý tưởng các bước như sau:

Bước 1: Thiết kế lại các bước trong giao thức gốc [27] bằng cách:

- Ở pha khởi tạo: chỉ yêu cầu mỗi U_i, V_i sử dụng $n_k + 1$ cặp khoá ($n_k = \lceil \sqrt{s_k} \rceil$), tương ứng CC chỉ tính s_k khoá công khai chung ($s_k = \left\lceil \frac{1}{2} + \sqrt{2m + \frac{1}{4}} \right\rceil$) dùng cho quá trình mã hoá các giá trị riêng tư của người dùng thay vì $2m + 1$ cặp khoá như giao thức gốc.

- Ở pha 1: mỗi U_i mã hoá các giá trị riêng tư sử dụng khoá công khai của mình và gửi các bản mã cho CC.

- Ở pha 2: chỉ yêu cầu mỗi V_i tính $Q_{11}^{(i,j)} = \frac{Q_1^{(i,j)}}{Q_2^{(i,j)}}$, $Q_{21}^{(i,j)} = Q_3^{(i,j)}$, gửi các giá trị này đến CC thay vì tính $Q_1^{(i,j)}$, $Q_2^{(i,j)}$, $Q_3^{(i,j)}$ và gửi các giá trị này đến CC như giao thức gốc.

- Ở pha 3: yêu cầu mỗi U_i tính $\frac{K_1{}^{(i,j)}}{K_2{}^{(i,j)}}$ và gửi giá trị duy nhất này đến CC thay vì tính riêng lẻ $K_1{}^{(i,j)}$, $K_2{}^{(i,j)}$ và gửi cho CC như giao thức gốc.

- Ở pha 4: CC chỉ thực hiện tính bản giải mã tổ hợp của các giá trị tần suất và thực hiện một lần tính các giá trị logarit rời rạc thay vì phải giải mã bản mã tổng và thực hiện tính nhiều lần các giá trị logarit rời rạc như giao thức gốc.

Bước 2: Chuyển đổi giao thức được thiết kế lại sang một biến thể dựa vào hệ mã hóa ElGamal trên đường cong elliptic.

Đề xuất này liên quan đến công trình [CT1, CT5].

2.3.2. Giao thức được đề xuất

Luận án thực hiện thiết kế lại giao thức trong công trình [27] theo ý tưởng cải tiến số 1 đã được đề xuất trong mục 2.3.1 và được giao thức tính tần suất bảo mật nhiều thành viên cải tiến được mô tả như trong Giao thức 2.3. Giao thức sử dụng hệ mật mã Elgamal với các tham số như được mô tả trong Mục 1.5.2.5.1.

Các giá trị $\sum_{i=1}^{n_1} u_{i,\lfloor j/m_v \rfloor} \cdot v_{i,j\%m_v}$ không quá lớn nên việc tính logarit rời rạc không khó.

2.3.3. Phân tích tính đúng đắn

Luận án đã chứng minh tính đúng đắn của giao thức đề xuất.

Đầu vào:

n: Số lượng người dùng trong hệ thống, trong đó mỗi miền dữ liệu U_i, V_i có n_1 người dùng, $i \in [1, n_1], n = 2n_1$.

$u_{i,j}$: Các giá trị nhị phân bí mật mà U_i nắm giữ $(0 \le j < m_u)$.

$v_{i,j}$: Các giá trị nhị phân bí mật mà V_i nắm giữ $(0 \le j < m_v)$, $m_u \le m_v$.

Output (đối với CC):

$SU = \{su_j = \sum_{i=1}^{n_1} u_{i,j\%m_v}.v_{i,\lfloor \frac{j}{m_v} \rfloor}\}_{j \in [0, m_u.m_v - 1)}$

*** Pha khởi tạo:**

- Mỗi U_i thực hiện:

1. $x_i \in_R \mathbb{Z}_q^*; X_i = g^{x_i}$;
2. for $(0 \le j < n_{k1})$
3. $ksu_{i,j} \in_R \mathbb{Z}_q^*; KPU_{i,j} = g^{ksu_{i,j}}$;
4. Gửi $\{KPU_{i,j}\}_{0 \le j < n_{k1}}, X_i$ cho CC;

- Mỗi V_i thực hiện:

5. $y_i \in_R \mathbb{Z}_q^*; Y_i = g^{y_i}$;
6. for $(0 \le j < n_{k1})$
7. $ksv_{i,j} \in_R \mathbb{Z}_q^*; KPV_{i,j} = g^{ksv_{i,j}}$;
8. Gửi $\{KPV_{i,j}\}_{0 \le j < n_{k1}}$ cho CC;

- CC thực hiện:

9. j=0;
10. for $(0 \le t < n_{k1})$
11. for $(0 \le k < n_{k1})$
12. $KP_j = \prod_{i=1}^{n_1}(KPU_{i,t}.KPV_{i,k})$; j++;
13. if $(j == n_k)$ break;
14. if $(j == n_k)$ break;
15. Gửi $\{KP_j\}_{0 \le j < n_k}$ cho $\{U_i, V_i\}_{1 \le i \le n}$;

*** Pha 1: Mỗi U_i thực hiện**

16. for $(0 \le j < m_u)$
17. $c_{i,j}^{(1)} \in_R \mathbb{Z}_q^*$;
18. $C_1^{(i,j)} = g^{u_{i,j}}.X_i^{c_{i,j}}; C_2^{(i,j)} = g^{c_{i,j}}$;
19. Gửi $\{C_1^{(i,j)}, C_2^{(i,j)}\}_{0 \le j < m_u}$ cho CC;

*** Pha 2: Mỗi V_i thực hiện**

20. Nhận $\{C_1^{(i,j)}, C_2^{(i,j)}\}_{0 \le j < m_u}, X_i$ từ CC;
29. j=0;
21. for $(0 \le t < n_k - 1)$
22. for $(t + 1 \le k < n_k)$
23. $c_{i,j}^{(2)} \in_R \mathbb{Z}_q^*$;
24. $Q_1^{(i,j)} = (C_1^{(i,\lfloor j/m_v \rfloor)})^{v_{i,j\%m_v}}.$ $KP_t^{ksv_{i,k\%n_{k1}}}.(C_2^{(i,\lfloor j/m_v \rfloor)})^{-c_{i,j}^{(2)}.y_i}.KP_k^{-ksv_{i,t\%n_{k1}}}$
25. $Q_2^{(i,j)} = Y_i^{c_{i,j}^{(2)}}.X_i^{-v_{i,j\%m_v}}$; j++;
26. if $(j == m_u.m_v - 1)$ break;
27. if $(j == m_u.m_v - 1)$ break;
28. Gửi $\{Q_1^{(i,j)}, Q_2^{(i,j)}\}_{0 \le j < m_u.m_v}$ cho CC;

*** Pha 3: Mỗi U_i thực hiện:**

29. Lấy $\{Q_1^{(i,j)}, Q_2^{(i,j)}\}_{0 \le j < m_u.m_v}$ từ CC;
30. j=0;
31. for$(0 \le t < n_k - 1)$
32. for $(t + 1 \le k < n_k)$
33. $A_{i,j} = Q_1^{(i,j)}.(Q_2^{(i,j)})^{c_{i,\lfloor j/m_v \rfloor}^{(1)}}.$ $KP_k^{-ksu_{i,\lfloor t/n_{k1} \rfloor}}.KP_t^{ksu_{i,\lfloor k/n_{k1} \rfloor}}$;
34. j++;
35. if $(j == m_u.m_v - 1)$ break;
36. if $(j == m_u.m_v - 1)$ break;
37. Gửi $\{A_{i,j}\}_{0 \le j < m_u.m_v}$ cho CC;

*** Pha 4: CC thực hiện:**

38. for $(0 \le j < m_u.m_v)$
39. $A_j = \prod_{i=1}^{n_1} A_{i,j}$;
40. $SU = Dlog(p, g, n_1, A)$;

//Sử dụng thuật toán vét cạn 1 lần

Giao thức 2.6. *Giao thức tính đồng thời nhiều giá trị tần suất có bảo mật trong mô hình 2PFD*

2.3.4. Phân tích tính riêng tư

Phần này đã chứng minh được giao thức được đề xuất có cùng mức độ an toàn với giao thức ban đầu [27] theo mệnh đề sau:

Mệnh đề 2.4. *Giao thức tính nhiều giá trị tần suất có bảo mật trong mô hình 2PFD bảo vệ tính riêng tư của mỗi người dùng trung thực trong mô hình bán trung thực. Trong trường hợp có sự thông đồng của các bên tham gia, giao thức bảo vệ tính riêng tư của người dùng trung thực chống lại trung tâm tính toán và tối đa n-2 người dùng không trung thực. Trong trường hợp chỉ có hai người dùng trung thực, điều này vẫn chính xác miễn là hai người dùng trung thực không sở hữu các giá trị thuộc tính của cùng một bản ghi.*

2.3.5. Phân tích tính hiệu quả

Trong phần này, luận án tiến hành phân tích và so sánh tính hiệu quả của giao thức đề xuất với giao thức gốc [27] và giao thức gần đây [50] về chi phí giao tiếp, chi phí tính toán. Kết quả cho giao thức đề xuất hiệu quả hơn các giao thức còn lại.

2.3.5.1. Đánh giá về chi phí truyền thông

2.3.5.2. Đánh giá về chi phí tính toán

2.3.6. Chuyển đổi giao thức đề xuất sử dụng hệ mật đường cong Elliptic

Để tiếp tục nâng cao hiệu suất của giao thức đề xuất, luận án chuyển đổi các giao thức được thiết kế lại sang một biến thể sử dụng hệ mật mã trên đường cong Elliptic theo ý tưởng cải tiến số 2 đã được đề xuất trong mục 2.3.1. Luận án cũng so sánh chi tiết về chi phí tính toán giữa giao thức đề xuất trước và sau khi chuyển đổi sử dụng hệ mật đường cong Elliptic. Từ kết quả so sánh trên cho thấy việc chuyển đổi giao thức đề xuất sang sử dụng hệ mật mã Elgamal trên đường cong Elliptic giúp cải thiện thêm đáng kể về cả chi phí truyền thông và chi phí tính toán của giao thức.

2.3.7. Đánh giá về mặt thực nghiệm

Luận án tiến hành thực thi giao thức đề xuất, giao thức gốc [27] và giao thức [50] để so sánh về thời gian thực thi và nhận thấy giao thức đề xuất có thời gian thực thi thấp hơn so với các giao thức còn lại.

2.3.8. Ví dụ minh hoạ

Phần này trình bày một ví dụ minh hoạ dưới dạng số cụ thể đối với giao thức được đề xuất.

Kết luận chương

Trong chương này, luận án đã phát triển 02 giao thức tính toán bảo mật hiệu quả: Giao thức tính đồng thời nhiều giá trị trung bình, độ tương tự có bảo mật và Giao thức tính đồng thời nhiều giá trị tần suất có bảo mật trong mô hình 2PFD.

Đồng thời cũng đã chứng minh rằng các giao thức đề xuất vừa đảm bảo được tính đúng đắn của kết quả đầu ra, vừa đảm bảo được độ an toàn trong mô hình bán trung thực. Các đánh giá về mặt lí thuyết và thực nghiệm cũng cho thấy tính hiệu quả cao của các giao thức này so với các giao thức tính trung bình, giao thức tính độ tương tự được trình bày trong tài liệu [7] và giao thức tính tần suất trong mô hình 2PFD được trình bày trong tài liệu [27], tương ứng.

Do đó, các giao thức đề xuất có khả năng áp dụng vào các bài toán thực tế. Cụ thể như: giao thức tính trung bình và độ tương tự có bảo mật được đề xuất đã được áp dụng vào trong bài toán đảm bảo tính riêng tư cho hệ tư vấn người dùng đã được công bố trong công trình [CT3, CT4, CT5], giao thức tính tần suất có bảo mật trong mô hình 2PFD [CT1] đã được áp dụng vào trong bài toán đảm bảo tính riêng tư trong quá trình huấn luyện cây quyết định đã được công bố trong [CT2].

Nội dung của chương này liên quan đến các công trình [CT1, CT3, CT4, CT5, CT6] trong Danh mục các công trình khoa học sử dụng trong Luận án.

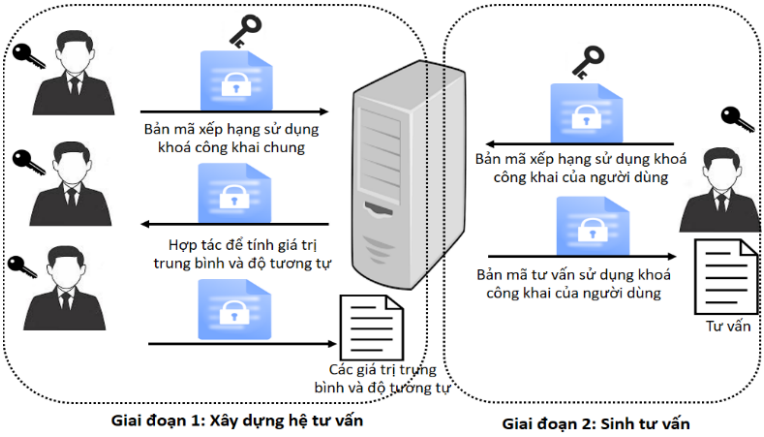
Chương 3. GIẢI PHÁP ĐẢM BẢO TÍNH RIÊNG TƯ CHO HỆ TƯ VẤN NGƯỜI DÙNG DỰA TRÊN CÁC GIAO THỨC TÍNH TOÁN BẢO MẬT

3.1. Mở đầu chương

Từ việc phát triển các giao thức tính toán bảo mật nhiều thành viên phục vụ cho việc tính các giá trị trung bình, độ tương tự và giao thức tính tần suất trong mô hình dữ liệu 2PDF ở chương trước. Các giao thức đề xuất này sẽ được ứng dụng trong giai đoạn đầu tiên của giải pháp Hệ gợi ý có đảm bảo tính riêng tư: giai đoạn tính giá trị trung bình xếp hạng của mục tin và độ tương tự của cặp mục tin. Những đề xuất này có liên quan đến các Công trình [CT3, CT4, CT6].

3.2. Bài toán tư vấn người dùng có đảm bảo tính riêng tư

Phần này thực hiện phân tích các giải pháp PPRS hiện có đáp ứng bài toán xem xét của luận và và xác định ý tưởng chung của các giải pháp được đề xuất trong Luận án được chia thành hai giai đoạn như Hình 3.1.



Hình 3.1. *Giải pháp PPRS [7]*

3.3. Giải pháp đảm bảo tính riêng tư cho Hệ tư vấn người dùng trong mô hình dữ liệu phân tán đầy đủ

3.3.1. Định nghĩa bài toán

Mục này mô tả bài toán hệ tư vấn trong mô hình dữ liệu phân tán đầy đủ.

3.3.2. Giải pháp đảm bảo tính riêng tư cho hệ tư vấn người dùng

Trong giải pháp mới này luận án thiết kế lại cả hai pha của giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư trong [7] nhằm tăng hiệu quả của giải pháp. Giải pháp đề xuất cụ thể sẽ như sau:

- Giai đoạn 1, để tính trung bình xếp hạng của các mục tin và độ tương tự giữa các cặp mục tin luận án sử dụng giao thức SMASC được đề xuất trong Mục 2.2.6.

- Giai đoạn 2, để sinh tư vấn có đảm bảo tính riêng tư cho người dùng mục tiêu, luận án đề xuất 02 phương thức: Sinh tư vấn dựa trên CBF và Sinh tư vấn dựa trên CF có đảm bảo tính riêng tư. Tuy nhiên, giải pháp đề xuất sử dụng hệ mật mã trên đường cong Elliptic thay cho hệ mật mã Elgamal như giải pháp ban đầu nhằm đồng bộ với giai đoạn 1 và nâng cao tính hiệu quả.

Giải pháp được trình bày cụ thể trong ***Giải pháp 3.1***.

3.3.3. Ví dụ minh hoạ

3.3.4. Đánh giá về tính đúng đắn

Phần này luận án chứng minh tính đúng đắn của giải pháp đề xuất.

3.3.5. Đánh giá về tính riêng tư

Luận án cho thấy giải pháp đề xuất có cùng mức độ an toàn với giải pháp của nhóm tác giả Verma và các cộng sự [75] và cao hơn giải pháp ban đầu [7].

3.3.6. Đánh giá tính hiệu quả

Luận án so sánh chi phí truyền thông, chi phí tính toán, thời gian thực thi giữa giải pháp được đề xuất với giải pháp ban đầu [7] và giải pháp của nhóm tác giả Verma và các cộng sự [75]. Kết quả cho thấy giải pháp đề xuất hiệu quả hơn các giải pháp còn lại.

3.3.6.1. Đánh giá về chi phí truyền thông

3.3.6.2. Đánh giá về chi phí tính toán

3.3.6.3. Đánh giá về mặt thực nghiệm

Giải pháp 3.1. *PPRS trong mô hình dữ liệu phân tán đầy đủ*

Đầu vào:
n: Số lượng người dùng.
m: Số lượng các giá trị trung bình cần tính.
RM: Ma trận xếp hạng của những người dùng đối với các mục tin, trong đó mỗi $r_{i,j}$ là giá trị xếp hạng riêng tư của người dùng thứ i đối với mục tin thứ j.
max: Giá trị lớn nhất trong các giá trị bí mật mà các U_i nắm giữ.
k: Chỉ số của mục tin mà người dùng U_i yêu cầu sinh tư vấn.
Đầu ra: (đối với U_i):
$P_{i,k}$: Điểm đề xuất được dự đoán xếp hạng cho mục tin thứ k của người dùng U_i, với $k = \{0,2,\ldots,m-1\}$ là thứ tự mục tin mà U_i đã yêu cầu sinh tư vấn.
Nếu sinh tư vấn dựa trên nội dung – CBF thì: $P_{i,k} = \frac{\sum_{j=1}^{m} r_{i,j} S(i_k,i_j)}{\sum_{j=1}^{m} S(i_k,i_j)}$
Nếu sinh tư vấn dựa trên lọc cộng tác – CF thì: $P_{i,k} = \frac{R_k \sum_{j=1}^{m} S(i_k,i_j) + \sum_{j=1}^{m} (r_{i,j} - R_j) S(i_k,i_j)}{\sum_{j=1}^{m} S(i_k,i_j)}$
R_k, R_j được tính theo công thức (1.7.13) và $S(i_k,i_j)$ được tính theo công thức (1.7.14).

Giai đoạn 1: Tính trung bình xếp hạng của các mục tin và độ tương tự giữa cặp mục tin
1. $\{R_j, S(i_j,i_k)\}_{0 \le j,k < m; j<k} =$ SMASC$(n,m,RM,n.\max^2)$;
/* Thực hiện giao thức SMASC
2. **for** $(0 \le j < m-1)$
3. **for** $(j+1 \le k < m)$
4. $S(i_k,i_j) = S(i_j,i_k)$;

Giai đoạn 2: Sinh tư vấn cho người dùng mục tiêu U_i
*** Pha 1.** U_i **thực hiện**
5. **for** $(0 \le j < m)$
6. $c_j^{(1)} \in_R \mathbb{Z}_d^*$;
7. $E(r_{i,j} \cdot G) = (C_j^{(1)} = r_{i,j} \cdot G + c_j^{(1)} \cdot KPU_{1,1}, C_j^{(2)} = c_j^{(1)} \cdot G)$;
/*E là thuật toán mã hoá của hệ mật mã đường cong Elliptic.*/
8. Gửi $\{E(r_{i,j} \cdot G)\}_{0 \le j < m}$ cho Rs;

*** Pha 2. Rs thực hiện**
9. **for** $(0 \le j < k < m)$
10. $c_k^{(2)} \in_R \mathbb{Z}_d^*$;
11. $E(\sum_{j=1,j\neq k}^{m} S(i_k,i_j) \cdot G) = (F_{3,k} = \sum_{j=1,j\neq k}^{m} S(i_k,i_j) \cdot G + c_k^{(2)} \cdot KPU_{1,1}, F_{4,k} = c_k^{(2)} \cdot G)$;
12. **if** (Phương thức sinh tư vấn = CBF)
13. $(F_{1,k}, F_{2,k}) = \sum_{j=1,j\neq k}^{m} S(i_k,i_j) \cdot E(r_{i,j} \cdot G)$;
14. **else**
15. $c_k^{(4)} \in_R \mathbb{Z}_d^*; c_k^{(5)} \in_R \mathbb{Z}_d^*$;
16. $E(R_k \sum_{j=1,j\neq k}^{m} S(i_k,i_j) \cdot G) = (R_k \sum_{j=1,j\neq k}^{m} S(i_k,i_j) \cdot G + c_k^{(4)} \cdot KPU_{1,1}, c_k^{(4)} \cdot G)$;
17. $E(R_j \cdot G) = (R_j \cdot G + c_k^{(5)} \cdot X_i, c_k^{(5)} \cdot G)$;
18. $(F_{1,k}, F_{2,k}) = E(R_k \sum_{j=1,j\neq k}^{m} S(i_k,i_j) \cdot G) + \sum_{j=1,j\neq k}^{m} (E(r_{i,j} \cdot G) - E(R_j \cdot G)) \cdot S(i_k,i_j)$;
19. Gửi cho U_i: $\{F_{1,k}, F_{2,k}, F_{3,k}, F_{4,k}\}_{0 \le k < m}$;

*** Pha 3.** U_i **thực hiện** 20. **for** $(0 \le k < m)$
21. $C_k^{(3)} = F_{1,k} - ksu_{1,1} \cdot F_{2,k}$;
22. $C_{k+m}^{(3)} = F_{3,k} - ksu_{1,1} \cdot F_{4,k}$;
23. $d^{(3)} = Dlog_{EC}(\mathbb{E}, G, \max^2, C^{(3)})$;
24. **for** $(0 \le k < m)$
25. $P_{i,k} = \frac{d_k^{(3)}}{d_{k+m}^{(3)}}$;

3.4. Giải pháp đảm bảo tính riêng tư cho Hệ tư vấn người dùng trong mô hình dữ liệu 2PFD

3.4.1. Định nghĩa bài toán

Hệ tư vấn người dùng có m mục tin, trong đó dữ liệu đánh giá về các mục tin này là sở hữu của từng cặp người dùng (u_i, v_i) thuộc hai miền dữ liệu khác nhau (dữ liệu thuộc hai tổ chức khác nhau) và có n_1 cặp người dùng tham gia đánh giá, tương ứng n người dùng trong hệ thống. Mỗi người dùng u_i sẽ đưa ra ý kiến đánh giá đối với tập con gồm m_1 mục dữ liệu ($m_1 < m$). Mỗi người dùng v_i sẽ đưa ra ý kiến đánh giá đối với tập con gồm $m - m_1$ mục dữ liệu còn lại.

3.4.2. Giải pháp đảm bảo tính riêng tư cho hệ tư vấn người dùng

Giải pháp đề xuất được trình bày cụ thể trong Giải pháp 3.2.

.

Trong đó: $n_{k1} = \left\lfloor \frac{1}{2} + \sqrt{m_1(m_1+5) + \frac{1}{4}} \right\rfloor$, $n_{k2} = \left\lfloor \frac{1}{2} + \sqrt{(m-m_1)(m-m_1+5) + \frac{1}{4}} \right\rfloor$, $n_{k3} = \left\lfloor \frac{1}{2} + \sqrt{2m_1(m-m_1) + \frac{1}{4}} \right\rfloor$

3.4.3. Ví dụ minh hoạ

Phần này trình bày một ví dụ minh hoạ dưới dạng số cụ thể đối với giao thức được đề xuất.

3.4.4. Đánh giá về tính đúng đắn

Phần này luận án chứng minh tính đúng đắn của giải pháp.

3.4.5. Đánh giá về tính riêng tư

Giải pháp đề xuất đảm bảo tính riêng tư của mỗi người dùng trong mô hình bán trung thực, hơn nữa còn bảo vệ tính riêng tư của mỗi người dùng trung thực trước sự thông đồng của máy chủ với tối đa $n - 2$ người dùng không trung thực sao cho hai người dùng trung thực không sở hữu các giá trị thuộc tính của cùng một bản ghi.

Giải pháp 3.2. *PPRS trong mô hình 2PFD*

Đầu vào:
n: Số lượng người dùng trong hệ thống, trong đó mỗi miền dữ liệu có n_1 người dùng ($n = 2n_1$).
m: Số lượng các mục tin trong cả hai miền dữ liệu.
m_1: Số lượng các mục tin thuộc miền dữ liệu đầu tiên (U_i).
RM_1: Ma trận xếp hạng của những người dùng đối với các mục tin thuộc miền dữ liệu đầu tiên, trong đó mỗi $r_{i,j}$ là giá trị xếp hạng riêng tư của người dùng U_i đối với mục tin thứ j, $j \in [0, m_1)$.
RM_2: Ma trận xếp hạng của những người dùng đối với các mục tin thuộc miền dữ liệu thứ hai (V_i), trong đó mỗi $r_{i,j}$ là giá trị xếp hạng riêng tư của người dùng V_i đối với mục tin thứ j, $j \in [0, m - m_1)$.
max: Giá trị xếp hạng lớn nhất trong thang xếp hạng.
Đầu ra (đối với U_i):
$P_{i,k}$: Điểm đề xuất được dự đoán xếp hạng cho mục tin thứ k của người dùng U_i, $k \in [0, m)$ là thứ tự mục tin mà người dùng mục tiêu đã yêu cầu sinh tư vấn.
Nếu sinh tư vấn dựa trên nội dung - CBF thì: $P_{i,k} = \frac{\sum_{j=1}^{m} r_{i,j} S(i_k, i_j)}{\sum_{j=1}^{m} S(i_k, i_j)}$
Nếu sinh tư vấn dựa trên lọc cộng tác – CF thì: $P_{i,k} = \frac{R_k \sum_{j=1}^{m} S(i_k, i_j) + \sum_{j=1}^{m} (r_{i,j} - R_j) S(i_k, i_j)}{\sum_{j=1}^{m} S(i_k, i_j)}$

Giai đoạn 1: Tính trung bình xếp hạng của các mục tin và độ tương tự giữa cặp mục tin

1. $\{R_j, \sum_{i=1}^{\frac{n}{2}} r_{i,j}^2, S(i_j, i_k)\}_{0 \le j,k < m_1; j \le k} = SMASC(\frac{n}{2}, m_1, RM_1, \max)$;

/* Thực hiện giao thức SMASC với tập $\frac{n}{2}$ bên tham gia U_i.*/

2. $\{R_j, \sum_{i=1}^{\frac{n}{2}} r_{i,j}^2, S(i_j, i_k)\}_{m_1 \le j,k < m; j \le k} = SMASC(\frac{n}{2}, m - m_1, RM_2, \max)$;

/* Thực hiện giao thức SMASC với tập $\frac{n}{2}$ bên tham gia V_i. Trong đó: Mỗi người dùng U_i có n_{k1} khóa bí mật và khóa công khai ($ksu_{i,j}, KPU_{i,j}$) và đã gửi khóa công khai cho máy chủ khi thực hiện giao thức SMASC lần đầu. Mỗi người dùng V_i có n_{k2} khóa bí mật và khóa công khai ($ksv_{i,j}, KPV_{i,j}$) và đã gửi khóa công khai cho máy chủ. */

/*Thực hiện tính các giá trị tần suất giữa các cặp mục tin thuộc hai miền dữ liệu*/

Pha khởi tạo: Rs thực hiện

3. $j = 0$;
4. for $(0 \le t < n_{k1})$
5. for $(0 \le k < n_{k2})$
6. $K_j = KU_t + KV_k$;
7. j++;
8. if $(j == n_{k3} - 1)$ break;
9. if $(j == n_{k3} - 1)$ break;

/* KU_t, KV_k lần lượt là các khóa công khai dùng chung của tập người dùng U_i, V_i đã được máy chủ tính trong quá trình thực thi các giao thức SMASC bên trên */

10. Gửi cho $\{U_i, V_i\}_{i=1}^{\frac{n}{2}}$: $\{K_j\}_{j=0}^{n_{k3}-1}$;

/*Thực hiện giao thức PPMFC có sửa đổi với tập n bên tham gia U_i, V_i*/

/*Trong đó không cần thực thi pha khởi tạo*/

11. $suv = \{\sum_{i=1}^{\frac{n}{2}} r_{i,j} r_{i,k}\}_{0 \le j < m_1 \le k < m} = PPMFC(n, RM_1, RM_2, \max)$
12. $l = 0$;
13. for $(0 \le j < m_1)$
14. for $(m_1 \le k < m)$
15. $S(i_j, i_k) = S(i_k, i_j) = \frac{suv_l}{(\sqrt{\sum_{i=1}^{\frac{n}{2}} r_{i,j}^2} \cdot \sqrt{\sum_{i=1}^{\frac{n}{2}} r_{i,k}^2})}$;
16. l++;

Giai đoạn 2: Sinh tư vấn cho U_i

// Giống với giải pháp đề xuất trong mục 3.3.2.

3.4.6. Đánh giá hiệu quả

Phần này đánh giá chi tiết về thời gian tính toán của mỗi người dùng, của máy chủ trong quá trình xây dựng hệ tư vấn người dùng. Kết quả cho thấy tính hiệu quả của giải pháp được đề xuất.

Kết luận chương

Chương này đã trình bày về hai giải pháp hệ tư vấn người dùng có đãm bảo tính riêng tư được đề xuất: Giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư cho mô hình dữ liệu phân tán đầy đủ và Giải pháp hệ tư vấn người dùng có đảm bảo tính rêng tư cho mô hình dữ liệu 2PFD. Trong các giải pháp này có sử dụng các giao thức đề xuất trong Chương 2 cho giai đoạn thứ nhất của hệ tư vấn người dùng, ở giai đoạn thứ hai và đề xuất hai phương pháp sinh tư vấn có đảm bảo tính riêng tư: sinh tư vấn dựa trên CBF và sinh tư vấn dựa trên CF với việc cải tiến giai đoạn này trong giải pháp gốc sang sử dụng hệ mật mã trên đường cong Elliptic để đồng bộ với giai đoạn thứ nhất của giải pháp đề xuất đồng thời cũng góp phần nâng cao được tính hiệu quả của giải pháp.

• Các giải pháp này không những đảm bảo tính đúng đắn mà còn đảm bảo tính riêng tư cho hồ sơ và lịch sử xếp hạng của người dùng khỏi bất kỳ bên thứ ba nào hoặc những người dùng khác.

• Kết quả thực nghiệm chứng minh rằng giải pháp đề xuất cho mô hình phân tán đầy đủ không những đáp ứng yêu cầu ứng dụng thực tế mà còn vượt trội so với các giải pháp trước đây [7, 75], giải pháp mới đề xuất cho mô hình 2PFD có hiệu suất tốt, hoàn toàn có thể áp dụng vào trong các hệ thống thực tế.

• Các giải pháp này có thể được triển khai thực tế dưới dạng mô hình máy chủ - máy trạm, đây là một mô hình phổ biến hiện nay.

Nội dung của chương này liên quan đến các công trình [CT3, CT4, CT6] trong Danh mục các công trình khoa học sử dụng trong Luận án.

KẾT LUẬN

Trong luận án này, nghiên cứu sinh đã nghiên cứu giải pháp đảm bảo tính riêng tư cho hệ tư vấn người dùng. Nghiên cứu sinh nhận thấy rằng công việc thiết kế các giải pháp như vậy không đơn giản ngay cả khi sử dụng các kỹ thuật mã hóa tiên tiến như mã hóa đồng cấu và tính toán bảo mật với mô hình dữ liệu phân tán, đặc biệt là mô hình dữ liệu phân tán đầy đủ và mô hình dữ liệu phân tán đầy đủ hai bên. Các giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư sử dụng mã hoá đồng cấu hay tính toán bảo mật có thể đảm bảo được tính đúng đắn của kết quả tính toán, vì thế đảm bảo chất lượng của hệ tư vấn tương đương với hệ tư vấn nguyên thuỷ. Bên cạnh đó các giải pháp này cũng đảm bảo tính riêng tư cho dữ liệu của người dùng ở mức độ cao. Tuy nhiên hiệu suất của các giải pháp này thường thấp do các hoạt động mã hoá. Do đó, trong luận án này, nghiên cứu sinh đã đề xuất giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư mới giúp đảm bảo được tính riêng tư dữ liệu của người dùng ở mức độ cao, đồng thời duy trì được tính đúng đắn của các tư vấn tương đương với các giải pháp điển hình hiện có và hiệu suất tốt hơn.

I. Kết quả đạt được

Kết quả chính mà Luận án được tóm lược như sau:

1. Phát triển 02 giao thức tính toán bảo mật:

- Giao thức tính đồng thời nhiều giá trị trung bình và độ tương tự có bảo mật trong mô hình dữ liệu phân tán đầy đủ bảo toàn tính đúng đắn của kết quả đầu ra tương đương với phép toán nguyên thuỷ khi chưa áp dụng biện pháp bảo mật, có hiệu quả tính toán và truyền thông tốt hơn các giao thức đã được công bố trước đây trong cùng mô hình tính toán, đồng thời việc chứng minh về mặt lý thuyết cũng đã chỉ ra mức độ an toàn tương đương với giao thức điển hình tính trung bình và độ tương tự điển hình có mức độ an toàn cao.

- Giao thức tính đồng thời nhiều giá trị tần suất có bảo mật trong mô hình 2PFD bảo toàn tính đúng đắn của kết quả đầu ra tương đương với phép toán nguyên thuỷ khi chưa áp dụng biện pháp bảo mật, có hiệu quả tính toán và truyền thông tốt hơn các giao thức đã được công bố trước đây trong cùng mô hình tính toán, đồng thời việc chứng minh về mặt lý thuyết cũng đã chỉ ra mức độ an toàn tương đương với các giao thức điển hình có mức an toàn cao.

2. Thiết kế giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư dựa trên các giao thức tính toán bảo mật đã phát triển:

- Giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư trong mô hình dữ liệu phân tán đầy đủ

- Giải pháp hệ tư vấn người dùng có đảm bảo tính riêng tư trong mô hình dữ liệu 2PFD

Các giải pháp này có thể dễ dàng được triển khai dưới dạng mô hình máy chủ - máy trạm, đây là một mô hình phổ biến hiện nay.

Đối với mỗi đề xuất:

+ Chứng minh tính đúng đắn của kết quả tư vấn được bảo toàn tương đương với giải pháp nguyên thuỷ khi chưa áp dụng biện pháp bảo mật.

+ Chứng minh được các giải pháp đề xuất đảm bảo tính riêng tư của các bên tham gia trung thực trong mô hình bán trung thực khi không có sự thông đồng giữa các bên, trong trường hợp có sự thông đồng của số lượng lớn người dùng bán trung thực thì tính riêng tư của các bên tham gia trung thực vẫn được đảm bảo. Tương đương với độ an toàn của giải pháp hiện có có độ an toàn cao.

+ Chứng minh được về mặt lý thuyết chi phí truyền thông và chi phí tính tính của các giải pháp đề xuất tốt hơn một số giải pháp điển hình hiện có.

Cuối cùng, chương trình mô phỏng hoạt động của các giao thức, giải pháp được triển khai cài đặt, xây dựng để thực nghiệm nhằm chỉ ra tính khả thi của các giao thức, giải pháp đã phát triển, thiết kế và so sánh kết quả với các công trình đã công bố trước đây và cho thấy các giao thức, giải pháp của luận án có hiệu suất tốt hơn cả về chi phí truyền thông và chi phí tính toán so với các giải pháp trong các công trình này. Vì vậy những giải pháp này hoàn toàn có thể được triển khai trong thực tế.

II. Các hướng nghiên cứu tiếp theo

Mặc dù các giao thức được phát triển có thể sử dụng cho một số ứng dụng hệ tư vấn người dùng có đảm bảo tính riêng tư. Tuy nhiên, vẫn còn một số vấn đề có thể được cải thiện trong tương lai:

1. Tiếp tục cải tiến để nâng cao mức độ an toàn của các giải pháp.

2. Nghiên cứu thêm về các hệ mật mã đồng cấu khác để có thể phát triển các giao thức SMC có khả năng kháng lượng tử, chỉ thực thi với dữ liệu có xác thực.

3. Tiếp tục nghiên cứu, cải tiến để có thể giảm bớt số lần tương tác giữa người dùng và trung tâm tính toán trong giải pháp tính tần suất được đề xuất cho mô hình 2PFD.

4. Nghiên cứu để mở rộng lĩnh vực ứng dụng thực tế của các giao thức đã được đề xuất trong Luận án.

5. Nghiên cứu để đề xuất giải pháp đảm bảo tính riêng tư cho hệ tư vấn người dùng sử dụng các kỹ thuật tư vấn phức tạp hơn khác.

6. Dựa trên các yêu cầu của các vấn đề thực tế hoặc các mô hình dữ liệu phân tán mới hoặc các mức độ an toàn khác nhau, các giao thức tính toán bảo mật mới cần được phát triển.