Đề xuất phương pháp bảo vệ mẫu có độ dài cố định sử dụng hệ mã mật khóa công khai Paillier

Nguyễn Thị Hồng Hà[1], Trương Phi Hồ[2*], Đặng Đức Trịnh[3], Lê Mạnh Hùng[1], Phạm Duy Trung[1]

[1]Khoa An toàn Thông tin, Học viện Kỹ thuật Mật mã, Thanh Trì, Hà Nội.
[2]Khoa CNTT-TCM, Trường Đại học Thông tin liên lạc, Nha Trang, Khánh Hòa;
[3]Khoa Toán-Tin, Học viện Quân y, Hà Đông, Hà Nội.
*Email: phihosqtt@gmail.com
Nhận bài: 12/5/2022; Hoàn thiện: 14/7/2022; Chấp nhận đăng: 10/8/2022; Xuất bản: 26/8/2022.
DOI: https://doi.org/10.54939/1859-1043.j.mst.81.2022.148-155

TÓM TẮT

Tính riêng tư về dữ liệu của khách hàng hiện nay đang rất được quan tâm, có rất nhiều vụ bê bối lợi dụng thông tin khách hàng để vụ lợi gây ra bởi nhiều công ty công nghệ khác nhau, có sở hữu cơ sở dữ liệu lớn thông tin về khách hàng. Vấn đề đặt ra là làm sao khi người dùng có thể cung cấp dữ liệu đáp ứng được thuật toán được sử dụng trong mô hình xác thực, cho ra kết quả chính xác mà vẫn không làm lộ thông tin riêng tư hoặc dữ liệu cá nhân. Nghiên cứu lý thuyết về hệ mật mã đồng cấu và mã hóa công khai có thể giải quyết những vấn đề về bảo mật tính riêng tư người dùng. Nội dung bài báo trình bày những vấn đề cơ bản về: phép đồng cấu và hệ mật mã công khai Paillier; tính khoảng cách Euclid đối với mẫu có độ dài cố định; kết quả thực nghiệm mã hóa mẫu bằng mã hóa Paillier sau đó kiểm tra tính đúng đắn bằng cách tính toán, so sánh độ dài với ngưỡng δ xác định để chọn loại bỏ mẫu không thích hợp và quyết định sự kết hợp của hệ thống. Từ đó, đề xuất phương án bảo vệ mẫu và bảo đảm tính riêng tư của người dùng đối với mẫu có độ dài cố định.

Từ khoá: Sự riêng tư; Mã hóa; Đồng cấu; Khoảng cách Euclid; Mã hóa Paillier; Vân tay.

1. MỞ ĐẦU

Quyền riêng tư của người dùng trên mạng hay trong một hệ thống xác thực tùy thuộc vào khả năng kiểm soát cả lượng thông tin cá nhân mà người dùng cung cấp cho đơn vị quản lý dữ liệu, và những người có quyền truy nhập thông tin đó. Vấn đề được đặt ra như sau: một ngân hàng (công ty) sở hữu một thuật toán với mục đích xác thực các khách hàng khi đăng nhập tài khoản, xác thực thông tin khi giao dịch. Thuật toán tại ngân hàng này chỉ sử dụng một phép tính đơn giản như sau:

$$result = X_1 + X_2 - X_3$$

Với X_1, X_2, X_3 là dữ liệu khách hàng cung cấp để xác thực là các số nguyên ví dụ như khoảng cách Euclid trong nhận dạng dấu vân tay. Đây là những dữ liệu quan trọng trong các giao dịch quan trọng liên quan đến lĩnh vực ngân hàng và một số ngành nghề khác. Khách hàng C (viết tắt Customer) rất muốn sử dụng dịch vụ của ngân hàng, có thể đưa dữ liệu thông qua hình ảnh hoặc lấy mẫu trực tiếp từ thiết bị của ngân hàng. Tuy nhiên, đối với những dữ liệu này đối với C có tính chất khá riêng tư, C không muốn công khai với người khác, cụ thể ở đây là công khai với nhân viên ngân hàng.

Để giải quyết vấn đề này, chúng ta xem xét một giải pháp dành cho C. Đầu tiên C tìm cách mã mật dữ liệu bằng mật mã khóa công khai dữ liệu X_1, X_2, X_3 để ngân hàng không có được những hình ảnh cụ thể của mẫu do C cung cấp, nhưng ngân hàng vẫn có thể làm việc được trên dữ liệu đã được mã hóa nhưng vẫn xác nhận đúng là C khi thực hiện các giao dịch thanh toán. Đây là ý tưởng cơ bản của một khái niệm trong lý thuyết mật mã hiện đại: hệ mật mã đồng cấu. Giả sử C sẽ mã hóa và giải mã dữ liệu thỏa công thức sau:

$$Decrypt(Encypt(x) + Encrypt(y)) = x + y$$

Giờ C sẽ gửi những dữ liệu này cho ngân hàng. Ngân hàng vẫn áp dụng các phép tính giống hết như tác động lên dữ liệu gốc:

$$Y_{result} = Encrypt(X_1) + Encrypt(X_2) - Encrypt(X_3) = X_1 + X_2 - X_3$$

Từ giả thuyết nêu trên, sử dụng thuật toán giải mã qua hàm Derypt: $Derypt(Y_{result}) = X_1 + X_2 - X_3$. Từ đó thấy rằng, đây cũng chính là kết quả mà ngân hàng nhận được nếu như thực hiện trên mẫu gốc từ C. Điều này có nghĩa là C sẽ không bị mất sự riêng tư của mình. Từ đó, có thể sử dụng phép toán cũng như hệ thống xác thực của ngân hàng. Ở đây, chúng ta nhận ra trong thuật toán trên hệ thống của ngân hàng đã sử dụng có tính đồng cấu. Nó có *cộng tính*, nghĩa là ta có thể làm việc với phép cộng trên dữ liệu mã mật bằng với đầu ra của hàm mật mã trong bài toán – và sau khi giải mã mật ta sẽ nhận được kết quả là giá trị khi thực hiện phép cộng đó trên bản rõ tương ứng. Đồng nghĩa với một hệ thống xác thực ta hoàn toàn sử dụng một hệ mã hóa có tính cộng (hoặc nhân) được trình bày trong 2.1 để mã hóa dữ liệu, tránh lộ lọt thông tin về các giá trị đầu vào của một hệ thống, bảo đảm tính riêng tư của người dùng.

Trong bài báo này, tác giả đề xuất hệ mã hóa khóa công khai Pailier và thực nghiệm tính toán khoảng cách Euclid đã được chứng minh [1] nhằm xác định tính khả thi trong việc bảo vệ mẫu, bảo đảm tính riêng tư của người dùng khi sử dụng mã hóa công khai Paillier.

Bài báo được trình bày gồm 3 phần chính: Mục 2 trình bày lý thuyết cơ bản về phép đồng cấu trong toán học và mã hóa công khai Paillier; tiếp theo mục 3 trình bày về cách tính tổng sử dụng khoảng cách Euclid và phương thức sử dụng mã hóa Paillier trong bài báo; sau đó là phần thực nghiệm để so sánh kết quả giữa 2 mẫu được bảo vệ và không được bảo vệ. Kết thúc bài báo đưa ra kết luận và đề xuất phương án bảo vệ mẫu cũng như quyền riêng tư người dùng.

2. LÝ THUYẾT NGHIÊN CỨU

Mã hoá đồng cấu có một số khả năng ứng dụng tức thời, chẳng hạn như bầu cử điện tử. Hệ mã hoá Paillier đồng cấu với phép cộng nên đã được dùng trong nhiều giao thức bầu cử điện tử. Mỗi người bỏ phiếu mã hoá phiếu bầu của mình như một con số và công bố nó với thế giới. Bất kỳ ai cũng có thể cộng các phiếu bầu để tạo nên kết quả cuối cùng (được mã hoá) khiến kẻ xấu khó có thể bỏ qua những phiếu bầu hợp lệ. Giải mã bản mã hoá kết quả cuối cùng sẽ chỉ cho biết tổng số phiếu bầu cho mỗi ứng viên nhưng không lộ phiếu bầu của từng cử tri. Khả năng ứng dụng của mã hoá đồng cấu còn rất lớn, nhất là với sự phổ biến của điện toán đám mây. Ngoài ra, mật mã đồng cấu có thể ứng dụng trong bỏ phiếu, y tế, tài chính [3, 7, 8].

Trong báo cáo này, nhóm tác giả sẽ trình bày rõ hơn về lý thuyết đồng cấu và hệ mã công khai Paillier ở phần 2.1.

2.1. Phép đồng cấu và mã hóa công khai Paillier

2.1.1. Phép đồng cấu

Để hạn chế những rủi ro về dữ liệu mang tính riêng tư của cá nhân, cần thực hiện nhiều giải pháp phù hợp, đầy đủ để xây dựng sự tin cậy của người dùng thông qua việc bảo vệ dữ liệu cá nhân bằng phép toán đồng cấu. Thuật ngữ "đồng cấu" xuất hiện sớm nhất từ năm 1892, bởi nhà toán học người Đức Felix Klein. Trong đại số, phép đồng cấu là một ánh xạ bảo toàn cấu trúc giữa hai cấu trúc đại số cùng loại (chẳng hạn như hai nhóm, hai vành, hoặc hai không gian vectơ). Phép đồng cấu của không gian vectơ còn được gọi là ánh xạ tuyến tính, và việc nghiên cứu của chúng là đối tượng trong môn học đại số tuyến tính.

Phép đồng cấu là một ánh xạ giữa hai cấu trúc đại số cùng loại, bảo toàn các phép toán của cấu trúc. Điều này có nghĩa là một ánh xạ $f: A \to B$ giữa hai tập A, B được trang bị cùng một cấu trúc thoả mãn, nếu là một phép toán của cấu trúc (để đơn giản hóa, ta giả sử nó là một phép toán hai ngôi), khi đó:

$f(x.y) = f(x).f(y)$ cho mọi cặp x, y trong các phần tử của A. Ta thường nói rằng f bảo toàn phép toán hoặc tương thích với phép toán.

Về mặt hình thức, một ánh xạ $f: A \to B$ phép toán μ của k ngôi được xác định trên cả hai A và B nếu: $f(\mu_A(a_1, \dots, a_k)) = \mu_B(f(a_1), \dots, f(a_k))$, với mọi $a_1, \dots, a_k$ trong A.

Các phép toán được bảo toàn bởi phép đồng cấu bao gồm các phép toán 0-ary (hoặc phép toán nullary), là các hằng số. Đặc biệt, khi cấu trúc yêu cầu phải bao gồm một phần tử đơn vị, phần tử đơn vị của cấu trúc đầu tiên phải được ánh xạ tới phần tử đơn vị tương ứng của cấu trúc thứ hai.

Chính vì tính chất đặc biệt của phép toán này nên thích hợp trong việc bảo đảm tính toàn vẹn, riêng tư của mẫu được dùng trong xác thực.

2.1.2. Hệ mã hóa công khai Paillier

Hệ thống mật mã Paillier, được phát minh bởi Pascal Paillier vào năm 1999 [2], là một thuật toán bất đối xứng xác suất cho mật mã khóa công khai. Các bước tạo khóa, mã hóa, giải mã mật mã công khai Pailler theo [5], được trình bày khái quát ở phần 2.2.1.

Một tính năng đáng chú ý của hệ thống mật mã Paillier là các thuộc tính đồng cấu của nó cùng với mã hóa không tất định của nó. Vì hàm mã hóa có tính đồng cấu cộng tính, các đặc tính như cộng tính hoặc nhân tính có thể được mô tả như sau:

Phép cộng đồng cấu của các bản rõ

Tích của hai bản mã sẽ giải mã thành tổng các bản rõ tương ứng của chúng,

$$D(E(m_1, r_1)E(m_2, r_2) \bmod n^2) = m_1 + m_2 \bmod n$$

Sản phẩm của một bản mã với một bản rõ nâng lên g sẽ giải mã thành tổng của các bản rõ tương ứng theo công thức sau:

$$D(E(m_1, r_1).g^{m_2} \bmod n^2) = m_1 + m_2 \bmod n$$

Phép nhân đồng cấu của các bản rõ

Một bản mã được nâng lên thành lũy thừa của một bản rõ sẽ giải mã thành tích của hai bản rõ

$$D(E(m_1, r_1)^{m_2} \bmod n^2) = m_1 m_2 \bmod n$$
$$D(E(m_2, r_2)^{m_1} \bmod n^2) = m_1 m_2 \bmod n$$

Nói một cách tổng quát hơn, một bản mã được nâng lên một hằng số k sẽ giải mã thành tích của bản rõ và hằng số:

$$D(E(m_1, r_1)^k \bmod n^2) = k m_1 \bmod n$$

Những phép toán vừa nêu sẽ được trình bày cụ thể hơn trong phần 3 đối với tính cộng. Chú ý rằng, với việc mã hóa Paillier của hai thông điệp, không có cách nào được biết để tính toán tích mã hóa của những thông điệp này mà không cần biết khóa cá nhân. Do đó, khóa cá nhân (private key) là cần thiết và phải được tính toán nếu sử dụng hệ mã hóa công khai Paillier.

2.2. Phương pháp tính toán điểm tương đồng sử dụng khoảng cách Euclid đối với các mẫu có độ dài cố định

Cách tính khoảng cách Euclid được trình bày trong [1, 6]. Trong bài báo này, tác giả giới thiệu cách tính khoảng cách Euclid để thực nghiệm theo cụ thể như sau:

2.2.1. Quy ước các ký hiệu

$\mathbf{T}_p = \{p_1, ..., p_f, ..., p_F\}$ và $\mathbf{T}_r = \{r_1, ..., r_f, ..., r_F\}$: $\mathbf{T}_p$ là mẫu thu được từ cảm biến của hệ thống, $\mathbf{T}_r$ là mẫu tham chiếu là mẫu không được bảo vệ (chưa mã hóa), bao gồm các F đặc trưng (T_r và T_p có F thành phần).

$S_{dist} = d_{dist}$ ($\mathbf{T}_p$, $\mathbf{T}_r$): Điểm giống nhau giữa hai mẫu $\mathbf{T}_p$ và $\mathbf{T}_r$, trong đó, d_{dist} là một hàm khoảng cách cụ thể, ví dụ như khoảng cách Euclid (viết tắt là euc).

$(\mathbf{T}_r)_{dist}$, $(\mathbf{T}_p)_{dist}$: Khoảng cách Euclid tính được của $\mathbf{T}_r$ và $\mathbf{T}_p$.

Cách sinh khóa, mã hóa và giải mã hệ mã Paillier theo https://en.wikipedia.org/wiki/Paillier_cryptosystem; tác giả chỉ giới thiệu sơ lượt và quy ước một số ký hiệu được sử dụng trong bài báo này như sau:

m và m^* : bản rõ và bản mã tương ứng.

$m^* = E_{pk}(m, s)$, trong đó E biểu thị hàm mã hóa, s là số ngẫu nhiên và pk khóa công khai, và E được định nghĩa bằng (1):

$$E_{pk}(m,s) = g^m.s^n \bmod n^2 \quad (1)$$

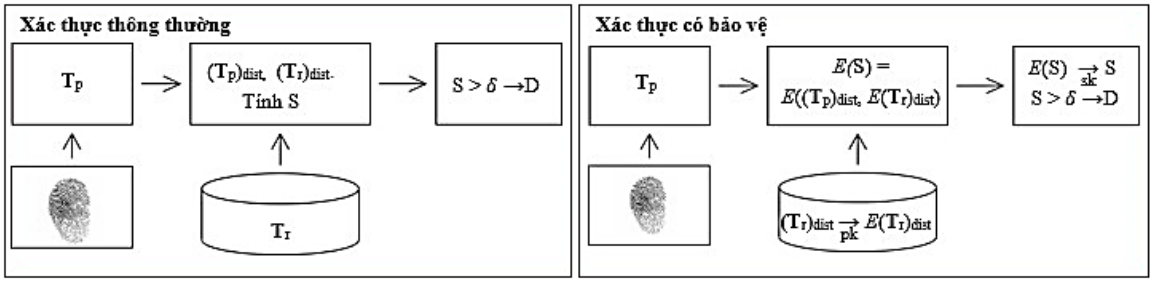
với $n = p \cdot q$ với p và q là hai số nguyên tố lớn sao cho $gcd\ (pq, (p - 1)\ (q - 1)) = 1$ và $g \in Z^*_{n^2}$ và m là thông điệp $\in Z_n$. Để tránh ký hiệu quá phức tạp, các giá trị được mã hóa $E_{pk}(m,s)$ sẽ được ký hiệu đơn giản là $E(m)$, mặc dù số ngẫu nhiên s và khóa công khai pk là cần thiết cho tính toán mã hóa.

$m = D_{sk}(m^*)$, trong đó D là hàm giải mã và sk khóa riêng tư. Mặt khác, khóa riêng tư được định nghĩa là $sk = (\lambda, \mu)$, trong đó $\lambda = lcm(p - 1, q - 1)$ và $\mu = L(g^\lambda \bmod n^2)^{-1} \bmod n$. Với $L(x)$ là hàm được định nghĩa $L(x) = (x\text{-}1)/n$ tính được D_{sk} theo (2):

$$D_{sk}(m^*) = \mathrm{L}((m^*)^\lambda) \bmod n^2).\mu \bmod n \quad (2)$$

$E(\mathbf{T}_r)_{dist}$: mẫu tham chiếu được mã hóa được định nghĩa trong (7), mẫu mã hóa $E(\mathbf{T}_r)_{dist}$ là khác nhau cho mỗi khoảng cách tính được. Lưu ý rằng, $E(\mathbf{T}_r) \neq \{E(r_1), ..., E(r_F)\}$. Ta thực hiện $E(\mathbf{T}_r)_{dist}$ cho mỗi thước đo khoảng cách, để kết quả có thể được tính trực tiếp trong miền được mã hóa theo hình 1(b).

$E(S_{dist})$: khoảng cách trung bình tương tự được mã hóa, được tính giữa $\mathbf{T}_p$ và $E(\mathbf{T}_r)_{dist}$ như được định nghĩa trong (6). Hàm $E(S_{dist})$ nhận đầu vào là $\mathbf{T}_p$ và $E(\mathbf{T}_r)_{dist}$, và xuất trực tiếp điểm số được mã hóa mà không có sự giải mã liên quan đến quá trình như hình 1(b).



(a) *(b)*

Hình 1. *Xác thực sinh trắc học (a): Mẫu không được bảo vệ (mã hóa); (b): Mẫu được bảo vệ.*

Trong trường hợp không được bảo vệ (a), một mẫu sinh trắc học được thu thập và các đặc trưng của mẫu được trích xuất ($\mathbf{T}_p$). Điểm tương tự đối với tham chiếu cảm biến, $\mathbf{T}_r$, được tính (S). Sau đó, đầu ra cuối cùng là quyết định kết hợp/không kết hợp, $D = (S > \delta)$. Trong tình huống mẫu được bảo vệ (b), tất cả dữ liệu hoặc luồng thông tin được mã hóa được mô tả bằng màu đỏ: $E(\mathbf{T}_r)$ và $E(S)$.

2.2.2. Khoảng cách Euclid được mã hóa

Cho trước hai F-mẫu $\mathbf{T}_p$ và $\mathbf{T}_r$ chưa được mã hóa, điểm số $S_{euc} = d^2_{euc}(\mathbf{T}_p, \mathbf{T}_r)$, có thể được tính toán một cách như (hình 1(a)):

$$S_{euc} = \sum_{f=1}^{F} p_f^2 + r_f^2 - 2p_f r_f \quad (3)$$

Sau đó, sử dụng:

$$D_{sk}(m_1^*.m_2^* \bmod n^2) = m_1 + m_2 \bmod n \quad (4)$$

Với: $m_1^* = E_{pk}(m_1, s_1)$ và $m_2^* = E_{pk}(m_2, s_2)$

$$D_{sk}\left((m_1^*)^l \bmod n^2\right) = m_1.l \bmod n \quad (5)$$

Điểm được mã hóa có thể được tính trực tiếp trong miền được mã hóa mà không cần thực hiện bất kỳ mã hóa nào trong máy khách (hình 1(b)) như (6), theo [9]:

$$E(S_{euc}) = \prod_{f=1}^{F} E(1)^{p_f^2}.E\left(r_f^2\right).E(rf)^{-2pf} = \prod_{f-1}^{F}(1^*)^{p_f^2}.euc_{1f}^*.(euc_{2f}^*)^{-2pf} \quad (6)$$

Do đó, mẫu tham chiếu được lưu trữ trong cơ sở dữ liệu được mã hóa được xác định bởi các mật mã sau, theo quy ước 1* là bản rõ mã có thông điệp là 1, theo [9]:

$$E(\mathbf{T}_r)_{euc} = \{1^*\} \cup \{euc1_f^*, euc2_f^*\}_{f=1}^{F} \quad (7)$$

Với $euc1_f^*$ và $euc2_f^* = E(r_f)$. Do đó, tất cả các bản mã có liên quan đến phương trình (6) được gửi bởi máy chủ và các sản phẩm và lũy thừa được tính trực tiếp trên máy khách.

Với tính chất đồng cấu của hệ thống mật mã Paillier, $E(1)$ có thể được tính toán và lưu trữ riêng biệt cho từng đối tượng tại thời điểm đăng ký, dẫn đến các giá trị được mã hóa khác nhau và do đó tăng tính bảo mật và quyền riêng tư của người dùng.

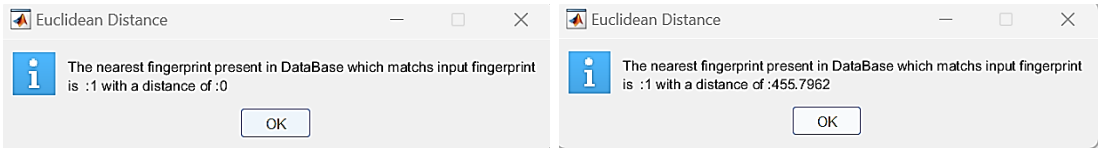
3. TÍNH TOÁN KHOẢNG CÁCH EUCLID KẾT HỢP MÃ HÓA PAILLIER, ĐÁNH GIÁ KẾT QUẢ

3.1. Số liệu đầu vào

Bài báo sử dụng cơ sở dữ liệu trong cuộc thi xác minh dấu vân tay năm 2002 có dấu vân tay tham chiếu duy nhất và bảy dấu vân tay khác có sai lệch về vị trí, mỗi tập con như vậy chứa 80 hình ảnh dấu vân tay. Trong báo cáo này, nhóm tác giả sử dụng bộ dữ liệu FCV2002DB1 được rút gọn gồm 48 mẫu chia thành 6 lớp mẫu vân tay khác nhau: mỗi lớp mẫu gồm 8 mẫu là hình ảnh cùng 1 vân tay nhưng được lấy mẫu có sự chênh lệch về vị trí. Sử dụng mẫu đầu tiên trong lớp là có ký hiệu 10*x*_1 (*x*: tên lớp mẫu với điều kiện $1 \le x \le 8$) làm mẫu tham chiếu $\mathbf{T}_r$.

3.2. Phương pháp, công cụ mô phỏng

Để đánh giá toàn diện hiệu quả của phương pháp mã hóa mẫu bằng mã hóa công khai Paillier và cũng như độ chính xác của phương pháp này đã được chứng minh, tác giả sử dụng bộ dữ liệu mẫu vân tay FVC2002DB1. Do trong báo cáo này tác giả chỉ chứng minh tính đúng đắn của phương pháp mã hóa được đề xuất nên chỉ dùng số mẫu rút gọn cho ra kết quả tính toán được trong bảng 2. Trong các thử nghiệm của mình, tác giả đã sử dụng ngôn ngữ lập trình Matlab để thực nghiệm và tính toán kết quả được thống kê như bảng 1 và bảng 2. Hình ảnh tính toán khoảng cách Euclid như hình 2.



(a) *(b)*

Hình 2. *Khoảng cách Euclid (a): Mẫu tham chiếu $\mathbf{T}_p$; (b): Mẫu từ cảm biến $\mathbf{T}_r$.*

3.3. Kết quả mô phỏng và bình luận

Bảng 1 thể hiện khoảng cách Euclid của $\mathbf{T}_p$ so sánh với mẫu đầu tiên trong lớp được chọn là $\mathbf{T}_r$. Giá trị trung bình là trung bình cộng của tất cả $(\mathbf{T}_p)_{dist}$. Quá trình mã hóa Paiilier tác giả chọn 2 số p = 47, q = 53 phù hợp với lý thuyết [5]. Để giới hạn sự lựa chọn trong 1 khoảng nhất định, tác giả chọn 2 số *g, r* là 2 số nguyên ngẫu nhiên thỏa điều kiện $0 \le g, r \le 150$ phù hợp với lý thuyết [5]. Tính toán mã hóa ra được giá trị $S=D_{sk}(E(S_{dist}))$ theo bảng 2. Thuật toán viết bằng

ngôn ngữ lập trình Matlab tính toán kết quả chính xác được thể hiện tại hình 4. Do hệ mã hóa Paiiler chỉ hoạt động được với thông điệp là số nguyên, vì vậy, đối với khoảng cách Euclid tính được tác giả đã làm tròn giá trị.

Bảng 1. *Tính toán khoảng cách của Euclid và giá trị trung bình của nó đối với mẫu chưa được bảo vệ.*

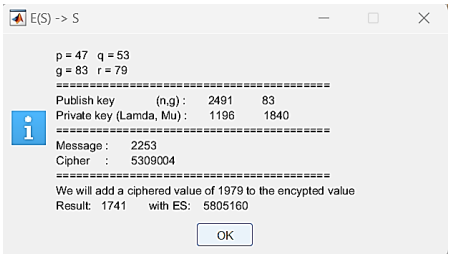
T_r	Tên mẫu	$(T_p)_{dist}$	Giá trị trung bình	T_r	Tên mẫu	$(T_p)_{dist}$	Giá trị trung bình
1	101_1	#	**1282.177**	3	104_1	2253.1952	**1862.716**
1	101_2	455.7962		4	104_2	2024.1596	
1	101_3	754.2149		4	104_3	1534.0680	
4	101_4	1994.914		4	104_4	1895.2493	
3	101_5	1494.0208		4	104_5	1591.589	
1	101_6	1737.6361		4	104_6	1952.2484	
1	101_7	1975.2713		4	104_7	1672.1864	
3	101_8	1845.5593		4	104_8	1979.0343	
1	102_1	2090.9072	**1361.316**	2	105_1	1699.2128	**1707.226**
2	102_2	529.9636		4	105_2	1902.3023	
2	102_3	1159.2189		4	105_3	1911.2314	
2	102_4	737.5271		5	105_4	1623.4203	
4	102_5	1591.9504		4	105_5	1656.3848	
4	102_6	2013.0897		6	105_6	1575.8337	
2	102_7	1076.3802		4	105_7	1811.5125	
4	102_8	1691.4902		4	105_8	1482.8015	
1	103_1	2909.2993	**1654.592**	3	106_1	1611.2106	**1748.947**
3	103_2	1153.5412		6	106_2	1281.1058	
4	103_3	1895.0744		6	106_3	1377.2865	
3	103_4	1205.6649		4	106_4	1893.1081	
4	103_5	1645.2424		4	106_5	2109.4199	
4	103_6	1610.0485		4	106_6	1956.7577	
3	103_7	1293.7116		6	106_7	2030.2207	
3	103_8	1524.1532		4	106_8	1732.4679	

Bảng 2. *Tính toán khoảng cách S_{dist} đối với mẫu được mã hóa kết hợp với mẫu tham chiếu từ đó tính ra giá trị δ.*

T_r	Tên mẫu	$(T_p)_{dist}$	S = $D_{sk}(E(S))$	T_r	Tên mẫu	$(T_p)_{dist}$	S = $D_{sk}(E(S))$
1	101_1	#	**1427**	3	104_1	**2253**	**1806**
1	101_2	455	455	4	104_2	2024	1786
1	101_3	754	754	4	104_3	1534	1296
4	101_4	1994	1994	4	104_4	1895	1657
3	101_5	1494	1494	4	104_5	1591	1353
1	101_6	1737	1737	4	104_6	1952	1714
1	101_7	1713	1713	4	104_7	1672	1434
3	101_8	1845	1845	4	104_8	1979	1741

1	102_1	**2090**	**855**	2	105_1	**1699**	**916**
2	102_2	529	128	4	105_2	1902	1110
2	102_3	1159	758	4	105_3	1911	1119
2	102_4	737	336	5	105_4	1623	831
4	102_5	1591	1190	4	105_5	1656	864
4	102_6	2013	1612	6	105_6	1575	783
2	102_7	1076	675	4	105_7	1811	1019
4	102_8	1691	1290	4	105_8	1482	690
1	103_1	**2909**	**1893**	3	106_1	**1611**	**888**
3	103_2	1153	1571	6	106_2	1281	401
4	103_3	1895	2313	6	106_3	1377	497
3	103_4	1205	1623	4	106_4	1893	1013
4	103_5	1645	2063	4	106_5	2109	1229
4	103_6	1610	2028	4	106_6	1956	1076
3	103_7	1293	1711	6	106_7	2030	1150
3	103_8	1524	1942	4	106_8	1732	852

Qua kết quả bảng 1 và bảng 2 ta thấy kết quả của các phép tính khoảng cách Euclid đối với mẫu được bảo vệ và mẫu không được bảo vệ thông qua mô tả ở hình 1 (a) và (b). Sau khi tính được S so sánh với δ từ đó đưa ra quyết loại mẫu không phù hợp hoặc kết hợp hay không. Từ kết quả cho thấy các mẫu có S > δ là những mẫu lựa chọn để kết hợp và có thể đưa thêm vào cơ sở dữ liệu hoặc kết luận đây là những mẫu không chính xác.



Hình 3. *Ảnh chụp nhanh phần mềm đã phát triển tính toán mã hóa giá trị* $\mathbf{T}_r$ *và tính S= D(E(S)).*

4. KẾT LUẬN

Thuật toán áp dụng đối với khoảng cách Euclid sử dụng hệ mã hóa Paiiler để mã hóa giá trị khoảng cách Euclid của mẫu nhằm bảo mật về giá trị của mẫu tham chiếu, bảo đảm sự riêng tư của người dùng. Kết quả tính toán khoảng cách Euclid giữa $\mathbf{T}_p$ và $\mathbf{T}_r$ từ đó tính giá trị trung bình để cho ra quyết định lựa chọn ngưỡng δ phù hợp. Có thể loại bỏ những mẫu có S > δ. Đây là những mẫu hệ thống không thể xác thực chính xác so với mẫu tham chiếu. Bảng 2 cho thấy tỉ lệ chính xác thuật toán do tác giả đề xuất còn phụ thuộc vào việc xác định ngưỡng δ. Với δ được chọn bằng giá trị trung bình của S_{dist} kết quả đạt được xấp xỉ 76%. Chọn $\delta = \min(S_1, ...,S_8)$, với $(S_1, ...,S_8)$ là những mẫu cho kết quả nhận dạng sai để loại bỏ hết các mẫu sai thì kết quả chính xác 100%.

Qua thực nghiệm nhóm tác giả cũng thấy rằng, phương pháp bảo vệ mẫu có độ dài cố định sử dụng mã hóa công khai Paillier vì hệ mã hóa Paillier có tính đồng cấu, thích hợp trong việc bảo vệ và mã hóa đảm bảo sự riêng tư của khách hàng. Cần lưu ý rằng không sử dụng phương pháp đối với mẫu có độ dài thay đổi như: chữ ký số, chữ viết tay,... do tính chất của mẫu thay đổi do người dùng. Hướng phát triển của bài báo có thể thực nghiệm và so sánh với các hệ mã khác nhau cùng có tính đồng cấu và dựa trên các khoảng cách khác như Cosine, Mahalanobis,...

TÀI LIỆU THAM KHẢO

[1]. Jadhav, Barbadekar, Patil, "*Euclidean Distance Based Fingerprint Matching,"* Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics, ISBN: 978-960-474-276-9.

[2]. Jonathan Katz, Yehuda Lindell, "*Introduction to Modern Cryptography: Principles and Protocols*," Chapman & Hall/CRC, (2007).

[3]. Khuất Thanh Sơn. "*Mã hóa đồng cấu đầy đủ và ứng dụng trong theo dõi sức khỏe an toàn dựa trên điện toán đám mây*". Luận văn Thạc sĩ, Đại học công nghệ - Đại học quốc gia Hà Nội, (2021).

[4]. Marta G´omez Barrero, "*Improving Security and Privacy in Biometric systems*," Ingeniero de Inform´atica y Licenciada en Matem´aticas Universidad Aut´onoma de Madrid, SPAIN, (2016).

[5]. Mark A. Will, Ryan K.L. Ko, "*A guide to homomorphic encryption*," in The Cloud Security Ecosystem, (2015).

[6]. Neeraj Bhargava, Anchal Kumawat, Ritu Bhargava, "*Fingerprint Matching of Normalized Image based on Euclidean Distance*," International Journal of Computer Applications (0975 – 8887)Volume 120 – No.24, (2015).

[7]. Hsin-Tsung Peng, William W.Y. Hsu, "*Homomorphic Encryption Application on FinancialCloud Framework*", The Ministry of Science and Technology of Taiwan.

[8]. Leon J. Helsloot; Gamze Tillem, Zekeriya Erkin, "*Privacy-preserving online behavioral advertising using homomorphic encryption*," IEEE Workshop on Information Forensics and Security, (2017).

[9]. Marta G´omez Barrero, *"Improving Security and Privacy in Biometric systems,"* Ingeniero de Inform´atica y Licenciada en Matem´aticas Universidad Aut´onoma de Madrid, Spain, (2016).

ABSTRACT

Proposed method of protection of fixed length samples through using Paillier cryptosystem

Privacy of databases is currently a very concern, there are many scandals of using customer information for personal gain caused by many different technology companies that own large databases of information about customers. The problem is how the user can provide a database that meets the algorithm used in the authentication model, giving accurate results without revealing private information or personal data. Theoretical research on homomorphic cryptosystems and public encryption can solve problems of user privacy security. The paper presents the basic issues of the homomorphic algorithm and the Paillier cryptosystem; computational similarity using the Euclidean distances for samples of fixed length. Sample experimental results coded by Paillier cryptosystem then check the correct calculation by calculating, comparing the length with a defined threshold to remove the inappropriate samples and decide the combination of the system. From there, propose a method of protection of fixed-length samples.

Keywords: Privacy; Cryptosystem; Homomorphic; Euclidean distances; Paillier cryptosystem; Fingerprint.