PROPOSE EUCLIDEAN DISTANCES CODECTION METHOD TO GUARANTEE USER PRIVACY IN FINGERPRINT RECOGNITION

Truong Phi Ho[1*], Nguyen Thi Hong Ha[2], Dang Xuan Bao[2], Pham Duy Trung[2]
[1]Telecommunications University, [2]Academy of Cryptography Techniques

ARTICLE INFO

Received: 17/5/2022
Revised: 24/6/2022
Published: 24/6/2022

KEYWORDS

Privacy
Cryptosystem
Homomorphic
Euclidean distances
Paillier cryptosystem
Fingerprint

ABSTRACT

As technology develops, password authentication contains many security risks and is outdated. In the current authentication methods, the fingerprint authentication method is based on the identification of features, thereby comparing and identifying an individual. One of the methods for fingerprint recognition is using Euclidean distance. The problem is how when the user can provide a database that meets the algorithm used in the authentication model, giving accurate results without revealing information about the Euclidean distance when system reference. The author's solution uses the Paillier public cryptosystem system; encrypt the Euclidean distance for the authentication sample before entering the system for comparison; Experimental results encode the sample by Paillier coding, then check the correctness by calculating, comparing the results with the usual calculation. From there, the conclusion of the proposed algorithm is made.

ĐỀ XUẤT PHƯƠNG PHÁP MÃ HÓA KHOẢNG CÁCH EUCLID BẢO ĐẢM TÍNH RIÊNG TƯ CỦA NGƯỜI DÙNG TRONG NHẬN DẠNG VÂN TAY

Trương Phi Hồ[1*], Nguyễn Thị Hồng Hà[2], Đặng Xuân Bảo[2], Phạm Duy Trung[2]
[1]Trường Đại học Thông tin Liên lạc, [2]Học viện Kỹ thuật Mật mã

THÔNG TIN BÀI BÁO

Ngày nhận bài: 17/5/2022
Ngày hoàn thiện: 24/6/2022
Ngày đăng: 24/6/2022

TỪ KHÓA

Riêng tư
Mã hóa
Đồng cấu
Khoảng cách Euclid
Mã hóa Paillier
Vân tay

TÓM TẮT

Công nghệ ngày càng phát triển, việc xác thực bằng mật khẩu ẩn chứa nhiều rủi ro bảo mật và đã lỗi thời. Trong các phương thức xác thực hiện nay, phương thức xác thực bằng dấu vân tay dựa vào việc nhận dạng các đặt trưng, từ đó đối chiếu và xác định danh tính một cá nhân. Một trong những phương pháp để nhận dạng vân tay là sử dụng khoảng cách Euclid. Vấn đề đặt ra là làm sao khi người dùng có thể cung cấp cơ sở 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 về khoảng cách Euclid khi tham chiếu với hệ thống. Giải pháp của nhóm tác giả sử dụng hệ mật mã công khai Paillier; mã hóa khoảng cách Euclid đối với mẫu xác thực trước khi đưa vào hệ thống để so sánh đối chiếu; kết quả thực nghiệm mã hóa mẫu bằng mã hóa Paillier sau đó kiểm tra tính đúng đắn bằng cách tính toán, so sánh kết quả với cách tính thông thường. Từ đó đưa ra kết luận đối với phương pháp được đề xuất.

DOI: https://doi.org/10.34238/tnu-jst.6000

* Corresponding author. *Email: phihosqtt@gmail.com*

1. Giới thiệu

Mã hoá đồng cấu có một số khả năng ứng dụng tức thời, chẳng hạn như bầu cử điện tử. Với E là hàm mã hóa, một hàm mã hoá (E) mà cả E(x + y) và E(x.y) đều có thể tính được dễ dàng từ E(x) và E(y) theo [1]. 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 được phát triển bởi [2].

Khả năng ứng dụng của mã hoá đồng cấu còn rất lớn, nhất là với sự phổ biến của điện toán đám mây. Chúng ta sẽ xem xét một khả năng ứng dụng mã hoá đồng cấu trong một số ngành:

Ngành y tế: Trong một hệ lưu trữ đảm bảo bí mật trên đám mây, dữ liệu về lịch sử khám chữa bệnh của các bệnh nhân được các cơ sở y tế mã hoá trước khi đẩy lên đám mây. Bệnh nhân kiểm soát việc chia sẻ và truy cập hồ sơ của mình bằng cách chia sẻ khoá bí mật với những cơ sở khám chữa bệnh nhất định với các tính năng bao gồm cấu trúc phân cấp của hồ sơ, hình ảnh [3], khả năng tìm kiếm dữ liệu mã hoá và các lựa chọn phân phối khoá. Khả năng mã hoá đồng cấu có thể cho phép nhà cung cấp dịch vụ đám mây tính toán trên dữ liệu mã hoá thay mặt bệnh nhân [4].

Ngành tài chính ngân hàng: Trong ngành tài chính có thể cần phải bảo vệ bí mật của cả dữ liệu và các hàm tính toán, như: dữ liệu về các doanh nghiệp, giá chứng khoán, hay các bảng cân đối kế toán [5]. Các hàm xử lý dữ liệu có thể cũng cần được bảo mật. Các hàm này dựa trên những mô hình dự đoán mới về giá chứng khoán và những mô hình đó có thể là sản phẩm của quá trình nghiên cứu lâu dài, vì thế cần giữ kín để đảm bảo lợi thế cạnh tranh. Với mã hoá đồng cấu đầy đủ, một số hàm có thể được tính toán một cách bí mật theo cách sau: khách hàng tải bản mã hoá của hàm lên đám mây theo [6], chẳng hạn như một chương trình mà một số phép tính liên quan tới các đầu vào mã hoá. Dữ liệu chuyển lên đám mây được mã hoá với khoá công khai của khách hàng. Dịch vụ đám mây thực hiện hàm bí mật bằng cách áp dụng bản mô tả được mã hoá của chương trình với đầu vào mã hoá mà họ nhận được. Sau khi xử lý, đám mây trả bản mã hoá của giá trị trả về cho khách hàng.

Ngành quảng cáo: có thể hình dung một công ty mỹ phẩm muốn sử dụng thông tin ngữ cảnh để gửi quảng cáo tới đúng những khách hàng tiềm năng thích hợp. Người tiêu dùng sử dụng điện thoại di động sẽ liên tục gửi thông tin ngữ cảnh về bản thân họ như vị trí, thời gian, các từ khoá trong thư điện tử và các hoạt động duyệt web. Thông tin tải lên còn có thể là hình ảnh thương hiệu, khuôn mặt, hay các thông tin định vị (như các đồ vật xung quanh, công sở, nhà riêng hay cửa hàng). Khi thông tin ngữ cảnh được tải lên đám mây, công ty mỹ phẩm có thể xử lý dữ liệu bằng một số hàm để xác định loại quảng cáo cần gửi tới điện thoại của khách hàng cụ thể đó. Các thông tin ngữ cảnh khác có thể là mức thu nhập, nghề nghiệp của khách, lịch sử mua sắm, lịch sử du lịch, địa chỉ nhà riêng,… Những thông tin đó mang tính riêng tư và việc thu thập chúng sẽ khiến khách hàng lo ngại. Nhưng mã hoá đồng cấu có thể mã hoá toàn bộ dữ liệu ngữ cảnh, hình ảnh [7] bằng khoá công khai của người dùng trước khi tải lên máy chủ; máy chủ tính toán trên dữ liệu mã hoá để xác định cần gửi đi loại quảng cáo nào (cũng được mã hoá bằng khoá công khai của người dùng); những hàm tính toán đó có thể bí mật hoặc công khai. Nếu nhà cung cấp dịch vụ đám mây không hợp tác với công ty muốn quảng cáo, mã hoá đồng cấu sẽ đảm bảo bí mật dữ liệu của người tiêu dùng đối với nhà cung cấp dịch vụ đám mây và các công ty quảng cáo [8].

Trong bài báo này, tác giả đề xuất sử dụng hệ mã hóa khóa công khai Paillier theo [1] và thực nghiệm tính toán khoảng cách Euclid đã được chứng minh [9], [10] 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 khái quát về cách tính toán khoảng cách Euclid trong nhận dạng dấu vân tay và phương thức sử dụng mã hóa

Paillier trong bài báo; sau đó là phần thực nghiệm để so sánh kết quả giữa cách tính thông thường và cách tính kết quả khi được mã hóa. Kết thúc bài báo đưa ra kết luận về cách tính đề xuất nhằm bảo vệ những dữ liệu (khoảng cách) riêng tư người dùng.

2. Mã hóa công khai Paillier và khoảng cách Euclid đối với mẫu có độ dài cố định

2.1. Phép đồng cấu và mã hóa công khai Paillier

2.1.1. Phép đồng cấu

Để hạn chế những rủi ro được nêu ở phần giới thiệu, 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 $f: A \to B$ bảo tồn phép toán μ của ngôi k, được xác định trên cả hai A và B nếu: $f\left(\mu_A(a_1, \dots, a_k)\right) = \mu_B\left(f(a_1), \dots, f(a_k)\right)$, với mọi $a_1, \dots, a_k$ trong A .

Các phép toán phải được bảo toàn bởi phép đồng cấu 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 [1], 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 Paillier theo [9].

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 xác định của nó. Tính chất đồng cấu vẫn đảm bảo khi thực hiện cộng thêm giá trị, các đặc điểm cộng hoặc nhân thêm 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,

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

Tích của bản mã m_1 với g lũy thừa m_2 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 thành lũy thừa của một hằng số k sẽ giải mã thành tích của bản rõ và hằng số:

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

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 mã hóa tích 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.

2.1.3. Ứng dụng mã hóa công khai Paillier trong bảo vệ tính riêng tư của mẫu

Do đã được mã hóa từ bản rõ, việc bảo mật thông tin của mẫu giúp chống lại các cuộc tấn công bản rõ. Do các thuộc tính đồng cấu đã nêu, hệ thống có thể linh hoạt chống lại các cuộc tấn công bản rõ, kể cả bản mã. Thông thường trong mật mã, khái niệm về tính linh hoạt không được coi là một ưu điểm lợi thế, tuy nhiên trong một số ứng dụng như bỏ phiếu điện tử thì tính năng này thực sự cần thiết. Một số ứng dụng tiêu biểu của thuật toán Paillier đã được ứng dụng hiện nay như sau:

Biểu quyết, bỏ phiếu điện tử: các tính chất đồng cấu của hệ mã Paillier có thể được sử dụng cho các hệ thống bỏ phiếu điện tử an toàn. Ví dụ một phiếu bầu nhị phân đơn giản ("ủng hộ" hoặc "phản đối"). Cho m cử tri bỏ phiếu 1 (ủng hộ) hoặc 0 (phản đối). Mỗi cử tri mã hóa lựa chọn của họ trước khi bỏ phiếu. Viên chức bầu cử lấy sản phẩm của m phiếu được mã hóa, sau đó giải mã kết quả và nhận được giá trị n , là tổng của tất cả các phiếu bầu. Sau đó, viên chức bầu cử biết rằng n người đã bỏ phiếu cho và $m\ x\ n$ người đã bỏ phiếu chống. Vai trò của r là số nguyên ngẫu nhiên đảm bảo rằng hai phiếu bầu tương đương sẽ mã hóa thành cùng một giá trị chỉ với khả năng xảy ra không đáng kể, do đó đảm bảo quyền riêng tư của cử tri.

Tiền điện tử: khả năng thay đổi một bản mã này thành một bản mã khác mà không làm thay đổi nội dung giải mã của nó. Ví dụ rằng chúng ta cần thanh toán cho một mặt hàng trực tuyến mà nhà cung cấp không cần biết số thẻ tín dụng, danh tính của khách hàng. Mục tiêu của cả tiền điện tử và bỏ phiếu điện tử là đảm bảo đồng tiền điện tử (tương tự như bỏ phiếu điện tử) hợp lệ, đồng thời không tiết lộ danh tính của người hiện đang liên kết với nó.

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

Chúng ta định nghĩa khái niệm dấu vân tay tham chiếu và dấu vân tay tìm ẩn dưới đây:

Dấu vân tay tham chiếu: Dấu vân tay được người dùng đăng ký trong điều kiện lý tưởng người dùng được gọi đến tại văn phòng cơ quan và dấu vân tay của họ được ghi lại với sự hướng dẫn được gọi là dấu vân tay tham chiếu. Ở đây, mức độ chính xác và tính khả dụng của các dấu vân tay tham chiếu là rất cao và khoảng cách Euclid lý tưởng bằng 0.

Dấu vân tay nhận dạng: Dấu vân tay được phát hiện lần đầu tiên (sử dụng các kỹ thuật hóa học có sẵn) từ hiện trường vụ án thực tế, người dùng dùng để xác thực và sau đó được ghi danh được gọi là dấu vân tay tiềm ẩn hoặc cơ hội dấu vân tay. Nhưng những dấu vân tay như vậy được tìm thấy trong điều kiện đứt gãy/hư hỏng tại hiện trường vụ án do vết bẩn, vết dầu loang, bề mặt ẩm ướt, tuyết, bụi,... Ở đây mức độ chính xác thấp nhưng khả năng có sẵn tại hiện trường vụ án là cao và khoảng cách Euclid càng lớn tương ứng với độ chính xác nhận dạng thấp. Cách tính khoảng cách Eculid theo [11], [12] trong bài báo này ta quy ước một số ký hiệu như sau:

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

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

Cho trước hai mẫu F là ma trận $\mathbf{T}_p$ và $\mathbf{T}_r$, điểm số $S_{euc}=d^2_{euc}(\mathbf{T}_p, \mathbf{T}_r)$, có thể được tính toán một cách hiệu quả như (1):

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

2.3. Phương pháp cộng 2 khoảng cách Euclid đối với mẫu có độ dài cố định sử dụng mã hóa Paillier

E biểu thị hàm mã hóa, s là số ngẫu nhiên và *pk* khóa công khai, và *E* được định nghĩa bằng (2) và D_{sk} là hàm giải mã với khóa riêng tư *sk*:

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

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 như (2):

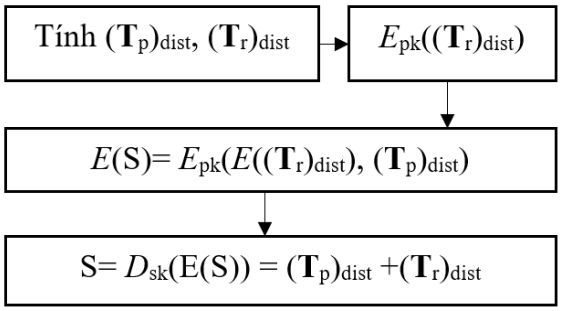
$$E(S_{euc}) = \prod_{f=1}^{F} E(1)^{p_f^2}.E(r_f^2).E(rf)^{-2pf} = \prod_{f-1}^{F}(1^*)^{p_f^2}.euc_{1f}^*.(euc_{2f}^*)^{-2pf} \quad (3)$$

Do đó, mẫu tham chiếu được lưu trữ trong cơ sở dữ liệu được mã hóa được xác định bởi các mật mã sau:

$$E(\mathbf{T}_r)_{euc} = \{1^*\} \cup \{euc1_f^*, euc2_f^*\}_{f=1}^{F} \quad (4)$$

Với $euc1_f^*$ và $euc2_f^* = E(r_f)$. Do đó, tất cả các bản mã có liên quan đến phương trình (3) được gửi bởi máy chủ và các tích và lũy thừa được tính trực tiếp trên máy khách.

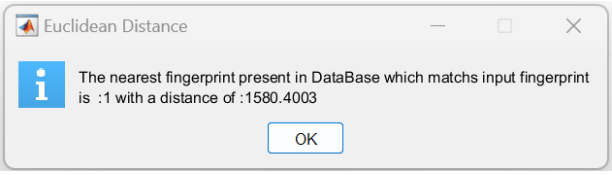
Với tính chất đồng cấu của hệ thống mật mã Paillier, *E*(1) có thể được tính toán và lưu trữ riêng biệt cho từng đối tượng tại thời điểm đăng ký, dẫn đến các giá trị được mã hóa khác nhau và do đó tăng tính bảo mật và quyền riêng tư của người dùng. Với $(\mathbf{T}_p)_{dist}$ và $(\mathbf{T}_r)_{dist}$ là khoảng cách Euclid được tính theo [11], ta có sơ đồ dưới đây:



Hình 1. *Sơ đồ tính toán phép cộng 2 khoảng cách Euclid dùng mã hóa Paillier*

3. Kết quả thực nghiệm tính khoảng cách Euclid kết hợp hệ mã Paillier để đánh giá kết quả

Để đánh giá toàn diện hiệu quả của phương pháp mã hóa khoảng cách Euclid 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 FVC2004_DB1. Do trong báo cáo này tác giả chỉ chứng minh tính đúng đắn đối với tính cộng đồng cấu của các bản rõ đối với khoảng cách Euclid 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 phần mềm tính toán khoảng cách điểm tương đồng Euclid như Hình 2.



Hình 2. *Khoảng cách điểm tương đồng Euclid của mẫu tham chiếu* $\mathbf{T}_p$ *và mẫu từ cảm biến* $\mathbf{T}_r$

3.1. Bộ dữ liệu thực nghiệm

Bộ dữ liệu FCV2004_DB1 được sử dụng trong chương trình gồm 32 mẫu chia thành 4 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 10x_1 (*x*: tên lớp mẫu với điều kiện $1 \leq x \leq 4$) làm mẫu tham chiếu $\mathbf{T}_r$.

3.2. Kết quả thực nghiệm

Bảng 1 thể hiện khoảng cách Euclid của $\mathbf{T}_p$ so sánh với mẫu đầu tiên trong lớp được chọn là $\mathbf{T}_r$. Đối với mẫu được chọn làm mẫu tham chiếu khoảng cách Euclid sẽ có giá trị bằng 0.

Bảng 1. *Tính toán khoảng cách của Euclid và tổng 2 khoảng cách không dùng mã hóa Paillier*

$\mathbf{T}_r$	Tên mẫu	$(\mathbf{T}_p)_{dist}$	$\boldsymbol{S}_{euc}$	$\mathbf{T}_r$	Tên mẫu	$(\mathbf{T}_p)_{dist}$	$\boldsymbol{S}_{euc}$
1	101_1	#	**0**	1	103_1	**3563.7955**	**0**
2	101_2	1746.9147	1746.9147	1	103_2	1786.4507	5350.2462
1	101_3	1580.4003	1580.4003	2	103_3	2210.5511	5774.3466
1	101_4	2063.6319	2063.6319	4	103_4	2148.2729	5712.0684
1	101_5	1065.8217	1065.8217	1	103_5	1972.7043	5536.4998
1	101_6	1917.1419	1917.1419	1	103_6	2064.226	5628.0215
2	101_7	2188.6997	2188.6997	1	103_7	2089.9011	5653.6966
3	101_8	1839.2283	1839.2283	4	103_8	1446.9414	5010.7369
1	102_1	**1843.8731**	**0**	1	104_1	**1789.5797**	**0**
1	102_2	1375.5933	3219.4664	1	104_2	1869.053	3658.6327
1	102_3	2111.7461	3955.6192	4	104_3	1164.4301	2954.0098
2	102_4	2171.6482	4015.5213	1	104_4	1978.5416	3768.1213
2	102_5	1203.8234	3047.6965	1	104_5	2218.4177	4007.9974
2	102_6	1453.2253	3297.0984	2	104_6	2064.7771	3854.3568
3	102_7	2644.5137	4488.3868	1	104_7	2078.0305	3867.6102
2	102_8	2397.1164	4240.9895	1	104_8	2130.5691	3920.1488

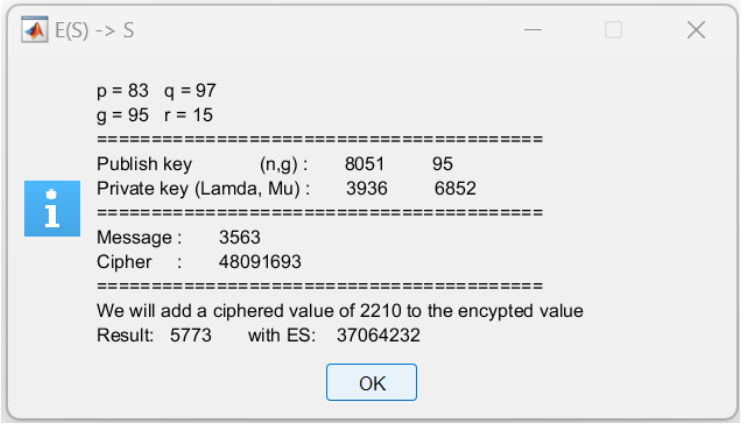
Bảng 2. *Tính toán khoảng cách S_{dist} sử dụng hệ mã Paillier, kết hợp với mẫu tham chiếu*

$\mathbf{T}_r$	Tên mẫu	$(\mathbf{T}_p)_{dist}$	S = $D_{sk}(E(S))$	$\mathbf{T}_r$	Tên mẫu	$(\mathbf{T}_p)_{dist}$	S = $D_{sk}(E(S))$
1	101_1	#	**0**	1	103_1	**3563**	**0**
2	101_2	1746	1746	1	103_2	1786	5349
1	101_3	1580	**1580**	2	103_3	2210	5773
1	101_4	2063	2063	4	103_4	2148	5711
1	101_5	1065	1065	1	103_5	1972	5535
1	101_6	1917	1917	1	103_6	2064	5627
2	101_7	2235	2235	1	103_7	2089	5652
3	101_8	1839	1839	4	103_8	1446	5009
1	102_1	**1843**	**0**	1	104_1	**1789**	**0**
1	102_2	1375	3218	1	104_2	1869	3658
1	102_3	2111	3954	4	104_3	1164	**2953**
2	102_4	2171	4014	1	104_4	1978	3767
2	102_5	1203	**3046**	1	104_5	2218	4007
2	102_6	1453	3296	2	104_6	2064	3853
3	102_7	2644	4487	1	104_7	2078	3867
2	102_8	2397	4240	1	104_8	2130	3919

Quá trình mã hóa Paillier tác giả chọn 2 số p = 83, q = 97 phù hợp với lý thuyết [9]. 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 $0 \leq g, r \leq 150$ phù hợp với lý thuyết. 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 3. Do hệ mã hóa Paillier chỉ hoạt động được với thông điệp là số nguyên vì vậy đối với khoảng cách Euclid tính được tác giả sử dụng làm tròn xuống.

Qua kết quả Bảng 1 và Bảng 2 ta thấy kết quả của các phép tính tổng 2 khoảng cách Euclid với 2 cách tính khác nhau thông qua mô tả ở Hình 1. Sau khi tính S và so sánh kết quả, ta thấy rằng kết quả giá trị S Bảng 1 xấp xỉ bằng giá trị S tại Bảng 2 với độ chênh lệch nhỏ hơn 2 (do các giá trị đã được làm tròn). Nếu chọn $\delta = \min(S_1,...,S_8)$ của những mẫu có $\mathbf{T}_r = x$ *với* $(1 \leq x \leq 4)$. Từ kết quả cho thấy các mẫu có $S > \delta$ lớn hơn giá trị trung bình là những mẫu cần được lựa chọn để kết hợp và đưa thêm vào cơ sở dữ liệu hoặc kết luận đây là những mẫu khó nhận biết hoặc nhận biết 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 sử dụng hệ mã hóa Paillier để mã hóa giá trị khoảng cách Euclid của mẫu nhằm giữ bí mật đối với giá trị thước đo Euclid của mẫu vân tay, thuộc lớp mẫu có độ dài cố định; bảo đảm sự riêng tư của người dùng. Kết quả tính toán S là tổng 2 độ dài Euclid của mẫu tham chiếu và mẫu thu được từ cảm biến xấp xỉ bằng so với kết quả tính toán khi kết hợp sử dụng hệ mã hóa công khai Paillier để mã hóa khoảng cách Euclid và tính S như đã trình bày ở phần trên.

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, đảm bảo giữ bí mật đối với dữ liệu là số nguyên. Cần lưu ý rằng phép đồng cấu chỉ có tính cộng, tính nhân nên không áp dụng với những thuật toán có phép trừ hoặc chia. Hướng phát triển của bài báo đề xuất phương pháp bảo vệ mẫu, tính riêng tư của người dùng, lựa chọn ngưỡng δ phù hợp mục đích loại bỏ những mẫu không thích hợp, hoặc cần cải thiện; ra quyết định kết hợp vào hệ thống với ngưỡng δ được lựa chọn đối với lớp mẫu có độ dài cố định.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1] J. Katz and Y. Lindell, *Introduction to Modern Cryptography: Principles and Protocols*. Chapman & Hall/CRC, 2007.

[2] Adida, *Helios: Web-based Open-Audit Voting*, Usenix Security Symposium, pp. 335-348, 2017.

[3] A. M. Vengadapurvaja, G. Nisha, R. Aarthy and N. Sasikaladevi, "An Efficient Homomorphic Medical Image Encryption Algorithm For Cloud Storage Security," *International Conference on Advances in Computing & Communications*, vol. 115, pp. 643-650, 2017.

[4] K. T. Son, "Full homomorphic coding and its application in electrically secure health monitoring cloud math," (in Vietnamese), Master's Thesis, Hanoi National University, 2021.

[5] H. -T. Peng, W. W. Y. Hsu, J. -M. Ho, and M. -R. Yu, "Homomorphic encryption application on FinancialCloud framework," *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-5, 2016, doi: 10.1109/SSCI.2016.7850013.

[6] I. Ahmad and A. Khandekar, "Homomorphic encryption method applied to cloud computing," *International Journal of Information & Computer Technology*, vol. 15, pp. 1519-1530, 2014.

[7] M. I. Wade, H. C. Ogworonjoy, and M.Gul, *Red Green Blue Image Encryption Based on Paillier Cryptographic System*, Department of Electrical Engineering and Computer Science, Howard University, Washington, 2018.

[8] L. J. Helsloot, G. Tillem, and Z. Erkin, "AHEad: Privacy-preserving online behavioural advertising using homomorphic encryption," *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, pp. 1-6, 2017, doi: 10.1109/WIFS.2017.8267662.

[9] M. A. Will and R. K. L. Ko, "A guide to homomorphic encryption," in *The Cloud Security Ecosystem*, 2015, doi: 10.1016/B978-0-12-801595-7.00005-7.

[10] A. Joss and A. Jain, "Biometric sensor interoperability," in *A case study in Fingerprints*, pp. 134-145, 2004, doi: 10.1007/978-3-540-25976-3_13.

[11] Jadhav, Barbadekar, and Patil, *Euclidean Distance Based Fingerprint Matching*, Recent Researches in Communications, Automation, Signal Processing, Nanotechnology, Astronomy and Nuclear Physics, 2011.

[12] N. Bhargava, A. Kumawat, and R. Bhargava, "Fingerprint Matching of Normalized Image based on Euclidean Distance," *International Journal of Computer Applications*, vol. 120, no. 24, pp. 20-23, 2015.