ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA

NGUYỄN THỊ ÁI THẢO

BẢO MẬT TRONG XÁC THỰC TỪ XA SỬ DỤNG ĐẶC TRƯNG SINH TRẮC

Ngành: Khoa học Máy Tính
Mã số ngành: 62480101

TÓM TẮT LUẬN ÁN TIẾN SĨ

TP. HỒ CHÍ MINH - NĂM 2021

Công trình được hoàn thành tại **Trường Đại học Bách Khoa – ĐHQG-HCM**

Người hướng dẫn 1: PGS. TS. Đặng Trần Khánh
Người hướng dẫn 2:

Phản biện độc lập 1:
Phản biện độc lập 2:

Phản biện 1:
Phản biện 2:
Phản biện 3:

Luận án sẽ được bảo vệ trước Hội đồng đánh giá luận án họp tại
.....
.....
vào lúc giờ ngày tháng năm

Có thể tìm hiểu luận án tại thư viện:

- Thư viện Trường Đại học Bách Khoa – ĐHQG-HCM
- Thư viện Đại học Quốc gia Tp.HCM
- Thư viện Khoa học Tổng hợp Tp.HCM

DANH MỤC CÔNG TRÌNH ĐÃ CÔNG BỐ

Tạp chí quốc tế

1. Thi Ai Thao Nguyen, Tran Khanh Dang, "Privacy Preserving Biometric-based Remote Authentication with Secure Processing Unit on Untrusted Server", *IET Biometrics*, The Institution of Engineering and Technology, United Kingdom, vol.8(1), pp. 79-91, ISSN 2047-4938, 2019. (**SCIE, Q2)**
2. Thi Ai Thao Nguyen, Tran Khanh Dang, "Protecting Biometrics using Fuzzy Extractor and Non-Invertible Transformation Methods in Kerberos Authentication Protocol", *Transactions on Large-Scale Data- and Knowledge-Centered Systems*, 31, pp. 47-66, LNCS 10140, ISBN 978-3-662-54172-2, Springer Verlag, 2017. (**Scopus)**
3. Thu Thi Bao Le, Tran Khanh Dang, Quynh Chi Truong, Thi Ai Thao Nguyen, "Protecting Biometric Features by Periodic Function-Based Transformation and Fuzzy Vault", *Transactions on Large-Scale Data- and Knowledge-Centered Systems*, 16, pp. 57-70, LNCS 8960, Springer Verlag, 2014. (**Scopus)**

Kỷ yếu hội nghị quốc tế

1. Thi Ai Thao Nguyen, Tran Khanh Dang, Dinh Thanh Nguyen, "Non-Invertibility for Random Projection based Biometric Template Protection Scheme", In Proceedings of *the 15th International Conference on Ubiquitous Information Management and Communication (IMCOM 2021)*, IEEE, Seoul, South Korea, January 4-6, 2021.
2. Phuong Thao Nguyen Le, Tran Khanh Dang, Tran Tri Dang, Thi Ai Thao Nguyen, "A 3-way energy efficient authentication protocol using Bluetooth Low Energy", In Proceedings of *the 7th International Conference on Future Data and Security Engineering (FDSE 2020 – Part I)*, Virtual (QNU, Binh Dinh, Vietnam), November 25-27, 2020, LNCS 12466, Springer Verlag.
3. Thi Ai Thao Nguyen, Tran Khanh Dang, Dinh Thanh Nguyen, "A New Biometric Template Protection using Random Orthonormal Projection and Fuzzy Commitment", In Proceedings of *the 13th International Conference on Ubiquitous Information Management and Communication*

(IMCOM 2019), pp. 723 – 733, LNCS, Springer Verlag, Phuket, Thailand, January 4-6, 2019.

4. Thi Ai Thao Nguyen, Tran Khanh Dang, Quynh Chi Truong, Dinh Thanh Nguyen, "Secure Biometric-based Remote Authentication Protocol using Chebyshev Polynomials and Fuzzy Extractor*"*, In Proceedings of *AUN/SEED-Net Regional Conference on Computer and Information Engineering (RCCIE2017)*, HCMUT, Ho Chi Minh City, Vietnam, VNUHCM Press, ISBN 978-604-73-5687-4, pp. 31-36, November 30-December 01, 2017.
5. Thi Ai Thao Nguyen, Tran Khanh Dang, "Combining Fuzzy Extractor in Biometric-Kerberos based Authentication Protocol", In Proceedings of *the 2015 International Conference on Advanced Computing and Applications (ACOMP 2015)*, pp. 1 – 6, IEEE CPS, ISBN-13: 978-1-4673-8234-2, Ho Chi Minh City, Vietnam, November 23-25, 2015
6. Thi Ai Thao Nguyen, Dinh Thanh Nguyen, Tran Khanh Dang, "A Multi-factor Biometric Based Remote Authentication Using Fuzzy Commitment and Non-invertible Transformation", In Proceedings of *Information & Communication Technology-EurAsia Conference 2015*, pp. 77 - 88, LNCS 9357, Springer-Verlag, October 4-7, 2015, Daejeon, Korea.

Đề tài nghiên cứu khoa học

1. Bảo mật trong xác thực từ xa bằng đặc trưng sinh trắc ứng dụng vi xử lý bảo mật, C2018-20-13/ĐHQG loại C, 2018-2019.
2. Phát triển kỹ thuật bảo vệ mẫu sinh trắc hiệu quả trong giao thức xác thực từ xa sử dụng đặc trưng sinh trắc, T-PTN-2017-91/Trường/Trường, 2017-2018.
3. Bảo mật cho xác thực từ xa sử dụng đặc trưng sinh trắc dựa trên giao thức Kerberos, T-KHMT-2016-60/Trường, 2016-2017.
4. Phát triển kỹ thuật sinh khóa sinh trắc hiệu quả dựa trên phương pháp thống kê để bảo vệ dữ liệu trên thiết bị di động, T-KHMT-2014-38/Trường, 2014-2015.
5. Bảo vệ tính riêng tư sinh trắc trong xác thực từ xa, TNCS-2014-KHMT-06/Học viên CH + NCS, 2014-2015.
6. Đồng xác thực và bảo vệ dữ liệu trên thiết bị di động sử dụng sinh trắc học, Bộ-ĐHQG-Sở, 2013-2015.

CHƯƠNG 1 GIỚI THIỆU

1.1 Giới thiệu đề tài

Những năm cuối thế kỉ XX và đầu thế kỉ XXI chứng kiến sự lớn mạnh vượt bậc của mạng Internet cả về quy mô và chất lượng. Internet được ứng dụng rộng rãi ở mọi ngành nghề, lĩnh vực kinh tế, xã hội và an ninh. Tính phổ biến rộng rãi khiến Internet đã và đang là nền tảng cơ sở cho các giao dịch thương mại toàn cầu và các ứng dụng của giao dịch điện tử tạo thành một hình thức "xã hội ảo" với các đặc trưng riêng biệt. Tuy nhiên, môi trường mạng không phải luôn an toàn. Đặc trưng của Internet là tính ảo và tính tự do, mọi người đều có thể tham gia và ít khi để lại dấu vết của mình. Việc xác thực mỗi cá nhân trên mạng thường khó khăn nên nguy cơ xảy ra giả mạo địa chỉ, giả danh người dùng hợp pháp và bị lừa đảo trực tuyến là rất cao. Đây vừa là điểm mạnh cũng chính là điểm yếu của các giao dịch qua mạng. Những năm gần đây các hình thức tội phạm sử dụng công nghệ cao đã phát triển nhanh chóng cùng với sự phát triển của công nghệ.

Mặc dù có các nguy cơ kể trên, tính phổ dụng và tiện lợi của công nghệ cao đã và đang làm thay đổi diện mạo cuộc sống và các giao dịch điện tử đang phát triển mạnh trên phạm vi toàn thế giới. Khi các dịch vụ thương mại điện tử được sử dụng rộng rãi như ngày nay, nhu cầu tất yêu để nó tồn tại lâu dài ngoài chất lượng dịch vụ còn phải kể đến tính bảo mật, đảm bảo sự an tâm cho khách hàng khi sử dụng dịch vụ. Phương pháp bảo mật đầu tiên phải kể đến để đảm bảo tính bảo mật cho các hệ thống thông tin là xác thực. Phương pháp xác thực truyền thống mà hầu hết các dịch vụ thương mại điện tử vẫn đang sử dụng là username/ password (tên tài khoản/ mật khẩu). Tuy nhiên, phương pháp này đang dần lộ những nhược điểm thuộc về bản chất của nó. Mật khẩu không thể phân biệt được người dùng hợp pháp với kẻ mạo danh có khả năng truy xuất mật khẩu của người dùng. Bên cạnh đó, về bản chất mật khẩu càng khó – càng bảo mật thì lại càng khó nhớ, hay nói một cách khác mật khẩu đúng nghĩa khó cho con người có thể nhớ nhưng lại dễ cho máy tính có thể đoán ra được. Đặc biệt, với sự phát triển của khoa học kĩ thuật ngày nay, khả năng của máy tính ngày càng được nâng cao,

đồng nghĩa với khả năng phá các loại mật khẩu cũng rất cao. Có hai vấn đề cần phải giải quyết ở đây là làm sao người sử dụng không cần phải nhớ mật khẩu của mình nữa và dữ liệu xác thực dù có bị kẻ tấn công đánh cắp cũng không thể nào sử dụng được. Với yêu cầu đó, phương pháp xác thực bằng đặc trưng sinh trắc ra đời, cùng với những ưu điểm cuả nó phương pháp này đang ngày dần thay thế phương pháp truyền thống cũ. Ưu điểm đầu tiên phải kể đến là đặc trưng sinh trắc (như khuôn mặt, giọng nói, tròng mắt, vân tay, dáng đi, chữ kí,…) phản ánh duy nhất một cá nhân cụ thể nên có thể ngăn chặn được việc sử dụng nhiều định danh của cùng một cá nhân. Hơn thế nữa, việc sử dụng đặc trưng sinh trắc thuận tiện hơn vì nó không đòi hỏi người sử dụng phải luôn ghi nhớ hay phải luôn mang nó bên mình.

Tuy vậy, tiện ích luôn đi kèm với những thử thách. Việc sử dụng đặc trưng sinh trắc trong xác thực đòi hỏi phải có các kĩ thuật để loại trừ nhiễu gây ra khi các bộ cảm ứng thu nhận những đặc trưng sinh trắc. Bên cạnh đó, một số vấn đề liên quan đến bảo mật và tính riêng tư cũng được đặt ra như: đặc trưng sinh trắc học khó có thể thay đổi và hủy bỏ khi bị đánh cắp dễ dàng như password, và một khi đã mất sẽ không có thể thu hồi lại được,… Đặc biệt trong kiến trúc xác thực từ xa, vấn đề bảo mật khi sử dụng đặc trưng sinh trắc càng quan trọng. Những khó khăn được đặt ra như là:

- Đặc trưng sinh trắc phản ánh bản thân người dùng, chứa đựng nhiều thông tin cá nhân. Trong kiến trúc xác thực từ xa, những thông tin đó cần được lưu trên cơ sở dữ liệu, và được tải đi trên đường truyền mạng. Làm sao bảo vệ những thông tin riêng tư đó trên đường truyền mạng luôn có nhiều mối đe dọa và ngăn chặn sự tò mò hay nghiệm trọng hơn là tấn công của chính người quản trị cơ sở dữ liệu.
- Số lượng dịch vụ thương mại điện tử mà người dùng sử dụng có thể rất nhiều trong khi số lượng đặc trưng sinh trắc của con người có hạn. Người dùng có xu hướng sử dụng một đặc trưng để xác thực cho nhiều dịch vụ.

Việc này rất nguy hiểm nếu các dịch vụ liên kết với nhau có thể biết được tất cả những hoạt động trên mạng của người dùng.

- Ngoài vấn đề về sinh trắc, những phương pháp bảo vệ thông điệp được truyền đi trên đường truyền mạng đảm bảo giao thức xác thực chạy đúng, cũng cần được nghiên cứu.

Trong đề tài này, tôi tập trung xây dựng một hệ thống xác thực từ xa sử dụng đặc trưng sinh trắc cung cấp sự tiện lợi cho người dùng. Đồng thời, hệ thống phải đảm bảo được tính bảo mật, bảo vệ được đặc trưng sinh trắc của người dùng trong ngữ cảnh các mối đe dọa tấn công có thể đến từ bất kì đâu, từ bên ngoài và có thể xuất phát từ bên trong hệ thống. Việc sử dụng bộ đồng xử lý bảo mật ở phía máy chủ nhằm giảm thiểu nguy cơ tấn công từ bên trong, đồng thời đảm bảo những thông tin nhạy cảm của người dùng được bảo vệ ở mức độ cao nhất có thể ở công nghệ hiện nay.

1.2 Mục tiêu và phạm vi luận án

1.2.1 Mục tiêu luận án

Với nội dung được giới thiệu tóm tắt trong phần trên, mục tiêu chính của luận án là đưa ra giải pháp bảo mật cho hệ thống xác thực từ xa sử dụng đặc trưng sinh trắc. Trong đó khía cạnh bảo mật chú trọng ở hai điểm chính: bảo vệ mẫu sinh trắc của người dùng, và thực hiện tính toán an toàn trong ngữ cảnh máy chủ xác thực không đáng tin cậy.

Bảo vệ mẫu sinh trắc là một lĩnh vực rất được quan tâm khi phát triển hệ thống xác thực bằng đặc trưng sinh trắc. Lý do chính khiến mẫu sinh trắc của người dùng trở thành dữ liệu nhạy cảm dễ bị dòm ngó bởi kẻ tấn công, do đặc tính phản ánh các đặc điểm sinh lý của người sở hữu nó; hơn nữa, các đặc trưng sinh trắc mà con người sở hữu thì hữu hạn nên việc sử dụng lặp lại dễ tạo ra lỗ hổng bảo mật. Vì thế, để bảo vệ được mẫu sinh trắc cho người dùng thì các mẫu này nên được chuyển đổi sang vùng an toàn trước khi được chuyển đi để xử lý. Việc chuyển đổi này giống như ta mã hóa mật khẩu trước khi lưu trữ nó vào cơ sở dữ

liệu. Tuy nhiên, đặc trưng sinh trắc là một loại dữ liệu có nhiễu nên việc chuyển đổi và thực hiện so trùng để xác thực sẽ khác so với việc dùng mật khẩu thông thường. Bên cạnh đó, mỗi loại đặc trưng sinh trắc lại có một cách biểu diễn khác nhau nên các phương pháp bảo vệ cũng cần phù hợp với từng loại. Để đảm bảo mục tiêu này, các bước cần thực hiện:

- Tìm hiểu về hệ thống xác thực, đặc biệt hệ thống xác thực sử dụng đặc trưng sinh trắc.
- Tìm hiểu đặc điểm của các loại đặc trưng sinh trắc, so sánh các loại sinh trắc với nhau để cho thấy loại sinh trắc nào phù hợp với hệ thống triển khai.
- Tìm hiểu các kĩ thuật bảo vệ đặc trưng sinh trắc. So sánh, phân tích ưu nhược điểm của từng kĩ thuật nhằm đề xuất kĩ thuật bảo vệ phù hợp.

Mục tiêu thứ hai trong luận án này liên quan tới tính toán an toàn trong ngữ cảnh máy chủ không đáng tin. Trong kiến trúc xác thực từ xa, người dùng gởi thông tin xác thực qua đường truyền mạng, tới máy chủ để được xử lý. Vấn đề bảo mật trong các kiến trúc dạng này được chú trọng ở việc bảo vệ dữ liệu nhạy cảm khi được truyền đi trong môi trường mạng, và khi đến được máy chủ liệu máy chủ có đủ mạnh để xử lý các dữ liệu đó một cách an toàn chống lại được các tấn công từ bên ngoài và đặc biệt từ chính bên trong của hệ thống. Để đáp ứng được mục tiêu này, các công việc cần thực hiện:

- Tìm hiểu các một số hệ thống xác thực từ xa. Một số các loại tấn công điển hình trên đường truyền mạng sẽ được khảo sát và các phương thức để chống lại chúng.
- Tìm hiểu các giải pháp để đáp ứng yêu cầu tính toán an toàn trong ngữ cảnh máy chủ không đáng tin cậy.
- Đề xuất giao thức xác thực từ xa sử dụng đặc trưng sinh trắc đảm bảo tính an toàn đối với việc tính toán ở phía máy chủ. Đồng thời kĩ thuật bảo vệ mẫu sinh trắc cũng sẽ được đưa vào trong giao thức này.

1.2.2 Phạm vi luận án

Đối với bài toán liên quan đến mẫu sinh trắc dùng để xác thực, luận án này có các giả định và giới hạn sau:

- Để có thể sử dụng trong hệ thống xác thực, mẫu sinh trắc cần qua các bước tiền xử lý, trước khi được rút trích đặc trưng quan trọng nhất để lưu trữ và so trùng. Trong phạm vi luận án này, chỉ tập trung xử lý dữ liệu đã được rút trích. Quá trình tiền xử lý dữ liệu hình ảnh thô sẽ không được đề cập tới. Dữ liệu sau khi tiền xử lý sẽ được rút trích, quá trình rút trích đặc trưng sẽ sử dụng kĩ thuật PCA. Lý do chọn kĩ thuật này và độ hiệu quả của nó trong quá trình xác thực sẽ không nằm trong phạm vi luận án.
- Một điểm đáng lưu ý nữa, do phạm vi nghiên cứu chỉ xét từ thời điểm mẫu đặc trưng đã được rút trích, nên các vấn đề bảo mật trước quá trình này ví dụ các kiểu tấn công đánh lừa máy cảm biến (làm giả giọng nói, vân tay, khuôn mặt,…) để xác thực sẽ không được đề cập đến.

Đối với bài toán liên quan tới tính toán an toàn về phía máy chủ, có rất nhiều kịch bản cho máy chủ không trung thực như: làm theo đúng các bước trong giao thức nhưng sẽ lợi dụng dữ liệu đầu vào để giả mạo người dùng, làm theo đúng các bước trong giao thức nhưng thay đổi kết quả cuối cùng, làm theo giao thức nhưng không thực sự tính toán nên cho kết quả không đáng tin ở các bước, và không làm theo giao thức chỉ xuất ra kết quả đánh lừa người dùng,… Đối với mỗi kịch bản tương ứng với mỗi cấp độ tấn công sẽ cần có những kĩ thuật để phát hiện mức độ không trung thực của máy chủ. Trong phạm vi luận án này, máy chủ giả định sẽ có khả năng tấn công ở mức bị động, có nghĩa là sẽ lợi dụng các thông tin và tài nguyên của hệ thống để mạo danh người dùng nhằm thực hiện các hành vi phi pháp chứ không tác động làm thay đổi hệ thống. Do đó, đối tượng cần được bảo vệ là dữ liệu sinh trắc của người dùng không cho kẻ tấn công bên ngoài và đặc biệt là bên trong xâm phạm.

CHƯƠNG 2 BẢO VỆ MẪU SINH TRẮC

2.1 Đặc điểm

Các kĩ thuật bảo vệ mẫu sinh trắc được đánh giá trên 3 tiêu chí:

- Tính khả đổi (Cancelability): Một mẫu sinh trắc bị lộ nên có thể được loại bỏ và có thể được thay thể bởi một mẫu mới. Vấn đề nằm ở chỗ, số lượng đặc trưng sinh trắc của con người rất giới hạn, không thể thay đổi liên tục như việc thay mật khẩu. Do đó, tiêu chí này yêu cầu thỏa ràng buộc việc thay đổi mẫu sinh trắc lưu trữ trong cơ sở dữ liệu không buộc người dùng phải thay thế mẫu sinh trắc gốc của họ. Bên cạnh đó, tiêu chí này còn yêu cầu các kĩ thuật bảo vệ không nên tạo ra các mẫu sinh trắc đã biến đổi giống nhau ở các ứng dụng khác nhau của cùng một người dùng. Điều này để phòng tránh dạng tấn công cross-matching.
- Bảo mật: khó có thể tính toán được đặc trưng sinh trắc gốc từ các đặc trưng sinh trắc biến đổi được lưu trong cơ sở dữ liệu. Việc này nhằm tránh việc kẻ gian có thể xây dựng lại mẫu giả dựa vào các mẫu đánh cắp được từ các cuộc tấn công vào hệ thống.
- Hiệu suất: khi áp dụng một kĩ thuật bảo vệ mẫu sinh trắc lên hệ thống xác thực, việc này không nên làm giảm độ chính xác nhận dạng của hệ thống. Tiêu chí này đề cập đến khả năng phân hóa (discriminability) của mẫu sinh trắc gốc, khả năng này nên được bảo tồn sau quá trình biến đổi.

2.2 Các hướng tiếp cận bảo vệ mẫu sinh trắc

Đề thỏa mãn các tiêu chí đó nhiều nhất có thể, rất nhiều kĩ thuật đã được đề xuất. Dựa vào đặc điểm từng kĩ thuật, ta có thể phân loại các kĩ thuật đó thành hai hướng tiếp cận: hướng biến đổi đặc trưng (Cancelable biometrics) và hướng mã hóa sinh trắc (Biometric cryptosystem).

Theo hướng tiếp cận biến đổi đặc trưng, mẫu sinh trắc được biến đổi bằng một hàm định nghĩa bởi các yếu tố do người dùng cung cấp như khóa, mật khẩu, chuỗi ngẫu nhiên… Mục đích của cách tiếp cận này là cung cấp tính đa dạng (diversity) và không có khả năng kết nối (unlinkability) bằng cách sử dụng nhiều hàm biến đổi khác nhau ở các ứng dụng khác nhau đối với một nhóm người dùng như nhau. Hướng tiếp cận này được phân ra làm hai loại:

- Salting: đây là kiểu biến đổi mà có thể suy ngược trở lại, tức là có thể tính được dữ liệu gốc từ dữ liệu đã biến đổi nếu biết hàm biến đổi. Do đó, nhân tố tạo hàm biến đổi (hay có thể coi là khóa xác thực) phải được giữ bí mật. Nhờ vào nhân tố khóa này, tiêu chí cancelability được đảm bảo, và nó cũng cho ra tỉ lệ chấp nhận sai thấp. Tuy nhiên, bất lợi lớn nhất của phương pháp này nằm cũng ở nhân tố khóa. Nếu khóa bị lộ, mẫu sinh trắc gốc cũng bị lộ theo.
- Non-invertible transform: các kĩ thuật thuộc hướng tiếp cận này sử dụng các hàm biến đổi không có khả năng suy ngược để bảo vệ mẫu sinh trắc gốc. Biến đổi bất khả ngược cần đến một hàm một chiều sao cho việc tính toán theo chiều thuận thì dễ dàng, nhưng để suy ngược thì hầu như không thể. Nhân tố khóa để sinh ra hàm này có thể được công khai. Thậm chí kẻ tấn công có thể biết được khóa này và mẫu sinh trắc đã biến đổi, nhưng khó có thể thực hiện việc tính toán ngược để phục hồi lại mẫu sinh trắc gốc. Chính yếu tố này giúp cho hướng tiếp cận này có độ bảo mật tốt hơn hướng tiếp cận salting. Tiêu chí khả đổi có thể dễ dàng đạt được bằng cách thay đổi nhân tố khóa sinh ra hàm biến đổi. Tuy nhiên, điểm bất lợi chính của phương pháp này ở chỗ chúng ta phải đánh đổi khả năng phân hóa (discriminability) của mẫu sinh trắc và độ khó của hàm biến đổi. Việc khó khăn của người thiết kế hàm biến đổi cần phải làm sao cho cân bằng giữa hai tính chất này cùng lúc. Hơn nữa, mỗi hàm thường chỉ thích hợp với một số đặc trưng sinh trắc nhất định.

Hướng tiếp cận mã hóa sinh trắc (biometric cryptosystem) ban đầu được phát triển cho việc bảo mật khóa mã hóa bằng cách sử dụng dữ liệu sinh trắc hoặc trực tiếp sinh ra khóa mã hóa từ dữ liệu sinh trắc. Tuy nhiên, nó cũng được sử dụng trong việc bảo vệ mẫu sinh trắc. Trong hướng tiếp cận này, một số dữ liệu trợ giúp công khai được lưu trữ trong cơ sở dữ liệu. Những dữ liệu trợ giúp này không tiết lộ bất cứ thông tin quan trọng nào về mẫu sinh trắc gốc, chúng được sử dụng trong suốt quá trình so trùng để rút trích ra khóa sinh trắc từ dữ liệu sinh trắc được cung cấp bởi người dùng. Hướng tiếp cận này có thể được phân loại thành hai nhánh: key-binding và key-generation

- Key-binding: ở cách tiếp cận này mẫu sinh trắc gốc được kết hợp với một khóa ngẫu nhiên. Kết quả của quá trình kết hợp đó được xem như là một dữ liệu hỗ trợ (helper data) và được lưu trữ trong cơ sở dữ liệu. Dữ liệu hỗ trợ này không tiết lộ thông tin về khóa xác thực hay về đặc trưng sinh trắc gốc của người dùng. Tuy nhiên, khuyết điểm lớn nhất của các tiếp cận này là thiếu đi tiêu chí khả đổi bởi rõ ràng nó không được thiết kế cho điều này. Thêm vào đó, quá trình so trùng của nó phụ thuộc vào lược đồ sửa lỗi mà nó sử dụng. Điều này có thể dẫn đến độ chính xác của hệ thống bị giảm.
- Key-generation: ở hướng tiếp cận này, khóa xác thực được rút trích trực tiếp từ mẫu sinh trắc. Do đó không cần phải quan tâm đến tính bảo mật của khóa xác thực. Tuy nhiên, hướng tiếp cận này có thường có khả năng phân biệt thấp, có thể hiểu khả năng này thông qua hai thuật ngữ *key stability* và *key entropy*. Key stability, độ ổn định của khóa liên quan tới khả năng khóa được tái sinh từ đặc trưng sinh trắc; và key entropy, độ bất định của khó liên quan tới số lượng khóa có khả năng được sinh ra. Ví dụ, nếu một lược đồ sinh ra cùng một khóa từ các đặc trưng sinh trắc khác nhau thì lược đồ đó có độ ổn định khóa cao nhưng độ bất định bằng không, điều này dẫn tới tỉ lệ chấp nhận sai cao. Mặt khác, nếu lược đồ

sinh ra các khóa khác nhau đối với một người dùng, lược đồ này có tính bất định cao nhưng độ ổn định bằng không, dẫn tới tỉ lệ từ chối sai cao. Do đó, hạn chế của cách tiếp cận này là khó có thể sinh ra một khóa mà vừa có tính ổn định và bất định cao được. Bên cạnh đó, hướng tiếp cận này cũng không đảm bảo được tiêu chí khả đổi.

Từ phân tích trên, có thể thấy rằng không một cách tiếp cận bảo vệ mẫu sinh trắc đơn lẻ nào có thể đồng thời thỏa mãn cả ba tiêu chí (khả đổi, bảo mật, và hiệu quả). Do đó, nhiều nghiên cứu gần đây có xu hướng tích hợp các ưu điểm của hai hướng tiếp cận trên đồng thời loại bỏ đi những hạn chế của chúng. Hướng tiếp cận lai là sự kết hợp của hai hay nhiều kĩ thuật để tạo ra một lược đồ bảo vệ mẫu sinh trắc duy nhất. Trong nội dung luận văn này, một phương pháp lai khác sẽ được áp dụng để bảo vệ mẫu sinh trắc trong hệ thống xác thực từ xa sử dụng đặc trưng sinh trắc, đó chính là sự kết hợp giữa kĩ thuật Cam kết mờ (Fuzzy commitment) và Phép chiếu trực giao ngẫu nhiên (Random Projection).

2.3 Lược đồ lai giữa Phép chiếu trực giao ngẫu nhiên và Cam kết mờ

2.3.1 Phép chiếu trực giao ngẫu nhiên

Phép chiếu trực giao ngẫu nhiên (Random Projection - RP) là phương pháp sử dụng ma trận trực giao để ánh xạ một điểm sang một miền không gian mới mà vẫn bảo toàn được khoảng cách điểm. Ban đầu, RP được đề xuất như một lược đồ bảo vệ mẫu sinh trắc riêng lẻ, và được xếp vào hướng tiếp cận salting. RP đôi khi cũng được sử dụng như một bước để sinh ra khóa sinh trắc từ dữ liệu sinh trắc nhằm đảm bảo tính khả đổi (cancelability), nhờ đó các khóa có thể thay đổi như mật khẩu trong quá trình xác thực. Quy trình thực hiện :

- Tạo m vector ngẫu nhiên thuộc không gian $\mathcal{R}^n$ (n là số chiều của vector đặc trưng sinh trắc) từ một số yếu tố do người dùng lựa chọn.
- Áp dụng quá trình trực giao hóa trên tập vector ngẫu nhiên trên để tạo ma trận trực giao $A[m \times n]$.

- Biến đổi vector đặc trưng sinh trắc X thành một vector đặc trưng sinh trắc biến đổi Y sử dụng ma trận trực giao A : $Y = A \times X$.

Để tạo ma trận trực giao A, ta sử dụng quy trình Hisham Al-Assam. Trước tiên ta xét ma trận có kích thước 2 x 2 như sau:

$$I_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Có thể nhận thấy rằng ma trận này luôn trực giao với mọi giá trị θ. Dựa vào tính chất này ta có thể tạo ra một ma trận $[2n \times 2n]$ từ những ma trận $[2 \times 2]$ ở trên. Al-Assam đề xuất phép xây dựng ma trận trực giao ngẫu nhiên như sau:

Từ một tập n các giá trị ngẫu nhiên $\{\theta_1, \theta_2 \ldots \theta_n\}$ là các số thực nằm trong $[0 \ldots 2\pi]$ được sinh ra ngẫu nhiên từ những yếu tố từ người dùng như khóa hoặc token. Mỗi θ_i ta tạo ra ma trận $I_{\theta i}$ như trên. Từ các ma trận trực giao $I_{\theta i}[2 \times 2]$, ta tạo ra một ma trận trực giao A $[2n \times 2n]$ đường chéo là các ma trận $[2 \times 2]$, các vị trí còn lại trong ma trận A có giá trị là 0.

$$A = \begin{bmatrix} I_{\theta 1} & 0 & \cdots & 0 \\ 0 & I_{\theta 2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{\theta n} \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & \ldots & \ldots & 0 \\ -\sin\theta_1 & \cos\theta_1 & \ldots & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cos\theta_n & \sin\theta_n \\ 0 & 0 & \cdots & -\sin\theta_n & \cos\theta_n \end{bmatrix}$$

Tóm lại, hệ thống xác thực sử dụng đặc trưng sinh trắc áp dụng phép chiếu trực giao ngẫu nhiên để bảo vệ mẫu sinh trắc gốc, đồng thời đảm bảo tính cancelability cho mẫu sinh trắc. Sử dụng quy trình Hisham Al-Asam trong hệ thống giúp hệ thống xác thực đảm bảo được tính hiệu quả về mặt tính toán, bên cạnh đó nó còn cải thiện tính an toàn trong khi vẫn duy trì các biến thể nội lớp (intra-class variation). Cụ thể, khi người dùng thấy nghi ngờ mẫu sinh trắc trong cơ sở dữ liệu bị lộ, họ chỉ cần tạo ra một ma trận trực giao mới để có được một bản mẫu sinh trắc hoàn toàn mới. Ta tạm gọi kết quả của quá trình biến đổi này một mẫu sinh trắc khả đổi (cancelable template).

2.3.2 Cam kết mờ

Để hiểu lược đồ cam kết mờ trong xác thực sinh trắc để xác thực, bước đầu ta cần có kiến thức về mã sửa lỗi (ECC – Error Correcting Code). Một mã sửa lỗi ECC bao gồm một tập hợp codeword $\mathsf{C} \subseteq \{0,1\}^n$. Và cặp bộ phận Mã hóa – Giải mã (Encode – Decode). Cho $\mathcal{M} = \{0,1\}^k$ $(k < n)$ là không gian của thông điệp M cần truyền đi. Bộ phận Mã hóa (Encode) chứa một hàm $g : \mathcal{M} \to \mathsf{C}$ là hàm chuyển đổi (hay hàm mã hóa), biểu diễn một ánh xa từ tập thông điệp $\mathcal{M}$ sang tập codeword C, trước khi thông điệp đó được truyền đi trên kênh nhiễu. Bộ phận Giải mã (Decode) chứa một hàm giải mã $f : \{0,1\}^n \to \mathcal{M}$. Chú ý rằng hàm f không phải là hàm nghịch đảo từ C đến $\mathcal{M}$, mà nó chỉ ánh xạ từ một chuỗi bit ngẫu nhiên chiều dài *n* đến codeword gần nhất trong không gian C. Nếu f có khả năng sửa tới *t* bit lỗi, ta gọi f có khả năng sửa lỗi là *t*.

Trong lược đồ cam kết mờ, dữ liệu sinh trắc được xem như là thông điệp bị nhiễu. Trong suốt quá trình đăng kí, một người dùng cung cấp mẫu sinh trắc B cho máy chủ. Tiếp đó, máy chủ lựa chọn ngẫu nhiên một codeword c và tính toán khoảng cách giữa B và c thông qua hàm $\delta = B \oplus c$. Máy chủ sẽ lưu trữ cặp $(\delta, Hash(c))$ trong cơ sở dữ liệu. Trong quá trình xác thực, người dùng cung cấp dữ liệu sinh trắc B' cho máy chủ. Khih đó, máy chủ sẽ thực hiện phép tính $c' = B' \oplus \delta$, tiếp đó tiến hành giải mã (sửa lỗi) *c'*. Nếu B' gần giống B, hay số bit lỗi nằm trong khả năng sửa lỗi của ECC được áp dụng, việc sửa lỗi cho c' sẽ cho ra kết quả là c. Do đó khi so sánh $Hash(c)$ và $Hash(Decode(c'))$, nếu kết quả khớp thì người dùng được xác thực, nếu không hệ thống sẽ từ chối truy cập.

CHƯƠNG 3 TÍNH TOÁN AN TOÀN

3.1 Mã hóa đồng hình

Mã hóa đồng hình là một dạng mã hóa có khả năng thực hiện một số phép tính trên dữ liệu mã hóa, và sinh ra kết quả ở dạng mã hóa, do đó kết quả của quá trình này khi được giải mã sẽ chính bằng kết quả của quá trình tính toán trên dữ liệu gốc. Chính vì vậy, mã hóa này đảm bảo được tính bí mật của dữ liệu cần được xử lý. Đây là lý do quan trọng cho thấy tính chất đồng hình của mã hóa trở thành một tính năng đáng trông chờ trong hệ thống thông tin hiện đại. Nhiều hệ thống bảo mật ứng dụng mã hóa đồng hình như: hệ thống tính toán đám mây an toàn, hệ thống bảo mật bầu cử điện tử, một số hệ thống xác thực dùng đặc trưng sinh trắc, hệ thống PIR (Private Information Retrieval – rút trích thông tin riêng tư),… Có hai loại mã hóa đồng hình: mã hóa đồng hình bán phần và mã hóa đồng hình toàn phần.

- Mã hóa đồng hình bán phần (Partially Homomorphic Encryption – PHE) bao gồm các hệ thống mã hóa: RSA, ElGamal, Pallier, Goldwasser Micali, Benaloh, … Những hệ thống mã hóa đó cho phép tính toán đồng hình ở một toán tử nhất định trên bản mã hóa (các toán hạng có thể là toán cộng, nhân, hàm bậc hai,…).
- Mã hóa đồng hình toàn phần (Fully Homomorphic Encryption – FHE): thuật ngữ này được giới thiệu bởi Rivest, Adle-man và Dertouzos vào năm 1978. FHE cho phép thực hiện chuỗi không giới hạn các phép toán trên không gian các toán hạng được mã hóa. Tuy nhiên cho đến này, các nhà nghiên cứu vẫn chưa tìm được hệ thống mã hóa nào thỏa mãn yêu cầu của mã hóa đồng hình toàn phần.

3.2 Bộ xử lý bảo mật

Đề cập đến vấn đề riêng tư của dữ liệu người dùng được lưu trữ trong cơ sở dữ liệu của máy chủ, nhiều nhà nghiên cứu đã nghĩ đến việc sử dụng thành phần thứ

ba đáng tin cậy trong các công trình của họ. Trong công trình trước đây, Đặng đã đề xuất một giao thức đảm bảo các yêu cầu bảo mật của mô hình trong đó máy chủ được thuê ở ngoài (outsourcing model) bằng cách dựa vào một thành phần thứ ba đáng tin cậy (Trust Third Party- TTP). Việc sử dụng TTP thay đổi hoàn toàn mô hình trên và tiếp cận được vấn đề khó giải ở mô hình đó về việc bảo vệ tính riêng tư cho dữ liệu người dùng. Trong ngữ cảnh được sử dụng, thuật ngữ TTP có thể được hiểu là một giao thức bảo mật mà trong đó người dùng có thể tin tưởng hoàn toàn rằng tính riêng tư về dữ liệu của mình được đảm bảo hoàn toàn. Tuy nhiên, việc xây dựng một giao thức bảo mật hoàn hảo thỏa mãn được tất cả các yêu cầu bảo mật cho một hệ thống cụ thể là một quá trình lâu dài và vẫn luôn là một thử thách được bỏ ngỏ cho các nhà nghiên cứu về bảo mật. Trong tiến trình đó, hướng tiếp cận về thành phần thứ ba đáng tin cậy dựa trên phần cứng đang rất được quan tâm, nó được xem như một bộ xử lý bảo mật được cài đặt trên máy chủ, nhiệm vụ chính là tính toán những dữ liêu. Về mặt công nghệ, IBM đã phát triển các dòng sản phẩm phục vụ cho mục đích như vậy, và được gọi là Bộ xử lý bảo mật Secure Processor.

Bộ xử lý bảo mật (hay còn gọi là vi xử lý mã hóa – Cryptographic Processor) là một bộ phận phần cứng bảo mật được thiết kế để ngăn chặn việc lạm dụng dữ liệu và các thành phần khóa của người dùng. Lịch sử phát triển của nó bắt nguồn từ cỗ máy mã hóa của quân đội và các thành phần bảo mật nơi mã hóa những số bí mật (PINs) của người dùng trong hệ thống xác thực ATM của ngân hàng. Kể từ đó, chúng được sử dụng rộng rãi nhằm bảo vệ khóa SSL (Secure Socket Layer) khỏi các web server, và bảo vệ các phần mềm độc quyền và giải thuật chống lại việc trộm cắp thông tin từ người quản lý hệ thống, hoặc tạo ra thẻ thông minh, chip bảo mật,…Trong thời đại thương mại điện tử, vi xử lý bảo mật cho phép nhiều ứng dụng được chấp nhận bởi đông đảo người dùng nhờ việc thực thi được các chương trình xác thực và bảo vệ tính riêng tư dù bị tấn công cả về phía phần cứng. Việc sử dụng các kĩ thuật mã hóa là tối cần thiết trong các ứng dụng thương mại điện tử hiện đại. Nhiều ứng dụng áp dụng việc tính toán mã hóa theo nhiều cách khác nhau để đảm bảo tính riêng tư, bảo mật và tính toàn vẹn của dữ liệu.

Về bản chất kĩ thuật, bộ xử lý bảo mật là một bộ phận phần cứng được bảo vệ và chỉ có thể truy cập vào trạng thái bên trong thông qua cổng giao tiếp I/O của nó. Điều kiện đó cho phép thành phần đó có thể lưu trữ những dữ liệu nhạy cảm mà không có bất kì nguy cơ rò rỉ nào. Đôi khi nó được sử dụng như một bộ đồng xử lý (coprocecssor). Thuật ngữ này được dùng để chỉ một bộ xử lý phụ giúp cung cấp thêm một số hàm nâng cao cho bộ xử lý chính. Những hàm bổ sung này có thể là quá trình xử lý tín hiệu số, hình ảnh, chuỗi, hoặc mã hóa, hoặc thiết kế giao diện I/O cho các thiết bị ngoại vi… Mục đích chính của một bộ đồng xử lý là tăng tốc độ xử lý của hệ thống bằng cách giảm tải của bộ xử lý chính.

Bộ đồng xử lý bảo mật là một bộ phận phần cứng gồm: CPU, bootstrap ROM, vùng nhớ bảo mật không bay hơi (secure non-volatile memory). Bộ phận này được bảo vệ về mặt phần cứng khỏi sử xâm nhập, và các cổng giao tiếp tới bộ phận này là con đường duy nhất để truy xuất vào trạng thái bên trong của bộ phận. Nếu khiêng bảo vệ bị phá vỡ, bộ đồng xử lý bảo mật sẽ xóa tất cả bộ nhớ quan trọng. Cụ thể hơn, kẻ tấn công có thể phá lớp vỏ và thấy được cấu trúc bên trong của bộ đồng xử lý bảo mật. Tuy nhiên, hắn không thể thấy được trạng thái bên trong hay thay đổi bất kì dữ liệu gì trừ khi thông qua kênh trao đổi duy nhất của bộ đồng xử lý bảo mật.

Hiện nay, có rất nhiều nhà sản xuất đã và đang đầu tư rất nhiều thời gian và tiền bạc vào nghiên cứu và phát triển các bộ đồng xử lý bảo mật bằng mã hóa (secure crypto-processor). Trong đó IBM đã liên tiếp trình làng những dòng chip xử lý bảo mật. Sản phẩm đầu tiên trong gia đình chip đồng xử lý bảo mật của IBM là IBM 4758 PCI Cryptographic Coprocessor (PCICC), các dòng kế tiếp là IBM e-business PCI Cryptographic Accelerator (PCICA), IBM PCI-X Cryptographic Coprocessor (4764/CEX2C/PCIXCC), và IBM PCIe Cryptographic Coprocessor (4765/CEX4S/CEX3C) – PCIeCC. Chip mới nhất là IBM PCIe Cryptographic Coprocessor version 2 (CEX5S).

IBM 4765 PCIe Cryptographic Coprocessor là một vi xử lý có thể lập trình được. Nó được sử dụng trong các toán tử mã hóa tốc độ cao và tối mật trên những dữ

liệu nhạy cảm mà không thể tiết lộ với những môi trường dùng chung không an toàn. Trong kỉ nguyên thương mại điện tử, đó là một sản phẩm xuất sắc cho phép các giao thức thương mại điện tử được thực hiện an toàn và phù hợp với nhiều ứng dụng mã hóa như phát hành và xác thực số PIN, ứng dụng Hạ tầng khóa công khai (PKI), ứng dụng web server, ứng dụng thẻ thông minh,… Việc sử dụng dòng Common Cryptography Architecture (CCA) của IBM như một chương trình phần mềm hỗ trợ giúp bộ xử lý chính có thể thực hiện những giải thuật mã hóa công nghiệp thông thường như DES,T-DES, SHA, HMAC, RSA, ECC..... IBM còn cung cấp các phần mềm hỗ trợ, còn gọi API. Tùy thuộc vào mỗi phiên bản, những API có thể được mở rộng và thay thể để tích hợp với các tính năng mã hóa và những yêu cầu đặc biệt từ máy chủ. Gioogns như những bộ đồng xử lý bảo mật khác, IBM 4765 có khiêng bảo vệ, bộ cảm ứng và mạch điều khiển để bảo vệ chip khỏi nhiều kiểu tấn công khác nhau vào hệ thống, từ bên ngoài vào bên trong, từ tấn công phần mềm hay thậm chí tấn công phần cứng. Hơn thế nữa, nó có chứa cặp khóa bí mật/ công khai, lưu trữ ngay bên trong thiết bị. Cặp khóa này được sinh ra ngay từ khi sản xuất chip tại nhà máy của IBM và đã được chứng thực.

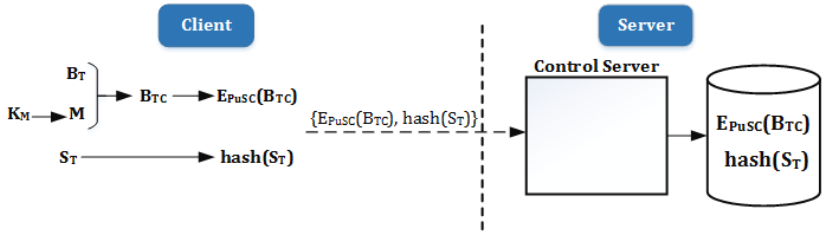
Trong đề tài này, để đảm bảo yêu cầu riêng tư của dữ liệu chống lại sự tò mò của máy chủ, tôi sử dụng một bộ phận vi xử lý bảo mật giả tưởng cài đặt trên máy chủ với chứng năng tương tự như bộ đồng xử lý bảo mật được phát hành bởi IBM. Tuy nhiên, thực tế các bộ đồng xử lý bảo mật nói chung đều bị hạn chế về khả năng tính toán và dung lượng bộ nhớ dẫn đến yêu cầu về việc giảm chi phí tính toán trong vi xử lý giả tưởng này được đưa về mức thấp nhất trong giao thức đề xuất trong đề tài này.

CHƯƠNG 4 GIAO THỨC ĐỀ XUẤT

4.1 Giao thức đề xuất

Các kí hiệu sử dụng trong giao thức:

- B_T, B: lần lượt là vector đặc trưng sinh trắc của người dùng trong quá trình đăng kí và xác thực.
- M: ma trận trực giao
- B_{TC}: mẫu sinh trắc đã được biến đổi được lưu trữ trong cơ sở dữ liệu.
- H(m): thông điệp m được mã hóa bởi hàm băm 1 chiều.
- BL: khóa sinh trắc (Biometric Lock) của người dùng.
- Pu & Pr: khóa công khai (public key) và khóa bí mật (private key).
- $E_{PuX}(m)$: thông điệp m được mã hóa bởi khóa công khai của X.
- K: khóa phiên xác thực được sinh ngẫu nhiên bởi người dùng.
- $E_K(m)$: mã hóa đối xứng thông điệp m sử dụng khóa bí mật *K*.
- S_T, S: lần lượt là số bí mật do người dùng cung cấp trong giai đoạn đăng kí và xác thực.
- C: người dùng.
- SC: bộ đồng xử lý bảo mật.
- CS: máy chủ điều khiển.
- PuSC & PrSC: tương ứng khóa công khai và khóa riêng phần của máy chủ điều khiển.



Hình 4.1: Giai đoạn đăng kí

4.1.1 Giai đoạn đăng ký

Trong giai đoạn đăng ký, người dùng sử dụng số ngẫu nhiên K_M (được lưu trữ trong thiết bị dùng để đăng ký) để sinh ra ma trận trực giao ngẫu nhiên *M*. Sau khi được trích xuất, vector đặc trưng sinh trắc B_T kết hợp với ma trận *M* để sinh ra một phiên bản mới có khả năng thay thế của B_T là B_{TC}. Tiếp đó B_{TC} được mã hóa bởi khóa công khai của *SC*. Bên cạnh đó, người dùng sẽ hash tham số S_T và có kết quả ($hash(S_T)$). Sau đó người dùng sẽ gởi gói tin có chứa $E_{PuSC}(B_{TC})$ và $hash(S_T)$ tới máy chủ. Nếu người dùng nghi ngờ rằng dữ liệu xác thực của anh ta bị lộ, anh ta có thể thay thế gói tin xác thực cũ bằng cách tạo ra một ma trận trực giao *M* khác thay vì phải thay đổi dữ liệu sinh trắc của bản thân anh ta. Người dùng tạo ra số ngẫu nhiên K_M mới để tạo ra ma trận *M* mới, từ đó sinh ra B_{TC} mới, mà không cần tạo ra bất kì thay đổi nào đối với vector đặc trưng gốc B_T . Trong giao thức đề xuất này, tôi sử dụng hệ thống xác thực bằng khuôn mặt để làm thí nghiệm. Trong đó, vector đặc trưng sinh trắc B_T được trích xuất bằng kỹ thuật Feature Extractor từ những hình ảnh khuôn mặt được lấy từ các cảm biến. Trong hầu hết các hệ thống xác thực người dùng bằng khuôn mặt, vector đặc trưng sinh trắc B_T có kích thước là $2n$. Do đó, ma trận trực giao *M* có kích thước là $2n \times 2n$. Và kích thước của số ngẫu nhiên K_M, được sử dụng để sinh ra ma trận trực giao *M*, là n.

4.1.2 Giai đoạn xác thực

Quá trình xác thực sẽ được thực hiện từng bước như sau:

Bước 1: Đầu tiên người dùng gởi yêu cầu tới máy chủ, máy chủ sau đó tạo ra số ngẫu nhiên N_A, sau đó nó được mã hóa bởi khóa công khai *PuC* của máy chủ.

Bước 2: Máy chủ gởi gói tin chứa thông tin *PuC* đã được mã hóa ($E_{PuC}(N_A)$) cho người dùng. Lưu ý là tất cả các thông tin truyền nhận giữa máy chủ và người dùng đều được bảo bệ bằng kỹ thuật mã hóa bất đối xứng (PKI- Public Key Infrastruture).

Bước 3: Ở phía người dùng, vector đặc trưng sinh trắc gốc, do cảm biến thu được từ người dùng trong giai đoạn xác thực, *B* sẽ được biến đổi bằng ma trận trực

giao *M* để có được phiên bản mới có thể thay thế (cancelable) B_C . Phiên bản biến đổi này được kết hợp với N_A (sau khi được giải mã) để sinh ra B_O. Bước này được thực hiện để đảm bảo mỗi lần người dùng gởi yêu cầu xác thực, một giá trị mới của B_O được sinh ra để tránh tấn công lặp lại. Sau đó B_O này, cùng với khóa xác thực *K* – yếu tố ngẫu nhiên do người dùng tạo ra, sẽ được đưa vào quá trình Fuzzy Commitment để sinh ra *BL*.

- ✓ *Bước 3.1:* Tiếp theo sau, *BL* được gởi tới máy chủ cho mục đích xác thực người dùng
- ✓ *Bước 3.*2: Bên cạnh đó, người dùng sẽ truy xuất số bí mật *S* (được lưu trữ trong thiết bị di động của người dùng). S được mã hóa bởi khóa *K,* sau đó được gởi tới máy chủ

Bước 4: Ở phía máy chủ, sau khi sinh ra N_A, máy chủ sẽ mã hóa N_A bằng khóa công khai *PuSC* của *SC*. Cần chú ý rằng tất cả những giải thuật mã hóa được áp dụng ở bước 4 phải đảm bảo tính chất mã hóa đồng hình. Máy chủ sẽ lấy thông tin $E_{PuSC}(B_{TC})$ từ cơ sở dữ liệu, sau đó áp dụng tính chất đồng hình của mã hóa được sử dụng để tính được $E_{PuSC}(B_{TO})$. B_{TO} là một phiên bản của mẫu đặc trưng sinh trắc và được tạo ra bằng sự kết hợp của B_{TC} và N_A.

Bước 5: CS gởi tới cho *SP* gói tin $E_{PuSC}(B_{TO})$

Bước 6: CS gởi tới *SP* các gói tin *BL* và $E_K(S)$

Sau khi nhận được các gói tiên từ *CS*, bộ đồng xử lý bảo mật *SP* được lập trình để theo sau các bước sau.

Bước 7: Sau khi nhận $E_{PuSC}(B_{TO})$ từ *CS, SP* sử dụng khóa riêng phần của nó để giải mã và có được B_{TO}.

Bước 8: Dữ liệu B_{TO} (từ bước 7) và *BL* (từ bước 6) được sử dụng để tái tạo lại khóa xác thực *K* áp dụng kĩ thuật Fuzzy Commitment;

Bước 9: SP sau đó sử dụng *K* để giải mã $E_K(S)$ (từ bước 6) và có được *S*;

Bước 10: SP băm *S* và có được kết quả *hash(S).*

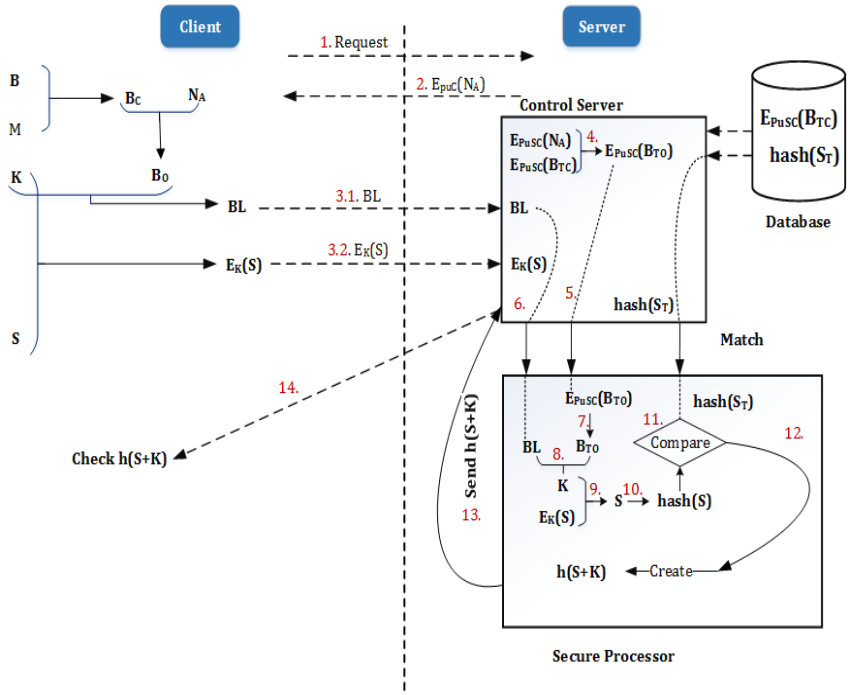
Bước 11: So sánh *hash(S)* vừa mới được sinh ra ở bước 10 và *hash(S_T)* được lấy từ cơ sở dữ liệu. Nếu kếu quả so sánh trùng khớp thì có nghĩa là dữ liệu đặc trưng sinh trắc (được người dùng cung cấp thông qua *BL)* sẽ trùng với mẫu đặc trưng sinh trắc B_{TO} (được người dùng gởi tới và lưu trữ trong cơ sở dữ liệu), và số xác thực bí mật *S* của người dùng cũng được thỏa. Kết quả của sự so sánh ở bước 11 có thể chứng minh được là người dùng được xác thực hay không. Nếu kết quả ở bước này là không khớp, CS sẽ báo kết quả lại người dùng và dừng quá trình xác thực tại đây. Nếu kết quả là khớp, quá trình tiếp tục thực hiện các bước sau.

Bước 12: Băm dữ liệu *S* và *K*, sinh ra *h(S+K)*

Bước 13: *SP* gởi cho *CS* gói tin *h(S+K).*

Chúng ta có thể thấy rằng toàn bộ dữ liệu nhạy cảm bao gồm dữ liệu đặc trưng sinh trắc gốc của người dùng và nhân tố bí mật *S* đều được tính toán thông qua *SP* và chỉ cho ra kết quả cuối cùng (việc tính toán này được bảo mật cao bởi *SP* và không bị can thiệp bởi các cuộc tấn công bên ngoài).

Bước 14: *CS* sẽ chuyển gói *h(S+K)* này đến người dùng. Bước này nhằm mục đích cho phép người dùng có khả năng xác thực ngược lại máy chủ. Người dùng sẽ so sánh giữa *h(S+K)* mới vừa nhận được từ máy chủ và *h(S+K)* do chính thiết bị người dùng tính toán (*S* lưu trữ trong thiết bị người dùng và *K* do người dùng tạo ra). Nếu hai chuỗi trùng với nhau, thì máy chủ đã được xác thực bởi người dùng. Người dùng có thể tin cậy máy chủ mà anh ta đang giao tiếp. Một khi việc xác thực lẫn nhau được hoàn thành thành công, *K* thì được sử dụng để bảo vệ giao tác giữa người dùng và máy chủ.



Hình 4.2: Quá trình xác thực.

4.2 Đánh giá

Hệ thống xác thực sử dụng đặc trưng sinh trắc được kiểm thử bởi cơ sở dữ liệu với 220 người, mỗi người có 20 ảnh biểu diễn các trạng thái khuôn mặt khác nhau. Tất cả các hình được sử dụng đều được chuẩn hóa, sau đó chiếu vào không gian Eigenface để rút trích được vector đặc trưng cho từng hình ảnh. Độ chính xác của lược đồ lai sử dụng trong hệ thống xác thực sinh trắc này được đánh giá thông qua 3 chỉ số: FAR (tỉ lệ chấp nhận sai), FRR (tỉ lệ từ chối sai), và EER (tỉ lệ lỗi cân bằng).

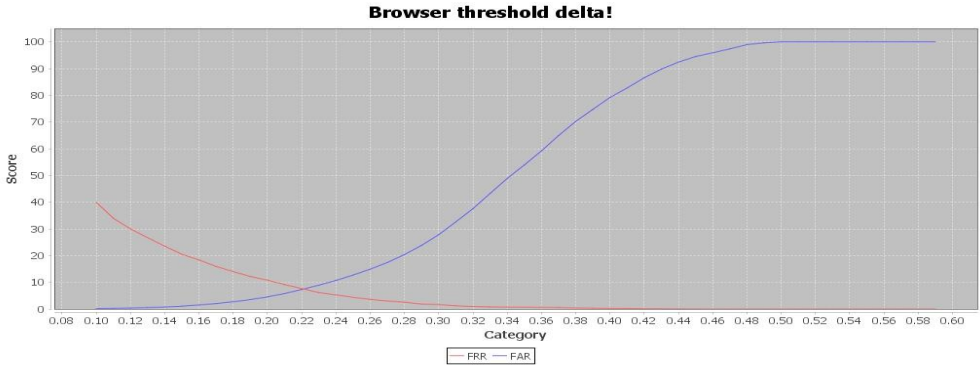
Để xác thực, ta tiến hành tính toán khoảng cách Euclide giữa vector đặc trưng thu được trong quá trình xác thực và vector đặc trưng của mẫu sinh trắc được lưu trữ trong quá trình đăng kí. Nếu kết quả này thỏa điều kiện nhỏ hơn một ngưỡng (threshold) nào đó, hệ thống sẽ cho phép truy cập. Như vậy, kết quả xác thực phụ thuộc vào việc xác định giá trị ngưỡng phù hợp. Để lựa chọn ngưỡng tốt nhất cho

một hệ thống xác thực bằng đặc trưng khuôn mặt như hệ thống này, cần thiết phải tính toán giá trị lỗi ở mỗi giá trị ngưỡng tương ứng. Định nghĩa các giá trị ngưỡng t được tính toán bởi công thức.

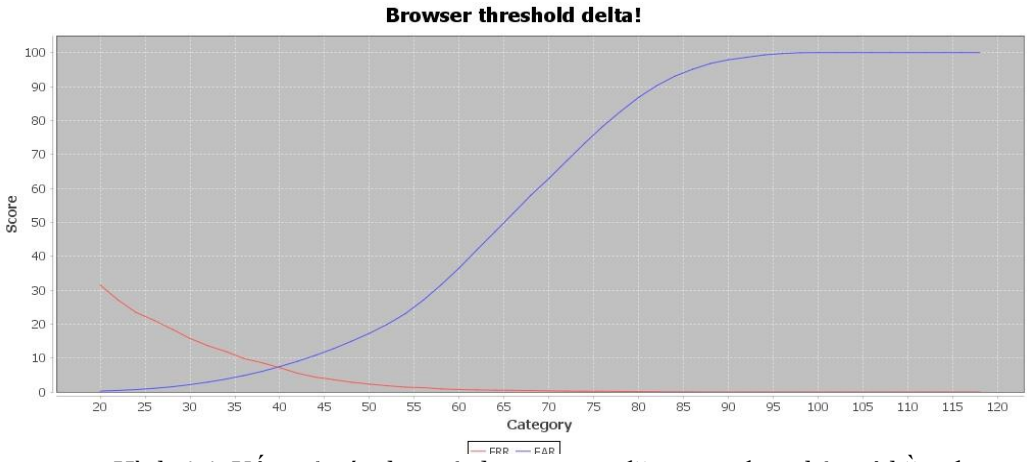
$$t_i = 0.1 + 0.01 \times i, \qquad với\ i \in \mathbb{N}, i\ \in [0, 49]$$

Với mỗi giá trị của ngưỡng, các chỉ số FAR và FRR sẽ được tính bằng phương pháp thống kê như sau:

- FAR: với mỗi người dùng, lấy bức ảnh đầu tiên so sánh với tất cả bức ảnh của 219 người còn lại trong tập kiểm thử. Điều này có nghĩa, không có bức ảnh nào của cùng người dùng đó được so sánh với nhau. Trong đó, mỗi người có 20 bức ảnh khác nhau. Vậy số lần thực hiện việc so sánh là $220 \times 219 \times 20 = 963600$ lần. Nếu kết quả so sánh là trùng khớp, thì hệ thống xác thực khuôn mặt của chúng ta đã mắc lỗi chấp nhận sai. Thống kê lại những trường hợp như vậy ta sẽ thu được tỉ lệ chấp nhận sai.
- FRR: với mỗi người dùng, lần lượt so sánh các bước ảnh của người dùng đó với nhau. Vậy số lần thực hiện so sánh ứng với một người dùng là C_2^{20}. Ta có 220 người dùng khác nhau thì tổng số lần thực hiện so sánh là $220 \times C_2^{20} = 41.800$ lần. Mỗi so sánh mà cho ra kết quả không trùng khớp, hệ thống xác thực khuôn mặt mắc lỗi từ chối sai.



Hình 4.3: Kết quả xác thực sử dụng vector đặc trưng gốc.



Hình 4.4: Kết quả xác thực sử dụng vector đặc trưng được bảo vệ bằng lược đồ lai giữa Fuzzy Commitment và Random Projection.

Các Hình 4.3, hình 4.4 cho thấy kết quả đánh giá độ chính xác của hệ thống thông qua chỉ số FAR và FRR trong hai trường hợp: các vector đặc trưng được so sánh trực tiếp không thông qua biện pháp bảo mật nào, và sử dụng lược đồ lai fuzzy commitment và random projection để bảo vệ vector đặc trưng gốc.

Kết quả của trường hợp đầu tiên thể hiện ở Hình 4.3 cho thấy FAR và FRR giao nhau ở ngưỡng $t \approx 0.22$. Tại đó, tỉ lệ lỗi là 7%.

Trong trường hợp cuối cùng, sử dụng mô hình lai kết hợp phép chiếu trực giao ngẫu nhiên và fuzzy commitment, được biểu diễn ở Hình 4.4, tại giao điểm của FAR và FRR tỉ lệ lỗi cũng là 7%. Hình này chứng tỏ rằng lược đồ lai được đề xuất cho kết quả khả quan với tỉ lệ xác thực chính xác vào khoảng 93%. Giá trị ERR không đổi. Điều đó cho thấy rằng hiệu suất nhận dạng của hệ thống áp dụng lược đồ lai này có thể cạnh tranh với các hệ thống nhận dạng khác không sử dụng kĩ thuật bảo vệ mẫu sinh trắc. Tóm lại, việc kết hợp random projection và fuzzy commitment để bảo vệ mẫu sinh trắc trong hệ thống xác thực từ xa có tính khả thi cao và có khả năng đưa vào thực tiễn.

CHƯƠNG 5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 Kết luận

Xác thực từ xa ngày càng đóng vai trò quan trọng trong việc triển khai các dịch vụ thương mại điện tử trong thực tiễn. Trong hoàn cảnh các ứng dụng, dịch vụ của các cơ quan, doanh nghiệp đang ngày càng phổ biến trên Internet, cần một nhu cầu là đảm bảo tính chính xác, hiệu quả và tin cậy trong xác thực người dùng. Trong khi các công trình nghiên cứu liên quan trước đó được đề xuất vẫn còn tồn đọng nhiều hạn chế đặc biệt trong việc đảm bảo an toàn cho đặc trưng sinh trắc người dùng lưu trữ trên hệ thống. Đề tài đã tiến hành nghiên cứu các cách thức xác thực từ xa, các kĩ thuật bảo vệ mẫu sinh trắc và yêu cầu bảo mật trong xác thực từ xa sử dụng đặc trưng sinh trắc để tiến hành tìm ra giải pháp, khắc phục những hạn chế đó. Về mặt khoa học, đề tài này sẽ thúc đẩy các nghiên cứu về ứng dụng sinh trắc học vào lĩnh vực bảo mật, nghiên cứu đưa mô hình xác thực từ xa vào trong các dịch vụ thương mại điện tử, các dịch vụ công. Thúc đẩy nghiên cứu để hoàn thiện các mô hình xác thực từ xa sử dụng đặc trưng sinh trắc của con người. Về mặt thực tiễn, qua nghiên cứu của để tài nếu được ứng dụng vào thực tiễn sẽ tăng cường mức độ tin cậy của người dùng về các dịch vụ trực tuyến, đẩy mạnh việc sử dụng các ứng dụng thương mại điện tử trên Internet.

Trong quá trình thực hiện, đề tài đã thực hiện được các công việc sau:

- Nghiên cứu lý thuyết về xác thực, xác thực từ xa sử dụng các phương pháp truyền thống và ưu nhược điểm của các hệ thống xác thực từ xa truyền thống.
- Nghiên cứu lý thuyết về sinh trắc, các yếu tố ảnh hưởng tới hệ thống sử dụng sinh trắc trong xác thực từ xa; cách rút trích đặc trưng sinh trắc của người dùng.
- Nghiên cứu về các cách tiếp cận để bảo vệ mẫu sinh trắc trong hệ thống xác thực nói chung. Trong đó tìm hiểu kĩ các kĩ thuật bảo vệ để phân tích

ưu và nhược điểm của từng kĩ thuật từ đó đưa ra giải pháp bảo vệ mẫu sinh trắc phù hợp.

- Nghiên cứu lý thuyết về cách thức bảo vệ dữ liệu nhạy cảm của người dùng khi được tính toán trong môi trường máy chủ không đáng tin. Phân tích các hướng tiếp cận và đề xuất hướng tiếp cận bảo mật trong phần cứng với bộ đồng xử lý bảo mật. Bên cạnh đó, tìm hiểu tổng quan một ví dụ thực tế của bộ đồng xử lý bảo mật (IBM 4765) về cách thức tương tác trong lập trình, các thư viện, các hàm cơ bản có thể được xử dụng trong quá trình xác thực từ xa.
- Nghiên cứu lý thuyết về các giao thức xác thực từ xa đã có, phân tích ưu nhược điểm của các giao thức, đề xuất giao thức dựa trên những phân tích có được.

Từ những công việc đã thực hiện, kết quả nghiên cứu đã giới thiệu một hệ thống xác thực từ xa sử dụng đặc trưng sinh trắc an toàn, bảo vệ được mẫu sinh trắc khi lưu trữ trong cơ sở dữ liệu của máy chủ hay được truyền trên mạng, đảm bảo an toàn tính toán cho các dữ liệu nhạy cảm khi được xử lý trên máy chủ không đáng tin cậy.

5.2 Hướng phát triển

Xác thực từ xa sử dụng đặc trưng sinh trắc học không phải là vấn đề mới, tuy nhiên những hệ thống hiện nay vẫn tiềm ẩn nhiều nguy cơ bị tấn công. Việc nhúng bộ đồng xử lý bảo mật vào bộ xử lý chính để chống lại hình thức tấn công xuất phát từ bên trong hệ thống vẫn đang còn khá mới, cần đi sâu nghiên cứu thêm về các khía cạnh có liên quan để có thể tăng hiệu quả về bảo mật cũng như giảm thời gian tính toán của hệ thống. Ngoài ra, để tăng cường tính bảo mật và cả độ hiệu quả nhận dạng cho hệ thống, một hướng nghiên cứu nữa cũng cần quan tâm đó là đồng xác thực. Đồng thời sử dụng chuyển đổi các đặc trưng sinh trắc với nhau cũng tăng sự tiện lợi cho người dùng và cả cho hệ thống phần cứng cung cấp nhiều dạng máy cảm biến khác nhau.