ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH
BỘ MÔN HỆ THỐNG THÔNG TIN
----------o0o----------

LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC

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

GVHD: Th.s Nguyễn Thị Ái Thảo
GVPB: Th.s Trương Quỳnh Chi
SVTH: Nguyễn Nam Tiệp
MSSV: 51103624

TP. HỒ CHÍ MINH, THÁNG 12 NĂM 2016

LỜI CAM KẾT

Báo cáo Luận văn tốt nghiệp của chúng tôi có tham khảo qua các tài liệu, kết quả nghiên cứu của một số tài liệu, trang web như đã trình bày ở phần "TÀI LIỆU THAM KHẢO". Chúng tôi xin cam doan ngoài những tài liệu trích dẫn trên, toàn bộ nội dung trong báo đề tài là do chúng tôi tự soạn thảo từ những kết quả nghiên cứu của chúng tôi, không sao chép bất kỳ tài liệu nào khác.

Chúng tôi hoàn toàn chịu trách nhiệm theo quy định nếu có bất kỳ sai phạm nào so với lời cam đoan. Nhóm sẽ chịu mọi trách nhiệm và xử phạt theo quy định trước Ban Chủ Nhiệm Khoa và Ban Giám Hiệu Nhà Trường.

TP.Hồ Chí Minh, Ngày Tháng Năm 2016

Nhóm sinh viên thực hiện đề tài

Nguyễn Nam Tiệp - 51103624

LỜI CẢM ƠN

Chúng tôi chân thành cảm ơn Khoa học và Kỹ thuật Máy Tính, trường đại học Bách Khoa Tp Hồ Chí Minh, đại học Quốc Gia Tp Hồ Chí Minh đã tạo điều kiện thuận lợi cho chúng tôi trong suốt quá trình học tập và thực hiện đề tài tốt nghiệp. chúng tôi xin chân thành cảm ơn quý thầy cô trong khoa Khoa học và Kỹ thuật Máy Tính đã tận tình giảng dạy, trang bị cho chúng tôi những kiến thức cần thiết trong suốt năm năm học qua.

Chúng tôi xin được gửi lời cảm ơnchân thành nhất đến cô Nguyễn Thị Ái Thảo là người đã hướng dẫn trực tiếp chúng tôi thực hiện đề tài. Cô cũng là người đã theo dõi, góp ý , sửa chữa những sai sót của nhóm trong quá trình thực hiện. Sau mời hai tuần thực hiện đề tài, bên cạnh nỗ lực của các cá nhân trong nhóm cùng với sự hỗ trợ nhiệt tình từ cô đã giúp nhóm chúng tôi rất nhiều trong việc bắt kịp tiến độ đã đề ra và hoàn thiện đề tài của mình một cách tốt nhất.

Trong quá trình thực hiện luận văn nhóm tôi đã cố gắn để hoàn thành đề tài với tất cả những nổ lực của bản thân, nhưng cũng không thể tránh khỏi những sai sót trong quá trình thực hiện đề tài. Vì vậy nhóm tôi rất mong nhận được sự thông cảm của quý thầy cô.

Cuối cùng, chúng tôi xin chân thành cảm ơn quý thầy cô và các bạn đã dành thời gian đọc tài liệu này.

Một lần nữa xin chân thành cảm ơn!

Tp. Hồ Chí Minh, ngày tháng năm .

Sinh viên

MỤC LỤC

Danh Sách Hình Vẽ

Danh Sách Bảng

TÓM TẮT LUẬN VĂN

Trong thời kỳ phát triển công nghệ như vũ bão hiện nay thì nó làm cho một số công nghệ cũ, đặc biệt là những công nghệ liên qua đến đến xác thực, bảo mật,... trở nên lạc hậu và xuất hiện những lỗ hổng lớn. Đặc biệt trong vài năm gần đây trên các trang thông tin đại chúng và trong giới công nghệ hện nay một chủ đề bàn tán sôi nổi nhất là việc các hacker dễ dàng có thể lấy được mật khẩu người dùng hoặc có thể lấy được mật khẩu trong thời gian khá ngắn, và để giải quyết vấn đề này thì các công ty lớn như Google, Yahoo, Facebook,... yêu cầu người dùng thường xuyên thay đổi pasword để bảo đảm an toàn cho account của họ, và mới đây trên truyền thông đại chúng có một tin cực kỳ gây sốc là công ty Facebook thường phải bỏ một số tiền lớn để mua lại password của người dùng đã bị hack ở chợ đen(black market) link. Như vậy có thể nói với công nghệ xác thực như hiện nay là không an toàn. Chính vì vậy vấn đền cấp bách nhất hiện nay là tìm ra một phương pháp mới thay thế cho công nghệ cũ, và một bước tiến lớn trong công nghệ hiện nay là xác thực bằng sinh trắc(những đặc điểm riêng biệt của từng người).

Một ưu điểm nổi bật của công nghệ này là khó đoán, khó giả mạo, khó tấn công và đặc biệt là người dùng không cần nhớ chúng. Với phương pháp cũ thì với sức mạnh công nghệ hiện nay việc tấn công bằng brute-force, hay tấn công bằng dictionary pasword,... không còn tốn nhiều thời gian nhưng khi sử dụng đặc trưng sinh trắc thì các phương pháp truyền thống gần như không còn tác dụng nữa vì không gian của sinh trắc của người dùng là gần như không có giới hạn và không có quy luật để có thể đoán được.

Chính những điểm mạnh như trên mà trong luận văn này chúng tôi trình bài một mô hình xác thực bằng sinh trắc. Nhưng trong trong mô hình có một số cải tiến đó là sự kết hợp của hai phương pháp fuzzy-commitmet và non-invertible transfomation để tăng khả năng bảo vệ mẫu sinh trắc, và bảo vệ trên đường truyền bằng đặc trưng sinh trắc. Và với phương pháp này thì FAR , FRR và cũng là ERR của hệ thống đạt 9% và kết quả này test trên một số lượng người khá lớn và thời gian chạy của hệ thống thì khoảng 5s tuy nhiên phương pháp PCA(sinh ra vector sinh trắc) chiếm hơn 4s, và phần còn lại của hệ thống chạy dưới 1s nên hệ thống này đạt hiệu suất khá tốt.

CHƯƠNG 1. GIỚI THIỆU

1.1. Giới thiệu đề tài

Sự phát triển của internet trong những năm gần đây là một cải cách công nghệ vượt bật, nó đã đem lại tiện ích không tưởng trong đời sống, đem toàn bộ thế giới đến ngôi nhà của mỗi người, kết nối mọi người với nhau,.... và đặc biệt trong công việc thương mại, kinh doanh, có thể nói ngoài lửa và bánh xe ra thì internet là một trong những phát minh quan trọng trong lịch sử loài người. Bây giờ gần như tất cả các giao dịch của người dùng đều thực hiện qua intener như chuyển tiền, mua hàng,... Tuy nhiên ngoài những điểm mạnh như đơn giản, tiện lợi và nhanh chóng thì nó cũng kèm theo rất nhiều vấn đề bảo mật và xác thực, nếu một hệ thống mà chức năng bảo mật hay xác thực người dùng không tốt thì rất dễ bị các kiểu tấn công như giả mạo định danh, ăn cắp thông tin, lừa đảo,... gây thiệt hại lớn cho người dùng và doanh nghiệp. Chính vì vậy nên yêu cầu xây dựng một hệ thống bảo mật thông tin, đảm bảo an toàn giao tiếp giữa những người dùng, có định danh và chống phủ nhận, đảm bảo riêng tư của người dùng là quan trọng nhất.

Chục năm qua chúng ta chứng kiến cuộc đua song mã giữa các hacker và các nhà bảo mật mạng, có khá nhiều ý tưởng, giải pháp để cải tiến hệ thống cũ tuy nhiên kết quả vẫn không khả quan và với tình hình đó có rất nhiều nhà nghiên cứu đã và đang tìm ra những giải pháp thay thế hiệu quả hơn so với phương pháp cũ. Hiện nay một phương pháp mà các nhà phát triển đang hướng tới đó là sử dụng đặc trưng sinh trắc của người dùng vào những hệ thống xác thực và bảo mật. Trên thực tế có khá nhiều sản phẩm đã đưa vào sử dụng như: thẻ ngân hàng sinh trắc, thẻ mua hàng, thẻ an ninh,.... Tuy nhiên những hệ thống này vẫn còn hạng chế và khó có thể áp dụng trên quy mô lớn, hơn thế nữa việc nghiên cứu liên quan đến con người luôn là vấn đề nhạy cảm có đặc thù trên từng quốc gia. Bởi vậy giải pháp này luôn được đặc biệt quan tâm và đang phát triển.

Một nguyên nhân quan trọng để chúng ta phải thay đổi hệ thống cũ là vấn đề sử dụng password, nếu password quá ngắn hoặc dễ thì kẻ gian rất dễ dò được và giả mạo người dùng, còn nếu nó dài và khó đoán thì người dùng rất khó để nhớ, còn chưa kể đến chuyện hiện nay với sự phát triển vượt bật về phần cứng máy tính và giá thành rẻ đi khiến cho việc thiết lập một máy tính với cấu hình cao để tấn công password không còn là chuyện xa vời, và với những password dài và khó thì nó có thể dò được trong khoảng thời gian vài ngày hoặc vài tuần.

Cho nên đã đến lúc chúng ta cần tìm một hệ thống mới thay thế. Và đó là mục đích mà trong luận văn này hướng đến. Trong bài này sẽ xây tìm hiểu và nghiên cứu một mô hình xác thực từ xa sử dụng đặc trưng sinh trắc, một điểm mạnh của việc sử dụng đặc trưng sinh trắc là người dùng không cần phải nhớ bất kỳ thứ gì, và thông tin sinh trắc này sẽ được biến đổi và bảo vệ đảm bảo tính riêng tư cho người dùng. Ngoài ra hệ thống sẽ đảm bảo tính an toàn trước các loại tấn công, cho dù hacker có lấy được dữ liệu hay lấy được thiết bị cũng không thể khai thác được thông tin gì. Chính vì vậy nhóm đã quyết định chọn đề tài để tìm hiểu và nghiên cứu là "**Bảo vệ mẫu sinh trắc trong xác thực từ xa**"

Ngoài ra còn có một số nguyên nhân cá nhân đó là sự hứng thú từ những đề tài liên quan đến hệ thống sinh trắc và bộ môn Hệ Thống Thông Tin. Với những hệ thống sinh trắc đặc biệt thì với hệ thống như thế này là một hệ thống xác thực giống như một con người. Đăng nhập vào một hệ thống giống như đi vào một căn nhà, muốn đi vào căn nhà thì người chủ phải đảm bảo họ biết người đó và phải tin tưởng người đó, và sẽ có rất nhiều cách để nhận diện đối phương, nhưng cách thông thường của một con người là nhận diện giọng nói, khuôn mặt,... , vậy sẽ rất thú vị nếu chúng ta có một hệ thống mà giống như người canh của vậy, thay vì đưa một đoạn mã(password) chúng ta cung cấp cho hệ thống đó hình ảnh khuôn mặt, giọng nói, vân tay,... để hệ thống đó xác thực chúng ta.

Sau khi tìm hiểu và được sự hướng dẫn nhiệt tình của giáo viên hướng dẫn nhóm cũng đã đạt được mục tiêu đề ra. Nhưng do thời gian hạn hẹp và kiến thức ở nhiều lĩnh vực còn rộng và sâu nên có thể trong báo cáo có chút sai sót, mong thầy cô và các bạn thông cảm, khính mong nhận được sự đóng góp của các bạn và thầy cô.

1.2. Mục tiêu đề tài

Nhiệm vụ: luận văn này trình bày cho người đọc những kiến thức cơ bản và chuyên sâu về một hệ thống xác thực từ xa có bảo vệ đặc trưng sinh trắc và bảo vệ trên đường truyền nhờ vào đặc trưng sinh trắc, xây dựng một hệ thống hoàn chỉnh và phân tích những đặc điểm của hệ thống đó.

Nội dung: đề tài sẽ tập trung xây dựng một hệ thống bảo vệ mẫu sinh trắc trong xác thực từ xa sử dụng đặc trưng sinh trắc. Có thể tóm tắt những nội dung cần làm trong luận văn này là:

- Hiện thực giao thức đã được đề xuất.

- Đọc hiểu hệ thống xác thực sinh trắc.
- Đọc hiểu kỹ thuật bảo vệ mẫu sinh trắc: feature transform và fuzzy commitment.
- Hiện thực lại hai kỹ thuật trên.
- Nhúng hai kỹ thuật trên vào hệ thống xác thực từ xa sử dụng đặc trưng sinh trắc.
- Đánh giá hệ thống.

Những lý thuyết phải tìm hiểu:

- Những kỹ thuật để rút trích vector đặc trưng.
- Fuzzy commitment và các loại kỹ thuật sữa lỗi phù hợp.
- Kỹ thuật Non-invertible transformation và các phép toán sử dụng trên ma trận.
- Các thành phần trong một hệ thống xác thực từ xa.
- Các kỹ thuật tấn công hiện nay và cách phòng chống.

1.3. Phạm vi đề tài

Hiện tại, phạm vi nghiên cứu của đề tài phần lớn tập trung vào việc xây dựng một hệ thống bảo vệ đặc trưng sinh trắc trong xác thực từ xa có sử dụng đặc trưng sinh trắc. Đề tài sẽ không quá tập trung vào các vấn đề nghiệp vụ của hệ thống, và một số thành phần khác không liên quan đến đề tài như mã hóa, hash,..., và cũng không cố gắn xây dựng một hệ thống hoàn chỉnh vì đây là quá trình lâu dài cần nhiều kỹ thuật và kinh nghiệm tốt trong việc thiết kế và lập trình.

Hơn nữa, đây chỉ là một đề tài nghiên cứu nhỏ nên sẽ không cố gắn làm một hệ thống hoàn chỉnh đưa vào ứng dụng, mà chỉ dừng lại ở việc tạo một demo nho nhỏ để thuyết trình.

1.4. Cấu trúc báo cáo

Bảng 1. Cấu trúc báo cáo.

Chương	Nội dung
Chương 1: Giới Thiệu.	Giới thiệu mục tiêu, động lực để làm luận văn.
Chương 2: Cơ Sở Lý Thuyết.	Nêu cơ sở lý thuyết cần thiết để áp dụng vào hệ thống, các vấn đề, hưởng giải quyết,...
Chương 3:Phương pháp rút trích đặc trưng	Phương pháp rút trích đặc trưng khuôn mặt từ ảnh.

Chương 4: Mô Hình Đề Xuất.	Dựa vào những cơ sở lý thuyết trên, đưa ra mô hình một hệ thống hoàn chỉnh.
Chương 5: Hiện Thực Hệ Thống.	Hiện thực một hệ thống hoàng chỉnh bằng ngôn ngữ lập trình phù hợp.
Chương 6: Đánh Giá Và Thử Nghiệm.	Đánh giá và phân tích độ an toàn của hệ thống, và làm những thử nghiệm để đưa ra nhận xét về hệ thống.
Chương 7: Kết Luận và Hướng Phát Triển.	Kết luận những điểm mạnh, điểm yếu của hệ thống, mục tiêu phát triển tiếp theo và thành quả đạt được.
Chương 8: Tài Liệu Tham Khảo.	Tài liệu tham khảo
Chương 9: Phụ Lục.	Phụ lục

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

2.1. Sinh trắc học là gì

2.1.1. Định nghĩa

Sinh trắc học là gì?: sinh trắc học[5](Biometrics) là công nghệ sử dụng những thuộc tính vật lý, đặc điểm sinh học riêng của mỗi cá nhân như vân tay, mống mắt, khuôn mặt... để nhận diện. Có hai loại sinh trắc chính là: vật lý và hành vi. Các sinh trắc vật lý thường liên quan đến khuôn mặt, vân tay, giọng nói,... Các sinh trắc hành vi thường liên quan đến hành vi con người như: chữ kí, dáng đi,.... Các phương thức sinh trắc phổ biến hiện nay là:

- **Khuôn mặt:** những hình ảnh tĩnh hoặc động của khuôn mặt được dùng để nhận dạng. Phương pháp trích xuất dựa trên vị trí, hình dạng và những mối quan hệ giữa các đặc điểm trên khuôn mặt như mắt, mũi, môi, cằm,... Tuy nhiên quá trình nhận dạng dựa trên khuôn mặt cũng gặp nhiều khó khăn do khuôn mặt sẽ thay đổi theo thời gian.
- **Dấu vân tay:** đây là đặc trưng sinh trắc được dùng phổ biến nhất hiện nay do độ tin cậy của nó vì vân tay của một người không bao giờ thay đổi. Phương pháp trích xuất đặc trưng của dấu vân tay thì tùy vào phương pháp như đa số là dựa trên những đặc điểm của các nếp gấp, hình dạng của các đường vân,.... Hiệu suất của các hệ thống sinh trắc sử dụng dấu vân tay thường cao và chính xác.

- **Tròng mắt:** là màng nhỏ có màu hình tròn bao bọc con ngươi đủ phức tạp để có ích trong việc nhận dạng. Hiệu suất hệ thống sử dụng phương pháp này đầy triển vọng, nhưng FMR và FNMR cao, và tròng mắt thay đổi theo thời gian, và các loại cảm biến để lấy tròng mắt còn bị nhiễu rất nhiều khi lấy.
- **Giọng nói:** giọng nói kết hợp trực tiếp các đặc tính sinh trắc và hành vi. Âm thanh con người nói dựa trên nhiều yếu tố vật lý của cơ thể(miệng, mũi, môi, thanh quản,...) và bị tác động bởi tuổi tác, cảm xúc, ngôn ngữ, cộng đồng xung quanh và sức khỏe. Chính vì vậy nên giọng nói còn đang được nghiên cứu và độ tin cậy chưa cao, chưa đủ khả năng để đưa vào ứng dụng thực tế.
- **Chữ kí:** đây là kiểu sinh trắc hành vi, vì chúng ta có thể thay đổi chữ ký được, nó phụ thuộc vào ý muốn con người nhiều hơn. Khi dùng chữ ký một vấn đề cực kỳ quan trọng là việc có thể giả được chữ ký. Chính vì vậy các hệ thống thông tin còn chưa tin dùng loại sinh trắc này.

2.1.2. Đặc tính

Để có thể áp dụng đặc trưng sinh trắc vào hệ thống thì cần phải có các tính chất sau:

- **Tính rộng rãi:** đa số mọi người đều có đặc trưng này.
- **Tính phân biệt:** đặc trưng sinh trắc giữa hai người bất kỳ phải khác nhau, đảm bảo sự tồn tại duy nhất của chủ thể.
- **Tính ổn định:** đặc trưng phải ổn định(không thay đổi) trong thời gian dài.
- **Tính dễ thu thập:** dễ dàng thu nhận bởi các loại cảm biến trong hệ thống.
- **Tính hiệu quả:** việc xác thực phải chính xác, nhanh chóng và tài nguyên cần sử dụng cho hệ thống phải ở mức chấp nhận được.
- **Tính chấp nhận được:** quá trình thu thập mẫu sinh trắc phải được người dùng đồng ý.
- **Chống giả mạo:** khó thực hiện việc giả mạo người dùng, như làm ngón tay giả, giả giọng nói,...

Hình 1: Các đặc điểm sinh trắc phổ biến.

Nhiều đặc trưng sinh trắc đã được đưa vào sử dụng. Mỗi loại có một điểm mạnh, điểm yếu khác nhau. Không có một đặc trưng nào thỏa mãn tất cả các yêu cầu như trên, có nghĩa là không có đặc trưng nào là tối ưu hoàn toàn, dưới đây là bảng biểu diễn đặc điểm của những loại đặc trưng.

Đặc trưng sinh trắc	Tính rộng rãi	Tính phân biệt	Tính ổn định	Tính dễ thu thập	Tính hiệu quả	Tính chấp nhận	Chống giả mạo
Vân bàn tay	M	M	M	M	M	M	L
Dạng hình học bàn tay	M	M	M	H	M	M	M
Vân tay	M	H	H	M	H	M	M
Dáng đi	M	L	L	H	L	H	M
Khuôn mặt	H	L	M	H	L	H	H
Võng mạc	H	H	M	L	H	L	L
Mống mắt	H	H	H	M	H	L	L
Chỉ tay	M	H	H	M	H	M	M
Giọng nói	M	L	L	M	L	H	H

Hình 2. Đặc điểm sinh trắc của từng loại.

2.2. Hệ thống xác thực

Hệ thống xác thực là hệ thống thực hiện đối sánh 1-1 giữa mẫu sinh trắc thu nhận được(Biometric sample) với mẫu dạng sinh trắc(biometric template) đã có trong hệ thống từ trước. Kết quả trả lời câu hỏi mẫu sinh trắc thu nhận được có liên quan tới mẫu dạng sinh trắc hay không, thông thường trong hệ thẩm định kết hợp với thông tin định danh chủ thể thực hiện chức năng xác thực thẩm định sinh trắc(Authentication). Trong hệ xác thực thẩm định đòi hỏi cao về độ chính xác để kết quả trả lời câu hỏi "sinh trắc sống thu nhận được(biometric sample) có phải là sinh trắc của chủ thể đã lưu trong hệ thống không?"

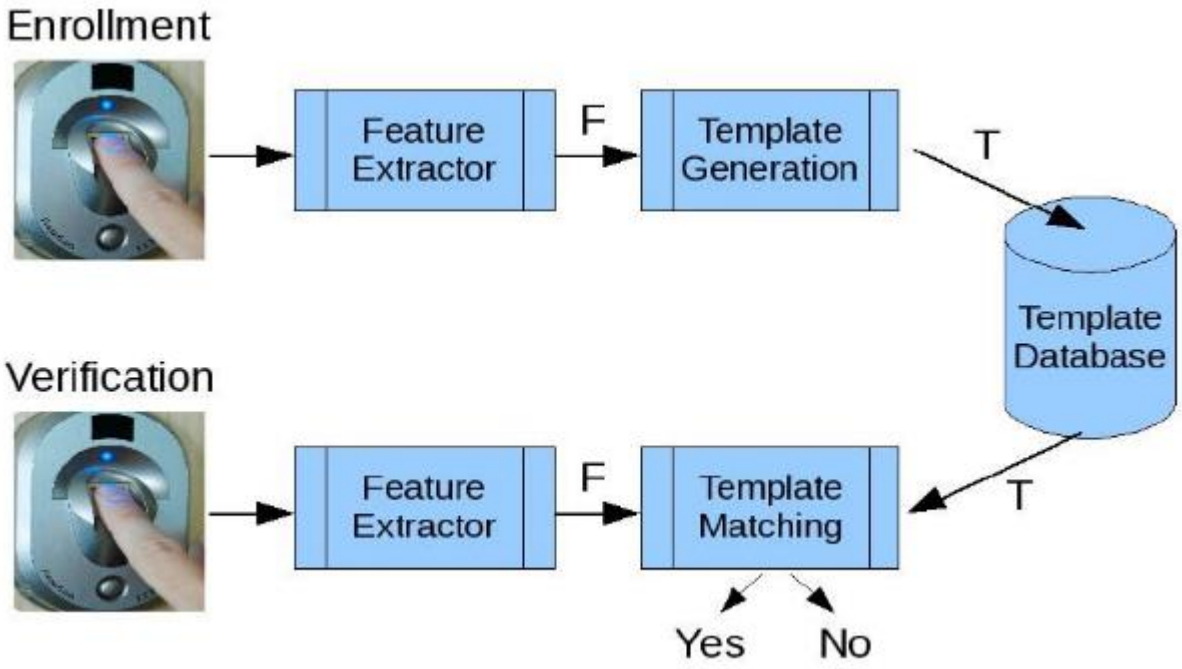
2.2.1. Thành phần

- **Thu nhận(Sensor, Capture):** thu nhận sinh trắc của con người và biểu diễn dưới dạng số hóa.
- **Trích xuất đặc trưng(Feature Extractor):** thực hiện các phép xử lý,phân tích và trích chọn đặc trưng từ mẫu sinh trắc.
- **Đối chiếu(Matching):** là thành phần chức năng thực hiện so sánh các đặc trưng vừa trích chọn với khuôn mẫu sinh trắc đã có trước đó.
- **Quyết định(Decision):** dựa trên kết quả đối sánh để trả lời câu hỏi đúng hay sai.

2.2.2. Cấu tạo hệ thống

Một hệ thống sinh trắc gồm có hai chức năng chính sau:

- **Ghi danh(Enrrollment):** giai đoạn cung cấp sinh trắc cho hệ thống sử lý và lưu lại.
- **Xác thực(Authentication):** giai đoạn thẩm định, nhận dạng mẫu sinh trắc của người đăng nhập với mẫu sinh trắc lưu trong cơ sở dữ liệu.



Hình 3. Hai chức năng chính của hệ thống sinh trắc.

2.2.3. Đặc điểm

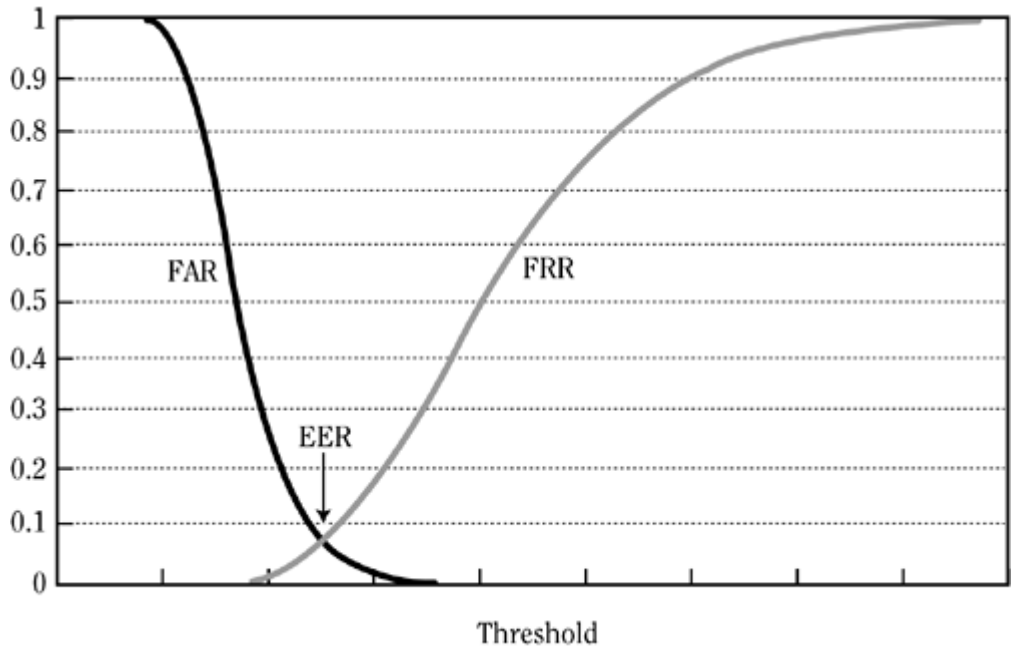
- Tính ổn định không cao: có hai nguyên nhân dẫn đến việc lấy hai mẫu sinh trắc khác nhau tuy cùng một người là do nhiễu từ môi trường và đặc tính sinh lý của con người bị thay đổi. Về nhiễu môi trường thì có thể do ánh sáng, môi trường, do cảm biến,..... Còn về đặc tính sinh lý của con người thì có thể do tuổi tác, sức khỏe, và các tác động khác lên con người làm họ thay đổi sinh lý.
- Tài nguyên cần thiết để thiết lập cho hệ thống là khá đắc đỏ: như chúng ta đã biết so với hệ thống cũ thì hệ thống sinh trắc cần rất nhiều thiết bị đi kèm để trích xuất đặc trưng người dùng, và cũng cần phải dùng những giải thuật đặc biệt để tính toán, và một số yêu cầu phần cứng để có thể đảm bảo hiệu suất cho ứng dụng,.. đây chính là trở ngại cực lớn trong việc thiết kế cho những hệ thống có số lượng người dùng lớn.

2.2.4. Cách đánh giá

Bất kỳ một hệ thống xác thực nào cũng dùng các chỉ số sau để đánh giá tính hiệu quả của nó:

- FAR(False Accept Ratio): tỷ số chấp nhận sai cho biết tỉ lệ trả lời đúng đối với dữ liệu vào là sai.

- FRR(False Rejection Ratio): tỷ số từ chối sai cho biết tỉ lệ trả lời sai đối với dữ liệu vào là đúng.
- Hai đại lượng này sẽ nghịch nhau: nếu FAR tăng thì FRR giảm và ngược lại. Tùy vào hệ thống mà chúng ta chấp nhận những giá trị này sao cho phù hợp. Thông thường người ta sử dụng giá trị này để tìm Threshold cho hệ thống. Như hình vẽ sau:



Hình 4. Đồ thị FRR và FAR.

Thường thì người ta lấy giao điểm để lấy Threshold cho hệ thống thực, tại điểm này FRR = FAR, và cũng là EER(Equal Error Ration), và người ta sẽ lấy giá trị Threshold cho hệ thống.

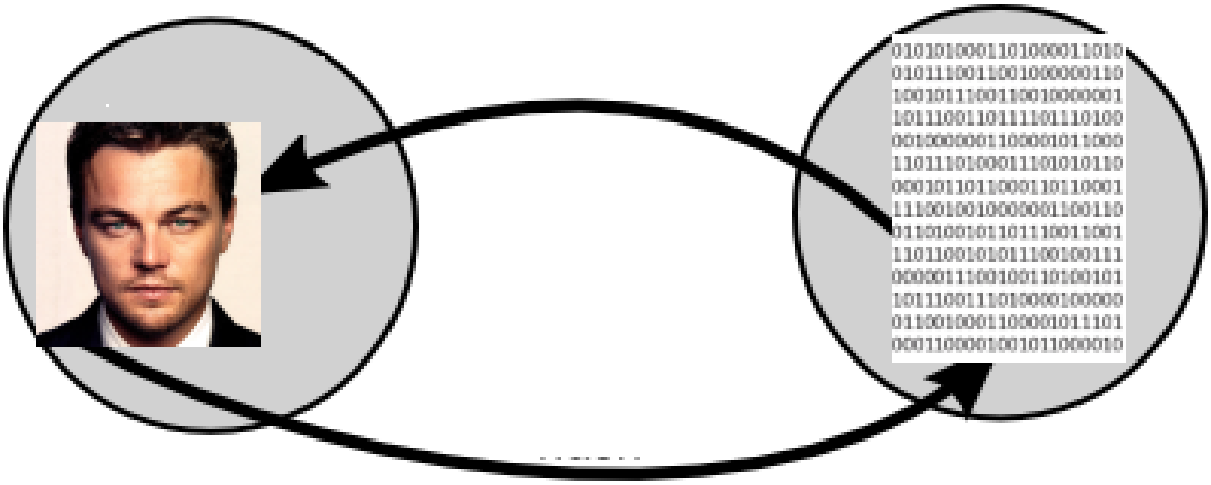
2.2.5. Hướng phát triển

Với sự đa dạng của loại sinh trắc trên cơ thể người mà có các hệ thống khác nhau tùy thuộc vào đặc tính của từng loại. Hiện nay gần như tất cả sinh lý của con người đều có giải thuật rút trích khá hiệu quả nên vấn đề mà nhiều nhà nghiên cứu là tập trung vào việc thiết lập hệ thống để đưa vào ứng dụng thực tiễn, không giống với hệ thống sử dụng mật khẩu thông thường, hệ thống sinh trắc sẽ sinh ra một chuỗi thông tin(vector, matrix,...) và thông tin này đa dạng và cách sử dụng khác nhau chính vì vậy nên với mỗi loại sinh trắc khác nhau thì có thể chỉ áp dụng được một số hệ thống nhất định. Chính vì vậy hướng phát triển các hệ thộng sử dụng sinh trắc là rất đa dạng và phát triển không ngừng.

2.3. Các loại tấn công trong hệ thống

Trong luận văn này tôi chỉ liệt kê những kiểu tấn công thông thường của một hệ thống, tôi sẽ không phân tích những lỗi của người dùng hay lỗi của phần cứng, hay những lỗi quá chuyên sâu bên hệ thống.... Sau đây là những kiểu tấn công một hệ thống.

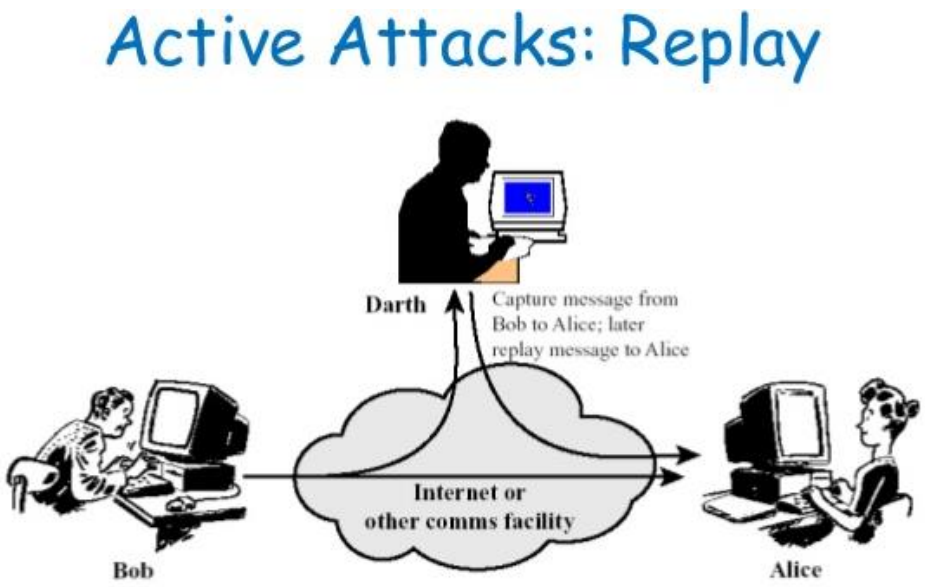
2.3.1. Biometric template attack



Hình 5. Biometric Template Attack.

Bất kỳ một hệ thống nào cũng cần phải lưu thông tin về sinh trắc của người dùng để sau này có thể truy suất và thực hiệp các phép so sánh đối chiếu để đưa ra quyết định xem đó có phải là người dùng hay không. Với kiểu tấn công này rõ ràng chúng ta cần phải thiết kế một hệ thống mà lưu sinh trắc đã được biến đổi hay đã được bảo vệ để kẻ khác khó có thể lấy và có thể truy ngược ra lại đặc trưng của người dùng. Một khó khăn trong hệ thống này là sinh trắc của con người gần như không thai đổi, chính vì vậy cần phải thiết kế một hệ thống mà việc thay đổi giá trị lưu trên server một cách dễ dàng mà không cần thay đổi sinh trắc người dùng.

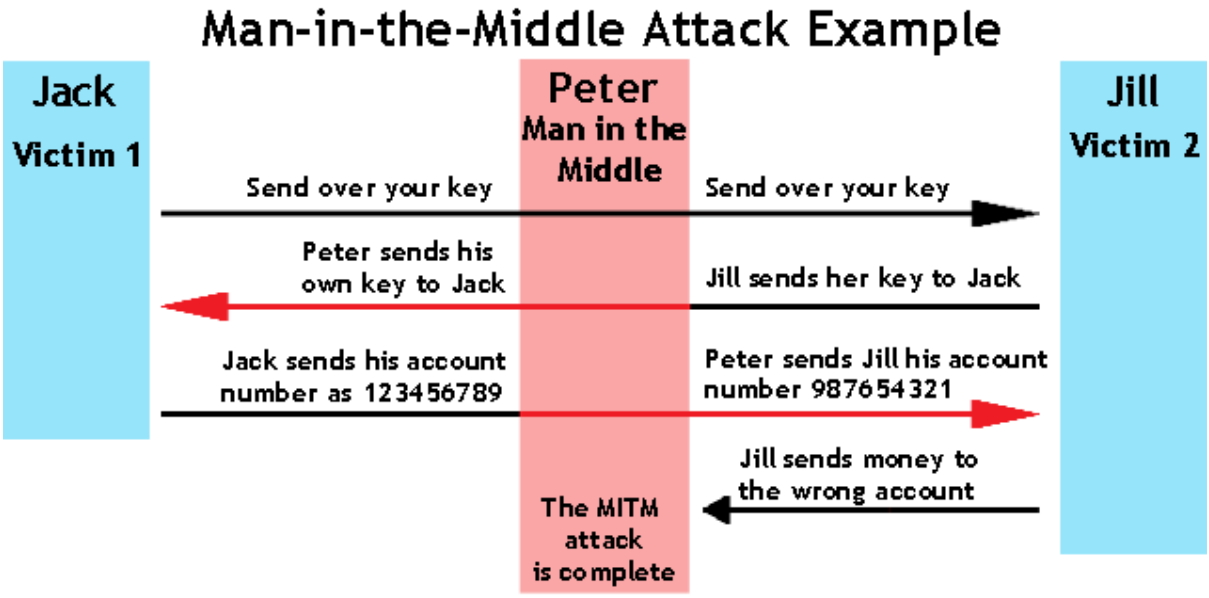
2.3.2. Replay Attack



Hình 6. Replay Attack.

Đây là kiểu tấn công mà kẻ xấu sử dụng lại dữ liệu cũ của lần đăng nhập trước đó giữa client và server để giả dạng client hay server và để đánh cắp thông tin. Loại tấn công này thường xảy ra với những hệ thống cũ, khi dữ liệu truyền giữa server và máy chủ không có phương pháp bảo vệ phù hợp và giống nhau trong các lần đăng nhập.

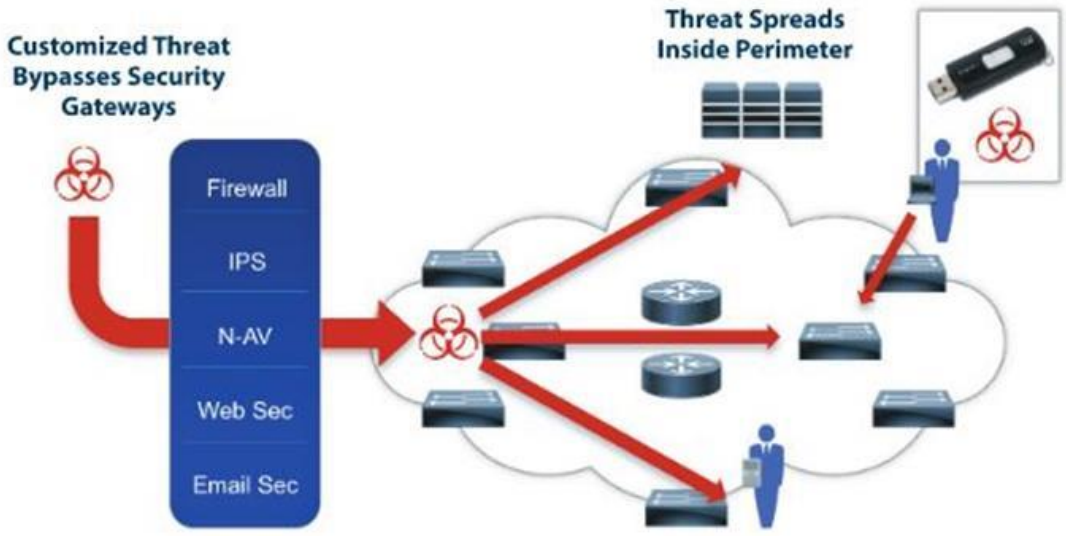
2.3.3. Man-in-midle attack



Hình 7. Man-in-the-middle Attack.

Đây là kiểu tấn công mà kẻ xấu sẽ ở giữa server và client để bắt các gói truyền dữ liệu của hai bên, sau đó đồng thời giả dạng cả hai bên, lúc đó client sẽ tưởng mình đang kết nối với máy chủ thật, còn server thì nghĩ mình đang kết nối với client thật, nhưng thật ra toàn bộ dữ liệu truyền giữa client và server đã bị thay dổi và kẻ xấu có thể gây hại cho cả hai bên.

2.3.4. Insider attack



Hình 8. Insider Attack.

Đây là kiểu tấn công ít gặp nhất trong các loại tấn công nhưng nếu nó xảy ra thì thường hậu quả cực kỳ nghiêm trọng. Kiểu tấn công này từ chính những người quản lý Server, những loại tấn công thường xảy ra như:

- Đánh cắp thông tin người dùng để giả dạng người dùng. Ví dụ đánh cắp password và username để đăng nhập thay cho người dùng.
- Tấn công vào module hệ thống, gây ra sai sót khi hệ thống đang chạy. Ví dụ như thay đổi password của người dùng để đăng nhập vào hệ thống, hay thay đổi module Matching của hệ thống gây ra sai sót.....
- Và còn rất nhiều loại tấn công khác vì kẻ tấn công lại là kẻ có nhiều quyền nhất trong hệ thống.

2.4. Các phương pháp bảo vệ mẫu sinh trắc

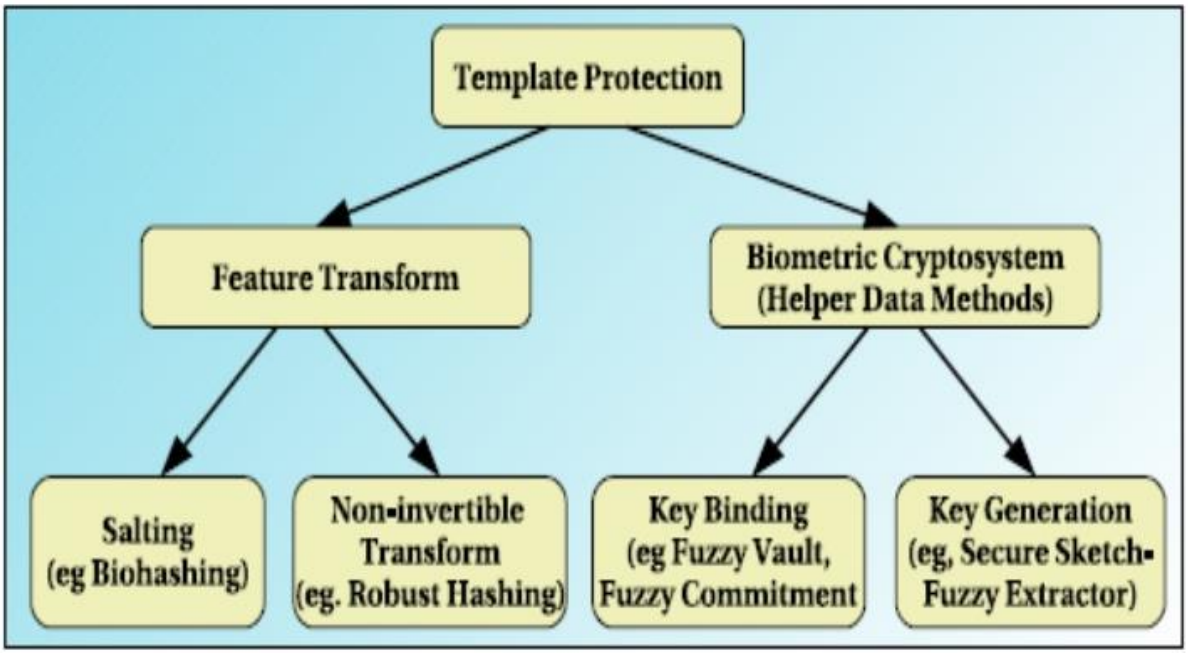
Các phương pháp bảo vệ mẫu sinh trắc phải đảm bảo 4 tiêu chí sau:

- **Đa dạng(Diversity):** mẫu đặc trưng bảo vệ không được giống nhau trên các hệ thống khác nhau, đây là yêu cầu cực kỳ quan trọng đối với các phương pháp bảo vệ để đảm bảo tính riêng tư cho người dùng, ví dụ: cùng một người và cùng sử

dụng khuôn mặt cho hai hệ thống sinh trắc khác nhau thì hai mẫu sinh trắc lưu trên database của hai hệ thống phải khác nhau.

- **Tính hủy bỏ(Revocability):** giống như password, người dùng có thể thay đổi những thông tin lưu trên database nhưng cùng một sinh trắc.
- **An toàn(Security):** một hệ thống phải đủ mạnh để khó có thể truy ngược lại các đặc điểm sinh trắc của người dùng với dữ liệu được lưu trong database.
- **Hiệu suất(Performance):** không được làm giảm khả năng nhận diện(FAR và FRR) của hệ thống sinh trắc.

Dưới đây sơ đồ tổng quan các phương pháp bảo vệ mẫu sinh trắc hiện nay:



Hình 9. Sơ đồ phân loại các phương pháp bảo vệ mẫu.

Như hình trên chúng ta có thể thấy một số phương pháp bảo vệ sinh trắc, mỗi phương pháp có mục đích và điểm mạnh riêng, vấn đề trong một hệ thống xác thực là làm sao kết hợp những phương pháp trên để có được một hệ thống có thể chống lại các loại tấn công như đã nêu trên. Dưới đây sẽ là điểm sơ một vài ý chính về đặc điểm các phương pháp bảo mật mẫu sinh trắc, một số phương pháp không sử dụng trong luận văn này chỉ nói một vài ý chính, riêng Non-invertible Transform và Fuzzy-commitment sẽ nói kỹ hơn. Có hai phương pháp bảo vệ sinh trắc chính là:

2.4.1. Feature Transform

Gọi F là một hàm chuyển đổi(transformation function) nó sẽ được áp dụng lên sinh trắc T, và kết quả của hàm này(F(T,K)) sẽ được lưu trên database của hệ

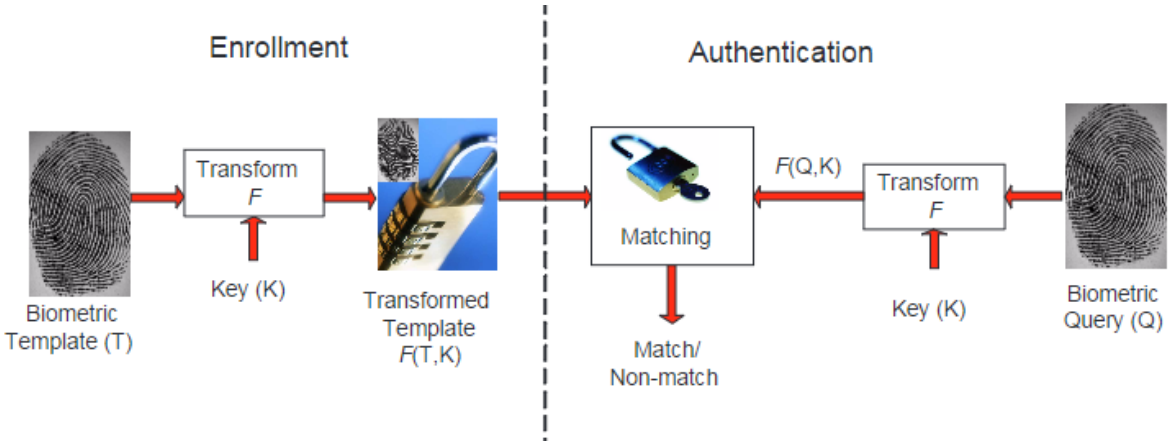
thống với K(có thể là random key hay password) là tham số của hàm F (Hình 6). Và hàm F này phải thỏa mãn một yêu cầu là kết quả sau khi tính toán sẽ được dùng giống như sinh trắc của người dùng. Ví dụ nếu T~T' thì F(T,K)~F(T',K) hay nói cách khác hàm F sẽ còn bảo toàn được độ lệch giữa hai sinh trắc sau khi chuyển. Phương pháp này sẽ bảo vệ mẫu sinh trắc gốc của người dùng, tức là với dữ liệu được lưu trong database thì không thể nào truy ngược lại được T. Tùy thuộc vào khả năng của hàm F mà có hai phương pháp con của nó là:

2.4.1.1. Salting

Hàm F có khả năng có thể phục hồi lại được đặc trưng gốc của người dùng nếu kết hợp với đúng K lúc chuyển đổi, như vậy khả năng bảo mật của phương pháp này phụ thuộc vào K.

2.4.1.2. Non-invertible

Không có khả năng phục hồi lại được sinh trắc của người dùng, phương pháp này còn có tên khác là biến đổi một chiều. Và rất khó để phục hồi lại đặc trưng sinh trắc gốc cho dù có biết K.



Hình 10. Sơ đồ ý tưởng của Feature Transform.

2.4.2. Biometric Cryptosystem

Phương pháp này được phân thành hai phương pháp: Key Binding và Key Generation tùy thuộc vào cách tính được dữ liệu phục hồi(Helper Data).

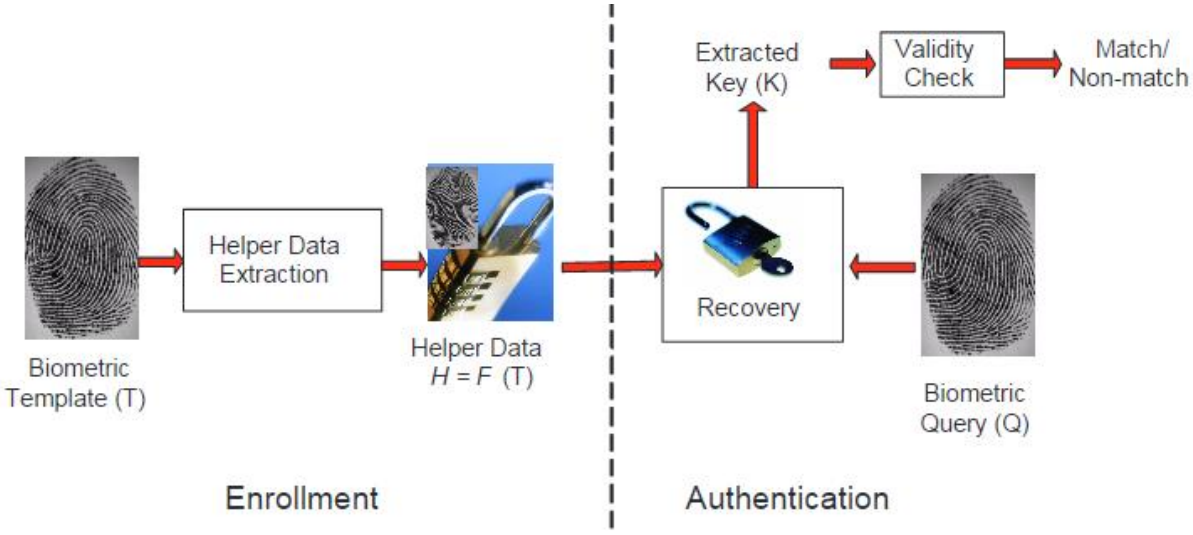
2.4.2.1. Key Binding

Phương pháp này sẽ kết hợp **key** với mẫu sinh trắc người dùng để tạo helper data, hệ thống này sẽ công khai helper data và cho dù có lấy được helper data thì cũng sẽ không khôi phục lại đặc trưng sinh trắc của người

dùng hay key. Muốn hồi phục lại khóa từ helper data thì cần phải cung cấp một mẫu sinh trắc phù hợp(cùng một người).

2.4.2.2. Key Generation

Với phương pháp này thì helper data được sinh ra từ mẫu sinh trắc gốc(biometric template) và khóa được sinh ra từ helper data và mẫu sinh trắc người dùng.



Hình 11. Sơ đồ ý tưởng của hệ thống Biometric Cryptosystem.

Bảng tóm tắt:

Bảng 2. Tóm tắt từng giải thuật.

Phương pháp	Thành phần bảo vệ mẫu sinh trắc	Đối tượng được lưu trữ	Cách xử lý biến thể intrauser
Salting	Bí mật của khóa K	Phạm vi công khai: mẫu chuyển đổi F(T,K) Bí mật: khóa K	Đối sánh trong phạm vi chuyển đổi M(F(T,K), F(Q,K))
Non-invertible tran-form	Không có khả năng khả nghịch từ hàm chuyển đổi F	Phạm vi công khai: mẫu chuyển đổi F(T,K), K	Đối sánh trong phạm vi chuyển đổi M(F(T,K), F(Q,K))
Key binding cryp-tosystem	Cấp độ an toàn phụ thuộc vào số lượng thông tin được tiết lộ ra bởi dữ liệu hỗ trợ H	Phạm vi công khai: hỗ trợ dữ liệu H = F(T,K)	Sửa lỗi và sử dụng lượng hóa cụ thể K = M(F(T,K),Q)
Key generation cryptosystem	Cấp độ an toàn phụ thuộc vào số lượng thông tin được tiết lộ ra bởi dữ liệu hỗ trợ H	Phạm vi công khai: hỗ trợ dữ liệu H = F(T,K)	Sửa lỗi và sử dụng lượng hóa cụ thể K = M(F(T),Q)

- F: Transformation function.
- T: Biometric when enrrollment phase.
- Q: Biometric when authentication phase.
- K: key.
- H:Helper data.
- M:The matcher that operates in the transfomated domain.

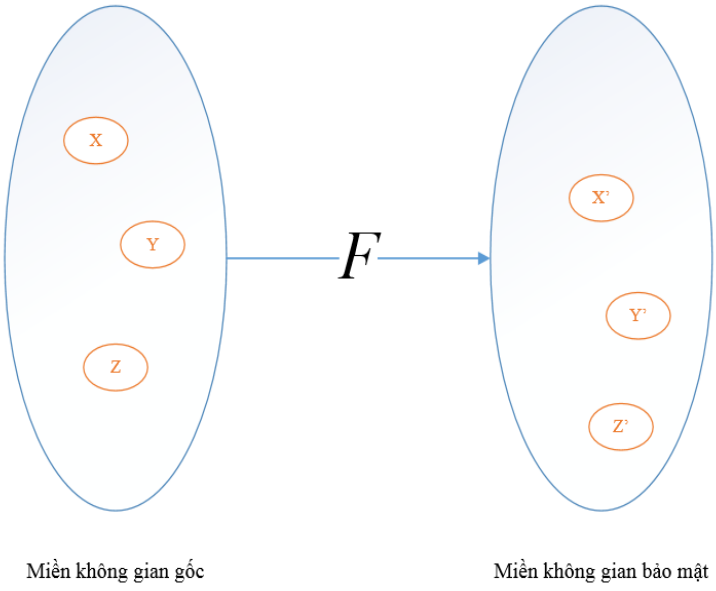
2.5. Các kỹ thuật bảo vệ chính

Trong phần này chỉ nêu lên hai kỹ thuật chính sử dụng trọng luận văn, đó là hai phương pháp bảo vệ: Non-invertible Transform và Fuzzy-commitment.

2.5.1. Non-invertible Transformaton

Mục đích của phương pháp này là bảo vệ đặc trưng sinh trắc người dùng, về tổng quát thì phương pháp này sử dụng một hàm F để ánh xạ các template ban đầu sang một miền không gian bảo mật. Những template đã được biến đổi này sẽ được sử dụng thay vì dùng template ban đầu. Và phương pháp này đảm bảo được tính hủy bỏ (revocability) nhờ việc có thể thay đổi hàm F khi cần.

Phương pháp này sử dụng phép chiếu ngẫu nhiên(Random projection – RP) trong việc thành lập hàm F. Phép chiếu ngẫu nhiên là: phương pháp sử dụng ma trận trực giao để ánh xạ một điểm sang một vùng không gian mới mà vẫn bảo toàn khoảng cách giữa các điểm.



Hình 12. Phép biến đổi F ánh xạ các vector đặc trưng vào miền không gian bảo mật.

Tùy theo hàm **F** mà có nhiều biến thể của giải thuật này, trong luận văn này tôi xây dựng hệ thống sinh trắc sử dụng kiểu Integer và vector sinh ra có dạng mảng một chiều nên tôi chọn hàm **F** được xây dựng dựa trên ma trận trực giao, chính vì vậy trong phần Non-invertible tôi chỉ giới thiệu phương pháp ma trận trực giao. Có nhiều phương pháp tạo ma trận trực giao như Gram-Schmidt hay Hisham Al-Assam.. nhưng trong luận văn này tôi chỉ giới thiệu phương pháp Hisham Al-Assam vì đây là phương pháp tối ưu, vì các bước tạo ra ma trận đơn giản hơn và phép nhân hai ma trận cũng đơn giản hơn rất nhiều, làm cho hệ thống tối ưu hơn, và độ phức tạp của hệ thống cũng không tăng lên đáng kể.
Quy trình thực hiện:

- Tạo **m** vector ngẫu nhiên thuộc không gian R^n (n là số chiều của vector đặc trưng sinh trắc x dựa vào khóa nền người dùng hoặc token).
- Áp dụng quy trình Hisham Al-assam để tạo ma trận trực giao $A[m \times n]$ từ các vector ngẫu nhiên trên.

– Biến đổi vector đặc trưng χ sử dụng ma trận trực giao A : ***y* = *Ax***

Phương pháp của Hisham Al-Assam:

trước tiên ta xét một ma trận 2×2 như sau:

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

Ta thấy ma trận trên trực gaio với mọi θ . Dựa vào tính chất này ta có thể tạo một ma trận $2n\times 2n$ từ những ma trận 2×2 . Hisham Al-Assam đề xuất phép chiếu ngẫu nhiên như sau:

– Tự tạo một tập ***n*** giá trại ngẫu nhiên {θ_1 , θ_2 ,.... θ_n} là các số thực trong khoảng [0..2π], ta tạo một ma trận trực giao $A[2n\times 2n]$ với các đường chéo là các ma trận đơn vị θ_i như trên.

$$\begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 & 0 & ... & ... & 0 & 0 \\ -\sin\theta_1 & \cos\theta_1 & 0 & 0 & ... & ... & 0 & 0 \\ 0 & 0 & \cos\theta_2 & \sin\theta_2 & ... & ... & 0 & 0 \\ 0 & 0 & -\sin\theta_2 & \cos\theta_2 & ... & ... & 0 & 0 \\ M & M & M & M & O & O & M & M \\ M & M & M & M & O & O & M & M \\ 0 & 0 & 0 & 0 & ... & ... & \cos\theta_n & \sin\theta_n \\ 0 & 0 & 0 & 0 & ... & ... & -\sin\theta_n & \cos\theta_n \end{bmatrix}$$

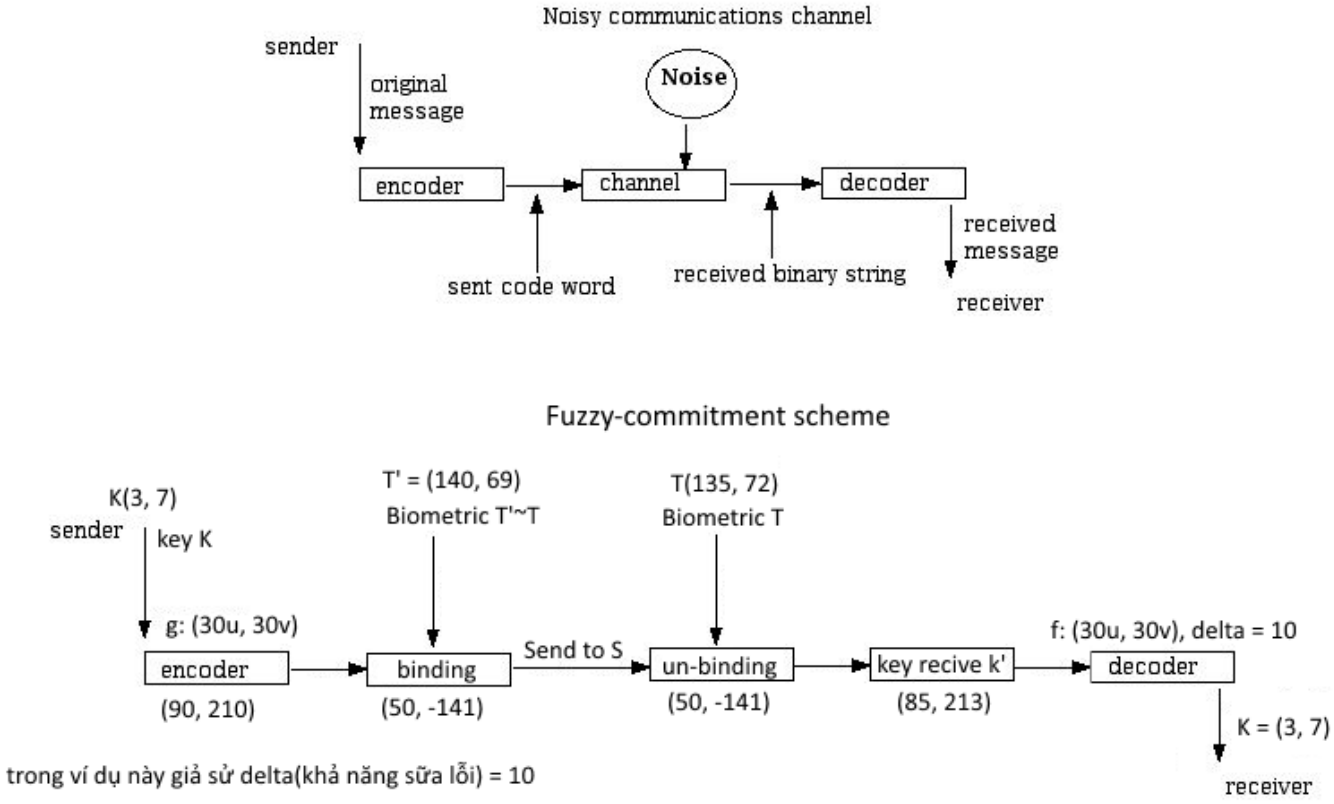
– Ánh xạ cac vector đặc trưng ***x*** sang miền bảo mật sử dụng ma trận ***A: y = A.x.***

Trong thực tế, ***y = A.x + b***, với vector *2n* chiều ***b*** nhằm tạo khả năng revocability cho mô hình.

2.5.2 Fuzzy-commitment

2.5.2.1 Ý tưởng

Ý tưởng chính của fuzzy-commitment là dựa trên thành quả của kỹ thuật sữa lỗi(Error Corection Code)[6]. Hình dưới đây sẽ giải thích tại sao fuzzy-commitment lại dựa trên ý tưởng của EEC.



Hình 13. Sơ đồ ý tưởng fuzzy-commiement.

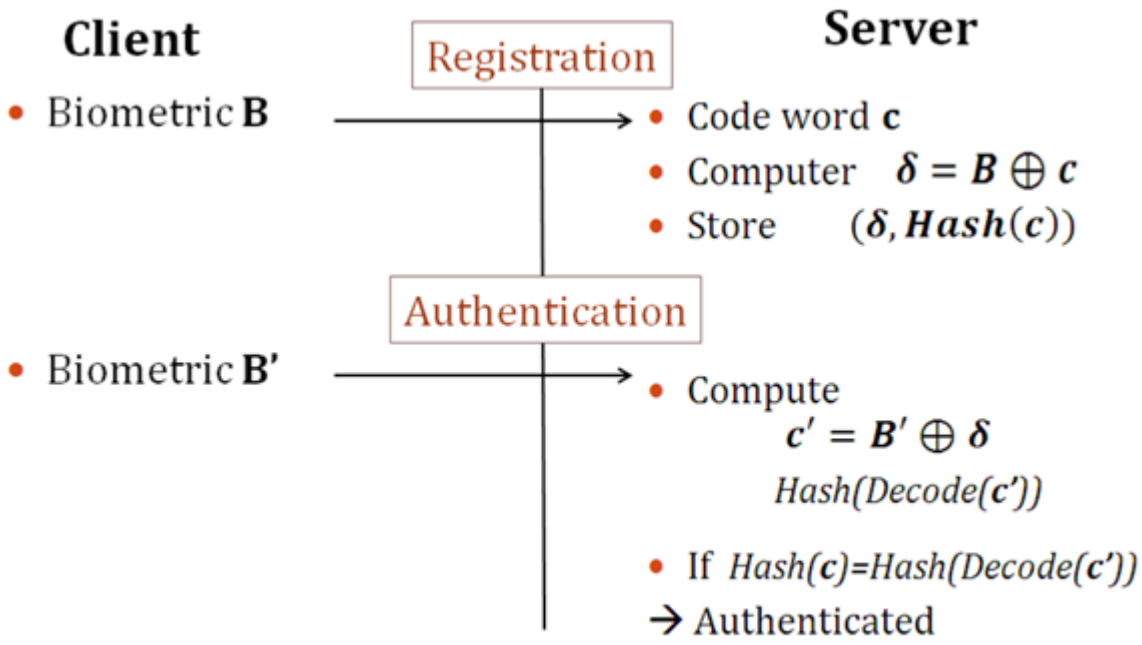
Nhình hình phía trên sẽ giải thích tại sao lại nói fuzzy-commitment lại dựa trên ý tưởng của EEC, có thể nói 'nhiễu' trong EEC được tác giả áp dụng cho độ lệch của biometric tác dụng lên giá trị key và sao đó sẽ thu được khóa k sau khi áp dụng hàm sữa lỗi. Hay nói một cách khác nhiễu do đường truyền được chuyển thành nhiễu trong quá trình trích xuất đặc trưng sinh trắc. Và để hiểu hơn về chi tiết giải thuật thì phần tiếp theo sẽ nói chi tiết toàn bộ giải thuật.

2.5.2.2 Chi tiết kỹ thuật

Fuzzy-commitment là phương pháp đã được đề cập rất trong tài liệu[3] đây được coi là một trong những phương pháp đầu tiên trong việc sử dụng biometric trong hệ thống bảo mật(biometric crptosystem). Fuzzy-commitment là một ý niệm tổng quát để xây dựng hệ thống bảo vệ khóa kết hợp với sinh trắc để có thể truyền dữ liệu này trên đường truyền, và như giới thiệu ở trên thì fuzzy-commitmen này dựa trên ý tưởng của hàm sữa lỗi trong ECC chính vì vậy nên mỗi loại hàm sữa lỗi thì mô hình này sẽ khác đi một chút ví dụ với hàm sữa lỗi trên kiểu số Integer thì hàm binding không thể là phép XOR mà phải là pháp khác (+,-,...) còn trên kiểu số Binary thì hàm binding là phép

XOR và giải thuật sữa lỗi phải áp dụng được với kiểu số đó(trong luận văn này thì về phần lý thuyết tôi xin giới thiệu về kiểu Binary và hàm sữa lỗi kiểu Binary, còn trong hệ thống thật tôi sử dụng kiểu Integer và kiểu sữa lỗi trên số Integer để các bạn có cái nhìn tổng quát hơn). Với các hệ thống bảo mật cũ như RSA , DES,.. thì các thông tin đầu vào để thiếp lập hệ thống như khóa,.. cần phải chính xác 100% tuy nhiên vấn đề tạo khóa, lưu khóa lại có một số rủi ro nhất định nếu kẻ gian lây được quyền sử dụng máy tính của người dùng, hay một số rủi ro khác trong quá trình lưu và sử dụng khóa. Và một giải pháp mà các nhà khoa học nghĩ đến là xây dựng một hệ thống bảo mật có sử dụng biometric làm 'khóa' để vận hành hệ thống tuy nhiên vấn đề lớn lại đặc ra đó là biometric của người dùng lại khác nhau trong những lần lấy khác nhau , như vậy thì làm sao đủ tiêu chuẩn để đưa vào các hệ thống bảo mật truyền thống được. Chính vì vậy mà người ta đa phát minh ra hệ thống mà 'khóa' của người dùng nhập vào có thể khác nhau, tất nhiên là khác nhau trong khoảng cho phép, và hệ thống mạng tên fuzzy-commitment ra đời , chính chính vì vậy nên có chữ 'fuzzy'. Đây là phương pháp kết hợp hai công nghệ nổi tiếng là Error Correcting Codes(ECC) và Cryptography. Cốt lõi của fuzzy-commitment là ECC, nó là phần trung tâm của hệ thống. Chính vì vậy nên có nhiều phiên bản của fuzzy-commitment, tùy thuộc vào giải thuật ECC mà có một số loại cải biến của fuzzy-commitment, nhưng sau khi qua tìm hiểu thì đa số người ta sử dụng Linear Error Correcting Codes. Trong ECC thì có hai phần quan trọng đó là Encode và Decode về vấn đề chi tiết của giải thuật ECC thì sẽ được đề cập trong tài liệu [3] mà ở đây chúng ta sẽ bàn về vấn đề làm sao để áp dụng giải thuật này vào trong mô hình của chúng ta. Thực tế chúng ta chỉ áp dụng 'một nữa' chức năng của ECC , chúng ta sử dụng chức năng Decodes nhưng không cần sử dụng hoàng toàn chức năng Encode mà chỉ sử dụng một phần. Nói sơ qua ECC. trong hệ thống ECC gồm có các thành phần chính sau $c \subseteq C$ (C là tập code-words), và một hàm có chức năng ánh xạ msg(thông điệp) vào một code-words nào đó trước khi truyền qua kênh bị nhiễu. Giả sử M là thông điệp cần chuyển và một hàm ánh xạ vào không gian code-words là hàm g(encoding function) như sau: $\mathbf{g}$: $\mathbf{M} \rightarrow \mathbf{C}$, và một hàm decode $\mathbf{f}$: $\mathbf{C} \rightarrow \mathbf{M}$, thực ra thì hàm $\mathbf{f}$ có thể hồi phục $\mathbf{M}$ từ một tập các chuỗi gần $\mathbf{C}$. Trong fuzzy-commitment thì đặc trưng sinh trắc của người dùng được xem như là C. Trong pha đăng ký thì người dùng cung cấp cho hệ thống mẫu sinh trắc B cho server, sau đó server sẽ

chọn một code-words c và tính δ = B ⊕ c, sau đó server sẽ lưu cặp giá trị (δ, Hash(c)) vào database. Trong pha xác thực thì người dùng sẽ cung cấp cho server mẫu đặc trưng sinh trắc B' khi đó server sẽ tính giá trị c' = B' ⊕ δ , sau đó áp dụng hàm Decode để decode c' , sau đó lấy giá trị Hash của kết quả sau decode và so sánh với giá trị Hash(c) lưu trong database trước đó, nếu cùng một người thì hai gí trị này giống nhau và ngược lại. Sau đây là ví dụ từng bước tính toán của fuzzy-commitmet(trong bài báo cáo thì tôi xin làm giống với nguyên văn của fuzzy-commitment là áp dụng mã sữa lỗi trên chuỗi binary ,nhưng trong hiện thực hệ thống vì sử dụng lại mã nguồn của nhóm trước để lấy đặc trưng sinh trắc mà đặc trưng sinh trắc này lại sử dụng kiểu integer nên trong hiện thực thì tôi sẽ dùng mã sữa lỗi kiểu integer). Dưới đây là mô hình của một hệ thống xác thực sử dụng fuzzy-commitment.



Hình 14. Sơ đồ hệ thống xác thực bằng fuzzy-commitment.

+ **Registration:** người dùng sẽ đưa vào một vector biometric B để đăng ký như sau B(01010,10101) và trong hệ thống sữa lỗi thì code-words sẽ được chọn ngẫu nhiên từ tập sau {00000,11111}, và sau khi nhận được vector biometric thì hệ thống sẽ bắc đầu bằng việc chọn ngẫu nhiên C ví dụ hệ thống chọn ra c(00000,11111) , sau đó nó tính δ = B ⊕ c = (01010, 01010) và server

sẽ lưu lại (δ,Hash(c)) và giả sử khả năng sữa lỗi của hệ thống này là t = 2 và đây cũng chính là giá trị sai lệch chấp nhận của cùng một người.

+ **Authentication:** người dùng sẽ phải cung cấp cho hệ thống một vector đặc trưng sinh trắc mới ví dụ B'(11010, 11101) (ta có thể thấy độ lệch Hamming giữa B và B' là 2), khi đó hệ thống sẽ tín C' = B'⊕ δ = (10000, 10110) sau đó hệ thống sử dụng hàm Decode để decoeding C' và đêm so sánh với dữ liệu trong database để quyết định xem có phải cùng một người không, và dễ dàng ta thấy Hash(decode(C')) == Hash(C), trong hệ thống này thì hàm decode là hàm sẽ làm như sau: nếu trong chuỗi đó số 1 nhiều hơn số 0 thì decode ra 11111 còn ngược lại sẽ ra 00000 như vậy decode(10000, 10110) = (00000,11111)).

Nhận xét: với hệ thống xác thực dựa trên fuzzycommitment thì ta có thể thấy một số ưu điểm như sau:

- không lưu trực tiếp biometric của người dùng lên server, mà thay vào đó chúng ta sẽ lưu (δ,Hash(c)) như vậy với dữ liệu này thì không có khả năng truy xuất được c cũng như B từ Hash(c) và δ, như vậy nếu áp dụng phương pháp này với mô hình đề xuất phía trên thì chỉ bảo mật được tấn công bên trong , nhưng không có khả năng chống lại tấn công bên ngoài vì biometric gửi trực tiếp lên server mà không có bất kỳ phương pháp bảo mật nào. Chính vì vậy mà chúng ta sẽ cần phải sử dụng thêm một phương pháp nữa trong bảo vệ đặc trưng sinh trắc để chống lại tấn công bên ngoài phương pháp đó là non-invertible transformation.

CHƯƠNG 3. PHƯƠNG PHÁP RÚT TRÍCH ĐẶC TRƯNG KHUÔN MẶT

3.1. Phương pháp xác định khuôn mặt

Đã có nhiều công trình nghiên cứu tiến hành để tìm ra phướng pháp xác định khuông mặt người từ ảnh xám đến ngày nay là ảnh màu. Dựa vào tính chất của các phương pháp xác định khuôn mặt, có 4 hướng tiếp cận chính.

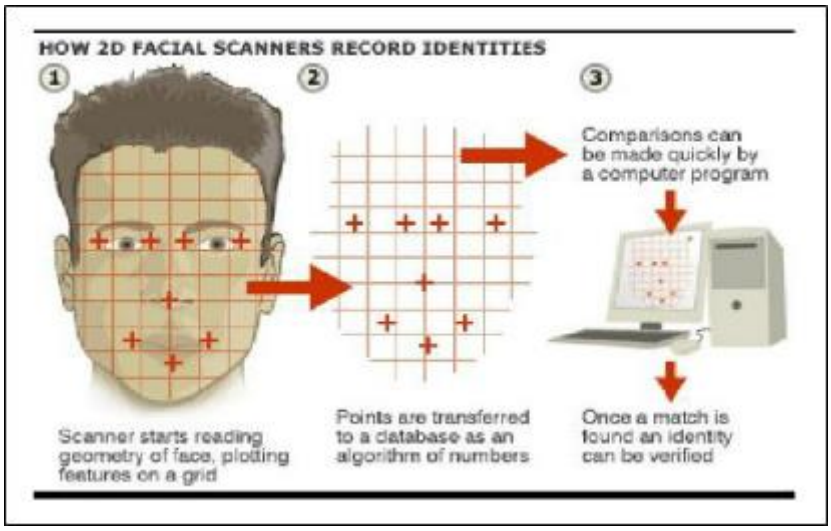
– **Dựa trên so khớp mẫu:** dùng các mẫu sinh trắc chuẩn của khuôn mặt người (các mẫu này được chọn lựa và lưu trữ) để mô tả cho khuôn mặt người hay các đặc trưng khuôn mặt. Trong so khớp mẫu, các mẫu chuẩn của khuôn mặt sẽ được xác định trước hoặc xác định các tham số thông qua một hàm. Từ một ảnh đưa vào, tính các giá trị tương quan so với các mẫu chuẩn về đường

viền khuôn mặt, mắt, mũi và miệng. Thông qua các giá trị tương quan này mà các tác giả quyết định có hay không tồn tại khuôn mặt trong ảnh. Tuy nhiên phương pháp này không hiệu quả đối với những ảnh có tỷ lệ chênh lệch cao. Dùng PCA(phân tích thành phần chính - Principal Compoment Analysis) để có một tập hình chiếu cơ bản từ các mẫu khuôn mặt, hình chiếu được mô tả như một mảng các bit. Dùng đặc trưng hình chiếu riêng biệt kết hợp biến đổi Hough để xác minh khuôn mặt người. Sau đó một phương pháp xác định dựa trên đa loại mẫu để xác định các thành phần của khuôn mặt người được trình bày. Phương pháp này định nghĩa một số giả thuyết để mo tả các khả năng của các đặc trưng khuôn mặt. Với một khuôn mặt sẽ có một tập giả thuyết, lý thuyết DepsterShafer. Dùng một nhân tố tin cậy để kiểm tra sự tồn tại hay không của các đặc trưng khuông mặt và kết hợp nhân tố tin cậy này với một độ do để xem xet có hay không có khuôn mặt.

- **Dựa trên diện mạo:** phương pháp này áp dụng kỹ thuật theo hướng xác suất thống kê và học máy để tìm những đặc tính liên quan của khuôn mặt. Các đặc tính đã được học ở trong hình thái các mô hình phân bố hay các hàm biệt số nên dùng, có thể dùng các tính này để xác định khuôn mặt người. Đồng thời, bài toán giảm số chiều được quan tâm để tăng hiệu quả tính toán cũng như hiệu quả xác định.
- **Dựa trên tri thức:** thực hiện mã hóa các hiểu biết của con người về các loại khuôn mặt người thành các luật. Thông thường là các luật mô tả quan hệ của các đặc trưng trên khuôn mặt. Trong hướng này , các luật sẽ phụ thuộc rất lớn vào tri thức của những tác giả nghiên cứu về bài toán xác định khuôn mặt người. Đây là hướng tiếp cận dễ dàng xây dựng các luận cơ bản để mô tả các đặc trưng của khuôn mặt và các quan hệ tương ứng.
- **Dựa trên đặc trưng không thay đổi:** đây là hướng tiếp cận theo các đặc trưng không thay đổi của khuôn mặt người để xác định khuôn mặt người. Trên thực tế con người dễ dàng nhận biết các khuôn mặt, các đối tượng trong các tư thế khác nhau và điều kiện ánh sáng khác nhau thì phải tồn tại các thuộc tính hay đặc trưng không thai đổi. Các đặc trưng này thường là lông mày, mắt, mũi, miệng, và các đường viền của tóc được trích xuất bằng phương pháp xác định cạnh.

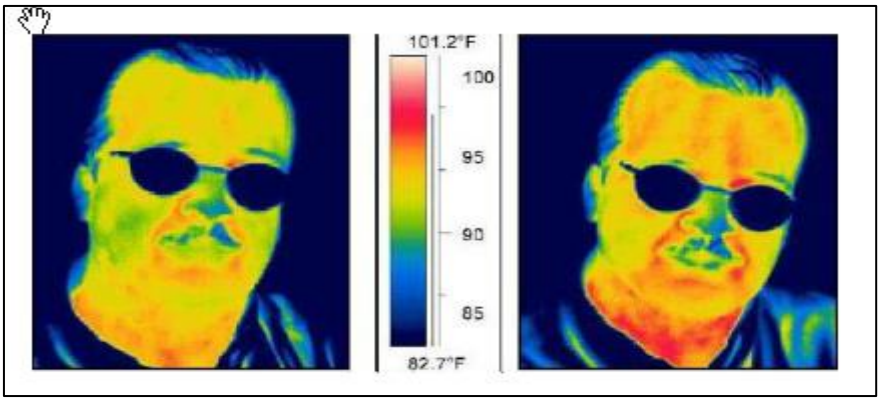
3.2. Các đặc trưng khuôn mặt

– **Đặc trưng hình học:** cấu trúc hình dạng và các thành phần trên khuôn mặt: miệng , mắt, mũi, lông mày. Khoảng cách giữa mắt, mũi , miệng và hàm đường bao các hốc mắt, các cạnh của miệng, vị trí mũi, hai mắt và các vùng xung quanh. Các thành phần khuôn mặt được rút trích để hình thành vector đặc trưng biểu diễn hình học khuôn mặt.



Hình 15. Điểm đặc trưng khuôn mặt.

– **Đặc trưng về diện mạo:** kết cấu da như các nếp nhăn trên khuôn mặt, biểu đồ nhiện khuôn mặt, các mẫu nhiệt khuôn mặt là duy nhất với mỗi người và đặc trưng nụ cười. Các đặc trưng về diện mạo có thể được rút trích trên cả khuôn mặt hoặc một phần nào đó trên khuôn mặt.

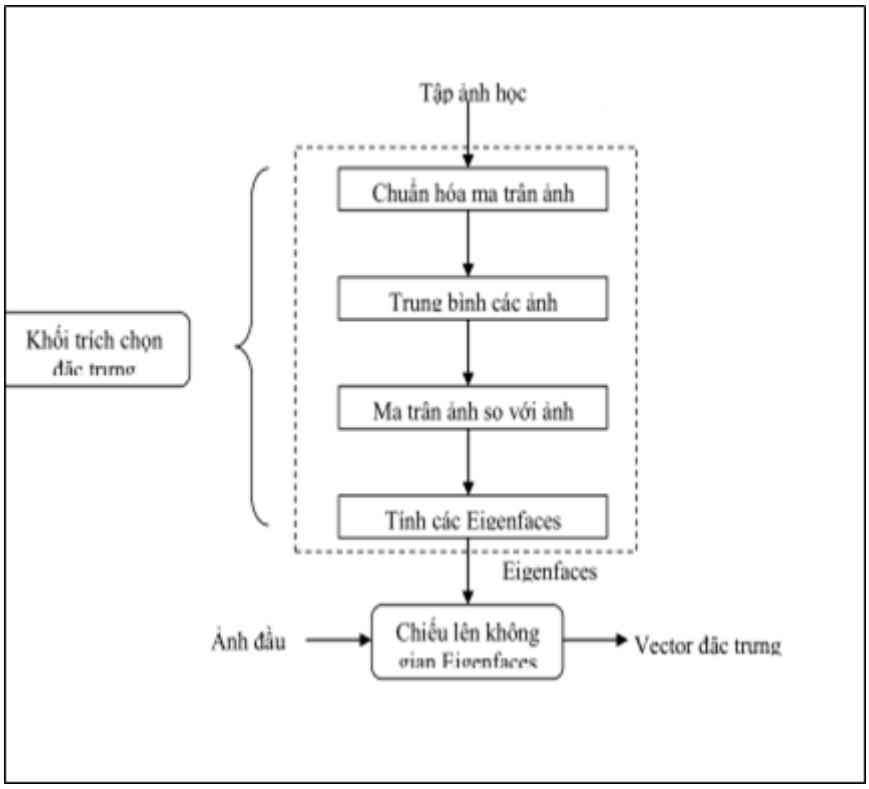


Hình 16. Diện mạo biểu hiện sự thay đổi khuôn mặt.

3.3. Phương pháp rút trích

Trích rút đặc trưng là kỹ thuật sử dụng các thuật toán để lấy ra những thông tin mang những đặc điểm riêng biệt của một người. Các khâu trong qua trình trích chọn đặc trưng:

- Đầu vào: ảnh đã được chuẩn hóa.
- Đầu ra: vector đặc trưng của ảnh đầu vào.



Hình 17. Phương pháp rút trích.

3.3.1 Phép biến đổi PCA

Ta có không gian dữ liệu quan sát S = {x} gồm n vector dữ liệu mẫu N chiều. Dữ liệu thô tồn tại sự tương quan ngẫu nhiên giữa các thành phần, do đó có sự dư thừa dữ liệu. Ý tưởng của phép biến đổi PCA là phân tích dữ liệu thành những không tương quan (gọi là các thành phần chính) để giảm độ dư thừa dữ liệu.

Phép biến đổi PCA được định nghĩa như sau: $u = \Psi^T x$

Trong đó $x = [x_1 x_2 x_3 \dots x_N]^T$, $u = [u_1 u_2 u_3 \dots u_N]^T$ là các thành phần thứ i và j không tương quan trong không gian mới.

Và ma trận Ψ là ma trận phép biến đổi với kích thước N^2 có dạng:

$$\Psi^T = [V_i]^T = \begin{matrix} V_{11} & \dots & V_{1N} \\ \dots & \dots & \dots \\ V_{N1} & \dots & V_{NN} \end{matrix}$$

Trong đó V_i là các vector riêng tương ứng với ma trận hiệp phương sai của các x quan sát được.

$$C = E[(x - \mu)(x - \mu)^T$$

Trong đó $\mu = \begin{bmatrix} \mu_1 \\ \cdots \\ \mu_N \end{bmatrix}$ trong đó $\mu_i = \frac{1}{n} \sum_{k=1}^{n} x_{ik}$

3.3.2 Eigenface

Eigenface đã được M. A. Turk và A. P. Pentland đề xuất năm 1991. Ý tưởng chính của eigenface là để lấy được các đặc trưng trong nghĩa toán học thay vì đặc trưng về mặt vật lý bằng cách sử dụng chuyển đổi toán học để được nhận dạng.

Giai đoạn đầu tiên, một số lượng lớn các hình ảnh khuôn mặt được đưa vào tập huấn luyện. Những hình ảnh được huấn luyện là đại diện tốt nhất cho tất cả các khuôn mặt có thể được gặp phải. Kích thước, hướng và cường độ ánh sáng nên được chuẩn hóa.

Chọn tập ảnh huấn luận là T_1, T_2, T_3, … T_M các bức ảnh này đã được chỉnh cùng kích thước và chỉnh tâm. Chúng ta biểu diễn mỗi bức ảnh T_i kích thước N x N bằng một vector T có kích thước N^2 chiều.

- *Bước 1: Tính vector trung bình*

  Khi đó ảnh trung bình được xác định theo công thức sau:

$$\Psi = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i$$

- *Bước 2: Tính* $\Phi_i = \Gamma_i - \Psi$ *với i = 1,2… N^2 là khác nhau của từng ảnh trong tập huấn luyện so với ảnh trung bình:*

  Ví dụ như tập ảnh được cho ở hình :

Hình 18. Tập ảnh huấn luyện.

Hình 19. Ảnh trung bình.

Khi đó ta được ảnh trung bình như ảnh bên trên

– *Bước 3: Tính ma trận phương sai C của các vector quan sát*

Sau đó ta tính được ma trận C:

$$C = \frac{1}{M}\sum_{i=1}^{M} \Phi_i \Phi_i^T = A A^T$$

Ma trận C có kích thước N^2 x N^2

Trong đó $A = [\Phi_1 \Phi_2 \dots \Phi_M]$ *kích thước M x M*

– *Bước 4: Tìm các vector riêng u_i của ma trận phương sai C*. Tuy nhiên ma trận này có kích thước là N^2 x N^2 quá lớn nên việc tìm vector này không khả thi.

Chúng ta xét ma trận $A A^T$ có kích thước là MxM

Tìm vector riêng V_i của ma trận A^T A, ta có:

$$A^T A v_i = \lambda_i v_i$$

Nhân 2 vế với ma trận A và rút gọn ta được mối quan hệ :

$u_i = A\ v_i$

Do A A^T và A^TA có cùng giá trị riêng và các vector riêng của chúng quan hệ với nhau theo:

$$u_i = A\ v_i$$

Chú ý: A^T A có thể có M giá trị riêng và vector riêng.

M giá trị riêng của A^T A cùng với vector riêng tưng ứng với M giá trị riêng lớn nhất của A A^T.

Sau đó tính M vector riêng ứng với M giá trị riêng lớn nhất của AA^T :

$$u_i = Av_i$$

đồng thời chuẩn hóa vector u_i sao cho

$$||u_i|| = 1.$$

– *Bước 5: Chọn K vector riêng ứng với K giá trị riêng lớn nhất.*

3.3.3 Biểu diễn khuôn mặt theo tập huấn luyện tìm được

Mỗi vector biểu diễn khuôn mặt (trừ đi vector trung bình)

$$\Phi_i - \text{mean} = \sum_{j=1}^{K} w_j v_j \quad i = 1,2\ldots M$$

Trong đó v_j gọi là các ảnh riêng, $w_j = u_j^T \Phi_i$ chính là thành phần chính thứ j trong không gian mới.

u_j là các vector ảnh riêng

Mỗi ảnh được chuẩn hóa trong tập luyện sẽ được biểu diễn trong cơ sở này bởi vector:

$$\Omega_i = \begin{matrix} w_1^i \\ w_2^i \\ \cdots \\ w_k^i \end{matrix} \quad \text{trong đó i = 1,2 …M}$$

CHƯƠNG 4. MÔ HÌNH ĐỀ XUẤT

4.1. Kiến thức tổng quát

Mục tiêu cuối cùng của chúng ta là làm sao có một mô hình hoàn thiện để có thể đưa vào thực tiễn. Một hệ thống mà có thể chống lại các loại tấn công hiện nay(trong luận văn này chúng ta sẽ không đề cập đến các loại tấn công dựa trên những sai lầm của người dùng mà chỉ tập trung vào các loại tấn công trên hệ thống). Khi xây dựng một hệ thống thì nhà thiết kế hệ thống phải giải quyết những câu hỏi như sau:

– Server sẽ lưu thông tin gì để lúc xác thực có thể lấy ra sử dụng?.

– Làm sao bảo vệ những thông tin người dùng nhập vào.

- Với loại thông tin đó(password với hệ thống cũ, mẫu sinh trắc với hệ thống bây giờ) thì áp dụng giải thuật nào cho phù hợp.
- Mô hình xác thực nào phù hợp với việc truyền dữ liệu trên đường truyền mạng không an toàn như hiện nay.

Và để trả lời cho những câu hỏi ở trên thì trong mô hình xác thực này sẽ áp dụng hai phương pháp bảo vệ mẫu sinh trắc đã đề cập ở phần lý thuyết là: Non-invertible Transformation và Fuzzy-commitment. Mỗi phương pháp sẽ đảm nhiệm một chức năng riêng trong hệ thống tuy nhiên có thể tóm tắt mục đích của hai phương pháp này như sau:

- **Non-invertible Transformation:** nó sẽ giả quyết vấn đề quan trọng nhất đó là không sử dụng trực tiếp mẫu sinh trắc của người dùng mà sử dụng một mẫu sinh trắc đã qua biến đổi và nó an toàn vì đây là phương pháp biến đổi một chiều nên không thể truy ngược lại được đặc trưng gốc(trong trường hợp xấu nhất nếu kẻ xấu biết được hàm F thì cũng rất khó để truy ngược lại), không những vậy phương pháp này còn giúp hệ thống giải quyết vấn đề khả năng hủy bỏ(Revocability) đó là chỉ cần thay đổi tham số hàm F là chúng ta có thể thay đổi được thông tin lưu trên database mà không cần thay dổi sinh trắc của người dùng.
- **Fuzzy-commitment:** đây là phương pháp thuộc loại Key Binding, chính vì vậy mục đích quan trọng nhất của phương pháp này là truyền một khóa K trên đường truyền mạng thông qua Helper data. Đây có thể được xem như là một hệ thống mã hóa sử dụng đặc trưng sinh trắc(Biometric Cryptosystem).

Tóm tắt ý tưởng xây dựng hệ thống: sử dụng Non-invertible để có thể lưu thông tin liên quan đến đặc trưng sinh trắc trên server giúp hệ thống có khả năng revocablity giúp hệ thống chống lại Biometric Template Attack, còn fuzzy-commitment dùng để truyền một khóa **k**(dùng một lần) cho hệ thống Matching sẽ giúp hệ thống chống lại một số loại tấn công như Man-in-Middle Attack hay Replay Attack, và vấn đề Insider Attack sẽ được chống lại khi kết hợp hai phương pháp này với hệ thống.

kí hiệu:

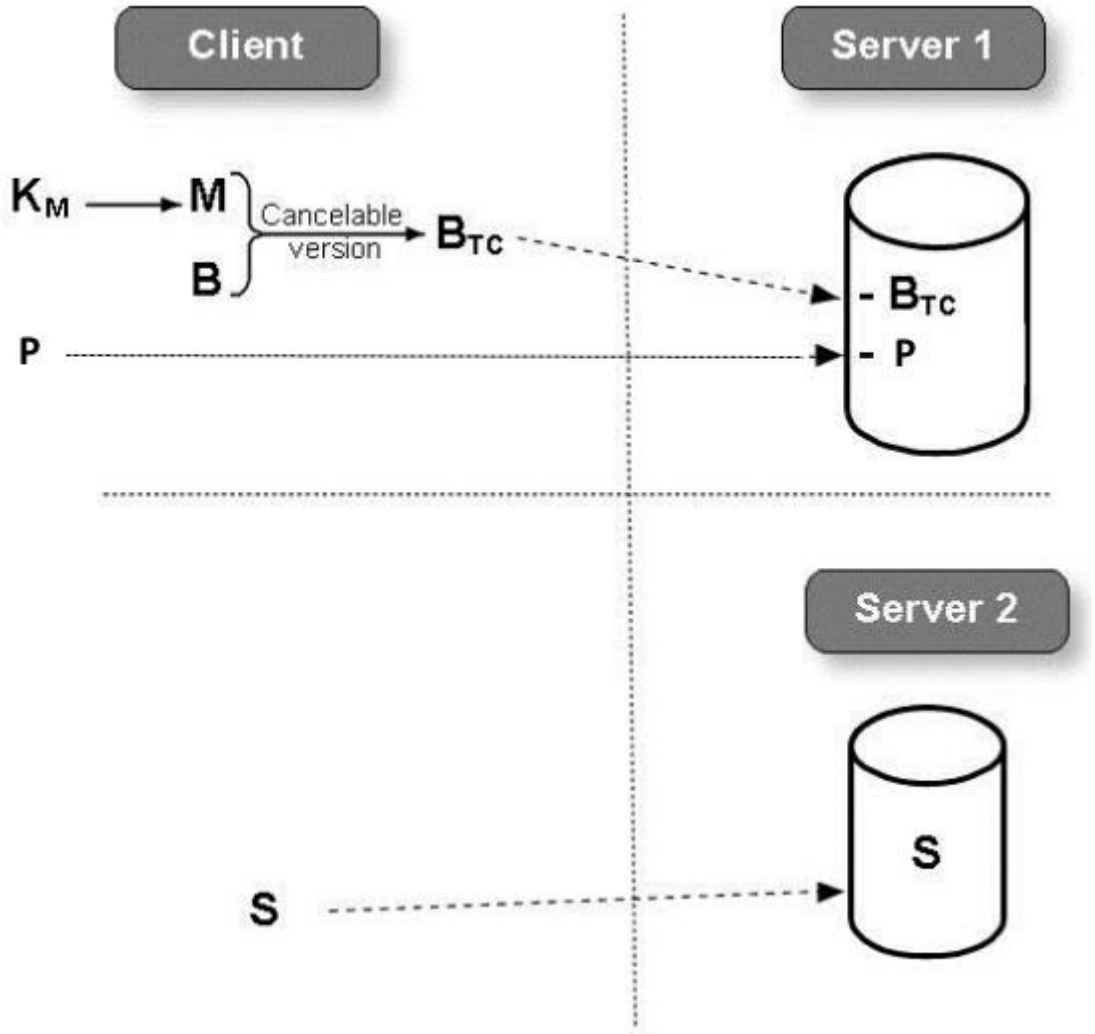
- B : đặc trưng sinh trắc của người dùng để nhập vào hệ thống.
- M : ma trận trực giao.
- B_{TC} : đặc trưng sinh trắc đã được biến đổi nhờ vào việc nhân với ma trận trực giao và sẽ dùng để lưu trên server.
- H(m) : hàm Hash thông điệp m.

- BL : biometric lock của người dùng.
- P : hoán vị
- Pu và Pr : lần lược là PublicKey và PrivateKey của hệ thống.
- E_{PuX} (m) mã hóa thông tin ***m*** với key ***X*** và mã hóa đối xứng.
- S : mobile serial number.
- S_T : mobile serial number lưu trong database của server.
- C : client.
- S_1, S_2 : lần lượt là server thứ nhất và server thứ hai.

4.2. Chi tiết hệ thống

Bất kỳ một hệ thống xác thực hoàn chỉnh nào cũng gồm hai phần chính là: Enrollment và Authentication. Enrollment là phiên đăng ký cho người dùng, trong pha này người dùng sẽ cung cấp một vài thông tin cần thiết cho hệ thống dùng cho pha Authentication. Trong pha Authentication thì người dùng cũng cần cung cấp một vài thông tin cần thiết để hệ thống nhận diện được người dùng. Và nhiệm vụ là chúng ta cần phải thiết kế hệ thống đảm bảo tính bảo mật cho người dùng bằng cách áp dụng hai phương pháp đã nêu ở phía trên.

4.2.1. Enrollment Phase



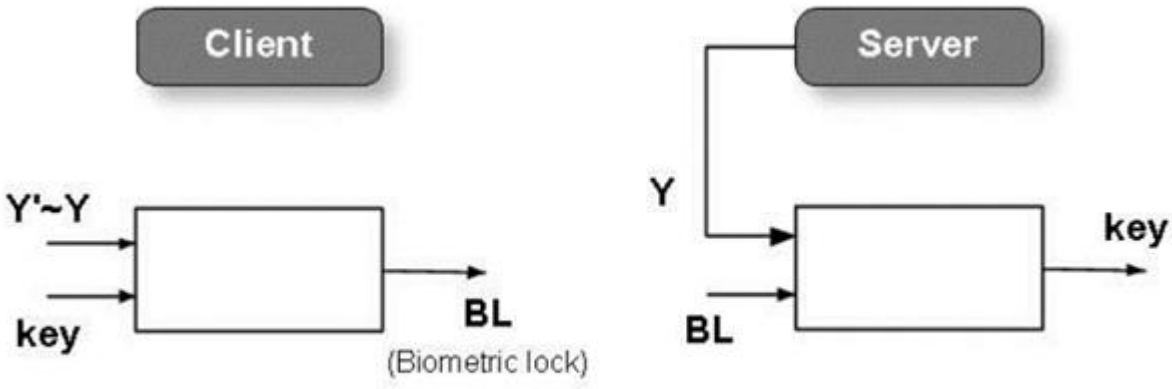
Hình 20. Enrollment Phase.

Trong pha này người dùng sẽ tạo ngẫu nhiên vector K_m và sẽ được lưu trong thiết bị của người dùng, và K_m sẽ được dùng để tạo ma trận trực giao M(dựa trên giải thuật của Hisham Al-Assham) . Sau khi người dùng cung cấp vector đặc trưng sinh trắc B thì sẽ kết hợp với ma trận trực giao(phép nhân hai ma trận) sẽ thu được B_{TC} (chuyển sang vùng bảo mật) và sẽ được gửi qua server S_1 và sẽ được lưu ở đó. Kèm theo đó phía client sẽ sinh ra một số bí mật P và gửi giá trị này đến server . Song song với những công việc trên thì phía client cũng sẽ gửi đến cho server S_2 số serial number S. Việc chia hai server sẽ chia bớt thông tin từ client gửi đến server để giảm quá tải, và điều quan trọng hơn là tăng khả năng an toàn của server , tránh trường hợp lưu những thông tin bảo mật trên cùng một

server. Như vậy kết thúc quá trình enrollment người dùng sẽ gửi các thông tin như sau: B_{TC}, H(PIN) ở S_1 và S(serial number) ở S_2, và như ta đã thấy ở pha này chúng ta chỉ sử dụng một phương pháp bảo mật là Non-invertible transformation.

4.2.2. Authentication Phase

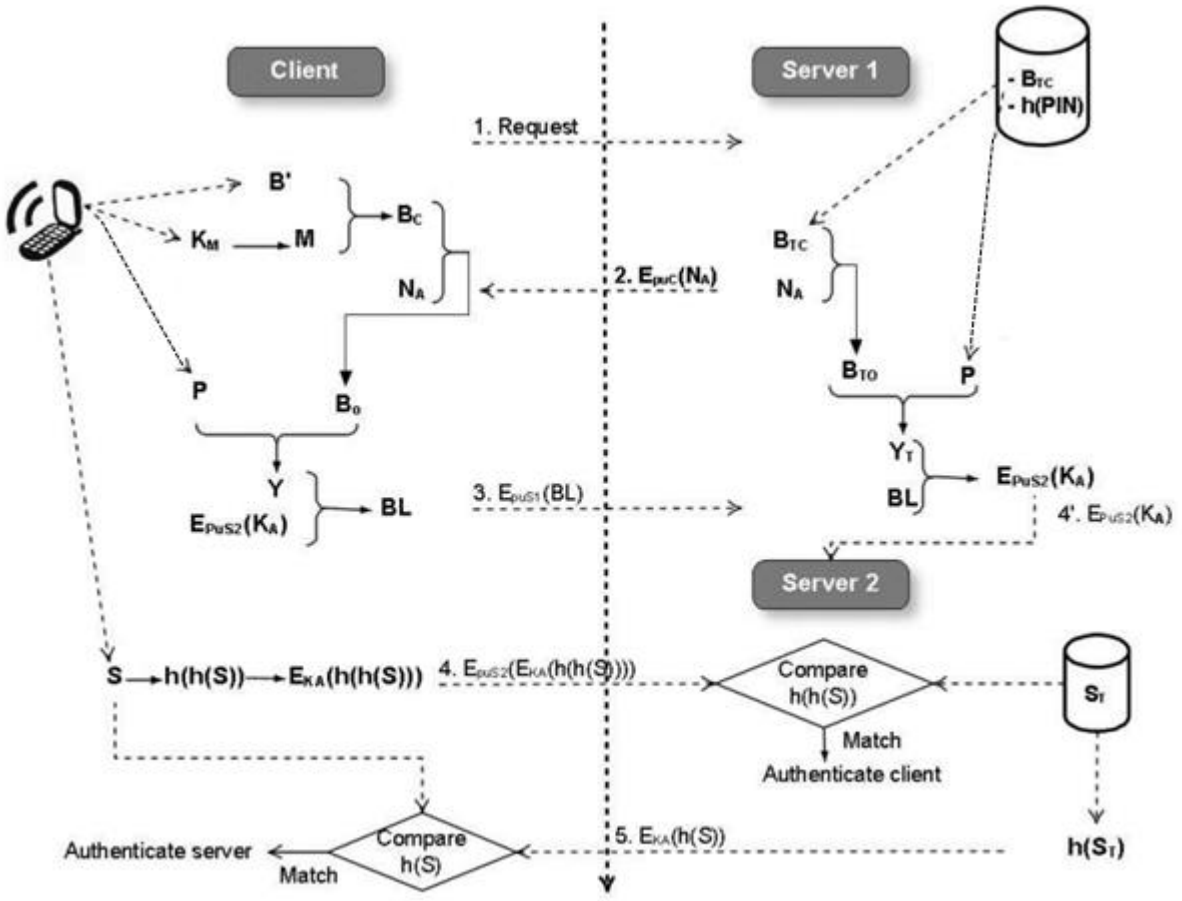
Trong pha này chúng ta sẽ áp dụng fuzzy-commitment để thực hiện việc xác thực từ xa dựa vào đặc trưng sinh trắc. Thay vào việc truyền trực tiếp giá trị biometric có thể đọc được trong đường truyền kém an toàn của hệ thống mạng như những hệ thống cũ, thì phía client sẽ gửi biometric-lock(BL) hay còn gọi là helper-data tới server. Tại phía server , BL sẽ được kết hợp với thành phần Y liên quan đến biometric của người dùng(Y có được là sau nhiều quá trình sử lý trên biometric của người dùng như cộng thêm một vector khác,...) có được lúc enrollment . Kết quả thu được sẽ là khóa xác thực của người dùng. Mô hình sẽ được biểu diễn bằng biểu đồ dưới đây.



Hình 21. Authentication based Fuzzy-commitment.

Bước thực thi: mục đích chính của quá trình này là truuyền **key** đến server để xác thực người dùng. Tại lúc này phía server đã có lưu B_{TC} sau đó sử dụng B_{TC} này qua các quá trình như hoán vị, và cộng thêm một số khác,..... sẽ có được Y về mặt ý nghĩa thì Y đại diện cho đặc trưng sinh trắc của người dùng nhưng đã được qua các bước bảo mật để có thể sử dụng. Bên phía client người dùng cũng sẽ cung cấp cho hệ thống biometric của họ, sau đó cũng qua các bước biến đổi tương tự như bên server và kết quả là Y'(nếu cùng một người thì Y'~Y), sau đó client sẽ kết hợp Y' với key để thu được BL, và BL này sẽ được gửi qua server , bên server sẽ lấy Y và kết hợp với BL để phục hồi lại key, và hệ thống sẽ dùng

key này cho pha xác thực ở S_2 và fuzzy-commitment này có sức mạnh ở chỗ có thể khôi phục lại khóa key nếu Y'~Y. Sau khi lấy được khóa key sẽ được chuyển qua S_2 thực hiện xác thực. Mô hình chi tiết sẽ được biểu diễn trong hình sau:



Hình 22. Authentication Phase.

Các bước cụ thể như sau: đầu tiên client gửi yêu cầu đến S_1, sau đó S_1 sẽ tạo ra một số ngẫu nhiên dùng một lần N_a, sau đó gửi lại phía client. Tất cả các thông tin liên lạc giữa server và client đều được mã hóa bởi hệ thống bất đối xứng(PKI- public Key infrastructure). Trong khoảng thời gian đó, client sẽ biến đổi biometric vào vùng an toàn bằng cách kết hợp với ma trận trực giao tạo ra B_C từ B' và ma trận trực giao M sẽ được sinh ra từ K_M. Sau đó B_C sẽ được kết hợp với N_A được gửi từ S_1 để tạo ra B_O, mục đích kết hợp vơi N_A là để chống lại replay attack. Sau đó B_O sẽ được thực hiện qua phép hoán vị P và tạo ra Y với phép hoán vị này đảm bảo độ bảo mật cho biometric. Sau đó Y sẽ kết hợp với khóa K_A đã được mã hóa của người dùng sẽ đi vào phương pháp fuzzy-

commitment và kết quả sẽ tạo ra BL sẽ được gửi sang phía S_1. Bên S_1 sau khi tạo ra N_A, thì đồng thời cũng lấy B_{TC} và cũng qua các quá trình tương tự như phái client như, kết hợp vơi N_A , qua phép hoán vị P để kết quả là Y_T, và sau khi nhận được BL từ client thì server sẽ kết hợp với Y_T để làm input cho quá trình fuzzy-commitment bên phái S_1, sau khi kết thúc quá trình này thì ta sẽ thu được giá trị mã hóa của khóa xác thực K_A để gửi qua S_2 thực hiện các phép so sánh như đã trình bày rất rõ trên hình trên(vì phần này đơn giản và mục đích của luận văn không làm phần này nên không trình bày rõ). Và sau khi xác thực được client , server sẽ gửi lại client $H(S_T)$ để phía client có thể xác thực lại server.

4.2.3. Ý tưởng bảo mật

- **Non-invertible Transformation**: tuy công việc của thành phần này khá đơn giản nhưng nó đóng vai trò cực kỳ quan trọng trong hệ thống. Đó là bảo vệ mẫu sinh trắc cho người dùng, chính vì vậy mà chức năng này có ở hai pha enrrollment và authentication. Sau khi lấy được sinh trắc của người dùng thì chức năng này thực thi ngay nhằm bảo vệ mẫu.
- **Fuzzy-commitment**: hệ thống này chỉ có ở pha authentication, đây là một hệ thống Biometric Cryptosystem, chính vì vậy nên nó đảm nhiệm chức năng bảo vệ khóa sử dụng đặc trưng, đây là thành phần cốt lõi của cả hệ thống, ý tưởng chính của hệ thống xác thực nằng ở chức năng này.

CHƯƠNG 5. HIỆN THỰC HỆ THỐNG

Một số yêu cầu trong thực hiện hệ thống:

- Ngôn ngữ: Java.
- Thư Viện ngoài:
  - Opencv.
  - Jai_cor.
  - Jai_corc.
  - Colt.
  - Jasypt.
- Công cụ lập trình: Eclipse.

Trong hệ thống này thì tôi có sử dụng lại mã nguồn của nhóm trước về phương pháp trích xuất đặc trưng sinh trắc người dùng PCA chính vì vậy nếu các thầy cô hay các bạn phát hiện có lỗi trong phần này thì mong bỏ qua, còn nếu phát hiện lỗi trong các phần còn lại thì hy vọng mọi người có thể góp ý để tôi hoàn thiệt tốt hơn, xin cảm ơn.

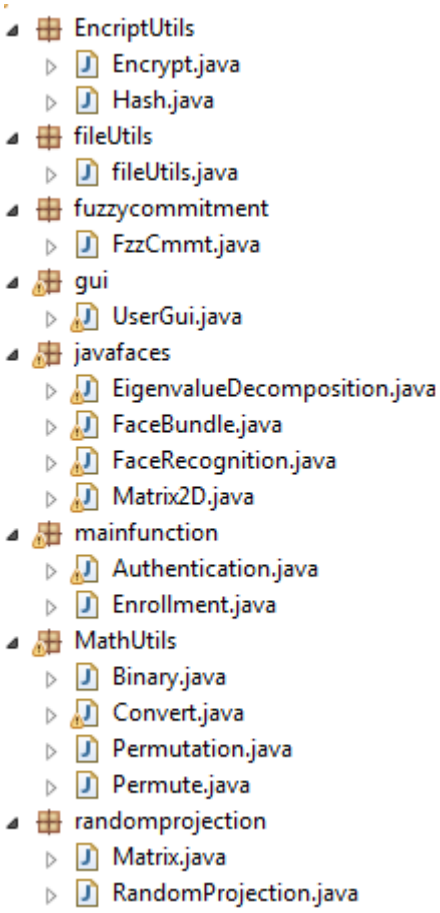
5.1. Cấu trúc mã nguồn

5.1.1. Project client

Trong cấu trúc bộ mã nguồn của client thì có 7 packet:

- EncriptUtils: chứa các class có chức năng hash và encrypt.
- fileUtils: chứa các class quản lý file(ghi và đọc file).
- Fuzzycommitment: hiện thực giải thuật fuzzy-commitment.
- Gui: giao diện chương trình.
- Javafaces: trích xuất đặc trưng sinh trắc của người dùng.
- Mainfunction: hai chức năng chính của chương trình enrollment và authentication.
- MathUtils: thực hiện các phét toán cơ bản.
- Randomprojection: hiện thực giải thuật Non-invertible transform.

Cấu trúc:

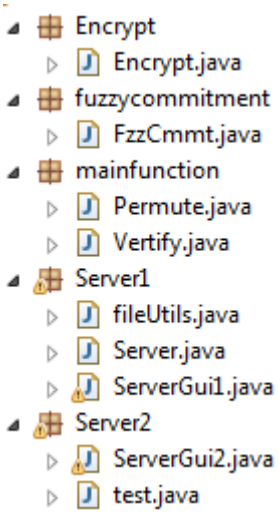


5.1.2. Project server

Trong cấu trúc bộ mã của server thì gồm ba packet chính:

- Fuzzycommitment: hiện thực giải thuật fuzzy commitment ở phía server.
- Mainfunction: thực hiện công việc authentication và enrollment.
- Server1: chứa giao diện và giao thức kết nối với client và thực hiện các chức năng của S_1.
- Server2: chứa giao diện và chức năng của S_2.
- Encrypt: chứa các class làm nhiệm vụ hash và mã hóa.

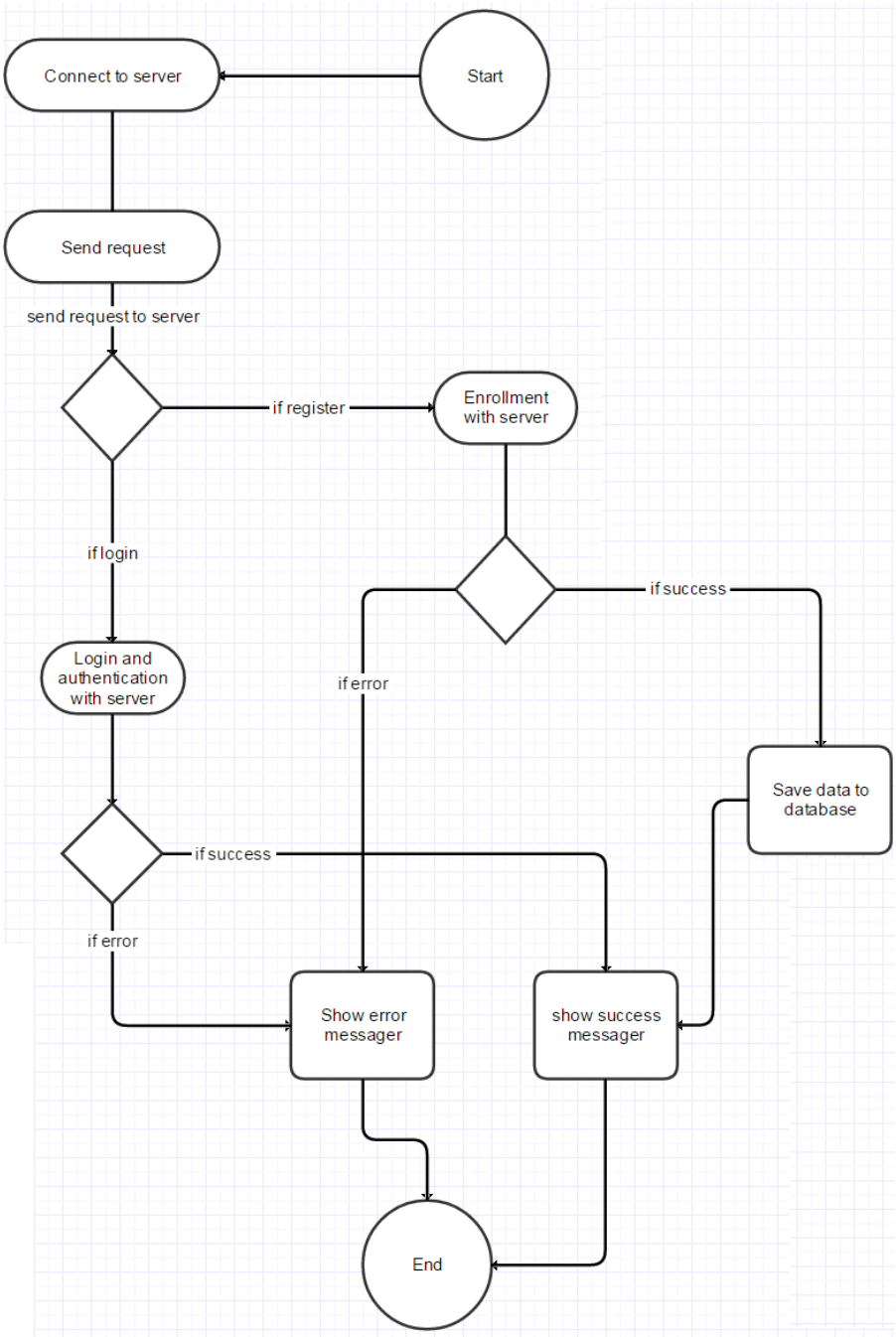
Cấu trúc:



5.2. Các thành phần của hệ thống

5.2.1. Client flowchart

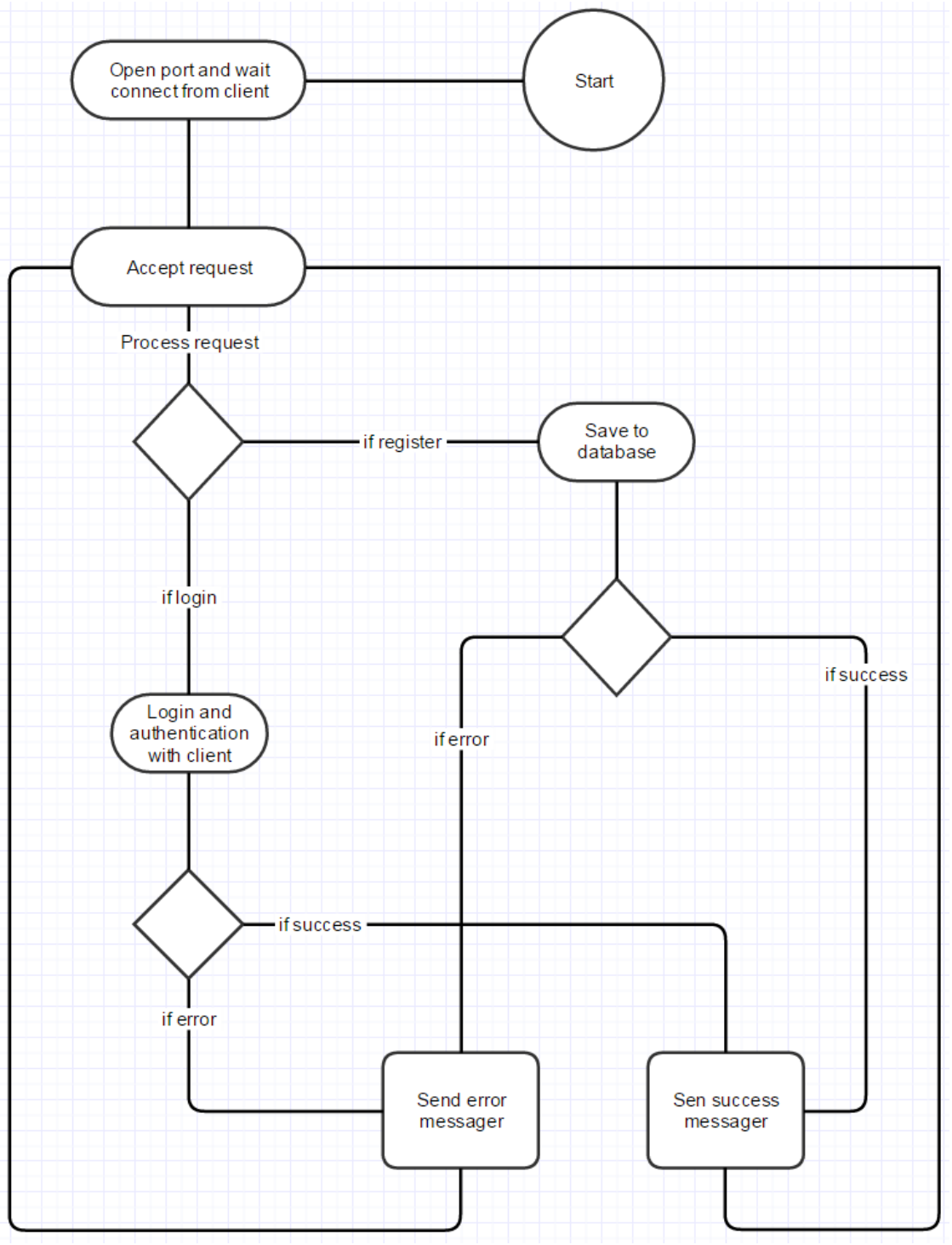
- Bước 1: kết nối với server.
- Bước 2: tùy theo yêu cầu người dùng mà client gửi yêu cầu enrrollment hay authentication tới server.
- Bước 3: nhận thông tin từ server và thông báo cho người dùng và lưu một số thông tin cần thiết.



Hình 23. Luồn thực thi client

5.2.2. Server flowchart

- Bước 1: khởi tạo server và các kết nối TCP/IP và đợi client kết nối.
- Bước 2: tùy thuộc vào yêu cầu từ phía client mà server sử lý yêu cầu cho client.
- Bước 3: phản hồi lại client và gửi một số thông tin về phái client.



Hình 24. Luồn thực thi server.

5.2.3. Chức năng tạo ma trận trực giao

Dựa theo mô hình của ma trận trực giao để chúng thực hiện chức năng này với các bước như sau:

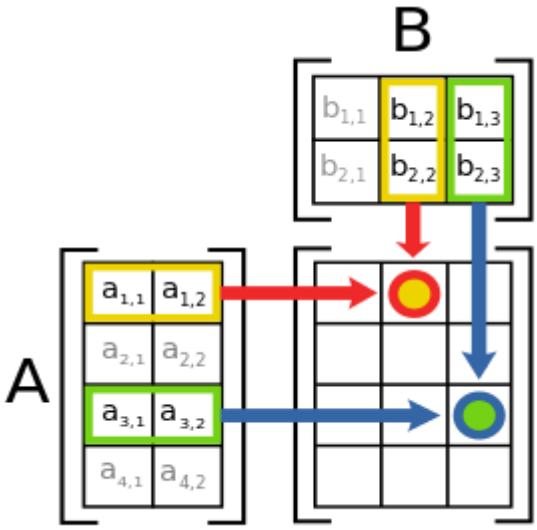
- Bước 1: tạo K_M bằng việc random một mảng một chiều có độ dài n giả sử vector sinh trắc có độ dài là 2n.

– Bước 2: theo cấu trúc của mảng một chiều mà thực hiện tính toán sao cho phù hợp.

Chức năng tạo ma trận trực giao là chức năng khởi đầu của hệ thống, các phép toán sử dụng không quá phức tạp, chủ yếu là phép sin, cos toán học thông thường và hàm random, hàm này có trong thư viện của chương trình nếu không xét đến độ phức tạp của các hàm sin hay cos thì độ phức tạp của việc tạo ma trận là O(n) do chỉ cần tạo đường chéo của ma trận.

5.2.4. Chức năng chuyển đổi đặc trưng(Non-invertible Transform)

Thật ra hàm F chỉ là phép nhân hai ma trận lại với nhau theo công thức: F x->y, y = A.x với x là đặc trưng sinh trắc gốc, A là ma trận trực giao đã tạo bước trên. Phép nhân hai ma trận này cũng rất đơn giản, nhờ vào điều đặc biệt của ma trận mà độ phức tạp của chương trình chỉ ở O(2n) vì các phần từ ngoài đường chéo đều bằng không, nên khi thực thi chương trình chúng ta chỉ quan tâm các phần tử trên đương chéo.



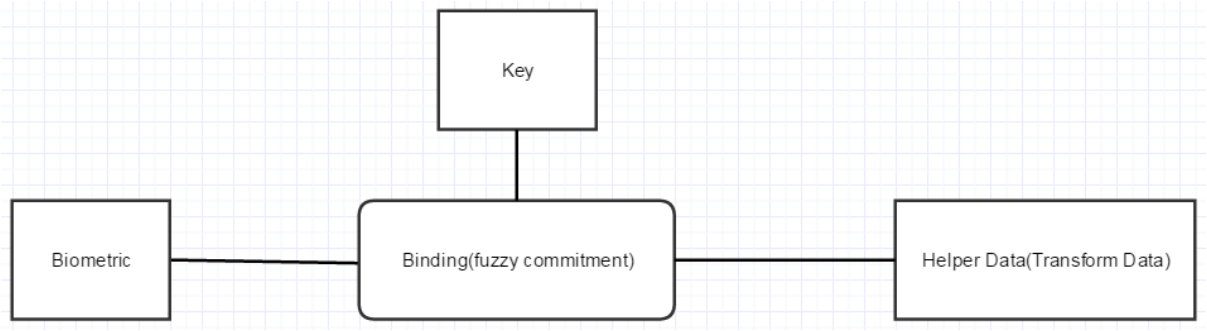
Hình 25. Ví dụ nhân hai ma trận 2x2.

5.2.5. Chức năng sinh khóa cho fuzzy-commitment

Vì đây là khóa chỉ sử dụng một lần nên việc sinh khóa cực kỳ đơn giản, sử dụng hàm ramdom của hệ thống. Công việc sinh khóa cho fuzzy-commitment rất đơn giản, chỉ cần sử dụng hàm random để sinh ra mảng một chiều với độ dài bằng với độ dài vector đặc trưng.

5.2.6. Chức năng binding

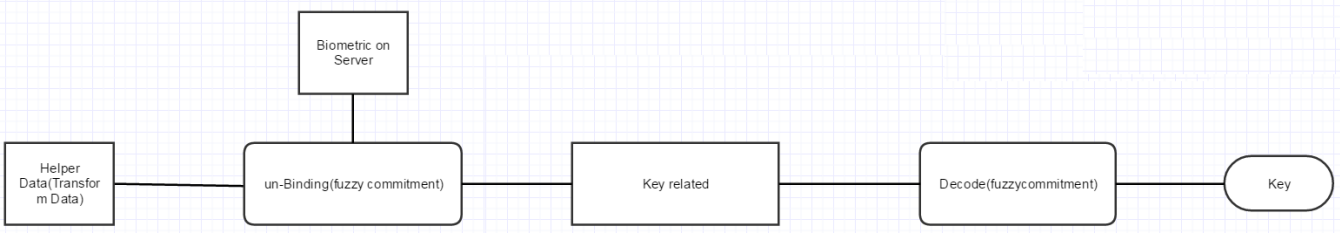
Với chức năng này thì đầu vào của nó gồm khóa K đã sinh ở trên và vector đặc trưng sinh trắc của người dùng sau đó qua hàm binding thì sẽ cho ra kết quả. Đối với dữ liệu là dạng binary thì hàm binding là hàm XOR, còn đối với dữ liệu là kiểu integer thì hàm binding này là phép Cộng.



Hình 26. Chức năng binding

5.2.7. Chức năng decode trong fuzzy-commitment

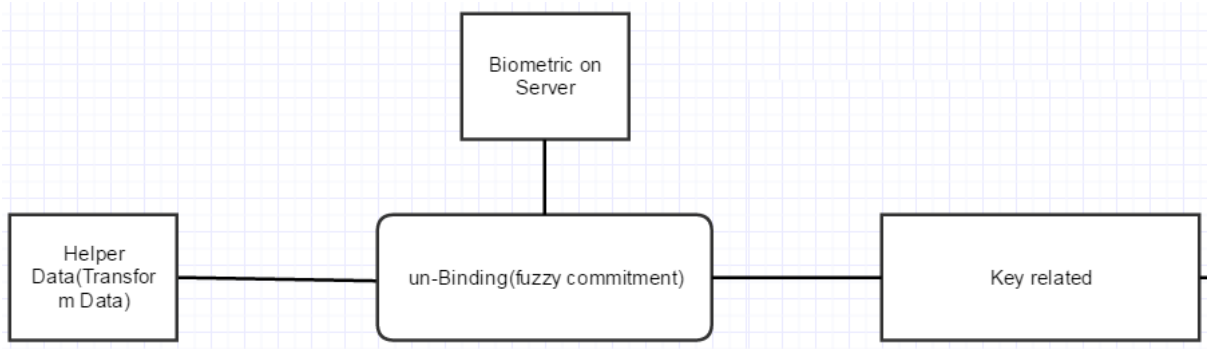
Đây là chức năng quan trọng nhất của fuzzycommitment, đầu vào của chức năng này là dữ liệu nhận từ phái client, sau đó tiếp tục thực hiện chức năng unbinding và kết quả của hàm này sẽ là một Key chưa hoàn chỉnh, muốn có một Key hoàn chỉnh chúng ta phải qua một bước nữa là decode. Dưới đây là sơ đồ hiện thực của giải thuật.



Hình 27. Chức năng khôi phục key

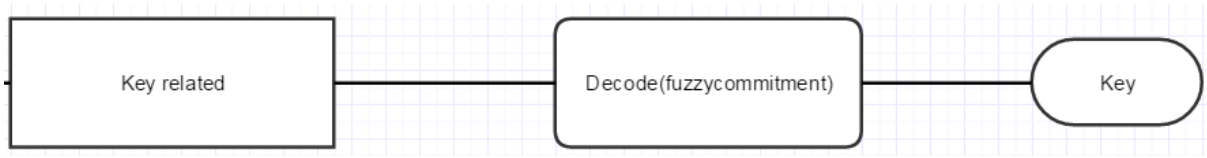
Để cho dễ hình dung thì chức này có hai chức năng con:

Un-bindind: sẽ thu được Key-related.



Hình 28. Chức năng un-binding

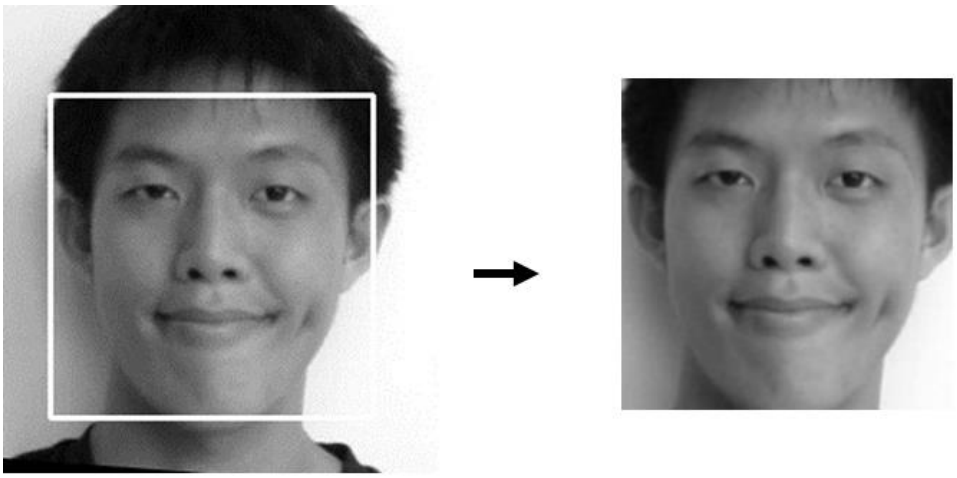
Decode: nếu thành công sẽ thu lại được Key của client lúc gửi.



Hình 29. Chức năng decode

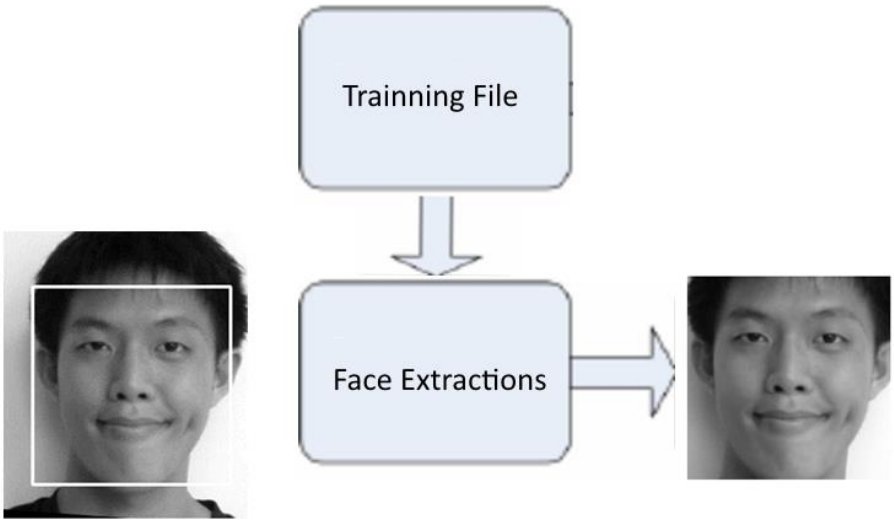
5.2.8. Chức năng trích xuất khuôn mặt

Để tăng khả năng tương tác với người dùng và dễ dàng trong sử dụng thì tôi thêm chức năng trích xuất khuôn mặt cho hệ thống, với chức năng này không những làm dễ dàng hơn cho người sử dụng mà còn làm cho hiệu suất của hệ thống tăng lên đáng kể.



Hình 30. Face Extractions.

Mô hình hoạt động của hệ thống này dựa trên chức năng detect object của opencv và kết hợp với tập training, đây là một trong những chức năng phân mảnh trong xử lý hình của opencv, dưới đây là mô hình của nó:



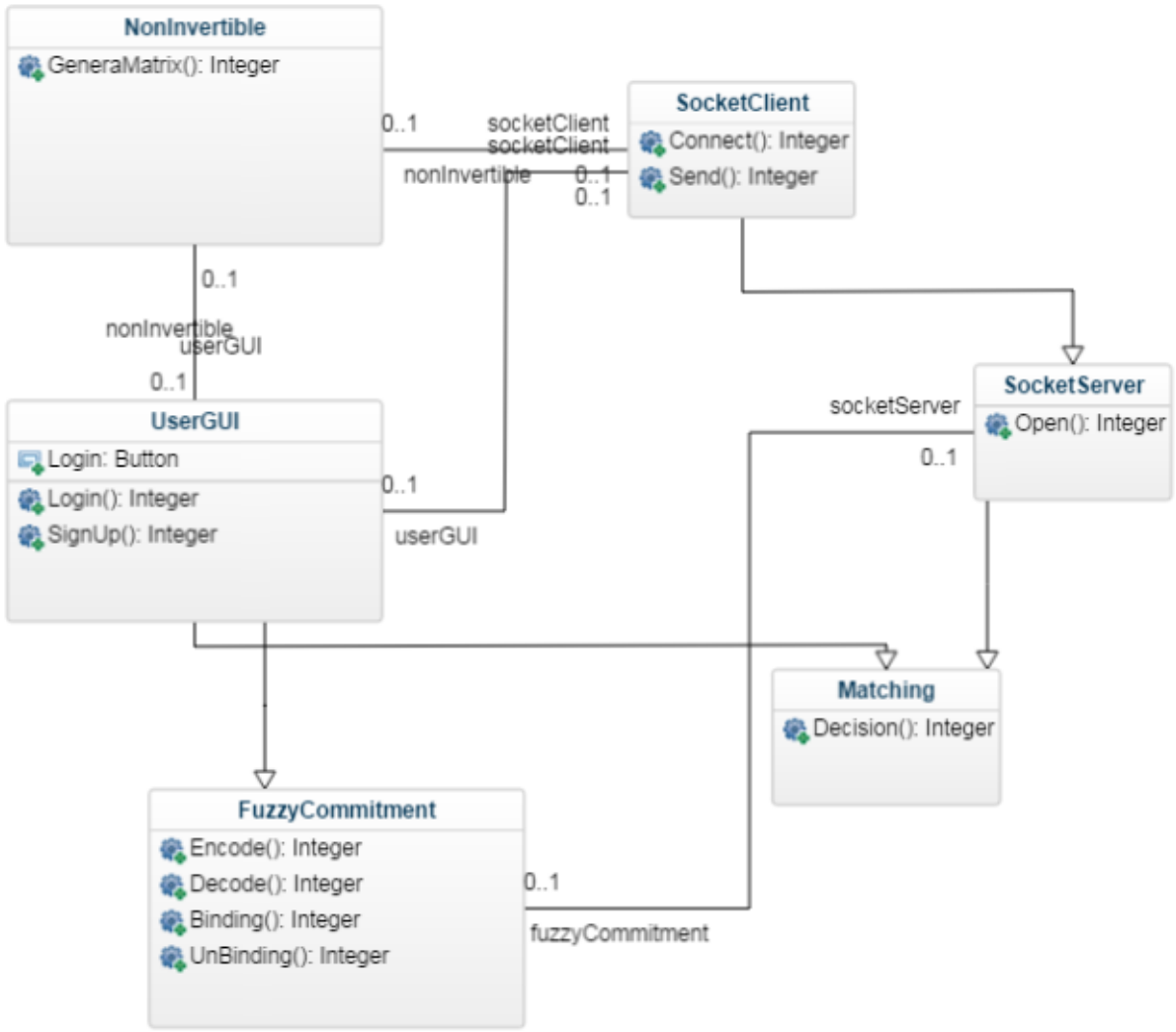
Hình 31. Face extraction diagram.

5.2.9. Cơ sở dữ liệu

Để đơn giản thì cơ sở dữ liệu sẽ lưu dưới dạng file text: bên phía client sẽ lưu lại KM sẽ có dạng một dòng cách nhau bởi dấu cách.

Còn bên server sẽ lưu lại sinh trắc của người dùng dưới dạng một dòng, cách nhau bởi dấu cách.

5.3. Các lớp chính trong chương trình



Hình 32. Các lớp chính và sự tương tác giữa các lớp trong chương trình.

Trong chương trình, ứng với mỗi chức năng sẽ có mỗi lớp đảm nhiệm chức năng đó, ngoài ra còn một số chức năng như mã hóa, Hash,.. tuy nhiên đó không phải là mục tiêu của đề tài nên không đề cập ở đây.

CHƯƠNG 6. ĐÁNH GIÁ VÀ THỬ NGHIỆM

6.1. Độ bảo mật của hệ thống

Với mô hình của một hệ thống như trên thì khả năng xác thực của hệ thống sẽ phụ thuộc vào đầu vào của các giá trị như sau:

- Đặc trưng sinh trắc của người dùng.
- Số N_A, số này được server sinh ra và sử dụng một lần.
- K_M để sinh ra ma trận trực giao.

- PIN dùng trong phép hoán vị.
- Mobile serial number S.

Dựa vào các thành phần trên chúng ta phân tích khả năng bảo mật của hệ thống, một hệ thống được gọi là có độ bảo mật cao khi đáp ứng được các yếu tố sau:

6.1.1 Biometric template attack

Như đã trình bày đặc trưng sinh trắc gốc của người dùng được bảo vệ bởi phép biến dổi một chiều (non-invertible transformation), và server sẽ lưu giá trị sau khi đã chuyển vào vùng an toàn thay vì sử dụng đặc trưng gốc, chính vì vậy nên không có khả năng phục hồi lại đặc trưng của người dùng với những gì đã lưu trên server. Và phương pháp này còn có một điểm mạnh nữa là khả năng hủy bỏ(revocability) giá trị biometric đang lưu trên server với chỉ một thao tác cực kỳ đơn giản là thay đổi K_M sau đó đăng ký lại với cùng người đó và khi đó giá trị bảo mật sẽ hoàng toàn thay đổi với cùng một đặc trưng của người dùng, việc này giống như thay passwords trong hệ thống truyền thống. Và để tăng khả năng bảo mật của hệ thống này trên biometric của người dùng, chúng ta còn áp dụng kỹ thuật hoán vị P, với kỹ thuật này thì mỗi lần lấy ra biometric lại hoàn toàn khác nhau, chính vì vậy việc truy vấn ngược lại đặc trưng của người dùng là không thể.

6.1.2 Replay attack

Replay attack là kiểu tấn công mà kẻ gian sử dụng lại giữ liệu cũ của lần đăng nhập trước đó để gửi lên server. Tuy nhiên kiểu tấn công lại bị ngăn chặn bởi việc dùng số N_A và session key K_A là những số chỉ sử dụng một lần. Như vậy với việc sử dụng kỹ thuật này thì kẻ gian không thể giả dạng người dùng được vì các giá trị BL luôn khác nhau rất nhiều giữa những lần đăng nhập khác nhau cho dù gí trị B có giống hoàn toàn đi chăng nữa. Tương tự với việc K_A chỉ dùng một lần thì cho dù kẻ gian có lấy được K_A của lần đăng nhập trước cũng không còn giá trị trong lần đăng nhập này.

6.1.3 Main-in-the-middle attack

Với phương pháp tấn công này thì không hiệu quả vì mô hình ở trên yêu cầu xác thực từ hai phía chứ không phải từ một phía. Và kết hợp với các phương pháp bảo mật trên đường truyền với việc dùng các số Key hoặc N_A chỉ dùng một lần nên kẻ tấn công không thể thay đổi hay chỉnh sữa được dữ liệu trên đường truyền.

6.1.4 Insider attack

Với kiểu tấn công này thì đã được giải quyết bằng phương pháp chia ra làm hai server: S_1 thì tuy lưu thông tin của người dùng nhưng chỉ làm nhiện vụ nhận khóa xác từ client, S_2 tuy không lưu thông tin người dùng nhưng làm nhiệm vụ xác thực, với hai chức năng hoàn toàn độc lập từ hai server khác nhau thì việc admin tấn không là không thể. Không những vậy các thông tin liên quan đến người dùng đều được bảo vệ ở mức cao nên cho dù admin có lấy được cũng không khai thác được gì.

6.2. Độ phức tạp của hệ thống

Vì hệ thống chỉ tập trung vào các giải thuật trong việc bảo vệ và xác thực người dùng nên chỉ liệt kê độ khó của hai giải thuật sử dụng trong hệ thống là fuzzy-commiment và non-invertible còn các giải thuật khác như Hash, Encryption,.... thì xin được bỏ qua. Dưới đây là bảng thống kê độ phức tạp của từng giải thuật.

6.2.1. Phương pháp Non-invertible

Bảng 3. Độ phức tạp của Non-invertible.

Bước	**Số bước tính toán**	**Độ khó giải thuật.**
1. tạo ra tập ***n*** giá trị ngẫu nhiên $\{\theta_1, \theta_2, \ldots. \theta_n\}$	Bước này nhằm tạo ra tập n giá trị ngẫu nhiên trong khoảng [0..2π] để chuẩn bị cho bước tạo ma trận trực giao. Trong tính toán thì ta cần dùng hàm random và hàm này sẽ chạy n lần.	**O(n).**
2. tạo ma trận trực giao.	Ma trận trực giao có dạng như hình (1) như vậy số lần tính hàm sin và hàm cos là n lần.	Gọi độ phức tạp của hàm sin là O(s) và độ phức tạp của hàm cos là O(c) thì độ phức tạp lúc này là : **O(nc)+O(ns) = O(2n.s) = O(2n.c).** Ví dụ hàm sin,cos được tính theo giải thuật "Arithmetic-geometric mean iteration" thì ta có s=c=$O(M(k) \log k)$ với k là số chữ số của số cần tính sin,cos và O(M(k)) là độ phức tạp khi chọn giải thuật để tính toán. Như vậy khi đó độ khó sẽ là: O(2n.M(k).log(k)). Có thể chọn k=10.
3. thực thi non-invertible	Đây là phép nhân ma trận tạo ra ở bước 2 với vector sinh trắc của người dùng. Gọi M(2n x 2n) là ma trận trực	Theo công thức độ phức tạp khi nhân hai ma trận là **$O(4n^2)$** nhưng M là ma trận trực giao chỉ có đường chéo

	giao, X(2nxp) là vector đặc trưng sinh trắc(đa số các vector đặc trưng sinh trắc người ta sẽ chuẩn hóa để p=1 cho dễ tính toán).	là các phần tử khác 0 nên với mỗi phần tử trong vector kết quả thì ta cần 2 phép nhân và một phép cộng nên độ phức tạp thực sự là: **O(n.(2k$^{1.465}$+log(k)))**.(với k là số chữ số của con số trong phép tính ,giả sử phép nhân được tính theo giải thuật của "3-way Toom–Cook multiplication"). Tùy theo giải thuật xác thực hay độ lệch chấp nhận của giải thuật fuzzy commitment thì chúng ta có thể chọn k = 10.

6.2.2. Phương pháp Fuzzy-commitment

Bảng 4. Độ phức tạp của fuzzy-commitment.

Bước	**Chi tiết**	**Độ khó**
1.thực hiện phép XOR		O(n).
2.thực hiện Decode	Sử dụng linear error correcting code để decode, với mỗi phần tử thì nó chạy giải thuật **Nearest Neighbor Algorithm** giải thuật này có một vòng lặp nhỏ với độ lặp tối đa là 2.delta với delta là độ sai cho phép của giải thuật.	Như vậy với mỗi phần tử thì ta sẽ dùng 1 phép trừ, 1 phép cộng , một hàm min, một hàm max và một vòng lặp có sử dụng phéo so sánh phía trong như vậy có thể tính được độ khó của nó như sau:O(2log(k)+2k+2.delta.log(k)), và giải thuật này áp dụng trên 2n phần tử của vector vậy độ khó tổng cộng là O(n(2log(k)+2k+2.delta.log(k))).
3.hàm hash	Theo một số tài liệu.	O($2^{n/2}$)

6.3. Độ hiệu quả của hệ thống

Về phần kết quả và thống kê xem hiệu suất của chương trình và mô hình này hiệu quả đến đâu thì nhóm sẽ test trên tập dữ liệu gồm 153 người và mỗi người có 20 ảnh, và phương pháp tìm độ lệch giữa các ảnh với nhau là có chút khác biệt so với hệ thống khác(euclidean distance, city block,...) chính vì vậy nên trước khi bước vào hiện thực thì nhóm cần test qua khả năng đáp ứng của phương pháp PCA này rồi sau đó mới hiện thực hoàn chỉnh chương trình. Chính vì vậy nên trong phần thống kê này

có thêm một phần nữa là 'đánh giá PCA' để xem liệu rằng phương pháp trích xuất đặc trưng sinh trắc của người dùng có phù hợp hay không. Về tập test và tập train thì hệ thống sẽ dùng giống với phương pháp đánh giá PCA sẽ được nên rõ trong phần này ngay phái dưới.

6.3.1. Đánh giá PCA

Như ta đã biết phương pháp với PCA thì đầu vào của nó là một ảnh và sau đó sinh ra một vector đặc trưng sinh trắc của ảnh đó. Và vector này luôn có độ dài là 200 và mỗi thành phần của vector nằm trong khoảng [0,1]. Và phương pháp trích xuất này là của nhóm trước hoàn thiện và họ sử dụng độ lệch Euclid giữa hai vector nhưng trong luận văn này thì phương pháp fuzzy-commitment sẽ sữa lỗi trên từng thành phần nên trong luận văn này nhóm đề xuất phương pháp tính độ lệch khác nên trước khi bắt đầu hoàn thiện chương trình cần phải đánh giá PCA này có phù hợp hay không. Như vậy điều kiện cần để có thể sử dụng phương pháp này trong luận văn là phải đảm bảo với một ảnh sẽ sinh ra một vector và nếu cùng một người thì độ lệch giữa các vector là không được quá lớn, thì đây chính là điều kiện tối thiểu mà phương pháp PCA cần phải thỏa mãn để có thể áp dụng được vào chương trình. Thì với phương pháp PCA này cần hai tập ảnh là tập train và tập test.

Tập train: để trainning PCA thì tôi sử dụng tập ảnh gồm 50 ảnh, trong đó có 9 người Châu Âu mỗi người 2 ảnh, 21 người Tây và Trung Á mỗi người 1 ảnh và 10 ảnh của sinh viên trường đại học BK Tp.HCM.

Tập test: như đã nói ở trên tập test gồm 153 người và mỗi người 20 ảnh.

Phương pháp thực hiện: như vậy mục đích của chúng ta là kiểm tra xem với cùng một người liệu rằng nó có thể sinh ra những vector sinh trắc có độ lệch nhỏ hay không, để thực hiện phép thống kê ta sẽ thực hiện công việc sau:

- Mỗi vector đặc trưng sinh trắc sẽ có N thành phần (với PCA thì với mỗi vector sẽ có 200 thành phần) và trong bài luận văn này tôi đề xuất phương pháp tính độ lệch giữa hai vector là giá trị độ lệch lớn nhất giữa các thành phần với nhau tức là:

- Gọi $X[x_1, x_2, x_3, x_4, \ldots\ldots x_N]$ và $X'[x'_1, x'_2, x'_3, x'_4, \ldots\ldots x'_N]$ lần lược là hai vector đặc trưng sinh trắc.

khi độ lệch giữa hai vector là:

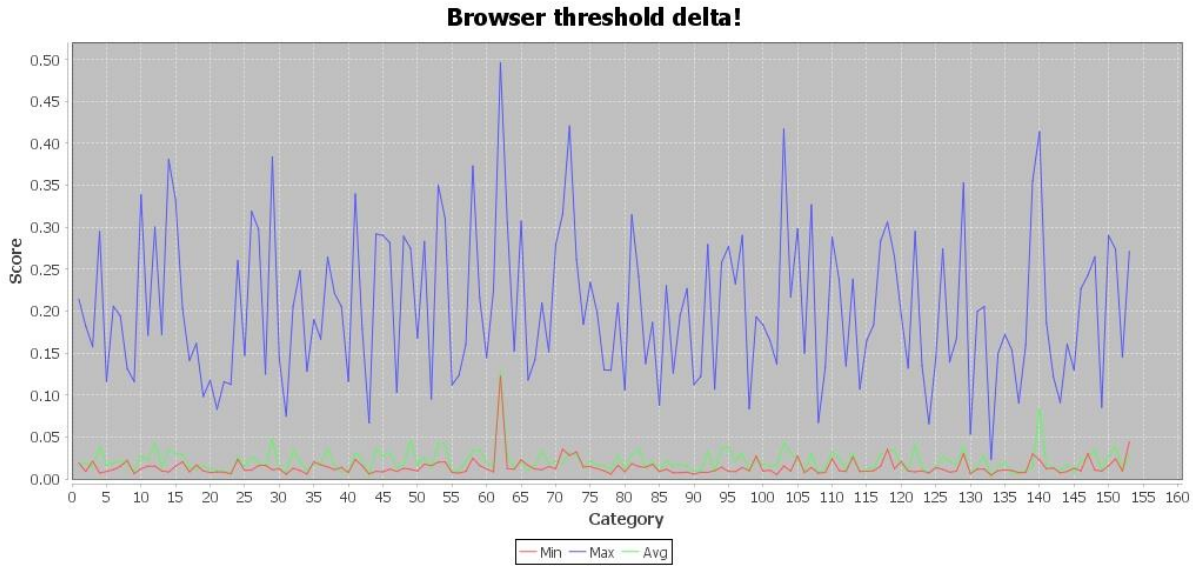
$$||X-X'|| = MAX((x_1 - x'_1), (x_2 - x'_2), (x_3 - x'_3), \ldots, (x_N - x'_N)) = (x_k - x'_k)$$

Giả sử tại thành phần thứ k thì đạt độ lệch lớn nhất. trong tập test chúng ta có 153 người , mỗi người 20 hình tức là mỗi người sẽ có 20 vector sinh trắc , chúng ta sẽ tính ra tập độ lệch giữa mỗi vector với các vector còn lại sau đó chúng ta lấy ra các giá trị **min**, **max** và **avg** với:

min: giá trị nhỏ nhất của các độ lệch.

max: giá trị lớn nhất của các độ lệch.

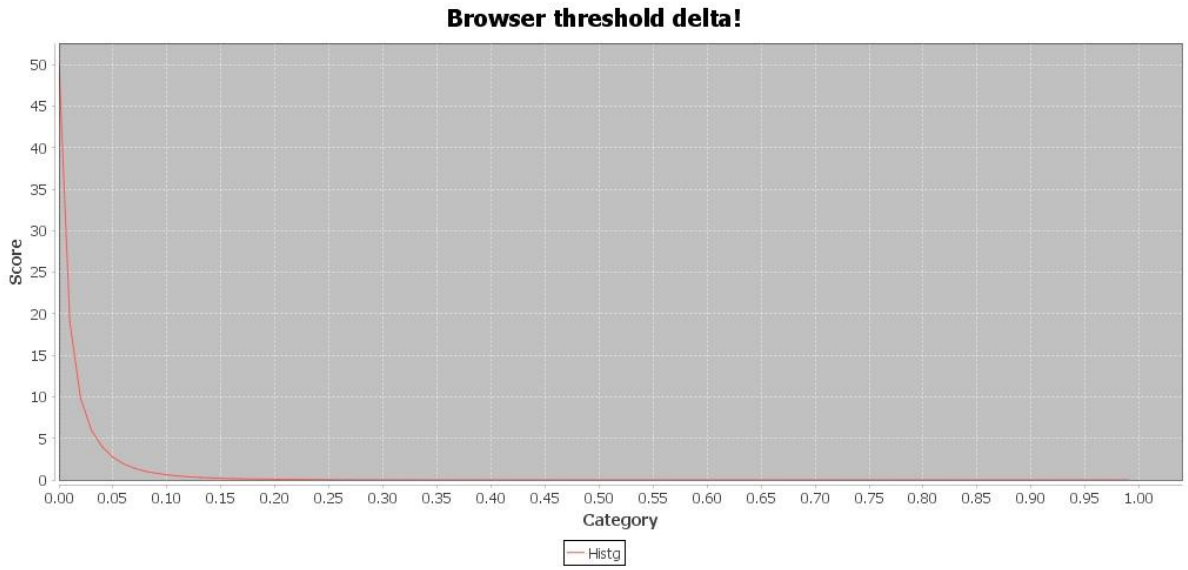
avg: giá trị trung bình giữa các độ lệch. Dưới đây là kết quả của 153 người.



Hình 33. Thống kê độ tin cậy của PCA

Từ đồ thị trên ta có thể rút ra nhận xét: giá trị **avg** và giá trị **min** khá nhỏ và đa số nhỏ hơn 0.05 và hai đường **min** và **avg** nằm sát nhau và tách xa với giá trị của đường **max** điều này chứng tỏ vector sinh trắc được sinh ra từ phương pháp PCA với cùng một người có độ lệch rất thấp và xác xuất sinh ra vector có độ lệch cao là rất ít.

Và để đảm bảo phương pháp này là đủ khả năng cho hệ thống thì tôi làm thêm một thống kê nữa, thống kê này giống như Histogram trong xử lý ảnh , phương pháp này sẽ vẽ ra đồ thị độ lệch của hai vector và số lượng độ lệch tập trung ở đâu là nhiều nhất:



Hình 34. Histogram PCA.

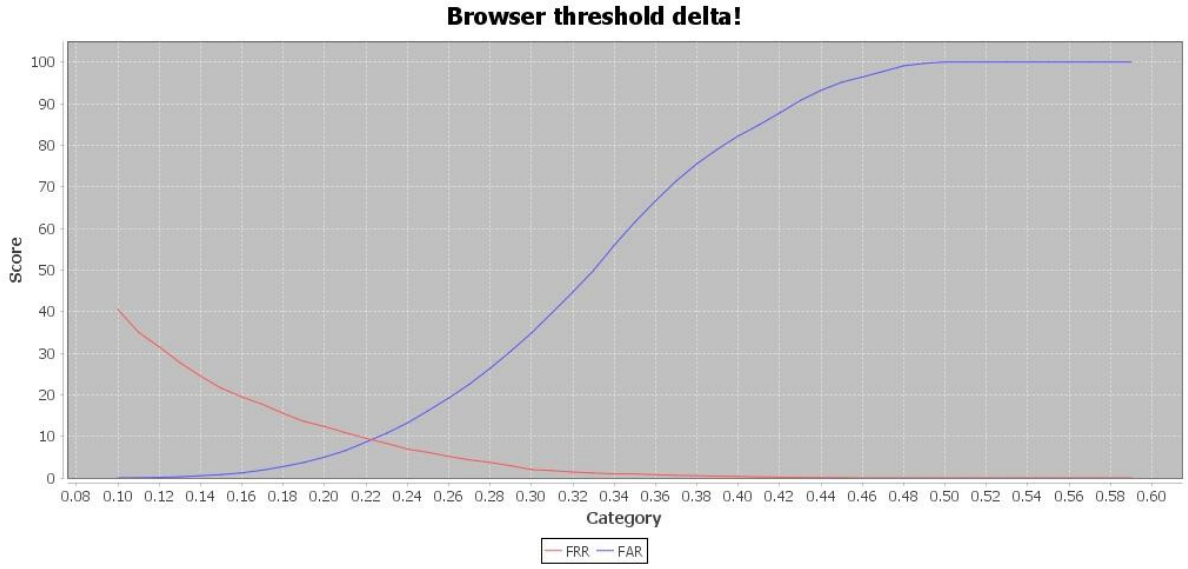
Nhận xét: với đồ thị trên ta có thể thấy với phương pháp PCA thì độ tương đương của hai vector là rất cao , với độ lệch nhỏ hơn 0.05 thì hơn 80% số lượng vector có độ lệch này. Vậy với phương pháp thống kê sẽ cho chúng ta thấy được sự tương thích của phương pháp PCA khi áp dụng phương pháp tính độ lệch mới, như vậy chúng ta có thể sử dụng phương pháp trích xuất này để hiện thực hệ thống, và từ hai biểu đồ trên chúng ta có thể dự đoán được giá trị threshold của hệ thống sẽ nằm trong khoảng [0.05, 0.3], và để đảm bảo cho việc lấy giá trị threshold này thì chúng ta sẽ thực hiện tính FRR và FAR sẽ được tính dưới đây.

6.3.2. Kiểm nghiệm hệ thống.

- **FAR (False Acceptance Rate)** là tỉ lệ chấp nhận sai, chấp nhận một truy cập khi người truy cập không hợp lệ. Xác xuất kẻ mạo danh đăng nhập nhưng thành công.
- **FRR (False Reject Rate)** là tỉ lệ từ chối bị sai, từ chối một truy cập khi người truy cập hợp lệ. Xác suất khách hàng đăng nhập nhưng bị từ chối.
- **EER (Error Rate)** là tỉ lệ lỗi, giao điểm của 2 đường FAR và FRR. Tại đó tỉ lệ chấp nhận bị sai bằng tỉ lệ từ chối bị sai.

Tập test là 153 người và mỗi người có 20 ảnh tuy nhiên chỉ có khoảng 4,5 kiểu khuôn mặt khác nhau nên tôi chỉ lấy ảnh thứ 1(với index là 0) làm ảnh enrrollment và các ảnh với thứ tự lần lược là 1,3,6,8,15,17 làm ảnh lúc vertify.

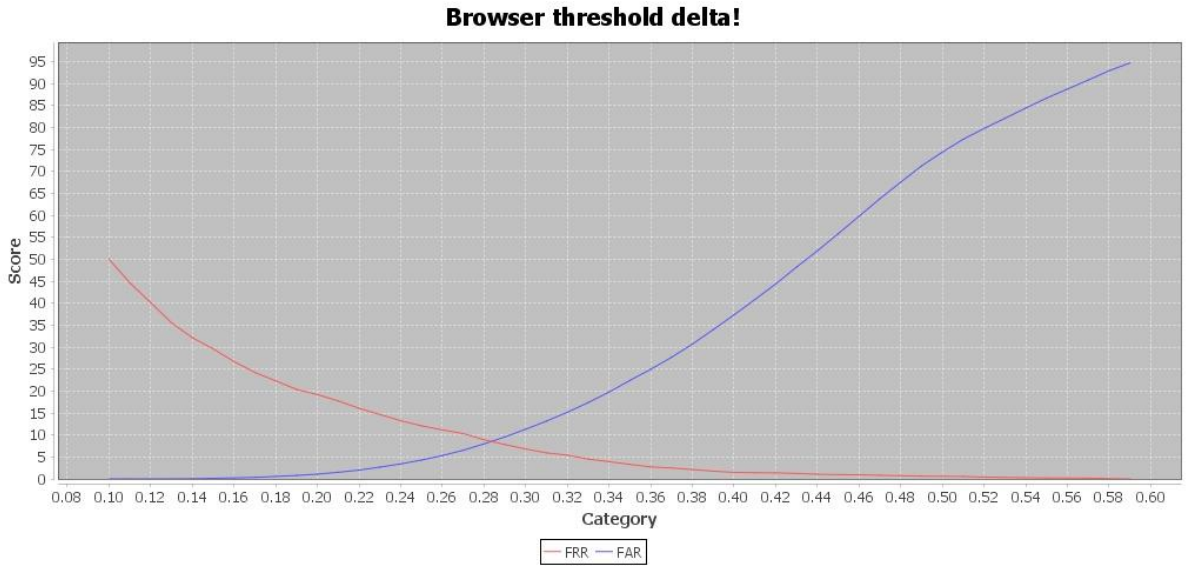
6.3.2.1. Khả năng trích xuất đặc trưng



Hình 35. FRR và FAR của PCA.

Nhận xét: chúng ta có thể thấy giá trị threshold là khoảng 0.22 và tại đây giá trị sai số của hệ thống khoảng 9%, như vậy có thể nói phương pháp trích xuất này khá tốt.

6.3.2.2. Kiểm nghiệm Non-invertible Transformation.

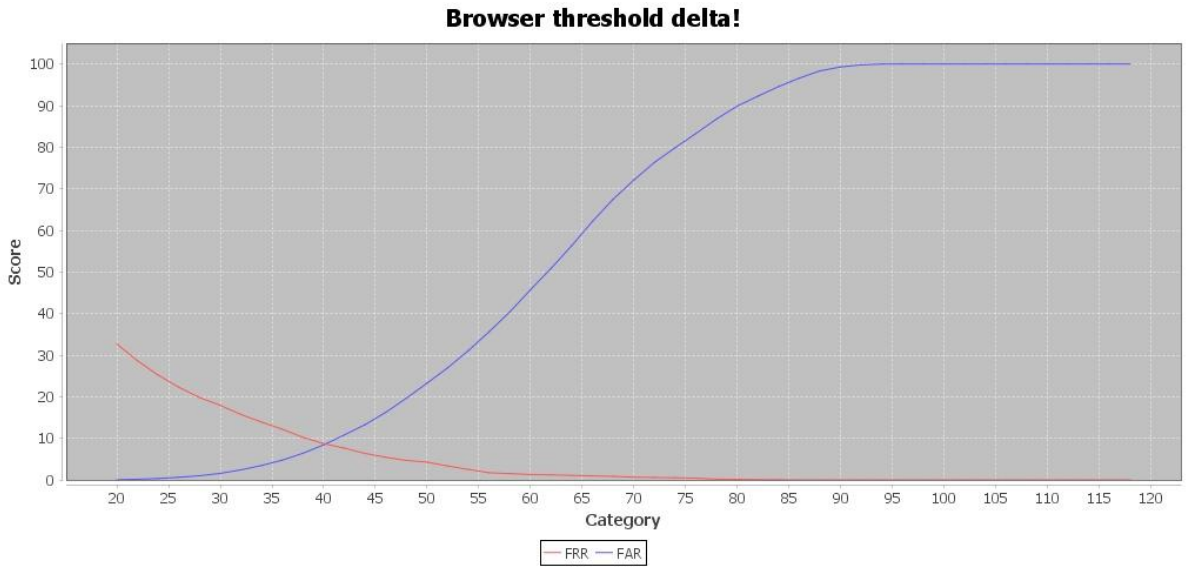


Hình 36. FRR và FAR sau khi biến đổi Non-invertible.

Nhận xét: sau khi biến đổi Non-invertible thì giá trị threshold thay đổi thừ 0.22 lên 0.29 thì điều này cũng rất dễ hiểu vì giá trị lúc đầu trong khoảng [0, 1] còn bây giờ giá trị nó đã lên [0, $\sqrt[2]{2}$], còn một điều rất đặc biệt sau khi biến đổi qua phép này là độ sai số của hệ thống đã giảm từ 9% xuống còn

8%, kết quả này trái với suy nghĩ bang đầu là sau khi qua các bước biến đổi thì hệ số sai số sẽ tăng lên nhưng không ngờ qua phép biến đổi này nó lại làm cho hệ số này thấp xuống và làm cho hệ thống tốt hơn, việc này có thể giải thích do phép nhân hai ma trận.

6.3.2.3. Kiểm nghiệm fuzzy-commitment và toàn hệ thống



Hình 37. FRR và FAR của toàn hệ thống.

Nhận xét chung: với vector sinh ra từ phương pháp PCA và sau khi sử dụng phương pháp nhân với ma trận trực giao và phương pháp fuzzycommitment thì điểm giao giữa hai đường FRR và FAR đều khoảng 9% điều này chứng tỏ phương pháp này cho ra kết quả khá tốt với khả năng đúng khoảng 91%(giá trị threshold thì còn phụ thuộc vào giá trị quatization nhưng kết quả này sẽ không thay đổi, trong hệ thống này tôi lấy giá trị quatization là 200), giá trị threshold khi chưa qua bất kỳ phương pháp bảo mật nào thì khoảng 0.22 và độ sai khoảng 10%, nhưng sau khi trải qua phương pháp nhân với ma trận trực giao(với độ scale của ma trận là $\sqrt[2]{2}$ thì giá trị threshold của nó tầm 0.285 = 0.22x$\sqrt[2]{2}$ nên giá trị 0.285 của threshold là dễ hiểu) nhưng độ sai của hệ thống khoảng 8%(rất tốt) không những sau khi nhân với ma trận trực giao làm tăng khả năng bảo vệ của hệ thống mà còn làm giảm độ sai của hệ thống(điều này có thể là do phép nhân hai mâ trận), còn sau khi kết hợp hai phương pháp nhân với ma trận trực giao và fuzzycommitment thì giá trị threshold khoảng 40(với độ sclare là 200 , 40 =

0.22x200) và độ lệch khoảng 9% tức là độ tin cậy của hệ thống là 91% -> khá tốt đối với hệ thống này.

CHƯƠNG 7. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

7.1. Kết luận

Trong bài luận văn này chúng tôi nghiên cứu một phương pháp có thể chống lại những loại tấn công khá tinh vi trên hệ thống. Mục đích chính của hệ thống này là sử dụng sinh trắc người dùng vào trong các hệ thống xác thực nhưng ở cấp độ bảo mật cao hơn và an toàn hơn cho người sử dụng. Và với sự kết hợp hoàn hảo của fuzzy-commitment và non-invertible transformation đã giúp chúng tôi có thể thiết lập một hệ thống có thể chống lại các loại tấn công trên các hệ thống sinh trắc hiện nay như biometric template attack, replay attack, main-in-midle attack. Điểm quan trọng nhất trong hệ thống này so với các hệ thống khác là khả năng chống lại tấn công bên trong(insider attack), trong trường hợp xấu nhất nếu người vận hành hệ thống(admintranslator) lấy những thông tin lưu trên server thì họ cũng không có khả năng truy ngược lại đặc trưng gốc của người dùng và hệ thống cũng được thiết kế để giảm tối thiểu sự can thiệp từ người quản trị để tránh loại tấn công khó ngừa này. Và nhờ sức mạnh của phương pháp Non-invertible transform giúp cho hệ thống có khả năng revocbility.

7.2. Hướng phát triển

Trong quá trình tìm hiểu và phát triển hệ thống tôi có một số ý tưởng để phát triển hệ thống như:

- Trong bài luận văn này tôi tập trung giới thiệu vào việc tạo ma trận trựng giao trong phép biến dổi Non-invertible Transform, tuy nhiên tôi muốn nhấn mạnh một lần nữa là Non-invertible Transform chỉ là ý tưởng chính vì vậy mà nó tùy thuộc vào mục đích của mỗi người mà chọn hàm biến đổi cho phù hợp(vì đặc trưng sinh trắc tôi nghiên cứu sử dụng độ lệc Euclid nên chọn phương pháp nhân với ma trận trực giao, nếu các bạn sử dụng các loại sinh trắc(vân tay, móng mắt,...) sử dụng độ lệch khác như Hamming kiểu Binary thì có thể phát triển nó dựa trên biến đổi CE(Coverage-Effort)[7]), còn đối với fuzzy-commitment thì sử dụng phép sữa lỗi trên kiểu Binary như Reed-Solomon.

Nếu đề tài phát triển tiếp theo hướng này thì có thể nói đây là một trong số ít những mô hình mà có thể áp dụng hầu hết các loại sinh trắc của con người.

- Một hướng phát triển nữa là áp dụng các mô hình bảo mật mới ví dụ như có thể thay phương pháp Non-invertible bằng phương pháp Salting hay có thể sử dụng những ý phát triển từ fuzzy-commitment như fuzzy-valt,.... Tùy thuộc vào hệ thống và khả năng chống lại các loại tấn công mà bạn có thể đề thay đổi cho phù hợp.
- Hướng phát triển theo tôi quan trọng nhất là phương pháp trích xuất đặc trưng. Thật sự mà nói hệ thống của tôi không thể đưa vào thực tiễn vì hệ số lỗi của toàn bộ chưa trình là 9%, cái này là do phương pháp PCA trích xuất vẫn còn chưa tốt và thời gian chạy của giải thuật này vẫn còn lâu. Nếu có thể giải quyết được vấn đề này thì mô hình này có thể đem ra thực tiễn và sử dụng một cách có hiệu quả.

CHƯƠNG 8. TÀI LIỆU THAM KHẢO

[1] Thi Ai Thao Nguyen, Dinh Thanh Nguyen and Tran Khanh Dang(2015). "A multi-factor Biometric Based Remote Authentication using Fuzzy Commitment and Non-invertible Transformation." In Proceeding of Infomation & Communication Technology-EurAsia Conference 2015, LNCS 9357, Springer-Verlag, October 4-7, 2015, Daejeon, Korea, pp. 77-88.

[2] Đề tài "Nhận dạng mặt người sử dụng đặc trưng PCA" – Vũ Mạnh Hùng (2013)

[3] Ari Juels and Martin Wattenberg (2013), *A Fuzzy Commitment Scheme*.

[4] Al-Assam, Hisham, Harin Sellahewa, and Sabah Jassin. "A liehtweight approach for biometric template protection." *SPIE Defense, Security, and Sensing*. Intenational Society for Optic and Photonics, 2009.

[5] Biometric Template Security of Anil K. Jain, Karthik Nandakumar and Abhishek Nagar.

[6] https://en.wikipedia.org/wiki/Error_detection_and_correction.

[7] Nguyen Thi Hoang Lan (2009), Hệ thống an ninh thông tin dựa trên sinh trắc học Bio-PKI, Đại học Bách Khoa Hà Nội.

[8] Yang G, Huang T S. Human (1994). *Face detection in complex background*. Pattern Recognition, pp 53-63

[9] Wikipedia – Crytpography hashing function http://en.wikipedia.org/wiki/Cryptographic_hash_function

[10] https://en.wikipedia.org/wiki/Biometrics

[11] Al-Assam, H., R. Rashid, and S. Jassim (*2013). Combining steganography and biometric cryptosystems for secure mutual authentication and key exchange,* The 8th International Conference for Internet Technology and Secured Transactions (ICITST).

[12] Bundesamt für Sicherheit in der Informationstechnik (2011*). Study of the Privacy and Accuracy of the Fuzzy Commitment Scheme*.

[13] Matthew A.Turk and Alex P.Pentland (1991). *Face Recognition Using Eigenfaces*, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586-591.

[14] Failla, P., Y. Sutcu, and M. Barni et al (2010). *A privacy-preserving fuzzy commitment scheme for authentication using encrypted biometrics*, in Proceedings of the 12th ACM workshop on Multimedia and security, ACM: Roma, Italy, p. 241-246.

CHƯƠNG 9. PHỤ LỤC

9.1. Cài đặt chương trình

- **Bước 1:** tải các một số thư viện cần thiết cho chương trình như

  Opencv tại link : http://opencv.org/downloads.html

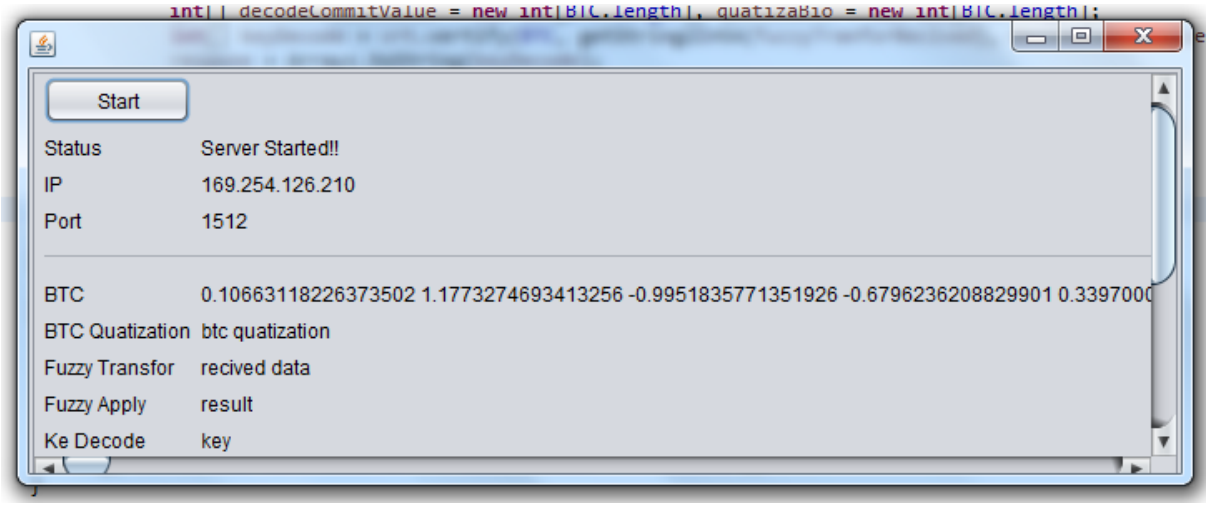
  Jfreechart tại link : http://www.jfree.org/jfreechart/download.html (có thể cài hoặc không tùy ý, mục đích thư viện này chỉ để thống kê và kiểm tra chương trình)

  Jasypt tại link: http://www.jasypt.org/download.html.
- **Bước 2:** import source code của project vào Eclipse và add các thư viện sau: colt.jar, jai_codec.jar, jai_core.jar, opencv.jar, jasypt-1.9.2.jar.
- **Bước 3:** build và chạy thử chương trình.

9.2. Chạy chương trình

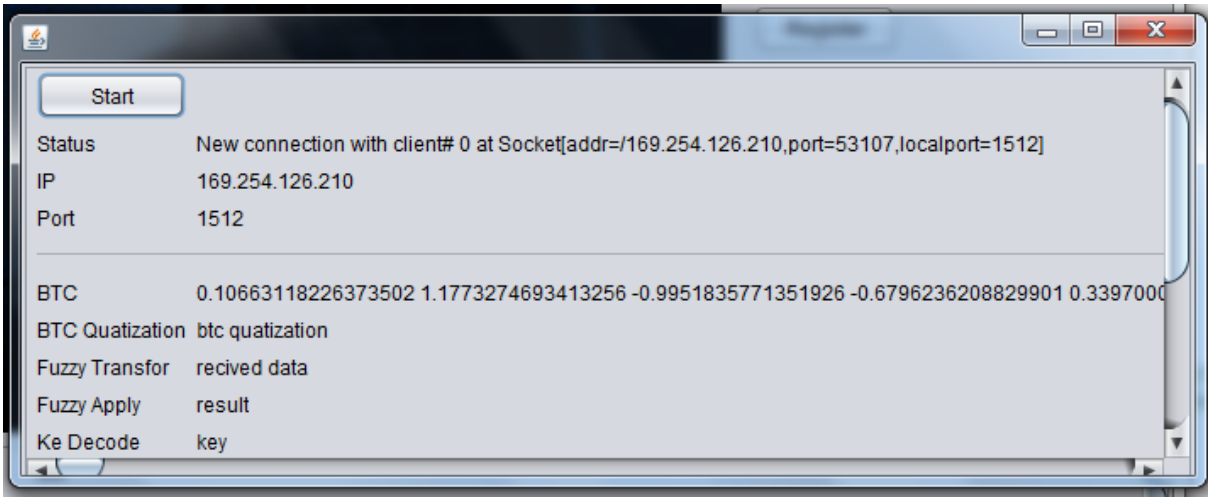
9.2.1 Server S1

Lúc start server thì nó sẽ hiện một vài thông tin như IP, Port và Biometric đang lưu trên cơ sở dữ liệu.



Hình 38. Server Demo

Và server sẽ đợi client đến kết nối, nếu có client nào kết nối thì server sẽ show trên status như sau:



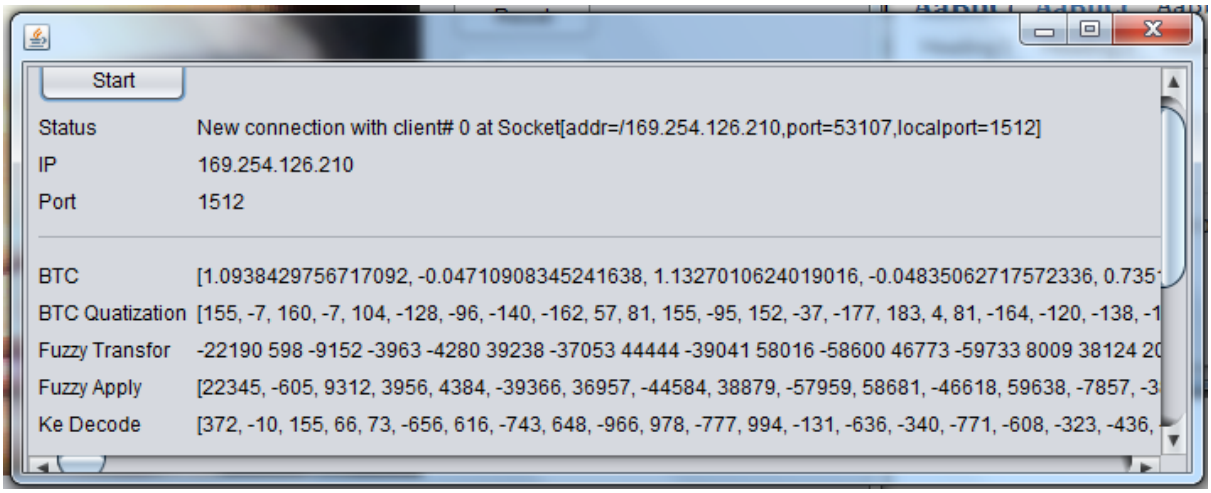
Hình 39. Server Demo

Và nếu client thực hiện enrroll thì giá trị BTC sẽ thay đổi.

Ví dụ:

BTC 1.0938429756717092 -0.04710908345241638 1.1327010624019016 -0.04835062717572336 0.73512

Và nếu người dùng thực hiện chức năng authentication thì nó sẽ show hết toàn bộ thông tin như sau:



Hình 40. Server Demo

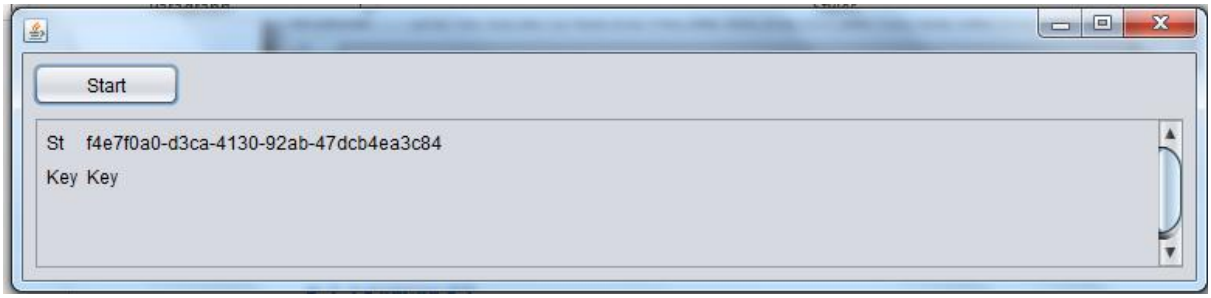
Fuzzy Transfor: giá trị nhận từ client hay còn gọi là helper data.

Fuzzy apply: sau khi un-binding.

Key decode: khóa sau khi decode.

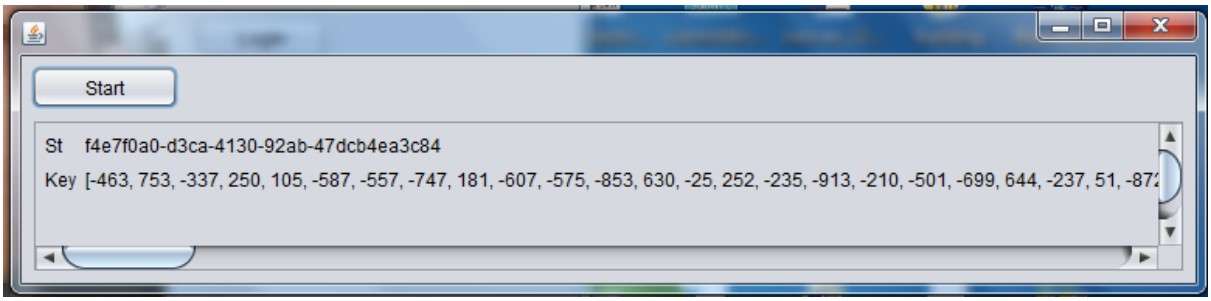
9.2.2 Server S2

Server S_2 chỉ làm nhiệm vụ xác thực người dùng nên chỉ cần lưu thông tin giá trị S và thực hiện các phép hash, encrypt và decrypt để so sánh và trả về kết quả.



Hình 41. Server S_2 start.

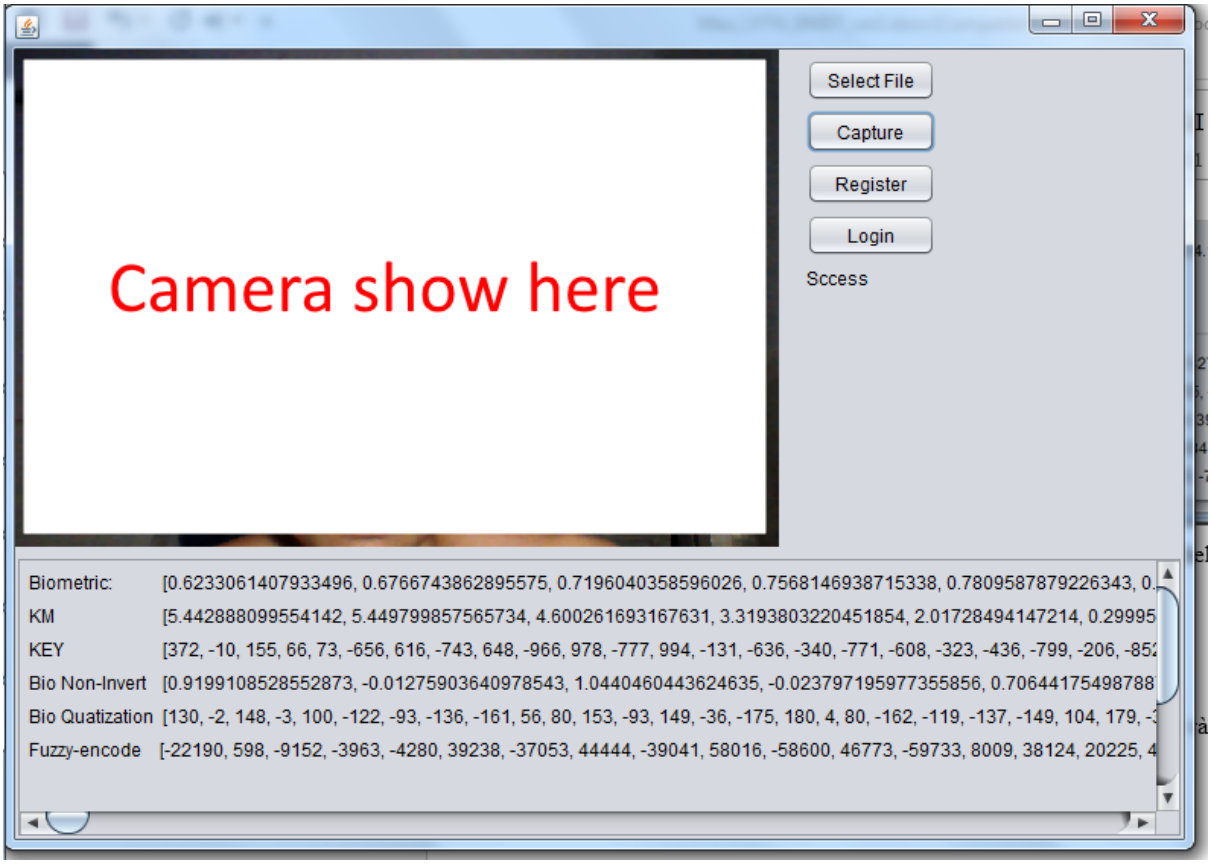
Và khi thực hiện việc Matching để đưa ra quyết định.



Hình 42. Server S_2 Matching.

9.2.3 Client

Phía client có 2 chức năng là đăng ký và đăng nhập. và phía client có nhận hai đầu vào là từ camera và ảnh lưu trong máy tính.

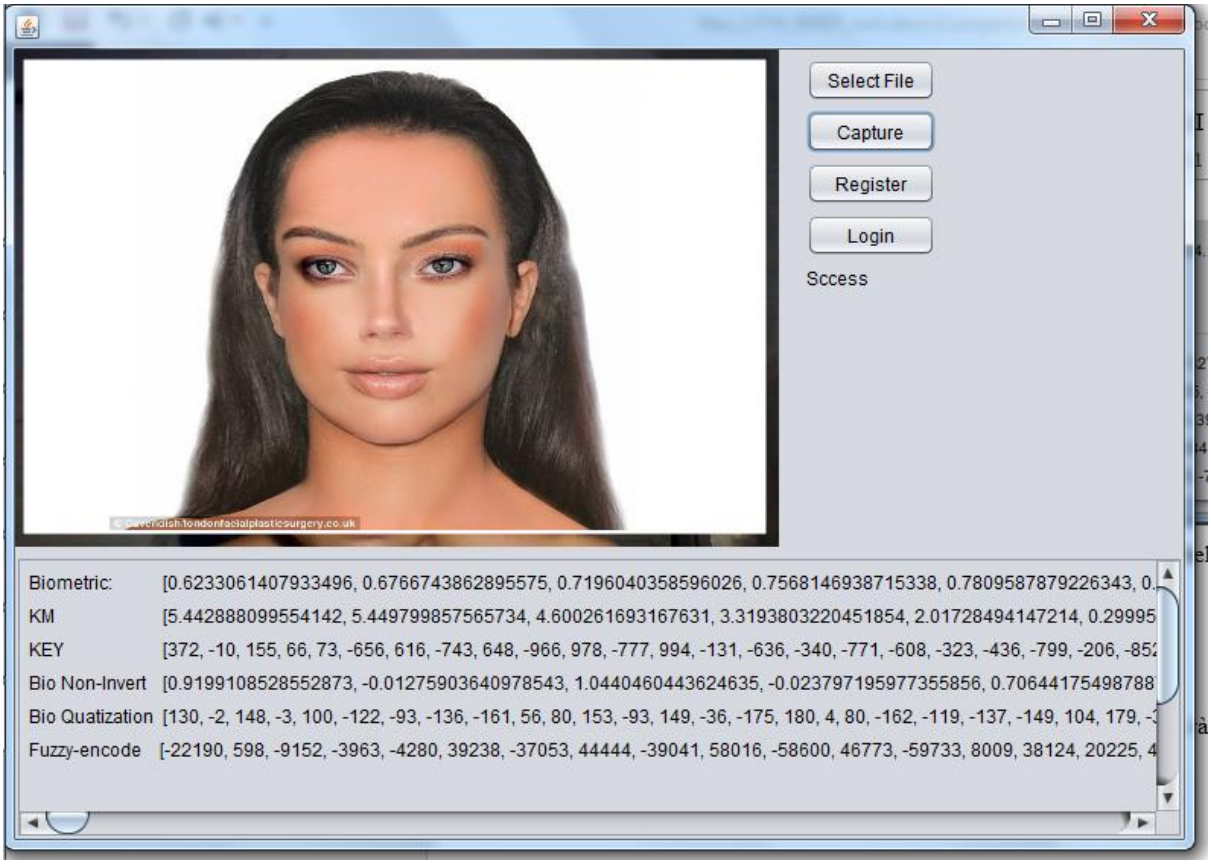


Hình 43. Client Demo

Và chỗ show camera nó cũng show hình load từ máy tính lên và các thông tin lẫn các Button để thực hiện chức năng.

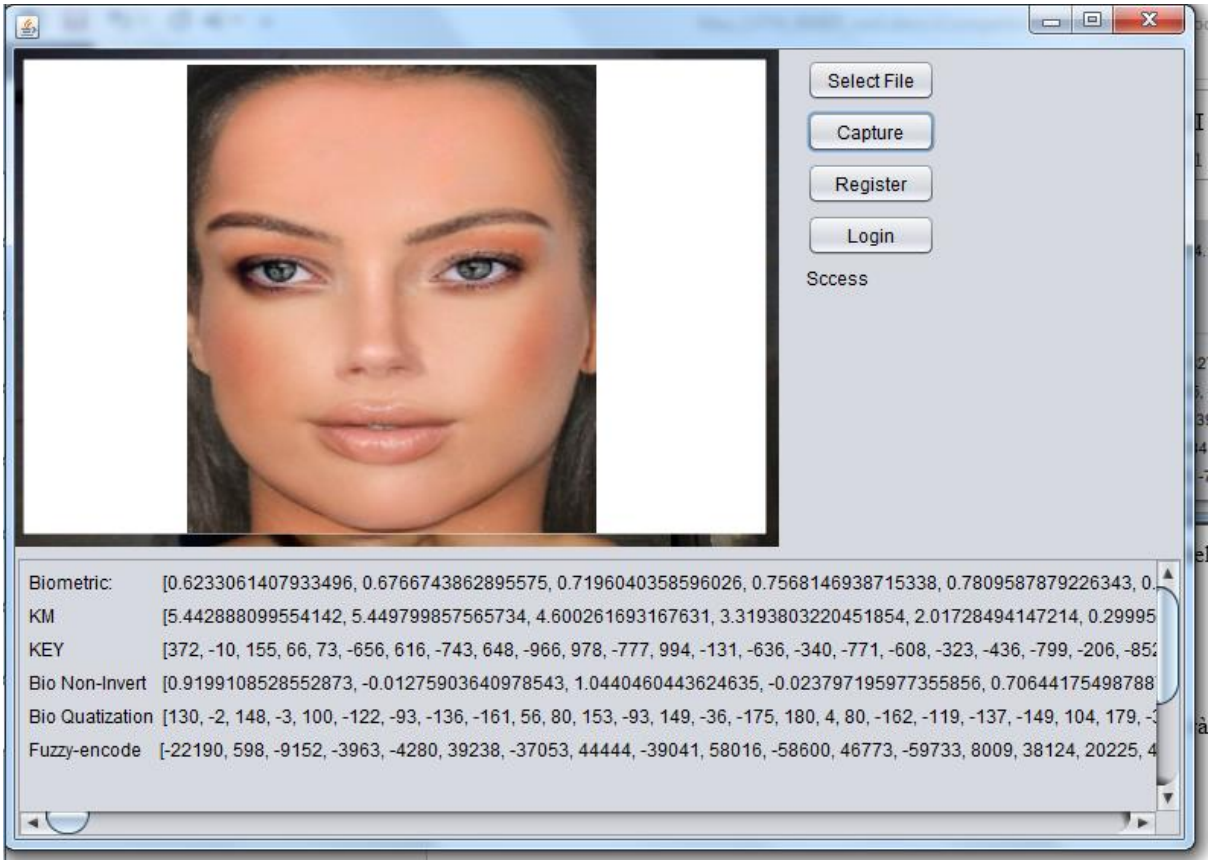
Bên phía client tôi có thêm một phần khá quan trọng để tăng khả năng làm việc cho app đó là trích xuất khuôn mặt từ camera ví dụ:

Khi trên camera có hình này:



Hình 44. Client Demo

Sau đó bạn nhấn capture thì chương trình sẽ tự động trích suất khuôn mặt của bạn và chỉ lấy khuôn mặt mà thôi, đây là kết quả sau khi nhấn Capture:



Hình 45. Client Demo