**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

TRƯỜNG ĐẠI HỌC BÁCH KHOA

KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH

**BỘ MÔN KHOA HỌC MÁY TÍNH**

---------------o0o---------------

****

**LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC**

**TIÊU ĐỀ LUẬN VĂN**

**GVHD: Th.s Nguyễn Thị Ái Thảo**

**SVTH: Nguyễn Nam Tiệp**

**MSSV: 51103624**

**TP. HỒ CHÍ MINH, THÁNG 12 NĂM 2016**

**MỤC LỤC**

**Lời cam kết**

**Lời cảm ơn**

**Tóm tắt luận văn**

**Danh sách hình vẽ**

**Danh sách bảng**

1. **Gới thiệu**
   1. **Tổng quan**
   2. **Nhiệm vụ luận văn**
2. **Lý thuyết**
   1. Fuzzy-commitment
   2. Non-invertible Transformation
   3. Thiết kế hệ thống xác thực hoàn chỉnh
   4. Phân tích độ bảo mật của hệ thống
3. **Hiện thực chương trình**
4. **Kết quả và thống kê**
   1. Tiền thống kê PCA
   2. Thống kê toàn bộ chương trình bằng phương pháp FRR, FAR, và ERR
5. **Kết luận và hướng phát triển**
   1. Kết luận
   2. Hướng phát triển
6. **Tài liệu tham khảo.**
7. **Phụ lục.**

**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 tồi 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**

**TÓM TẮT LUẬN VĂN**

Trong thời kỳ phát triển công nghệ như vũ bảo hiện nay thì nó làm cho một số công nghệ cách đó vài năm cũng trở nên lạc hậu và không còn phù hợp trong những hệ thống hiện nay. 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 giật gân 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](http://www.forbes.com/sites/leemathews/2016/11/11/why-is-facebook-buying-stolen-passwords-on-the-black-market/#3b16d99d2cd1). 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 hiện nay, 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. Chính vì vậy hệ thống này không những thành công về mặt lý thuyết mà còn ở cả thực tiễn.

Luận văn này ý tưởng chính tham khảo dựa trên nghiên cứu[1]. Một lần nữa xin chân thành cảm ơn tác giả đã cung cấp tài liệu để hoàn thành luận văn này.

# GIỚI THIỆU

* 1. **Giới thiệu đề tài:** Bộ môn hệ thống thông tin luôn là một môn hấp dẫn và mới mẻ, trong sự phát triển của công nghệ đặc biệt là bộ môn này, các giảng viên của trường ĐH Bách Khoa TP.HCM luôn tạo nên sự mới mẻ và đột phá bằng việc tìm tòi và nguyên cứu những công nghệ mới đón đầu xu thế, chính vì vậy nên khi nhận đề tài luận văn em luôn tìm những đề tài của bộ môn này để làm một nghiên cứu nho nhỏ trước khi ra trường. Trong những đề tài thứ luôn làm em cảm thấy thú vị nhất là nghiên cứu về những hệ thống hoạt động gần giống với con người, chính vì vậy nên em chọn đề tài này vì với hệ thống như thế này thì có thể nói là một hệ thống giả lập 50% cong 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 đương nhiên 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) vô tri vô giác, chúng ta cung cấp cho hệ thống đó hình ảnh khuôn mặt, giọng nói, vân tay,... để hệ thống đó xác thực chúng ta. Chưa xét đến độ bảo mật của hệ thống, với những hệ thống xác thực cũ thì người dùng phải nhớ một mật mã(password) để có thể đăng nhập vào hệ thống cùng với username, nếu nó quá ngắn thì dễ bị người khác phát hiện, còn nếu nó quá dài thì thậm chí đôi lúc người đó cũng không thể nhớ được. Vậy khi so với một hệ thống mà họ không cần nhớ bất cứ thứ gì, chỉ cần cung cấp đặc điểm của mình cho hệ thống đó, thì đây là một bước tiến về ý tưởng khi đêm lại sự tiện dụng cho người dùng.

Tuy nhiên yếu tố tiên quyết để đem một hệ thống vào sử dụng là độ tin cậy và bảo mật của hệ thống. Với các hệ thống xác thực cũ, thì có rất nhiều phương pháp để bảo mật password(encrypt, hash,...) tuy nhiên với sức mạnh của phần cứng máy tính như hiện nay thì việc dò password không còn là bất khả thi nữa. Chính vì vậy nên xu hướng của thế giới là tìm một giải pháp cho hệ thống xác thực để thay thế cho hệ thống cũ, đó là sử dụng đặc trưng sinh trắc của người dùng và đặc trưng của người dùng gần như là vô hạn, đôi khi có hai người giống nhau nhưng khả năng này rất thấp, và khi sử dụng đặc trưng thì các phương pháp tấn công cũ như bruce-force,.. không sử dụng được, nhưng khi sử dụng đặc trưng của người dùng thì nó lại pháp sinh một số vấn đề về lưu trữ hay bảo mật trên đường truyền.

Chính vì hai vấn đề nói trên mà trong luận văn này nhóm sẽ tìm hiểu **mô hình xác thực từ xa** **sử dụng đặc trưng sinh trắc** và các **phương pháp bảo vệ sinh trắc** và **bảo vệ trên đường truyền có sử dụng đặc trưng sinh trắc**. Mục tiêu cụ thể sẽ được giới thiệu ở phần dưới.

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 thòi 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.

## 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 dụng 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.

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. **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. **Cấu trúc báo cáo.**

# CƠ SỞ LÝ THUYẾT.

# Sinh trắc học(biometric). Sinh trắc học là gì?: sinh trắc học[](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 đí,.... 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 đầu 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 và 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.

# Nói tóm lại để có thể áp dụng đặc trưng sinh trắc vào hệ thống thì cần phải có các yêu cầu 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.

# 

# Hình 2. Đặc điểm sinh trắc của từng loại.

# Hệ thống xác thực(Authentication system) Hệ thống xác thực là hệ thống thực hiện đối sánh 1-1 giữa mẫu sinh trắc thu nhận được(Biometric sample) với mẫu dạn 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?”.

# Các thành phần của hệ thống sinh trắc.

# Thu nhận(Sensor, Capture): thu nhận sinh trắc của con người và biểu diễn dưới dạng số hóa.

# Xử lý và tríc xuất đặc trưng(Feature Extraction): Thực hiện các phép xử lý phân tích và trích chọn các 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.

# Ra quyết định(Decision): dựa trên kết quả đối sánh sẽ khẳng định đây có phải là cùng một người hay không.

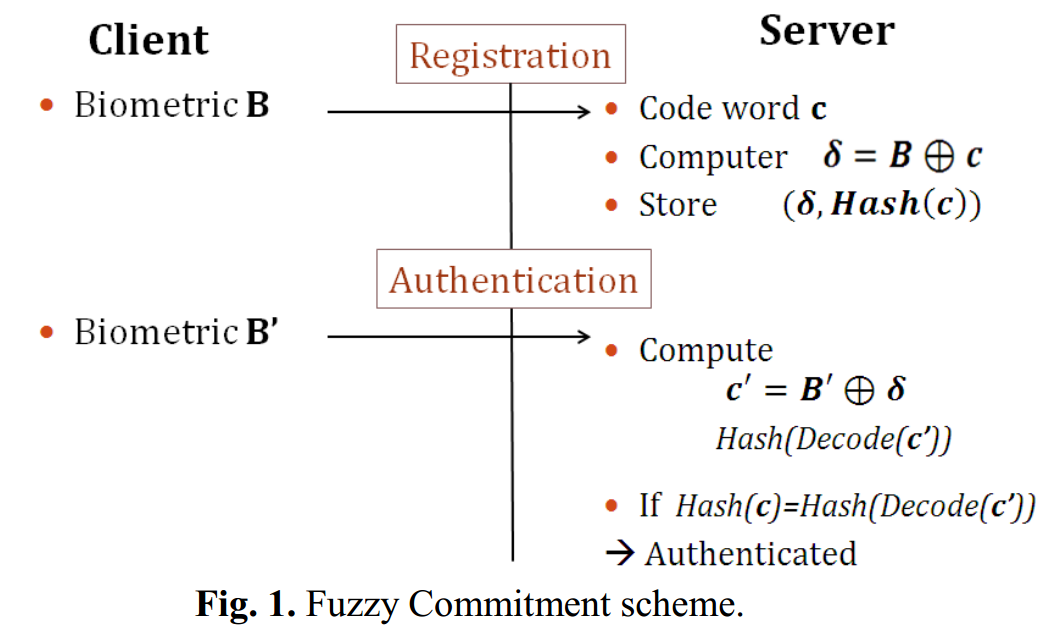
# Cơ chế hoạt động của hệ thống. Một hệ thống sinh trắc gồm có hai thành phần chính sau:

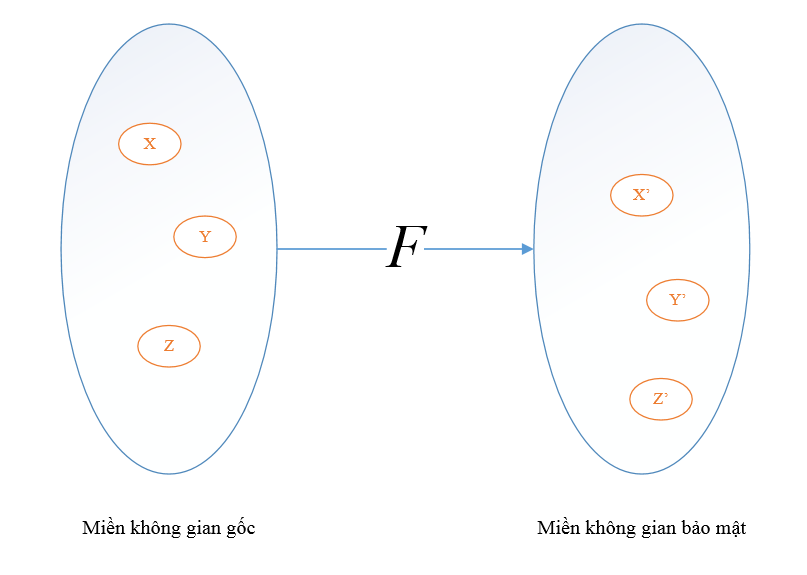
# Ghi danh(Enrollment): Gai đoạn cung cấp sinh trắc của người dùng cho hệ thống sử lý và lưu lại.

# Xác thực(Authentication): Giai đoạn thẩm định, nhận dạng mẫu sinh trắc vừa trích chọn với khuôn mẫu sinh trắc đã có trước.

# Image result for authentication system a Hình 3. Hai thành phần chính của hệ thống sinh trắc.

Như đã giói thiệu ở trên luận văn sẽ kết hợp hai phương pháp fuzzy-commitment và non-invertible transformation để bảo vệ mẫu đặc trưng sinh trắc trong hệ thống xác thực từ xa xử dụng đặc trưng sinh trắc. Trước khi vào chi tiết hệ thống thì chúng ta cần tìm hiểu hai phương pháp cốt lõi trong hệ thống này đó là fuzzy-commitment và non-invertible transformation. Mỗi phương pháp sẽ giải quyết một vấn đề trong hệ thống xác thực từ xa, đầu tiên em xin liệt kê những vấn đề lớn cần giải quyết của hệ thống xác thực người dùng sử dụng đặc trưng sinh trắc là:

* Bảo vệ đặc trưng sinh trắc được lưu trên máy chủ đề ngăn chặn khả năng tấn công bên trong, vấn đề này sẽ được giải quyết bằng phương pháp non-invertiable transformation.
* Bảo vệ trên đường truyền mạng có sử dụng đặc trưng sinh trắc để ngăn chặn tấn công bên ngoài , vấn đề này sẽ được gải quyết bằng phương pháp fuzzy-commitment.
  1. **Fuzzy-commitment.**Fuzzy-commitment là phương pháp đã được đề cập rất trong tài liệu[] đây được coi là một trong những phương pháp đầu tiên trong việc sử dụng biometric trong hệ thống bảo mật(biometric crptosystem approach). 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 [] 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⊆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: **g**: **M→C**, và một hàm decode **f**: **C→M**, thực ra thì hàm **f** có thể hồi phục **M** từ một tập các chuỗi gần **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 em 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 em sẽ dùng mã sữa lỗi kiểu integer). Dưới đây là mô hình của một hệ thống xác thực sử dụng 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 phí 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.
  2. **Non-invertible Transfomation.**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 chiều 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.

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

có nhiều phương pháp tạo ma trận trực giao như Gram-Schmidt hay Hisham Al-Assam.. nhưng trong luận văn này em 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  ( là số chiều của vector đặc trưng sinh trắc  dựa vào khóa nền người dùng hoặc token).
    - Áp dụng quy trình Hisham Al-assam để tạo ma trận trực giao từ các vector ngẫu nhiên trên.
    - Biến đổi vector đặc trưng  sử dụng ma trận trực giao : ***y = Ax***

**Phương pháp của Hisham Al-Assam:**trước tiên ta xét một ma trận như sau:



Ta thấy ma trận trên trực gaio với mọi . Dựa vào tính chất này ta có thể tạo một ma trận  từ những ma trận . Hisham Al-Assam đề xuật phép chiếu ngẫu nhiên như sau:

* + - Tự tạo một tập ***n*** giá trại ngẫu nhiên {*θ1* , *θ2* ,*…. θn*} là các số thực trong khoảng [0..2π], ta tạo một ma trận trực giao **** với các đường chéo là các ma trận đơn vị *θi* như trên.



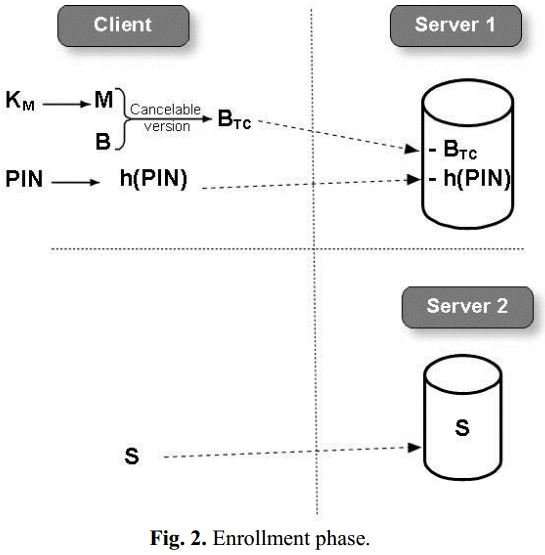
* + - Ánh xạ cac vector đặc trưng ***x*** sang miền bảo mật sử dụng ma trận ***A: y = A.x.***

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

**Như đã trình bày ở trên fuzzy-commitment và non-invertible transfomation là hai phương pháp cốt lõi của một hệ thống xác thực từ xa, tuy nhiên muốn áp dụng vào hệ thống thì xây dựng lại một hệ thống mới để có thể áp dụng được hai phương pháp này và trong phần tiếp theo em sẽ trình bài một mô hình dựa trên nghiên cứu của nhóm Thạc Sĩ Nguyễn Thị Ái Thảo.**

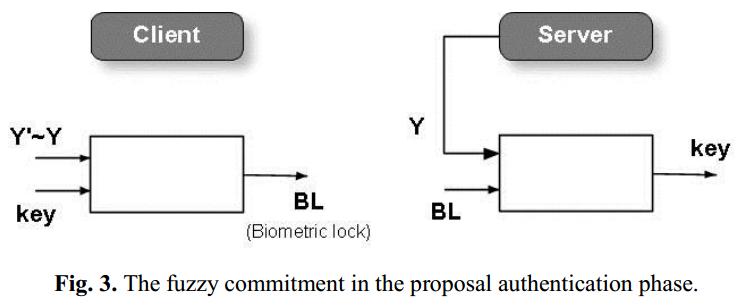
***kí hiệu:***

* B : đặc trưng sinh trắc của người dùng để nhập vào hệ thống.
* M : ma trận trực giao.
* BTC : 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.
* EPuX (m) mã hóa thông tin ***m***  với key ***X*** và mã hóa đối xứng.
* S : mobile serial number.
* ST : mobile serial number lưu trong database của server.
* C : client.
* S1, S2 : lần lượt là server thứ nhất và server thứ hai.
  1. **Thiết kế hệ thống xác thực hoàn chỉnh.**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.  
       
       
       
       
       
       
       
     **Enrollment Phase:**

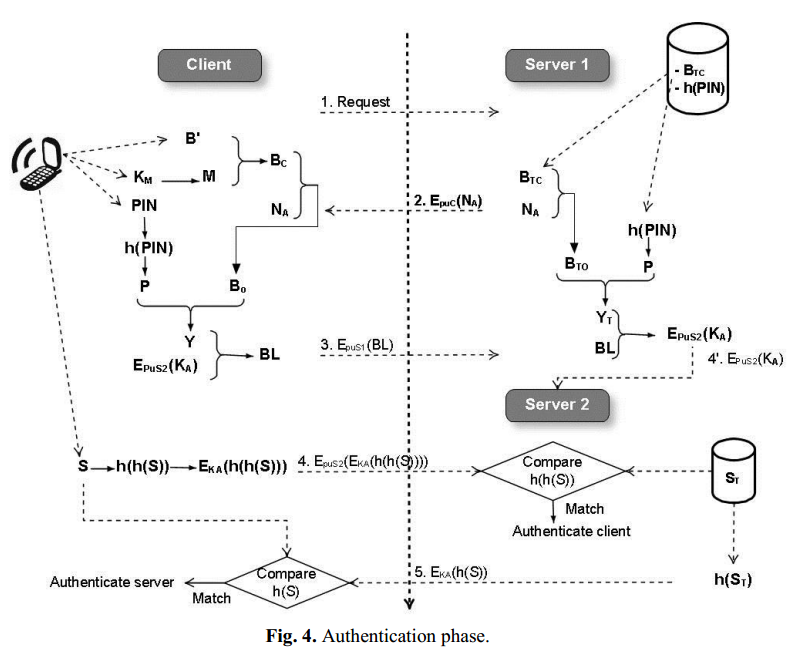


Trong pha này người dùng sẽ tạo ngẫu nhiên vector Km và sẽ được lưu trong thiết bị của người dùng, và Km sẽ được dùng để tạo ma trận trực giao M(dựa trên giải thuật của Hisham Al-Assham) . Sau khi người dùng cung cấp vector đặc trưng sinh trắc B thì sẽ kết hợp với ma trận trực giao(phép nhân hai ma trận) sẽ thu được BTC (chuyển sang vùng bảo mật) và sẽ được gửi qua server S1 và sẽ được lưu ở đó. Kèm theo đó phía client sẽ sinh ra một số bí mật PIN và gửi giá trị H(PIN) đến server . Song song với những công việc trên thì phía client cũng sẽ gửi đến cho server S2 số serial number S. Việc chia hai server sẽ chia bớt thông tin từ client gửi đến server để giảm quá tải, và điểu 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: BTC, H(PIN) ở S1 và S(serial number) ở S2, và như ta đã thấy ở pha này chúng ta chỉ sử dụng một phương pháp bảo mật là Non-invertible transformation.

**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 help-date 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 quá thu được sẽ là khóa xác thực của người dùng. Mô hình sẽ được biểu diễn bằng biểu đồ dưới đây.



**Có thể tóm tắt một cách dễ hiểu như sau:** mục đích chính của quá trình này là truyên **key** đến server để xác thực người dùng. Tại lúc này phía server đã có lưu BTC sau đó sử dụng BTC 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 S2 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 S2 thực hiện xác thực. Mô hình chi tiết sẽ được biểu diễn trong hình sau:



**Các bước cụ thể như sau:** đầu tiên client gửi yêu cầu đến S1, sau đó S1 sẽ tạo ra một số ngẫu nhiên dùng một lần Na, 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 BC từ B’ và ma trận trực giao M sẽ được sinh ra từ KM. Sau đó BC sẽ được kết hợp với NA được gửi từ S1 để tạo ra BO, mục đích kết hợp vơi NA là để chống lại replay attack. Sau đó BO 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 KA 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 S1. Bên S1 sau khi tạo ra NA, thì đồng thời cũng lấy BTC và cũng qua các quá trình tương tự như phái client như, kết hợp vơi NA , qua phép hoán vị P để kết quả là YT, và sau khi nhận được BL từ client thì server sẽ kết hợp với YT để làm input cho quá trình fuzzy-commitment bên phái S1, sau khi kết thúc quá trình này thì ta sẽ thu được giá trị mã hóa của khóa xác thực KA để gửi qua S2 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(ST) để phía client có thể xác thực lại server.

* 1. **Phân tích 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 NA, số này được server sinh ra và sử dụng một lần.
     + KM 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:

* + 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 KM 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.
    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 NA và session key KA 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 KA chỉ dùng một lần thì cho dù kẻ gian có lấy được KA của lần đăng nhập trước cũng không còn giá trị trong lần đăng nhập này.
    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.
    4. **Insider attack:**  
       Với kiểu tấn công này thì đã được giải quyết bằng phương pháp chia ra làm hai server: S1 thì tuy lưu thông tin của người dùng nhưng chỉ làm nhiện vụ nhận khóa xác từ client, S2 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.

# HIỆN THỰC CHƯƠNG TRÌNH.

# KẾT QUẢ VÀ THỐNG KÊ. 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 ‘tiền thống kê’ để xem xem khả năng đáp ứng tối thiểu của PCA.

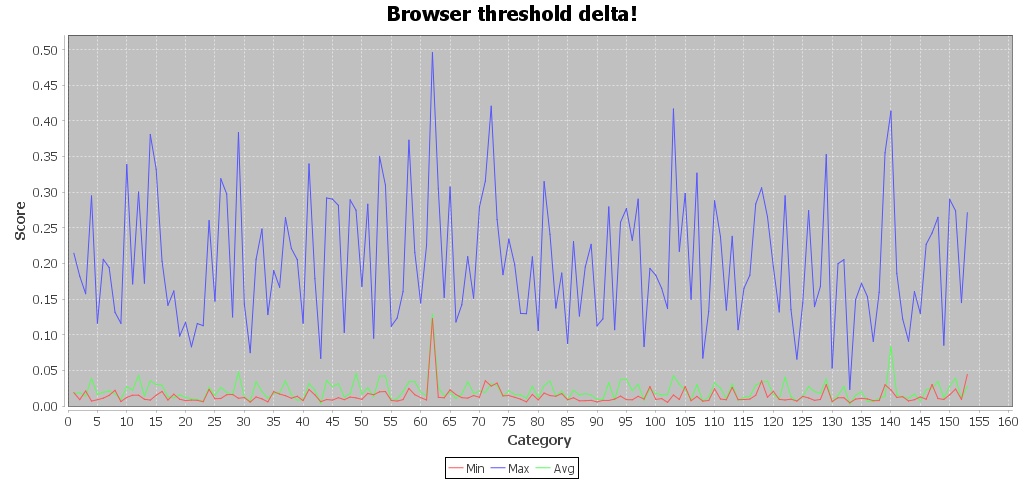
* 1. **Tiền thống kê.**Như ta đã biết phương pháp PCA thì đầu vào của nó là một ảnh và sau đó sinh ra một vector đặc trưng sinh trắc của ảnh đó. Và vector này luôn có độ dài là 200 và mỗi thành phần của vector nằm trong khoảng [0,1]. Như vậy để đảm bảo với phương pháp PCA thì một ảnh sẽ sinh ra một vector 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 em 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 em 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:

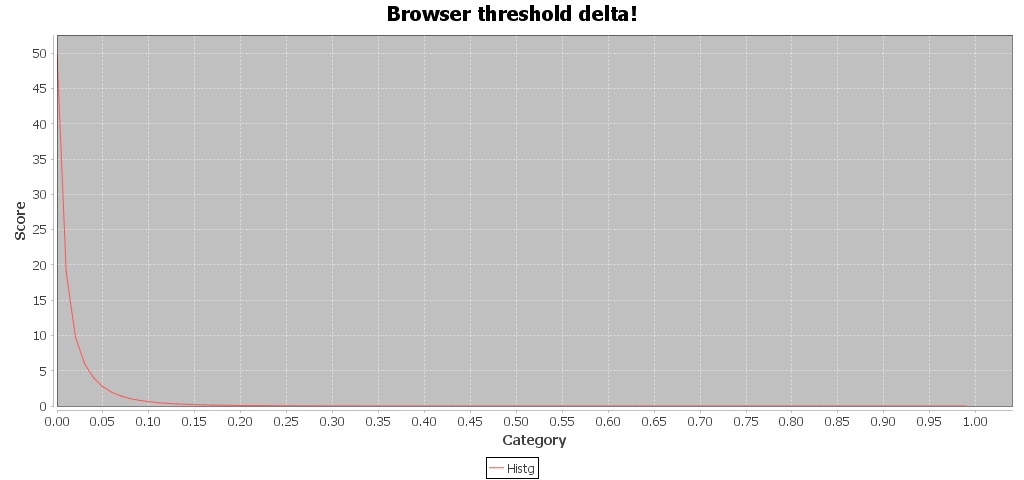
Gọi X[x1, x2, x3, x4,...... xN] và X’[x’1, x’2, x’3, x’4,...... x’N] lần lược là hai vector đặc trưng sinh trắc.   
 khi độ lệch giữa hai vector là:

||X-X’|| = MAX( (x1 – x’1), (x2 – x’2), (x3 – x’3), ....,( xN – x’N) ) = (xk – 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.



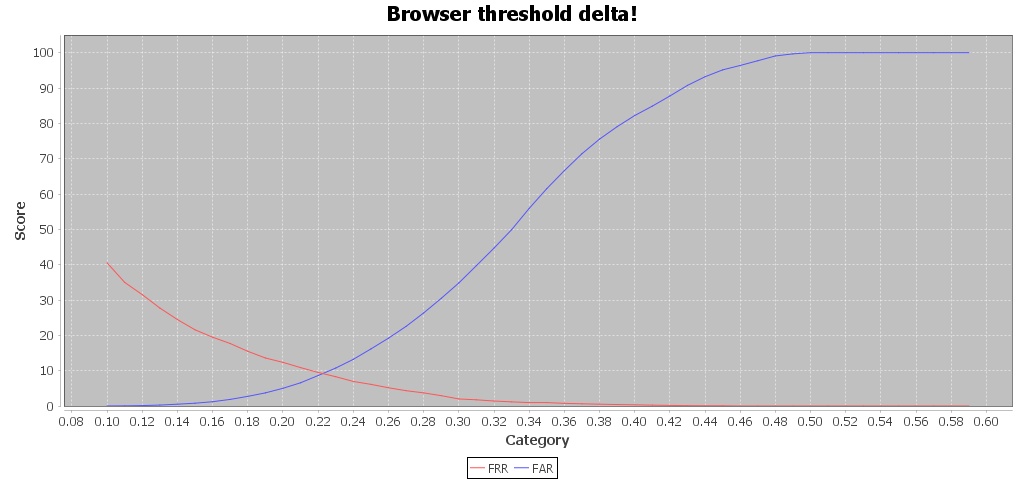
Từ đồ thị trên ta có thể rút ra nhận xét: giá trị **avg** và giá trị **min** khá nhỏ và đa số nhỏ hơn 0.05 và hai đường **min** và **avg** nằm sát nhau và tách xa với giá trị của đường **max** điều này chứng tỏ vector sinh trắc được sinh ra từ phương pháp PCA với cùng một người có độ lệch rất thấp và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



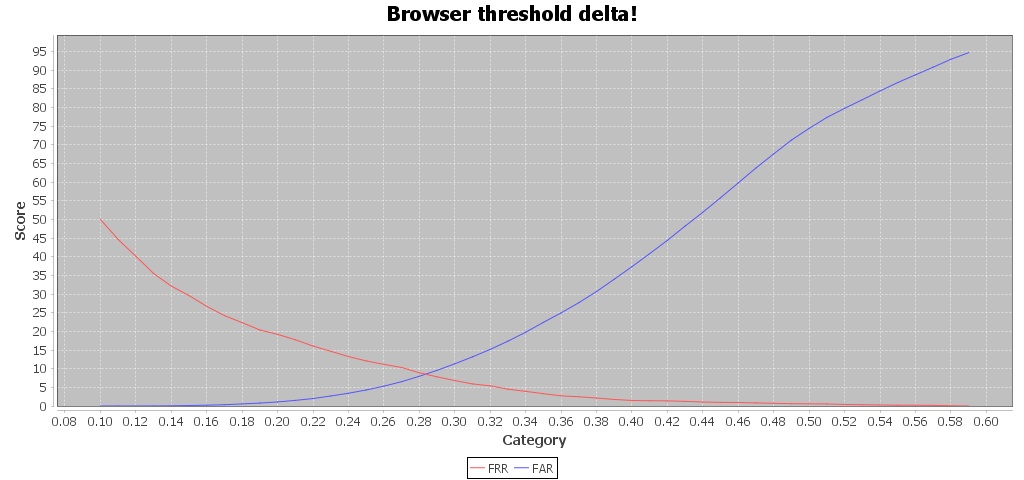
**Nhận xét:** với đồ thị trên ta có thể thấy với phương pháp PCA thì độ tương đương của hai vector là rất cao , với độ lệch nhỏ hơn 0.05 thì hơn 80% số lượng vector có độ lệch này. **Với phương pháp thống kê sẽ cho chúng ta cái nhìn tổng quát về các vector và có thể dự đoán được khoảng threshold để áp dụng cho hệ thống tuy nhiên chúng ta cần phải tìm được giá trị threshold cho hệ thống nên dưới đây em làm thêm phương pháp FRR và FAR để tìm giá trị threshold này.**

* 1. **Thống kê toàn bộ chương trình bằng phương pháp (FRR và FAR).**
* **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.

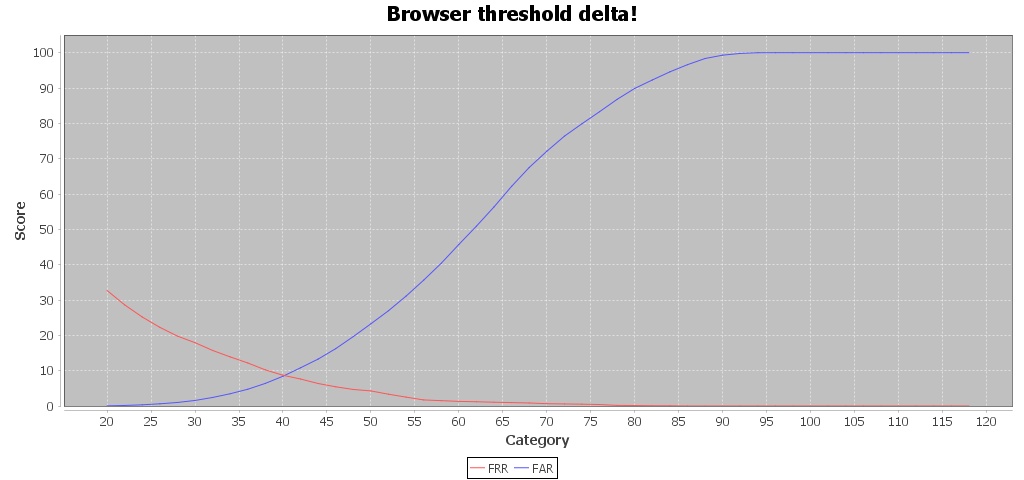
Tập test là 153 người và mỗi người có 20 ảnh tuy nhiên chỉ có khoảng 4,5 kiểu khuôn mặt khác nhau nên em chỉ lấy ảnh thứ 1( với index là 0) làm ảnh enrrollment và các ảnh với thứ tự lần lược là 1,3,6,8,15,17 làm ảnh lúc vertify.



**FRR và FAR với vector sinh ra từ phương pháp PCA.**



**FRR và FAR sau khi nhân với ma trận trực giao.**



**FRR và FAR của toàn hệ thống khi kết hợp với phương pháp nhân với ma trận trực giao và Fuzzycommitment.**

**Nhận xét:** 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 em 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 thì giá trị threshold của nó tầm 0.285 = 0.22x 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.

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## Kết luận

## Hướng phát triển

# TÀI LIỆU THAM KHẢO

# PHỤ LỤC