NÂNG CAO TÍNH BẢO MẬT TRONG XÁC THỰC NGƯỜI DÙNG WEB SỬ DỤNG ĐẶC TRƯNG SINH TRẮC HỌC

Nguyễn Hữu Nội[1*], Vũ Thanh Nhân[2], Trần Nguyên Ngọc[1]

Tóm tắt: *Bài báo này nghiên cứu sử dụng hành vi gõ bàn phím trong xác thực người dùng hướng tới mục đích bảo mật thông tin. Bài báo cũng đề xuất việc xây dựng cơ chế xác thực kết hợp giữa mật khẩu thông thường dạng text với mật khẩu sinh trắc học trên nền ứng dụng Web. Các thông tin của người dùng được thu thập thông qua việc gõ bàn phím (máy tính, điện thoại) và được tổ chức thành các vec-tơ đặc trưng sau đó sẽ được gửi lên phía máy chủ để xử lý.*

Từ khóa: Mật khẩu sinh học, Xác thực, Sinh trắc học.

1. ĐẶT VẤN ĐỀ

Hiện nay, với sự phát triển của khoa học công nghệ thì việc sử dụng mật khẩu gồm những chuỗi ký tự (gồm số, chữ cái, ký tự đặc biệt) khó nhớ đang dần được thay thế bằng các phương pháp khác, chẳng hạn như theo vân tay, hình dáng khuôn mặt, nhịp tim, hình dáng tai [1,2,3,14]… Những phương pháp nhận diện này được gọi chung là sinh trắc học (biometrics). Trong tương lai, thì nhận diện sinh trắc học sẽ ngày càng được sử dụng rộng rãi trong việc xác định danh tính.

Sử dụng vân tay là nhận dạng sinh trắc học phổ biến nhất, nó đã được hàng loạt các hãng công nghệ áp dụng trên các sản phẩm của họ, từ di động cho đến máy tính, chẳng hạn như Apple đã nhúng cảm biến vân tay vào nút "Home" của iPhone 5S [15]. Công nghệ này hoạt động theo nguyên tắc khi đặt ngón tay lên trên một thiết bị đọc dấu vân tay, ngay lập tức thiết bị này sẽ quét hình ảnh ngón tay đó và đưa vào hệ thống. Hệ thống sẽ xử lý dấu vân tay, chuyển sang dạng dữ liệu số rồi đối chiếu các đặc điểm của vân tay đó với dữ liệu đã được lưu trữ trong hệ thống. Nếu dấu vân tay khớp với dữ liệu thì hệ thống sẽ cho phép các chức năng tiếp theo.

Cũng như dấu vân tay, công nghệ nhận diện khuôn mặt hiện nay cũng được dùng khá phổ biến, bằng cách sử dụng các máy ảnh được trang bị sẵn trên các thiết bị (điện thoại, máy tính, máy tính bảng) để chụp lại khuôn mặt của người dùng, sau đó sử dụng các công cụ phần mềm để xử lý hình ảnh thu được với các mẫu khuôn mặt có sẵn trong cơ sở dữ liệu (CSDL) để nhận dạng người. Gần đây, công ty Facebook đã công bố một dự án nhận diện khuôn mặt riêng với tên gọi DeepFace [13], có khả năng nhận diện rất chính xác các khuôn mặt, thậm chí ngay cả khi khuôn mặt đó không được chụp chính diện.

Mặc dù vậy, các kỹ thuật trích chọn thông tin sinh trắc học đều cần đến các thiết bị đặc chủng, ví dụ cần có máy quyét vân tay, camera giúp nhận dạng khuôn mặt, thiết bị sõi võng mạc trong nhận dạng tròng mắt... Việc sử dụng đặc trưng sinh trắc học đối với người dùng Web thường không cho phép yêu cầu bắt buộc người dùng phải sử dụng các thiết bị đó. Vì thế, trong nghiên cứu này chúng tôi hướng tới việc sử dụng những thông tin đơn giản nhất mà hầu như bất kỳ người dùng web nào cũng có thể cung cấp để hỗ trợ nâng cao tính bảo mật trong xác thực người dùng.

Phép nhận dạng khá đơn giản đó là sử dụng các thông tin thu được từ các thao tác gõ bàn phím của người dùng (Keystroke Dynamics – KD) [1, 2, 7]. Về bản chất KD là một dạng đặc trưng sinh trắc học cho phép mô tả thao tác người dùng khi gõ bàn phím máy tính, nhấn phím trên điện thoại di động (kể cả bàn phím cảm ứng ảo trên các dòng diện thoại thông minh) [10]. Ở đây, cần lưu ý rằng, với đa số các trang web hiện nay đều có khả năng phân biệt người dùng trên điện thoại di động hay máy tính cá nhân để đưa ra giao diện tương tác phù hợp, do vậy, việc khai thác đặc trưng sinh trắc học cũng có thể tiếp cận lợi thế này để biết trước thông tin thu được là từ bàn phím máy tính hay thiết bị di động.

Việc sử dụng KD trong đảm bảo an toàn thông tin có ưu điểm nổi bật là không cần sử dụng thêm các thiết bị phần cứng phụ trợ ngoại trừ bàn phím (Keyboard, Keypad). Việc sử dụng KD sẽ làm mạnh hơn sự xác thực thông tin người dùng, ngay cả trong trường hợp các thông tin đăng nhập (tên đăng nhập, mật khẩu) bị lộ lọt.

Trong nghiên cứu này, chúng tôi tiếp cận bài toán trên cơ sở sử dụng các kết quả nghiên cứu trước đó đã công bố tại [11,16] để xây dựng cơ chế xác thực cho người dùng trên nền ứng dụng Web. Bố cục bài báo ở các phần tiếp theo được tổ chức như sau: trong phần 2 tổng hợp kết quả của một số công trình nghiên cứu trước đó, các thuật toán tính khoảng cách và kiểm tra trên các bộ dữ liệu có sẵn [12]; cách tính ngưỡng xác thực; phần 3 trình bày về mô hình tương tác của ứng dụng, tính toán và thảo luận; phần 4 trình bày kết luận và các hướng nghiên cứu, phát triển tiếp theo của nhóm tác giả.

2. THUẬT TOÁN PHÂN LOẠI DỮ LIỆU GÕ BÀM PHÍM

2.1. Xây dựng lý thuyết

Trong phần này chúng ta sẽ xem xét một cách cụ thể về mật khẩu sinh học và các phương pháp phát hiện bất thường thông qua phân tích thời gian gõ mật khẩu.

2.1.1. Cách trích chọn vec-tơ dữ liệu đặc trưng

Đặc trưng dữ liệu KD được trích chọn dựa trên các thông tin về thời gian khi người dùng thao tác với bàn phím [3, 11]. Với các sự kiện bàn phím như: nhấn phím (key-press), nhả phím (key-release), chúng ta sẽ tính được các khoảng thời gian tương ứng. Giả sử có hai phím X, Y được nhấn, khi đó, chúng ta sẽ thu được các dữ liệu tương ứng là:

- H.X – là thời gian giữ phím X (H – Hold), tính từ khi phím được nhấn cho đến khi được thả ra.
- DD.X.Y – là thời gian tính từ thời điểm phím X được nhấn (X – Down) đến thời điểm phím Y được nhấn (Y – Down); X, Y được nhấn liên tiếp nhau.
- UD.X.Y – là thời gian tính từ lúc phím X được thả ra (X – Up) đến khi phím Y được thả ra (Y – Down); thời gian này có thể mang giá trị âm.

Do vậy, nếu một mật khẩu là một chuỗi ký tự có độ dài n - $\{k_1, k_2, \ldots k_n\}$ thì vec-tơ đặc trưng cho mỗi quá trình gõ bàn phím sẽ được xác định như sau:

$\bar{f} = (H.k_1, DD.k_1.k_2, UD.k_1.k_2, \ldots, H.k_n.R, DD.k_n.R, UD.k_n.R, H.R)$ – với chiều dài $3n + 1$, trong đó, R là phím Enter/Return chỉ việc kết thúc thao tác nhập mật khẩu.

Bài toán đặt ra lúc này là từ một tập hợp dữ liệu các vec-tơ đặc trưng $\bar{f}_1, \bar{f}_2 \ldots \bar{f}_n$ và vec-tơ định nhãn tương ứng (xác định vec-tơ đặc trưng f_i là của người dùng có ID là l_i) $\bar{L} = (l_1, l_2, \ldots l_n)$, $l_i \in [1, k], k \leq n$ (n – số lượng người dùng) cần xây dựng một thuật toán cho phép phân loại (xác định) vec-tơ đặc trưng $\bar{f}^*$ ($\bar{f}^*$ được thu thập mới và không nằm trong số $\bar{f}_1, \bar{f}_2 \ldots \bar{f}_n$) liệu có nhãn tương ứng là l^* ($l^* \in [1, k]$) hay không?

Điều này có nghĩa là từ bộ dữ liệu thu được qua quá trình huấn luyện dữ liệu nhận được từ người dùng ($\bar{f}_1, \bar{f}_2 \ldots \bar{f}_n$) chúng ta cần phải tính được một vec-tơ đặc trưng cho người dùng đó, ta coi tập dữ liệu này là tập huấn luyện; sau đó với mỗi lần người dùng xác thực, các dữ liệu mới sẽ được gửi lên (mỗi lần dữ liệu được gửi lên được coi như tập kiểm thử) ta sẽ tiến hành so sánh với vec-tơ đặc trưng để xác định xem đó có phải là người dùng đó hay là một người khác đang cố gắng truy cập vào hệ thống.

2.1.2. Tiêu chí đánh giá

Để đánh giá mức độ tin cậy của bài toán xác thực sử dụng thông tin sinh trắc học, đa phần các nghiên cứu [6, 11,16] đều dựa vào việc sử dụng một bộ dữ liệu kiểm thử (độc lập với bộ dữ liệu dùng để huấn luyện) và áp dụng tiêu chí EER (Equal Error Rate) trên bộ dữ liệu đó.

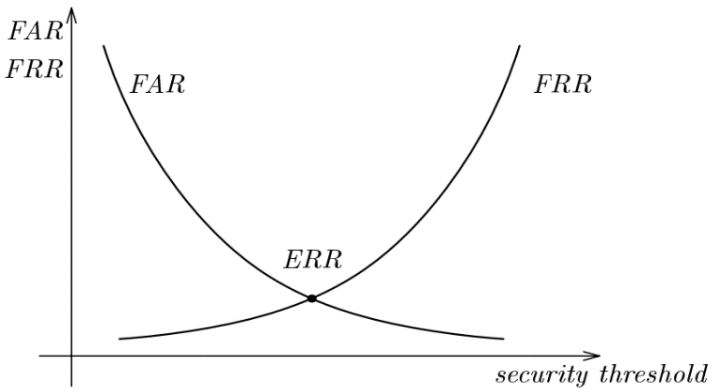
Trước hết, ký hiệu:

- P là tổng số các đối tượng có nhãn L được mang đi phân loại;
- N là tổng số các đối tượng không có nhãn L được mang đi phân loại;
- TP (True Possitive) là số lượng các đối tượng có nhãn L được phân loại đúng, mục tiêu là tăng độ lớn của TP;
- FP (False Possitive) là số lượng các đối tượng không có nhãn L nhưng được phân loại nhầm là có nhãn L, mục tiêu là giảm FP;
- FN (False Negative) là số lượng các đối tượng có nhãn L nhưng là được phân loại là không phải, mục tiêu là giảm FN.

Khi đó, hai tiêu chí tỉ lệ chấp nhận sai – FAR (False Accept Rate) và tỉ lệ từ chối sai – FRR (False Rejection Rate) được xác định như sau:

$$FAR = \frac{FP}{N}; FRR = 1 - \frac{TP}{P}$$

Trong các thuật toán, người ta cố gắng điều chỉnh tham số của bộ phân loại để FAR và FRR có giá trị trùng nhau, khi đó EER = FAR = FRR là giá trị cần tìm. Giá trị của EER càng thấp chứng tỏ hệ thống có độ tin cậy càng cao hay thuật toán hoạt động càng tốt. Minh họa qua hình 1.



Hình 1. *Ví dụ cách xác định giá trị EER.*

2.1.3. Thuật toán tính khoảng cách và kết quả kiểm tra trên các bộ dữ liệu có sẵn

Đã có khá nhiều khoảng cách khác nhau được sử dụng để so sánh giữa vec-tơ dữ liệu huấn luyện với vec-tơ đặc trưng để từ đó xác định. Giả sử ta có các vec-tơ như sau:

$\bar{x} = (x_1, x_2, ...x_N)$ – vec-tơ đặc trưng đại diện cho nhãn L được xác định từ tập dữ liệu huấn luyện;

$\bar{y} = (y_1, y_2, \dots y_N)$ – vec-tơ dữ liệu nhập vào được sử dụng để kiểm tra đăng nhập;

$\bar{a} = (a_1, a_2, \dots a_N)$ – vec-tơ độ lệch chuẩn được tính toán từ tập huấn luyện.

Khi đó, trong các nghiên cứu [8, 9, 11, 12] một số khoảng cách sau được áp dụng để tính toán:

a. *Khoảng cách Euclid*

$$d(\bar{x}, \bar{y}) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \quad (1)$$

b. *Khoảng cách Mahalanobis*

$$d(\bar{x}, \bar{y}) = \sqrt{\sum_{i=1}^{N}\frac{(x_i - y_i)^2}{a_i^2}} \quad (2)$$

c. *Khoảng cách Manhattan*

$$d(\bar{x}, \bar{y}) = |x_i - y_i| \quad (3)$$

d. *Khoảng cách Manhattan-scaled*

$$d(\bar{x}, \bar{y}) = \sum_{i=1}^{N}\frac{|x_i - y_i|}{a_i} \quad (4)$$

e. *Khoảng cách Logarit cải tiến*

$$d(\bar{x}, \bar{y}) = \sum_{i=1}^{N} ln\left(1 + \frac{|x_i - y_i|}{a_i}\right) \quad (5)$$

Trong nghiên cứu [11] cũng đã đưa ra so sánh kết quả làm việc của các thuật toán trên bộ dữ liệu CMU.

Bảng 1. *Kết quả kiểm tra hoạt động các thuật toán trên bộ dữ liệu CMU [12].*

Phương pháp sử dụng khoảng cách	***EER***	***STD (Standard Deviation)***
Manhattan	0.153	0.0925
Euclidean	0.171	0.095
Manhattan scaled	0.0961	0.0693
Logarit cải tiến	**0.0693**	**0.0588**

Từ bảng so sánh ta thấy rõ ràng với khoảng cách mới được thiết kế, EER có giá trị nhỏ hơn cả và độ lệch chuẩn (STD) cũng nhỏ hơn cả. Trong phần thực nghiệm, chúng tôi chọn khoảng cách này là tiêu chí để đánh giá việc xác thực của người dùng, sẽ được mô tả mở mục 2.2.

2.2. Chuẩn bị thực nghiệm

2.2.1. Phương pháp xác định ngưỡng xác thực

Sau khi huấn luyện dữ liệu xong thu được vec-tơ đặc trưng của người dùng là $\bar{x} = (x_1, x_2, \dots x_N)$ và giả sử người dùng tiến hành đăng nhập m lần, khi đó bộ dữ liệu tương ứng là $y = (y_1, y_2, \dots, y_m)$ với $y_i = (y_i^1, y_i^2, \dots, y_i^N), (i = 1, \dots, m)$ với N là độ dài vec-tơ dữ liệu nhận được theo mục II.1. Với mỗi vec-tơ y_i sử dụng công thức (1) ta sẽ thu được khoảng cách tương ứng là d_i. Toàn bộ vec-tơ khoảng cách là $d = (d_1, d_2, \dots, d_m)$.

Giá trị trung bình (Mean) của vec-tơ *d* được tính theo công thức sau:

$$T = Mean(d) = \frac{\sum_{i=1}^{m} d_i}{m}, \quad (6)$$

Độ lệch chuẩn các giá trị của vec-tơ *d*, ký hiệu σ (sigma) thì σ được xác định như sau:

$$\sigma = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(d_i - T)^2}, \quad (7)$$

Trong ứng dụng của mình, chúng tôi tạm thời sử dụng ngưỡng xác thực là giới hạn cho giá trị của khoảng cách *d* là (T-σ, T+σ) để kiểm tra quá trình đăng nhập của người dùng.

Khi người dùng tiến hành đăng nhập, nếu khoảng cách nằm trong khoảng ngưỡng giá trị nói trên thì ta có thể kết luận là người dùng đó đã được ghi nhận trên hệ thống và đăng nhập thành công, ngược lại thì quá trình xác thực sẽ bị loại bỏ.

2.2.2. Xây dựng ứng dụng thử nghiệm

Chúng tôi đã xây dựng một ứng dụng Web để kiểm tra hoạt động và để xem tính đúng đắn của thuật toán tính khoảng cách và tính ngưỡng cũng như độ lệch ngưỡng (theo các công thức (5), (6), (7)) (xem thêm [8]).

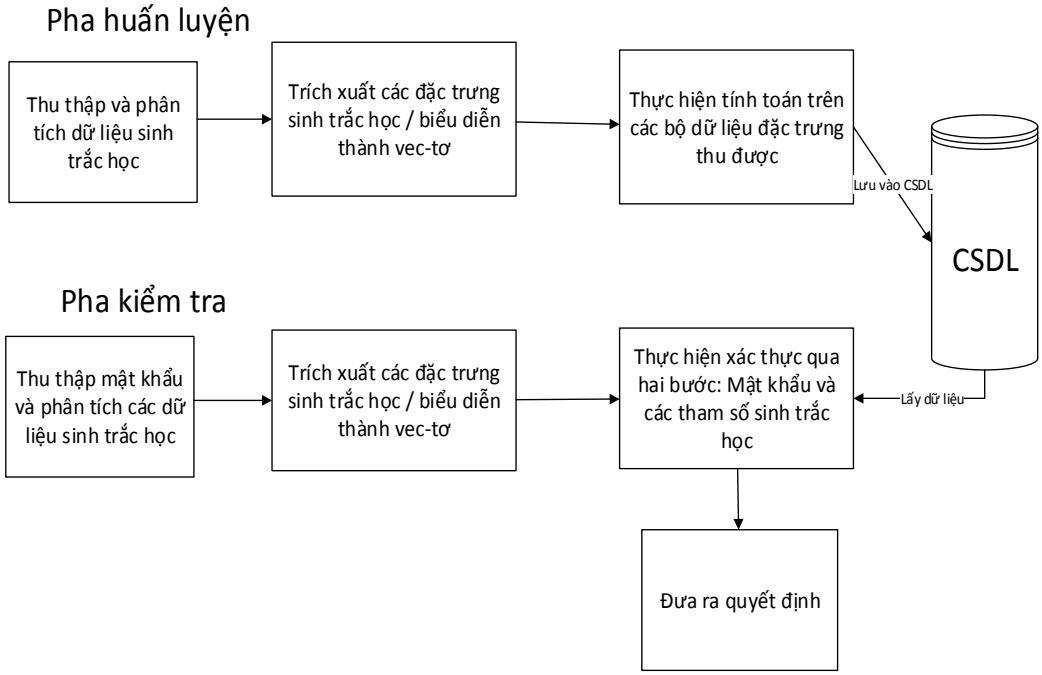
Hoạt động của ứng dụng được chia thành hai pha: ***pha huấn luyện*** và ***pha kiểm tra***.

Pha 1 – Pha thu thập và huấn luyện dữ liệu. Tại pha này, người dùng sau khi đăng ký sẽ được yêu cầu nhập mật khẩu để tiến hành quá trình huấn luyện. Dữ liệu được gửi lên gồm mật khẩu người dùng (plain text) và toàn bộ các thông tin về thời gian gõ bàn phím. Người dùng được yêu cầu phải nhập ít nhất 50 lần trở lên để

đảm bảo khoảng cách cũng như độ lệch ngưỡng sẽ có độ hội tụ cao hơn. Sau khi đủ số lần nhập cần thiết thì hệ thống sẽ tiến hành tính toán theo các công thức (1), (2), (3) để ra được vec-tơ đặc trưng của người dùng (vec-tơ median) và các giá trị ngưỡng, độ lệch ngưỡng.

Pha 2 – Pha kiểm tra. Tại pha này người dùng sẽ tiến hành đăng nhập, toàn bộ dữ liệu của người dùng sẽ được gửi lên trên máy chủ. Tại máy chủ sẽ diễn ra quá trình xác thực hai bước:

- Kiểm tra mật khẩu xem có tồn tại trong CSDL hay không? (normal password);
- Xác thực sinh trắc học, bước này chỉ được tiến hành sau khi đã vượt qua được bước đầu tiên. Sau khi tính toán khoảng cách và đối chiếu với khoảng giá trị của ngưỡng (mục 2.2.1), máy chủ sẽ đưa ra quyết định (decision maker) xem người dùng có đăng nhập thành công hay không.



Hình 2. *Mô hình huấn luyện dữ liệu và kiểm tra.*

3. MÔ PHỎNG, TÍNH TOÁN, THẢO LUẬN

Để đánh giá hiệu quả hoạt động của thuật toán cũng như hệ thống, chúng tôi tiến hành kiểm tra theo một số kịch bản xác định.

Kịch bản đầu tiên là chúng tôi tiến hành huấn luyện dữ liệu đồng loạt cho nhiều người dùng, chỉ khác nhau ở tên đăng nhập (tên đăng nhập trùng với tên người

dùng) và giống nhau ở mật khẩu "**.cntT2016@**". Kiểm tra này giúp chúng tôi nắm bắt được sự khác biệt về thói quen gõ bàn phím của mỗi người dùng. Những tình nguyện viên được yêu cầu nhập ít nhất 50 lần mật khẩu trên.

Bảng 2. *Kết quả sau khi huấn luyện người dùng.*

STT	Tên đăng nhập	Mật khẩu	Ngưỡng (T)	Độ lệch ngưỡng (σ)
1	cong	.cntT2016@	7.842	1.782
2	giap	.cntT2016@	12.734	2.947
3	longtv	.cntT2016@	9.234	1.943
4	ntngu	.cntT2016@	11.527	2.576
5	quy	.cntT2016@	11.854	2.385
6	can123	.cntT2016@	10.942	2.793
7	Tu	.cntT2016@	7.818	1.729
8	Hoai	.cntT2016@	8.860	1.788

Từ bảng dữ liệu thu được và đối chiếu với khả năng gõ bàn phím của mỗi người chúng tôi nhận thấy, những người có thói quen gõ bàn phím (được xác định bởi khả năng nhớ phím, tốc độ gõ phím) sẽ có ngưỡng nhỏ hơn những người khác, và tương ứng là độ lệch ngưỡng.

Sau đó quá trình kiểm tra đăng nhập được thực hiện. Và kết quả kiểm tra như sau.

Bảng 3. *Kết quả kiểm tra đăng nhập với cùng mật khẩu.*

STT	Tên đăng nhập	Số lần gõ	Thành công	Tỉ lệ (%)
1	cong	10	10	100
2	giap	10	9	90
3	longtv	10	9	90
4	ntngu	10	8	80
5	quy	10	8	80
6	can123	10	8	80

7	Tu	10	10	100
8	Hoai	10	7	70

Từ bảng kết quả chúng ta nhận thấy, có hai người thực hiện đăng nhập thành công 10/10 (=25%), có hai người thực hiện đăng nhập thành công 9/10 (=25%), có 3 người có tỉ lệ đăng nhập thành công 8/10 (=37%) và một người có tỉ lệ đăng nhập thành công 7/10 (=13%).

Kịch bản thứ hai là chúng tôi cho một người dùng tự huấn luyện cho tài khoản của mình "**Nghia/nghia123**" và có các giá trị sau khi huấn luyện tương ứng là T = 5.021, σ = 1.776. Rõ ràng là mật khẩu của người dùng này khá đơn giản, không chứa các ký tự in hoa hay các ký tự đặc biệt. Lúc này, những cộng tác viên được yêu cầu cùng thử đăng nhập bằng tài khoản này. Bản thân người dùng trên cũng tiến hành đăng nhập 20 lần vào tài khoản đó, kết quả thu được là:

Bảng 4. *Kết quả kiểm tra đăng nhập với tài khoản của người dùng.*

Bảng kết quả			
Người dùng chính		Người dùng khác	
Số lần đăng nhập	Số lần thành công	Số lần đăng nhập	Số lần thành công
20	17	15	0

Với người dùng chính: Tỉ lệ thành công đăng nhập là 17/20 = 85%; không thành công là 3/20 = 15%. Với người dùng khác (attacker): Tỉ lệ thành công là 0/15 = 0%.

Từ kết quả trên nhận thấy, rõ ràng đề có thể có một "*thói quen gõ bàn phím*" gần giống với một ai đó là điều không đơn giản. Dù có bị lộ các thông tin về tài khoản cá nhân thì việc có thể đăng nhập bằng tài khoản đó cũng sẽ khó thành công (ngay cả trong trường hợp mật khẩu của người dùng khá đơn giản như trong thí nghiệm); Tất nhiên, không loại trừ khả năng là sau khá nhiều lần thử, sẽ có một lần thành công, nhưng nhìn chung, con số này là không nhiều.

4. KẾT LUẬN

Trong nghiên cứu này chúng tôi đã đề xuất phương pháp tích hợp độ đo khoảng cách sử dụng cho mật khẩu sinh học trên nền ứng dụng Web. Đây là một kỹ thuật không phải là quá mới, song với việc áp dụng các kết quả nghiên cứu trước đó

chúng tôi đã bước đầu thu được những kết quả khả quan. Bên cạnh sử dụng các thiết bị hiện đại thì thì phương pháp xác thực KD sử dụng tính năng các thiết bị đơn giản (bàn phím máy tính, keypad của các thiết bị di động cảm ứng) vẫn chứng tỏ được sự hiệu quả và đảm bảo được tính bảo mật cần thiết. Trong tương lai, việc áp dụng mật khẩu sinh trắc học sẽ khá rộng rãi, nhất là trong giai đoạn hiện nay khi vấn đề bảo mật thông tin đang trở nên hết sức cấp thiết.

Việc sử dụng các đặc trưng sinh trắc học trong xác thực sẽ tăng cường tính bảo mật cho các máy chủ Web trong quá trình xác thực. So với phương pháp xác thực chỉ sử dụng tên đăng nhập và mật khẩu thì rõ ràng phương pháp xác thực này có tính bảo mật cao hơn nhiều. Do mỗi người dùng đều có những thói quen sử dụng máy tính, gõ bàn phím là khác nhau, do đó, những điều này sẽ tạo nên đặc trưng cho mỗi người dùng, và những điều đó có thể được sử dụng để bảo vệ họ khỏi những rủi ro trong quá trình thao tác trên mạng, đặc biệt là khi thực hiện các giao dịch phức tạp, cần độ bảo mật cao.

Từ những kết quả đã thu được, trong thời gian tới chúng tôi sẽ tiến hành ứng dụng trên các dòng thiết bị khác (điện thoại thông minh, máy tính bảng) và tiến hành thử nghiệm nhiều hơn từ các bộ dữ liệu thực tế thu được để tiến hành tối ưu về phương pháp lấy mẫu cũng như tốc độ tính toán khoảng cách cũng như tăng độ chính xác của quá trình xác thực này. Dữ liệu cũng sẽ được mã hóa trước khi gửi đi, đảm bảo việc hoàn toàn bí mật cho dữ liệu của người dùng, hướng tới việc áp dụng cho các hệ thống thực tế để tăng cường tính bảo mật mà vẫn tiết kiệm được chi phí trong việc nâng cấp cơ sở hạ tầng.

TÀI LIỆU THAM KHẢO

[1]. Haider, Sajjad, Ahmed Abbas, and Abbas K. Zaidi (2000). "*A multi-technique approach for user identification through keystroke dynamics.*" Systems, Man, and Cybernetics, 2000 IEEE International Conference on. Vol. 2. IEEE.

[2]. Monrose, Fabian, and Aviel D. Rubin (2000). "*Keystroke dynamics as a biometric for authentication.*" Future Generation computer systems 16.4. pp 351-359.

[3]. Yu, Enzhe, and Sungzoon Cho (2003). "*GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification*." Neural Networks, 2003. Proceedings of the International Joint Conference on. Vol. 3. IEEE.

[4]. Kang, Pilsung, Seong-seob Hwang, and Sungzoon Cho (2007). "*Continual retraining of keystroke dynamics based authenticator*." Advances in Biometrics. Springer Berlin Heidelberg. pp 1203-1211.

[5]. Lee, Jae-Wook, Sung-Soon Choi, and Byung-Ro Moon (2007). "*An evolutionary keystroke authentication based on ellipsoidal hypothesis space*." Proceedings of the 9th annual conference on Genetic and evolutionary computation. ACM.

[6]. Kevin S. Killourhy and Roy A. Maxion (2009). "*Comparing Anomaly Detectors for Keystroke Dynamics*." in Proceedings of the 39th Annual International Conference on Dependable Systems and Networks (DSN-2009), pages 125-134, Estoril, Lisbon, Portugal, June 29-July 2, 2009. IEEE Computer Society Press, Los Alamitos, California.

[7]. Mrs, D. Shanmugapriya, G. Padmavathi (2009). "*A Survey of Biometric keystroke Dynamics: Approaches, Security and Challenges*". International Journal of Computer Sciene and Information Security, Vol. 5, No. 1.

[8]. Giot, Romain, Mohamad El-Abed, and Christophe Rosenberger (2012). "*Web-based benchmark for keystroke dynamics biometric systems: A statistical analysis*."Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on. IEEE.

[9]. Zhong, Yu, Yan Deng, and Anubhav K. Jain (2012). "*Keystroke dynamics for user authentication.*" Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE.

[10]. Antal, Margit, László Zsolt Szabó, and Izabella László (2015). "*Keystroke dynamics on android platform*." Procedia Technology 19. pp 820-826.

[11]. Trần Nguyên Ngọc, Nguyễn Hữu Nội (2015). "*Mật khẩu sinh học – Hướng tiếp cận mới cho thao tác gõ bàn phím*". ICT 2015, 35-38.

[12]. http://www.cs.cmu.edu/~keystroke/

[13]. Parkhi O. M., Vedaldi A., Zisserman "*A Deep face recognition*". British Machine Vision Conference. – 2015. – T. 1. – №. 3. – C. 6.

[14]. Zirjawi N., Kurtanovic Z., Maalej W. "*A survey about user requirements for biometric authentication on smartphones*". Evolving Security and Privacy Requirements Engineering (ESPRE), 2015 IEEE 2nd Workshop on. – IEEE, 2015. – C. 1-6.

[15]. De Luca "*A. et al. I Feel Like I'm Taking Selfies All Day!: Towards Understanding Biometric Authentication on Smartphones*". Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. – ACM, 2015. – C. 1411-1414.

[16]. Nguyen Ngoc Tran. *"Distance-based classification of keystroke dynamics."*First International Workshop on Pattern Recognition. International Society for Optics and Photonics, 2016.

ABSTRACT

USER AUTHENTICATION USING KEYSTROKE DYNAMICS

This paper presents a study of using keystroke dynamics for user authentication towards the information security purpose. In this paper, a method for authentication combined normal text and bio-password based on Web-application is also proposed. The user information collected through keystroke dynamics on devices (computer, smartphone) were ordered into feature vectors and sent to the server for further processing. Then, the distance metrics were calculated with other parameters (threshold – T and threshold deviation – σ) on server side for user authentication.

Keywords: Authentication, Biometrics, Keystroke dynamics.

Nhận bài ngày 31 tháng 02 năm 2017
Hoàn thiện ngày 14 tháng 4 năm 2017
Chấp nhận đăng ngày 01 tháng 5 năm 2017

Địa chỉ: [1] Học viện Kỹ thuật quân sự, Bộ Quốc phòng;
[2] Cục Quản lý xuất nhập cảnh / Bộ Công an;
[*] Email: huunoidq@gmail.com