Đánh giá độ mạnh mật khẩu sử dụng ngôn ngữ tiếng Việt dựa trên ước lượng entropy

Hoàng Thu Phương, Trần Sỹ Nam

***Tóm tắt*— Mật khẩu là một trong những nhân tố được sử dụng phổ biến nhất trong hệ thống xác thực. Vai trò của mật khẩu là đảm bảo người dùng có quyền hợp lệ với dữ liệu mà họ đang muốn truy cập. Hầu hết các hệ thống đều cố gắng thực thi bảo mật bằng cách bắt buộc người dùng tuân theo các chính sách tạo mật khẩu thông qua đánh giá độ mạnh mật khẩu. Bài báo này giới thiệu một số phương pháp đánh giá độ mạnh mật khẩu trong đó tập trung vào phương pháp đánh giá dựa trên ước lượng entropy, từ đó đề xuất phát triển một công cụ đánh giá độ mạnh mật khẩu có thể ứng dụng được trong các phần mềm xác thực người dùng dựa trên mật khẩu sử dụng ngôn ngữ tiếng Việt.**

***Abstract*— Password is one of the most common means of authentication systems. The role of the password is to ensure that the user has legal right to the data they are trying to access. Most systems try to enforce security by requiring their users to follow some password generation policies with evaluating password strength. This paper introduces a number of methods for evaluating password strength, particularyly the entropy estimation based method, then, it is proposed to develop a password strength evaluation tool that can be applied in password-based user authentication software using Vietnamese language.**

***Từ khóa:* — độ mạnh mật khẩu; xác thực; đánh giá; ngôn ngữ tiếng Việt.**

***Keywords*: password strength; authentication; evaluation; Vietnamese language.**

Bài báo được nhận ngày 15/11/2018. Bài báo được gửi nhận xét và được chấp nhận đăng bởi phản biện thứ nhất vào ngày 04/12/2018 và 26/12/2018. Bài báo được gửi nhận xét và được chấp nhận đăng bởi phản biện thứ hai vào ngày 05/12/2018 và 28/12/2018.

I. GIỚI THIỆU

Việc sử dụng mật khẩu đã được biết đến từ xa xưa. Trong thời hiện đại, tên người dùng và mật khẩu thường được sử dụng trong quá trình đăng nhập vào các hệ điều hành máy tính, điện thoại di động, bộ giải mã truyền hình cáp, máy rút tiền tự động (ATM), v.v... Một người dùng máy tính thông thường có mật khẩu cho nhiều mục đích: đăng nhập vào tài khoản, lấy e-mail, truy cập các ứng dụng, cơ sở dữ liệu, mạng, trang web và thậm chí đọc báo buổi sáng trực tuyến [1].

Trong các vấn đề về an toàn phổ biến của mật khẩu [2], độ mạnh mật khẩu là vấn đề được quan tâm hàng đầu. Độ mạnh của mật khẩu là một thước đo hiệu quả khả năng chống lại các tấn công đoán hoặc vét cạn của mật khẩu. Nói một cách đơn giản, nó ước lượng số lần thử nghiệm trung bình mà kẻ tấn công sẽ cần để đoán chính xác mật khẩu đó. Thông thường, độ mạnh mật khẩu được xác định bằng entropy của thông tin và được đo bằng bit [3]. Thay vì số lần đoán cần thiết để tìm ra mật khẩu một cách cụ thể, giá trị logarit cơ số 2 của số đó được đưa ra, đó là số "bit entropy" của mật khẩu đó.

Căn cứ vào cách thiết lập, mật khẩu có thể được phân thành hai loại: *Mật khẩu ngẫu nhiên* và *mật khẩu được người dùng lựa chọn*. *Mật khẩu ngẫu nhiên* là mật khẩu bao gồm một chuỗi các ký tự có độ dài xác định được lấy từ một tập hợp các ký tự sử dụng một quy trình lựa chọn ngẫu nhiên, trong đó mỗi ký tự có khả năng được lựa chọn như nhau. Để lựa chọn ngẫu nhiên một mật khẩu có độ dài k bit thì sẽ có thể có 2^k khả năng và mật khẩu đó được coi là có k bit entropy. Nếu mật khẩu có độ dài l ký tự được chọn ngẫu nhiên từ b ký tự trong một bảng ký tự nào đó thì entropy của mật khẩu đó là b^l [4].

Tuy nhiên, đối *với mật khẩu mà người dùng lựa chọn*, việc ước lượng entropy là khó khăn hơn nhiều, vì chúng không được chọn ngẫu nhiên và không có phân phối ngẫu nhiên đồng nhất. Các mật khẩu được người dùng lựa chọn

có thể phản ánh một cách khái quát các mẫu từ và phân phối tần suất của các ký tự trong văn bản tiếng Anh thông thường và được người dùng lựa chọn để họ có thể ghi nhớ chúng. Thực nghiệm còn cho thấy rằng nhiều người dùng còn chọn mật khẩu dễ đoán và thậm chí là mật khẩu xuất hiện trong các từ điển thông dụng, rất dễ dàng để có thể bẻ khóa thành công.

Các nhà mật mã học đã đưa ra khái niệm thay thế của entropy, "entropy ước lượng" để làm thước đo độ khó trong việc đoán hay xác định mật khẩu được người dùng lựa chọn hoặc khóa [4].

Trong nghiên cứu này, trên cơ sở tìm hiểu các phương pháp đánh giá độ mạnh mật khẩu đã được công bố, chúng tôi tập trung vào phân tích phương pháp dựa trên ước lượng entropy và từ đó đề xuất một phương pháp đánh giá độ mạnh mật khẩu. Chúng tôi đã nghiên cứu và thiết lập các ngưỡng đánh giá độ an toàn dựa trên giá trị entropy ước lượng của mật khẩu đồng thời xây dựng một từ điển tiếng Việt để sử dụng làm cơ sở dữ liệu đánh giá độ mạnh của mật khẩu dựa trên tiếng Việt. Ngoài ra, chúng tôi cũng đã sử dụng phương pháp được đề xuất này để đánh giá một số danh sách mật khẩu nổi tiếng và thử nghiệm cài đặt mô-đun đánh giá độ mạnh mật khẩu vào một số phần mềm để chứng minh khả năng ứng dụng của nó trong các phần mềm có xác thực người dùng dựa trên mật khẩu.

Trong các phần tiếp theo của bài báo, trước tiên, mục II giới thiệu các nghiên cứu có liên quan trong lĩnh vực độ mạnh mật khẩu; sau đó, phương pháp đánh giá độ mạnh mật khẩu dựa trên ước lượng entropy sẽ được thảo luận trong mục III; mục IV trình bày đề xuất phương pháp đánh giá độ mạnh mật khẩu có thể ứng dụng trong các phần mềm xác thực người dùng dựa trên mật khẩu có sử dụng ngôn ngữ tiếng Việt; một số kết quả thử nghiệm phương pháp được đề xuất được trình bày trong mục V và cuối cùng mục VI là một số kết luận

II. CÁC NGHIÊN CỨU LIÊN QUAN ĐẾN ĐỘ MẠNH MẬT KHẨU

A. Các tấn công lên mật khẩu

Các mối đe dọa đối với độ an toàn của mật khẩu có thể xuất phát từ các yếu tố liên quan đến con người hoặc hệ thống xác thực được sử dụng. Tấn công đoán mật khẩu có thể diễn ra dưới hai hình thức là *tấn công trực tuyến* và *tấn công ngoại tuyến*. Trong *tấn công trực tuyến*, kẻ tấn công sẽ cố gắng đoán mật khẩu trong giai đoạn đăng nhập. Trong khi đó, *tấn công ngoại tuyến* là hình thức tấn công trong đó kẻ tấn công đánh cắp giá trị băm của mật khẩu được lưu trữ trong một máy chủ và thử đoán với số lần không giới hạn mà không cần tương tác trực tiếp với máy chủ. Tài liệu [5] đã trình bày một số tấn công đoán mật khẩu quan trọng như sau:

Tấn công vét cạn. Kẻ tấn công cố gắng đoán mật khẩu thông qua một tìm kiếm toàn diện trên tập hợp các chuỗi kết hợp có thể được tạo ra.

Tấn công từ điển. Khác với tấn công vét cạn, tấn công này tập trung vào thử các từ trong một danh sách dài gọi là từ điển được xây dựng sẵn để tăng tốc độ tìm kiếm. Các cơ sở dữ liệu mật khẩu bị rò rỉ và các từ điển được xây dựng từ các từ và cụm từ quen thuộc thường được kẻ tấn công sử dụng trong các tấn công từ điển.

Tấn công từ điển có chủ đích. Đây có thể được coi là một trường hợp đặc biệt của tấn công từ điển, trong đó từ điển được sử dụng có chứa thông tin cá nhân (ví dụ: tên người dùng, tên và họ, ngày sinh, v.v…) của người dùng. Kẻ tấn công sử dụng các thông tin này nhằm làm giảm số lần đoán cần thiết để tìm ra mật khẩu.

Tấn công từ điển dựa trên quy tắc. Kẻ tấn công sử dụng các quy tắc để biến đổi mật khẩu ví dụ như thêm tiền tố hoặc hậu tố trước khi thực hiện tấn công từ điển để xác định mật khẩu phức tạp hơn.

Tấn công từ điển kết hợp. Phương pháp này là gần nhất với những gì kẻ tấn công thực hiện trên thực tế, trong đó, tấn công vét cạn được áp dụng khi tấn công từ điển không phát huy hiệu quả.

Gần đây, các phương pháp tấn công tiên tiến hơn đã được đề xuất để cải thiện khả năng của các công cụ bẻ khóa mật khẩu. Các phương pháp này có thể được phân loại thành bốn kiểu dưới đây [5]:

Phương pháp dựa trên mô hình Markov. Các mô hình Markov được sử dụng để thu hẹp không gian tìm kiếm khi tấn công vét cạn cần được sử dụng [6].

Phương pháp ngữ pháp không phụ thuộc vào bối cảnh xác suất (Probabilistic Context-Free Grammars - PCFG). Phương pháp này

xem xét cấu trúc của mật khẩu trong đó xác suất của mật khẩu được lựa chọn với một cấu trúc nhất định cao hơn các mật khẩu khác có cấu trúc khác [7].

Phương pháp dựa trên học máy. Gần đây, các mạng nơ-ron đã được sử dụng trong dự đoán mật khẩu. Phương pháp này đã được thử nghiệm và cho thấy tính hiệu quả cao hơn so với các phương pháp được đề cập trước đó trong phần này [8].

Phương pháp dựa trên cá nhân hóa. Những cải tiến gần đây đối với một số tấn công kể trên xem xét thông tin cá nhân trong các mô hình và thuật toán của họ để cải thiện tính hiệu quả trong việc bẻ khóa mật khẩu, ví dụ như OMEN+ [9], Personal-PCFG [10].

B. Các công cụ đánh giá độ mạnh mật khẩu

Công cụ đánh giá mật khẩu là phần mềm được sử dụng để kiểm tra độ mạnh của mật khẩu đã cho nhằm phát hiện và/hoặc ngăn chặn việc sử dụng mật khẩu yếu [5].

Chức năng cơ bản của công cụ đánh giá mật khẩu là cung cấp phản hồi ngay lập tức về độ mạnh của mật khẩu mà người dùng đang nhập để người dùng có thể đưa ra quyết định sáng suốt hơn về việc mật khẩu hiện tại có đủ an toàn để được sử dụng hay không. Các phản hồi này thường được đưa ra một cách trực quan trên màn hình thiết bị máy tính của người dùng.

Trong Bảng 1 là 18 công cụ đánh giá mật khẩu được sử dụng rộng rãi và tổng kết một số yêu cầu cũng như tính năng phổ biến của các công cụ này.

BẢNG 1. CÁC YÊU CẦU VỀ MẬT KHẨU VÀ ĐẶC ĐIỂM CỦA CÁC CÔNG CỤ ĐÁNH GIÁ MẬT KHẨU

Phân loại	**Tên dịch vụ**	**Phạm vi đánh giá**	**Giới hạn độ dài**	**Yêu cầu về bộ ký tự**	**Thông tin người dùng**	**Vị trí cho phép ký tự khoảng trắng**	
						Bên ngoài	**Bên trong**
Phía máy khách trên nền tảng Web	Dropbox	5 mức	6 - 72	Không có	1 phần	√	√
	Drupal	4 mức	6 - 128	Không có	Không có	×	√
	Fed Ex	5 mức	8 - 35	1+ viết thường, 1+ viết hoa, 1+ số	Không có	×	×
	Intel	2 mức	> 1	Không có	Không có	√	√
	Microsoft	4 mức	> 1	Không có	Không có	×	×
	QQ	3 mức	6 - 16	Không có	Không có	×	×
	Twitter	6 mức	6 - >1000	Không có	1 phần	√	√
	Yahoo!	4 mức	6 - 32	Không có	1 phần	√	√
	12306.cn	3 mức	6 - 25	1+ bộ ký tự	Không có	×	×
Phía máy chủ trên nền tảng Web	eBay	4 mức	6 - 20	2 bộ ký tự bất kỳ	1 phần	×	√
	Google	4 mức	8 - 100	Không có	Không có	×	√
	Skype	3 mức	6 - 20	2 bộ ký tự hoặc viết hoa hoàn toàn	Không có	×	×
Kết hợp trên nền tảng Web	Apple	3 mức	8 - 32	1+ viết thường, 1+ viết hoa, 1+ số	1 phần	×	×
	PayPal	3 mức	8 - 20	2 bộ ký tự bất kỳ	Không có	×	×
Trên nền tảng ứng dụng	1Password	6 mức	> 1	Không có	Không có	√	√
	KeePass	0-128 bit	> 1	Không có	Không có	√	√
	LastPass	0-100%	> 1	Không có	1 phần	√	√
	RoboForm	3 mức	6 - 49	Không có	Không có	√	√

Yêu cầu về độ dài và tập hợp các ký tự. Hầu hết các công cụ đều đặt ra yêu cầu về số ký tự tối thiểu, một số công cụ còn đặt ra một giới hạn về độ dài tối đa. Ngoài một số yêu cầu sử dụng bộ ký tự nhất định, việc sử dụng ký tự khoảng trắng cũng có nhiều quy định khác nhau như không được phép sử dụng, được phép dưới dạng ký tự bên ngoài (ở đầu hoặc cuối mật khẩu) hoặc được phép sử dụng dưới dạng ký tự bên trong. Một số công cụ cũng không cho phép các ký tự liên tiếp giống hệt nhau (ví dụ: 3 ký tự đối với Apple và 4 đối với FedEx).

Thang đánh giá độ mạnh và nhãn. Thang độ mạnh và nhãn được sử dụng bởi các công cụ đánh giá mật khẩu cũng khác nhau. Ví dụ: cả Skype và PayPal chỉ có 3 mức đánh giá độ mạnh của mật khẩu (Yếu-Trung bình-Mạnh), trong khi Twitter có 6 mức (Quá ngắn-Rõ ràng-Không đủ an toàn-Có thể an toàn hơn-Ổn-Hoàn hảo).

Thông tin người dùng. Một số công cụ có xem xét các tham số liên quan đến người dùng, chẳng hạn như tên thật/tên tài khoản hoặc địa chỉ email. Lý tưởng nhất là mật khẩu chứa thông tin đó phải được coi là yếu (hoặc ít nhất là bị trừ trong cách tính điểm).

Phân loại. Dựa vào vị trí việc đánh giá được thực hiện, các công cụ dựa trên nền tảng web được phân loại thành: các công cụ phía máy khách (ví dụ: Dropbox, Drupal, FedEx); các công cụ phía máy chủ (ví dụ: eBay, Google và Skype); và

các công cụ kết hợp: kết hợp cả hai hình thức trên (ví dụ: Apple và PayPal). Ngoài ra còn có các công cụ được xây dựng trên nền tảng ứng dụng như 1Password, KeePass, LastPass, RoboForm.

Tính đa dạng. Mỗi dịch vụ web và các trình quản lý mật khẩu đều cung cấp công cụ đánh giá độ mạnh mật khẩu riêng, tuy nhiên không có bất kỳ lời giải thích nào về cách thức hoạt động của chúng hay cách các tham số độ mạnh được quy định.

Ước lượng entropy và danh sách đen. Hầu hết các công cụ đánh giá mật khẩu đều sử dụng một bộ tính toán entropy tùy chỉnh dựa trên độ phức tạp và độ dài của mật khẩu. Các tham số của mật khẩu thường được xem xét để tính toán entropy/điểm bởi các công cụ khác nhau bao gồm: chiều dài, tập hợp các ký tự được sử dụng và các mẫu phổ biến. Một số công cụ còn so sánh mật khẩu đã cho với một từ điển các mật khẩu phổ biến (dưới dạng danh sách đen) và giảm đáng kể điểm của mật khẩu nếu nó xuất hiện trong các danh sách này.

C. Các phương pháp đánh giá độ mạnh mật khẩu

Gần đây, với sự phát triển nhanh chóng của các kỹ thuật tấn công mật khẩu, vấn đề về độ mạnh mật khẩu đã và đang nhận được nhiều sự quan tâm. Các chuyên gia mật khẩu phần lớn đồng ý rằng độ mạnh của một mật khẩu nên tương ứng với nỗ lực tối thiểu cần thiết để bẻ khóa mật khẩu đó [5].

Có một số phương pháp ước lượng độ mạnh của mật khẩu đã được đề xuất. Các phương pháp này có thể được phân loại thành *phương pháp dựa trên thống kê* và *phương pháp tham số hóa* (hay còn gọi là phương pháp dựa trên xác suất). Các phương pháp dựa trên thống kê tập trung vào dự đoán độ mạnh tổng thể của mật khẩu trong khi đó trọng tâm của các phương pháp dựa trên xác suất là xác định khả năng dự đoán được của mật khẩu dựa trên một kỹ thuật bẻ khóa cụ thể nào đó.

Các phương pháp dựa trên thống kê. Các phương pháp dựa trên thống kê được phân thành hai loại là các *phương pháp dựa trên entropy* và các *phương pháp dựa trên tính có thể dự đoán được*. *Phương pháp dựa trên entropy* là một phép đo cơ bản được sử dụng bởi nhiều trang web để kiểm tra độ mạnh của mật khẩu dựa trên các quy tắc đơn giản liên quan đến định dạng mật khẩu, chẳng hạn như độ dài mật khẩu và các kiểu ký tự khác nhau được sử dụng. Trong các phương pháp thuộc loại này, zxcvbn [11] của Dropbox được đánh giá là sử dụng các thuật toán phức tạp hơn cả và cho kết quả chính xác hơn.

Phương pháp dựa trên tính có thể dự đoán được là phương pháp được một số nghiên cứu đề xuất dựa trên một tấn công lý tưởng trong đó, mật khẩu có xác suất cao nhất đã được thử đầu tiên, và do đó, tính có thể dự đoán được được xác định bằng số lần thử không thành công trước khi đoán đúng mật khẩu. Tuy nhiên, phương pháp này có hai vấn đề lớn: (1) hiệu quả của chúng không thể được khái quát hóa vì nó

sẽ bị ảnh hưởng bởi các cài đặt cấu hình và tấn công được lựa chọn; (2) hiệu suất về mặt thời gian của chúng sẽ bị giảm do phải cùng lúc thực hiện tấn công. Vì những hạn chế này, các phương pháp này có thể không phù hợp để được sử dụng trong các công cụ kiểm tra mật khẩu chủ động.

Phương pháp tham số hóa. Các phương pháp này cố gắng khắc phục hạn chế của các phương pháp dựa trên thống kê ở trên, thay vì các tấn công lý tưởng, chúng xem xét đánh giá độ an toàn của mật khẩu chống lại các tấn công thực sự như đã được trình bày trong phần II.A.

Tuy nhiên, một nghiên cứu của tác giả S. Ji và cộng sự đã cho thấy rằng không có cơ chế bẻ khóa nào là lý tưởng đối với tất cả các trường hợp vì hiệu suất của chúng bị ảnh hưởng bởi các yếu tố khác nhau, bao gồm cơ sở dữ liệu huấn luyện và thuật toán được sử dụng [12]. Các nghiên cứu tương tự cũng đã nhấn mạnh rủi ro khi tin tưởng vào một thuật toán bẻ khóa duy nhất để xác định độ mạnh của mật khẩu.

Do các hạn chế và rủi ro đã được chứng minh của các phương pháp xác suất dựa trên tính không thể dự đoán được và các phương pháp tham số hóa như đã được đề cập ở trên, phần tiếp theo của bài báo sẽ tập trung phân tích phương pháp đánh giá độ mạnh mật khẩu bằng xác suất dựa trên entropy của thuật toán zxcvbn, phương pháp mà theo các phân tích ở trên là phù hợp để có thể được triển khai trong các công cụ kiểm tra mật khẩu.

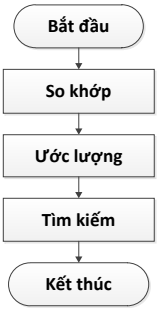
III. PHƯƠNG PHÁP ĐÁNH GIÁ ĐỘ MẠNH MẬT KHẨU DỰA TRÊN ENTROPY – THUẬT TOÁN ZXCVBN

Phần này trình bày chi tiết cơ sở lý thuyết của thuật toán zxcvbn, trong đó mục A) giới thiệu các bước trong mô hình tổng quát và các mục tiếp theo lần lượt đi sâu phân tích từng bước trong mô hình này.

A. Mô hình

Zxcvbn thực hiện ước lượng độ mạnh của một mật khẩu đầu vào thông qua ba bước: so khớp, ước lượng và tìm kiếm được thực hiện lần lượt [13] như được thể hiện trong Hình 1.

So khớp. Bước này liệt kê tất cả các mẫu (có thể trùng lặp) mà nó có thể phát hiện như so khớp từ điển, mẫu tổ hợp ký tự bàn phím, chuỗi lặp lại (aaa), chuỗi có trình tự (123, gfedcba), năm và ngày, tháng. Đối với tất cả các từ điển, so khớp được thực hiện cả với các thay thế leet đơn giản (leet là hệ thống các cách viết thay thế ký tự dựa trên sự giống nhau về hình dáng hoặc các sự tương đồng khác của chúng, ví dụ như sử dụng chuỗi "1337" thay thế cho "leet").



Hình 1. Lưu đồ thuật toán ước lượng độ mạnh mật khẩu của phương pháp zxcvbn

Ước lượng. Bước này tính entropy của từng mẫu đã được so khớp ở bước trên, độc lập với phần còn lại của mật khẩu.

Tìm kiếm. Với tất cả các nhóm các so khớp trùng lặp có thể xảy ra, bước tìm kiếm sẽ tìm chuỗi không trùng lặp đơn giản nhất (entropy tổng của các mẫu tạo thành chuỗi là thấp nhất).

B. Bước so khớp

Các kiểu mẫu so khớp được zxcvbn tìm kiếm được thể hiện trong Bảng 2 [11]:

BẢNG 2. CÁC KIỂU SO KHỚP CỦA ZXCVBN

Mẫu	**Ví dụ**
token	logitech, 10giT3CH, ain't, parliamentarian, 1232323q
token đảo ngược	DrowssaP
chuỗi	123, 2468, jklm, ywusq
lặp lại	zzz, abababab, 10giT3CH10giT3CH
tổ hợp ký tự bàn phím	qwertyuio, qAzxcde3, diueoa
ngày tháng	7/8/1847, 8.7.47, 781947, 4778, 7-21-2011, 72111, 11.7.21
ngẫu nhiên	X$JQhMzt

Trình so khớp token sẽ chuyển mật khẩu đầu vào thành chuỗi viết thường hoàn toàn và sau đó kiểm tra xem từng chuỗi con của nó có xuất hiện trong các từ điển được xếp hạng theo tần suất hay không. Ngoài ra, nó còn thực thi các thay thế leet đơn giản.

Trình so khớp chuỗi tìm kiếm các chuỗi trong đó cho phép cách quãng, chẳng hạn như

trong chuỗi "7531", và nhận ra các chuỗi nằm ngoài bảng chữ cái La Mã và các chữ số Ả Rập, chẳng hạn như các chuỗi Kirin (bảng chữ cái được sử dụng cho nhiều ngôn ngữ ở miền Đông Âu, Bắc và Trung Á.

Trình so khớp lặp lại tìm kiếm các khối lặp lại của một hoặc nhiều ký tự, sử dụng cả các biểu thức chính quy thông thường kiểu greedy "/(.+)\1+/" và kiểu lazy "/(.+?)\1+/" để tìm kiếm các vùng lặp lại bao gồm nhiều ký tự nhất. Trình so khớp lặp lại thực thi vòng so khớp – ước lượng – tìm kiếm theo cách đệ quy trên đơn vị lặp lại được chọn trước đó của nó.

Trình so khớp tổ hợp ký tự bàn phím tìm kiếm các chuỗi phím liền kề theo từng sơ đồ bàn phím liền kề của nó. Các biểu đồ này được biểu diễn dưới dạng ánh xạ giữa mỗi phím đến một danh sách phím lân cận có thể theo chiều kim đồng hồ của nó. Trình so khớp đếm chiều dài chuỗi, số đường đi và số ký tự được kết hợp với phím shift. Zxcvbn có thể so khớp các mẫu theo bàn phím QWERTY, DVORAK và sơ đồ bàn phím của Windows và Mac.

Trình so khớp ngày tháng sẽ xem xét các vùng chữ số gồm 4 đến 8 ký tự và tìm kiếm các phân tách có thể sao cho năm có hai hoặc bốn chữ số, năm không ở giữa, tháng bao gồm từ 1 đến 12, và ngày bao gồm từ 1 đến 31. Với nhiều phân tách hợp lệ, năm gần nhất với năm tham chiếu (năm hiện tại) sẽ được ưu tiên lựa chọn. Năm được ghi bởi hai chữ số được khớp với các năm trong thế kỷ 20 hoặc 21, tùy thuộc vào năm nào gần với năm tham chiếu hơn.

C. Bước ước lượng

Ở bước này, ước lượng số lần đoán cho mỗi so khớp sẽ được xác định [11].

Đối với các token, giá trị ước lượng chính là vị trí xếp hạng của từ đó trong từ điển mà nó được tìm thấy, bởi vì đó là số lần thử ít nhất mà kẻ tấn công đoán token theo thứ tự phổ biến sẽ cần. Token đảo ngược được ước lượng nhân đôi, vì kẻ tấn công sẽ cần thử hai lần đoán (bình thường và đảo ngược) đối với mỗi token.

Ở đây, nếu trong token đó có ký tự viết hoa được sử dụng, một điểm cộng sẽ được thêm vào tùy theo vị trí của ký tự được viết hoa. Nếu token này sử dụng các lược đồ viết hoa phổ biến (ký tự viết hoa là ký tự đầu tiên hoặc là ký tự cuối cùng hoặc viết hoa tất cả các ký tự) thì chỉ làm tăng gấp đôi không gian tìm kiếm của kẻ tấn công tương tự như trường hợp token đảo ngược, do đó ước lượng của mẫu này cũng được xác định là hai lần vị trí xếp hạng của token đó. Còn nếu token này sử dụng ký tự viết hoa theo cách khác thì số lần đoán trung bình được xác định theo công thức dưới đây:

$$\frac{1}{2}\sum_{i=1}^{\min(U,L)} C_{U+L}^{i} \quad (1)$$

Trong đó, U và L lần lượt là số lượng các ký tự viết hoa và viết thường trong token, i là số ký tự viết hoa trong lần đoán hiện tại. 1/2 là để chuyển đổi tổng không gian đoán thành nỗ lực trung bình cần thiết, giả sử rằng mỗi lược đồ viết hoa đều có khả năng như nhau. Thuật ngữ min() có ý nghĩa là nếu số lượng ký tự được viết hoa nhiều hơn số lượng ký tự viết thường thì token sẽ được chuyển đổi thành viết hoa hoàn toàn trước khi tính toán ước lượng.

Ví dụ, để đoán "paSswOrd" – token bao gồm 8 ký tự trong đó có 2 ký tự viết hoa (U=2), 6 ký tự viết thường (L=6), trước tiên, kẻ tấn công sẽ thử đoán token này có 1 ký tự viết hoa, với i =1, số vị trí có thể của ký tự viết hoa sẽ là $C_8^1 = 8$, sau đó tiếp tục đoán token này có 2 ký tự viết hoa, với i=2, số cách phân bổ 2 ký tự viết hoa vào 8 vị trí có thể là $C_8^2 = 28$. Như vậy, số lần đoán trung bình cho lược đồ viết hoa của token "paSswOrd" là: $\frac{1}{2}(8 + 28) = 18$.

Ngoài ra, nếu token có sử dụng thay thế leet, một điểm cộng khác cũng sẽ được tính tương tự công thức (1) trong đó U và L lần lượt là số ký tự được thay thế và không được thay thế leet trong token.

Đối với các mẫu tổ hợp ký tự bàn phím, số lần đoán được ước lượng bằng công thức sau:

$$\frac{1}{2}\sum_{i=2}^{L}\sum_{i=1}^{\min(T,i-1)} C_{i-1}^{j-1} SD^{j} \quad (2)$$

Trong đó, L là chiều dài (số ký tự) của mẫu, T là số lượng đường đi, D là số phím liền kề trung bình của mỗi phím (phím "~" có 1 phím liền kề trên bàn phím QWERTY, phím "a" có 4) và S là số lượng phím trên bàn phím. Đối với một mẫu tổ hợp bàn phím có độ dài L và T đường đi, giả sử rằng kẻ tấn công sẽ bắt đầu bằng việc đoán các mẫu có ít ký tự hơn và số đường đi ít hơn trước, bắt đầu từ chiều dài bằng 2. Thuật ngữ min() là để tránh xem xét nhiều lượt hơn có thể đối với các mẫu có độ dài ngắn hơn. Với i là chiều dài và j là số đường đi của

mẫu trong lần đoán hiện tại, bắt đầu từ một ký tự đầu tiên, có C_{i-1}^{j-1} cách lựa chọn ký tự chuyển hướng (với −1 được thêm vào mỗi biến vì đường đi đầu tiên được xác định là xảy ra trên ký tự đầu tiên). Vì chuỗi có thể đã bắt đầu trên bất kỳ phím nào trong S và mỗi đường đi có thể là bất kỳ cách nào trong D, do đó có SD^j.

Công thức trên được sử dụng để ước lượng số lần đoán của chuỗi "kjhgfdsa" trên bàn phím QWERTY như sau: Ta có L=8, T=1, S=47, $D = \frac{7\times6+1\times4}{8} = 5{,}75$ (số ký tự liền kề của ký tự "a" là 4 còn của các ký tự còn lại là 6). Vì $\min(1, i-1) = 1$, nên j chỉ nhận giá trị duy nhất là 1. Với $j = 1$, $C_{i-1}^{j-1} \times 47 \times 5{,}75^j = C_0^0 \times 47 \times 5{,}75^0 = 270{,}25$. Do đó, số lần đoán trung bình của mẫu này là: $\frac{1}{2}\sum_{i=2}^{8}\sum_{j=1}^{\min(1,i-1)} C_{i-1}^{j-1} \times 47 \times 5{,}75^j = \frac{1}{2} \times 7 \times 270{,}25 = 945{,}875 \approx 10^3$. Công thức (1) cũng sẽ được áp dụng để tính điểm cộng nếu trong mẫu có sử dụng các phím được kết hợp với phím Shift, khi đó L và U trở thành số lượng các phím có kết hợp và không kết hợp với phím Shift.

Các đối tượng so khớp lặp lại được phân tích thành một cơ sở được lặp lại n lần, trong đó các bước so khớp – ước lượng – tìm kiếm đệ quy trước đó được thực hiện lại với một số lần đoán cơ sở g. Do đó, số lần đoán mẫu lặp lại được ước lượng là $g \times n$.

Các mẫu chuỗi được tính điểm theo công thức $s \times n \times |d|$, trong đó s là số ký tự bắt đầu có thể, n là độ dài của chuỗi và d là độ lệch (ví dụ: $d = -2$ trong mẫu "9753").

Đối với mẫu ngày tháng, công thức tính số lần đoán là $365 \times |\text{reference year} - year|$.

Cuối cùng, *các so khớp bruteforce* có độ dài l được gán một hằng số C = 10 lần đoán cho mỗi ký tự, mang lại giá trị ước lượng tổng là C^l.

D. Bước tìm kiếm

Với một mật khẩu chuỗi và một tập hợp các so khớp trùng lặp tương ứng δ, bước cuối cùng là tìm kiếm chuỗi kết hợp S liền kề không trùng lặp bao gồm hoàn toàn mật khẩu và thỏa mãn biểu thức sau [1]:

$$\arg\min_{S\subseteq\delta} D^{|S|-1} + |S|! \prod_{m\in S} m.guesses \quad (3)$$

Trong đó, |S| là độ dài của chuỗi S, D là một hằng số. Giả sử rằng chuỗi được chọn S được cấu thành từ các chuỗi con m với số lần đoán cần thiết của mỗi chuỗi con là *m.guesses*, phép Π tính số lần đoán mà người đó cần thực hiện trong trường hợp xấu nhất. $|S|!$ được thêm vào vì kẻ tấn công chỉ biết số lượng mẫu mà không biết thứ tự xuất hiện của chúng. Ví dụ, nếu mật khẩu bao gồm một từ phổ biến c, một từ không phổ biến u và một ngày d, có *3!* thứ tự có thể thử: cud, ucd, v.v…

$D^{|S|-1}$ mô hình hóa một kẻ tấn công không biết về độ dài của chuỗi mẫu. Giả sử rằng trước khi thử các chuỗi có độ dài |S|, kẻ tấn công sẽ cố gắng thử các chuỗi mẫu có độ dài ngắn hơn trước với giá trị tối thiểu là D lần đoán đối với mỗi mẫu, và sẽ mất tổng cộng $\sum_{i=1}^{|S|-1} D^i = D^1 + D^2 + \cdots + D^{|S|-1} \approx D^{|S|-1}$ lần đoán. Ví dụ, nếu một mật khẩu bao gồm token mật khẩu t phổ biến thứ 20 với một chữ số d ở cuối –một chuỗi gồm 2 mẫu – và kẻ tấn công biết D = 10000 mật khẩu phổ biến nhất và td không ở trong danh sách 10000 này, D^1 biểu thị một kẻ tấn công thử 10000 lần đoán trong danh sách này trước khi thử đoán hai mẫu.

zxcvbn được đánh giá là xem xét thành phần của mật khẩu kỹ lưỡng hơn tất cả các công cụ kiểm tra khác trong các thử nghiệm của [3], dẫn đến việc đánh giá thực tế hơn về độ phức tạp của mật khẩu cho trước. Tuy nhiên, có quá nhiều mẫu khác nhau người dùng có thể sử dụng để tạo mật khẩu mà zxcvbn không thể nhận biết hết được, ví dụ như các từ đã bị bỏ đi chữ cái đầu, các từ không có nguyên âm, các từ đánh vần sai, n-gram, mã zip của các khu vực, các tổ hợp bàn phím cách xa nhau ví dụ như qzwxec, v.v… [13]. Nhóm phát triển bộ công cụ này cũng đã đưa ra khuyến nghị rằng việc bổ sung các mẫu ví dụ như các bộ từ điển và cụm từ thông dụng trong các ngôn ngữ khác ngoài tiếng Anh, các mẫu tổ hợp bàn phím theo các sơ đồ bàn phím khác, v.v.. trong bước so khớp chính là một cách hiệu quả để cải thiện tính hiệu quả cho bộ công cụ này.

IV. ĐỀ XUẤT PHƯƠNG PHÁP ĐÁNH GIÁ ĐỘ MẠNH MẬT KHẨU

Phần này sẽ trình bày đề xuất phương pháp đánh giá độ mạnh mật khẩu được phát triển dựa trên ước lượng entropy theo thuật toán zxcvbn đã được phân tích ở trên.

Mã nguồn chương trình zxcvbn đã được phát triển thêm một số nội dung để biến nó trở

thành một công cụ đánh giá độ mạnh mật khẩu sử dụng được trong các hệ thống xác thực người dùng dựa trên mật khẩu có sử dụng tiếng Việt. Các nội dung đã được phát triển thêm bao gồm: tính toán ngưỡng an toàn theo entropy của mật khẩu và tích hợp mô-đun đánh giá độ mạnh mật khẩu theo thang điểm; và tích hợp từ điển tiếng Việt vào dữ liệu bước so khớp.

Các mục tiếp theo đây sẽ lần lượt phân tích chi tiết các nội dung đã được bổ sung vào mã nguồn chương trình zxcvbn trên.

A. Tính toán ngưỡng an toàn theo entropy của mật khẩu và tích hợp mô-đun đánh giá độ mạnh mật khẩu theo thang điểm

Zxcvbn cung cấp giá trị ước lượng entropy của một mật khẩu đầu vào. Tuy nhiên, với một người dùng thông thường, giá trị entropy này cũng không thể giúp họ đánh giá được mật khẩu mà họ lựa chọn đã đủ an toàn hay chưa. Để cụ thể hơn, cần phải có một mô-đun đánh giá mức độ mạnh yếu của mật khẩu theo thang điểm. Căn cứ đánh giá độ mạnh của một mật khẩu theo giá trị entropy được dựa vào là RFC 4086 – "Các yêu cầu về tính ngẫu nhiên đối với bảo mật" [14] năm 2004. Tài liệu này đã trình bày một số mô hình tấn công lên các giá trị bí mật (mật khẩu, khóa mật mã) và entropy cần thiết đối với từng mô hình. Kết quả là để chống lại các tấn công trực tuyến, một mật khẩu có 29 bit entropy được đánh giá là đủ an toàn. Tuy nhiên, đây là giá trị cần thiết tại năm 2004. Theo định luật Moore, mỗi năm giá trị này chỉ cần thêm *2/3* bit, có nghĩa là đến năm 2019, giá trị entropy tối thiểu để mật khẩu được đánh giá là đủ an toàn trước các tấn công trực tuyến phải là: $29 + \frac{2}{3} \times (2019 - 2004) = 39$.

Vì $2^{39} \approx 10^{12}$ nên các ngưỡng đánh giá độ mạnh của mật khẩu dựa trên số lần đoán cần thiết để tìm ra một mật khẩu được thiết lập như sau:

- Rất yếu (0 điểm) đối với các mật khẩu có thể bị bẻ khóa với chưa đến 10^3 lần đoán;
- Yếu (1 điểm) nếu số lần đoán cần thiết nằm trong khoảng 10^3-10^6;
- Trung bình (2 điểm) nếu cần từ 10^6-10^9 lần đoán;
- Khá mạnh (3 điểm) nếu cần từ 10^9-10^{12}lần và;
- Đủ mạnh (4 điểm) đối với các mật khẩu cần trên 10^{12} lần đoán.

B. Bổ sung tiếng việt vào dữ liệu so khớp từ điển

Một trong những điểm yếu của zxcvbn là ở chỗ bộ từ điển của nó chỉ bao gồm các từ tiếng Anh, do đó, nó không thể nhận ra các từ hoặc cụm từ phổ biến trong các ngôn ngữ khác. Ở đây, một từ điển tiếng Việt đã qua xử lý đã được bổ sung vào dữ liệu so khớp từ điển của chương trình này.

Đặt giả thiết rằng các hệ thống thường chỉ cho phép người dùng thiết lập mật khẩu không dấu hoặc nếu có cho phép thì cũng rất ít người dùng sử dụng tiếng Việt có dấu cho mật khẩu của mình, do đó, kẻ tấn công khi muốn bẻ khóa một mật khẩu sẽ thử đoán bằng một từ điển bao gồm các từ không dấu trước và các mật khẩu là các từ có dấu được xem là khó đoán hơn hay mạnh hơn. Một danh sách các từ tiếng việt được lấy từ dự án Từ điển tiếng Việt miễn phí (The Free Vietnamese Dictionary Project [15]) sau đó được chuyển đổi thành định dạng các từ tiếng Việt không dấu và viết thường rồi đưa vào dữ liệu từ điển so khớp của chương trình.

Hình 2 mô tả định dạng các từ trong từ điển tiếng Việt nguyên gốc được lấy từ [15] (bên trái) và sau khi được chuyển đổi thành định dạng phù hợp để sử dụng trong cơ sở dữ liệu so khớp của chương trình (bên phải).

```
1  a   QN  P5
2  A Bung  QYJK    x
3  A Di    QYJ7    B7
4  A Di Đà kinh    QYL2    Ep
5  A Di Đà Phật    QYQf    E8
6  A Di Đà Tam Tôn QYVb    D4
7  A Dơi   QYcd    0
8  a dua   gG  Ca
9  A-đam   QYdR    Cy
10 A-đi-xơn    QYgD    FM
11 A Đới   QY1P    6
12 a hoàn  ig  BO
13 a hoàn  PwSP    y
14 A-la    QYnO    Bi
15 A-la-hán    QYow    GA
16 A Lù    QYuw    q
17 A Lưới  QYva    Bp
18 A Mron  QYxD    p
19 A Mú Sung   QYxs    v
20 A Ngo   QYyb    Bf
```

```
1  a
2  abung
3  ac
4  accam
5  acchien
6  ach
7  acmong
8  acnghiet
9  acqui
10 acta
11 actam
12 acthu
13 acvang
14 adam
15 adao
16 adi
17 adidakinh
18 adidaphat
19 adidatamton
20 adixon
```

Hình 2. Định dạng dữ liệu tiếng Việt trước và sau khi được xử lý

V. MỘT SỐ KẾT QUẢ THỬ NGHIỆM PHƯƠNG PHÁP ĐÁNH GIÁ ĐỘ MẠNH MẬT KHẨU ĐƯỢC ĐỀ XUẤT

Để đánh giá hoạt động của phương pháp được đề xuất, một số danh sách mật khẩu dưới đây đã được lựa chọn để thử nghiệm đánh giá độ mạnh mật khẩu.

Danh sách top 10.000 mật khẩu từ nghiên cứu của Mark Burnett được công bố trên trang web xato.net [16]: Danh sách 10.000 mật khẩu phổ biến này được xếp hạng theo số lượng người dùng sử dụng cùng một mật khẩu và đã được chuyển đổi thành tất cả các chữ cái viết thường [17].

Danh sách mật khẩu người dùng của diễn đàn phpBB.com [18]. Năm 2009, diễn đàn này đã bị xâm nhập do một ứng dụng bên thứ ba và một cơ sở dữ liệu chứa mật khẩu băm đã bị rò rỉ. Do nền tảng kỹ thuật của người dùng đăng ký trên trang web này, mật khẩu có xu hướng phức tạp hơn một chút so với các từ điển khác. Thành phần mật khẩu: 41,24% chỉ bao gồm chữ cái viết thường, 35,7% chỉ bao gồm chữ cái viết thường và chữ số, 11,24% chỉ bao gồm chữ số, 4,82% có sử dụng cả chữ cái viết hoa và viết thường và chữ số, 2,68% sử dụng chữ cái hỗn hợp và phần còn lại được tạo thành từ các tổ hợp ký tự khác nhau [2].

Danh sách mật khẩu người dùng của trang web MySpace.com. Trang web này đã bị hack vào ngày 11 tháng 6 năm 2013 bởi công cụ LeakedSource, một công cụ tìm kiếm có khả năng tìm kiếm, tổng hợp dữ liệu từ hàng trăm nguồn khác nhau. Danh sách đầy đủ bao gồm tổng cộng 427.484.128 mật khẩu của gồm 360.000.000 người dùng [19], một danh sách thu nhỏ bao gồm 37.126 mật khẩu được cung cấp trong [20] đã được sử dụng để thử nghiệm công cụ này.

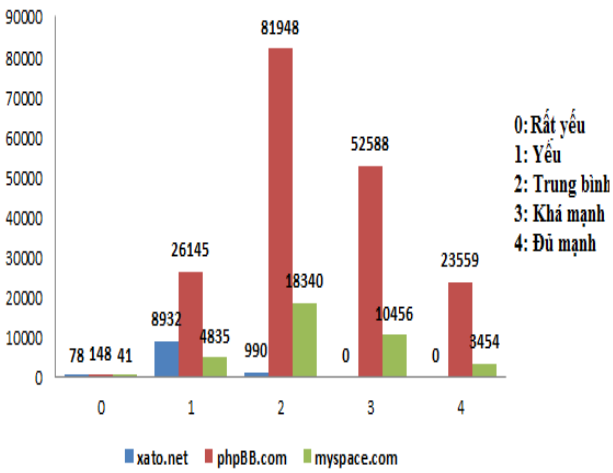
Danh sách mật khẩu người dùng của diễn đàn vnzoom.com. Diễn đàn chia sẻ công nghệ này đã bị tấn công vào ngày 31/5/2012 và để bị rò rỉ ra một cơ sở dữ liệu bao gồm mật khẩu của sáu triệu người dùng trên diễn đàn này. Danh sách sáu triệu mật khẩu này [21] đã được đưa vào để thử nghiệm nhằm mục đích đánh giá dữ liệu từ điển tiếng Việt đã được đưa tích hợp thêm vào dữ liệu so khớp của bộ công cụ đánh giá.

Kết quả thử nghiệm đánh giá các danh sách mật khẩu trên được thể hiện trong Bảng 3.

BẢNG 3. KẾT QUẢ THỬ NGHIỆM CÔNG CỤ ĐÁNH GIÁ ĐỘ MẠNH MẬT KHẨU

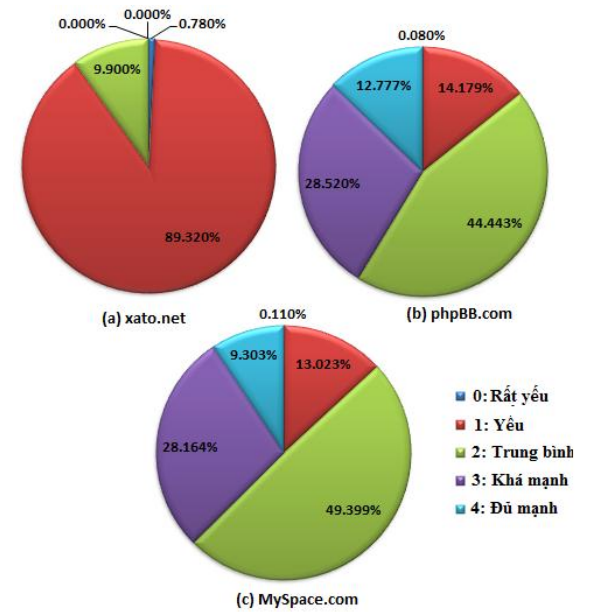
Điểm	xato	phpBB	MySpace	vnzoom
0	78	148	41	220
1	8932	26145	4835	96895
2	990	81948	18340	1117424
3	0	52588	10456	2359467
4	0	23559	3454	2470234
Tổng số	10000	184388	37126	6044240
Điểm TB	1.09	2.40	2.34	3.19

Bảng trên cho thấy, trong các danh sách mật khẩu dựa trên tiếng Anh, xato.net cung cấp danh sách 10.000 mật khẩu phổ biến, thường có mặt trong các từ điển đối chiếu của các công cụ đánh giá độ mạnh, do đó, điểm đánh giá độ mạnh đạt thấp nhất, chỉ khoảng hơn 1 điểm và không có mật khẩu nào đạt được điểm 3 hoặc 4. Trong khi đó, mật khẩu thực sự được người dùng sử dụng (phpBB.com và MySpace.com) có độ mạnh trung bình cao hơn, nằm trong khoảng giữa 2 và 3 điểm, và có cả mật khẩu đủ an toàn (4 điểm).



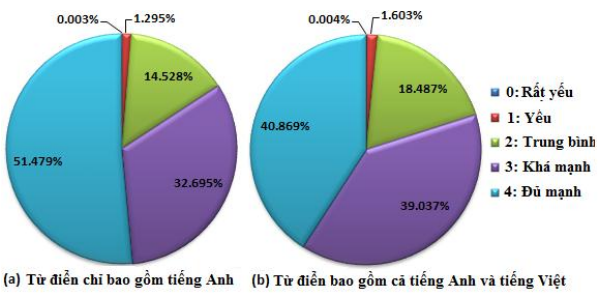
Hình 3. Biểu đồ phân bố độ mạnh mật khẩu của các danh sách mật khẩu dựa trên tiếng Anh

Hình 3 và 4 minh họa cụ thể hơn sự phân bố mật khẩu dựa trên ngôn ngữ tiếng Anh của ba danh sách đầu tiên (xato.net, phpBB.com, MySpace.com) theo thang điểm độ mạnh.



Hình 4. Biểu đồ phân bố độ mạnh mật khẩu của từng danh sách mật khẩu dựa trên tiếng Anh

Những danh sách mật khẩu ở các trang web trên chỉ chứa ngôn ngữ tiếng Anh. Để đánh giá tác động của từ điển tiếng Việt đối với bộ công cụ được đề xuất, chúng tôi đã thực hiện đánh giá mật khẩu của vnzoom.com trong hai trường hợp: từ điển chỉ bao gồm tiếng Anh và từ điển bao gồm cả tiếng Anh và tiếng Việt. Kết quả thử nghiệm hai trường hợp này lần lượt được minh họa trong hình 5(a) và 5(b). Kết quả này cho thấy rằng khi có thêm từ điển tiếng Việt vào dữ liệu so khớp, số lượng mật khẩu được đánh giá là đủ mạnh giảm hơn 10% (từ 51,479% xuống còn 40,869%). Điều này chứng tỏ rằng việc đánh giá độ mạnh mật khẩu phụ thuộc nhiều vào ngôn ngữ mà người sử dụng lựa chọn để thiết lập mật khẩu, và việc bổ sung từ điển tiếng Việt đã giúp đánh giá độ mạnh mật khẩu của người dùng sử dụng ngôn ngữ tiếng Việt một cách chính xác hơn.



Hình 5. Biểu đồ phân bố độ mạnh mật khẩu của vnzoom.com

Hình 6 minh họa kết quả tích hợp mô-đun đánh giá độ mạnh mật khẩu vào phần mềm.

Hình 6. Tích hợp mô-đun đánh giá mật khẩu

Các kết quả đánh giá độ mạnh mật khẩu có thể giúp các nhà quản trị hệ thống đưa ra chính sách tạo mật khẩu phù hợp, ví dụ như chỉ cho phép các mật khẩu có độ mạnh đạt một giá trị tối thiểu nhất định. Mô-đun đánh giá độ mạnh mật khẩu được trình bày trong bài báo này cũng đã được thử nghiệm cài đặt vào một số hệ thống và cho thấy được tính hiệu quả.

VI. KẾT LUẬN

Bài báo này đã phân tích và đề xuất phương pháp đánh giá độ mạnh mật khẩu dựa trên entropy và thử nghiệm đánh giá hoạt động của phương pháp đánh giá độ mạnh mật khẩu này. Kết quả thử nghiệm sử dụng phương pháp được đề xuất để đánh giá độ an toàn mật khẩu của người dùng một số hệ thống trên thực tế cho thấy rằng ngay cả các hệ thống mà người dùng có trình độ công nghệ thông tin thì cũng chưa đến một nửa số người dùng lựa chọn được một mật khẩu đủ an toàn. Vì vậy, vấn đề về độ an toàn của mật khẩu cần phải đặt ra ngay cả đối với những người dùng am hiểu về công nghệ.

Ngoài ra, kết quả thử nghiệm cài đặt mô-đun đánh giá độ mạnh mật khẩu trong một số hệ thống cũng cho thấy khả năng công cụ này có thể được triển khai một cách đơn giản và hiệu quả trong các hệ thống xác thực người dùng dựa trên mật khẩu.

TÀI LIỆU THAM KHẢO

[1]. https://en.wikipedia.org/wiki/Password
[2]. Xavier De Carne De Carnavalet, Mohammad Mannan (2015) "A Large-Scale Evaluation of High-Impact Password Strength Meters".
[3]. https://en.wikipedia.org/wiki/Password_strength
[4]. NIST Special Publication 800-63: Electronic Authentication Guideline, 2004.
[5]. Nouf Mohammed D. Aljaffan, "Password Security and Usability: From Password Checkers To a New Framework For User Authentication", 2017.
[6]. A. Narayanan and V. Shmatikov, "Fast dictionary attacks on passwords using time-space tradeoff", 2005.
[7]. M. Weir, S. Aggarwal, B. de Medeiros, and B. Glodek, "Password cracking using probabilistic context-free grammars", 2009.
[8]. W. Melicher, B. Ur, S. M. Segreti, S. Komanduri, L. Bauer, N. Christin, and L. F. Cranor, "Fast, lean and accurate: Modeling password guessability using neural networks", 2016.
[9]. C. Castelluccia, A. Chaabane, M. Durmuth, and D. Perito, "When privacy meets security: Leveraging personal information for password cracking", 2013.
[10]. Y. Li, H. Wang, and K. Sun, "A study of personal information in human-chosen passwords and its security implications", 2016.
[11]. Daniel Lowe Wheeler (2016), "zxcvbn: Low-Budget Password Strength Estimation".
[12]. S. Ji, S. Yang, T. Wang, C. Liu, W.H. Lee, and R. Beyah, "PARS: A uniform and open-source password analysis and research system", 2015.
[13].https://blogs.dropbox.com/tech/2012/04/zxcvbn-realistic-password-strength-estimation/
[14]. https://tools.ietf.org/html/rfc4086
[15]. https://www.informatik.unileipzig.de/~duc/Dict
[16].https://github.com/danielmiessler/SecLists/blob/master/Passwords/xato-net-10-million-passwords-100000.txt
[17].https://xato.net/10-000-top-passwords 6d6380716fe0
[18].https://raw.githubusercontent.com/danielmiessler/SecLists/master/Passwords/Leaked-Databases/phpbb.txt
[19]. https://leakedsource.ru/blog/myspace
[20].https://raw.githubusercontent.com/danielmiessler/SecLists/master/Passwords/Leaked-Databases/myspace.txt
[21].http://www.mediafire.com/file/tttc26nlemi8ntb/6tr-user-vn-zoom.rar

SƠ LƯỢC VỀ TÁC GIẢ

CN. Hoàng Thu Phương

Đơn vị công tác: Viện KH-CN mật mã, Ban Cơ yếu Chính phủ.

Email: thuphuonghoang306@gmail.com

Quá trình đào tạo: Nhận bằng kỹ sư tại Đại học Thượng Hải 2012.

Hướng nghiên cứu hiện nay: Bảo mật trên thiết bị di động

ThS. Trần Sỹ Nam

Đơn vị công tác: Viện KH-CN mật mã, Ban Cơ yếu Chính phủ.

Email: transynam1989@gmail.com

Quá trình đào tạo: Nhận bằng kỹ sư chuyên ngành An toàn thông tin và Hệ thống mạng tại Học viện FSO, Nga và nhận bằng thạc sĩ chuyên ngành Kỹ thuật mật mã tại Học viện Kỹ thuật mật mã 2018

Hướng nghiên cứu hiện nay: Bảo mật mạng, dữ liệu lưu trữ