BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Nguyễn Trọng Hưng

NGHIÊN CỨU CÁC GIẢI PHÁP PHÁT HIỆN TẤN CÔNG WEB SỬ DỤNG WEB LOG VÀ NỘI DUNG KẾT HỢP ẢNH MÀN HÌNH TRANG WEB

TÓM TẮT LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN
Mã số: 9 48 01 04

Hà Nội - 2024

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ,
Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học:
Người hướng dẫn 1: PGS.TS. Hoàng Xuân Dậu, Học viện Công nghệ và BCVT
Người hướng dẫn 2: PGS.TS. Nguyễn Đức Dũng, Viện Công nghệ thông tin

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

Luận án được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện họp tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi giờ , ngày tháng năm 2024.

Có thể tìm hiểu luận án tại:
1. Thư viện Học viện Khoa học và Công nghệ
2. Thư viện Quốc gia Việt Nam

DANH MỤC CÁC BÀI BÁO ĐÃ XUẤT BẢN
LIÊN QUAN ĐẾN LUẬN ÁN

1. Hoang Xuan Dau, Ninh Thi Thu Trang, **Nguyen Trong Hung**,"*A Survey of Tools and Techniques for Web Attack Detection"*. Journal of Science and Technology on Information security, Special Issue CS (15) 2022, pp. 109-118.
2. Xuan Dau Hoang, **Trong Hung Nguyen**, "*Detecting common web attacks based on supervised machine learning using web logs*", Journal of Theoretical and Applied Information Technology Vol.99. No 6, 31st March 2021, Scopus Q4.
3. **Trong Hung Nguyen**, Xuan Dau Hoang, Duc Dung Nguyen, *"Detecting Website Defacement Attacks using Web-page Text and Image Features"*, Article Published in International Journal of Advanced Computer Science and Applications(IJACSA), Volume 12 Issue 7, 2021, Scopus Q3.
4. Hoang Xuan Dau, **Nguyen Trong Hung**, *"Phát hiện tấn công web thường gặp dựa trên học máy sử dụng web log",* Hội nghị khoa học quốc gia về "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin" FAIR 2020.8.
5. **Trong Hung Nguyen**, Dau Hoang, Nguyen Duc Dung, Vu Xuan Hanh. *"Phát hiện tấn công thay đổi giao diện trang web sử dụng đặc trưng văn bản"*, Hội nghị KHCN Quốc gia lần thứ XVII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin(FAIR), Hà Nội, 8/2024.
6. Xuan Dau Hoang, **Trong Hung Nguyen**, Hoang Duy Pham, *"A Novel Model for Detecting Web Defacement Attacks Using Plain Text Features"* Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 2024, Scopus Q3 *(Đã nhận được thư chấp nhận đăng)*.

MỞ ĐẦU

1. Tính cấp thiết của luận án

Do tính chất nguy hiểm của tấn công web đối với các cơ quan, tổ chức và cá nhân, nhiều giải pháp đã được nghiên cứu, phát triển và triển khai để phát hiện, phòng chống tấn công web, như sử dụng tường lửa web (WAF), hệ thống phát hiện xâm nhập web (Web IDS), kiểm thử xâm nhập [5] [6] [7]. Nói chung, hiện nay có hai hướng tiếp cận chính trong phát hiện tấn công web: (1) phát hiện dựa trên dấu hiệu, chữ ký và (2) phát hiện dựa trên bất thường [7] [8] [9].

Theo hướng tiếp cận (2), luận án nghiên cứu về việc sử dụng kỹ thuật phát hiện tấn công web dựa trên bất thường, Cụ thể hơn, luận án tập trung nghiên cứu theo hai hướng chính: (i) phát hiện các dạng tấn công web cơ bản, bao gồm *SQLi, XSS, duyệt đường dẫn, CMDi* và (ii) là phát hiện tấn công thay đổi giao diện trang web. Theo hướng (i), qua khảo sát chưa có nhiều công trình sử dụng bộ dữ liệu từ web log và các nghiên cứu này thường chỉ thực hiện phát hiện được một hình thức tấn công trên một tập dữ liệu thử nghiệm cụ thể. *Do đó, luận án này tiếp tục nghiên cứu phát hiện đồng thời các dạng tấn công web thường gặp, bao gồm SQLi, XSS, duyệt đường dẫn, CMDi dựa trên dữ liệu web log sử dụng các mô hình học máy có giám sát.* Theo hướng (ii), qua khảo sát, đánh giá hầu hết các nghiên cứu đã có chỉ tập trung sử dụng một loại đặc trưng liên quan đến nội dung trang web mà chưa có sự kết hợp các loại đặc trưng điển hình, gồm nội dung và hình ảnh của của trang web bị tấn công thay đổi giao diện. *Do vậy, luận án tập trung nghiên cứu phương pháp phát hiện tấn công thay đổi giao diện trang web sử dụng các thuật toán học sâu và kết hợp các đặc trưng văn bản/nội dung và hình thức thể hiện - là ảnh chụp màn hình trang web để cải thiện độ chính xác, tốc độ và thời gian tính toán.*

2. Mục tiêu nghiên cứu của luận án

- Nghiên cứu, đánh giá, các phương pháp, kỹ thuật, giải pháp, công cụ phát hiện tấn công web.

- Nghiên cứu đề xuất mô hình phát hiện các dạng tấn công web thường gặp dựa trên kỹ thuật học máy có giám sát sử dụng dữ liệu web log, nhằm nâng cao độ chính xác, giảm cảnh báo sai, đồng thời cho phép phát hiện nhiều loại tấn công web.

- Nghiên cứu đề xuất mô hình phát hiện tấn công thay đổi giao diện trang web dựa trên kỹ thuật học sâu và kết hợp hai loại đặc trưng văn bản và hình ảnh của trang web, nhằm nâng cao độ chính xác, giảm cảnh báo sai.

- Cài đặt, thử nghiệm và đánh giá các mô hình phát hiện tấn công web đã đề xuất sử dụng các tập dữ liệu đã được công bố và tập dữ liệu thu thập thực tế.

3. Các nội dung nghiên cứu chính của luận án

Chương 1. Tổng quan về phát hiện tấn công web giới thiệu khái quát về web và dịch vụ web, các lỗ hổng bảo mật web theo OWASP, các dạng tấn công web thường gặp, một số giải pháp và công cụ phát hiện tấn công web. Tiếp theo, chương này giới thiệu khái quát về học máy, học sâu và mô tả một số giải thuật học máy có giám sát và học sâu sử dụng trong các mô hình phát hiện tấn công web được đề xuất trong chương 2 và chương 3. Phần cuối của chương chỉ ra hai vấn đề sẽ được giải quyết trong luận án.

Chương 2. Phát hiện tấn công web dựa trên học máy sử dụng web log giới thiệu khái quát về web log, một số đề xuất phát hiện tấn công web sử dụng học máy, đánh giá ưu nhược điểm của các đề xuất. Phần cuối của chương này thực hiện việc xây dựng, cài đặt, thử nghiệm và đánh giá mô hình phát hiện tấn công web thường gặp dựa trên học máy sử dụng web log.

Chương 3. Phát hiện tấn công thay đổi giao diện trang web giới thiệu khái quát về tấn công thay đổi giao diện, các phương pháp phát hiện tấn công thay đổi giao diện, so sánh các phương pháp phát hiện thay đổi giao diện sử dụng đặc trưng ảnh chụp màn hình trang web. Phần cuối của chương thực hiện việc xây dựng, cài đặt, thử nghiệm và đánh giá mô hình phát hiện tấn công thay đổi giao diện trang web dựa trên học sâu sử dụng kết hợp đặc trưng ảnh chụp màn hình và đặc trưng nội dung văn bản của trang web.

CHƯƠNG 1: TỔNG QUAN VỀ PHÁT HIỆN TẤN CÔNG WEB

1.1. Khái quát về web và dịch vụ web

Dịch vụ web (Web service): Tổ chức World Wide Web Consortium (W3C) định nghĩa Dịch vụ web là hệ thống phần mềm cho phép các máy khác nhau tương tác với nhau thông qua mạng. Các dịch vụ web đạt được nhiệm vụ này với sự trợ giúp của các tiêu chuẩn mở, bao gồm XML, SOAP, WSDL và UDDI [29]. *Ứng dụng web (Web application)* là một phần mềm ứng dụng chạy trên nền web [30]. Ứng dụng web cũng được vận hành dựa trên giao thức HTTP theo mô hình khách chủ (Client/Sever). *Website* là tập hợp của các trang web được cài đặt và chạy (host) trên máy chủ web. Trang web (Web page) là một phần của một website cung cấp một đầu mục nội dung hay một tính năng cụ thể của website. Ngôn ngữ thường dùng để tạo các trang web là HTML.

1.2. Tổng quan về tấn công web

Tấn công web, hay tấn công ứng dụng web là việc lợi dụng những điểm yếu, lỗ hổng tồn tại trên hệ thống website, ứng dụng web để thực hiện các hành vi khai thác, đánh cắp dữ liệu nhạy cảm tồn tại trên hệ thống [32]. Cũng theo [32], gần đây có tới 75% cuộc tấn công mạng được thực hiện ở cấp độ ứng dụng web.

Có thể kể đến các dạng tấn công, xâm nhập phổ biến vào các website, ứng dụng web (gọi tắt là tấn công web), bao gồm tấn công chèn mã SQL (SQLi – SQL injection), tấn công XSS (Cross-Site Scripting), tấn công CSRF (Cross-site Request Forgery), tấn công chèn dòng lệnh (CMDi – Command injection), tấn công duyệt đường dẫn, tấn công DoS/DDoS và tấn công thay đổi giao diện [33] [31] [35].

1.3. Phát hiện tấn công web

Nói chung, có 3 hướng tiếp cận phòng thủ đối với các cuộc tấn công này, bao gồm (1) kiểm tra, xác thực tất cả dữ liệu đầu vào, (2) giảm các bề mặt tấn công và (3) sử dụng chiến lược "phòng thủ theo chiều sâu" [33] [48] [49]. Cụ thể, hướng tiếp cận (1) yêu cầu tất cả dữ liệu đầu vào cho các ứng dụng web phải được kiểm tra kỹ lưỡng sử dụng các bộ lọc dữ liệu đầu vào và chỉ những đầu vào hợp pháp mới được chuyển sang các bước tiếp theo để xử lý. Mặt khác, hướng tiếp cận (2) yêu cầu chia ứng dụng web thành nhiều phần và sau đó áp dụng các biện pháp điều khiển truy cập phù hợp để hạn chế quyền truy cập của người dùng. Đối với hướng tiếp cận (3), một số biện pháp phòng thủ được triển khai trong các lớp kế tiếp nhau để bảo vệ các trang web, ứng dụng web và người dùng web.

Các giải pháp và công cụ phát hiện tấn công web: Có nhiều giải pháp, công cụ phát hiện tấn công web được phát triển và triển khai ứng dụng trên thực tế, như [50][51][52][53][54][55][56]. Các kỹ thuật phát hiện tấn công web: Có nhiều kỹ thuật phát hiện tấn công web được đề xuất và ứng dụng trong những năm qua. Tuy nhiên, có 2 nhóm kỹ thuật phát hiện tấn công web sử dụng phổ biến, bao gồm (1) phát hiện dựa trên chữ ký, mẫu hoặc tập luật [59] và (2) phát hiện dựa trên bất thường [60].

1.4. Hướng nghiên cứu của luận án

Hướng nghiên cứu của luận án là phát hiện tấn công web thường gặp và tấn công thay đổi giao diện web dựa trên bất thường do phương pháp này có khả năng phát hiện các dạng tấn công web mới, đồng thời có khả năng tự động hóa việc xây dựng mô hình phát hiện. Trên cơ sở khảo sát, phân tích các ưu điểm và hạn chế của các đề xuất đã có, luận án tập trung nghiên cứu, giải quyết các vấn đề sau: (1) Đề xuất mô hình phát hiện tấn công web thường gặp dựa trên học máy sử dụng web log và (2) Đề xuất mô hình phát hiện tấn công thay đổi giao diện trang web dựa trên học sâu sử dụng kết hợp dữ liệu văn bản nội dung trang web và ảnh chụp màn hình trang web. Lý do thực hiện (1) là do một số kỹ thuật phát hiện dựa trên bất thường chỉ phát hiện được một loại tấn công trên một tập dữ liệu cụ thể, mà không phát hiện được đồng thời nhiều loại tấn công web, như: XSS, SQLi, duyệt đường dẫn, CMDi. Ngoài ra, một số đề xuất phát hiện dựa trên bất thường có tỷ lệ phát hiện đúng còn thấp và tỷ lệ cảnh báo sai còn cao. Tương tự, việc thực hiện (2) nhằm nâng cao tỷ lệ phát hiện đúng và giảm tỷ lệ cảnh báo sai cho mô hình phát hiện tấn công thay đổi giao diện sử dụng dữ liệu đầu vào kết hợp giữa dữ liệu văn bản nội dung trang web và ảnh chụp màn hình trang web.

CHƯƠNG 2: PHÁT HIỆN TẤN CÔNG WEB DỰA TRÊN HỌC MÁY SỬ DỤNG WEB LOG

2.1. Phát hiện tấn công web dựa trên học máy

Kết quả nghiên cứu và khảo sát nhận thấy, các giải pháp đề xuất phát hiện tấn công web dựa trên dữ liệu web log là một hướng hiệu quả. Đặc biệt, hướng nghiên cứu sử dụng sử dụng học máy là nhánh có triển vọng do mô hình phát hiện đơn giản, có thể được xây dựng tự động từ tập dữ liệu huấn luyện. Đây cũng chính là nhánh nghiên cứu của luận án chọn thực hiện.

Một số vấn đề cần tiếp tục nghiên cứu như: (1) một số đề xuất tuy sử dụng cơ chế đơn giản, nhưng chỉ cho độ chính xác phát hiện cao với tập dữ liệu cụ thể hoặc với một loại tấn công web cụ thể, và số lượng đặc trưng quá ít hoặc quá nhiều điển hình như các nghiên cứu của Sharma và cộng sự [20], Saleem và cộng sự [21]; (2) một số đề xuất sử dụng mô hình học sâu hoặc sử dụng bộ công cụ giám sát máy chủ nên đòi hỏi chi phí tính toán lớn cho quá trình xây dựng mô hình, cũng như quá trình giám sát phát hiện và điều này làm giảm khả năng triển khai ứng dụng trên các hệ thống thực[15][16]; và (3) một số đề xuất sử dụng mô hình học sâu, đòi hỏi nhiều tài nguyên tính toán, nhưng không phát hiện được nhiều hình thức tấn công web (SQLi, XSS, CMDi, duyệt đường dẫn), như [7][15].

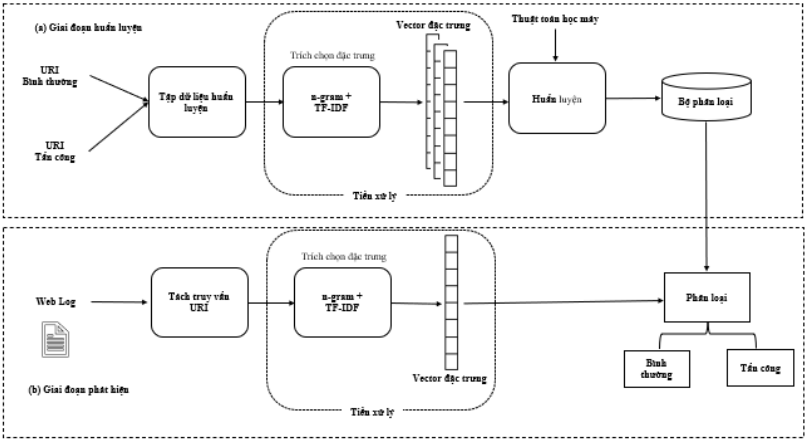
2.2. Xây dựng và thử nghiệm mô hình phát hiện tấn công web dựa trên học máy sử dụng web log

2.2.1. Mô tả mô hình phát hiện

2.2.1.1. Giới thiệu mô hình

Mô hình phát hiện tấn công web đề xuất được triển khai trong 2 giai đoạn: (a) giai đoạn huấn luyện và (b) giai đoạn phát hiện. Giai đoạn huấn luyện như biểu diễn trên Hình 2. 4.

Trong giai đoạn huấn luyện dữ liệu URI tấn công và bình thường được thu thập, tiếp theo sẽ tiến hành tiền xử lý dữ liệu nhằm trích xuất các đặc trưng cho quá trình huấn luyện. Trong bước huấn luyện, các thuật toán học máy có giám sát, như Naïve bayes, SVM, Cây quyết định, Rừng ngẫu nhiên được áp dụng để học ra bộ phân loại, thuật toán cho kết quả tốt nhất sẽ được sử dụng cho mô hình phát hiện. Trong giai đoạn Phát hiện, các truy vấn URI sẽ được trích lọc từ dữ liệu weblog, qua quá trình tiền xử lý như giai đoạn Huấn luyện và đến bước phân loại sử dụng Bộ phân loại từ giai đoạn Huấn luyện để xác định truy vấn Bình thường hay Tấn công.



Hình 2.4. Mô hình phát hiện tấn công web dựa trên dữ liệu weblog

2.2.1.2. Tiền xử lý dữ liệu, huấn luyện và phát hiện

Quá trình tiền xử lý dữ liệu web log dựa trên kỹ thuật n-gram, TF-IDF và giảm chiều được thực hiện theo các bước như sau:

Bước 1: Tách các truy vấn **?query_string** trong các truy vấn URI

Bước 2: Từ các truy vấn này thực hiện tách các đặc trưng n-gram

Bước 3: Tính giá trị cho các đặc trưng n-gram sử dụng phương pháp TF-IDF [84].

Bước 4: Giảm chiều dữ liệu sử dụng phương pháp hệ số tương quan, phương pháp Information Gain, hoặc phương pháp PCA.

Các thuật toán học máy được sử dụng bao gồm: naive bayes, SVM, cây quyết định và rừng ngẫu nhiên. Đối với mỗi thuật toán, lấy ngẫu nhiên 80% dữ liệu dùng cho quá trình huấn luyện để xây dựng mô hình phát hiện, sau đó sử dụng 20% dữ liệu để kiểm thử cho kết quả của các độ đo đánh giá.

2.2.2. Tập dữ liệu thử nghiệm

HTTP Param Dataset [91] có các truy vấn bình thường được lọc từ bộ dữ liệu HTTP CISC 2010 [69] và các truy vấn tấn công SQLi, XSS, CMDi, duyệt đường dẫn được thực hiện từ các môi

trường tấn công SQLmap, XSSya, Vega Scaner, FuzzDB repository. Bộ dữ liệu này gồm 31.067 chuỗi truy vấn **?query_string** trong URI của các yêu cầu web, bao gồm độ dài và nhãn của truy vấn. Có 2 loại nhãn truy vấn là Norm (Bình thường) và Anom (Tấn công). Nhãn Anom lại gồm 4 loại tấn công cụ thể: SQLi, XSS, CMDi và duyệt đường dẫn.

2.2.3. Thử nghiệm và kết quả

2.2.3.1. Kịch bản thử nghiệm

Kịch bản 1: Đánh giá ảnh hưởng của các tham số 2-gram, 3-gram, 4-gram, 5-gram trên mô hình đề xuất với thuật toán học máy Rừng ngẫu nhiên từ đó lựa chọn tham số n-gram cho kết quả tốt nhất. trong kịch bản này luận án giữ nguyên tập đặc trưng và không sử dụng phương pháp giảm chiều dữ liệu.

Kịch bản 2: Đánh giá ảnh hưởng của ba phương pháp giảm chiều dữ liệu là PCA, Information Gain, Hệ số tương quan lên tập đặc trưng thu được từ Kịch bản 1 *(thuật toán Random Forest sử dụng với n-gram cho kết quả tốt nhất)*. Từ đó lựa chọn được phương pháp giảm chiều dữ liệu cho kết quả tốt nhất.

Kịch bản 3: Đánh giá kết quả của mô hình huấn luyện sử dụng các thuật toán học máy có giám sát Navie Bayes và SVM, Cây quyết định, Rừng ngẫu nhiên (10, 30, 50, 60 cây) với 3-gram và phương pháp giảm chiều dữ liệu PCA từ kết quả Kịch bản 1 và Kịch bản 2, từ đó lựa chọn thuật toán cho kết quả tốt nhất, sẽ được sử dụng cho quá trình phát hiện.

Kịch bản 4: Đánh giá mô hình đề xuất với thuật toán học máy có giám sát cho kết quả tốt nhất từ Kịch bản 3 với các nghiên cứu liên quan.

2.2.3.2. Kết quả thử nghiệm

Bảng 2. 4. Kết quả đánh giá Kịch bản 1

Thuật toán	n-gram	PPV	TPR	FPR	FNR	ACC	F1	Time(s)

Thuật toán	n-gram	PPV	TPR	FPR	FNR	ACC	F1	Time(s)
Rừng ngẫu nhiên	2-gram	98,94	99,32	0,64	0,68	99,34	99,13	17,90
	3-gram	100	99,14	0	0,86	99,68	99,57	92,99
	4-gram	99,91	99,1	0,05	0,9	99,63	99,51	132,56
	5-gram	100	98,80	0	1,20	99,55	99,40	135,23

Kết quả từ Bảng 2.4 cho thấy với thuật toán Rừng ngẫu nhiên khi sử dụng đặc trưng 3-gram cho độ chính xác chung ACC và độ đo F1 cao nhất so với khi sử dụng các đặc trưng 2-gram, 4-gram và 5-gram

Bảng 2. 5. Kết quả đánh giá Kịch bản 2

Thuật toán	PP Giảm chiều	PPV	TPR	FPR	FNR	ACC	F1
Rừng ngẫu nhiên	PCA	98,97	98,72	0,62	1,28	99,13	98,84
	Information Gain	99,28	94,53	0,41	5,47	97,68	96,85
	Hệ số tương quan	99,59	92,77	0,23	7,23	97,14	96,06

Kết quả từ Bảng 2.5 cho thấy sau khi thực hiện giảm chiều dữ liệu, phương pháp giảm chiều với PCA cho kết quả ACC, F1, Recall là cao nhất so với phương pháp giảm chiều với Information Gain và Hệ số tương quan.

Bảng 2.6. Kết quả kịch bản 3

Thuật toán	PPV	TPR	FPR	FNR	ACC	F1
NavieBayes	89,48	96,41	6,84	3,59	94,38	92,82
SVM	99,87	98,50	0,08	1,50	99,09	99,18
Cây quyết định	96,48	98,42	2,17	1,58	98,05	97,44
Rừng ngẫu nhiên - 10	98,13	98,85	1,14	1,15	98,86	98,49
Rừng ngẫu nhiên - 30	98,68	98,80	0,80	1,20	99,05	98,80
Rừng ngẫu nhiên - 50	**98,97**	**98,72**	**0,62**	**1,28**	**99,13**	**98,84**
Rừng ngẫu nhiên - 60	98,80	98,76	0,72	1,24	99,08	98,78

Kết quả tại Bảng 2.6 cho thấy khi sử dụng thuật toán Rừng ngẫu nhiên (50 cây) với đặc trưng 3-gram kết hợp phương pháp giảm chiều dữ liệu PCA cho kết quả về độ đo ACC và F1 tốt nhất, thuật toán NavieBayes cho kết quả thấp nhất.

Bảng 2.7. Kết quả kịch bản 4

Thuật toán	PPV	TPR	FPR	FNR	ACC	F1	Thời Gian huấn luyện	Thời gian phát hiện
Đề xuất - Rừng ngẫu nhiên (50 cây)	**98,97**	**98,72**	**0,62**	**1,28**	**99,13**	**98,84**	**27,52**	**1.49**
Liang và cộng sự[11]	99,04	96,88	1,13	3,12	97,78	97,95	1177,20	5,67
Ming Zhang và cộng sự [45]	98,59	93,35	1,37	6,65	96,49	95,92	151,00	4,18
Saiyu Hao cùng cộng sự [7]	98,77	93,71	0,62	6,29	97,41	96.17	13063,56	15,05
Pan và cộng sự [12]	90,60	92,80				91,80		
S. Sharma và cộng sự[16]	99,60	91,52	0,20	8,48	96,91	95,39		

Bảng 2.7 cho thấy thuật toán Rừng ngẫu nhiên (50 cây) dùng cho mô hình đề xuất cho kết quả với các độ đo ACC, F1, Recall tốt hơn các đề xuất [15][16][49][7][20].

Bảng 2.8. Tỷ lệ phát hiện (DR) cho các cuộc tấn công web trên thuật toán học máy

Algorithms	SQLi(%)	XSS(%)	CMDi(%)	Path(%)	Average (%)
Rừng ngẫu nhiên	99,90	98,68	82,02	98,62	99,67

Bảng 2.8 thể hiện kết quả của mô hình đề xuất trong phát hiện từng loại tấn công cụ thể: SQLi, XSS, CMDi, duyệt đường dẫn và tỷ lệ phát hiện trung bình Có thể thấy, mô hình cho tỷ lệ phát hiện tấn công SQLi cao nhất và tỷ lệ phát hiện tấn công CMDi là thấp nhất.

2.3. Nhận xét

Luận án sẽ đánh giá về hiệu suất phát hiện của mô hình đề xuất dựa trên các khía cạnh sau: (1) ảnh hưởng của việc phân bố số lượng tấn công web đến tỷ lệ phát hiện, (2) hiệu suất phát hiện của mô hình dựa trên các thuật toán học máy khác nhau và (3) so sánh giữa mô hình đề xuất với các đề xuất trước đó. Số lượng các loại tấn công web cụ thể trong tập dữ liệu phân bố không cân bằng do đó ảnh hưởng tới hiệu suất phát hiện với từng loại tấn công web. Điều này

dẫn đến tỷ lệ phát hiện cho tấn công SQLi là cao nhất và tỷ lệ phát hiện cho tấn công CMDi là thấp nhất. Về hiệu suất phát hiện, mô hình đề xuất cho độ chính xác ACC, F1, Recall cao nhất so [15][16][20][49][7] đồng thời thời gian huấn luyện của mô hình đề xuất cũng nhanh hơn nhiều so với các nghiên cứu [15][20][7].

CHƯƠNG 3: PHÁT HIỆN TẤN CÔNG THAY ĐỔI GIAO DIỆN TRANG WEB

3.1. Tấn công thay đổi giao diện, các phương pháp phát hiện tấn công thay đổi giao diện

Tấn công thay đổi giao diện trang web là việc khai thác các lỗ hổng trên trang web hoặc máy chủ web nhằm thực hiện mã khai thác làm thay đổi giao diện hoặc xóa, thay đổi nội dung của trang web thông qua văn bản, hình ảnh hoặc cả hai [92].

Từ các nghiên cứu đánh giá được khảo sát tại mục *1.3.3.2.* của luận án có thể thấy các giải pháp đã đề xuất cho phát hiện tấn công thay đổi giao diện trang web đang tồn tại những vấn đề sau:

- Các giải pháp như kiểm tra checksum, so sánh DIFF và phân tích cây DOM chỉ có thể hoạt động tốt với các trang web tĩnh.

- Một số đề xuất yêu cầu sử dụng nhiều tài nguyên tính toán do chúng sử dụng các mô hình phát hiện có độ phức tạp cao, điển hình như trong [70][28]

- Một số đề xuất khác có mức độ cảnh báo sai cao, trong khi hiệu suất phát hiện lại phụ thuộc vào việc lựa chọn các ngưỡng phát hiện, điển hình là công trình nghiên cứu [27]

- Nhiều đề xuất chỉ có thể xử lý nội dung văn bản của các trang web. Các thành phần trang web quan trọng khác, như mã JavaScript, CSS, các tệp ảnh nhúng không được xử lý hoặc chỉ xử lý bằng kỹ thuật đơn giản, chẳng hạn như kiểm tra tính toàn vẹn dựa trên hàm băm, điển hình là các nghiên cứu [18][19][17]

- Các nghiên cứu [17][18][19][28] sử dụng tập dữ liệu nhỏ hoặc rất nhỏ với khoảng 300 đến hơn 1000 dữ liệu trang web bị tấn công và bình thường. Tập dữ liệu thử nghiệm nhỏ ảnh hưởng đến độ tin cậy của kết quả phát hiện.

- Hầu hết các nghiên cứu chỉ tập trung nghiên cứu các đặc trưng tệp tin HTML và kiểm tra mã hàm băm của các ảnh nhúng trong nội dung trang web, chưa có nghiên cứu nào tập trung vào nội dung văn bản thuần trong tệp HTML kết hợp sử dụng ảnh chụp màn hình của trang web.

Theo nghiên cứu [96] việc kết hợp dữ liệu các cuộc tấn công và dữ liệu văn bản, dữ liệu hình ảnh là nguồn dữ liệu phong phú cho phát hiện rộng quy mô hoạt động tấn công thay đổi giao diện. Mặt khác, theo Mao và cộng sự [97] và thống kê các cuộc tấn công thay đổi giao diện trang web từ [98] cho thấy sau khi bị thay đổi giao diện thì trang web có giao diện mới chỉ thuần một giải màu *(màu đen – trắng, đỏ - đen,..)* hoặc chứa tin nhắn, hình ảnh nhúng hoặc logo và video không liên quan tới tiêu đề của trang web. Do đó, luận án đề xuất mô hình dựa trên việc kết hợp hai đặc trưng hình ảnh chụp màn hình trang web và nội dung văn bản của trang web.

3.2. Thu thập dữ liệu thử nghiệm

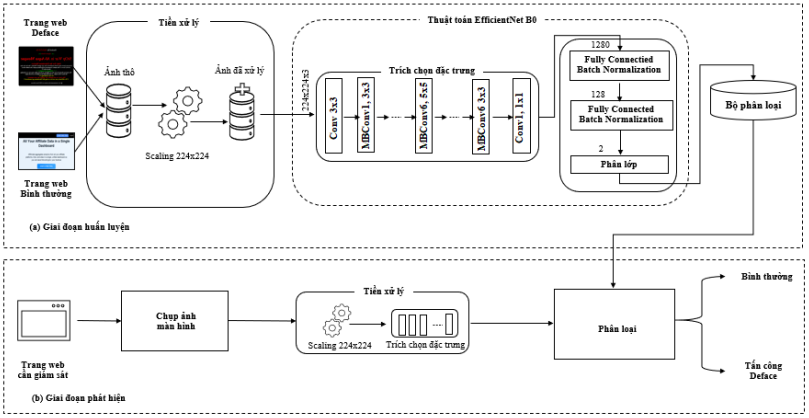
Các nghiên cứu [28] [18] [19] [17] [71] đều sử dụng tập dữ liệu tương đối nhỏ từ 100 đến khoảng 4000 mẫu trang web bình thường và bị tấn công. Các bộ dữ liệu nhỏ như vậy không thực sự phù hợp với các thuật toán học sâu được sử dụng trong các mô hình huấn luyện và phát hiện. Đồng thời, các nghiên cứu này thu thập tập dữ liệu dán nhãn tấn công từ các nguồn mở, cơ sở lưu trữ các trang web bị tấn công thay đổi giao diện, như zone-h.org, zone-xsec.com và đều không công bố tập dữ liệu sử dụng. Do đó, luận án sẽ tiến hành thu thập dữ liệu tấn công với nhãn “Defaced” từ nguồn zone-

h.org - đây là một kho lưu trữ dữ liệu về tấn công thay đổi giao diện trang web được thành lập từ năm 2002 với nhiều thông số được lưu trữ, trong đó có hình ảnh chụp màn hình trang web tại thời điểm bị tấn công thay đổi giao diện. Dữ liệu các trang web bình thường dán nhãn "Normal" được thu thập từ tập 1 triệu trang web do Alexa cung cấp [99].

3.3. Phát hiện thay đổi giao diện sử dụng ảnh chụp màn hình trang web

3.3.1. Mô tả mô hình phát hiện

Mô hình sử dụng dữ liệu đầu vào là ảnh chụp màn hình của trang web bình thường và trang web bị tấn công thay đổi toàn bộ giao diện. Mô hình được xây dựng dựa trên cơ sở phân tích đặc điểm nhận dạng của tấn công thay đổi giao diện trang web, khi trang web bị tấn công thay đổi giao diện thì toàn bộ nội dung trang web bị thay đổi và giao diện trang web cũng bị thay đổi. Mô hình đề xuất được triển khai trong 2 giai đoạn: (1) giai đoạn huấn luyện và (2) giai đoạn phát hiện.



Hình 3.9. Mô hình phát hiện tấn công thay đổi giao diện trang web sử dụng ảnh chụp màn hình trang web

3.3.2. Tiền xử lý dữ liệu và huấn luyện

Từ dữ liệu thu thập được của các trang web bình thường và trang web bị tấn công thay đổi giao diện, chương trình thực hiện chụp ảnh màn hình của từng trang web chuẩn bị tiền xử lý, chuẩn hóa làm dữ liệu đầu vào cho thuật toán EfficientNet(B0) thực hiện trích chọn đặc trưng và huấn luyện. Cụ thể quá trình thực hiện như sau:

Bước 1: Sử dụng kỹ thuật Scaling để chuẩn hóa ảnh mầu thô ban đầu về đúng kích thước 224x224 - kích thước chuẩn đầu vào của thuật toán EfficientNet(B0) [81].

Bước 2: Sử dụng mô hình EfficientNet(B0) với cấu trúc cơ bản như Bảng 3.2 *(trong nội dung luận án)*, qua từng lớp của thuật toán thu được vector đặc trưng có kích 1280.

Bước 3: Sau khi thu được tập đặc trưng 1280 từ mô hình EfficientNet(B0), tiếp đến sử dụng một lớp BatchNormalization giúp chuẩn hóa dữ liệu, tránh nhiễu ở các đặc trưng. Sau đó là 1 lớp Dense với 128 node sử dụng hàm kích hoạt là softmax kết hợp với 1 lớp BatchNormalization để chuẩn hóa ngay sau đó, giá trị đầu ra của lớp này sẽ đưa vào làm giá trị đầu vào để tính ra kết quả cuối cùng ở lớp đầu ra với 2 node là 2 giá trị xác suất cho việc hình ảnh đầu vào là bị tấn công thay đổi giao diện hay là trang web bình thường.

3.3.3. Tập dữ liệu thử nghiệm

Dữ liệu trong quá trình thử nghiệm được sử dụng như trong mô tả tại mục *3.2. (trong nội dung luận án).*

Tập dữ liệu được chia ngẫu nhiên thành 3 phần là tập huấn luyện (Training set), tập xác thực (Validation set) và tập kiểm tra (Testing set) theo tỷ lệ như sau:

- Tập huấn luyện chiếm 60% được sử dụng để làm đầu vào mô hình và tinh chỉnh tham số mô hình;

- Tập xác thực chiếm 20% được dùng để kiểm tra độ chính xác của mô hình trong quá trình huấn luyện mô hình nhằm điều chỉnh tham số mô hình tránh việc quá khớp trong quá trình huấn luyện;

- Tập kiểm tra chiếm 20% dùng để đánh giá mô hình sau khi mô hình đã được huấn luyện xong.

3.3.4. Thử nghiệm và kết quả

Bảng 3. 4. Hiệu suất của mô hình phát hiện với các thuật toán học sâu

Kỹ thuật học sâu	**ACC(%)**	**PPV(%)**	**TPR(%)**	**F1(%)**	**FPR(%)**	**FNR(%)**
EfficientNet(B0)	94.12	94.60	90.71	92.62	3.55	9.29
Xception	94.01	93.98	91.21	92.58	4.05	8.79
Inception	89.91	89.37	84.78	87.02	6.69	15.22
Bi-LSTM	89.18	87.73	85.22	86.46	8.13	14.78

Bảng 3.4 cung cấp hiệu suất của mô hình đề xuất dựa trên EfficientNet(B0) và các mô hình dựa trên Xception, Bi-LSTM và Inception. Kết quả cho thấy mô hình đề xuất dựa trên EfficientNet(B0) cho độ đo ACC và F1 tốt nhất tiếp sau là mô hình dựa trên Xception, Inception.

Bảng 3. 6. Hiệu suất mô hình đề suất với các thuật toán học sâu và mô hình trước đó

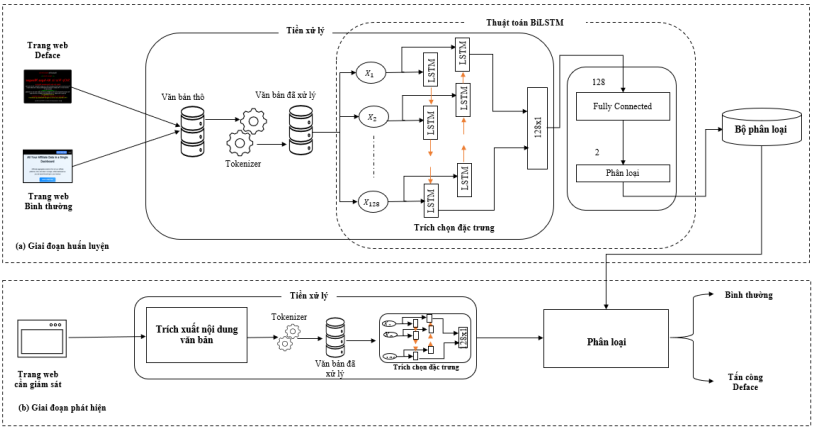
Mô hình phát hiện	Đặc trưng	ACC(%)	PPV(%)	TPR(%)	F1(%)	FPR(%)	FNR(%)
Naïve Bayes Hoang [13]	Văn bản	82,54	78,12	79,26	78,69	15,21	20,74
Cây quyết định Hoang [13]	Văn bản	87,33	84,4	84,4	84,4	10,67	15,6
Rừng ngẫu nhiên Hoang [12]	Văn bản	93,88	93,81	90,76	92,26	4,03	9,24
Xception	Ảnh	94,01	93,98	91,21	92,58	4,05	8,79
Inception	Ảnh	89,91	89,37	84,78	87,02	6,69	15,22
Bi-LSTM	Ảnh	89,18	87,73	85,22	86,46	8,13	14,78
EfficientNet(B0)	Ảnh	94,12	94,60	90,71	92,62	3,55	9,29

Bảng 3. 6. cung cấp số liệu tổng hợp so sánh hiệu suất của mô hình đề xuất với các thuật toán học sâu và các mô hình phát hiện dựa trên các thuật toán học máy Naïve Bayes, cây quyết định và Rừng ngẫu nhiên đề xuất trong Hoang [17] và Hoang [18]. Có thể thấy mô hình đề xuất cho hiệu suất tốt hơn so với các đề xuất của Hoang và cộng sự [17] và tốt hơn nhiều so với Hoang và cộng sự [18].

Hạn chế: mặc dù độ chính xác chung của mô hình đề xuất cao hơn đáng kể so với các mô hình đã có, tuy nhiên tỷ lệ cảnh báo sai, gồm FPR và FNR vẫn trên 10% là tương đối cao và với đặc trưng ảnh mô hình phát hiện kém với những ảnh có ít thay đổi về màu sắc như Hình 3.5. Do đó, phần tiếp theo của Luận án sẽ đưa ra mô hình phát hiện tấn công thay đổi giao diện trang web có thể giải quyết được các vấn đề còn tồn tại phía trên bằng cách sử dụng mô hình phát hiện sử dụng đặc trưng văn bản trong các tệp HTML.

3.4. Phát hiện thay đổi giao diện sử dụng nội dung văn bản

3.4.1. Giới thiệu mô hình



Hình 3. 16. Mô hình huấn luyện, phát hiện tấn công thay đổi giao diện với đặc trưng văn bản

Hình 3.16 biểu diễn mô hình đề xuất cho phát hiện tấn công thay đổi giao diện trang web sử dụng đặc trưng văn bản qua 2 giai đoạn: (a) giai đoạn huấn luyện và (b) giai đoạn phát hiện. Trong giai đoạn huấn luyện, tập dữ liệu huấn luyện được thu thập từ trích xuất nội dung văn bản trong các trang web bị tấn công và trang web bình thường, sau đó được tách thành các từ với kỹ thuật Tokenizer, sau đó huấn luyện với thuật toán học sâu BiLSTM để tạo ra bộ phân loại. Trong giai đoạn phát hiện, trang web giám sát được trích xuất nội dung văn bản, qua quá trình tiền xử lý dữ liệu như giai đoạn huấn luyện và đến bước phân loại sử dụng bộ phân loại từ giai đoạn Huấn luyện để xác định trạng thái là bình thường hay bị thay đổi giao diện.

3.4.2. Tiền xử lý dữ liệu và huấn luyện mô hình

Bước 1: Từ các trang web bình thường và trang web bị thay đổi giao diện, sử dụng một chương trình tự viết bằng python trích xuất các nội dung văn bản làm dữ liệu cho quá trình huấn luyện.

Bước 2: Từ tập dữ liệu văn bản thu được sử dụng sử dụng kỹ thuật Tokenizer [100] để tách các từ trong văn bản và mỗi từ này được ánh xạ thành một số nguyên dương.Tiếp đó lựa chọn 128 từ đầu tiên liên tiếp nhau làm đầu vào cho thuật toán BiLSTM.

Bước 3: Sử dụng lớp Embedding để giúp mô hình hiểu được mối quan hệ ngữ nghĩa của các từ thông qua vector đầu vào của mô hình. Kết quả là một vector 128x128 thể hiện đặc trưng của các từ và mối quan hệ giữa các từ với nhau trong tập dữ liệu, giúp tăng khả năng hiểu nội dung văn bản của mô hình.

Bước 4: Sử dụng lớp GlobalMaxPooling để giảm chiều dữ liệu còn 128.

Bước 5: Lớp kết nối đầy đủ cuối cùng chuyển hóa 128 đặc trưng về giá trị phân loại của mô hình, sử dụng hàm kích hoạt softmax để tính xác suất phát hiện tấn công hay bình thường.

3.4.3. Tập dữ liệu thử nghiệm

Dữ liệu trong quá trình thử nghiệm được sử dụng như trong mô tả tại mục *3.2. (trong nội dung luận án).*

Tập dữ liệu được chia ngẫu nhiên thành 3 phần là tập huấn luyện (Training set), tập xác thực (Validation set) và tập kiểm tra (Testing set) như sau:

- Tập huấn luyện chiếm 60% được sử dụng để làm đầu vào mô hình và tinh chỉnh tham số mô hình;

- Tập xác thực chiếm 20% được dùng để kiểm tra độ chính xác của mô hình trong quá trình huấn luyện mô hình nhằm điều chỉnh tham số mô hình tránh việc quá khớp trong quá trình huấn luyện;

- Tập kiểm tra chiếm 20% dùng để đánh giá mô hình sau khi mô hình đã được huấn luyện xong.

3.4.4. Thử nghiệm và kết quả

Luận án lựa chọn thử nghiệm mô hình phát hiện đề xuất dựa trên thuật toán Bi-LSTM và các mô hình phát hiện đề xuất bởi [18] (Naive Bayes, Decision Tree) và [17] (Rừng ngẫu nhiên) chỉ sử dụng dữ liệu văn bản trích xuất từ trang web để so sánh, đánh giá.

Bảng 3.8. Kết quả thử nghiệm các mô hình phát hiện dựa trên các thuật toán học máy chỉ sử dụng đặc trưng văn bản

Mô hình phát hiện	**Đặc trưng**	**ACC(%)**	**PPV(%)**	**TPR(%)**	**F1(%)**	**FPR(%)**	**FNR(%)**
Naïve Bayes Hoang [13]	Văn bản	82,54	78,12	79,26	78,69	15,21	20,74
Cây quyết định Hoang [13]	Văn bản	87,33	84,4	84,4	84,4	10,67	15,6
Rừng ngẫu nhiên Hoang [12]	Văn bản	93,88	93,81	90,76	92,26	4,03	9,24
BiLSTM	Văn bản	96,54	96,93	94,43	95,66	2,03	5,57

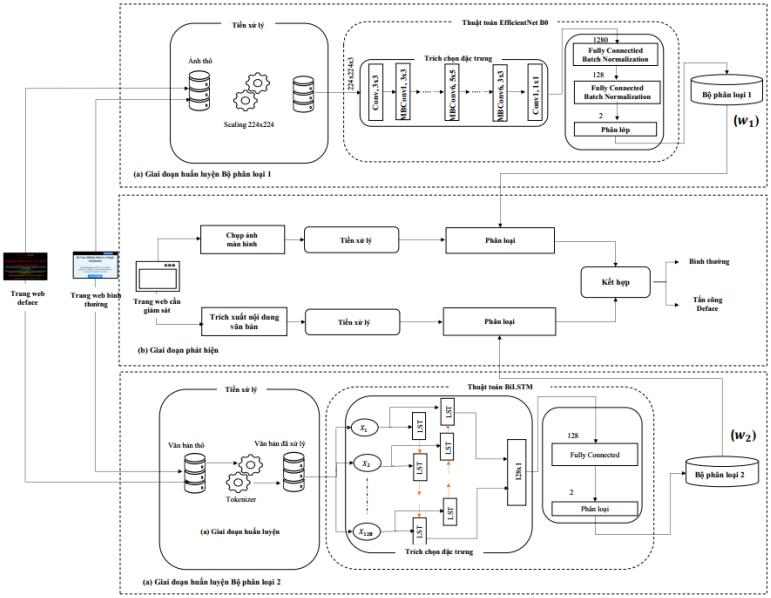
Kết quả thử nghiệm cho trên Bảng 3.8. Có thể thấy độ chính xác ACC và độ đo F1 của mô hình phát hiện dựa trên BiLSTM cao

nhất so với các mô hình đề xuất bởi [18] (Naive Bayes, Decision Tree) và [17] (Rừng ngẫu nhiên).

3.5. Phát hiện thay đổi giao diện sử dụng kết hợp nội dung văn bản và ảnh chụp màn hình trang web

3.5.1. Mô tả mô hình phát hiện

Mô hình đề xuất phát hiện tấn công thay đổi giao diện trang web sử dụng kết hợp đặc trưng hình ảnh và văn bản bao gồm: (a) giai đoạn huấn luyện và (b) giai đoạn phát hiện. Giai đoạn huấn luyện đã được thực hiện tại các mô hình nhánh tại mục 3.3. Phát hiện thay đổi giao diện sử dụng ảnh chụp màn hình trang web với kết quả thu được là Bộ phân loại 1 và mục 3.4. Phát hiện tấn công thay đổi giao diện sử dụng nội dung văn bản với kết quả thu được là Bộ phân loại 2.



Hình 3. 20. Mô hình phát hiện tấn công thay đổi giao diện kết hợp đặc trưng văn bản và hình ảnh trang web

Cả hai bộ phân loại đều được sử dụng cho giai đoạn phát hiện như mô tả trên Hình 3. 20. Trong giai đoạn phát hiện, trang web cần giám sát được tách hai đặc trưng là hình ảnh chụp màn hình và dữ liệu văn bản, lần lượt thực hiện với các mô hình nhánh tương ứng, kết quả cuối cùng của mỗi nhánh sau khi được phân loại sẽ được kết hợp bằng phương pháp học kết hợp (tính trung có trọng số kết quả từ các mô hình nhánh) từ đó có bộ phân loại cuối cùng giúp phát hiện trạng thái bình thường hoặc bị tấn công thay đổi giao diện của trang web cần giám sát.

3.5.2. Tiền xử lý dữ liệu, huấn luyện và phát hiện

Các tập dữ liệu sau tiền xử lý được huấn luyện để xây dựng các mô hình phát hiện thành phần. Dữ liệu văn bản thuần được huấn luyện sử dụng thuật toán học sâu BiLSTM và dữ liệu ảnh chụp màn hình được huấn luyện sử dụng thuật toán học sâu EfficientNet.

3.5.3. Tập dữ liệu thử nghiệm

Dữ liệu trong quá trình thử nghiệm được sử dụng như trong mô tả tại mục *3.2.(trong nội dung luận án)*. Tập dữ liệu thử nghiệm cũng được chia ngẫu nhiên thành 3 tập con: 60% cho tập huấn luyện (Training Set), 20% tập xác thực (Validation Set) và 20% cho tập kiểm tra (Testing Set).

3.5.4. Thử nghiệm và kết quả

Bảng 3. 11. Kết quả thực nghiệm mô hình kết hợp

Kỹ thuật học sâu và kết hợp	**Đặc trưng**	**ACC(%)**	**PPV(%)**	**TPR(%)**	**F1(%)**	**FPR(%)**	**FNR(%)**
Naïve Bayes Hoang[14]	Văn bản	82,54	78,12	79,26	78,69	15,21	20,74
Cây quyết định Hoang [14]	Văn bản	87,33	84,4	84,4	84,40	10,67	15,6
Rừng ngẫu nhiên Hoang [13]	Văn bản	93,88	93,81	90,76	92,26	4,03	9,24
SVM Siyan Wu[67]	Văn bản	95,34	95,37	95,34	95,32		

Kỹ thuật học sâu và kết hợp	Đặc trưng	ACC(%)	PPV(%)	TPR(%)	F1(%)	FPR(%)	FNR(%)
EfficientNet(B0)	Ảnh	94,12	94,60	90,71	92,62	3,55	9,29
BiLSTM	Văn bản	96,54	96,93	94,43	95,66	2,03	5,57
BiLSTM+ EfficientNet (Kết hợp)	Văn bản và Ảnh	98,12	98,83	96,49	97,65	0,78	3,51

Kết quả trên Bảng 3.11. cho thấy phương pháp kết hợp của hai nhánh mô hình với các thuật toán BiLSTM và EfficientNet(B0) cho kết quả là tốt hơn kết quả cho bởi các nhánh độc lập; và cũng tốt hơn các mô hình đã có Naïve Bayes Hoang[18], Cây quyết định Hoang[18], Rừng ngẫu nhiên Hoang[17], SVM Siyan Wu [95].

KẾT LUẬN

Luận án này tập trung giải quyết hai vấn đề: (1) nghiên cứu, đề xuất mô hình phát hiện tấn công web dựa trên học máy có giám sát sử dụng dữ liệu web log, nhằm tăng tỷ lệ phát hiện đúng và giảm tỷ lệ cảnh báo sai, đồng thời mô hình có khả năng phát hiện 4 kiểu tấn công web nguy hiểm bao gồm SQLi, XSS, CMDi và duyệt đường dẫn; và (2) nghiên cứu, đề xuất các đặc trưng và lựa chọn sử dụng phương pháp học sâu phù hợp với các đặc trưng cụ thể cho xây dựng mô hình phát hiện tấn công thay đổi giao diện trang web, nhằm xây dựng mô hình phát hiện cho phép phát hiện hiệu quả tấn công thay đổi giao diện trang web. Vấn đề (1) được giải quyết bởi đóng góp thứ nhất của luận án, còn vấn đề thứ (2) được giải quyết bởi đóng góp thứ hai của luận án.

NHỮNG ĐÓNG GÓP CỦA LUẬN ÁN

Đóng góp thứ nhất của luận án là đề xuất mô hình phát hiện các dạng tấn công web dựa trên học máy sử dụng các đặc trưng ký tự trong dữ liệu truy vấn URI trích xuất từ web log. Các thuật toán học máy có giám sát được sử dụng gồm Rừng ngẫu nhiên, Cây quyết định, Navie Bayes và SVM. Mô hình đề xuất có khả năng phát hiện

hiệu quả bốn dạng tấn công web thường gặp nguy hiểm nhất, bao gồm SQLi, XSS, CMDi và duyệt đường dẫn. Các thử nghiệm trên tập dữ liệu mẫu được dán nhãn và tập dữ liệu web log thực, khẳng định mô hình đề xuất dựa trên thuật toán Rừng ngẫu nhiên cho hiệu suất tốt hơn các mô hình phát hiện dựa trên học sâu[15][16]. Ngoài hiệu suất phát hiện cao, mô hình đề xuất còn có một số ưu điểm so với các đề xuất trước đó: (i) mô hình đề xuất được xây dựng sử dụng các thuật toán học máy có giám sát truyền thống với chi phí tính toán thấp nhưng vẫn đạt được kết quả cao, điều này rất quan trọng khi triển khai thực tế vì hệ thống phát hiện tấn công web thường phải xử lý một lượng web log rất lớn và (ii) mô hình đề xuất có thể được xây dựng tự động từ dữ liệu huấn luyện và không yêu cầu cập nhật thường xuyên.

Đóng góp thứ hai của luận án là đề xuất ba mô hình phát hiện tấn công thay đổi giao diện trang web dựa trên học sâu sử dụng đặc trưng hình ảnh chụp màn hình của trang web, đặc trưng văn bản trích xuất từ trang web và kết hợp các đặc trưng văn bản trích xuất từ trang web kết hợp với các đặc trưng hình ảnh chụp màn hình của trang web. Các kết quả thử nghiệm cho thấy mô hình phát hiện nhánh dựa trên EfficientNet, mô hình phát hiện nhánh dựa trên BiLSTM và mô hình kết hợp đều cho hiệu suất phát hiện cao hơn các mô hình đề xuất bởi các nghiên cứu đi trước và mô hình dựa trên các thuật toán học sâu khác. Đặc biệt, mô hình phát hiện dựa trên kết hợp hai đặc trưng hình ảnh và văn bản của trang web có hiệu suất phát hiện vượt trội so với kết quả đề xuất bởi Hoang và cộng sự [17] và Hoang[18], cũng như các mô hình dựa trên các thuật toán học sâu Xception, Inception và Bi-LSTM và EfficientNet chỉ các đặc trưng hình ảnh chụp màn hình của trang web.

Các vấn đề hay tồn tại của các đề xuất trong luận án cũng chính là các hướng mở cho tiếp tục nghiên cứu, bổ sung. Cụ thể:

- Vấn đề thứ nhất là các truy vấn URI chỉ có thể trích xuất từ web log nếu phương thức HTTP sử dụng là GET. Nếu phương thức sử dụng là POST, dữ liệu gửi từ máy khách đến máy chủ không được lưu trong log. Một hướng giải quyết vấn đề này là triển khai mô hình phát hiện dưới dạng 1 tường lửa ứng dụng web (WAF) để bắt và xử lý tất cả các yêu cầu truy cập từ người dùng.

- Mô hình phát hiện tấn công thay đổi giao diện dựa trên BiLSTM và EfficientNet đòi hỏi chi phí tính toán lớn cho huấn luyện do phải xử lý một lượng lớn ảnh chụp màn hình và nội dung văn bản trang web sử dụng các thuật toán học sâu. Một hướng khắc phục vấn đề này là việc huấn luyện mô hình có thể thực hiện offline nên không ảnh hưởng nhiều đến thời gian phát hiện. Ngoài ra, mô hình có thể kết hợp sử dụng phát hiện dựa trên chữ ký với các dạng tấn công đã biết nhằm giảm thời gian phát hiện.

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

[CT1] Hoang Xuan Dau, Ninh Thi Thu Trang, **Nguyen Trong Hung**. *"A Survey of Tools and Techniques for Web Attack Detection".* Journal of Science and Technology on Information security, Special Issue CS (15) 2022, pp. 109-118.

[CT2] Xuan Dau Hoang, **Trong Hung Nguyen**, *" Detecting commonweb attacks based on supervised machine learning using web logs"*, Journal of Theoretical and Applied Information Technology Vol.99. No 6, 31st March 2021, Scopus Q4.

[CT3] **Trong Hung Nguyen,** Xuan Dau Hoang, Duc Dung Nguyen, *"Detecting Website Defacement Attacks using Web-page Text and Image Features",* Article Published in International Journal of Advanced Computer Science and Applications(IJACSA), Volume 12 Issue 7, 2021, Scopus Q3.

[CT4] Hoàng Xuân Dậu, **Nguyễn Trọng Hưng**, *"Phát hiện tấn công web thường gặp dựa trên học máy sử dụng web log*", Hội nghị khoa học quốc gia về "Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin" FAIR 2020.

[CT5] **Trong Hung Nguyen**, Dau Hoang, Nguyen Duc Dung, Vu Xuan Hanh. *"Phát hiện tấn công thay đổi giao diện trang web sử dụng đặc trưng văn bản"*, Hội nghị KHCN Quốc gia lần thứ XVII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin(FAIR), Hà Nội, 8/2024.

[CT6] Xuan Dau Hoang, **Trong Hung Nguyen**, Hoang Duy Pham, "A Novel Model for Detecting Web Defacement Attacks Using Plain Text Features" Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 2024, Scopus Q3 *(Đã nhận được thư chấp nhận đăng).*