**ĐẠI HỌC BÁCH KHOA HÀ NỘI**

# ĐỒ ÁN TỐT NGHIỆP

## Phát hiện tấn công web chưa biết sử dụng mạng thích nghi sâu

**NGUYỄN QUỐC HƯNG**
hung.nq200294@sis.hust.edu.vn

**Ngành: Kỹ thuật máy tính**

**Giảng viên hướng dẫn:** TS. Tống Văn Vạn _____

Chữ kí GVHD

**Ngành:** Kỹ thuật máy tính

**Trường:** Công nghệ Thông tin và Truyền thông

**HÀ NỘI, 06/2024**

# LỜI CAM KẾT

Họ và tên sinh viên: Nguyễn Quốc Hưng
Điện thoại liên lạc: 0982947213
Email: hung.nq200294@sis.hust.edu.vn
Lớp: Kỹ thuật máy tính 01
Hệ đào tạo: Cử nhân

Tôi – *Nguyễn Quốc Hưng* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *TS.Tống Văn Vạn*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

*Hà Nội, ngày 30 tháng 6 năm 2024*

Tác giả ĐATN

*Nguyễn Quốc Hưng*

# LỜI CẢM ƠN

Tôi xin gửi lời cảm ơn chân thành đến T.S Tống Văn Vạn, người đã tận tâm hỗ trợ và hướng dẫn tôi trong suốt quá trình nghiên cứu và hoàn thành đồ án này. Thầy đã luôn sẵn sàng chia sẻ kiến thức sâu rộng và những lời khuyên quý báu, giúp tôi vượt qua nhiều thử thách khó khăn và đạt được những kết quả đáng khích lệ. Không có sự giúp đỡ và sự kiên nhẫn của thầy, tôi sẽ không thể hoàn thành được đồ án này.

Tôi cũng muốn bày tỏ lòng biết ơn sâu sắc đến bố mẹ và bạn gái tôi. Họ đã luôn đứng bên cạnh, ủng hộ và động viên tôi trong suốt hành trình đầy thách thức này. Sự yêu thương và khích lệ không ngừng nghỉ của họ là nguồn động lực to lớn, giúp tôi vượt qua những thời điểm khó khăn nhất. Sự kiên nhẫn và sự tin tưởng của họ đã là động lực để tôi không ngừng phấn đấu.

Ngoài ra, tôi xin gửi lời cảm ơn chân thành đến các bạn bè thân thiết, những người đã luôn đồng hành và chia sẻ cùng tôi trên con đường học tập và nghiên cứu. Các bạn đã hỗ trợ tôi không chỉ về mặt tinh thần mà còn trong công việc nghiên cứu, giúp tôi có được những góc nhìn và ý tưởng mới mẻ. Sự sẻ chia và đồng cảm của các bạn đã giúp tôi vượt qua nhiều trở ngại và mang lại niềm vui, sự hứng khởi trong suốt quá trình học tập và làm việc. Sự động viên và giúp đỡ của các bạn là một phần không thể thiếu trong thành công của tôi.

# TÓM TẮT NỘI DUNG ĐỒ ÁN

Với sự phát triển nhanh chóng của công nghệ thông tin, các ứng dụng web đã trở thành một phần không thể thiếu trong cuộc sống hàng ngày, đặc biệt khi ngày càng nhiều người chuyển ứng dụng và dữ liệu nhạy cảm lên đám mây. Sự phổ biến của các ứng dụng web và lượng lớn dữ liệu nhạy cảm mà chúng xử lý làm cho chúng dễ trở thành mục tiêu của các cuộc tấn công. Do đó, việc bảo vệ các ứng dụng web khỏi các mối đe dọa này là vô cùng quan trọng. Các cuộc tấn công phổ biến như SQL Injection, XSS, và CSRF đã được OWASP xác định là những mối đe dọa nghiêm trọng. Sự phát triển không ngừng của các ứng dụng web dẫn đến sự xuất hiện của các biến thể tấn công mới, đòi hỏi các hệ thống phát hiện không chỉ nhận dạng các mô hình tấn công đã biết mà còn phải thích nghi với những mô hình mới chưa biết. Đề tài này tập trung vào phát triển một hệ thống phân loại lỗ hổng bảo mật mới với độ chính xác cao và thời gian xử lý hợp lý, có khả năng xử lý đồng thời nhiều yêu cầu. Chúng tôi đề xuất sử dụng mạng thích ứng miền sâu (Deep Adaptation Network) để giải quyết các hạn chế của các phương pháp hiện tại, bao gồm cả những phương pháp dựa trên chữ ký như ModSecurity và các kỹ thuật học máy, vốn gặp khó khăn khi phải đối mặt với các lỗ hổng mới (zero-days). Hệ thống của chúng tôi áp dụng các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP), mô hình BERT, học sâu, và chiến lược học chuyển giao để phát hiện và phân loại các cuộc tấn công web chưa biết. Cụ thể, chúng tôi sử dụng SecBERT để trích xuất đặc trưng từ các yêu cầu HTTP và áp dụng tỷ lệ chuyển giao động cho mạng thích nghi sâu nhằm cải thiện độ chính xác và tốc độ hội tụ của mô hình. Kết quả thí nghiệm cho thấy phương pháp đề xuất đạt F1-score cao, đặc biệt là 99% trong phân loại các lỗ hổng bảo mật đã biết và 71% đối với các lỗ hổng mới với tỷ lệ chuyển giao động. Hệ thống khi tích hợp với WAF có khả năng phân loại với độ chính xác cao và thời gian phản hồi trung bình khoảng 18ms, đánh dấu một bước tiến quan trọng trong việc tăng cường an ninh cho các ứng dụng web.

Sinh viên thực hiện

*(Ký và ghi rõ họ tên)*

# ABSTRACT

With the rapid advancement of information technology, web applications have become an indispensable part of daily life, especially as more people move their applications and sensitive data to the cloud. The widespread use of web applications and the vast amount of sensitive data they process make them prime targets for attacks. Therefore, protecting web applications from these threats is of paramount importance. Common attacks such as SQL Injection, XSS, and CSRF have been identified by OWASP as serious threats. The continuous development of web applications has led to the emergence of new attack variants, requiring detection systems not only to recognize known attack patterns but also to adapt to new, unknown ones. This project focuses on developing a new security vulnerability classification system with high accuracy and reasonable processing time, capable of handling multiple requests simultaneously. We propose using a Deep Domain Adaptation Network to address the limitations of current methods, including signature-based approaches like ModSecurity and machine learning techniques, which struggle with new vulnerabilities (zero-days). Our system employs Natural Language Processing (NLP) techniques, the BERT model, deep learning, and transfer learning strategies to detect and classify unknown web attacks. Specifically, we use SecBERT to extract features from HTTP requests and apply a dynamic transfer rate for the deep domain adaptation network to improve the model's accuracy and convergence speed. Experimental results show that the proposed method achieves high F1-scores, notably 99% in classifying known security vulnerabilities and 71% for new vulnerabilities using a dynamic transfer rate. When integrated with WAF, the system can classify with high accuracy and an average response time of just 18ms, marking a significant advancement in enhancing the security of web applications.

# MỤC LỤC

# DANH MỤC HÌNH VẼ

# DANH MỤC BẢNG BIỂU

# DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

| Thuật ngữ | Ý nghĩa |
|---|---|
| BOW | Bag-of-words |
| CMD | Central Moment Discrepancy |
| CNN | Mạng thần kinh tích chập(Convolutional Neural Network) |
| Coral | Correlation Alignment |
| CRS | Core Rule Set |
| DA | Domain Adaptation |
| DAN | Mạng thích nghi sâu (Domain Adaptation Network) |
| DDC | Deep Domain Confusion |
| DL | Deep Learning |
| HTML | Ngôn ngữ đánh dấu siêu văn bản (HyperText Markup Language) |
| HTTP | Hypertext Transfer Protocol |
| LSTM | Long Short-Term Memory |
| MK-MMD | Multi-kernel Maximum Mean Discrepancy |
| ML | Machine Learning |
| MLP | Multi-layer Perceptron |
| MMD | Maximum Mean Discrepancy |
| NLP | Natural Language Processing |
| OWASP | Open Web Application Security Project |
| SQLi | SQL injection |
| SWD | Sliced Wasserstein Distance |
| TF-IDF | Term Frequency Inverse Document Frequency |
| UDA | Unsupervised Domain Adaptation |
| WAFs | Web Application Firewalls |
| XSS | Cross-Site Scripting |