THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút): *https://youtu.be/AmMtdnIGm98*
- Link slides (dạng .pdf đặt trên Github): *https://github.com/tuananh8232/CS2205.CH181/blob/main/SQL-INJECTION-ATTACK-DETECTION-SYSTEM-USING-BERT-AND-MACHINE-LEARNING-TECHNIQUES_SLIDE.pdf*
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

• Họ và Tên: Lê Tuấn Anh • MSSV: 230202001 	• Lớp: CS2205.CH181 • Tự đánh giá (điểm tổng kết môn): 8/10 • Số buổi vắng: 1 • Số câu hỏi QT cá nhân: • Link Github: https://github.com/tuananh8232/CS2205.CH181

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

HỆ THỐNG PHÁT HIỆN TẤN CÔNG TIÊM NHIỄM SQL SỬ DỤNG BERT VÀ CÁC KỸ THUẬT MÁY HỌC

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

SQL INJECTION ATTACK DETECTION SYSTEM USING BERT AND MACHINE LEARNING TECHNIQUES

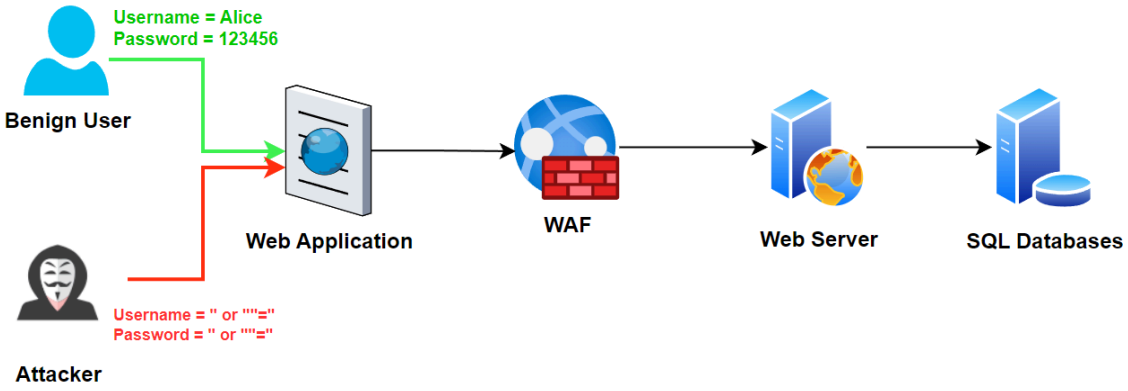
TÓM TẮT *(Tối đa 400 từ)*

An ninh mạng là một trong những vấn đề được quan tâm hàng đầu trong thời đại công nghệ hiện nay. Các ứng dụng web ngày càng phổ biến, trở thành mục tiêu tiềm năng cho các cuộc tấn công mạng. Một trong những loại tấn công nguy hiểm nhất là tấn công tiêm nhiễm SQL (SQL injection - SQLi). SQLi cho phép kẻ tấn công thực thi các lệnh trái phép trên máy chủ cơ sở dữ liệu, dẫn đến rò rỉ dữ liệu nhạy cảm, phá hoại hệ thống. Nghiên cứu này nhằm mục tiêu phát triển một hệ thống phát hiện tấn công SQLi sử dụng kết hợp mô hình Bidirectional Encoder Representations from Transformers (BERT) và các mô hình phân loại máy học. Hệ thống có khả năng tự phân biệt các truy vấn SQL bình thường và truy vấn SQL độc hại với độ chính xác cao, giúp tăng cường khả năng bảo mật cho các ứng dụng web. Nghiên cứu được thực hiện dựa trên việc tìm hiểu về phương pháp kết hợp, tinh chỉnh mô hình BERT và các mô hình phân loại máy học như Naive Bayes, SVM, XGBoost để đạt được kết quả tốt nhất. Nghiên cứu được thực nghiệm trên tập dữ liệu được lấy từ Kaggle với hơn 30000 câu lệnh SQL. Nghiên cứu được đánh giá dựa trên các thông số như độ chính xác, hiệu suất và các thông số khác, từ đó tối ưu và đưa ra mô hình tối ưu nhất có thể.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Các cuộc tấn công SQLi rất đa dạng về hình thức và có thể rất tinh vi. Chúng có thể khai thác kết hợp cùng nhiều hình thức tấn công khác nhau, làm cho việc phát hiện tấn công và phòng thủ trở nên khó khăn. Trong một cuộc tấn công SQLi, kẻ tấn công

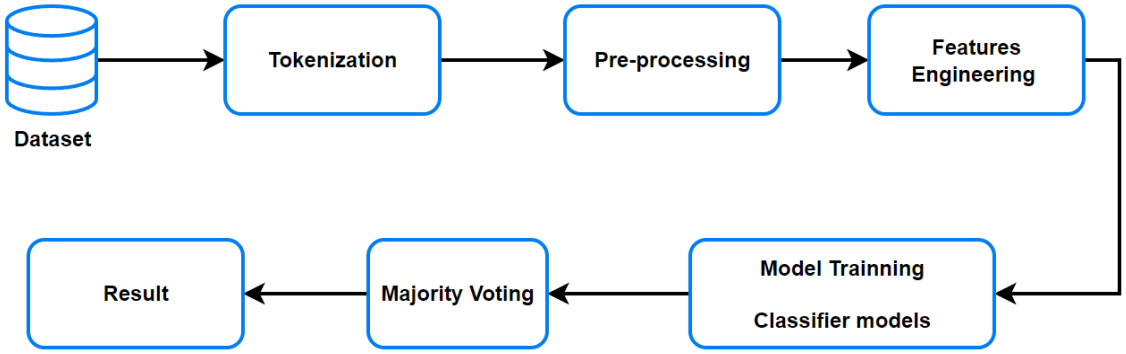
khai thác lỗ hổng của các ứng dụng từ đó chèn các lệnh SQL độc hại vào câu truy vấn, từ đó khai thác các thông tin nhạy cảm. Mặc dù hệ thống tường lửa ứng dụng web (WAF) dựa trên quy tắc được sử dụng rất phổ biến ngày nay, nhưng các kẻ tấn công thường có thể vượt qua các quy tắc do người quản trị đặt ra bằng cách tạo ra các câu lệnh tinh vi hơn. Do đó WAF dựa trên quy tắc thường tạo có tỷ lệ dương tính giả và khó khăn trong việc phát hiện tấn công.



Hình 1. Mô phỏng tấn công SQLi

Từ những hạn chế trên, nghiên cứu này được thực hiện nhằm mục tiêu tạo ra hệ thống phát hiện tấn công SQLi sử dụng mô hình BERT kết hợp với các thuật toán phân loại máy học, giúp cho việc phát hiện tấn công hiệu quả hơn. Nghiên cứu sử dụng kết hợp mô hình BERT và các mô hình phân loại máy học như Naive Bayes, SVM, XGBoost và một số kĩ thuật nhằm tạo ra mô hình tối ưu nhất cho bài toán.

Mô hình đề xuất:



Input: Câu lệnh truy vấn SQL

Output: Phân loại là truy vấn tấn công hay truy vấn bình thường

MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

- Nghiên cứu về mô hình BERT và các mô hình phân loại máy học, kết hợp được các mô hình với nhau, áp dụng được vào bài toán phát hiện tấn công SQLi.
- Tạo ra được mô hình phát hiện tấn công SQLi với độ chính xác và hiệu suất tốt, đáp ứng được yêu cầu của bài toán đặt ra nhưng cũng đảm bảo về độ chính xác cũng như hiệu năng của mô hình.
- Đưa ra các so sánh, đánh giá với các phương pháp hiện có.

NỘI DUNG VÀ PHƯƠNG PHÁP

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

Nội dung:

- Nghiên cứu về một số hình thức tấn công Web phổ biến, tấn công SQLi.
- Nghiên cứu về mô hình BERT, các mô hình phân loại máy học Naive Bayes, SVM, XGBoost, các kĩ thuật Majority Voting, Random Search.
- Nghiên cứu, tìm ra phương pháp áp dụng, kết hợp các mô hình đã đề xuất áp dụng vào bài toán. Xây dựng được kiến trúc mô hình.
- Thu thập dataset về tấn công SQLi
- Huấn luyện mô hình, đánh giá mô hình.
- Xây dựng hệ thống phát hiện tấn công SQLi dựa trên mô hình đã huấn luyện được.

Phương pháp:

- Tìm hiểu cơ chế, kiến trúc của các mô hình BERT, các mô hình phân loại máy học: Naive Bayes, SVM, XGBoost, các kĩ thuật Majority Voting, Random Search.
- Khảo sát tổng luận, tìm hiểu về các nghiên cứu liên quan qua các nguồn tạp chí, hội nghị uy tín.
- Thu thập dữ liệu, tìm hiểu các bộ dữ liệu từ Kaggle, đánh giá, tiền xử lý dữ liệu cho phù hợp với Input của mô hình.

- Thực hiện xây dựng kiến trúc mô hình, huấn luyện mô hình trên tập dữ liệu đã thu thập, thực hiện tinh chỉnh mô hình, tối ưu tham số nhằm cải thiện độ chính xác. Đánh giá mô hình thông qua các thông số như: Accuracy, Precision, Recall, F1 score, Confusion Matrix.
- So sánh tính hiệu quả so với các phương pháp, nghiên cứu liên quan, đưa ra đánh giá, kết luận.
- Phát triển hệ thống thử nghiệm với mô hình đã huấn luyện được.

KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

- Báo cáo về kiến trúc, kỹ thuật, và phương pháp áp dụng vào bài toán của các mô hình đã sử như BERT, các mô hình phân loại máy học Naive Bayes, SVM, XGBoost và các kỹ thuật Majority Voting, Random Search.
- Kết quả thực nghiệm, đánh giá, kết luận và so sánh giữa phương pháp đã thực nghiệm với các phương pháp khác.
- Hệ thống phát hiện tấn công SQLi sử dụng mô hình đã thử nghiệm.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1]. Ignacio Samuel Crespo-Martínez, Adrián Campazas Vega, Ángel Manuel Guerrero-Higueras, Virginia Riego-Del Castillo, Claudia Álvarez-Aparicio, Camino Fernández Llamas: SQL injection attack detection in network flow data. Comput. Secur. 127: 103093 (2023)

[2]. Chen, Yuhong, et al. "Request-Response Network Traffic Packets: Enhancing SQL Injection Attack Detection with a Transformer-Based Model." 2023 9th International Conference on Computer and Communications (ICCC). IEEE, 2023.

[3]. Kasim Tasdemir, Rafiullah Khan, Fahad Siddiqui, Sakir Sezer, Fatih Kurugollu, Sena Busra Yengec-Tasdemir, Alperen Bolat: Advancing SQL Injection Detection for High-Speed Data Centers: A Novel Approach Using Cascaded NLP. CoRR abs/2312.13041 (2023)

[4]. Janet Zulu, Bonian Han, Izzat Alsmadi, Gongbo Liang: Enhancing Machine

Learning Based SQL Injection Detection Using Contextualized Word Embedding. ACM Southeast Regional Conference 2024: 211-216