ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
VNUHCM - UIT

PHÁT HIỆN TẤN CÔNG SQL INJECTION SỬ DỤNG MÔ HÌNH XỬ LÝ NGÔN NGỮ TỰ NHIÊN VÀ MẠNG SINH ĐỐI KHÁNG

Đồng Thị Ngọc Trâm - 230201057

Tóm tắt

- Lớp: CS2205.MAR2024
- Link Github:
- Link YouTube video: https://youtu.be/D9p1ekTGkgc
- Ảnh + Họ và Tên: Đồng Thị Ngọc Trâm
- Tổng số slides: 10

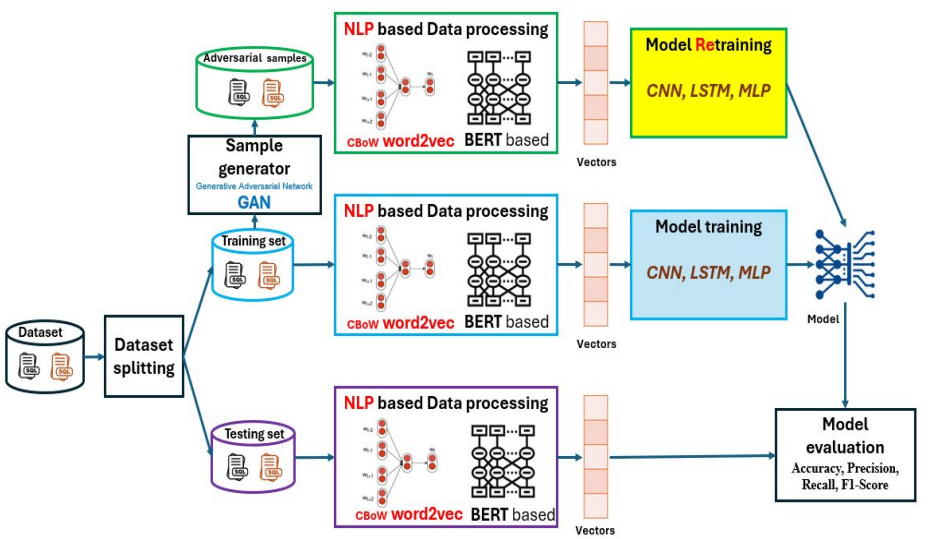
Giới thiệu

Tấn công SQL injection luôn đứng vị trí số một trong nhóm mười nguy cơ bảo mật phổ biến nhất đối với các ứng dụng web theo đánh giá của tổ chức OWASP (Open Web Application Security Project)[1]. Có nhiều nghiên cứu trong lĩnh vực phát hiện tấn công SQL injection. Một số dùng kỹ thuật phát hiện từ khóa, một số dùng mô hình học máy. Tuy nhiên:

- Các nghiên cứu hiện tại chỉ dùng một trong các mô hình xử lý ngôn ngữ tự nhiên để rút trích đặc trưng → cần đánh giá **nhiều mô hình NLP** cho bài toán SQL injection
- Bộ dataset hiện tại làm các mô hình có khả năng bị overfit → cần bổ sung **thêm số lượng** mẫu thử cũng như **đa dạng các câu truy vấn** độc hại để tránh overfit mô hình → cần dùng kỹ thuật hiện đại như **GAN**.

[1] https://owasp.org/www-project-top-ten/

Mục tiêu



- Xây dựng được mô-đun rút trích đặc trưng bằng mô hình xử lý ngôn ngữ tự nhiên. Đánh giá, so sánh các mô hình để chọn mô hình phù hợp cho từng loại môi trường triển khai.
- Xây dựng được mô-đun tạo các câu truy vấn SQL bằng mạng sinh đối kháng.
- Xây dựng được mô hình phát hiện tấn công SQL injection bằng các mô hình học sâu như CNN, LSTM, MLP trên bộ dữ liệu thử nghiệm gốc từ Kaggle và bộ dữ liệu được tạo từ GAN trong nghiên cứu .

Nội dung và Phương pháp

- **Nội dung 1: Mô-đun rút trích đặc trưng bằng mô hình xử lý ngôn ngữ tự nhiên**

- Tìm hiểu cấu trúc các lệnh SQL, các kỹ thuật tấn công SQL injection như Boolean based, Union based, error based,... để phục vụ cho việc tiền xử lý trước khi thực hiện vector hóa đặc trưng.
- Khảo sát và tổng hợp các công trình nghiên cứu liên quan (literature review) về các phương pháp phát hiện tấn công SQL injection để xác định các khoảng trống nghiên cứu (research gaps).
- Nghiên cứu cách thức vector hóa các đặc trưng bằng các mô hình xử lý ngôn ngữ tự nhiên như word2vec, BERT, DistilBERT.
- Thử nghiệm các cách thức vector hóa khác nhau, đánh giá về thời gian, bộ nhớ cần thiết cho từng mô hình.

Nội dung và Phương pháp

- **Nội dung 2: Mô-đun tạo các câu truy vấn SQL bằng mạng sinh đối kháng**

- Nghiên cứu và triển khai mô hình mạng sinh đối kháng phổ biến như DCGAN.
- Từ các vector được rút trích từ nội dung 1, thực hiện tạo vector tùy chỉnh bằng GAN.
- Từ vector tùy chỉnh này thực hiện tạo câu truy vấn SQL tùy chỉnh từ GAN.
- Kiểm thử chất lượng mẫu thử được tạo từ GAN.
- Chuẩn hóa dataset mới được tạo từ GAN để cung cấp cho các nghiên cứu liên quan trong tương lai.

Nội dung và Phương pháp

- **Nội dung 3: Mô hình phát hiện tấn công SQL injection bằng các mô hình học sâu**

- Nghiên cứu Xây dựng các mô hình nhận diện tấn công SQL injection dựa vào các mô hình học sâu như CNN, LSTM, MLP bằng cách huấn luyện từ bộ dữ liệu gốc và bộ dữ liệu thử nghiệm phát sinh từ GAN.
- Tinh chỉnh các siêu tham số của các mô hình để xác định bộ siêu tham số tối ưu
- Sử dụng các độ đo như Accuracy, F1-Score, Recall, Precision để đánh giá mô hình được huấn luyện từ dataset gốc và dataset được tạo từ GAN.

Kết quả dự kiến

- Xây dựng được hệ thống phát hiện tấn công SQL injection sử dụng các mô hình xử lý ngôn ngữ tự nhiên khác nhau và có khả năng tái huấn luyện dựa vào mẫu thử phát sinh từ mạng sinh đối kháng. Độ chính xác mô hình tăng khi tái huấn luyện với dataset mới.
- Xây dựng được mô-đun phát sinh các câu truy vấn SQL làm mẫu thử bằng cách dùng mạng sinh đối kháng.

Tài liệu tham khảo

- OWASP. (2024, May 01). OWASP Top Ten. Available: https://owasp.org/www-project-top-ten/
- S. Lakhani, A. Yadav, and V. Singh, "Detecting SQL Injection Attack using Natural Language Processing," in 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2022, pp. 1-5.
- J. Devlin, M. Chang, K. Lee, and K. Toutanova, "Bidirectional encoder representations from transformers," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, vol. 1, pp. 4171-4186.
- H. Zhang and M. O. Shafiq, "Survey of transformers and towards ensemble learning using transformers for natural language processing," Journal of big Data, vol. 11, no. 1, p. 25, 2024.
- Q. Jiao and S. Zhang, "A brief survey of word embedding and its recent development," in 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021, vol. 5, pp. 1697-1701: IEEE.

Tài liệu tham khảo

- OWASP Y. Natarajan, B. Karthikeyan, G. Wadhwa, S. A. Srinivasan, and A. S. P. Akilesh, "A Deep Learning Based Natural Language Processing Approach for Detecting SQL Injection Attack," Cham, 2023, pp. 396-406: Springer Nature Switzerland.
- B. Gogoi, T. Ahmed, and A. Dutta, "Defending against SQL Injection Attacks in Web Applications using Machine Learning and Natural Language Processing," in 2021 IEEE 18th India Council International Conference (INDICON), 2021, pp. 1-6.
- Y. Fang, J. Peng, L. Liu, and C. Huang, "WOVSQLI: Detection of SQL Injection Behaviors Using Word Vector and LSTM," presented at the Proceedings of the 2nd International Conference on Cryptography, Security and Privacy, Guiyang, China, 2018. Available: https://doi.org/10.1145/3199478.3199503
- F. Zhong, X. Cheng, D. Yu, B. Gong, S. Song, and J. Yu, "MalFox: Camouflaged adversarial malware example generation based on conv-GANs against black-box detectors," IEEE Transactions on Computers, 2023.