PHÁT HIỆN MÃ ĐỘC PE SỬ DỤNG ĐẶC TRƯNG ĐƯỢC TỐI ƯU HÓA

Lê Văn Minh - 230202011

Tóm tắt

- Lớp: CS2205.CH181
- Link Github: https://github.com/minh240899/CS2205.MAR2024
- Link YouTube video: https://youtu.be/lHFBZgpaWsE
- Lê Văn Minh

Tóm tắt

- Sự tiến bộ vượt bậc của công nghệ thông tin: Máy tính trở thành một phần không thể thiếu, đồng thời mở ra các cuộc tấn công mạng nguy hiểm.
- Malware (phần mềm độc hại): Cho phép kẻ tấn công xâm nhập hệ thống, ăn cắp dữ liệu, kiểm soát hệ thống.
- Nghiên cứu phát hiện malware PE trên Windows: Sử dụng kỹ thuật học máy và học sâu như LSTM và XGBoost.
- Kết quả mong đợi: So sánh hiệu suất của LSTM và XGBoost trong việc phát hiện mã độc, góp phần bảo vệ hệ thống.

Giới thiệu

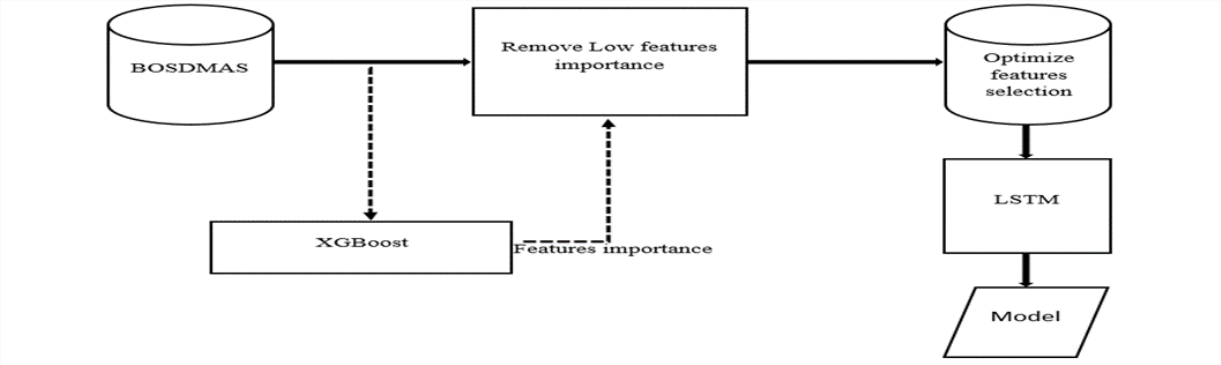
- Sự phát triển nhanh chóng của công nghệ thông tin dẫn đến sự gia tăng các mối đe dọa an ninh mạng.
- Malware là một loại phần mềm độc hại được thiết kế để gây hại cho hệ thống máy tính.
- Malware PE là loại malware được lưu trữ dưới dạng tệp PE (Portable Executable) trên Windows.
- Việc phát hiện malware PE là rất quan trọng để bảo vệ hệ thống máy tính khỏi các cuộc tấn công mạng.
- Học máy và học sâu là những lĩnh vực trí tuệ nhân tạo có khả năng học hỏi từ dữ liệu và đưa ra dự đoán chính xác.

Mục tiêu

- Nghiên cứu về mã độc PE và đặc trưng của nó.
- Sử dụng máy học và học sâu để rút trích đặc trưng từ mã độc PE.
- Phân loại mã độc PE bằng LSTM và XGBoost.

Nội dung và Phương pháp

- Nghiên cứu về mã độc PE và các đặc trưng:
  - Cách thức hoạt động và đặc trưng của mã độc PE.
  - Tìm hiểu các công trình khoa học đã công bố.



Nội dung và Phương pháp

- Xây dựng và đánh giá mô hình phát hiện mã độc PE:
  - Xây dựng mô hình bằng LSTM và XGBoost: Thu thập, chuẩn bị dữ liệu, tiền xử lý và trích xuất đặc trưng.
  - Huấn luyện và đánh giá mô hình: Đánh giá trên dữ liệu kiểm tra độc lập.
  - Điều chỉnh và cải tiến mô hình: Điều chỉnh siêu tham số và áp dụng các kỹ thuật bổ sung.
  - Đánh giá trên cơ sở dữ liệu BODMAS: Sử dụng dữ liệu BODMAS để kiểm tra và so sánh hiệu suất.

Nội dung và Phương pháp

- Sử dụng XGBoost để chọn lọc đặc trưng:
  - XGBoost chọn lọc đặc trưng thông qua "feature importance".
- Sử dụng LSTM để phân loại mã độc PE:
  - LSTM học các mẫu và mối quan hệ phức tạp giữa các chuỗi dữ liệu.
- Kết hợp XGBoost và LSTM:
  - Tận dụng khả năng xử lý dữ liệu chuỗi của LSTM và khả năng chọn lọc đặc trưng của XGBoost.

Kết quả dự kiến

- Xây dựng mô hình phát hiện mã độc PE: Bằng LSTM và XGBoost.
- Đánh giá trên cơ sở dữ liệu BODMAS: Kết quả thực nghiệm cao và thời gian xử lý tốt hơn.
- So sánh mô hình: Phương pháp đề xuất có khả năng nhận diện mã độc PE tốt và thực hiện nhanh.

Tài liệu tham khảo

[1] T. Rezaei and A. Hamze, "An Efficient Approach For Malware Detection Using PE Header Specifications," 2020 6th International Conference on Web Research (ICWR), Tehran, Iran, 2020, pp. 234-239, doi: 10.1109/ICWR49608.2020.9122312.

[2] C. Galen and R. Steele, "Evaluating Performance Maintenance and Deterioration Over Time of Machine Learning-based Malware Detection Models on the EMBER PE Dataset," 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), Paris, France, 2020, pp. 1-7, doi: 10.1109/SNAMS52053.2020.9336538.

[3] Muhamad Malik Matin and B. Rahardjo, "A Framework for Collecting and Analysis PE Malware Using Modern Honey Network (MHN)," 2020 8th International Conference on Cyber and IT Service Management (CITSM), Pangkal, Indonesia, 2020, pp. 1-5, doi: 10.1109/CITSM50537.2020.9268810.

[4] M. Kim, "Research on Malware Detection System Using Artificial Intelligence," 2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science (BCD), Danang, Vietnam, 2022, pp. 211-213, doi: 10.1109/BCD54882.2022.9900792.

[5] P. Singh, S. K. Borgohain and J. Kumar, "Performance Enhancement of SVM-based ML Malware Detection Model Using Data Preprocessing," 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), Patna, India, 2022, pp. 1-4, doi: 10.1109/ICEFEET51821.2022.9848192.