PHÂN LOẠI MÃ ĐỘC PE SỬ DỤNG MACHINE LEARNING

Lê Xuân Hiếu -230202026

Tóm tắt

- Lớp: CS2205.MAR2024
- Link Github: https://github.com/hieulx/CS2205.MAR2024/
- Link YouTube video: https://youtu.be/GgQTKuZkY8Y
- Họ và Tên:Lê Xuân Hiếu

Giới thiệu

- Phần mềm antivirus truyền thống sử dụng phương pháp signatures để phát hiện mã độc, nhận diện mã độc dựa trên một cơ sở dữ liệu lớn đã được phân tích từ các mã độc được phát hiện trước đó

- Sự phát triển của các cuộc tấn công hiện đại và tinh vi, phần mềm antivirus truyền thống đang khó khăn trong việc phát hiện các malware mới.

=> Việc nghiên cứu sử dụng các kỹ thuật tiên tiến để chọn ra các đặc trưng có liên quan, cải thiện độ chính xác trong việc phát hiện phần mềm độc hại, đồng thời giảm kích thước dữ liệu đã và đang được nghiên cứu và cải thiện.

=> Tôi sẽ sử dụng các thuật toán học máy phân loại khác nhau để phát hiện malware.

Giới thiệu (tiếp)

- Input: tập tin PE malware và benign (lành tính) .

  Sử dụng dataset BODMAS và mô hình học máy: K-Nearest Neighbor, Decision Tree, Logistic Regression và Ensemble (kết hợp các mô hình trên).

- Output: phân loại được mẫu đưa vào là malware hay benign, hiệu suất của các mô hình học máy sử dụng trong phát hiện và phân loại malware

Mục tiêu

- Phát hiện được tập tin PE đưa vào là malware hay benign
- Phân loại tập tin malware thuộc họ nào của malware
- Đánh giá hiệu suất phân loại malware tốt nhất trong các mô hình học máy Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Tree và Ensemble

Nội dung và Phương pháp

❖ Nội dung:

- Các công trình liên quan và hạn chế
- Sử dụng bộ dataset BODMAS để huấn luyện và kiểm thử .
- Tiền xử lử các dữ liệu trong bộ dataset, trích xuất các đặc trưng và đánh nhãn . Tạo ra file metadata (timestamp, family) để tăng hiệu quả đánh giá
- Chia các dữ liệu để huấn luyện và kiểm thử (80%-20%) cho các mô hình LR, KNN, Decision Tree và Ensemble (kết hợp các mô hình trên)
- Đánh giá hiệu suất của các mô hình khi có và không sử dụng metadata, trong cả phân loại nhị phân và phân loại đa nhãn

Nội dung và Phương pháp (tiếp)

❖ Phương pháp thực hiện:

- Tìm hiểu các bài báo có nội dung liên quan tới malware classification sử dụng machine learning, hỏi giảng viên hướng dẫn
- Download bộ dataset BODMAS [1] của tác giả Yang Limin[1].
- Trích xuất đặc trưng dựa trên dự án LIEF [2], kỹ thuật trích xuất giống với dự án Ember [3] (ByteHistogram, ByteEntropyHistogram, StringExtractor, GeneralFileInfo,HeaderFileInfo, SectionInfo, ImportsInfo, ExportsInfo, DataDirectories) , đưa ra dưới dạng vector.

- Tệp đưa vào bodmas.npz chưa hai phần, đặc trưng và nhãn. Tạo thêm file metatada.csv (timestamp, family) để đưa vào huấn luyện và kiểm thử các mô hình
- Chọn tiêu chí đánh giá hiệu suất mô hình Accuracy score, Precision score, Recall score, sử dụng hàm đánh giá của thư viện Sklearn
- Xuất kết quả ra dạng bảng biểu và tiến hành so sánh khi có và không sử dụng metadata, trong trường hợp phân loại nhị phân và phân loại đa nhãn

Kết quả dự kiến

- Áp dụng thành công mô hình LR, KNN, Decision Tree, Ensemble trong việc phân loại mã độc.
- Bản biểu đồ tổng quan mô hình thực nghiệm.
- Xuất được bảng biểu hiển thị thông tin hiệu suất mô hình qua các tiêu chí đánh giá, trong phân loại nhị phân và phân loại đa nhãn (có và không sử dụng metadata)

Tài liệu tham khảo

- [1]. Y. Limin, A. Ciptadi, I. Laziuk, A. Ahmadzadeh and G. Wang, “BODMAS,” 2019. [Online]. Available: https://liminyang.web.illinois.edu/slides/DLS21_BODMAS_LiminYang.pdf.
- [2] R. Thomas, “LIEF - Library to Instrument Executable Formats,” April 2017. [Online]. Available: https://lief.re/.
- [3] H. Anderson and P. Roth, “EMBER: An Open Dataset for Training Static PE Malware Machine Learning Models,” April 2018. [Online]. Available: https://github.com/elastic/ember/blob/master/ember/features.py.