


THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):
(<https://youtu.be/vCcF62oRfB4>)
- Link slides (dạng .pdf đặt trên Github):
(<https://github.com/dddecemberrr/CS2205.MAR2024>)
- 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

<ul style="list-style-type: none">• Họ và Tên: Phạm Trường Chinh• MSSV: 230202033 	<ul style="list-style-type: none">• Lớp: CS2205.MAR2024• Tự đánh giá (điểm tổng kết môn): 7.5/10• Số buổi vắng: 0• Số câu hỏi QT cá nhân:• Link Github: https://github.com/dddecemberrr/CS2205.MAR2024/
---	---

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI: ỨNG DỤNG TẠO CÁC MẪU MÃ ĐỘC ĐỐI NGHỊCH TRÊN MÔI TRƯỜNG WINDOWS SỬ DỤNG GENERATIVE ADVERSARIAL NETWORKS

TÊN ĐỀ TÀI TIẾNG ANH: GENERATING ADVERSARIAL MALWARE EXAMPLES APPLICATION IN WINDOWS ENVIRONMENT USING GENERATIVE ADVERSARIAL NETWORKS

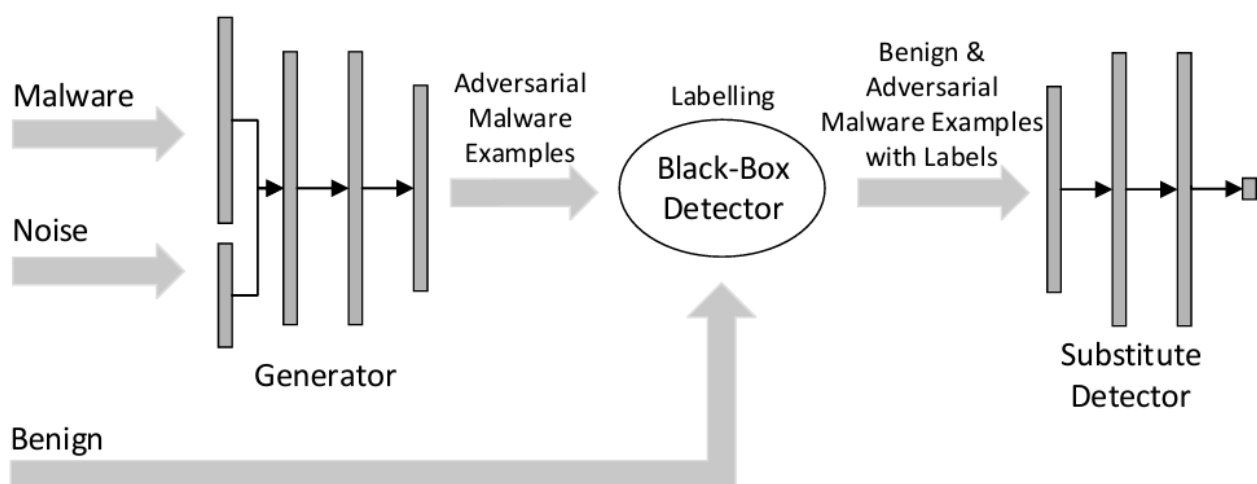
TÓM TẮT

Trong những năm gần đây, việc sử dụng ML (Machine Learning) và DL (Deep Learning) để phát hiện, phân loại các phần mềm độc hại đang càng ngày được ưa chuộng, ví dụ như Kaspersky, Bitdefender, McAfee Endpoint Security, và tỏ ra có hiệu quả trong việc chống lại các loại mã độc. Tuy nhiên, các nghiên cứu gần đây chứng minh rằng các mô hình ML/DL hiện tại vốn dễ bị tổn thương trước các cuộc tấn công đối nghịch dưới dạng các mẫu đối nghịch được tạo ra bằng cách gây nhiễu mà thêm các đầu vào hợp lệ để gây nhầm lẫn cho các mô hình ML/DL [1]. Các hệ thống phát hiện mã độc hiện tại thường sử dụng mô hình ML/DL để có thể nâng cao hiệu suất và khả năng phát hiện mã độc của mình. Tuy nhiên, các mã độc sử dụng tấn công đối nghịch đã được thiết kế để làm cho các mô hình máy học hoạt động không chính xác bằng cách thêm vào hoặc thay đổi một số thông tin không rõ ràng trong dữ liệu đầu vào, dẫn tới các mô hình ML/DL đưa ra dự đoán hoặc nhận diện sai lầm thành các phần mềm không độc hại, nhưng lại hoạt động giống hệt phần mềm độc hại gốc. Vì vậy việc phát triển một ứng dụng có thể tạo ra các mẫu mã độc đối nghịch là cần thiết để huấn luyện các mô hình máy học thích nghi với các cuộc tấn công như vậy. Ở đây chúng tôi tập trung vào phần mềm độc hại với định dạng tệp thực thi di động (PE) trong hệ điều hành Windows bởi người dùng Windows đang chiếm 70 - 80% thị phần, vì vậy, việc phòng chống mã độc trên môi trường Windows cần được đặt lên hàng đầu trong bối cảnh hiện nay.

GIỚI THIỆU (Tối đa 1 trang A4)

Để có thể tạo ra đủ nhiều các mẫu mã độc đối nghịch phục vụ cho việc đào tạo máy học là một việc rất tốn công sức và thời gian, chưa kể các mẫu này cần phải đảm bảo sự khác biệt, thay đổi nhưng vẫn cần đảm bảo các tính năng vốn có của mã độc.

Trong đề tài này, chúng tôi nghiên cứu thuật toán MalGAN [3] (một thuật toán dựa theo GAN [2]) để tạo ra các mẫu mã độc đối nghịch phục vụ cho việc huấn luyện các mô hình máy học phát hiện các phần mềm độc hại. Trong MalGAN [3], một generator được sử dụng để chuyển đổi các mẫu mã độc thành các mẫu đối nghịch bằng cách cộng thêm một noise vector, sau đó mẫu đối nghịch sẽ được đánh giá bằng Black-box Detector và dán nhãn là mã độc hay không.



Hình 1: Mô hình thuật toán MalGAN

Input:

- Tập dữ liệu mã độc thu thập từ các nguồn như: MDR (Malware Dataset Repository), Kaggle, VirusShare,...
- Các mẫu chương trình lành tính trên Windows.

Output: Các mẫu đối nghịch của các mã độc trong tập dữ liệu.

MỤC TIÊU

- Nghiên cứu thuật toán MalGAN [3] hiện có và áp dụng vào việc tạo ra các mẫu mã độc đối nghịch của các mẫu mã độc thu thập được.
- Xây dựng một database lớn các mẫu mã độc đối nghịch có định dạng PE.
- Xây dựng ứng dụng tạo mã độc trên hệ điều hành Windows

PHẠM VI

- Các mẫu mã độc được sử dụng để tạo mẫu đối nghịch là các mẫu mã độc có ảnh hưởng tới hệ điều hành Windows với định dạng tệp thực thi di động (PE).

NỘI DUNG

- Nghiên cứu thuật toán MalGAN [3] để tạo ra các mẫu mã độc đối nghịch.
- Tự xây dựng bộ dữ liệu các mã độc ảnh hưởng tới môi trường Windows để làm input cho thuật toán.
- Huấn luyện mô hình MalGAN [3] đề cập ở trên sử dụng bộ dữ liệu đã thu thập để tạo ra các mẫu mã độc đối nghịch đạt yêu cầu.
- Xây dựng ứng dụng minh họa

PHƯƠNG PHÁP

- Thu thập các mẫu mã độc trên các nguồn như MDR (Malware Dataset Repository), Kaggle, VirusShare,... để làm dataset.
- Thu thập các mẫu chương trình lành tính từ các nguồn.
- Nghiên cứu phương pháp tạo mẫu mã độc đối nghịch trong thuật toán MalGAN [3].
- Nghiên cứu cơ chế phát hiện của Black-box detector trong MalGAN [3].
- Huấn luyện thuật toán MalGAN [3] chạy trên bộ dữ liệu đã thu thập, so sánh và đánh giá dựa trên dán nhãn đầu ra.
- Xây dựng chương trình trên hệ điều hành Windows để người dùng có thể sử dụng tạo ra các mẫu mã độc.

KẾT QUẢ MONG ĐỢI

- Báo cáo phương pháp và kỹ thuật của thuật toán MalGAN [3] được sử dụng trong bài toán tạo các mẫu mã độc đối nghịch. Kết quả thực nghiệm và đánh giá thuật toán.
- Tập dữ liệu gồm các mẫu mã độc đối nghịch đã được tạo ra.
- Chương trình tạo ra các mẫu mã độc đối nghịch chạy trên hệ điều hành Windows.

TÀI LIỆU THAM KHẢO

- [1]. Nicholas Carlini, David Wagner: Towards Evaluating the Robustness of Neural Networks. IEEE Symposium on Security and Privacy, 2017.
- [2]. Ian Goodfellow, Jean Pouget Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio: Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.
- [3]. Weiwei Hu, Ying Tan: Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN. arXiv preprint arXiv:1702.05983v1, 2017