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

Tác giả: Phạm Trường Chinh

Trường Đại học Công nghệ Thông tin-Đại học Quốc gia

What ?

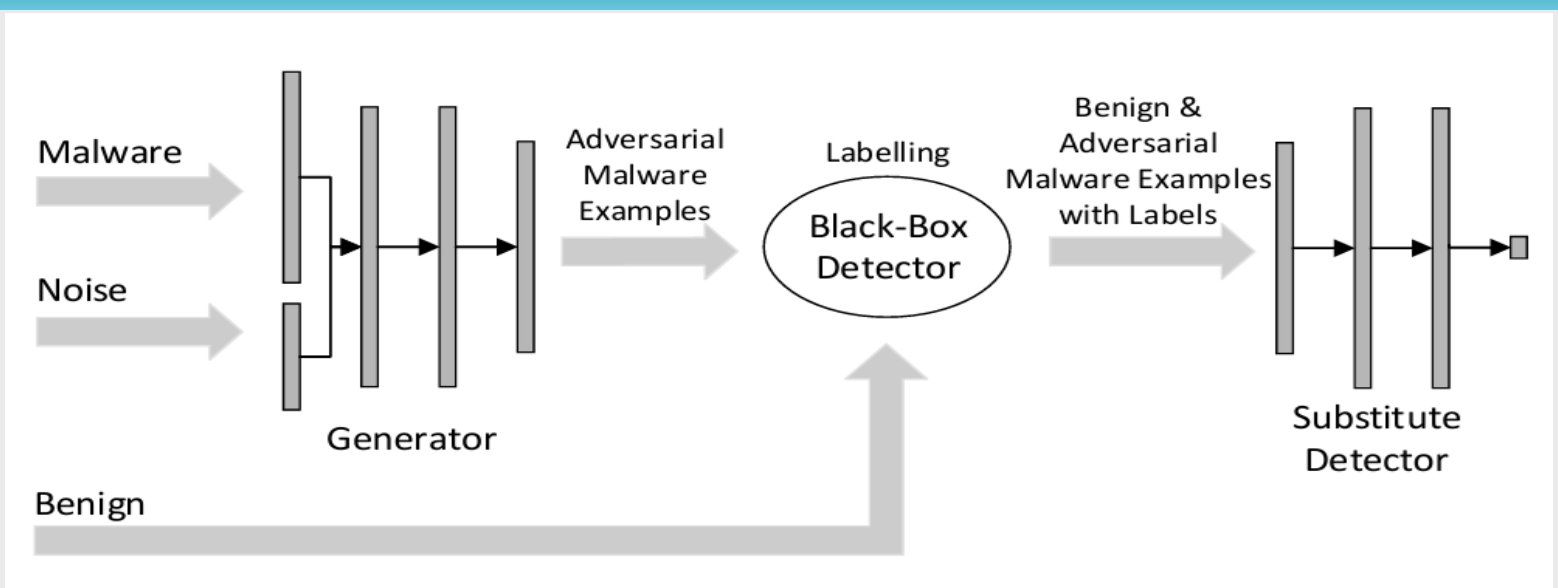
Chúng tôi nghiên cứu thuật toán MalGAN với mục đích:

- Nghiên cứu một phương pháp để tạo ra các mẫu mã độc đối nghịch phục vụ huấn luyện các mô hình ML/DL.
- Xây dựng một database lớn các mẫu mã độc đối nghịch có định dạng
- Xây dựng ứng dụng tạo các mẫu mã độc đối nghịch trên Windows

Why ?

- Việc sử dụng ML (Machine Learning) và DL (Deep Learning) để phát hiện, phân loại các phần mềm độc hại đang càng ngày phổ biến vì tính hiệu quả của nó. 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.
- Chúng tôi tập trung vào mã độc định dạng PE trong hệ điều hành Windows bởi người dùng Windows đang chiếm 70 - 80% thị phần.

Overview



Description

1. Tấn công đối nghịch

- Tấn công đối nghịch là việc các phần mềm độc hại được sửa đổi bằng cách thêm vào các vector gây nhiễu, từ đó khiến các mô hình ML/DL nhầm lẫn là các phần mềm sạch
- Các phần mềm độc hại được chỉnh sửa nhưng vẫn đảm bảo các tính năng vốn có

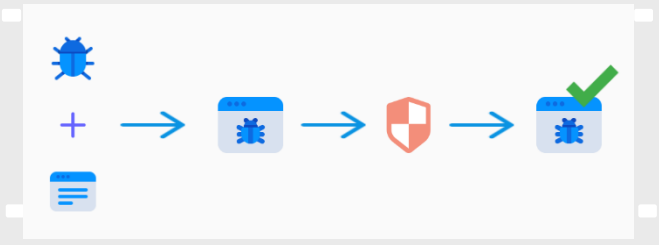


Figure 1. Cơ chế tấn công đối nghịch

2. Thuật toán MalGAN

- Là một thuật toán dựa theo thuật toán GAN (Generative Adversarial Networks)
- Thuật toán MalGAN sử dụng một generator với đầu vào là phần mềm độc hại và các vector gây nhiễu. Generator này tạo ra các mẫu đối nghịch của phần mềm độc hại. Mẫu này sẽ được đưa qua Black-Box detector cùng với các mẫu phần mềm sạch để đánh giá và dán nhãn. Substitute Detector sẽ phát hiện và phân loại dựa vào nhãn dán.

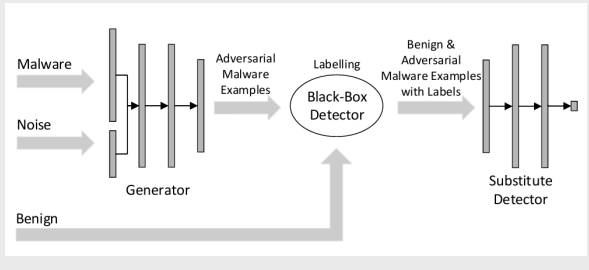


Figure 2. Thuật toán MalGAN

3. Ứng dụng tạo mã độc

- Dựa vào thuật toán MalGAN đã nghiên cứu, chúng tôi sẽ tạo ra một ứng dụng dựa trên hệ điều hành WIndows để tạo ra các mẫu mã độc đối nghịch phục vụ cho việc huấn luyện các mô hình ML/DL.

- Ứng dụng có khả năng tùy chọn cho đầu vào như tùy chọn loại phần mềm độc hại, tùy chọn vector gây nhiễu,... để xác định phần mềm độc hại đầu ra.

NII

Phạm Trường Chinh – Trường Đại học Công nghệ Thông tin-Đại học Quốc gia

TEL : ******* Email : ***********