TÌM HIỂU PHƯƠNG PHÁP PHÁT SINH MÃ ĐỘC ANDROID BẰNG MẠNG SINH ĐỐI KHÁNG

Đoàn Minh Trung - 220202016

Tóm tắt

- Lớp: CS2205.CH1702
- Link Github: https://github.com/cometofruition/CS2205.CH1702.git
- Link YouTube video: https://youtu.be/kbm-rwZyT5A
- Ảnh + Họ và Tên:

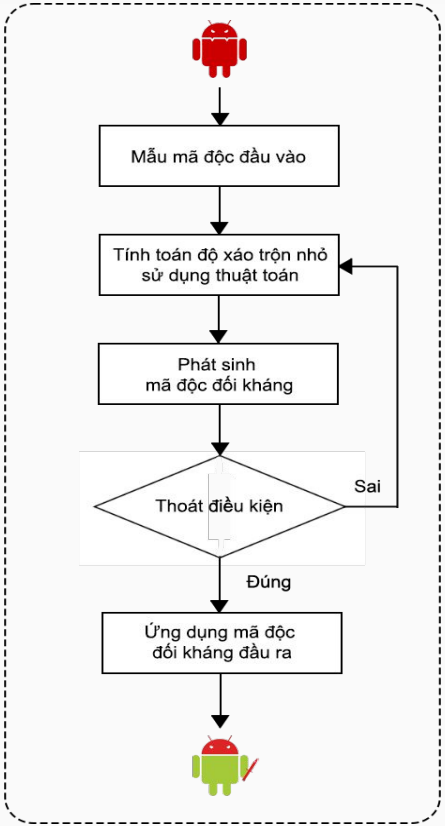
Đoàn Minh Trung

Giới thiệu

- Việc xây dựng các hệ thống phát hiện mã độc Android dựa trên học máy (Machine Learning - ML) đã **thu hút sự quan tâm lớn** từ cộng đồng nghiên cứu trên thế giới.
- Các thuật toán học máy hiện nay đã có những bước tiến quan trọng trong việc phát hiện mã độc. Tuy nhiên, vẫn dễ bị ảnh hưởng bởi **các cuộc tấn công đối kháng (Adversarial Example - AE)**.
- Phần lớn các nghiên cứu chỉ tập trung vào khả năng vượt mặt của AE mà không chú trọng đến việc **duy trì chức năng**, **khả năng thực thi** và **hành vi ban đầu** của chúng.

Giới thiệu

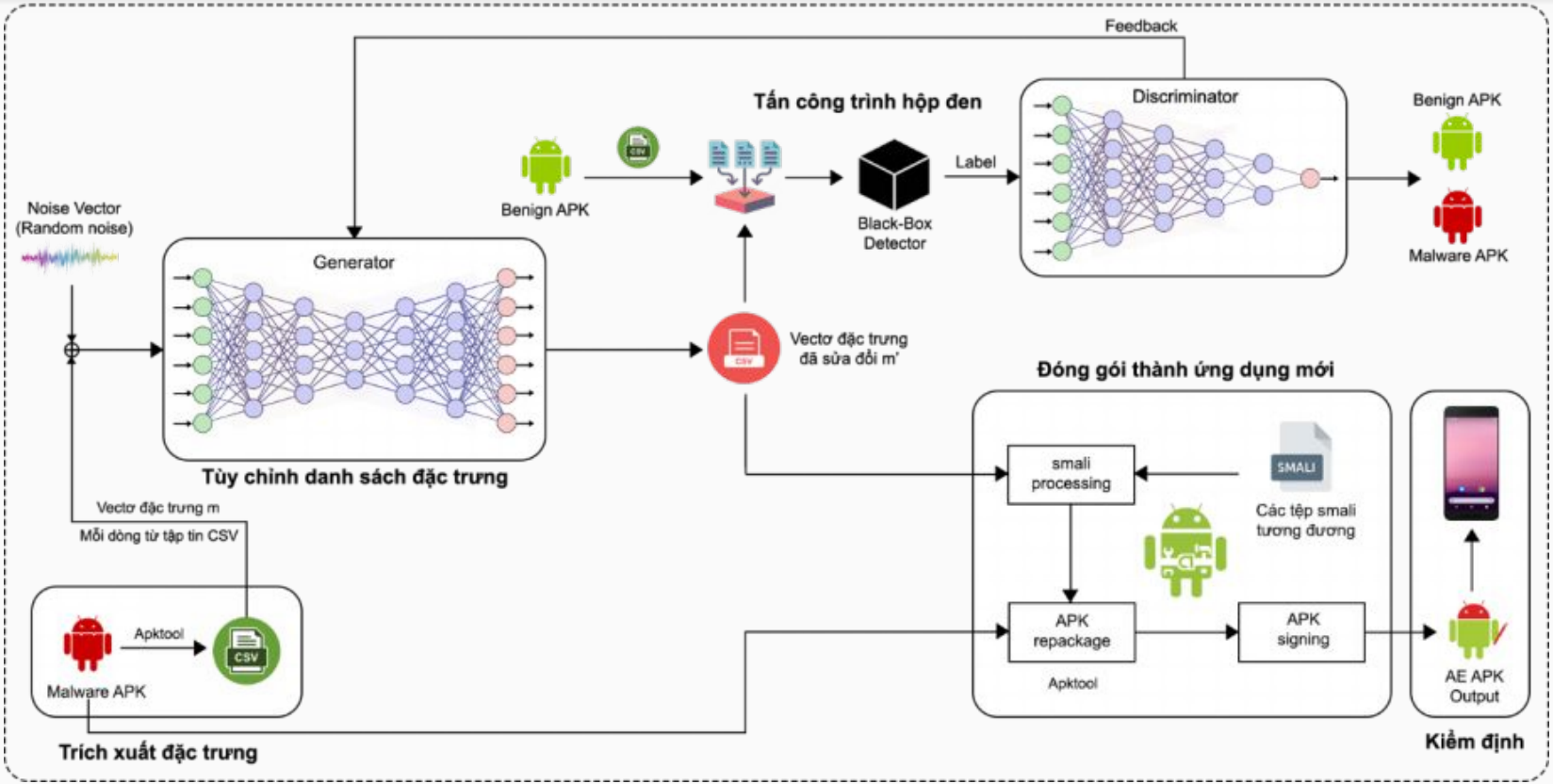
- Hệ thống phát sinh mã độc Android sử dụng Mạng sinh đối kháng:
  - Đầu vào: Ứng dụng mã độc Android gốc.
  - Đầu ra: Ứng dụng mã độc đối kháng có khả năng vượt mặt các mô hình phát hiện và vẫn duy trì được các chức năng ban đầu.



Mục tiêu

- Tiến hành xây dựng một hệ thống phát sinh mã độc đối kháng, có **khả năng gây nhầm lẫn** và khiến cho các mô hình phát hiện mã độc **phân loại các mẫu mã độc thành lành tính**.
- Các mẫu **có tỉ lệ đánh lừa thành công cao** sẽ được hệ thống này tự động đóng gói lại thành các ứng dụng mã độc mới và được đem đi kiểm định xem có còn duy trì được các chức năng, khả năng thực thi và hành vi ban đầu không.
- Tăng số lượng các ứng dụng mã độc đối kháng **duy trì được các chức năng**, **khả năng thực thi** và **hành vi ban đầu**, tiến hành xây dựng mô hình phát sinh mới dựa trên các đặc trưng có thể thay đổi được và giữ lại các đặc trưng không thể thay đổi để bảo toàn các hành vi độc hại của chúng.

Phương pháp


Feedback
Tấn công trình hộp đen
Discriminator
Benign APK
Label
Black-Box Detector
Benign APK
Malware APK
Noise Vector (Random noise)
Generator
Vecto đặc trưng đã sửa đổi m'
Tùy chỉnh danh sách đặc trưng
Đóng gói thành ứng dụng mới
smali processing
SMALI
Các tệp smali tương đương
APK repackage
APK signing
Apktool
AE APK Output
Vecto đặc trưng m
Mỗi dòng từ tập tin CSV
Apktool
Malware APK
Trích xuất đặc trưng
Kiểm định

Kết quả dự kiến

- Hệ thống phát sinh mã độc đối kháng dựa trên học máy đạt tỉ lệ tấn công thành công vào các mô hình phát hiện.
- Dự kiến tỉ lệ thành công các mẫu mã độc đối kháng duy trì được chức năng vào khoảng > 50%.

Tài liệu tham khảo

[1]. Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, ‘Explaining and harnessing adversarial examples’, in International Conference on Learning Representations, (2015).

[2]. Xingxing Wei, Siyuan Liang, Ning Chen, and Xiaochun Cao, ‘Transferable adversarial attacks for image and video object detection’, in IJCAI, pp. 954–960, (2019).

[3]. Jianhua Wang, Xiaolin Chang, Yixiang Wang, Ricardo J Rodríguez, and Jianan Zhang. Lsgan-at: enhancing malware detector robustness against adversarial examples. Cybersecurity, 4:1–15, 2021.

[4]. Xiao Chen, Chaoran Li, Derui Wang, Sheng Wen, Jun Zhang, Surya Nepal, Yang Xiang, and Kui Ren. Android hiv: A study of repackaging malware for evading machine-learning detection. IEEE Transactions on Information Forensics and Security, 15:987–1001, 2020.

[5]. Hamid Bostani and Veelasha Moonsamy. Evadedroid: A practical evasion attack on machine learning for black-box android malware detection. arXiv preprint arXiv:2110.03301, 2021.