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

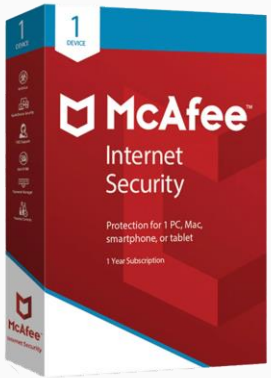
Phạm Trường Chinh - 230202033

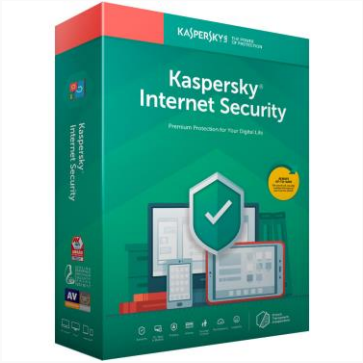
Tóm tắt

- Lớp: CS2205.MAR2024
- Link Github: *https://github.com/dddecemberrr/CS2205.MAR2024*
- Link YouTube video: *https://youtu.be/vCcF62oRfB4*
- Ảnh + Họ và Tên: Phạm Trường Chinh
- Tổng số slides không vượt quá 10

Tóm tắt

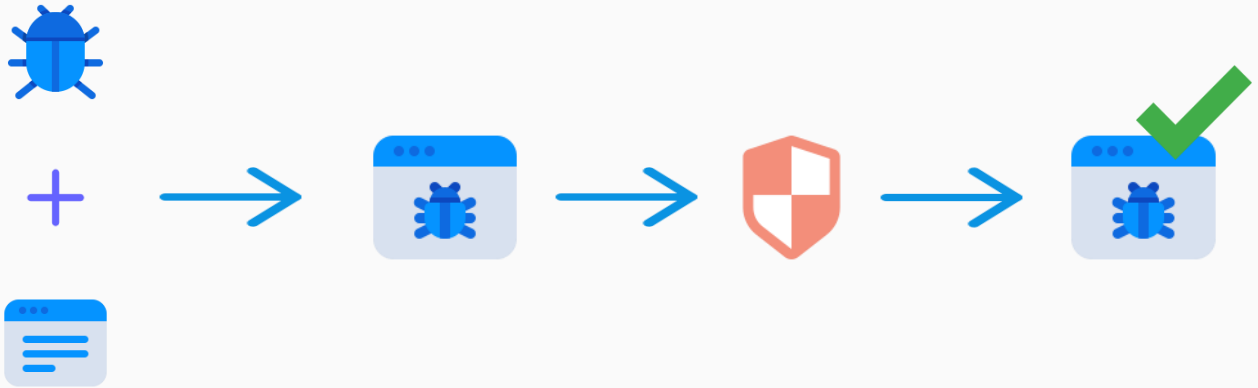
- Gia tăng sử dụng ML/DL trong các phần mềm anti-virus.
- Có hiệu quả trong việc phát hiện mã độc





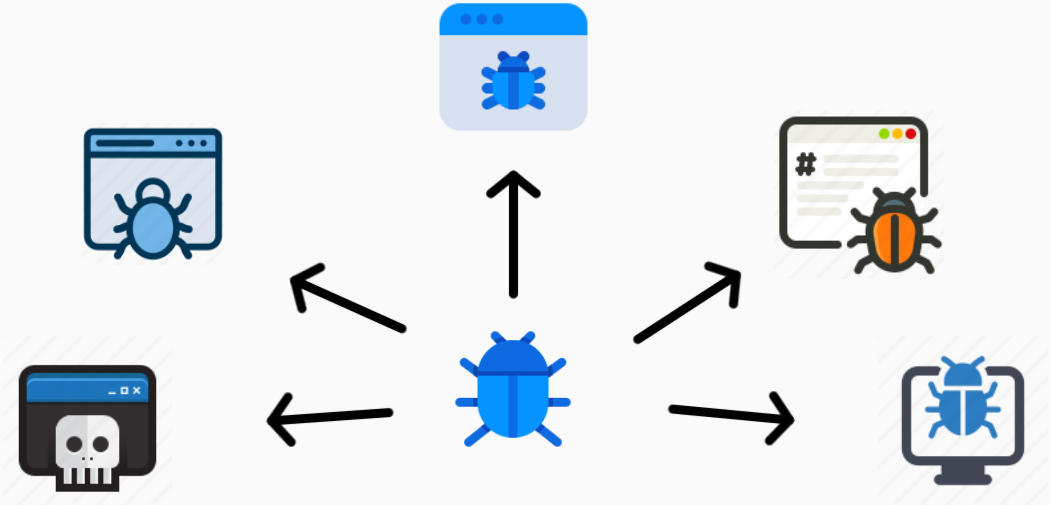
Tóm tắt

- Nghiên cứu chỉ ra các mô hình ML/DL dễ bị tổn thương trước các cuộc tấn công đối nghịch
- Tấn công đối nghịch là gì?



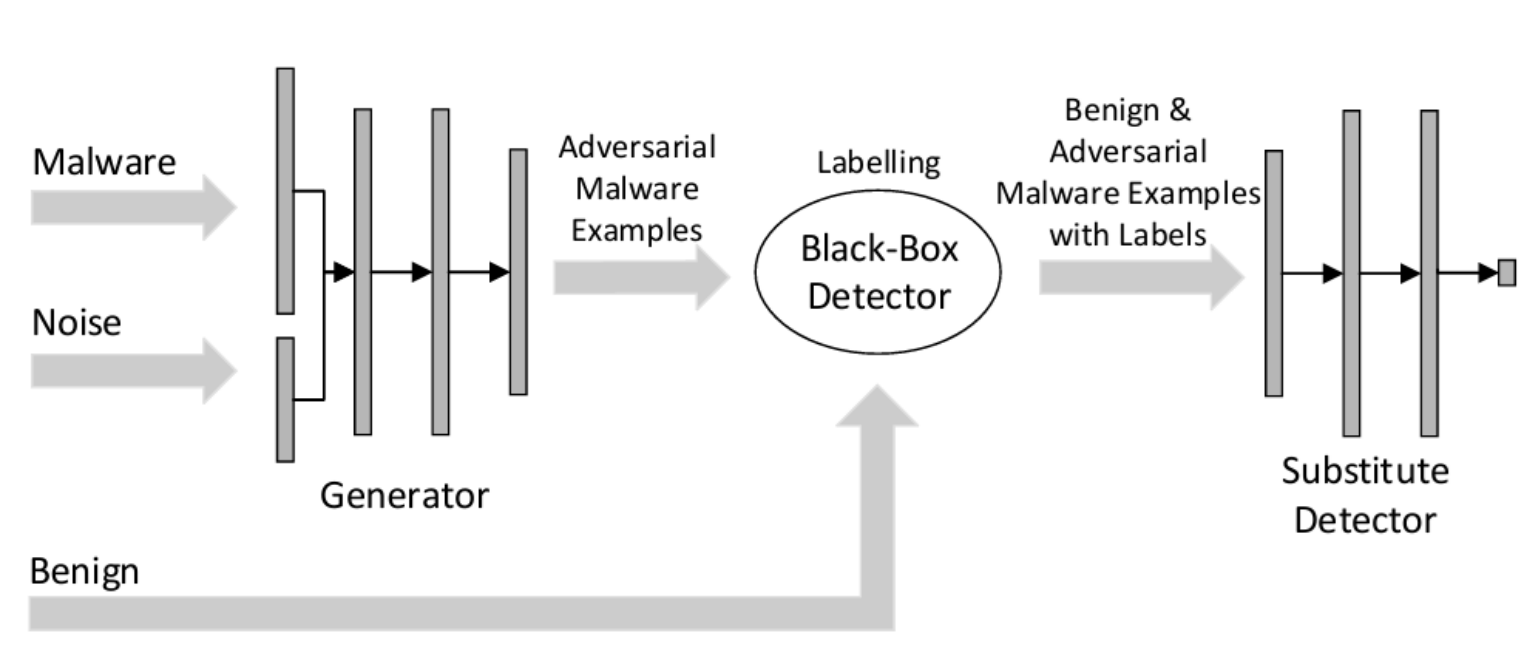
Giới thiệu

- Việc tạo ra số lượng lớn các mẫu mã độc khác nhau là việc khó khăn



Giới thiệu

- Thuật toán MalGAN



Giới thiệu

- Input:
  - Tệp dữ liệu mã độc với định dạng PE
  - Tệp các chương trình lành tính trên Windows
- Output:
  - Các mẫu đối nghịch của các mã độc trong tệp dữ liệu

Mục tiêu

- Nghiên cứu thuật toán MalGAN 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.

Nội dung và Phương pháp

- Thu thập các mẫu mã độc trên các nguồn như MDR (Malware Dataset Repository), Kaggle, VirusShare,..
- 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
- Nghiên cứu cơ chế phát hiện của Black-box detector trong MalGAN
- Huấn luyện thuật toán MalGAN
- Xây dựng chương trình trên hệ điều hành Windows

Kết quả dự kiến

- Báo cáo phương pháp và kỹ thuật của thuật toán MalGAN, kết quả thực nghiệm và đánh giá thuật toán
- Tệp 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

- Nicholas Carlini, David Wagner: Towards Evaluating the Robustness of Neural Networks. IEEE Symposium on Security and Privacy, 2017.
- 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.
- Weiwei Hu, Ying Tan: Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN. arXiv preprint arXiv:1702.05983v1, 2017