# PHÁT HIỆN MÃ ĐỘC ANDROID DÙNG FEDERATED LEARNING

Nguyễn Vương Thịnh - 220201025

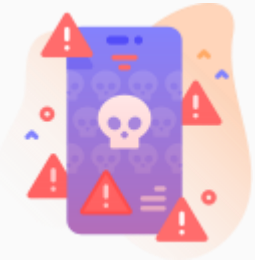
# Tóm tắt

- Lớp: CS2205.APR2023
- Link Github: https://github.com/vuongthinh209/CS2205.APR2023
- Link YouTube video: https://youtu.be/rEwJ8I_Qfes
- Họ và Tên: Nguyễn Vương Thịnh
- Tổng số slides: 8

# Giới thiệu

- Thiết bị di động Android trở thành mục tiêu của các cuộc tấn công mã độc (malware)
- Tác hại của malware

Trộm thông tin người dùng



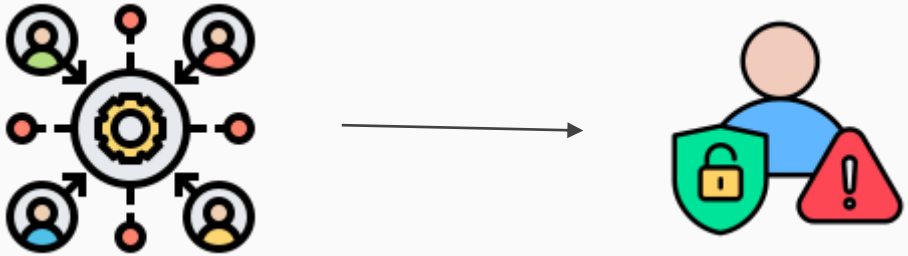
Hack điện thoại

Vô hiệu hóa điện thoại

# Giới thiệu

- Machine learning rất hiệu quả để phát hiện và ngăn chặn mã độc Android
- Nhưng, hầu hết hướng tiếp cận yêu cầu dữ liệu phải được gửi về một máy chủ trung tâm để xử lý **[1]**

➔ Gây ra mối lo ngại về bảo mật và thông tin người dùng

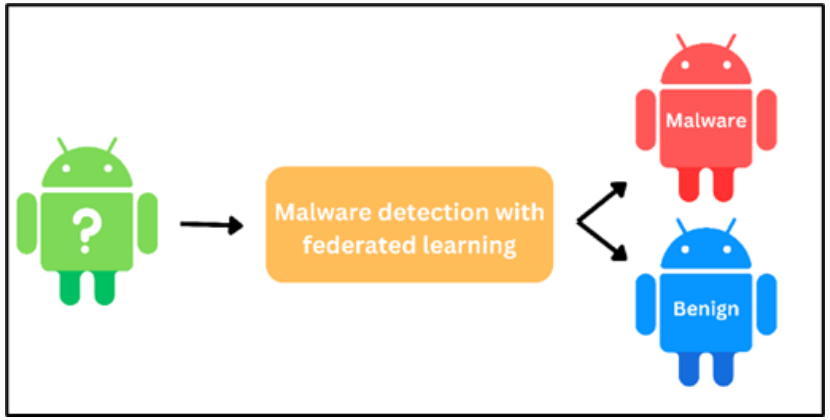


Hình 1. Hướng xử lý dữ liệu tập trung gây vấn đề bảo mật

# Giới thiệu

**Đề xuất ứng dụng học liên kết (Federated learning) để phát hiện Android malware**

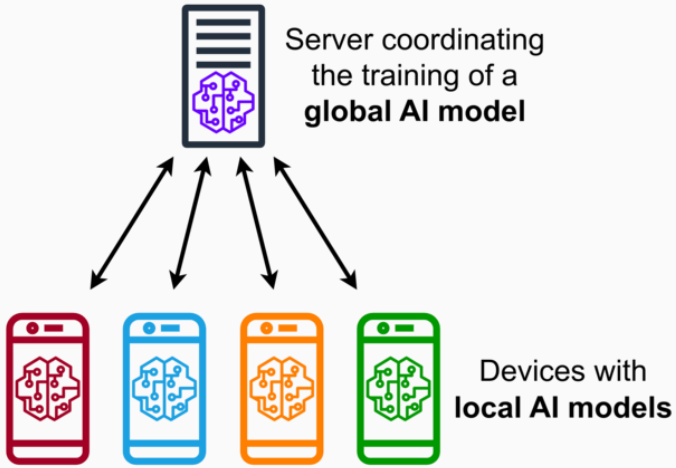
- **Phát biểu bài toán phát hiện mã độc Android bằng Federated learning**
  - Input: một phần mềm Android
  - Output: kiểm tra phần mềm có mã độc (malware) hay không có mã độc (benign)




Hình 2. Input và Output của phát hiện mã độc malware dùng federated learning

# Giới thiệu

**Federated learning** là kỹ thuật học máy phi tập trung. Nó sẽ huấn luyện model cục bộ ở thiết bị di động và gửi bản update của các model cục bộ đến model trung tâm ở máy chủ để tổng hợp.




Hình 3*. Ảnh mô tả federated learning

*Nguồn ảnh: https://en.wikipedia.org/wiki/Federated_learning

# Mục tiêu

Xây dựng mô hình phát hiện mã độc Android sử dụng kỹ thuật Federated learning

Giải quyết mối lo ngại liên quan vi phạm quyền riêng tư, bảo mật người dùng

# Nội dung và Phương pháp

1. Tìm hiểu các công trình về bài toán phát hiện mã độc Android bằng học máy
2. Tìm hiểu về các bài toán liên quan đến Federated learning
3. Thu thập bộ dữ liệu huấn luyện mô hình học máy
4. Thiết lập các loại mô hình học máy như: Decision tree **[2]**, Random forests **[3]**, SVM **[4]**, Naive Bayes **[5]**
5. Thiết lập Federated learning framework
6. Thực nghiệm và đánh giá kết quả
    a. Thử nghiệm từng loại mô hình học máy vào Federated learning framework
    b. Đánh giá bằng thang đo Accuracy, Recall, Precision, và F1-score

# Kết quả dự kiến

Một mô hình phát hiện mã độc trên hệ điều hành Android sử dụng kỹ thuật học liên kết (Federated learning)

Một bảng số liệu so sánh kết quả giữa các loại mô hình máy học sử dụng Federated learning để phát hiện mã độc Android

# Tài liệu tham khảo


1. Rasha Al-Huthaifi, Tianrui Li, Wei Huang, Jin Gu, Chongshou Li: Federated learning in smart cities: Privacy and security survey

1. Aqil Zulkifli, Isredza Rahmi A. Hamid, Wahidah Md Shah, Zubaile Abdullah: Android Malware Detection Based on Network Traffic Using Decision Tree Algorithm. SCDM 2018: 485-494

1. Mohammed S. Alam, Son Thanh Vuong: Random Forest Classification for Detecting Android Malware. GreenCom/iThings/CPScom 2013: 663-669

1. Tanuvir Singh, Fabio Di Troia, Corrado Aaron Visaggio, Thomas H. Austin, Mark Stamp: Support vector machines and malware detection. J. Comput. Virol. Hacking Tech. 12(4): 203-212 (2016)

1. Fengjun Shang, Yalin Li, Xiaolin Deng, Dexiang He: Android malware detection method based on naive Bayes and permission correlation algorithm. Clust. Comput. 21(1): 955-966 (2018)