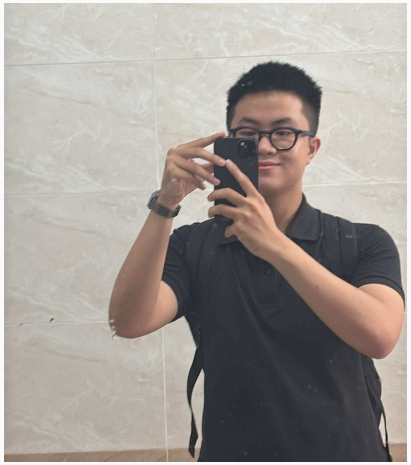
# ENHANCING ATTACK DETECTION ACCURACY ON NON-IID DATA IN FEDERATED LEARNING

Dương Việt Huy - 22520540
Lê Bình Nguyên - 22520969

# Tóm tắt

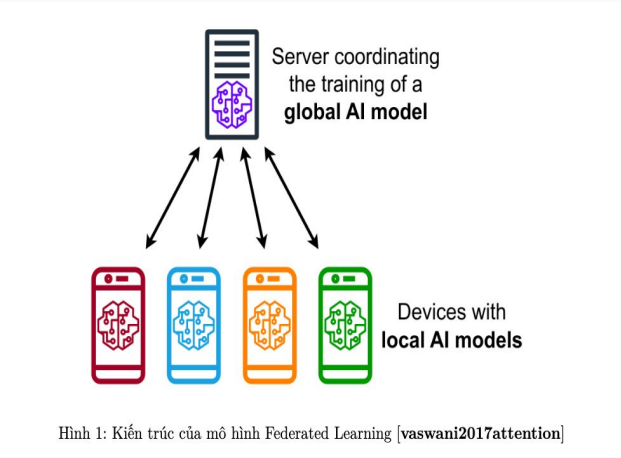
Dương Việt Huy
22520540

Lê Bình Nguyên
22520969


- Lớp: CS519.O21.KHTN
- Link Github của nhóm: https://github.com/lbngyn/CS519.O21.KHTN
- Link YouTube video: https://youtu.be/edcWCrDG5Yg

# Giới thiệu

- Mô hình học liên kết được dùng trong phát hiện các cuộc tấn công mạng.
- Thách thức chính là sự không đồng nhất trong dữ liệu (Non-IID).
- Có hai loại mất cân bằng lớp:
    - Mất cân bằng cục bộ (Local Non- IID).
    - Mất cân bằng toàn cầu (Global Non-IID).




Hình 1: Kiến trúc của mô hình Federated Learning [vaswani2017attention]

Trong nghiên cứu này, chúng tôi cung cấp một framework hoạt động trên Federated Learning, nhằm giải quyết Global Non-IID bằng cách sử dùng GAN

# Mục tiêu

- Xác định được mô hình học liên kết tối ưu cho việc phát hiện lỗ hổng bảo mật dựa trên các tính chất của dữ liệu không đồng nhất.
- Đạt được và thể hiện trực quan mức độ cải thiện đáng kể về thời gian xử lý và độ chính xác trong việc phát hiện lỗ hổng bảo mật so với các mô hình, phương pháp trước đó.
- Cung cấp báo cáo chi tiết về hiệu suất và khả năng áp dụng của mô hình, bao gồm khuyến nghị cho các tổ chức về việc triển khai mô hình để tăng cường bảo mật hệ thống

# Nội dung và Phương pháp

## Khảo Sát và Phân Tích Các Mô Hình Học Liên Kết Để Phát Hiện Lỗ Hổng Bảo Mật

**Phương pháp:**

Tìm hiểu và đánh giá sâu các mô hình học liên kết hiện có và cách thức họ giải quyết vấn đề mất cân bằng dữ liệu.

Ở đây chúng tôi chọn 2 kỹ thuật đó là:

- Data Label Distribution Strategies
- Multi-Model Learning

# Nội dung và Phương pháp

## Ứng dụng GAN cho dữ liệu dạng bảng để giải quyết mất cân bằng dữ liệu trong Federated Learning

**Phương pháp:**

Nghiên cứu và thử nghiệm các kỹ thuật mới như học tăng cường, học không giám sát, và các mô hình dựa trên GAN để xem xét hiệu suất của chúng trong việc cải thiện mô hình.

Các kĩ thuật với mô hình sinh có thể sử dụng :

- CTGAN ( mô hình sinh cho dữ liệu dạng bảng )
- Free Data training

Áp dụng các thuật toán tối ưu hóa mới để giảm thời gian huấn luyện và cải thiện độ chính xác của mô hình.

# Nội dung và Phương pháp

## Đánh Giá Hiệu Suất và Tính Khả Thi của Mô Hình Trong Môi Trường Thực Tế

**Phương pháp:**

- Thực hiện các thử nghiệm trong điều kiện dữ liệu lớn và phức tạp hơn, thực hiện đo đạc thời gian và yêu cầu phần cứng ở các máy khách.
- So sánh và phân tích kết quả thu được với kết quả từ các mô hình, phương pháp trước đó để đánh giá các cải tiến.

# Kết quả dự kiến

- Cung cấp một framework mới đạt được hiệu suất tốt hơn.
- Báo cáo Nghiên Cứu Chi Tiết.
- Phân tích so sánh hiệu quả của mô hình.
- Các tài liệu hướng dẫn triển khai mô hình.
- Khuyến nghị và hướng phát triển tiếp theo.

# Tài liệu tham khảo

[1] Sara Babakniya, Zalan Fabian, Chaoyang He, Mahdi Soltanolkotabi, Salman Avestimehr:
**A Data-Free Approach to Mitigate Catastrophic Forgetting in Federated Class Incremental Learning for Vision Tasks**. NeurIPS 2023
[2] Sijia Chen, Baochun Li:
**Towards Optimal Multi-Modal Federated Learning on Non-IID Data with Hierarchical Gradient Blending**. INFOCOM 2022: 1469-1478
[3]Lixu Wang, Shichao Xu, Xiao Wang, Qi Zhu:
**Addressing Class Imbalance in Federated Learning**. AAAI 2021: 10165-10173
[4] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, Kalyan Veeramachaneni:
**Modeling Tabular data using Conditional GAN**. NeurIPS 2019: 7333-7343
[5] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, Vikas Chandra:
**Federated Learning with Non-IID Dat**a. CoRR abs/1806.00582 (2018)
[6] Q Tan, S Wu, and Y Tao. **Privacy-Enhanced Federated Learning for Non-IID Data**. 2023. url: https://doi.org/10.3390/math11194123.
[7] W.-C. and Chung. FedISM: **Enhancing Data Imbalance via Shared Model in Federated Learning**. 2023. url: https://doi.org/10.3390/math11102385.