Tên học viên: Đinh Phúc Hưng

MSHV: 2370118

Báo cáo demo cho chủ đề thuyết trình Federated Learning

Các thư viện Python sử dụng:

A screen shot of a computer

Description automatically generated

1. Mô phỏng việc luyện mô hình bằng thư viện Tensorflow Federated

Mã nguồn cho việc mô phỏng này nằm ở đường link [này](https://github.com/dphung2000/Database-Security/blob/master/src/shared/test.ipynb).

Tập dữ liệu sử dụng là tập chữ viết tay MNIST được chia đều ra 3387 ID khác nhau, mô phỏng việc mỗi máy khách có tập dữ liệu không độc lập và không có phân phối giống nhau.

A yellow square with blue and yellow numbers

Description automatically generated

A yellow squares with numbers

Description automatically generated

A group of numbers in yellow squares

Description automatically generatedCác hình trên là tính trung bình các mẫu chữ viết tay của mỗi người dùng. Ở đây có thể thấy kiểu chữ của mỗi người dùng tuân thủ quy tắc không độc lập và không đồng nhất.

Khi luyện mô hình, em đề ra một số thông số cơ bản trong việc luyện mô hình:

Số lượng người dùng (tương ứng số tập dữ liệu rời rạc) tham gia huấn luyện: 300

Số epoch cho mỗi lần luyện ở mỗi người dùng: 5

Số điểm dữ liệu truyền vào mỗi epoch: 20

Số lần cập nhật mô hình trung tâm (tương ứng với số lần luyện mô hình): 10

Kết quả luyện:

* 1. sparse categorical accuracy: 0.34273288, loss: 2.042317
  2. sparse categorical accuracy: 0.36498317, loss: 1.9654884
  3. sparse categorical accuracy: 0.38597083, loss: 1.8934559
  4. sparse categorical accuracy: 0.4078844, loss: 1.8259594
  5. sparse categorical accuracy: 0.42821267, loss: 1.7626241
  6. sparse categorical accuracy: 0.44854096, loss: 1.7031678
  7. sparse categorical accuracy: 0.4679784, loss: 1.6474884
  8. sparse categorical accuracy: 0.48734567, loss: 1.5953704
  9. sparse categorical accuracy: 0.50597644, loss: 1.5465877
  10. sparse categorical accuracy: 0.5228255, loss: 1.5009215

1. Mô phỏng việc tấn công mô hình bằng việc đảo nhãn (label flipping attack)

Mã nguồn cho việc mô phỏng này nằm ở link [này](https://github.com/dphung2000/Database-Security/blob/master/src/)

A diagram of a server

Description automatically generated

Ở phần mô phỏng này em thiết lập hệ thống như hình. Hệ thống mô phỏng sẽ gồm 4 máy ảo ,1 máy có vai trò làm server, đưa mô hình tới các 3 máy khách, mỗi máy sẽ luyện mô hình dựa trên dữ liệu tại mỗi máy và gửi trọng số về máy chủ, máy chủ sẽ cập nhật mô hình mới và lặp lại điều này. Một máy khách có vai trò tấn công sẽ có nhãn dữ liệu bị đổi. Một số thông số cơ bản của lần mô phỏng này:

* Dữ liệu tại 3 máy khách là dữ liệu nhân tạo với 5 cột và 10000 hàng, trong đó có một cột là nhãn của dữ liệu.
* Máy khách sẽ luyện mỗi vòng từ 2-3 epoch, số lượng epoch được điều khiển bởi máy chủ. Vì dữ liệu là dữ liệu nhân tạo và hoàn toàn sạch nên độ chính xác hội tụ rất nhanh, để có thể thấy được tác động của việc tấn công, em lựa chọn epoch thấp.

Các bước thực hiện mô phỏng và kết quả:

1. Chạy lệnh docker build -t federated:latest . tại thư mục chính để tạo ảnh.
2. Chạy lệnh docker compose up -d để chạy hệ thống
3. Truy cập vào máy server bằng giao diện người dùng của docker và chạy lệnh python server/server.py
4. Mở một terminal ngoài máy chính và chạy lệnh curl localhost:5000/start để bắt đầu việc chạy hệ thống Federated Learning
5. Kiểm tra log của các máy client-1 và client-2, ta thấy được độ lỗi rất thấp và hội tụ nhanh từ epoch đầu tiên

A screenshot of a computer screen

Description automatically generated

1. Tắt chương trình tại server và thêm giá trị 'http://fl-client-3:5000' tại mảng clients ở dòng 28 trong file server/server.py
2. Thực hiện lại bước 3, 4 và 5. Ở bước 5 ta thấy độ chính xác của mô hình tại mỗi máy khách bị tác động, thấp hơn khi không có máy tấn công tham gia luyện mô hình

A screenshot of a computer screen

Description automatically generated

1. Tham khảo:
   1. [Federated Learning for Image Classification  |  TensorFlow Federated](https://www.tensorflow.org/federated/tutorials/federated_learning_for_image_classification)
   2. [TensorFlow 2 quickstart for beginners  |  TensorFlow Core](https://www.tensorflow.org/tutorials/quickstart/beginner)
   3. [PPML Series #2 - Federated Optimization Algorithms - FedSGD and FedAvg | Shreyansh Singh](https://shreyansh26.github.io/post/2021-12-18_federated_optimization_fedavg/)