ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

BÁO CÁO TỔNG KẾT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ SINH VIÊN NĂM 2021

Tên đề tài tiếng Việt:

Trình phát hiện tấn công dựa trên học cộng tác trong mạng khả lập trình

Tên đề tài tiếng Anh:

Federated learning-based intrusion detection in SDN-aware network

Khoa/ Bộ môn: **Mạng máy tính và truyền thông**

Thời gian thực hiện: **Tháng 3/2021 – tháng 6/2021**

Cán bộ hướng dẫn: **Ths. Phan Thế Duy**

Tham gia thực hiện

TT	Họ và tên, MSSV	Chịu trách nhiệm	Điện thoại	Email
1.	Nguyễn Hồng Hà, 17520419	Chủ nhiệm	0966503812	17520419@gm.uit.edu.vn
2.	Trần Văn Hùng, 17520554	Tham gia	0868337213	17520554@gm.uit.edu.vn

Thành phố Hồ Chí Minh – Tháng 03 /2022

ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

Ngày nhận hồ sơ	
Mã số đề tài	
(Do CQ quản lý ghi)	

BÁO CÁO TỔNG KẾT

Tên đề tài tiếng Việt:

Trình phát hiện tấn công dựa trên học cộng tác trong mạng khả lập trình

Tên đề tài tiếng Anh:

Federated learning-based intrusion detection in SDN-aware network

Ngày 18 tháng 03 năm 2022
Cán bộ hướng dẫn
(Họ tên và chữ ký)

Ngày 18 tháng 03 năm 2022
Sinh viên chủ nhiệm đề tài
(Họ tên và chữ ký)

THÔNG TIN KẾT QUẢ NGHIÊN CỨU

1. Thông tin chung:

- Tên đề tài: Trình phát hiện tấn công dựa trên học cộng tác trong mạng khả lập trình

- Mã số:

- Chủ nhiệm: Nguyễn Hồng Hà

- Thành viên tham gia: Trần Văn Hùng

- Cơ quan chủ trì: Trường Đại học Công nghệ Thông tin.

- Thời gian thực hiện: Tháng 3/2021 – tháng 6/2021

2. Mục tiêu:

- Tận dụng phương pháp Học Cộng Tác (Federated Learning) trên các hệ thống phát hiện xâm nhập IDS nhằm mục đích đảm bảo tính riêng tư của dữ liệu, và đảm bảo kết quả phát hiện những hành vi bất bình thường lẫn giữa những hành vi bình thường trong lưu lượng mạng khả lập trình (SDN).
- Nghiên cứu kỹ thuật tấn công nhiễm độc (Poisoning attack) lên IDS dùng phương pháp học cộng tác để đánh giá được những rủi ro ở các mô hình IDS học máy; từ đó tìm ra hướng phòng thủ để ngăn chặn mối đe dọa này.

3. Tính mới và sáng tạo:

- Đảm bảo tính riêng tư của dữ liệu tại các thiết bị người dùng mà không chia sẻ ra bên ngoài.
- Tránh tình trạng không đủ bộ nhớ vì quá nhiều dữ liệu để huấn luyện nếu tất cả dữ liệu để nằm tại một máy.
- Không cần nhiều diện tích và chi phí lắp đặt các thiết bị phần cứng trong hệ thống mạng.

4. Tóm tắt kết quả nghiên cứu:

4.1 Mô hình, ý tưởng thực hiện

- Trong đề tài này, mô hình IDS máy học, có tên gọi là FL-IDS, được huấn luyện cộng tác với bộ dữ liệu riêng của từng đối tượng tham gia. FL-IDS đưa ra dự đoán cho các mẫu lưu lượng mạng hỗ trợ việc phát hiện các cuộc xâm nhập, tấn công mạng; mà vẫn giữ được quyền riêng tư dữ liệu của các bên tham gia. Các IDS của từng hệ thống mạng hoặc tổ chức được ảo hóa và triển khai trên SDN. Mô hình triển khai ý tưởng được biểu diễn như hình 2.

Phương pháp huấn luyện mô hình IDS theo cách tiếp cận Federated learning bao gồm hai thành phần chính:

- Aggregation Server được tạo ra từ sự đồng thuận giữa những người đứng đầu quản lý SDN của mỗi tổ chức. Server tin cậy, thực hiện trao đổi các tham số của các client gửi lên, tổng hợp các tham số này thành tham số cục bộ. Server sau đó sẽ gửi tham số mới này lại cho các client.
- Training Client là một máy chủ đại diện cho mỗi mạng tham giao vào quá trình huấn luyện nội bộ mô hình học máy trên dữ liệu riêng của chính nó. Sau khi huấn luyện hoàn thành, client sẽ gửi các tham số mới lên server để server tiếp nhận và tổng hợp lại.

Qui trình huấn luyện FL:

- Bước 1: Tại máy chủ, tạo ra một mô hình chung trong đó kiến trúc mạng thần kinh được xây dựng, tại giai đoạn này số lượng layer, neuron, epoch … được xác định.
- Bước 2: Mô hình học máy được tải xuống bởi các thiết bị muốn sử dụng mô hình, cho dù có đóng góp hay không trong quá trình FL
- Bước 3: Các thiết bị được lựa chọn dữ liệu cục bộ riêng tư và sử dụng chúng trên thiết bị để nâng cao mô hình đang được nghiên cứu.
- Bước 4: Chỉ các siêu tham số của mô hình được chia sẻ với máy chủ trung tâm thay vì gửi dữ liệu nhạy cảm và xâm phạm quyền riêng tư của từng thiết bị.
- Bước 5: Máy chủ tổng hợp trọng số từ các mô hình ở các thiết bị khác nhau sau khi đã nhận được tất cả các bản cập nhật và tạo ra một mô hình cập nhật mới. Để tổng hợp, thuật toán Federated Averaging được sử dụng.
- Bước 6: Máy chủ đẩy trở lại các thông số mô hình đã cập nhật cho các thiết bị
- Bước 7: Mỗi thiết bị sử dụng các thông số mô hình được cập nhật và cải thiện chúng dựa trên dữ liệu mới được tạo ra từ chính thiết bị đó.
  Các bước 4,5,6 và 7 được lặp lại để liên tục học hỏi và cải thiện mô hình.

Differential Privacy (DP) để cải thiện quyền riêng tư của dữ liệu trong quá trình tổng hợp tham số của mô hình huấn luyện. Trong trường hợp này, nhiễu được thêm vào bước 4 mục đích là thêm nhiễu vào quá trình huấn luyện để tăng sự riêng tư cho dữ liệu cá nhân người dùng. DP sẽ đảm bảo dữ liệu của mỗi tác tử học máy không thể bị truy ngược lại nguồn gốc khi thực hiện huấn luyện.

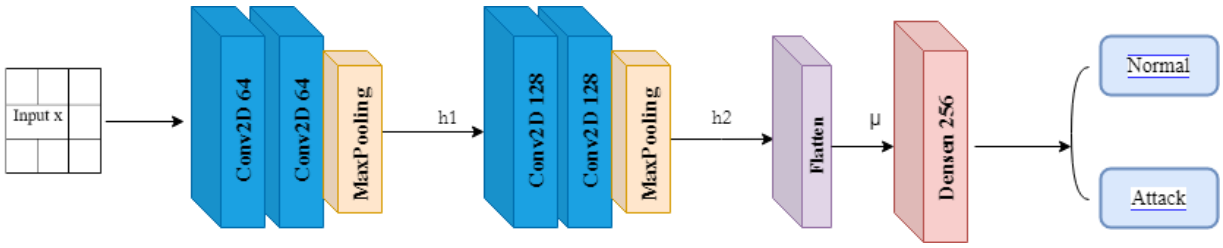
Các nhiệm vụ chức năng của các phần tử SDN:

- SDN Controller là bộ điều khiển và quản lý tập trung mô hình mạng. Triển khai các luật (Access Control List) để cho phép và ngăn chặn các lưu lượng mạng xác định đi vào hoặc ra các thiết bị mạng.
- OpenvSwitch là các Switch ảo giao tiếp với SDN Controller. SDN Controller điều hướng lưu lượng mạng từ Containernet sang IDS.
- IDS nằm ở các OpenvSwitch có chức năng tiếp nhận lưu lượng mạng từ các Host, cập nhật mô hình huấn luyện từ Training Client khi có cập nhật mới. IDS sẽ sử dụng mô hình mới cập nhật để phân loại các lưu mạng bình thường và các lưu lượng mạng độc hại.
- Các Host/devices/equipments đóng vai trò là các thiết bị vận hành trong hệ thống, phát sinh lưu lượng mạng.

Sau khi mô hình huấn luyện FL hoàn tất file Model đã được huấn luyện ở từng máy Training Client được lưu lại và gửi xuống vị trí của Security Gateway bằng thư viện Socket của python.

Các mô hình ML sử dụng trong FL-IDS

- Trong đề tài này chúng tôi sẽ sử dụng mô hình học máy VGG-16 để huấn luyện tại các IDS. Mô hình học máy được huấn luyện tại IDS chỉ là một nửa của mô hình VGG-16 vì dữ liệu đầu vào quá nhỏ so với kích thước hình ảnh quy định của bộ VGG-16. Mô hình VGG-16 được tinh chỉnh như hình 1



Hình 1. Mô hình VGG-16 được tinh chỉnh

So sánh hiệu năng của FL-IDS với các cách tiếp cận khác nhau

Trong đề tài này, chúng tôi sẽ thực hiện so sánh hiệu năng của FL-IDS với các cách tiếp cận khác nhau với mục đích đánh giá được FL-IDS có hiệu năng như thế nào đối với các cách tiếp cận khác. Những cách tiếp cận mà chúng tôi sẽ thực hiện trong khóa luận như sau:

Local model: Với model Vgg-16 đã tinh chỉnh lại, được sử dụng để đào tạo dữ liệu ngay trên chính thiết bị sinh dữ liệu.

Ideal model: Với model Vgg-16 đã tinh chỉnh lại, được sử dụng để đào tạo toàn bộ dữ liệu được các thiết bị sinh dữ liệu tại một thiết bị tập hợp.

FL (Federated Learning) model: Với model Vgg-16 đã tinh chỉnh lại, được triển khai theo phương pháp Federated learning với việc đào tạo các dữ liệu ngay tại thiết bị của mình rồi gửi tham số lên thiết bị có chức năng thu thập để tính giá trị trung bình, sau đó thiết bị đó có chức năng gửi tham số sau khi được đào tạo lại về với các thiết bị sinh dữ liệu.

DPFL (Differential Privacy Federated Learning) model: Với model Vgg-16 đã tinh chỉnh lại, được triển khai giống với FL model nhưng có áp dụng DP. Sử dụng thư việc Tensorflow Privacy để thực hiện DPFL.

Tiền xử lý tập dữ liệu

- Chúng tôi sử dụng tập dữ liệu CICDDOS2019 cho mô hình VGG-16 đã tinh chỉnh.
- Chúng tôi quyết định lấy 6 bộ dữ liệu con, tương ứng với 4 nhãn NetBIOS, MSSQL, Syn, BENIGN trong tập dữ liệu CICIDDOS2019, trong đó 3 bộ dữ liệu để đánh giá, 3 bộ dữ liệu dùng để huấn luyện, mỗi bộ dữ liệu tương ứng có 2 nhãn BENIGN cộng với nhãn của loại 1 tấn công DDOS. Chúng tôi chia đều tập dữ liệu cho các worker trong mô hình theo tỉ lệ 1:1.
- Đầu tiên xóa đi các trường dữ liệu không phải là số (NaN) và các giá trị vô cực (Inf). Các nhãn "BENIGN" sẽ được gán bằng 0, tất cả các nhãn còn lại được gán bằng 1.
- Sau đó thì dữ liệu được chuyển về định dạng mảng, dữ liệu sau khi chuyển sang mảng có kích thước là 25x25, bước tiếp theo chuyển về dạng RBG.

4.2 Kết quả

Hiện nay, trong lĩnh vực huấn luyện học máy, bảo mật và đảm bảo tính riêng tư của dữ liệu người dùng khi huấn luyện là một vấn đề đang rất được chú ý. Bên cạnh đó, với sự tiến bộ của công nghệ và khoa học, lưu lượng mạng truy cập của người dùng ngày càng tăng. Điều này khiến cho hệ thống phát hiện xâm nhập phải chịu quá nhiều tải vì lượng lớn các truy cập. Để giải quyết các vấn đề này, nhóm chúng tôi đã thực hiện giải pháp huấn luyện mô hình máy học bằng phương pháp học công tác trên mạng khả lập trình.

Trong đề tài này nhóm chúng tôi đã hoàn thành nghiên cứu trình phát hiện tấn công dựa trên học cộng tác trong mạng khả lập trình. Phương pháp học cộng tác trên nhiều tác tử đã cho ta thấy nhiều lợi ích của phương pháp này mang lại.

Những lợi ích mà phương pháp học cộng tác mang lai:

+ Đảm bảo tính riêng tư của dữ liệu người dùng trên thiết bị.

+ Các máy huấn luyện không bị quá tải ở ổ cứng về dữ liệu thu thập được và không bị quá tải khi tiền xử lý quá nhiều dữ liệu một lúc.

+ Khi nạp bộ dữ liệu vào để huấn luyện thì thời gian nạp vào sẽ ít hơn và tốn ít RAM hơn so vơi mô hình huấn luyên truyền thống. Việc huấn luyện sẽ tốn ít tài nguyên CPU hơn.

Với những mục tiêu được đề ra ở mục 3, sau quá trình thực nghiệm mô hình đề xuất ở mục 4.1, nhóm chúng tôi tổng kết được các kết quả đã đạt được:

1. Kết quả tại mô hình học công tác trên bộ dữ liệu kiểm tra chung của tất cả các tác tử cho ra được kết quả cao hơn so với mô hình học máy trên bộ dữ liệu chỉ có trên máy huấn luyện đó. Kết quả mô hình học máy truyền thống (tất cả dữ liệu tập trung tại một máy) cho ra kết quả gần như giống kết quả của mô hình học cộng tác.
2. Kết quả của mô hình học cộng tác khi áp dụng DP so với mô hình học cộng tác bình thường.
3. Vận hành SDN điều hướng các lưu lượng mạng từ containernet đến các IDS để nhận dạng và dự đoán.

	Label attack	Syn	MSSQL	NetBios
Local Model	Precision	**0.551**	**0.9993**	**0.9777**
	Recall	**0.7480**	**0.9992**	**0.9805**
	F-score	**0.6346**	**0.9992**	**0.9792**
Federated Learning Model	Precision	**0.9581**	**0.9933**	**0.9466**
	Recall	**0.8996**	**0.974**	**0.9745**
	F-score	**0.9279**	**0.9835**	**0.9603**
Ideal Model	Precision	**0.9833**	**0.9704**	**0.95613**
	Recall	**0.81065**	**0.9814**	**0.97815**
	F-score	**0.8886**	**0.9993**	**0.9669**

Bảng 1. Kết quả huấn luyện theo 3 mô hình

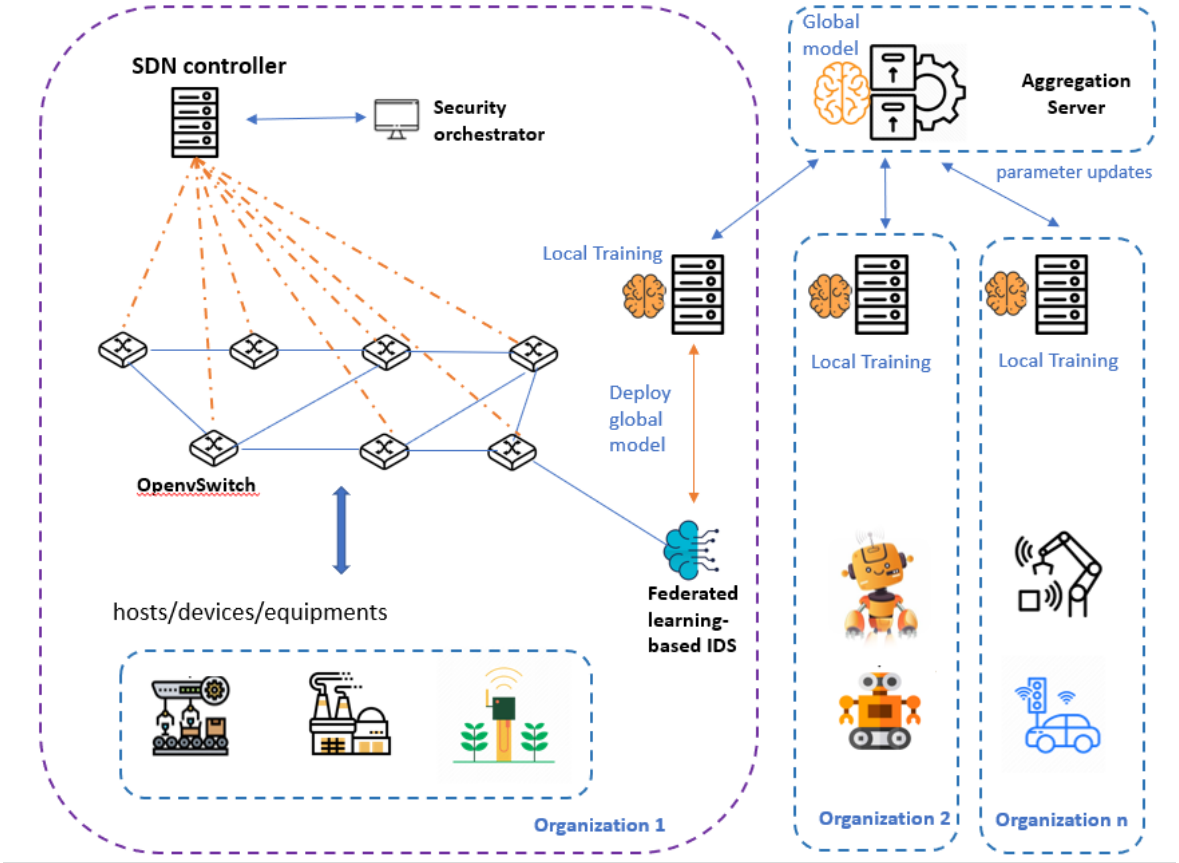
- **Kết quả mô hình FL:** kết quả cao trên cả 3 tập dữ liệu với nhãn là 3 cuộc tấn công khác nhau.
- **Kết quả mô hình tại mỗi máy nội bộ:** kết quả huấn luyện tại mỗi máy nội bộ lại xấp xỉ mô hình FL thậm chí còn thấp hơn ở tập dữ liệu với nhãn là Syn.
- **Kết quả mô hình tại máy tập trung:** Có độ chính xác tương đương với mô hình FL nhưng cao hơn bên mô hình Local.

5. Tên sản phẩm: Trình phát hiện tấn công dựa trên học cộng tác trong mạng khả lập trình.

6. Hiệu quả, phương thức chuyển giao kết quả nghiên cứu và khả năng áp dụng:

- Khả năng áp dụng mô hình học cộng tác trong mạng khả lập trình trong việc dự đoán tấn công mạng bằng mô hình máy học mà vẫn giữ được tính riêng tư của dữ liệu người dùng. Bên cạnh đó, việc triển khai và quản lý hạ tầng mạng đơn giản và nhanh chóng hơn khi áp dụng kiến trúc hạ tầng SDN.

7. Hình ảnh, sơ đồ minh họa chính



Hình 2. Cấu trúc mô hình phát hiện xâm dựa trên học cộng tác trong SDN

Cơ quan Chủ trì	Chủ nhiệm đề tài
(ký, họ và tên, đóng dấu)	*(ký, họ và tên)*