XÂY DỰNG HỆ THỐNG PHÁT HIỆN XÂM NHẬP DỰA TRÊN HỌC SÂU KHẢ DIỄN GIẢI SỬ DỤNG GIÁ TRỊ SHAPLEY

Huỳnh Thái Thi - 230202032[1]

[1] Trường Đại học Công nghệ Thông tin, ĐHQG TPHCM

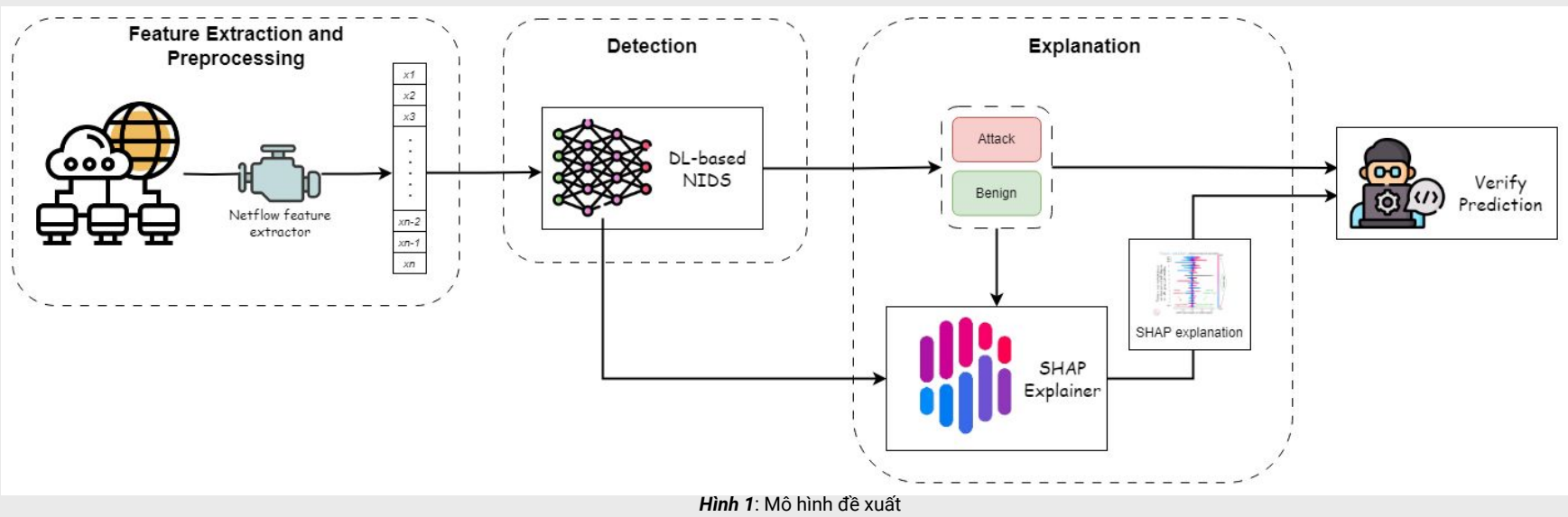
What ?

- Xây dựng các mô hình phát hiện xâm nhập dựa trên các mạng học sâu như CNN, LSTM, GRU, … sử dụng bộ dataset NF-ToN-IoT.
- Diễn giải được dự đoán của mô hình phát hiện xâm nhập sau áp dụng bộ khung SHAP theo cách mà các nhà phân tích có thể hiểu được.
- So sánh két quả diễn giải từ bộ khung SHAP với các thuật toán khác như LIME, Anchors, ….

Why ?

- **Hệ thống phát hiện xâm nhập mạng** dựa trên **học sâu** đang là xu hướng phát triển.
- Các **mạng nơ-ron** cấu thành nên các mô hình học sâu được xem là "hộp đen" bởi tính khó diễn giải của chúng.
- Các thuật toán **học máy khả giải** (**Giá trị Shapley**, LIME, etc.) nhiều tiềm năng nhưng còn hạn chế trong việc ứng dụng trong lĩnh vực an toàn thông tin.

Overview



Hình 1: Mô hình đề xuất

Description

1. Thu thập và tiền xử lý dữ liệu

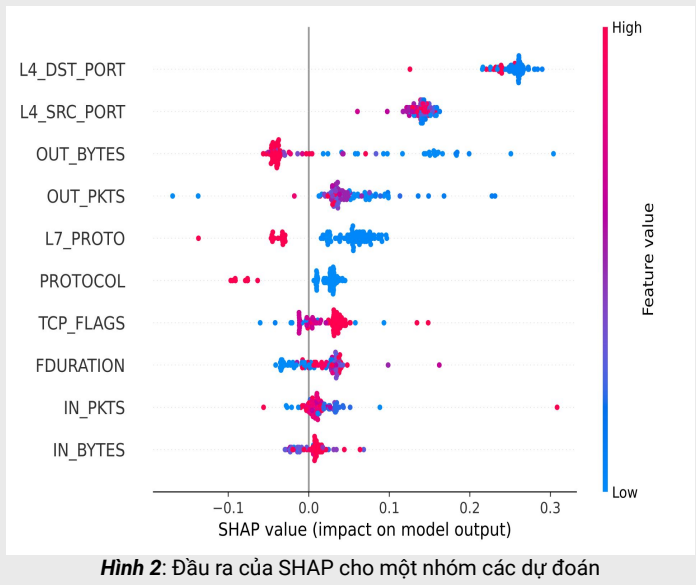
- Sử dụng bộ dữ liệu NF-ToN-IoT, bao gồm lưu lượng mạng được ghi lại từ các thiết bị IoT bị xâm nhập và lưu lượng mạng bình thường đã được trích xuất thuộc tính dựa trên giao thức NetFlow.
- Tiền xử lý dữ liệu để đảm bảo tính nhất quán và phù hợp với định dạng đầu vào của các mô hình học sâu. Các bước tiền xử lý có thể bao gồm:
  - Chuẩn hóa dữ liệu
  - Xử lý giá trị thiếu
  - Xử lý ngoại lệ
  - Chia dữ liệu thành tập huấn luyện, tập kiểm tra và tập đánh giá.

2. Xây dựng mô hình học sâu

- Tham khảo nghiên cứu và thiết kế đi trước để xây dựng nên các mô hình phát hiện xâm nhập dựa trên các mạng học sâu như CNN, LSTM, GRU, …
- Huấn luyện mô hình trên tập huấn luyện và đánh giá hiệu quả mô hình trên tập NF-ToN-IoT đã tiền xử lý.

3. Áp dụng bộ khung SHAP để diễn giải mô hình

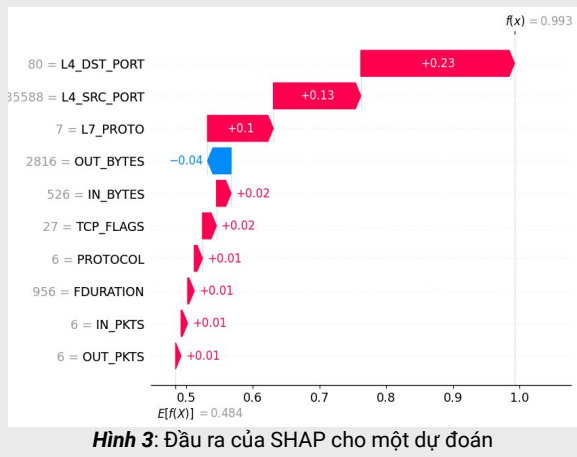
- Sử dụng bộ khung SHAP để tính toán giá trị Shapley (thể hiện mức độ ảnh hưởng của từng thuộc tính đối với dự đoán của mô hình) cho mỗi thuộc tính đầu vào như **Hình 2** và **Hình 3**.
- Biểu diễn giá trị Shapley bằng các phương pháp trực quan như biểu đồ thanh, biểu đồ nhiệt để giúp người dùng dễ dàng hiểu được.



Hình 2: Đầu ra của SHAP cho một nhóm các dự đoán

3. So sánh độ chính xác của SHAP với các thuật toán học máy khả giải khác.

- Áp dụng các thuật toán học máy khả giải khác như LIME, Anchors vào mô hình phát hiện xâm nhập.
- So sánh độ chính xác của các thuật toán học máy khả giải trong việc giải thích dự đoán của mô hình.
- Đánh giá ưu và nhược điểm của từng thuật toán để lựa chọn phương pháp phù hợp nhất cho bài toán cụ thể.



Hình 3: Đầu ra của SHAP cho một dự đoán

NII

Huỳnh Thái Thi – Trường Đại học Công nghệ thông tin

TEL : 0384 895 923 Email : thiht.18@gra.uit.edu.vn