PHÁT HIỆN TẤN CÔNG MẠNG DỰA TRÊN PHƯƠNG PHÁP HỌC MÁY

Huỳnh Thị Xuân Thịnh - 230202015

Tóm tắt

- Lớp: CS2205.CH181
- Link Github: https://github.com/xunthin99/CS2205.CH181_PPLNCKH
- Link YouTube video: https://youtu.be/DxieAHhZUVo
- Họ và Tên: Huỳnh Thị Xuân Thịnh
- Mã học viên: 230202015

Giới thiệu

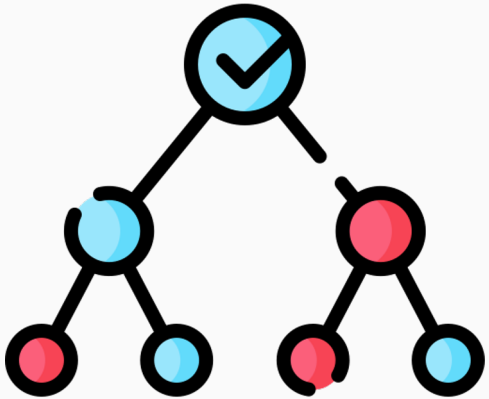
- An ninh mạng ngày nay đang đối mặt với những thách thức lớn khi kẻ tấn công ngày càng tinh vi, nguy hiểm.
- Các hệ thống phát hiện xâm nhập truyền thống chưa phát hiện được các cuộc tấn công mới, chưa có signature.
- Tạo ra nhiều cảnh báo giả.

Mục tiêu

- Xây dựng mô hình phát hiện tấn công mạng dựa trên phương pháp máy học
- Tăng cường tính hiệu quả và khắc phục được những điểm yếu của mô hình truyền thống
- Phát triển mô hình đã có và có thể thực hiện khả năng học liên tục và tiến hóa trong thuật toán
- Góp phần phát hiện nhanh những đợt tấn công mạng đã có sẵn và mới nhất

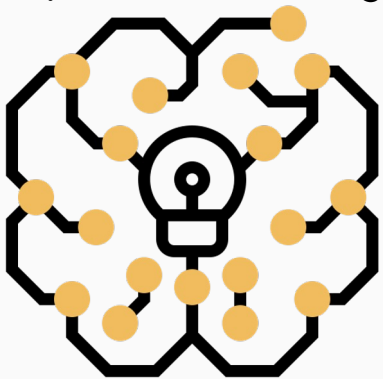
Nội dung và Phương pháp

- **Nội dung 1:** Nắm rõ được những khái niệm về tấn công mạng, phương thức mà các công cụ NIDS phát hiện tấn công mạng.
- **Nội dung 2:** Tìm hiểu về cách hoạt động và xây dựng cây quyết định (Decision Tree) từ thuật toán ID3



Nội dung và Phương pháp

- **Nội dung 3:** Lựa chọn, thu thập Dataset
- **Nội dung 4:** Thực hiện training cho mô hình và tinh chỉnh các số liệu cho phù hợp.
- **Nội dung 5:** Dựa trên các kết quả từ môi trường thử nghiệm và môi trường thực tế, từ đó cho ra được những đánh giá thiết thực về mô hình nghiên cứu.

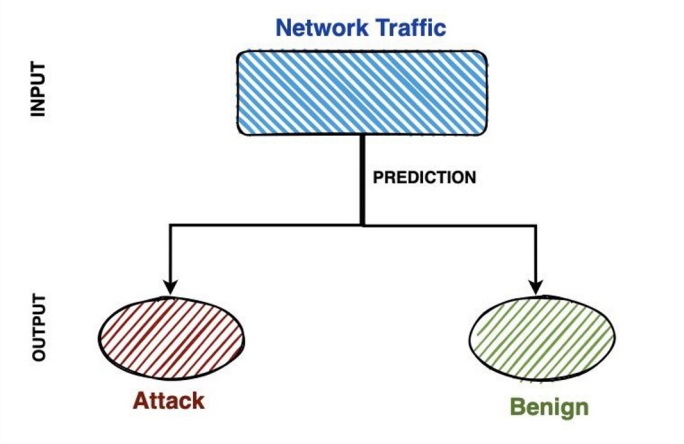


Nội dung và Phương pháp

- Xây dựng một mô hình NIDS cơ bản có hiệu quả. Phát hiện được các cuộc tấn công cơ bản, đã có signature được công bố trên cộng động.
- Phân tích nghiên cứu về cấu trúc hoạt động của Decision Tree, tổng hợp các cách và thứ tự xây dựng. Các model của Decision Tree bao gồm CART, C4.5 và ID3. ID3 sẽ được áp dụng trong đề tài này bởi tính tiện dụng và hiệu quả.
- Trong thuật toán ID3 cần phải xác định thứ tự của các thuộc tính cần xem xét tại mỗi bước. Thứ tự của các thuộc tính được phân chia theo thuộc tính tốt nhất sẽ lần lượt được chọn ra dựa theo một tiêu chuẩn nào đó (ví dụ 1 lưu lượng đến từ quốc gia lạ).
- Để đánh giá được chất lượng thì, sau mỗi lần phân chia dữ liệu vào từng thuộc tính, ta thực hiện xem xét dữ liệu được đổ hoàn toàn vào cùng một thuộc tính hay không, hoặc là dữ liệu bị phân chia lẫn vào nhau. Từ đó đưa ra được điểm cần điều chỉnh và tiến hành xây dựng node quyết định

Nội dung và Phương pháp

- Bộ dữ liệu NSL-KDD tập hợp các bản ghi lưu lượng mạng được sử dụng trong đề tài này. NSL-KDD datatset bao gồm có 4 tập dữ liệu con là DDTest+, KDDTest-21, KDDTrain+ và KDDTrain+_20Percent trong đó KDDTrain+ sẽ được ưu tiên dùng để huấn luyện mô hình và KDDTest+ dùng để kiểm tra.
- Huấn luyện mô hình trên bộ dữ liệu được chọn và so sánh kết quả.
- Xây dựng các cảnh báo phát hiện tấn công xâm nhập mạng và bỏ qua với những lưu lượng cho là bình thường.



Kết quả dự kiến

- Có thể thực hiện kết hợp xây dựng mô hình NIDS và học máy để phát hiện tấn công mạng
- Bộ dữ liệu sử dụng áp dụng tốt cho mô hình
- Có khả năng thu thập dữ liệu, học hỏi và tiến hóa phát hiện chính xác tấn công mạng
- Phát hiện tốt và nhanh chóng các cuộc tấn công mạng xảy ra
- Áp dụng được trong thực tiễn.

Tài liệu tham khảo

[1] Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., Ahmad, F.: Network intrusion detection system: a systematic study of machine learning and deep learning approaches. Trans. Emerg. Telecommun. Technol. 32(1), e4150 (2021)

[2] Ring, M., Wunderlich, S., Scheuring, D., Landes, D., Hotho, A.: A survey of network-based intrusion detection data sets. Comput. Secur. 86, p. 147–167 (2019)

[3] Manghnani, T., Thirumaran, T.: Computational CBGSA – SVM model for network based intrusion detection system. In: Applications and Techniques in Information Security, pp. 185–191. Singapore (2019)

[4] Zhou, Y., Cheng, G., Jiang, S., Dai, M.: Building an efficient intrusion detection system based on feature selection and ensemble classifier. Comput. Netw. 174, 107247 (2020)

[5] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[6] https://github.com/shramos/Awesome-Cybersecurity-Datasets#network-traffic