THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
  (ví dụ: https://www.youtube.com/watch?v=AWq7uw-36Ng)
- Link slides (dạng .pdf đặt trên Github của nhóm):
  (ví dụ: https://github.com/mynameuit/CS2205.xxx/TenDeTai.pdf)
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*
- *Lớp Cao học, mỗi nhóm một thành viên*

- Họ và Tên: Phạm Xuân Bách
- MSHV: 240202002

- Lớp: CS2205.CH183
- Tự đánh giá (điểm tổng kết môn): 8.5/10
- Số buổi vắng: 1
- Số câu hỏi QT cá nhân: 4
- Số câu hỏi QT của cả nhóm: 4
- Link Github: https://github.com/pxuanbach/Final_Project-CS2205.CH183

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

Hệ thống phát hiện xâm nhập tự phục hồi tự động cho môi trường điện toán đám mây

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

Autonomous Self-Healing IDS for Cloud Computing Environments

TÓM TẮT *(Tối đa 400 từ)*

GIỚI THIỆU

Với sự bùng nổ của các dịch vụ web, điện toán đám mây đã trở thành nền tảng quan trọng trong hạ tầng công nghệ thông tin của các tổ chức và doanh nghiệp. Với những lợi ích về khả năng mở rộng, tính linh hoạt và tối ưu chi phí, các mô hình **Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS)** và **Software-as-a-Service (SaaS)** ngày càng phổ biến, thúc đẩy sự phát triển của nền kinh tế số. Tuy nhiên, cùng với sự gia tăng của các dịch vụ đám mây, các mối đe dọa an ninh mạng cũng trở nên phức tạp và tinh vi hơn. Các cuộc tấn công từ chối dịch vụ phân tán (**DDoS**), khai thác lỗ hổng (**Zero-day Exploits**), tấn công nội gián (**Insider Threats**) và xâm nhập trái phép vào hệ thống đám mây đã gây ra những hậu quả nghiêm trọng về tổn thất tài chính, rò rỉ dữ liệu và suy giảm hiệu suất hệ thống.

Để đối phó với các mối đe dọa này, các hệ thống phát hiện xâm nhập (**Intrusion Detection System – IDS**) đã được phát triển nhằm phát hiện, phân tích và phản hồi trước các hành vi bất thường trong hệ thống mạng. Tuy nhiên, IDS truyền thống có nhiều hạn chế, bao gồm tỷ lệ cảnh báo sai cao (False Positives), phản ứng chậm, không tự động phục hồi khi bị tấn công, dẫn đến giảm hiệu quả trong môi trường đám mây động và phân tán. Vì vậy, **hệ thống phát hiện xâm nhập tự phục hồi (Autonomous Self-Healing IDS)** đã trở thành một nhu cầu cấp bách.

Nghiên cứu này đề xuất một hệ thống IDS tự phục hồi dựa trên **Machine Learning**

(ML) và Deep Learning (DL), giúp cải thiện khả năng phát hiện tấn công, phản ứng nhanh và tự động điều chỉnh hệ thống khi gặp sự cố. Với việc sử dụng CICIDS2017 làm tập dữ liệu huấn luyện, mô hình này có thể phát hiện chính xác các mối đe dọa mạng phổ biến, đồng thời tối ưu hóa chiến lược phục hồi bằng học tăng cường (Reinforcement Learning).

MỤC TIÊU

Trong nghiên cứu này, chúng tôi hướng tới các mục tiêu chính sau:

- Phát triển mô hình IDS tự phục hồi có khả năng phát hiện tấn công mạng có độ chính xác cao và tự động khôi phục hệ thống mà không cần sự can thiệp của con người.
- Ứng dụng các kỹ thuật học máy và học sâu để nâng cao khả năng phát hiện bất thường và phân loại mối đe dọa với độ chính xác cao hơn IDS truyền thống.
- Phân tích tác động của IDS tự phục hồi lên hiệu suất hệ thống đám mây, bao gồm chi phí tính toán, mức độ tiêu thụ tài nguyên, độ trễ hệ thống và mức độ ảnh hưởng đến dịch vụ.

NỘI DUNG VÀ PHƯƠNG PHÁP

Nội dung 1: Tìm hiểu về các kiến thức nền tảng và các nghiên cứu hiện có của hệ thống phát hiện xâm nhập tự phục hồi

Tìm hiểu nguyên lý cơ bản, ưu, nhược điểm của IDS truyền thống so với IDS tự phục hồi.

Khảo sát, phân tích các nghiên cứu hiện có về chủ đề IDS tự phục hồi:

- Hệ thống các thuật toán ML/DL đã được sử dụng và kết quả đánh giá (độ chính xác, tỷ lệ cảnh báo sai).
- Các bộ dữ liệu đã được sử dụng để huấn luyện các mô hình ML/DL.
- Tìm hiểu thuật toán học tăng cường để đưa ra quyết định tự phục hồi cho hệ thống.

Nội dung 2: Thiết kế IDS tự phục hồi ứng dụng Machine Learning (ML) và Deep

Learning (DL)

Hệ thống IDS tự phục hồi cần **khả năng học hỏi, thích nghi và phản ứng tự động** trước các cuộc tấn công mạng. Để đạt được cơ chế đó, hệ thống này sẽ bao gồm ba thành phần chính:

- **Phát hiện tấn công**: Các mô hình ML/DL giúp phát hiện bất thường trong lưu lượng mạng và nhận diện các hành vi tấn công **chưa từng thấy trước đó**.
- **Phân loại mối đe dọa**: Học sâu giúp phân biệt giữa các loại tấn công, giảm thiểu tỷ lệ cảnh báo sai (False Positives) và nâng cao độ chính xác của IDS.
- **Cơ chế tự phục hồi**: Các thuật toán học tăng cường (Reinforcement Learning - RL) giúp hệ thống **tự động đưa ra quyết định**, như cô lập máy chủ bị tấn công, điều chỉnh firewall, hoặc khôi phục hệ thống về trạng thái an toàn.

Các thuật toán ML/DL được nhóm đề xuất là Random Forest (RF), Support Vector Machine (SVM), Recurrent Neural Networks (RNNs) và Long Short-Term Memory (LSTM).

Nội dung 3: Phương pháp thực hiện

1. **Thông tin bộ dữ liệu**

Nghiên cứu này sẽ sử dụng bộ dữ liệu **CICIDS2017** – một trong những tập dữ liệu an ninh mạng tiêu chuẩn do **Canadian Institute for Cybersecurity** cung cấp. Với hơn 3 triệu bản ghi, 80 thuộc tính liên quan đến lưu lượng mạng và các loại tấn công đa dạng như Brute Force, DoS, DDoS, Botnet, SQL Injection, XSS, Infiltration, PortScan và Heartbleed.

2. **Tiền xử lý dữ liệu CICIDS2017**

Bộ dữ liệu có sự mất cân bằng giữa các lớp, nên cần tiền xử lý trước khi sử dụng trong ML/DL. Trong đó, chúng tôi sẽ chuyển đổi dữ liệu của các thuộc tính phi số thành số hóa và loại bỏ các giá trị trống. Tiếp theo, chúng tôi dự định tách thành 3

phần với tỉ lệ 60:20:20 tương ứng là tập dữ liệu huấn luyện, kiểm thử và đánh giá. Cuối cùng, với bộ dữ liệu đã qua tiền xử lý, nhóm sẽ chuẩn hóa dữ liệu bằng MinMaxScaler để đảm bảo dữ liệu có phân phối hợp lý cho các thuật toán ML/DL.

3. Đánh giá hiệu suất hệ thống

Hệ thống IDS tự phục hồi sẽ được đánh giá theo bốn tiêu chí chính:

- Độ chính xác phát hiện tấn công (Accuracy, Precision, Recall, F1-score).
- Thời gian phản ứng của hệ thống trong từng trường hợp bị tấn công.
- Ảnh hưởng đến tài nguyên hệ thống, phần trăm sử dụng CPU, RAM so với việc không áp dụng IDS và IDS truyền thống.
- Tỷ lệ tự phục hồi thành công (phần trăm số lần phản ứng đúng khi bị tấn công trên số lần bị tấn công).

KẾT QUẢ MONG ĐỢI

- Tài liệu báo cáo về một IDS có khả năng tự phục hồi ứng dụng ML/DL.
- Đưa ra các đánh giá thực nghiệm và so sánh với các IDS hiện có.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1]. Zamani, Mahdi and Movahedi, Mahnush. (2013). Machine learning techniques for intrusion detection. *arXiv preprint arXiv:1312.2177*.

[2]. Khraisat, Ansam and Gondal, Iqbal and Vamplew, Peter and Kamruzzaman, Joarder. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity, 2, 1--22.

[3]. Sharafaldin, Iman and Lashkari, Arash Habibi and Ghorbani, Ali A and others. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, *1*, 108--116.

[4]. Ahmed, Nisher and Hossain, Md Emran and Rishad, SSI and Mohiuddin, Arafath Bin and Sarkar, Md Imran and Hossain, Zakir. (2023). Leveraging Reinforcement Learning for Autonomous Cloud Management and Self-Healing Systems. *JURIHUM: Jurnal Inovasi Dan Humaniora*, *1*, 678--689.

[5]. Dai, Wenbin and Riliskis, Laurynas and Wang, Peng and Vyatkin, Valeriy and Guan, Xinping. (2018). A cloud-based decision support system for self-healing in distributed automation systems using fault tree analysis. *IEEE Transactions on Industrial Informatics*, *14*, 989--1000.

[6]. Kushal, Sauharda, Shanmugam, Bharanidharan, Sundaram, Jawahar, & Thennadil, Suresh. (2024). Self-healing hybrid intrusion detection system: an ensemble machine learning approach. *Discover Artificial Intelligence*, *4*, 28.