ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG PHÁT HIỆN HÌNH THỨC TẤN CÔNG TỪ CHỐI DỊCH VỤ

Nguyễn Thế Cường[1], Nguyễn Thu Hương[2]

TÓM TẮT

Các cuộc tấn công từ chối dịch vụ phân tán tạo ra các mối đe dọa lớn đối với mạng Internet và các thực thể trong mạng. Nhiều cơ chế phòng thủ đã được đề xuất để có thể chống lại hình thức tấn công này. Tuy nhiên, các công cụ hỗ trợ tấn công luôn được nâng cấp để có thể vượt qua các hệ thống bảo mật. Việc phát hiện sớm được dạng tấn công này có thể giúp các hệ thống giảm thiểu được rủi ro, đặc biệt là với các hệ thống cung cấp dịch vụ. Ứng dụng trí tuệ nhân tạo vào quá trình giám sát mạng nhằm phát hiện các nguy cơ xuất hiện của hình thức tấn công từ chối dịch vụ hiện đang thu hút được nhiều sự quan tâm. Trong bài báo này, chúng tôi đề xuất giải pháp ứng dụng trí tuệ nhân tạo trong việc phát hiện hình thức tấn công từ chối dịch vụ dựa trên các dữ liệu giám sát mạng.

Từ khóa: *Tấn công mạng, trí tuệ nhân tạo, từ chối dịch vụ, mạng dịch vụ.*

1. TỔNG QUAN VỀ TỪ CHỐI DỊCH VỤ PHÂN TÁN

Tấn công từ chối dịch vụ phân tán (Distributed Denial of Service - DDoS) là hành động làm cho máy cung cấp dịch vụ không thể phục vụ các yêu cầu sử dụng dịch vụ của người dùng hợp pháp. Kết quả, sau cuộc tấn công, hệ thống cung cấp dịch vụ ngừng hoạt động trong một khoảng thời gian đủ dài. Các tấn công DDoS có thể thực hiện từ các hệ thống máy tính với các địa chỉ IP khác nhau trên mạng Internet.

Các cuộc tấn công DDoS gây ra mối đe dọa lớn đối với mạng Internet và có nhiều cơ chế phòng thủ đã được đề xuất để chống lại vấn đề này. Lĩnh vực nghiên cứu về tấn công DDoS ngày càng trở nên phức tạp hơn và trở nên khó khăn đến mức khó có thể xác định được mức độ tấn công. Những kẻ tấn công liên tục sửa đổi các công cụ của họ để vượt qua các hệ thống bảo mật được xây dựng, trong khi đó, các nhà nghiên cứu liên tục phải sửa đổi các phương pháp tiếp cận của họ để xử lý các cuộc tấn công mới. Một cuộc tấn công từ chối dịch vụ được đặc trưng bởi một nỗ lực rõ ràng nhằm ngăn chặn việc sử dụng hợp pháp một dịch vụ được triển khai trên mạng. Một cuộc tấn công DdoS có thể dùng nhiều đối tượng, phương thức để đạt được mục tiêu này.

Thông thường, có một số cách để thực hiện một cuộc tấn công DDoS:

Cách thứ nhất, kẻ tấn công có thể gửi một luồng gói tin đến nạn nhân; luồng này tiêu thụ một số tài nguyên quan trọng, do đó khiến tài nguyên không khả dụng cho các khách hàng hợp pháp của nạn nhân.

[1] *Phòng Đảm bảo chất lượng và Khảo thí, Trường Đại học Hồng Đức; Email: nguyenthecuong@hdu.edu.vn*

[2] *Trường Cao đẳng Y tế Thanh Hóa*

Cách thứ hai, kẻ tấn công gửi một vài gói tin không đúng định dạng gây nhầm lẫn cho một ứng dụng hoặc một giao thức trên máy nạn nhân và buộc máy tính đó phải đóng băng hoặc khởi động lại.

Cách thứ ba, chiếm quyền sử dụng các máy trong mạng nạn nhân và tiêu thụ một số tài nguyên quan trọng để các máy khách hợp pháp từ cùng một mạng không thể nhận được một số dịch vụ bên trong hoặc bên ngoài.

Nhìn chung, có nhiều cách khác nhau để từ chối dịch vụ trên internet, một số cách mà chúng ta không thể đoán trước được và những cách này sẽ chỉ được phát hiện sau khi chúng đã bị khai thác trong một cuộc tấn công lớn.

2. CÁC NGHIÊN CỨU LIÊN QUAN ĐẾN DỰ ĐOÁN TẤN CÔNG TỪ CHỐI DỊCH VỤ

Trong phần này, tác giả trình bày một số nghiên cứu có liên quan đến việc chống tấn công DDoS dựa trên các kỹ thuật học máy và học sâu.

Các giải pháp bảo mật như Hệ thống ngăn chặn xâm nhập (IPS) và Hệ thống phát hiện xâm nhập (IDS) được sử dụng để đảm bảo an ninh mạng. Với các thuật toán học máy, hệ thống IDS đã có được khả năng đưa ra các nhận xét và dự đoán tương đối chính xác về các tấn công. Pérez-Díaz và cộng sự [1] đề xuất một giải pháp kiến trúc mới để phát hiện các cuộc tấn công DDoS tỷ lệ thấp (LR-DDoS) và giảm thiểu độ thiệt hại của các cuộc tấn công trong các hệ thống SDN. Giải pháp kiến trúc bao gồm các mô-đun IPS và IDS được đặt trên bộ điều khiển. Việc phát hiện tấn công được thực hiện bằng cách sử dụng các phương pháp học máy và học sâu được đào tạo khác nhau thông qua giao diện lập trình ứng dụng nhận dạng (API) được đặt trong mô-đun IDS. Kết quả thử nghiệm cho thấy, thuật toán Multi-Layer Perceptron (MLP) cho kết quả tốt nhất với độ chính xác 95% trong số 6 thuật toán học máy khác nhau được sử dụng. Shoo và cộng sự [2] đã giới thiệu một mô hình tiến hóa mới để phân loại lưu lượng tấn công DDoS trong môi trường SDN. Mô hình sử dụng thuật toán SVM kết hợp để phân loại các lưu lượng độc hại với các lưu lượng bình thường. Các thuật toán di truyền (GA) đã được sử dụng để tối ưu hóa SVM khi xác định thành phần chính của nhân (KPCA) như một phương pháp lựa chọn thuộc tính để cải thiện hiệu suất phân loại của mô hình. Kết quả thực nghiệm cho thấy độ chính xác của phương pháp kết hợp đề xuất là 98,9%.

Kyaw, Aye Thandar và các cộng sự [3] đã sử dụng hai thuật toán học máy (polynomial SVM và linear SVM) để phát hiện các cuộc tấn công UDP Flood trong môi trường SDN. Các tác giả đã sử dụng công cụ Scapy để tạo gói lưu lượng. Hệ thống đã thu thập các dòng tĩnh thông qua bộ chuyển mạch OpenFlow. Sau giai đoạn trích xuất đặc trưng, các tác giả so sánh hiệu suất phân loại của các mô hình SVM tuyến tính và đa thức. Kết quả thử nghiệm cho thấy thuật toán Polynomial SVM có tỷ lệ báo động sai thấp hơn 34% với độ chính xác tốt hơn 3%.

Janarthanam, S và các cộng sự [4] đã đề xuất khung bảo mật phát hiện các cuộc tấn công DDoS trên môi trường SDN. Khung dựa trên mô hình học tập thích ứng sử dụng tập dữ liệu lịch sử để phân loại lưu lượng. Các tác giả đã sử dụng phương pháp xác nhận chéo để có kết quả phân loại hiệu quả. Tan, Liang và cộng sự [5] đề xuất một mô hình bảo mật mới cho các cuộc tấn công DDoS trong môi trường SDN. Mô hình gồm 2 mô-đun: Mô-đun xử lý dữ liệu sử dụng thuật toán K-Means để lựa chọn tính năng tốt nhất và mô-đun phát hiện sử dụng thuật toán k-láng giềng gần nhất (kNN) để phát hiện các luồng tấn công. Phương pháp có độ chính xác 98,85% với tỷ lệ đúng đắn lên đến 98,47%.

Deepa, V. và các cộng sự [7] đã đề xuất một kỹ thuật tổng hợp để phát hiện các cuộc tấn công từ chối dịch vụ (DDoS). Các tác giả đã sử dụng 4 mô hình học máy khác nhau để phát hiện lưu lượng truy cập đáng ngờ trong môi trường SDN. Thuật toán SVM-SOM cho kết quả tốt hơn so với các thuật toán ML khác với độ chính xác 98,12%. Các tác giả trong [8] đã giới thiệu một hệ thống phát hiện tấn công DDoS cho SDN. Hệ thống sử dụng 2 giai đoạn bảo mật. Thứ nhất, họ sử dụng Snort để phát hiện các cuộc tấn công dựa trên chữ ký. Sau đó, họ sử dụng trình phân loại SVM và mô hình học máy mạng nơ ron sâu (DNN) để phân loại tấn công. Kết quả thực nghiệm đã chứng minh DNN có tỷ lệ chính xác phân loại tốt hơn SVM là 92,30%.

Bài báo tập trung vào việc tiếp cận với các cơ sở dữ liệu mới nhất từ các mạng SDN và sử dụng một số giải thuật trí tuệ nhân tạo, học máy để xây dựng mô hình dự đoán.

3. CÁC KỸ THUẬT TRÍ TUỆ NHÂN TẠO SỬ DỤNG TRONG DỰ ĐOÁN

3.1. Cây quyết định

Cây quyết định [9] là một cấu trúc biểu diễn dưới dạng cây. Trong đó, mỗi node trong (internal node) biểu diễn một thuộc tính, mỗi nhánh (branch) biểu diễn giá trị có thể có của thuộc tính, mỗi lá (leaf node) biểu diễn các lớp quyết định và đỉnh trên cùng của cây gọi là gốc (root).

Trong lĩnh vực học máy, cây quyết định là một kiểu mô hình dự báo (predictive model), nghĩa là một ánh xạ từ các quan sát về một sự vật/hiện tượng tới các kết luận về giá trị mục tiêu của sự vật/hiện tượng. Mỗi nút trong (internal node) tương ứng với một biến; đường nối giữa nó với nút con của nó thể hiện giá trị cụ thể cho biến đó. Mỗi nút lá đại diện cho giá trị dự đoán của biến mục tiêu, cho trước các giá trị dự đoán của các biến được biểu diễn bởi đường đi từ nút gốc tới nút lá đó. Kỹ thuật học máy dùng trong cây quyết định được gọi là học bằng cây quyết định, hay chỉ gọi với cái tên ngắn gọn là cây quyết định.

3.2. Mô hình Logistic Regression

Các mô hình phân loại đều tìm cách chia dữ liệu thành các nhóm dữ liệu có tính chất phân biệt nhau. Hồi quy logistic cũng tìm kiếm một đường biên để phân chia tập dữ liệu trong bài toán nhị phân thành hai nhóm 0 và 1, tương đương với việc xác định tính có hay không một tính chất nào đó của đối tượng.

Trong hồi quy logistic [3] cần một hàm số có tác dụng chiếu giá trị dự báo lên không gian xác suất nằm trong khoảng [0,1] và đồng thời tạo ra tính phi tuyến cho phương trình hồi quy nhằm giúp tạo ra đường biên phân chia giữa hai nhóm tốt hơn. Việc xác định được các phương trình hồi quy tốt sẽ tăng được hiệu quả của mô hình phân lớp, nhằm tăng tính hiệu quả trong quá trình dự đoán.

3.3. Mô hình Random Forest

Mô hình rừng cây (Random Forest) được huấn luyện dựa trên sự phối hợp giữa luật kết hợp (ensembling) và quá trình lấy mẫu tái lặp (boostrapping). Về cơ bản thuật toán này tạo ra nhiều cây quyết định mà mỗi cây quyết định được huấn luyện dựa trên nhiều mẫu con khác nhau và kết quả dự báo là bầu cử (voting) từ toàn bộ những cây quyết định. Kết quả dự báo là

kết quả được tổng kết từ các mô hình dự báo khác nhau nên kết quả của mô hình không bị lệch và phương sai của mô hình tổng hợp nhỏ hơn so với việc sử dụng một mô hình độc lập. Điều này giải quyết được vấn đề quá khớp (overfit) trong quá trình huấn luyện dữ liệu.

3.4. Mô hình Support Vector Machine

Thuật toán Support Vector Machine [10] là thuật toán khá hiệu quả trong lớp các bài toán phân loại nhị phân và dự báo của học có giám sát. Thuật toán này có ưu điểm là hoạt động tốt đối với những mẫu dữ liệu có kích thước lớn và thường mang lại kết quả vượt trội so với lớp các thuật toán khác trong học có giám sát.

4. XÂY DỰNG CƠ SỞ DỮ LIỆU HUẤN LUYỆN VÀ THỬ NGHIỆM MÔ HÌNH

Tác giả sử dụng các mô hình đã được xây dựng, sau đó thu thập, xây dựng bộ dữ liệu huấn luyện cho các mô hình nhằm tìm kiếm mô hình tốt nhất cho quá trình dự đoán.

4.1. Mô tả dữ liệu

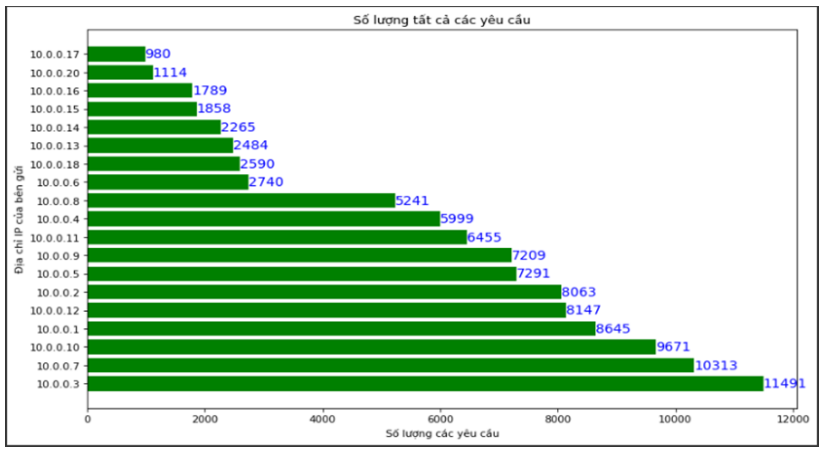
Dữ liệu huấn luyện là tập dữ liệu được tạo ra bằng cách sử dụng trình giả lập mạng con (mininet). Mô phỏng mạng được triển khai cho lưu lượng TCP, UDP và ICMP lành tính và lưu lượng độc hại, đây là kết quả của các tấn công TCP Syn, UDP Flood và ICMP.

Bảng 1. Danh sách các đặc trưng của dữ liệu

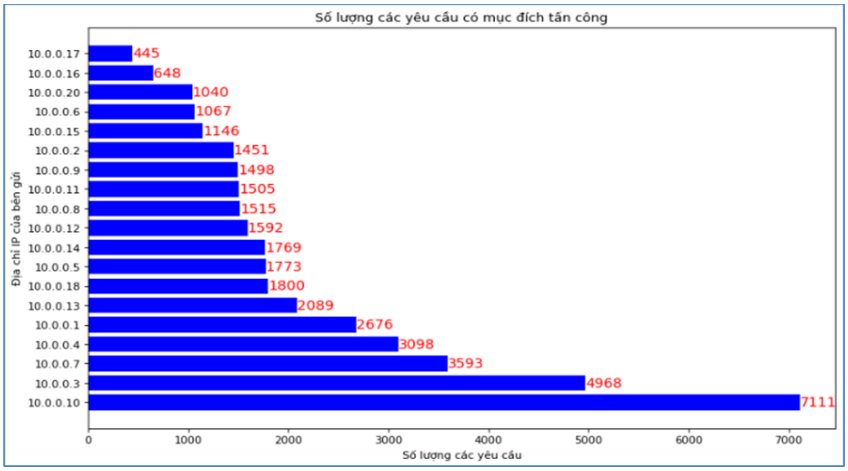
TT	Tên	Mô tả	TT	Tên	Mô tả
1	dt	Ngày giờ được chuyển đổi thành số	13	byteperflow	Số byte trên mỗi luồng đơn
2	switch	Mã số của bộ chuyển mạch	14	ptkrate	Số gói tin truyền đi trong mỗi giây
3	src	Địa chỉ nguồn	15	Pairflow	Số cặp luồng
4	dst	Địa chỉ đích	16	Protocol	Giao thức được thực hiện
5	ptkcount	Số lượng gói tin	17	port_no	Số cổng truyền tin
6	bytecount	Số lượng byte	18	tx_bytes	Số lượng gói tin truyền đi từ các cổng bộ chuyển mạch
7	dur	Thời lượng được tính theo giây	19	rx_bytes	Số lượng gói tin nhận về ở cổng bộ chuyển mạch
8	dur_nsec	Thời lượng được tính theo nano giây	20	tx_kbps	Thông lượng truyền đi qua cổng bộ chuyển mạch
9	tot_dur	Tổng thời lượng của dur và dur_nsec	21	rx_kbps	Thông lượng nhận về qua cổng bộ chuyển mạch
10	flows	Số luồng	22	tot_kbps	Tổng thông lượng
11	packetins	Số tin nhắn	23	label	Nhãn (lành tính, độc hại)
12	ptkperflow	Số gói trên mỗi luồng đơn			

4.2. Phân tích các đặc trưng của dữ liệu

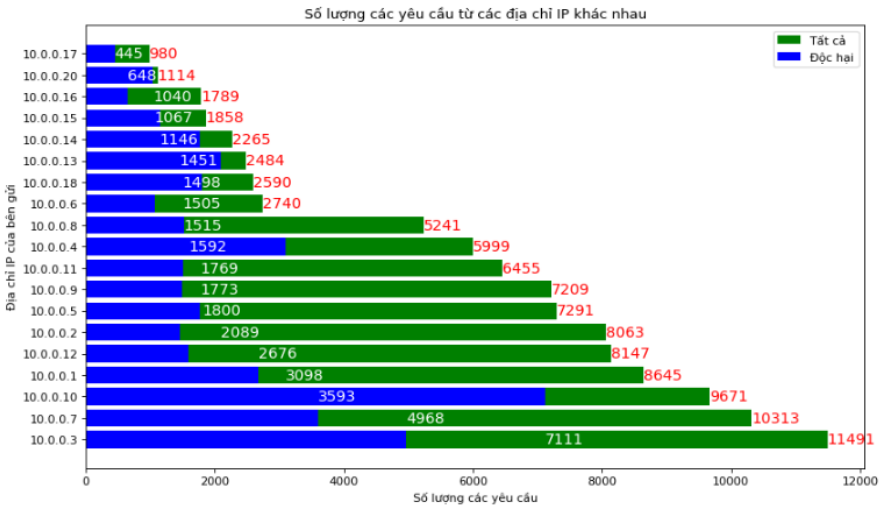
Tác giả đánh giá, phân tích các đặc trưng của dữ liệu để tìm ra các mối quan hệ giữa các đặc trưng của dữ liệu huấn luyện.



Hình 1. Kết quả thống kê các yêu cầu theo các địa chỉ máy gửi



Hình 2. Kết quả thống kê các yêu cầu có mục đích tấn công theo các địa chỉ máy gửi



Hình 3. Kết quả thống kê, so sánh về tỉ lệ các yêu cầu có mục đích tấn công

4.3. Lựa chọn đặc trưng dữ liệu

Trong bài báo này, tác giả sử dụng giải thuật NCA (Neighbourhood Component Analysis) để tìm kiếm các đặc trưng phù hợp nhất cho việc xây dựng mô hình phân lớp của hơn 100.000 bản ghi, trong đó có 22 đặc trưng liên quan đến mạng SDN.

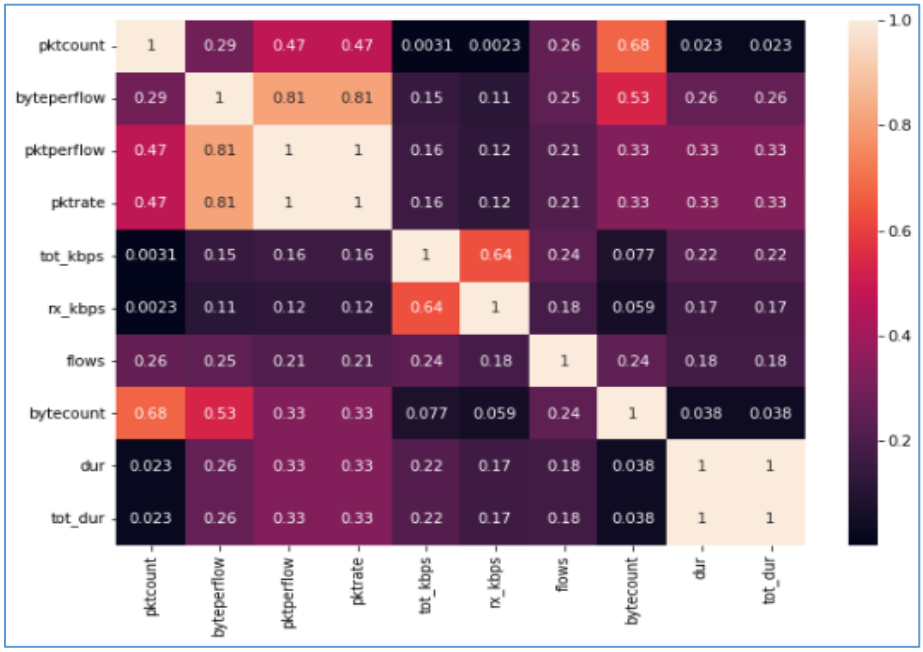
Sau khi phân tích 22 đặc trưng mạng bằng thuật toán NCA, quá trình phân loại đầu tiên được thực hiện với 8 đặc trưng có giá trị chỉ số lớn hơn 9. Trong nghiên cứu thử nghiệm thứ 2, 14 đặc trưng hiệu quả đã được chọn và đưa ra làm dữ liệu đầu vào cho các thuật toán học máy. Danh sách 14 đặc trưng và giá trị trọng số hiệu quả nhất được NCA lựa chọn.

Bảng 2. Danh sách các đặc trưng và giá trị trọng số tương ứng

TT	Đặc trưng	Trọng số được tính theo NCA	TT	Đặc trưng	Trọng số được tính theo NCA
1	src	17,87	8	rx_kbps	9,66
2	pktcount	15,16	9	flows	8,95
3	dst	13,64	10	bytecount	4,92
4	byteperflow	12,97	11	dt	2,33
5	pktperflow	11,35	12	protocol	1,31
6	pktrate	11,35	13	dur	1,11
7	tot_kbps	9,68	14	tot_dur	1,11

Bởi vì các đặc trưng như src, dst, dt là không cần thiết để triển khai mô hình nên ta loại bỏ các đặc trưng này khỏi danh sách các đặc trưng lựa chọn.

Sau đó, tác giả tính mức độ tương quan giữa các đặc trưng nhằm xác định mức độ phụ thuộc của các đặc trưng.



Hình 4. Biểu đồ mức độ tương quan giữa các đặc trưng

Khi xem xét bảng tương quan, ta thấy có một số đặc trưng là trùng nhau hoặc có độ tương quan lớn như "dur" và "tot_dur", "pktperflow" và "pktrate", do vậy, ta loại bỏ 3 đặc trưng là "dur", "pktrate", "pktperflow" ra khỏi bảng dữ liệu. Khi đó ta có:

Hình 5. Biểu đồ tương quan giữa các đặc trưng sau khi lựa chọn đặc trưng

4.4. Lựa chọn mô hình dự báo

Ứng dụng kỹ thuật lựa chọn đặc trưng vào cơ sở dữ liệu, khi đó, tác giả thực hiện việc xây dựng mô hình với 7 đặc trưng có ý nghĩa quan trọng với các thông tin về mạng SDN. Đó là các đặc trưng: pktcount, byteperflow, tot_kbps, rx_kbps, flows, bytecount, tot_dur. Kết quả triển khai cho thấy, độ chính xác cao nhất khi sử dụng mô hình Random Forest, với độ chính xác là 99,42%.

Sử dụng kỹ thuật Logistic Regression

Accuracy: **75.20%**

Best solver is : sag

	precision	recall	f1-score	support
0	0.85	0.77	0.81	20965
1	0.60	0.72	0.65	10187
accuracy			0.75	31152

Sử dụng kỹ thuật Support Vector Machine

Accuracy of SVM model **92.0%**

Best kernel is : rbf

	precision	recall	f1-score	support
0	0.90	0.96	0.93	17757
1	0.95	0.86	0.90	13395
accuracy			0.92	31152

Sử dụng kỹ thuật Random Forest

Accuracy of RF is : **99.42%**

	precision	recall	f1-score	support
0	0.99	1.00	1.00	18922
1	1.00	0.99	0.99	12230
accuracy			0.99	31152

Sử dụng kỹ thuật Decision Tree

The Accuracy is : **94.19%**

	precision	recall	f1-score	support
0	0.91	1.00	0.95	17287
1	1.00	0.87	0.93	13865
accuracy			0.94	31152

5. KẾT LUẬN

Trong bài báo, tác giả đã giới thiệu các kỹ thuật trí tuệ nhân tạo thường được sử dụng để xây dựng các mô hình phân lớp, dự đoán. Tác giả sử dụng 4 giải thuật bao gồm cây quyết định, Logistic Regression, Random Forest và Support Vector Machine cùng với giải thuật lựa chọn đặc trưng để xây dựng mô hình dự đoán khả năng xảy ra hiện tượng tấn công DDoS trong các mạng dịch vụ SDN. Kết quả cho thấy, mô hình Random Forest cho kết quả dự đoán tốt nhất, đặc biệt là có sự hỗ trợ của giải thuật lựa chọn đặc trưng NCA.

TÀI LIỆU THAM KHẢO

[1] J. A. Perez-Diaz, I. A. Valdovinos, K. K. R. Choo, D. Zhu (2020), *A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning*, IEEE Access, vol.8, p.155859-155872.

[2] K. S. Sahoo et al. (2020), *An Evolutionary SVM Model for DDOS Attack Detection in Software Defined Networks*, IEEE Access, vol.8, p.132502-132513.

[3] A. T. Kyaw, M. Z. Oo, C. S. Khin (2020), *Machine-Learning Based DDOS Attack Classifier in Software Defined Network*, in 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 431-434.

[4] S. Janarthanam, N. Prakash, M. Shanthakumar (2020), *Adaptive Learning Method for DDoS Attacks on Software Defined Network Function Virtualization*, EAI Endorsed Transactions on Cloud Systems, 6(18), p.166286.

[5] L. Tan, Y. Pan, J. Wu, J. Zhou, H. Jiang, Y. Deng (2020), *A New Framework for DDoS Attack Detection and Defense in SDN Environment*, IEEE Access, vol. 8, p.161908-161919.

[6] L. Wang, Y. Liu (2020), *A DDoS Attack Detection Method Based on Information Entropy and Deep Learning in SDN*, in 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), p.1084-1088.

[7] V. Deepa, K. M. Sudar, P. Deepalakshmi (2019), *Design of Ensemble Learning Methods for DDoS Detection in SDN Environment*, in 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), p.1-6.

[8] K. B. V., N. D. G., P. S. Hiremath (2018), *Detection of DDoS Attacks in Software Defined Networks*, in 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), p.265-270.

[9] K. Adhikary, S. Bhushan, S. Kumar, K. Dutta (2020), *Decision Tree and Neural Network Based Hybrid Algorithm for Detecting*, International Journal of Innovative Technology and Exploring Engineering, 9(5).

[10] Y. Zhao, B. Li, X. Li, W. Liu, S. Ren (2005), *Customer Churn Prediction using improved one-class Support Vector Machine*, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), p.300-306.

APPLYING ARTIFICIAL INTELLIGENCE IN DETECTING DENIAL OF SERVICE ATTACKS

Nguyen The Cuong, Nguyen Thu Huong

ABSTRACT

Distributed denial-of-service (DDoS) attacks pose significant threats to the Internet and entities within networks. Various defense mechanisms have been proposed to counteract this form of attack. However, attack support tools are constantly upgraded to bypass security systems. Early detection of these attacks can help systems minimize risks, especially for service-providing systems. The application of artificial intelligence in network monitoring to detect the emergence of DDoS attacks is gaining significant attention. In this paper, we propose a solution that applies artificial intelligence to detect DDoS attacks based on network monitoring data.

Keywords: *Cyber attacks, artificial intelligence, denial of service, service networks.*

** Ngày nộp bài: 15/5/2023; Ngày gửi phản biện:20/5/2023; Ngày duyệt đăng: 10/12/2023*