HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

BÙI ĐIỀN PHONG

XÂY DỰNG CÔNG CỤ PHÁT HIỆN XÂM NHẬP MẠNG MÁY TÍNH

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SĨ

TPHCM - NĂM 2022

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: **TS. NGUYỄN ĐỨC THÁI**

Phản biện 1: ……

Phản biện 2: ……

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng .... năm .....

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

PHẦN MỞ ĐẦU

1. Lý do chọn đề tài

Tên đề tài: Xây dựng công cụ phát hiện xâm nhập mạng máy tính.

Năm 1986 mạng NSFnet chính thức được thiết lập, kết nối năm trung tâm máy tính. Đây cũng là năm có sự bùng nổ kết nối, đặc biệt là ở các trường đại học. Như vậy là NSF và ARPANET song song tồn tại theo cùng 1 giao thức, có kết nối với nhau.

Năm 1990, với tư cách là 1 dự án ARPANET dừng hoạt động nhưng mạng do NSF và ARPANET tạo ra đã được sử dụng vào mục đích dân dụng, đó chính là tiền thân của mạng Internet ngày nay. Đến lúc này đối tượng sử dụng Internet chủ yếu là những nhà nghiên cứu và dịch vụ phổ biến nhất là email và FTP. Internet là 1 phương tiện đại chúng.

Ngày nay Internet gần như được sử dụng phổ biến mọi lúc mọi nơi, không chỉ trên máy tính và còn sử dụng đại đa số trên di động thông minh và kể cả trong Y học. Mọi hồ sơ sau khi khám, chữa bệnh phải đưa lên cổng thông tuyến của Bảo hiểm Xã hội Việt Nam, Bộ Y tế.

Vì vậy, an ninh mạng là một vấn đề lớn rất quan trọng cho người quản trị để đảm bảo an ninh trong môi trường làm việc của họ. Có thể nói sự tổn thất cho một người dùng có thể sẽ không quá lớn hoặc không quan trọng với họ, nhưng đối với các doanh nghiệp, tổ chức lớn có thể tổn thất sẽ lên tới hàng triệu đô la hoặc các cơ quan tổ chức, nhà nước thì sẽ bị lộ các thông tin bí mật nhà nước. Hàng loạt các cuộc tấn công có thể nhắm vào mọi thứ như là dữ liệu cá nhân

hoặc tổ chức, tài khoản ngân hàng, phần mềm, tài khoản người dùng, mạng cục bộ, … Đó là lý do tại sao các công cụ bảo mật phát triển ngày càng nhiều để đáp ứng các dạng phần mềm nguy hiểm, phần mềm độc hại và tin tặc ngày nay.

Bảo mật thông tin hay an toàn an ninh mạng là những yếu tố được quan tâm hàng đầu trong các doanh nghiệp. Đã có những doanh nghiệp thực hiện việc thuê một đối tác thứ 3 với việc chuyên bảo mật hệ thống mạng và bảo mật thông tin cho đơn vị mình, cũng có những doanh nghiệp đưa ra các kế hoạch tính toán chi phí cho việc mua sản phẩm phần cứng hoặc phần mềm để nhằm đáp ứng việc đảm bảo an toàn dữ liệu của đơn vị mình. Tuy nhiên đối với những giải pháp đó các cơ quan doanh nghiệp đều phải thực hiện cân đối về chính sách tài chính hằng năm với mục đích làm sao cho giải pháp an toàn 2 là tối ưu và có được chi phí rẻ nhất đảm bảo việc trao đổi thông tin được an toàn, bảo vệ thông tin của đơn vị mình trước những mối nguy cơ tấn công của các tội phạm công nghệ.

Do đó đề tài "Xây dựng công cụ phát hiện xâm nhập mạng máy tính" được phát triển nhằm giúp một phần nào yêu cầu của các cơ quan, tổ chức, doanh nghiệp đảm bảo được an toàn thông tin và bảo mật hệ thống mạng của đơn vị mình.

Luận văn nghiên cứu về phương pháp phát hiện và phòng chống xâm nhập mạng máy tính và xây dựng công cụ phát hiện xâm nhập mạng máy tính.

Mục đích đề tài nhằm nhằm xây dựng một hệ thống phát hiện xâm nhập và phòng chống các cuộc tấn công từ internet và áp dụng

giải pháp vào trong thực tiễn công việc tại Trung tâm Y tế huyện Gò Dầu.

2. Tổng quan về vấn đề nghiên cứu

Xây dựng hệ thống phát hiện xâm nhập theo thời gian thực. Thêm vào đó, hệ thống có chức năng nhận dạng và phân tích các cuộc xâm nhập trái phép vào hệ thống. Sau khi thu được kết quả sẽ tổng hợp và kết xuất dữ liệu báo cáo ra file để tổng hợp. Viết báo cáo tổng kết luận văn.

3. Mục đích nghiên cứu

Tập trung nghiên cứu các loại xâm nhập mạng, phân tích và phân loại thành các mức độ nguy hiểm khác nhau.

Nghiên cứu cơ chế hoạt động của một hệ thống chống xâm nhập mạng, từ đó đưa ra tiếp thu và đưa ra các giải pháp phù hợp trong việc xây dựng hệ thống.

Nghiên cứu các công cụ hỗ trợ phân tích các luồng thông tin ra vào mạng máy tính kết hợp với những thuật toán phân lớp hoặc phân cụm để theo dõi và truy vết dấu hiệu hợp pháp và bất hợp pháp, sau đó gửi tính hiệu cảnh báo cho quản trị mạng biết những dấu hiệu xâm nhập trái phép.

4. Đối tượng và phạm vi nghiên cứu

Đề tài nghiên cứu đưa ra cách nhìn tổng quan nhất về một hệ thống phát hiện xâm nhập mạng máy tính, các phương thức tấn công mạng và các giải pháp bảo mật hệ thống mạng. Bên cạnh đó xây dựng một hệ thống giám sát và cảnh báo bằng email đến quản trị viên, giúp cho việc quản trị hệ thống mạng trở nên cơ động và an toàn hơn.

Đề tài sẽ được ứng dụng ngay tại Trung tâm Y tế huyện Gò

Dầu của học viên hoặc có thể mở rộng trong toàn ngành nơi mà học viên đang công tác và đáp ứng nhu cầu của cơ quan.

5. Phương pháp nghiên cứu

Tìm hiểu về cách thức xâm nhập trái phép trong mạng máy tính.

Nghiên cứu lý thuyết về các khả năng tấn công mạng.

Phân tích các khả năng phát hiện xâm nhập và phòng chống tấn công.

Nghiên cứu các thuật toán, đặt biệt là các thuật toán Support Vector Machine (SVM), Decision Tree (DT), K Nearest Neighbor để phát hiện xâm nhập trái phép.

Nghiên cứu các ứng dụng đang sử dụng hiện nay để phát hiện xâm nhập trái phép.

Tiến hành hệ thống thử nghiệm.

6. Bố cục luận văn

Ngoài phần mở đầu, mục lục, kết luận và tài liệu tham khảo, nội dung chính của luận án được chia thành 4 chương, cụ thể như sau:

Chương 1 Tổng quan các phương pháp phát hiện và phòng chống xâm nhập mạng

Chương 2 Các công trình liên quan

Chương 3 Hệ thống phát hiện và phòng chống xâm nhập mạng

Chương 4 Xây dựng và triển khai hệ thống phát hiện xâm nhập mạng dựa vào học máy cho hệ thống mạng trung tâm y tế huyện Gò Dầu.

PHẦN NỘI DUNG

CHƯƠNG 1. TỔNG QUAN CÁC PHƯƠNG PHÁP PHÁT HIỆN VÀ PHÒNG CHỐNG XÂM NHẬP MẠNG

Tổng quan đề tài

Phát hiện xâm nhập là một tập hợp các kỹ thuật và phương pháp dùng để dò tìm những hoạt động đáng nghi ngờ trên mạng. Một hệ thống phát hiện xâm nhập được định nghĩa là một tập hợp các công cụ, phương thức, và tài nguyên giúp người quản trị xác định, đánh giá, và báo cáo hoạt động không được phép trên mạng.

Phát hiện xâm nhập được xem là một tiến trình được quyết định khi một người không xác thực đang cố gắng để xâm nhập hệ thống mạng trái phép. Hệ thống phát hiện xâm nhập sẽ kiểm tra tất cả các gói tin đi qua hệ thống và quyết định gói tin đó có vấn đề khả nghi hay không. Hệ thống phát hiện xâm nhập được trang bị hàng triệu tình huống để nhận dạng tấn công và được cập nhật thường xuyên. Chúng thực sự quan trọng và là lựa chọn hàng đầu để phòng thủ trong việc phát hiện và phòng chống xâm nhập mạng.

Việc nghiên cứu xây dựng hệ thống phát hiện và phòng chống xâm nhập (IDS/IPS) đang được phát triển mạnh và còn phát triển mạnh mẽ trong thời gian tới. Các sản phẩm thương mại trên thị trường có chi phí rất lớn, vượt quá khả năng đầu tư của nhiều doanh nghiệp. Bên cạnh đó, các nghiên cứu về mã nguồn mở cũng đã được đầu tư nghiên cứu và triển khai. Có nhiều đề tài trong nước nghiên cứu liên quan đến IDS/IPS bằng mã nguồn mở chủ yếu tập trung vào Snort. Nhìn chung chưa được áp dụng rộng rãi, còn tồn tại nhiều hạn chế như: do chương trình mã nguồn mở nên hầu hết không có giao diện thân thiện; thành

phần báo động không được tích hợp sẵn, hoặc nếu có cũng chỉ qua giao diện console, hoặc qua giao diện Web chưa tạo được sự linh động và tiện dụng cho người quản trị mạng; phần mềm mang tính đơn lẻ (chỉ tập trung nghiên cứu về Snort) trong khi nhu cầu tích hợp nhiều tính năng giám sát khác để nâng cao hiệu quả sử dụng chưa được chú trọng và phát triển. Hơn nữa, các dấu hiệu của các kiểu tấn công ngày một tinh vi phức tạp đòi hỏi hệ thống phát hiện và phòng chống xâm nhập (IDS/IPS) phải được thường xuyên cập nhật những dấu hiệu mới. Người quản trị mạng còn có thể dựa vào những phân tích khác như những dấu hiệu bất thường về lưu lượng ra vào hệ thống, hoạt động của CPU, RAM... để có những phản ứng kịp thời. Bên cạnh đó, hệ thống báo động cũng cần triển khai mang tính chất đa dạng nhiều hình thức, linh động, tiện dụng thực sự hỗ trợ thiết thực cho người quản trị mạng. Các nghiên cứu đã chứng minh rằng hầu hết các hệ thống có đặc điểm chung là tính đa dạng và thay đổi. Việc nghiên cứu và triển khai một hệ thống giám sát mạng, phát hiện và phòng chống xâm nhập với các yếu tố: chính xác, nhanh chóng, trực quan, linh động và tiện lợi là vấn đề cấp thiết trong thực tế.

Phát triển hệ thống giám sát trực quan theo dõi các diễn biến trên mạng như lưu lượng ra vào một Server, Switch, … hay hoạt động của CPU, bộ nhớ,… giúp người quản trị mạng có những phân tích để đưa ra ứng phó kịp thời.

Hệ thống phát hiện xâm nhập dựa vào những mẫu dấu hiệu tấn công triển khai để giúp phát hiện nhanh các cuộc tấn công mạng. Hệ thống phát hiện này kết hợp với tường lửa sẽ chống lại các cuộc

tấn công xâm nhập. Tuy nhiên, các dấu hiệu của các kiểu tấn công ngày một tinh vi phức tạp thì hệ thống phát hiện phải được thường xuyên cập nhật những dấu hiệu mới. Để có thể phát hiện nhanh chóng các bất thường trên mạng, người quản trị mạng còn có thể dựa vào những đồ thị trực quan về lưu lượng ra vào hệ thống để có những phản ứng kịp thời.

Hệ thống báo động cũng cần triển khai để thông báo cho người quản trị trong một số trường hợp: Server ngưng hoạt động, một dịch vụ mạng ngưng hoạt động hay có tấn công mạng. Hệ thống báo động có thể được triển khai qua nhiều hình thức để phát báo động như: bằng Web, E-mail hay qua tin nhắn SMS đến người quản trị mạng.

1.2 Các hình thức tấn công mạng

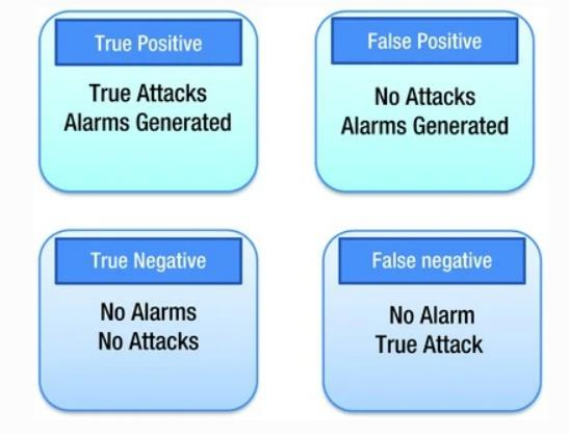
Có nhiều hình thức tấn công mạng khác nhau, tuy nhiên một số loại tấn công vẫn còn được sử dụng đến thời điểm hiện nay như:

- **Tấn công điểm cuối (Endpoint attacks)**:
- **Tấn công bằng các phần mềm độc hại (Malware attacks)**:
- **Các lỗ hổng, khai thác và tấn công (Vulnerabilities, exploits and attacks)**:.
- **Các mối đe dọa dai dẳng nâng cao (Advanced persistent threats)**,
- **Truy cập trái phép (Unauthorized access)**:
- **Từ chối dịch vụ (Distributed Denial of Service - (DDoS)): Những kẻ tấn**
- **Tấn công Man-in-the-Middle (MitM)**:
- **Tấn công SQL Injection**:

- **Tấn công leo thang đặc quyền (Privilege Escalation)**:
- **Tấn công nội bộ (Insider threats)**:

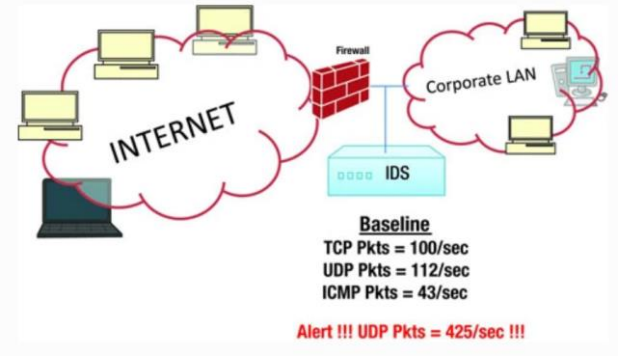
  khác.

1.3.2 Dựa trên các cảnh báo từ hệ thống IDS



Hình 1.1: Định nghĩa các cảnh báo trong hệ thống IDS [1]

1.3.3 Phát hiện dựa trên dòng dữ liệu bất thường



Hình 1.2: Phát hiện bất thường [1]

Các loại bất thường

Giao thức bất thường

Phát hiện bất thường thống kê - DDO thống kê

Phát hiện phân tích giao thức trạng thái

1.4 Giải pháp phát hiện và phòng chống xâm nhập

1.4.1 Phân chia mạng

1.4.2 Điều chỉnh quyền truy cập Internet qua máy chủ proxy

1.4.3 Đặt thiết bị bảo mật chính xác

1.4.4 Sử dụng NAT (Network Address Translation)

1.4.5 Giám sát lưu lượng mạng

1.4.6 Sử dụng công nghệ "đánh lừa"

1.5 Các thuật toán học máy trong hệ thống phát hiện xâm nhập mạng

1.5.1 Decision Tree (DT)

1.5.2 K-Nearest Neighbor (KNN)

1.5.3 Support vector machine

1.5.4 K-mean clustering

1.5.5 Artificial neural network

1.5.6 Ensemble methods

CHƯƠNG 2: CÁC CÔNG TRÌNH LIÊN QUAN

2.1 Một số công trình nghiên cứu tại Việt Nam

2.2 Một số công trình nghiên cứu trên thế giới

2.3 Kết luận chương

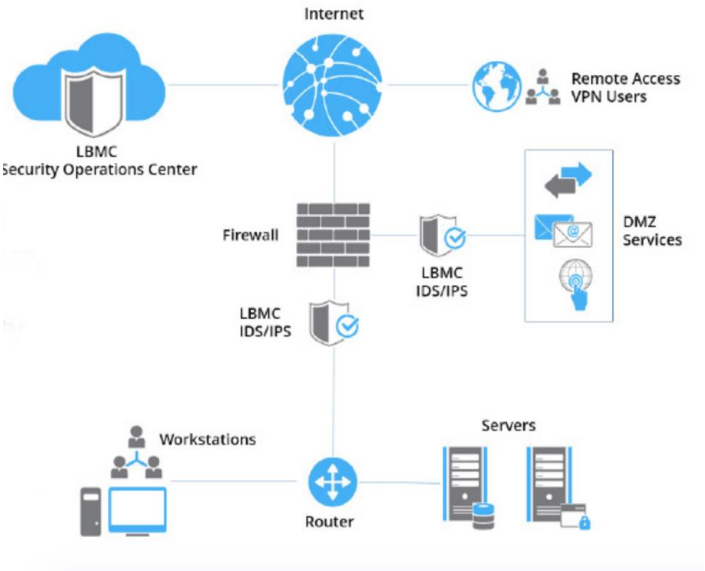
CHƯƠG 3. HỆ THỐNG PHÁT HIỆN VÀ PHÒNG CHỐNG XÂM NHẬP MẠNG

3.1 Vai trò và chức năng của hệ thống phát hiện và phòng chống xâm nhập

3.1.1 Quá trình phát triển

3.1.2 Vai trò và chức năng của hệ thống xâm nhập

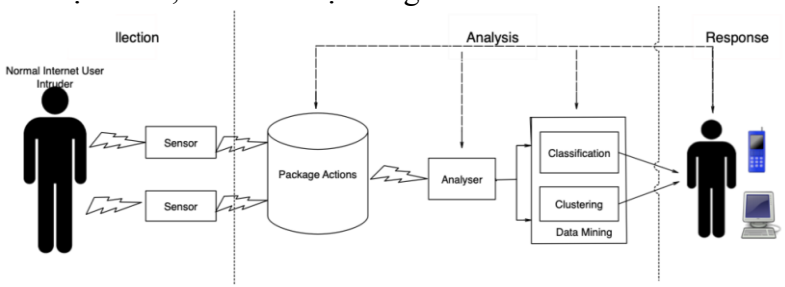
Cách thức hoạt động của IDS



Hình 0.1: Cách thức hoạt động của hệ thống IDS

Chức năng IDS

3.2 Đặc điểm, kiến trúc hệ thống của IDS



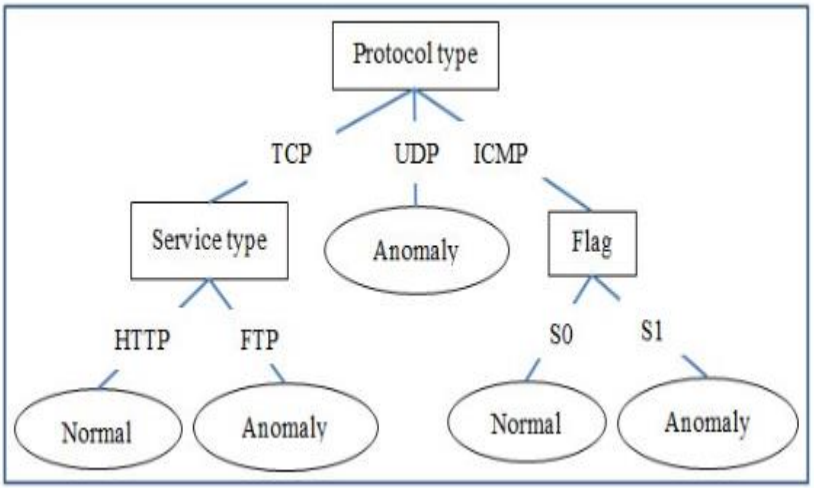
Hình 0.2: Kiến trúc của hệ thống phát hiện xâm nhập [15]

3.3 Các mô hình sử dụng cho hệ thống IDS

3.3.1 Mô hình Decision Tree

Một số lợi thế của cây quyết định là:

Những điểm yếu của cây quyết định bao gồm:

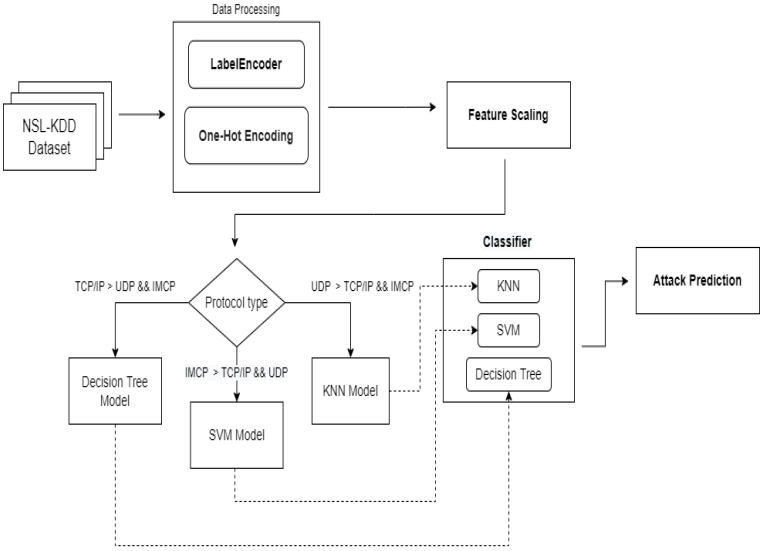


Hình 0.3: Sơ đồ cây quyết định [16]

3.3.2 Mô hình KNN

3.3.3 Mô hình máy Vector hỗ trợ (SVM)

4 Mô hình IDS đề xuất



Hình 0.4: Mô hình đề xuất của luận văn

5 Xử lý dữ liệu

Label Encoding [21]

Hình 0.5: Cách hoạt động của Label Encoding

One hot encoding [22]

Type
AA
AB
CD
AA

Onehot encoding →

Type	AA_Onehot	AB_Onehot	CD_Onehot
AA	1	0	0
AB	0	1	0
CD	0	0	1
AA	0	0	0

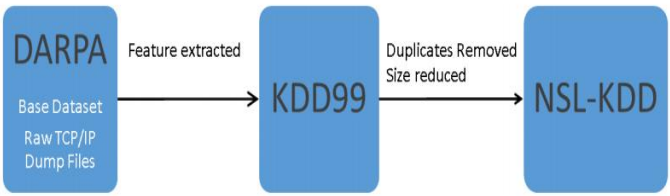
Hình 0.6: Cách One hot encoding biến đổi dữ liệu

CHƯƠNG 4 XÂY DỰNG VÀ TRIỂN KHAI HỆ THỐNG PHÁT HIỆN XÂM NHẬP MẠNG DỰA VÀO HỌC MÁY CHO HỆ THỐNG MẠNG TRUNG TÂM Y TẾ HUYỆN GÒ DẦU

Mô tả bộ dữ liệu sử dụng NSL-KDD

Quá trình hình thành

Trong quá trình nghiên cứu về lĩnh vực bảo mật thì cái tên DARPA không quá xa lạ với mọi người. DARPA là một bộ dữ liệu thô cơ sở. Trong khi đó, KDD99 là phiên bản trích xuất tính năng của bộ dữ liệu DARPA. Tiếp đó, NSL-KDD là phiên bản được loại bỏ và giảm kích thước của bộ dữ liệu KDD99, đồng thời là bộ dữ liệu chuẩn cho dữ liệu Internet hiện nay.



Hình 0.1: Mối quan hệ giữa các bộ dữ liệu cho hệ thống IDS: DARPA, KDD99, NSL-KDD

Mô tả bộ dữ liệu

NSL-KDD

4.2 Môi trường mô phỏng quá trình thực nghiệm

Dựa vào dữ liệu và các thuật toán có sẵn được sử dụng cho hệ thống IDS, luận văn đề xuất mô hình phân loại các thuật toán dựa trên các giao thức truyền dẫn thông tin. Tiến hành áp dụng các thuật toán học máy trên bộ dữ liệu NSL-KDD, đánh giá các kết quả đạt được và từ đó đưa ra sự phân loại phù hợp cho các tập dữ liệu.

Cài đặt thuật toán SVM, KNN, Decision Tree trên môi trường Google Collab và kiểm nghiệm kết quả.

5 Kết quả thực nghiệm

Như đã trình bày ở trên, luận văn sẽ sử dụng ba thuật toán phân lớp là KNN, Decision Tree và SVM để tiến hành thực nghiệm trên bộ dữ liệu NSL-KDD. Ngoài ra, kết quả thu được sẽ tiến hành phân tích các độ đo đánh giá hiệu quả mô hình, từ đó kết hợp với quá trình chọn mô hình thích hợp cho bộ dữ liệu dựa vào số lượng của từng loại giao thức trong bộ dữ liệu, cụ thể là trường protocol_type.

Trước tiên, ta chạy thực nghiệm thuật toán KNN với bộ dữ liệu NSL-KDD, kết quả được ghi nhận như sau:

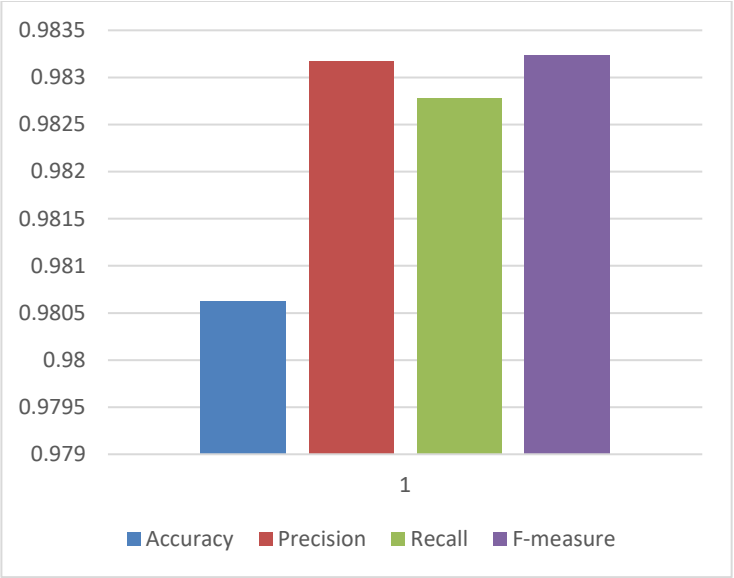
Bảng 0.1: Các độ đo accuracy, precision, recall, f-measure, traning time, testing time của thuật toán KNN trên tập dữ liệu NSL-KDD

k value	Accuracy	Precision	Recall	F-measure	Training time (s)	Testing time (s)
k = 3	0.95968	0.95506	0.97506	0.96495	0.380	11.477
k = 5	0.95755	0.95021	0.97662	0.96322	0.383	11.191
k = 10	0.95484	0.94720	0.97506	0.96091	0.390	12.849

Bảng 4.1 ghi nhận các kết quả thu được sau khi thực nghiệm tập dữ liệu với thuật toán KNN. Kết quả cho thấy thuật toán KNN có đầu ra tốt nhất tại n = 3 với độ chính xác 95.97%. Các độ đo hiệu quả mô hình còn lại cũng đạt các giá trị cao tương tự Accuracy.

Bảng 0.2 Các độ đo accuracy, precision, recall, f-measure, traning time, testing time của thuật toán Decision Tree trên tập dữ liệu NSL-KDD

Accuracy	Precision	Recall	F-measure	Training time (s)	Testing time (s)
0.98062	0.98317	0.98278	0.98324	0.557	0.004



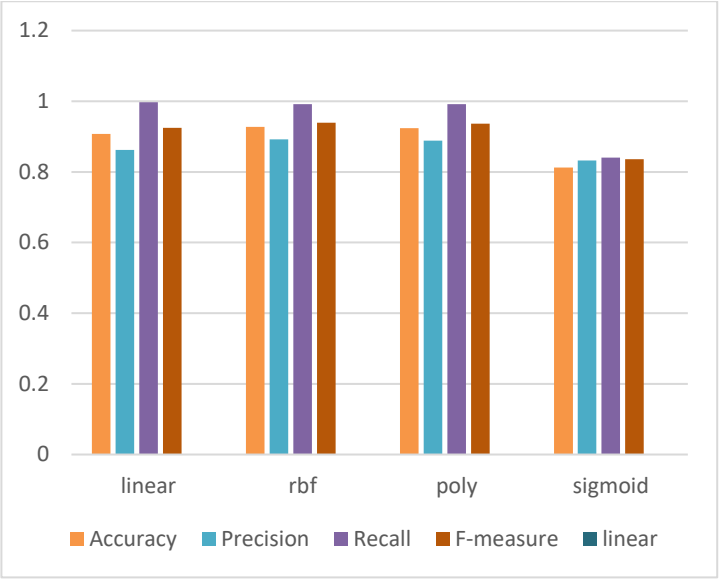
Hình 0.2: Biểu đồ thể hiện độ đo khi áp dụng thuật toán Decision Tree

trên tập dữ liệu NSL-KDD

Thuật toán Decision Tree ghi nhận các kết quả vượt trội, với độ chính xác (Accuracy) đạt 98.06%, hơn khoảng 2.09% so với thuật toán KNN. Ngoài ra, thời thử nghiệm (Testing time) của thuật toán cũng ghi nhận kết quả ấn tượng chỉ với 40 mili giây.

Bảng 0.3 Các độ đo accuracy, precision, recall, f-measure, traning time, testing time của thuật toán SVM trên tập dữ liệu NSL-KDD

Measure	Kernel			
	linear	**rbf**	**poly**	**sigmoid**
Accuracy	0.90751	0.92716	0.92366	0.81263
Precision	0.86203	0.89232	0.88810	0.83241
Recall	0.99735	0.99182	0.99182	0.84018
F-measure	0.92472	0.93942	0.93663	0.83622
Training time (s)	449.304s	91.381	87.186	465.570s
Testing time (s)	11.112s	11.467	6.117	24.432s

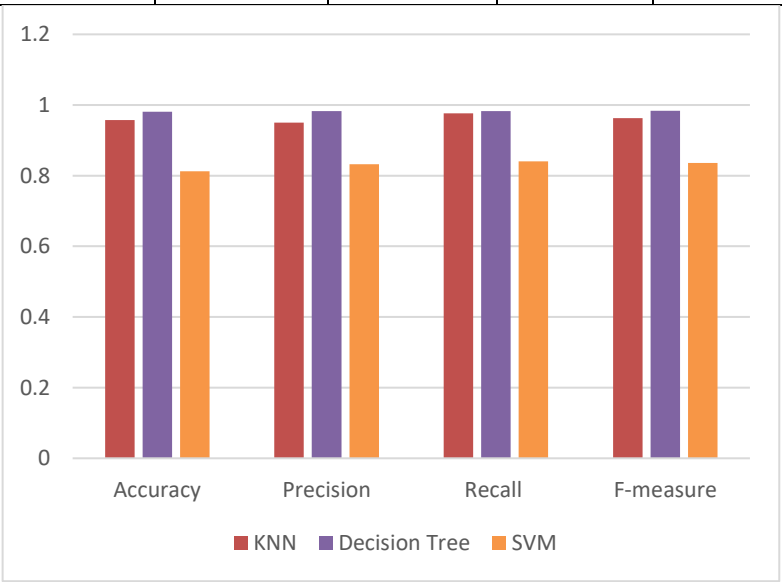


Hình 0.3: Biểu đồ thể hiện độ đo khi áp dụng thuật toán SVM trên tập dữ liệu NSL-KDD

Bảng 4.3 mô tả các kết quả thu được với từng loại kernel khác nhau của thuật toán SVM. Qua đó có thể thấy rõ được rằng ở kernel rbf ghi nhận kết quả tốt nhất với độ chính xác đạt 92.71%. Tuy nhiên, nếu vừa xét về thời gian và độ chính xác thì với kernel là poly vượt trội hơn so với rbf. Thời gian training và testing ở kernel poly lần lượt là 87.186 và 6.117 giây, đồng thời độ chính xác chỉ kém 0.35% so với kernel rbf.

Bảng 0.4: Các độ đo accuracy, precision, recall, f-measure, traning time, testing time tổng hợp từ ba thuật toán trên tập dữ liệu NSL-KDD

Classifier	Accuracy	Precision	Recall	F-measure
KNN	0.95755	0.95021	0.97662	0.96322
Decision Tree	0.98062	0.98317	0.98278	0.98324
SVM	0.81263	0.83241	0.84018	0.83622



Hình 0.4: Biểu đồ thể hiện tổng hợp độ đo khi áp dụng các thuật toán trên tập dữ liệu NSL-KDD

Bảng 4.4 cho thấy kết quả tổng hợp của ba thuật toán KNN,

Decision Tree và SVM. Kết quả tổng hợp được lấy kết quả tốt nhất mỗi lần chạy của từng thuật toán. Dựa vào bảng tổng hợp kết quả, có thể thấy được Decision Tree ghi nhận các thông số đầu ra vượt trội hơn so với hai thuật toán còn lại. Dựa các kết quả này, mô hình đề xuất sẽ có thể triển khai và áp dụng vào việc lựa chọn mô hình cho các bộ dữ liệu sau.

Xét bộ dữ liệu NSL-KDD, trường dữ liệu protocol_type có tổng cộng ba loại giao thức là TCP, UDP và IMCP. Trong đó, số lượng của từng loại giao thức được mô tả cụ thể như sau: TCP, UDP, IMCP có số lượng lần lượt là 18880, 2621 và 1043. Theo kết quả có được, với dữ liệu có số lượng giao thức TCP lớn hơn số lượng các loại giao thức còn lại thì Decision Tree sẽ là mô hình thích hợp cho bộ dữ liệu. Tương tự như vậy, đối với dữ liệu có số lượng UDP lớn hơn số lượng các giao thức còn lại thì KNN sẽ là mô hình phù hợp. Và cuối cùng là ứng với dữ liệu có số lượng IMCP lớn hơn số lượng các giao thức còn lại thì SVM sẽ là mô hình phù hợp.

6 Kết luận chương

Chương 4 đã giới thiệu và mô tả quá trình thực nghiệm trên bộ dữ liệu NSL-KDD - bộ dữ liệu được lấy làm chuẩn để giúp các nhà nghiên cứu so sánh các phương thức phát hiện xâm nhập khác nhau. Các kết quả được nhận xét, phân tích và đánh giá. Đồng thời, chương 4 cũng đã đề xuất phương pháp chọn lọc hệ thống phù hợp với trường dữ liệu protocol_type của mỗi dịch vụ, xem xét chọn trường dữ liệu phù hợp, từ đó chọn mô hình tốt nhất cho bộ dữ liệu.

KẾT LUẬN

1. Kết quả nghiên cứu của đề tài

Trên thế giới, các công trình nghiên cứu về lĩnh vực an ninh, an toàn thông tin được tiến hành bởi các trường đại học về khoa học máy tính, công nghệ thông tin, nhưng những công trình nổi bật và được ứng dụng rộng rãi là những công trình nghiên cứu được tiến hành bởi các công ty bảo mật. Một số công ty bảo mật hàng đầu phát triển các bộ giải pháp về an ninh mạng có thể kể đến là: Cisco System, Juniper Network, TrendMicro.. Sản phẩm của nghiên cứu mà các công ty bảo mật tiến hành có thể là những sản phẩm riêng rẽ như Cisco IDS, Juniper IPS, TrendMicro Server Protect, hoặc cũng có thể là cả một bộ giải pháp với nhiều sản phẩm phần cứng phần mềm khác nhau.

Về việc nghiên cứu phát triển các hệ thống phát hiện xâm nhập, có một số xu hướng nghiên cứu về lĩnh vực này đã được thực hiện. Đó là phát hiện xâm nhập dựa trên dấu hiệu xâm nhập; phát hiện xâm nhập dựa vào khả năng tự học của hệthống; phát hiện xâm nhập bằng cách kết hợp cả hai phương pháp trên. Đối với mỗi phương pháp phát hiện xâm nhập sẽ dẫn đến rất nhiều nghiên cứu, như những nghiên cứu về trí tuệ nhân tạo, hệ chuyên gia, nghiên cứu về phương pháp đặt luật so sánh trong phân tích lưu thông mạng.

Như vậy, an ninh thông tin là lĩnh vực ngày càng có thêm nhiều công trình nghiên cứu cũng như các sản phẩm được tạo ra từ các công trình đó, bởi trình độ của các hacker cũng ngày càng tiến bộ. Tuy rằng các sản phẩm an ninh thông tin mà các công ty nước ngoài phát triển hoạt động hiệu quả, nhưng các sản phẩm này có mức giá cao và hàng năm phải tốn chi phí để cập nhật. Việc áp dụng các sản phẩm này

một cách thụ động cũng là một nhược điểm. Vì khi đó đội ngũ quản trị mạng sẽ không thực sự hiểu bản chất hệ thống hoạt động như thế nào, hệ thống phân tích những gì ở mức dưới của hệ thống thông tin, dẫn đến việc không linh hoạt trong nghiệp vụ quản trị bảo mật.

Với việc hoàn thành các sản phẩm của luận văn là hệ thống phát hiện xâm nhập mạng dựa trên nền tảng mã nguồn mở và bộ quy trình phòng ngừa và ngăn chặn xâm nhập mạng, hy vọng rằng việc ứng dụng các sản phẩm này sẽ góp phần cải thiện những điểm yếu trong hệ thống an ninh thông tin của Trung tâm Y tế huyện Gò Dầu. Bên cạnh đó, nó sẽ mở ra những hướng phát triển tiếp theo trong nghiên cứu và ứng dụng hệ thống phát hiện xâm nhập mạng, giúp nền công nghệ thông tin nước ta có những bước tiến trong ứng dụng và làm chủ những sản phẩm công nghệ về an ninh thông tin.

2. Hạn chế của luận văn

Bên cạnh các kết quả đạt được, luận văn cũng còn một số hạn chế nhất định. Trong đó, dữ liệu sử dụng chưa phải là dữ liệu thực tế từ hệ thống IDS, từ đó chưa thể áp dụng được mô hình đề xuất vào bộ dữ liệu thực tế để kiểm tra cũng như đánh giá kết quả một cách trực quan nhất. Ngoài ra, hiện nay một số hệ thống IDS đã áp dụng các thuật toán học sâu (một hướng nghiên cứu phát triển từ học máy) tối ưu hơn, nhằm mục đích cải thiện các tốc độ xử lý dữ liệu cũng như phát hiện được những loại tấn công mạng nguy hiểm hơn.

3. Hướng phát triển của luận văn

Áp dụng các kỹ thuật cũng như thuật toán tối ưu hơn về tốc độ xử lý, thời gian huấn luyện và thử nghiệm cho mô hình. Ngoài ra, dữ liệu sử dụng sẽ thay thế bằng các tập dữ liệu thực tế được lấy từ

một hệ thống IDS đang hoạt động bất kỳ, từ đó có thể đưa ra các kết quả cũng như đánh giá khách quan nhất về mức độ hiệu quả của hệ thống phát hiện xâm nhập mạng.