**Đề Tài 8: PHÁT HIỆN XÂM NHẬP IOT BẰNG PHƯƠNG PHÁP HỌC SÂU**

## Tổng quan đề tài

### Xâm nhập mạng là gì?

Xâm nhập mạng [1] đề cập đến bất kỳ hoạt động trái phép nào trên mạng. Xâm nhập mạng có hai loại, một là xâm nhập mạng để kiểm thử hay còn gọi là pentest, hai là tấn công mạng. Kiểm thử xâm nhập là hành động xâm nhập nhằm tìm ra lỗ hổng bảo mật trong hệ thống để khắc phục. Sau khi tìm ra các lỗ hổng nghiêm trọng hoặc nguy hiểm, người kiểm thử sẽ đưa ra giải pháp để vá lỗ hổng nhằm bảo vệ tài nguyên của các tổ chức được bảo mật. Ngược lại, tấn công mạng là hành động lợi dụng mạng Internet để xâm nhập trái phép vào hệ thống website của cá nhân, tổ chức hay máy tính hoặc nguy hiểm hơn là cơ sở dữ liệu, … Mục đích chính của việc này là để đánh cắp thông tin, mã hóa dữ liệu hoặc làm gián đoạn các hoạt động vận hành của tổ chức cũng như doanh nghiệp. Những đối tượng của các cuộc tấn công này có thể là cá nhân, tổ chức, doanh nghiệp lớn nhỏ hoặc các cơ quan nhà nước. Đặc biệt, đối tượng phổ biến nhất bị nhắm đến là các doanh nghiệp bởi vì mục tiêu chính của những kẻ tấn công này là lợi nhuận. Những tin tặc tấn công chủ yếu nhằm trục lợi phi pháp, tống tiền các doanh nghiệp. Ngoài ra, cũng tồn tại một số doanh nghiệp cạnh tranh không lành mạnh tấn công mạng đối thủ của mình bằng cách đánh sập website của họ. Điều này gây ảnh hưởng ít nhiều đến hình ảnh và uy tín của công ty, thương hiệu.

### Phát hiện xâm nhập mạng

Phát hiện xâm nhập mạng [1] là tập hợp các phương pháp và kỹ thuật dùng để dò tìm những hoạt động có dấu hiệu bất thường, đáng nghi ngờ trên mạng. Một tập hợp các phương thức, công cụ, và tài nguyên giúp cho người quản trị xác định, đánh giá, và báo cáo các hoạt động không được cho phép trên mạng được định nghĩa là hệ thống phát hiện xâm nhập (IDS). Phát hiện xâm nhập được hiểu là một tiến trình mà nó được quyết định khi một người dùng chưa được xác thực cố gắng xâm nhập trái phép vào hệ thống mạng. IDS với hàng triệu tình huống được trang bị để nhận dạng

tấn công cũng như cập nhật thường xuyên sẽ kiểm tra tất cả các gói tin đi qua nó và quyết định liệu gói tin này có vấn đề gì khả nghi hay không. Vì thế, hệ thống này đã trở nên thực sự quan trọng, đồng thời là lựa chọn hàng đầu trong việc phát hiện và phòng chống xâm nhập mạng.

Công cuộc nghiên cứu và xây dựng hệ thống ngăn chặn và phát hiện xâm nhập (IDS/IPS) đang được nhiều người quan tâm cũng như phát triển và sẽ không ngừng phát triển mạnh mẽ trong thời gian tới. Những sản phẩm thương mại hiện tại trên thị trường đều có chi phí cao, vượt ngoài khả năng đầu tư của nhiều doanh nghiệp lớn nhỏ dẫn đến hệ quả là những nghiên cứu mã nguồn mở về IDS/IPS được đầu tư nghiên cứu và triển khai. Nhìn chung, những nghiên cứu trong nước liên quan đến IDS/IPS bằng mã nguồn mở tuy nhiều nhưng lại tập trung vào Snort. Ngoài ra, các nghiên cứu này vẫn chưa được áp dụng rộng rãi, tồn đọng nhiều hạn chế như: không có giao diện thân thiện; chưa được tích hợp sẵn thành phần báo động, hoặc có thì cũng chỉ giới hạn qua giao diện console hay qua giao diện Web, chính vì thế nó chưa tạo được sự linh động và tiện dụng cho người dùng; do chỉ tập trung nghiên cứu về Snort làm cho các phần mềm mang tính chất đơn lẻ mặc dù nhu cầu tích hợp các tính năng giám sát khác nhằm đạt được hiệu quả sử dụng cao hơn chưa thật sự được chú trọng và phát triển.

Mặt khác, IDS/IPS phải luôn trong tình trạng cập nhật những dấu hiệu tấn công mới vì các kiểu tấn công mạng đang ngày một tinh vi và phức tạp kéo theo những dấu hiệu của chúng liên tục thay đổi. Chuyên viên quản trị mạng có thể dựa vào các phân tích khác chẳng hạn như dấu hiệu bất thường của lưu lượng vào hoặc ra khỏi hệ thống, hoạt động của RAM, CPU, ... để phản ứng kịp thời. Thêm vào đó, hệ thống báo động cần triển khai phải mang tính chất đa dạng với nhiều hình thức, tiện dụng và linh động sẽ mang lại sự hỗ trợ thiết thực cho nhân viên quản trị mạng. Hầu hết các hệ thống có hai đặc điểm chung là tính đa dạng và thay đổi, điều này đã được các nhà nghiên cứu tìm hiểu và chứng minh. Công việc nghiên cứu và triển khai hệ thống giám sát, phát hiện và phòng chống xâm nhập mạng tồn tại các yếu tố: nhanh chóng, chính xác, trực quan, tiện lợi và linh động là một vấn đề cấp thiết trong thực tế.

Phát triển một hệ thống giám sát trực quan để theo dõi các diễn biến trên mạng như: lưu lượng ra/vào một Server, Switch hoặc hoạt động của CPU, bộ nhớ,… sẽ giúp

người quản trị mạng có những phân tích rõ ràng, nhanh chóng để đưa ra quyết định ứng phó kịp thời khi bị hệ thống bị xâm nhập.

IDS căn cứ vào các dấu hiệu tấn công được triển khai có thể giúp phát hiện một cách nhanh chóng các cuộc tấn công mạng. Khi kết hợp hệ thống với tường lửa sẽ có thể chống lại các cuộc tấn công xâm nhập. Dù vậy, dấu hiệu của từng kiểu tấn công mỗi ngày một tinh vi phức tạp hơn thì hệ thống càng phải được cập nhật thường xuyên những dấu hiệu mới để có thể phát hiện nhanh chóng các bất thường trên mạng, người quản trị mạng 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 thiết phải triển khai để gửi thông báo cho người quản trị trong một số trường hợp đặc biệt nghiêm trọng như: Server bỗng nhiên ngưng hoạt động, một dịch vụ mạng nào đó ngưng hoạt động hay có tấn công. Hệ thống có thể được triển khai qua nhiều hình thức khác nhau để phát thông báo như: qua Web, E-mail, tin nhắn SMS đến người quản trị mạng.

## Các hình thức tấn công mạng

### Phân loại tấn công mạng

Có nhiều loại tấn công mạng khác nhau theo đó mức độ nguy hiểm cũng khác nhau với từng loại, tuy nhiên có bốn loại tấn công nguy hiểm nhất mang sức mạnh phá hủy hạ tầng mạng một doanh nghiệp như [2]:

## Tấn công từ chối dịch vụ (DDoS)

Đây là loại tấn công rất phổ biến và cũng là một trong những tấn công mạng có khả năng phá hoại website doanh nghiệp lớn nhất. Những cuộc tấn công bằng DDoS đe dọa trực tiếp và ngay lập tức đến sự hoạt động của các trung tâm dữ liệu của các doanh nghiệp. Những kẻ tấn công mạng sẽ có nhiều cơ hội để xây dựng các mạng botnet rộng lớn và sau đó kích hoạt các tấn công DDoS toàn cầu bởi sự gia tăng không ngừng của các thiết bị có kết nối IoT với bảo mật kém. Một chứng minh rõ ràng của điều này là sự cố "tắc nghẽn mạng" vào tháng 6, 2019 đã làm cho Google Cloud phải ngừng hoạt động và nó cũng làm ít nhất 16 sản phẩm vệ tinh của Google phải ngừng hoạt động theo, các sản phẩm này chính là: Gmail, G-Suite, Google Cloud, Google Drive, Google Docs và YouTube…

## Mã hóa dữ liệu tống tiền (tấn công Ransomeware)

Cơ sở hạ tầng doanh nghiệp là mục tiêu được những tội phạm mạng nhắm vào bằng ransomware vì thiệt hại của nó có thể ảnh hưởng lâu dài trên diện rộng. Dẫn chứng cho sự tấn công này là công ty lưu trữ Nayana (Hàn Quốc) bị tấn công bằng ransomware làm cho hàng nghìn website của khách hàng lưu trữ trên máy chủ của công ty bị đình trệ hoạt động suốt nhiều tuần liền. Mặc dù đã chi trả một triệu đô la Mỹ tiền chuộc, song vẫn còn nhiều dịch vụ chưa thật sự được phục hồi. Ngoài việc đe dọa dữ liệu của khách hàng trên máy chủ nhà cung cấp dịch vụ, loại tấn công này còn làm giảm sút lòng tin của khách hàng vào doanh nghiệp kinh doanh. Không chỉ có nguy cơ bị phát tán dữ liệu một cách công khai ở khắp mọi nơi, dữ liệu còn bị thay đổi, làm đe dọa đến tính toàn vẹn vốn có của dữ liệu.

## Tấn công từ bên thứ ba

Vì có thêm một bên thứ ba khác mà các nhà cung cấp dịch vụ bảo mật thường hay gặp phải sự cố khi quản lý bảo mật cho trung tâm dữ liệu của khách hàng. Với mục đích gây hại cho các doanh nghiệp, những kẻ tấn công thường nhắm đến mục tiêu là những nhà cung cấp dịch vụ bảo mật. Cụ thể là vào năm 2019, tổ chức của NordVPN (công ty cung cấp dịch vụ mạng riêng ảo có tính năng bảo mật hàng đầu được đông đảo các doanh nghiệp sử dụng để bảo vệ những dữ liệu nhạy cảm) đã thừa nhận việc một trong số những trung tâm dữ liệu của họ bị tấn công vào năm 2018. Sự việc xảy ra khi họ cài đặt hệ thống truy cập từ xa bởi một bên thứ ba mà chưa thông báo nó cho khách hàng làm dẫn đến tình trạng máy chủ thiếu an toàn.

## Tấn công vào ứng dụng web do mật khẩu yếu

Tuy không gây ảnh hưởng trực tiếp đến những dịch vụ của trung tâm dữ liệu nhưng bù lại những tấn công vào các ứng dụng web hoặc máy chủ, cụ thể như các form đăng nhập vào ứng dụng web vẫn có thể gây ra gián đoạn các dịch vụ web. Việc đặt mật khẩu yếu, kém hoặc mật khẩu dễ đoán của người dùng là nguyên nhân cho những loại tấn công này. Chúng có thể được thực hiện thông qua các cuộc tấn công đến từ ứng dụng web. Với những mục được nhắm từ trước việc tấn công tốn ít băng thông hơn nhưng nó vẫn có thể khiến các dịch vụ web ngừng hoạt động một cách dễ dàng.

Ngoài ra vẫn còn một số loại tấn công khác được sử dụng đến thời điểm hiện nay như:

* **Tấn công điểm cuối (Endpoint attacks)**: Các cuộc tấn công điểm cuối đã đạt được quyền truy cập trái phép vào các thiết bị người dùng, máy chủ hoặc các điểm cuối khác. Loại hình tấn công này gây ảnh hưởng đến các thiết bị khác bằng cách lây nhiễm bằng phần mềm độc hại.
* **Tấn công bằng các phần mềm độc hại (Malware attacks)**: loại tấn công này lây nhiễm tài nguyên CNTT với phần mềm độc hại, cho phép kẻ tấn công thỏa hiệp với các hệ thống, đánh cắp dữ liệu và gây tổn hại đến hại thống. Chúng bao gồm các cuộc tấn công Ransomware (mã độc tống tiền), Spyware (phần mềm gián điệp), Virus, Worm (phần mềm độc hại lây lan với tốc độ nhanh).
* **Các lỗ hổng, khai thác và tấn công (Vulnerabilities, exploits and attacks)**: các lỗ hổng khai thác trong phần mềm được sử dụng trong tổ chức, để có quyền truy cập trái phép, thỏa hiệp hoặc hệ thống phá hoại.
* **Chiến dịch tấn công sử dụng kỹ thuật cao (Advanced persistent threats)**, đây là những mối đe dọa nhiều lớp phức tạp, không chỉ bao gồm các cuộc tấn công mạng mà còn cả các loại tấn công khác.

Trong một cuộc tấn công mạng, những kẻ tấn công tập trung vào việc thâm nhập chu vi mạng của công ty và đạt được quyền truy cập vào các hệ thống nội bộ. Thông thường, khi kẻ tấn công vào được bên trong hệ thống chúng sẽ kết hợp các loại tấn công khác, ví dụ như ảnh hưởng đến điểm cuối, lan truyền phần mềm độc hại hoặc khai thác lỗ hổng trong một hệ thống trong mạng.

### Các kỹ thuật tấn công

Bên cạnh các loại tấn công trên, các kỹ thuật tấn công mạng cũng chuyển biến khôn lường từ các kỹ thuật cơ bản, tuy nhiên mức độ nguy hiểm khi bị tấn công khó lường trước được và gây ra những hậu quả nghiêm trọng nếu không được phát hiện kịp thời. Một số các kỹ thuật tấn công mạng phổ biến như [3]:

* **Truy cập trái phép (Unauthorized access)**: Truy cập trái phép đề cập đến những kẻ tấn công truy cập mạng mà không nhận được sự cho phép. Các nguyên nhân của cuộc tấn công truy cập trái phép thường là do mật khẩu yếu,

thiếu bảo vệ các tài khoản xã hội, các tài khoản bị xâm nhập trước đó và các mối đe dọa nội bộ.

* **Từ chối dịch vụ (Distributed Denial of Service - DDoS):** Những kẻ tấn công xây dựng botnet và sử dụng chúng để hướng lưu lượng truy cập sai vào mạng hoặc máy chủ. DDoS có thể xảy ra ở cấp độ mạng, ví dụ bằng cách gửi khối lượng khổng lồ của các gói SYN / ACK có thể áp đảo một máy chủ hoặc ở cấp độ ứng dụng.
* **Tấn công Man-in-the-Middle (MitM)**: loại tấn công này liên quan đến những kẻ tấn công chặn lưu lượng, giữa các trang web và trang bên ngoài hoặc bên trong mạng. Nếu các giao thức truyền thông không được bảo mật hoặc kẻ tấn công tìm cách vượt qua bảo mật, họ có thể đánh cắp dữ liệu đang được truyền đi, có được thông tin đăng nhập của người dùng và chiếm quyền điều khiển tài khoản của họ.
* **Tấn công cơ sở dữ liệu (SQL Injection)**: Nhiều trang web chấp nhận đầu vào của người dùng và không xác thực các đầu vào đó. Những kẻ tấn công sau đó có thể điền vào một biểu mẫu hoặc thực hiện cuộc gọi API, truyền mã độc thay vì các giá trị dữ liệu dự kiến. Mã được thực thi trên máy chủ và cho phép kẻ tấn công lợi dụng nó.
* **Tấn công leo thang đặc quyền (Privilege Escalation)**: Khi kẻ tấn công xâm nhập mạng, họ có thể sử dụng sự leo thang đặc quyền để mở rộng tầm với của họ. Sự leo thang đặc quyền ngang liên quan đến những kẻ tấn công đạt được quyền truy cập vào các hệ thống bổ sung, liền kề và leo thang thẳng đứng có nghĩa là những kẻ tấn công có được một mức đặc quyền cao hơn cho cùng một hệ thống.
* **Tấn công nội bộ (Insider threats)**: hay còn gọi là mối đe dọa nội bộ. Đây là mối đe dọa đối với tổ chức đến từ những người thuộc tổ chức đó, chẳng hạn như nhân viên, nhân viên cũ, nhà thầu hoặc cộng sự kinh doanh, những người có thông tin nội bộ liên quan đến các phương thức bảo mật, dữ liệu và hệ thống máy tính của tổ chức. Các mối đe dọa nội bộ có thể khó phát hiện và bảo vệ chống lại, vì người trong cuộc không cần thâm nhập mạng để gây hại. Các công nghệ mới như phân tích hành vi người dùng (UEBA) có thể giúp xác

định hành vi đáng ngờ hoặc bất thường của người dùng nội bộ, có thể giúp xác định các cuộc tấn công nội bộ.

## Dấu hiệu nhận diện một cuộc tấn công mạng

Theo tài liệu [3] một cuộc tấn công mạng có thể được nhận biết thông qua các dấu hiệu sau

### Dựa vào các gói tin (packets)

Nguồn gốc và biểu hiện của một gói tin có thể giúp dự đoán những kết quả của một gói tin được phân phối. Một gói có thể mang một đoạn mã không giống phần mềm độc hại, nhưng khi được lắp lại với các thành phần khác, tạo thành một chương trình phần mềm độc hại. Hoặc một gói có thể chứa dữ liệu để đột nhập vào hệ thống với ID giả có thể không bị phát hiện trong nhiều tuần hoặc vài tháng. Người sở hữu phần mềm độc hại sử dụng tất cả các loại thủ thuật để tránh bị phát hiện.

Bằng cách nhìn vào lưu lượng gói, có thể theo dõi hành vi của gói đó. Ngoài ra, có thể xác định một số loại gói nhất định đang "đánh hơi xung quanh" các khu vực cụ thể vài phút mỗi ngày hoặc vài giờ một tuần. Nếu để ý và quan sát các gói tin, sẽ có thể dễ dàng nhận thấy một cuộc tấn công sắp xảy ra. Thêm vào đó, có thể xác định một gói có khả năng đáng ngờ nếu nó đang giao tiếp với một máy chủ proxy ẩn danh hoặc với các máy chủ nằm ở các quốc gia giả mạo. Lĩnh vực nghiên cứu này, được gọi là phân tích hành vi, đang ngày càng trở nên quan trọng trong an ninh mạng. Các chuyên gia an ninh mạng phải xem xét các kỹ thuật mới để bảo vệ dữ liệu của tổ chức, các nhà cung cấp dịch vụ mạng toàn cầu cần có sự nhìn nhận về chuyển động của một lượng lưu lượng truy cập khổng lồ.

Nghiên cứu các gói tin và lưu lượng truy cập đang trở nên quan trọng hơn khi ngày càng có nhiều các đối tượng, thiết bị và máy móc được kết nối thông qua Internet of Things (IoT). Ngoài máy tính, những kẻ tấn công có thể tìm cách chiếm quyền điều khiển các bộ định tuyến, thiết bị và các thiết bị khác để thực hiện các cuộc tấn công mạng.

Bất cứ ai cũng có thể là một đồng phạm vô tình trong một cuộc tấn công. Một người dùng ngồi vui vẻ trên một tài khoản FaceTime hoặc Skype với đồng nghiệp thậm chí có thể không nhận ra điện thoại thông minh hoặc máy tính xách tay của họ đã bị biến thành một bot tấn công. Người dùng có thể nhận thấy rằng đột nhiên hình

ảnh hoặc video bị mờ đi hoặc dừng lại, hoặc âm thanh bị cắt xén. Trong một số trường hợp, những trục trặc này là tác động của một cuộc tấn công đang được tiến hành.

Tin tặc là chuyên gia tìm kiếm các lỗ hổng trong các thiết bị hàng ngày thậm chí trước khi các nhà sản xuất sản phẩm nhận thấy những sai sót. Và một khi tin tặc tìm thấy lỗ hổng, họ sẽ tìm cách khai thác nó. Khi các nhà phân tích an ninh mạng phát hiện các chỉ số về thỏa hiệp (IOC), họ thông báo cho các mục tiêu - bao gồm các tổ chức không phải là khách hàng. Các công ty không biết họ đang bị tấn công cho đến khi các nhà phân tích an ninh mạng nói với họ rằng, giả sử dữ liệu đang được chuyển từ mạng của họ đến một máy chủ ở Đông Âu hoặc một số điểm đến ngoài ý muốn khác.

### Dựa trên các cảnh báo từ hệ thống IDS

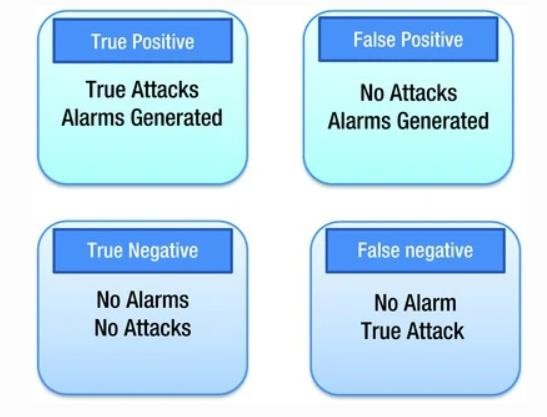
Ngay cả khi có một hệ thống tốt trong việc ghi nhận lịch sử tất cả các lưu lượng truy cập vào mạng nội bộ, thì việc kiểm tra một loạt các lịch sử này chưa mang lại hiệu quả cao. Không thể phân biệt thủ công giữa một gói độc hại và gói mạng tốt. Ngay cả với sự giúp đỡ của máy tính, đây là một công việc chuyên sâu đòi hỏi nhiều sức mạnh xử lý. Trong những năm qua, trong thế giới được kết nối chủ yếu thông qua nhiều phương tiện khác nhau bao gồm máy tính bảng và điện thoại di động, những người xấu với ý định xấu nhắm vào các tập đoàn khác nhau cũng như các cá nhân. Vì không thể phát hiện các cuộc tấn công như vậy theo cách thủ công để ngăn chặn hoặc giảm thiểu chúng, nên bắt buộc phải có một công cụ tự động để giúp giám sát hệ thống cho các cuộc tấn công. IDS đã trở thành một công cụ hữu ích để cung cấp giám sát này. Một số các thuật ngữ cảnh báo trong hệ thống IDS:

Dương tính thật (True Positives): Đây là những cảnh báo rằng một cái gì đó đúng và nó thật sự đúng. Ví dụ: IDS tìm thấy một gói có chứa mã độc và thực sự đúng là gói có mã độc, như được xác nhận bởi điều tra.

Âm tính thật (True Negatives): Đây là những cảnh báo rằng một cái gì đó là không đúng và nó thật sự không đúng. Ví dụ: IDS tìm thấy một gói là không có vấn đề gì và nó thực sự không có vấn đề gì.

Dương tính giả (False Positives): Đây là những cảnh báo chỉ ra rằng một cái gì đó đúng với một gói tin nhưng thực chất nó là không đúng. Ví dụ: IDS tìm thấy một gói có mã độc nhưng nó thực sự là một mã không độc.

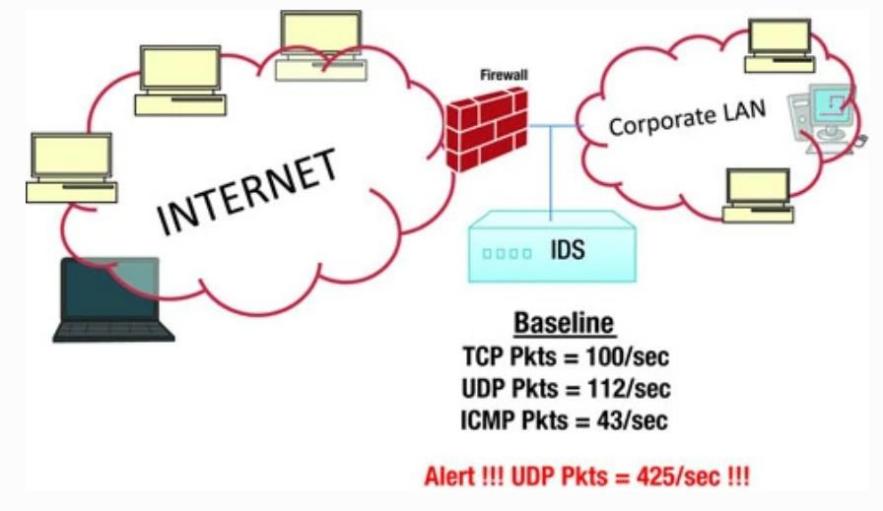
Âm tính giả (False Negatives): Đây là những cảnh báo rằng một cái gì đó là không đúng khi thực chất nó là đúng. Ví dụ: IDS thấy rằng một gói không có bất kỳ mã độc nào nhưng nó thực sự chứa một mã độc, như được tìm thấy thông qua điều tra.



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

### Phát hiện dựa trên dòng dữ liệu bất thường

Phát hiện dựa trên bất thường bảo vệ chống lại các mối đe dọa chưa biết. Nếu bất kỳ lưu lượng truy cập nào được tìm thấy là bất thường từ đường cơ sở, thì một cảnh báo được kích hoạt bởi các ID bị nghi ngờ là sự xâm nhập. IDP trước tiên tạo ra một cấu hình cơ sở thể hiện hành vi bình thường của lưu lượng truy cập. Hồ sơ cơ sở được tạo bằng cách cho phép hệ thống IDS tìm hiểu lưu lượng truy cập trong một khoảng thời gian để IDP có thể nghiên cứu hành vi di chuyển trong giờ cao điểm, giờ không cao điểm, giờ đêm, giờ đầu kinh doanh. Sau khi học, sự di chuyển của lưu lượng được thu thập trong khoảng thời gian được nghiên cứu thống kê và một hồ sơ cơ sở được tạo ra. Khi IDS được thay đổi từ chế độ học tập sang chế độ phát hiện/phòng ngừa, nó bắt đầu so sánh lưu lượng thông thường với cấu hình được tạo và nếu tìm thấy bất kỳ sự bất thường hoặc sai lệch nào so với hồ sơ cơ sở sự xâm nhập sẽ được ngăn chặn nếu nó được cấu hình cho chế độ phòng ngừa. Cấu hình tùy chỉnh cũng có thể được tạo cho hành vi lưu lượng truy cập cụ thể, chẳng hạn như số lượng e-mail được gửi bởi các nỗ lực truy cập của người dùng. Hình 1.2 dưới đây mô tả sự bất thường được phát hiện trên dòng dữ liệu.



**Hình 1.2: Phát hiện bất thường dựa trên dòng dữ liệu [1]**

Một số dấu hiệu về các hành vi đáng ngờ: quá nhiều phiên đăng nhập Telnet trong một ngày, lưu lượng truy cập HTTP trên một cổng không chuẩn, lưu lượng SNMP bị quá tải.

Để phát hiện xâm nhập hiệu quả, IDS phải có hồ sơ cơ sở mạnh mẽ bao gồm toàn bộ mạng của tổ chức và các phân khúc. Nó sẽ bao gồm hành vi di chuyển bình thường của tất cả các thành phần nhằm mục đích được bao phủ bởi hệ thống phát hiện và phòng ngừa xâm nhập. Hồ sơ cơ sở có thể thay đổi về độ phức tạp từ một nội dung đơn giản đến toàn diện, tùy thuộc vào các đặc điểm của mạng và các thành phần của nó. Ví dụ: một hồ sơ có thể bao gồm các dữ liệu sau:

* Một ứng dụng web đã đăng nhập từ xa bởi một bộ người dùng cụ thể
* Một ứng dụng có thiết kế mật khẩu được chấp nhận cụ thể
* Lưu lượng truy cập trong giờ cao điểm và giờ không cao theo quy định của tổ chức
* Mẫu kết nối từ mạng đối tác bên ngoài
* Kết nối từ một bộ thiết bị di động với máy chủ cơ sở dữ liệu

Thách thức của phương pháp phát hiện dựa trên sự bất thường là tạo ra một hồ sơ hiệu quả. Hồ sơ ban đầu, đôi khi được gọi là "hồ sơ đào tạo", được tạo bằng cách nghiên cứu mô hình lưu lượng trong một khoảng thời gian. Yếu tố thời gian có thể thay đổi từ tổ chức này sang tổ chức khác. Nó có thể là một vài giờ đến vài ngày. Khi cấu hình này được tạo, ID được đưa vào chế độ phát hiện và mỗi khi có một gói tin, một mẫu được khớp với cấu hình cơ sở. Đường cơ sở này có thể được thay đổi khi được yêu cầu dựa trên hành vi di chuyển. Nếu bất kỳ hoạt động độc hại nào đã tồn tại

ngay từ đầu, trong khi xây dựng hồ sơ cơ bản, hoạt động này cũng sẽ trở thành một phần của hồ sơ cơ sở và loại hoạt động độc hại như vậy sẽ không bị phát hiện. Do đó, phát hiện bất thường không nhất thiết phải phát hiện từng cuộc tấn công chưa biết. Giới hạn dựa trên hồ sơ cơ sở được tạo.

#### Các loại bất thường

Hệ thống phòng ngừa và phát hiện xâm nhập dựa trên dị thường (IDP) bảo vệ khỏi sự bất thường gây ra do vi phạm các giao thức và tải trọng ứng dụng. Nó cũng bảo vệ chống lại các cuộc tấn công từ chối dịch vụ và các cuộc tấn công tràn bộ đệm.

#### Giao thức bất thường

Giao thức bất thường đề cập đến sự bất thường trong định dạng giao thức và hành vi giao thức liên quan đến các tiêu chuẩn và thông số kỹ thuật của Internet. Có nhiều khía cạnh trong giao thức TCP và IP cần được theo dõi, ví dụ, các cờ khác nhau, SYN, ACK, FIN, sự kết hợp của chúng và các cờ dự trữ. Cách phân mảnh IP khi lắp ráp lại được thực hiện theo tiêu chuẩn. Nếu sự bất thường này không được phát hiện bởi ID, máy chủ cuối có thể không xử lý được các gói này và điều này có thể dẫn đến sự cố của hệ thống. Ở cấp độ ứng dụng, IDP phải có khả năng thực hiện phân tích giao thức sâu hơn để hiểu sự bất thường về giao thức ứng dụng. Nó cũng đòi hỏi sự hiểu biết sâu sắc về ngữ nghĩa ứng dụng để phát hiện sự bất thường của tải trọng ứng dụng. Một số ví dụ khác bao gồm: Phân đoạn TCP bất thường và các cờ TCP kết hợp, kiểm tra hoạt động bất thường, cờ phân mảnh IP không chính xác, các lệnh giao thức bất hợp pháp và việc sử dụng nó, chạy giao thức trên cổng không chuẩn, sử dụng sai các dịch vụ giao thức.

#### Phát hiện bất thường thống kê

Từ chối dịch vụ (DoS) và từ chối dịch vụ phân tán (DDoS) dẫn đến một loạt lưu lượng truy cập trên mạng diễn ra bất bình. Để khắc phục loại tấn công này, các cấu hình cơ bản được tạo ra trên lưu lượng di chuyển bình thường như được mô tả trước đó là dựa trên mô hình thống kê như Naïve Bayes để xác định các gói bất thường trên mạng. Trong khi học hành vi của lưu lượng mạng, chức năng của mô hình thống kê là tính toán điểm xác suất cho từng gói dữ liệu được coi là lưu lượng bình thường. Điểm số được tính toán dựa trên dữ liệu được lấy mẫu trong một khoảng thời gian và được lưu trữ trong một hồ sơ cơ sở. Một ngưỡng được đặt cho mỗi bộ giao thức và

người dùng. Khi IDS ở chế độ giám sát, dữ liệu được kiểm tra so với đường cơ sở và ngưỡng. Bất cứ khi nào một gói bất thường được phát hiện và điểm số trên ngưỡng, thì một cảnh báo được kích hoạt. Quá trình cảnh cáo sẽ chỉ báo khi dữ liệu được tìm thấy là bất thường trong một khoảng thời gian đủ, nếu không IDP sẽ chỉ đơn giản bỏ qua dấu vết. Ngưỡng có thể được đặt cho các cấu hình khác nhau, cho các giao thức khác nhau và cho người dùng khác nhau.

Khi IDP ở chế độ giám sát, nếu có bất cứ điều gì bất thường đối với đường cơ sở, hệ thống sẽ tạo ra một cảnh báo nhưng kết quả phân tích xác nhận rằng cảnh báo được tìm thấy là dương tính giả. Là một quản trị viên bảo mật, người ta mong đợi một loại hành vi tương tự xuất hiện mỗi ngày và để giảm thiểu chi tiêu. Lưu lượng vẫn được coi là bình thường và bất cứ điều gì vượt quá ngưỡng được thiết lập sẽ được xem là một sự xâm nhập. Ngưỡng cũng có thể được đặt cho một tập hợp người dùng hoặc tập hợp các giao thức.

Hồ sơ dựa trên các biện pháp thống kê có thể phát hiện một số dị thường DoS dựa trên các phân phối hoặc các vụ nổ ngắn hạn của lưu lượng đỉnh (tức là cao). Các cấu hình cơ sở liên tục được học trong khi hệ thống ở chế độ phát hiện và đường cơ sở được tạo lại để điều chỉnh mô hình lưu lượng thay đổi để tránh dương tính giả.

Bằng cách tạo các hồ sơ khác nhau, các cuộc tấn công DoS có thể được ngăn chặn. Ví dụ [1], đối với mỗi cuộc tấn công DoS, một hồ sơ có thể được tạo. Lỗ hổng hệ thống được nhắm mục tiêu bởi các phân đoạn TCP khi cờ SYN được bật sẽ tạo không gian cho cuộc tấn công DoS, khi đó những tấn công dạng này sẽ được gọi là cuộc tấn công SYN flood. Bất cứ khi nào có lưu lượng truy cập SYN flood trên mạng, các cảm biến IDS có thể phát hiện cuộc tấn công SYN flood bằng cách so sánh lưu lượng mạng với hồ sơ SYN flood do đó cảnh báo một cuộc tấn công SYN flood. Tương tự, hồ sơ UDP flood, hồ sơ phân đoạn dữ liệu TCP hoặc hồ sơ ICMP flood có thể được phát hiện và cảnh báo.

Mặc dù IDS dựa trên bất thường có lợi thế là phát hiện các cuộc tấn công chưa biết, việc xác định các quy tắc cho nó là lại trở nên khó khăn. Mỗi giao thức phải được phân tích, xử lý và so sánh với đường cơ sở. Bất kỳ giao thức tùy chỉnh nào đều làm cho nó trở nên không phù hợp.

Một cạm bẫy lớn khác của phát hiện bất thường là xác định lưu lượng bình thường trong khi tạo ra một đường cơ sở. Lưu lượng bình thường phải được sạch sẽ và không nên có bất kỳ hoạt động độc hại nào trong mạng. Trong trường hợp có bất kỳ hoạt động độc hại nào trong quá trình học tập, thì hồ sơ cơ sở tìm hiểu điều này và khiến việc phát hiện sự xâm nhập này khó khăn hơn hoặc thậm chí có thể không phát hiện sự xâm nhập của lưu lượng độc hại như vậy. Ví dụ, các cuộc tấn công trinh sát như dấu vân tay hoặc thư mục, tuân thủ giao thức mạng, dễ dàng không được chú ý vì nó tuân thủ các giới hạn giao thức và tải trọng.

#### Phát hiện phân tích trạng thái giao thức

Phương pháp này tương tự như phát hiện dựa trên bất thường, ngoại trừ các hồ sơ được tạo bởi các nhà cung cấp cung cấp thiết bị cảm biến (IDP). Các hồ sơ được xác định trước và được tạo thành từ hoạt động lưu lượng mạng an toàn được chấp nhận chung theo quy định của các tiêu chuẩn. "Trạng thái" có nghĩa là IDP có khả năng theo dõi trạng thái của giao thức cả trong lớp mạng và lớp ứng dụng. Ví dụ, trong trường hợp trạng thái thiết lập kết nối TCP, ID nên nhớ tất cả các trạng thái kết nối. Tương tự, trong trường hợp xác thực, phiên kết nối ban đầu ở trạng thái trái phép và IDS nên nhớ các trạng thái này. Sau khi trao đổi một số thông tin giữa hai bên máy khách và máy chủ, người dùng được xác thực và được cho phép truy cập vào mạng. Trong giai đoạn này, việc di chuyển là an toàn và IDP nên nhớ trạng thái hoặc nó sẽ dẫn đến dương tính thật.

Phương pháp phát hiện bất thường trong giao thức trạng thái sử dụng các cấu hình đã được tạo dựa trên các tiêu chuẩn và thông số kỹ thuật được chỉ định bởi nhà cung cấp thường tuân thủ hầu hết các giao thức từ các cơ quan tiêu chuẩn (Lực lượng đặc nhiệm kỹ thuật Internet). Nếu bất kỳ nhà cung cấp nào đã thực hiện các giao thức, với sự thay đổi của các tiêu chuẩn, nó sẽ gây khó khăn cho IDP trong việc phát hiện và phân tích các trạng thái. Trong các trường hợp như vậy, các mô hình giao thức IDPS cũng cần được cập nhật cho các thay đổi giao thức tùy chỉnh.

Hạn chế chính của phương pháp này là chúng rất tốn kém về quá trình và bộ nhớ như nhiều giao thức và IDP phải theo dõi đồng thời trạng thái của chúng. Một vấn đề khác là nếu một cuộc tấn công nằm trong hành vi giao thức thường được chấp

nhận, thì nó có thể đi qua. Nếu việc triển khai giao thức thay đổi từ hệ điều hành thì IDP có thể không hoạt động tốt trong việc phát hiện các xâm nhập.

## Giải pháp phát hiện và phòng chống xâm nhập

Ngoài việc nắm rõ các dấu hiệu của cuộc tấn công, ta cũng cần phải trang bị những giải pháp phát hiện và phòng chống chống xâm nhập mạng [3] như:

### Phân chia mạng

Một phần cơ bản của việc tránh các mối đe dọa bảo mật mạng là phân chia mạng thành các khu vực dựa trên các yêu cầu bảo mật. Điều này có thể được thực hiện bằng cách sử dụng các mạng con trong cùng một mạng hoặc bằng cách tạo các mạng cục bộ ảo (VLANS), mỗi loại hoạt động giống như một mạng riêng biệt. Phân khúc giới hạn tác động tiềm ẩn của một cuộc tấn công vào một khu vực và yêu cầu kẻ tấn công thực hiện các biện pháp đặc biệt để thâm nhập và có quyền truy cập vào các khu vực mạng khác.

### Điều chỉnh quyền truy cập Internet qua máy chủ proxy

Không cho phép người dùng mạng truy cập Internet không được kiểm tra. Vượt qua tất cả các yêu cầu thông qua một proxy trong suốt và sử dụng nó để kiểm soát và giám sát hành vi của người dùng. Đảm bảo rằng các kết nối bên ngoài thực sự được thực hiện bởi một con người và không phải là một bot hoặc cơ chế tự động khác. Các tên miền danh sách trắng để đảm bảo người dùng công ty chỉ có thể truy cập các trang web đã được phê duyệt một cách rõ ràng.

### Đặt thiết bị bảo mật chính xác

Tường lửa thường đặt tại vùng biên giới của hệ thống máy tính hay hệ thống mạng. Hoặc tùy vào loại tường lửa như tường lửa cá nhân (Internal Firewall) hay tường lửa hệ thống (System Firewall) mà vị trí đặt cũng như chức năng sẽ có sự khác nhau. Với Internal Firewall thì sau khi được cài đặt thì nó sẽ chiếm giữ việc quản lý tất cả các thông tin đi ra hay đi vào máy tính cá nhân của người dùng. System Firewall sẽ được lắp đặt ngay sau thiết bị kết nối WAN, như Router sử dụng các kênh thuê riêng (leased-line), hay Rounter ADSL … Nếu không thể triển khai tường lửa chính thức ở khắp mọi nơi, hãy sử dụng chức năng tường lửa tích hợp của các thiết bị chuyển mạch và bộ định tuyến, sử dụng hardware firewall hoặc thiết bị chống DDoS

để chống tấn công DDoS… Xem xét cẩn thận nơi đặt các thiết bị chiến lược như bộ cân bằng tải - nếu chúng nằm ngoài vùng Demilitarized Zone (DMZ) – vùng nằm giữa LAN và internet (chứa các server và cung cấp các dịch vụ cho các host trong LAN cũng như các host từ các LAN ở bên ngoài), chúng sẽ không được bảo vệ bởi bộ máy bảo mật mạng.

### Sử dụng NAT (Network Address Translation)

NAT cho phép dịch địa chỉ IP nội bộ thành địa chỉ có thể truy cập trên các mạng công cộng. Có thể sử dụng nó để kết nối nhiều máy tính với Internet bằng một địa chỉ IP duy nhất. Điều này cung cấp thêm một lớp bảo mật, bởi vì bất kỳ lưu lượng truy cập nào cũng phải đi qua thiết bị NAT.

### Giám sát lưu lượng mạng

Đảm bảo có khả năng hiển thị hoàn toàn lưu lượng truy cập đến, đi và mạng nội bộ, với khả năng tự động phát hiện các mối đe dọa và hiểu bối cảnh và tác động của họ. Kết hợp dữ liệu từ các công cụ bảo mật khác nhau để có được một hình ảnh rõ ràng về những gì đang xảy ra trên mạng, nhận ra rằng nhiều cuộc tấn công trải rộng trên nhiều hệ thống CNTT, tài khoản người dùng.

Đạt được mức độ hiển thị này có thể khó khăn với các công cụ bảo mật truyền thống. CYNET 360 là một giải pháp bảo mật tích hợp cung cấp các phân tích mạng tiên tiến, liên tục theo dõi lưu lượng mạng, tự động phát hiện hoạt động độc hại và phản ứng với nó tự động hoặc vượt qua thông tin về ngữ cảnh cho nhân viên bảo mật.

### Sử dụng công nghệ “đánh lừa”

Không có biện pháp bảo vệ mạng nào thành công 100% và những kẻ tấn công cuối cùng sẽ thành công trong việc thâm nhập mạng. Nhận ra điều này và đặt công nghệ đánh lừa để tạo ra những mồi nhử trên mạng, những kẻ tấn công bị cám dỗ sẽ "tấn công" và nó cho phép chuyên viên quản trị quan sát kế hoạch và kỹ thuật của kẻ tấn công. Có thể sử dụng mồi nhử để phát hiện các mối đe dọa trong tất cả các giai đoạn của vòng đời tấn công: Tệp dữ liệu, thông tin xác thực và kết nối mạng.

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

Học máy (Machine Learning - ML) là một tập hợp con của AI bao gồm tất cả các phương thức và thuật toán cho phép các máy để tìm hiểu tự động bằng các mô

hình toán học để trích xuất thông tin hữu ích từ các bộ dữ liệu lớn. ML phổ biến nhất (còn gọi là học tập nông) được sử dụng cho IDS là [4] cây quyết định, K-nearest neighbors (KNN), mạng thần kinh nhân tạo (ANN), máy vectơ hỗ trợ (SVM), cụm K-Mean, và mạng học tập nhanh.

### Decision Tree (DT)

DT là một trong những thuật toán ML có giám sát cơ bản được sử dụng để phân loại và hồi quy bộ dữ liệu đã cho bằng cách áp dụng hàng loạt các quyết định (quy tắc). Mô hình có cấu trúc cây thông thường với các nút, cành và lá. Mỗi nút biểu thị một thuộc tính hoặc một tính năng. Một nhánh đại diện cho một quyết định hoặc quy tắc trong khi mỗi lá đại diện cho một kết quả hoặc nhãn lớp. Thuật toán DT sẽ tự động chọn các tính năng tốt nhất để xây dựng nên cây quyết định và sau đó thực hiện thao tác cắt tỉa để loại bỏ các nhánh không liên quan từ cây để tránh sự phù hợp. Các mô hình DT phổ biến nhất là Cart, C4.5 và ID3. Nhiều thuật toán học tập nâng cao như rừng ngẫu nhiên (Random Forest) và XGBoost (eXtreme Gradient Boosting) được xây dựng từ nhiều cây quyết định.

### K-Nearest Neighbor (KNN)

KNN là một trong những thuật toán ML có giám sát đơn giản nhất sử dụng ý tưởng về "điểm dữ liệu tương tự tồn tại gần nhau trong một không gian" để dự đoán lớp của một mẫu dữ liệu nhất định. Nó xác định đầu ra một mẫu dựa trên thông tin của K điểm dữ liệu trong tập huấn luyện gần nó nhất (hay còn gọi là K-lân cận). Trong thuật toán KNN, tham số K ảnh hưởng đến hiệu suất của mô hình. Nếu giá trị của K nhỏ, mô hình có thể dễ bị overfitting khi đó mô hình không học được gì từ dữ liệu. Trong khi, khi lựa chọn giá trị K lớn có thể dẫn đến việc phân loại sai đối tượng mẫu. Karatas đã so sánh hiệu suất của các thuật toán ML khác nhau bằng cách sử dụng tập dữ liệu CSE-CIC-IDS2018. Họ giải quyết vấn đề về sự mất cân bằng dữ liệu bằng bằng cách sử dụng kỹ thuật tăng kích thước mẫu (Smote), dẫn đến việc cải thiện tỷ lệ phát hiện cho các mẫu thuộc lớp thiểu số.

### Support vector machine

SVM là một thuật toán ML có giám sát dựa trên ý tưởng về một siêu mặt phẳng (hay còn gọi là hyper lane) để phân tách các điểm của dữ liệu. Siêu mặt phẳng này sẽ

chịu trách nhiệm chia không gian thành các miền khác nhau và mỗi miền này sẽ chứa một loại dữ liệu. SVM được sử dụng làm giải pháp của cả các vấn đề tuyến tính và phi tuyến tính. Đối với các vấn đề phi tuyến tính, các hàm kernel được sử dụng. Ý tưởng là đầu tiên ánh xạ một vectơ đầu vào có kích thước thấp thành không gian tính năng có kích thước cao bằng hàm kernel. Tiếp theo, một mặt phẳng cận biên tối đa và tối ưu thu được sẽ hoạt động như một ranh giới quyết định sử dụng các vectơ hỗ trợ.

### K-mean clustering

Việc phân cụm là một ý tưởng phân chia dữ liệu thành các cụm có ý nghĩa (hoặc nhóm), bằng cách đặt dữ liệu có tính chất giống nhau vào cùng một cụm. K- Mean Clustering là một trong những thuật toán ML không giám sát phổ biến dựa trên centroid (tâm của cụm dữ liệu). K đại diện cho số lượng centroid trong một tập dữ liệu. Khoảng cách từ mỗi điểm dữ liệu tới tâm được tính toán để chỉ định một số điểm dữ liệu nhất định sẽ vào cùng một cụm nào. Ý tưởng chính của thuật toán là làm giảm tổng khoảng cách giữa các điểm dữ liệu và centroid tương ứng của chúng trong một cụm.

Yao [5] đã đề xuất một khung mô hình phát hiện xâm nhập nhiều cấp độ có tên multilevel semi-supervised ML (MSML) để vấn đề về sự mất cân bằng nghiêm trọng của lưu lượng mạng trong các danh mục khác nhau và sự phân bổ không đồng nhất giữa tập huấn luyện và tập kiểm tra trong không gian đặc trưng. Giải pháp được đề xuất bao gồm bốn mô-đun như trích xuất cụm thuần túy, phát hiện mẫu, phân loại fine-grained và cập nhật mô hình. Ý tưởng là nếu một cuộc tấn công không được dán nhãn trong một mô-đun thì nó được chuyển tiếp đến phần tiếp theo để phát hiện. Phương pháp đề xuất đã được thử nghiệm bằng cách sử dụng bộ dữ liệu KDD Cup'99. Kết quả thử nghiệm cho thấy sự vượt trội của mô hình khi dùng để phát hiện các cuộc tấn công ngay cả với các trường hợp nhãn có tỉ lệ thấp trong bộ dữ liệu.

### Artificial neural network

ANN cũng là một thuật toán ML có giám sát và được truyền cảm hứng từ hệ thống thần kinh của bộ não con người. ANN được tạo thành từ các yếu tố xử lý gọi là tế bào thần kinh (nút) và các kết nối giữa chúng (cạnh). Các nút này được sắp xếp trong một lớp đầu vào, nhiều lớp ẩn và một lớp đầu ra. Thuật toán lan truyền ngược

hay còn gọi là Backpropagation được sử dụng như một kỹ thuật học tập cho ANN. Ưu điểm chính của việc sử dụng kỹ thuật ANN là khả năng thực hiện mô hình hóa phi tuyến tính bằng cách học tập từ các bộ dữ liệu lớn. Tuy nhiên, vấn đề chính khi đào tạo mô hình với ANN là mức tiêu thụ thời gian cao do tính chất phức tạp của nó, làm chậm quá trình học tập và để đạt được một giải pháp dưới mức tối ưu.

Để khắc phục những hạn chế của ANN, Huang đã đề xuất một ANN mới gọi là một máy học tập cực đoan (Extreme learning machine - ELM). ELM là một mạng thần kinh truyền thẳng với một lớp ẩn duy nhất, sử dụng ngẫu nhiên các trọng lượng (weight) đầu vào và độ lệch (bias) lớp ẩn mà không điều chỉnh và xác định trọng lượng đầu ra theo cách phân tích. Dựa trên ý tưởng của ELM, Li đã đề xuất một mạng học nhanh (Fast learning network - FLN). FLN dựa trên việc kết nối mạng thần kinh chuyển tiếp nhiều lớp và một mạng thần kinh chuyển tiếp dữ liệu một lớp song song. FLN cho thấy hiệu suất và sự ổn định khá hợp lý bằng cách sử dụng một số ít các nút ẩn và sử dụng ít thời gian hơn.

Ali [6] đã đề cập đến vấn đề IDS bằng cách đề xuất một mô hình dựa trên mạng học nhanh và tối ưu hóa bầy hạt được đặt tên là PSO-FLN, mô hình đã thử nghiệm bằng cách sử dụng bộ dữ liệu KDD Cup'99. Thử nghiệm được diễn ra bằng cách so sánh FLN với các thuật toán tối ưu hóa khác nhau. Kết quả cho thấy PSO- FLN vượt trội hơn các mô hình FLN khác với các thuật toán tối ưu hóa khác nhau làm thuật toán di truyền, tối ưu hóa tìm kiếm Harmoney và tối ưu hóa dựa trên việc học. Họ cũng chứng minh rằng việc tăng số lượng tế bào thần kinh trong lớp ẩn làm tăng độ chính xác.

### Ensemble methods

Ý tưởng chính đằng sau các phương thức Ensemble là tổng hợp các kết quả dự đoán của nhiều model thành một model cuối cùng. Vì mỗi bộ phân loại có một số điểm mạnh và điểm yếu. Một số người có thể thực hiện tốt để phát hiện một loại tấn công cụ thể và cho thấy hiệu suất kém về các loại tấn công khác. Cách tiếp cận Ensemble là kết hợp các phân loại yếu bằng cách đào tạo nhiều phân loại và sau đó tạo thành một bộ phân loại mạnh hơn bằng cách chọn bằng thuật toán đã được bỏ phiếu.

Shen đã đề xuất nghiên cứu sử dụng phương pháp hòa đồng bằng cách chọn Elm làm bộ phân loại cơ sở. Để tối ưu hóa phương pháp đề xuất, thuật toán tối ưu hóa BAT được sử dụng trong giai đoạn cắt tỉa. Mô hình đã được thử nghiệm bằng cách sử dụng bộ dữ liệu KDD Cup'99, NSL-KDD và Kyoto. Kết quả thử nghiệm cho thấy nhiều Elms kết hợp vượt trội hơn về mặt hiệu suất so với các phương pháp khác.

Gao đã đề xuất một mô hình nghiên cứu của mình bằng cách sử dụng một số phân loại cơ sở dưới dạng DT, RF, KNN, mạng thần kinh sâu (DNN) và chọn ra bộ phân loại tốt nhất bằng thuật toán bỏ phiếu. Phương pháp được đề xuất đã được xác minh bằng cách thực hiện các thí nghiệm trên bộ dữ liệu NSL-KDD. Kết quả thí nghiệm đã chứng minh hiệu suất cao của mô hình đề xuất bằng cách so sánh với các mô hình khác

## Các mô hình sử dụng cho hệ thống IDS

### Mô hình Decision Tree

Cây quyết định [22] là một phương pháp học có giám sát được sử dụng để phân loại và hồi quy. Mục tiêu là tạo một mô hình dự đoán giá trị của biến mục tiêu bằng cách học các quy tắc quyết định đơn giản được suy ra từ các tính năng dữ liệu. Một cây có thể được xem là một xấp xỉ không đổi.

#### Một số lợi thế của cây quyết định là:

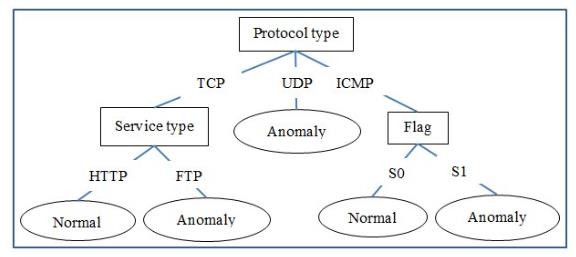
* Đơn giản để hiểu và dễ dàng thực hiện.
* Yêu cầu chuẩn bị ít dữ liệu. Các kỹ thuật khác thường yêu cầu chuẩn hóa dữ liệu, các biến giả cần được tạo và loại bỏ các giá trị trống.
* Chi phí sử dụng cây (tức là dữ liệu dự đoán) là logarit trong số điểm dữ liệu được sử dụng để đào tạo cây.
* Có thể xử lý cả dữ liệu số và dữ liệu phân loại. Tuy nhiên, hiện tại việc triển khai scikit-learning không hỗ trợ các biến phân loại. Các kỹ thuật khác thường chuyên về phân tích tập dữ liệu chỉ có một loại biến.
* Có khả năng xử lý các vấn đề đa đầu ra.
* Sử dụng mô hình hộp trắng. Nếu một tình huống nhất định có thể quan sát được trong một mô hình, thì lời giải thích cho điều kiện đó dễ dàng được giải thích bằng logic boolean. Ngược lại, trong mô hình hộp đen (ví dụ: trong mạng nơ-ron nhân tạo), kết quả có thể khó giải thích hơn.
* Có thể xác nhận một mô hình bằng cách sử dụng các bài kiểm tra thống kê. Điều đó làm cho nó có thể tính đến độ tin cậy của mô hình.
* Hoạt động tốt ngay cả khi các giả định của nó phần nào bị vi phạm bởi mô hình thực mà từ đó dữ liệu được tạo ra.

#### Những điểm yếu của cây quyết định bao gồm:

* Cây quyết định có thể tạo ra cây quá phức tạp, không tổng quát hóa dữ liệu tốt. Các cơ chế như thiết lập số lượng mẫu tối thiểu cần thiết tại một nút lá hoặc thiết lập độ sâu tối đa của cây là cần thiết để tránh vấn đề này.
* Cây quyết định có thể không ổn định vì các biến thể nhỏ trong dữ liệu có thể dẫn đến việc tạo ra một cây hoàn toàn khác. Vấn đề này được giảm thiểu bằng cách sử dụng cây quyết định trong một tập hợp.
* Các dự đoán của cây quyết định không trơn tru cũng không liên tục, mà là các phép gần đúng không đổi.
* Một cây quyết định tối ưu khi hoàn chỉnh NP dưới một số khía cạnh của tính tối ưu và ngay cả đối với các khái niệm đơn giản. Do đó, các thuật toán học cây quyết định thực tế dựa trên các thuật toán heuristic như thuật toán greedy, trong đó các quyết định tối ưu cục bộ được thực hiện tại mỗi nút với mục tiêu tìm được sự tối ưu trên toàn cục. Điều này có thể được giảm thiểu bằng cách

huấn luyện nhiều cây theo nhóm, trong đó các tính năng và mẫu được lấy ngẫu nhiên để thay thế.

* Có những khái niệm phức tạp, vì cây quyết định không thể hiện chúng một cách dễ dàng, chẳng hạn như vấn đề XOR, chẵn lẻ hoặc bộ ghép kênh.



**Hình 3.7: Sơ đồ cây quyết định [23]**

### Mô hình KNN

K-Nearest Neighbor [24] là một thuật toán khai phá dữ liệu và đồng thời là thuật toán học có giám sát. Đầu ra của biến mục tiêu được dự đoán bằng cách tìm k lân cận gần nhất, bằng cách tính khoảng cách Euclide [24] [25]. Đây là một kỹ thuật phân loại phi tham số không đưa ra bất kỳ giả định nào về dữ liệu cơ bản. Một số ưu điểm của KNN là:

* Dễ thực hiện và dễ hiểu.
* Sẽ rất hiệu quả và hiệu quả nếu dữ liệu huấn luyện rất lớn.
* Nó mạnh mẽ đối với dữ liệu nhiễu.
* Nó liên tục phát triển và dễ dàng thích nghi với môi trường mới.
* Dễ thực hiện cho bài toán nhiều lớp

Thuật toán KNN có công thức tổng quát như sau:

𝑛

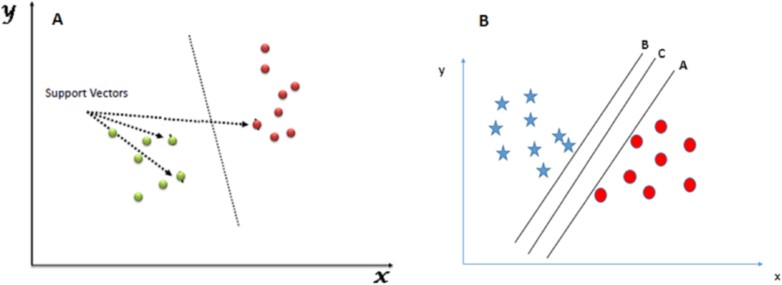
𝑑(𝑥, 𝑦) = √∑(𝑥𝑖 − 𝑦𝑖)2

𝑖=1

Trong đó, Xi là phần tử đặc trưng thứ i của đối tượng x, Yi là phần tử đặc trưng thứ I của đối tượng y và n là tổng số đặc trưng của bộ dữ liệu.

### Mô hình máy Vector hỗ trợ (SVM)

Máy hỗ trợ vectơ (SVM) [26] là một tập hợp các phương pháp học được giám sát được sử dụng để phân loại, hồi quy và phát hiện ngoại lệ. Ngoài ra, SVM cũng là một trong những thuật toán phân loại đáng tin cậy nhất trong máy học tập vì nó hỗ trợ quá trình dự đoán nhanh chóng và đơn giản so với các thuật toán khác. SVM phân loại các điểm dữ liệu dựa trên các vectơ hỗ trợ trong kho dữ liệu để tạo một siêu phẳng phân chia các nhãn lớp thành các lớp liên quan của chúng.

SVM phân loại các vectơ hỗ trợ dựa trên giá trị "gamma" được cung cấp cho thuật toán. Ví dụ: nếu gamma = 0, SVM dự đoán siêu phẳng là một đường cong. Nếu gamma = auto, SVM chỉ đơn giản là dự đoán siêu phẳng theo đầu vào dữ liệu đã cho.

**Hình 3.8: Phân lớp với SVM. (A) Kỹ thuật phân lớp SVM. (B) Kỹ thuật lựa chọn siêu phẳng SVM**

Hình 3.3 thể hiện tập hợp dữ liệu gồm các ngôi sao xanh và vòng tròn đỏ, nơi cần tìm ra siêu mặt phẳng phù hợp để phân loại chúng một cách chính xác. Ở đây, có ba siêu phẳng A, B và C. Nhưng siêu phẳng thực tế cần xem xét là siêu phẳng C vì nó có cùng khoảng cách từ cả hai điểm dữ liệu.

#### Một số ưu điểm của thuật toán SVM là:

* Hiệu quả trong không gian chiều cao.
* Vẫn có hiệu quả trong trường hợp số thứ nguyên lớn hơn số lượng mẫu.
* Sử dụng một tập hợp con các điểm huấn luyện trong hàm quyết định (được gọi là vectơ hỗ trợ), vì vậy nó cũng hiệu quả về bộ nhớ.
* Đa năng: các chức năng Kernel khác nhau có thể được chỉ định cho chức năng quyết định. Các nhân chung được cung cấp, nhưng cũng có thể chỉ định các nhân tùy chỉnh.
* Những nhược điểm của máy vector hỗ trợ bao gồm:
* Nếu số lượng tính năng lớn hơn nhiều so với số lượng mẫu, hãy tránh việc lựa chọn các chức năng của Kernel và thuật ngữ chính quy hóa là rất quan trọng.

SVM không trực tiếp cung cấp các ước tính xác suất, những ước tính này được tính toán bằng cách sử dụng xác thực chéo năm lần đắt tiền.

### Phân lớp và dự đoán

Mô-đun này sẽ sử dụng các mô hình tương ứng bao gồm KNN, SVM và cây quyết định, cùng với đặc trưng đã được chọn từ mô-đun 2 để để tiến hành phân lớp. Kết quả dự đoán sau khi áp dụng các mô hình sẽ bao gồm hai phân lớp: bị tấn công hoặc không bị tấn công. Cụ thể, sau khi quá trình xử lý dữ liệu kết thúc, từng mô hình KNN, Decision Tree, và SVM sẽ được áp dụng vào dữ liệu, kết quả thu được ở mỗi mô hình sẽ được phân tích dựa trên các độ đo hiệu quả mô hình, cụ thể là: Accuracy, Precision, F - measure, Recall.

Độ chính xác (Accuracy): Là tỷ lệ phần trăm của các phiên bản được phân loại chính xác trên tổng số phiên bản.

𝐴𝐶 =

𝑇𝑃 + 𝑇𝑁

𝑇𝑃 + 𝑇𝑁 + 𝐹𝑃 + 𝐹𝑁

Độ chính xác (Precision): Là tỷ lệ giữa các cá thể có liên quan theo các phiên bản được truy xuất.

𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 =

𝑇𝑃

𝑇𝑃 + 𝐹𝑃

Recall là tỷ lệ của các phiên bản có liên quan đã được truy xuất trên tổng số lượng các phiên bản có liên quan.

𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 =

𝑇𝑃

𝑇𝑃 + 𝐹𝑃

Độ đo F1: Là giá trị trung bình có trọng số của độ chính xác và độ thu hồi.

2 ∗ (𝑅𝑒𝑐𝑎𝑙𝑙 ∗ 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛)

𝐹1 𝑚𝑒𝑎𝑠𝑢𝑟𝑒 =

𝑅𝑒𝑐𝑎𝑙𝑙 + 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛

* TP (True positive): là số lượng các trường hợp tích cực được phân loại sửa chữa.
* TN (True negative): là số lượng các trường hợp âm được phân loại chính xác.
* FP (False positive): là số thực thể dương bị phân loại sai.
* FN (False negative): là số lượng các trường hợp phủ định bị phân loại sai.

# DANH MỤC TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | Umesh Hodeghatta Rao & Umesha Nayak, Intrusion Detection and Prevention  Systems, 2014. |
| [2] | Vnetwork, 3 November 2020. [Online]. Available: https://vnetwork.vn/news/4-  loai-tan-cong-mang-nguy-hiem-nhat-hien-nay. |
| [3] | R. Daş, A. Karabade and G. Tuna, "Common network attack types and defense mechanisms,"," *23nd Signal Processing and Communications Applications*  *Conference (SIU),* 2015. |
| [4] | Aleksey A. Titorenko, Alexey A. Frolov, “Analysis of Modern Intrusion  Detection System,” 2018. |
| [5] | H. Yao, D. Fu, P. Zhang, M. Li and Y. Liu, "MSML: A Novel Multilevel Semi- Supervised Machine Learning Framework for Intrusion Detection System,"  *IEEE Internet of Things Journal,* vol. 6, pp. 1949-1959, 2019. |
| [6] | M. H. Ali, B. A. D. Al Mohammed, A. Ismail and M. F. Zolkipli, "A New Intrusion Detection System Based on Fast Learning Network and Particle  Swarm Optimization," *IEEE Access,* vol. 6, pp. 20255-20261, 2018. |
| [7] | Kim-Hung Le, Minh-Huy Nguyen, Trong-Dat Tran, Ngoc-Duan Tran, “IMIDS:  An Intelligent Intrusion Detection System against Cyber Threats in IoT,” 10 February 2022. |
| [8] | Xuan-Ha Nguyen, Xuan-Duong Nguyen, Hoang-Hai Huynh, Kim-Hung Le, “Realguard: A Lightweight Network Intrusion Detection System for IoT  Gateways,” 7 January 2022. |
| [9] | Tran Ngoc Thinh, Tran Hoang Quoc Bao, Duc-Minh Ngo, Cuong Pham-Quoc,  “High-performance anomaly intrusion detection system with ensemble neural networks on reconfigurable hardware,” 26 April 2021. |
| [10] | Navaporn Chockwanich, Vasaka Visoottiviseth, “Intrusion Detection by Deep  Learning with TensorFlow,” trong *International Conference on Advanced Communications Technology(ICACT)*, 2019. |

|  |  |
| --- | --- |
| [11] | Jiadong Ren, Jiawei Guo, Wang Qian, Huang Yuan, Xiaobing Hao , and Hu  Jingjing, “Building an Effective Intrusion Detection System by Using Hybrid Data Optimization Based on Machine Learning Algorithms,” 16 June 2019. |
| [12] | Kinan Ghanem, Francisco J. Aparicio-Navarro, Konstantinos G. Kyriakopoulos, Sangarapillai Lambotharan, Jonathon A. Chambers, “Support Vector Machine for Network Intrusion and Cyber-Attack Detection,” trong  *Sensor Signal Processing for Defence (SSPD)*, London, 2017. |
| [13] | Anar Ahady, Ali Ghubaish, Tara Salman, Devrim Ünal, “Intrusion Detection  System for Healthcare Systems Using Medical and Network Data: A Comparison Study,” February 2017. |
| [14] | Isra Al-Turaiki, Najwa Altwaijry, “A Convolutional Neural Network for Improved Anomaly-Based Network Intrusion Detection,” *Big Data,* tập 9, p.  233–252, 2021. |
| [15] | Nuno Oliveira, Isabel Praça, Eva Maia, Orlando Sousa , “Intelligent Cyber Attack Detection and Classification for Network-Based Intrusion Detection  Systems,” 23 November 2021. |
| [16] | Oqbah Ghassan Abbas, Khaldoun Khorzom, Mohammed Assora, “Machine Learning based Intrusion Detection System for Software Defined Networks,” *International Journal of Engineering Research & Technology (IJERT),* tập 9,  số 09, 2020. |
| [17] | Mohannad Zead Khairallah, “Network Attacks Detection using Deep neural  network,” 2021. |
| [18] | Roza Dastres, Mohsen Soori, “A Review in Recent Development of Network Threats and Security Measures,” *International Journal of Computer and*  *Information Engineering,* tập 15, 2021. |
| [19] | E. Lundin, E. Jonsson, "Survey of research in the intrusion detection area,"  Department of Computer Engineering, Goteborg, 2002. |
| [20] | Ahmadian Ramaki, Ali & Ebrahimi Atani, Reza, "A survey of IT early warning  systems: architectures, challenges, and solutions," *Security and Communication Networks,* 2016. |

|  |  |
| --- | --- |
| [21] | Jack Koziol, Intrusion Detection with Snort, 2003. |
| [22] | Ratul Chowdhury, Pallabi Banerjee, Soumya Deep Dey, Banani Saha, Samir Kumar Bandyopadhyay, “A Decision Tree Based Intrusion Detection System  for Identification of Malicious Web Attacks,” 9 July 2020. |
| [23] | D. Singh, “A Collaborative IDS Framework for Cloud,” Research Gate, 2013. |
| [24] | M.Nikhitha, M.A.Jabbar, “K Nearest Neighbor Based Model for Intrusion Detection System,” *International Journal of Recent Technology and*  *Engineering (IJRTE),* tập 8, số 2, July 2019. |
| [25] | B. Basaveswara Rao, K.Swathi, “Fast kNN Classifiers for Network Intrusion  Detection System,” *Indian Journal of Science and Technology,* tập 10, 2017. |
| [26] | Arushi Agarwal, Purushottam Sharma, Mohammed Alshehri, Ahmed A. Mohamed, , Osama Alfarraj, “Classification model for accuracy and intrusion  detection using machine learning approach,” p. 20, 7 April 2021. |
| [27] | A. Chugh, “GreeksforGreeks,” 24 Sep 2021. [Trực tuyến]. Available:  [https://ww](http://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/)w.geeks[forgeeks.org/ml](http://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/)-[label-encoding-of-datasets-in-python/.](http://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/) |
| [28] | A. Fawcett, “Educative,” 11 Feb 2021. [Trực tuyến]. Available:  [https://ww](http://www.educative.io/blog/one-hot-encoding)w.educa[tive.io/blog/on](http://www.educative.io/blog/one-hot-encoding)e[-hot-encoding.](http://www.educative.io/blog/one-hot-encoding) |
| [29] | NM Shanono, Zulkiflee M, NA Abu, W. Yassin, MA Faizal, “Intrusion  Detection System Architecture: Issues and Challenges,” tập 62, August 2020. |