1. **TỔNG QUAN VỀ TÌNH HÌNH AN NINH MẠNG HIỆN NAY:**

Trong bối cảnh cuộc Cách mạng công nghiệp lần thứ tư và tiến trình chuyển đổi số quốc gia đang diễn ra mạnh mẽ, việc chủ động tham gia, ứng dụng công nghệ số đã trở thành yêu cầu tất yếu nhằm thúc đẩy phát triển kinh tế - xã hội và nâng cao năng lực cạnh tranh của đất nước. Tuy nhiên, song hành với quá trình số hóa là những thách thức ngày càng lớn về an ninh mạng, khi không gian mạng trở thành hạ tầng quan trọng gắn liền với mọi hoạt động của Chính phủ, doanh nghiệp và người dân. Sự phụ thuộc sâu vào công nghệ số đồng nghĩa với việc nguy cơ bị tấn công, xâm nhập và đánh cắp dữ liệu ngày càng gia tăng, đòi hỏi các biện pháp bảo vệ chủ động, linh hoạt và hiệu quả hơn.

Trong năm 2024, theo thống kê của Bộ Thông tin và Truyền thông, Việt Nam đã ghi nhận hơn 659.000 sự cố tấn công mạng, tăng gần 12% so với năm 2023. Các sự cố này ảnh hưởng đến hàng chục nghìn cơ quan, tổ chức và doanh nghiệp, đặc biệt trong các lĩnh vực tài chính, viễn thông và dịch vụ công trực tuyến. Các hình thức tấn công phổ biến gồm mã độc tống tiền (ransomware), tấn công từ chối dịch vụ (DDoS), xâm nhập chiếm quyền điều khiển hệ thống, đánh cắp dữ liệu và giả mạo danh tính người dùng. Sang quý I năm 2025, xu hướng tấn công có chủ đích (APT) vẫn tiếp tục gia tăng, nhiều chiến dịch được ghi nhận nhắm vào hạ tầng công nghệ thông tin trọng yếu, hệ thống quản lý dữ liệu công dân và các dịch vụ tài chính – ngân hàng.

Bên cạnh các mối đe dọa kỹ thuật, tình trạng phát tán thông tin sai lệch và tin giả trên không gian mạng trong năm 2024–2025 có xu hướng bùng phát mạnh hơn, đặc biệt trong các vấn đề kinh tế, chính trị và xã hội. Các nội dung xấu độc lan truyền qua mạng xã hội không chỉ gây nhiễu loạn thông tin mà còn tiềm ẩn nguy cơ tác động đến ổn định xã hội và lòng tin của người dân.

Sự phát triển nhanh của trí tuệ nhân tạo (AI), điện toán đám mây (Cloud Computing) và Internet vạn vật (IoT) trong giai đoạn 2024–2025 mang lại nhiều lợi ích nhưng đồng thời cũng mở rộng đáng kể bề mặt tấn công mạng. Việc người dùng cấp quyền truy cập rộng rãi cho các ứng dụng di động, mạng xã hội và nền tảng thương mại điện tử dẫn đến nguy cơ rò rỉ dữ liệu cá nhân, chiếm đoạt tài khoản và lạm dụng thông tin riêng tư. Các hình thức lừa đảo trực tuyến, cho vay “tín dụng đen”, đầu tư tài chính ảo hoặc đánh bạc qua mạng tiếp tục gia tăng với thủ đoạn ngày càng tinh vi, gây thiệt hại nghiêm trọng về tài sản và uy tín cho người dân, doanh nghiệp.

Để đối phó với tình hình này, trong năm 2025, Nhà nước đang đẩy mạnh việc xây dựng và hoàn thiện Luật An ninh mạng (sửa đổi), tập trung vào quản lý hệ thống thông tin trọng yếu, định danh điện tử và bảo vệ dữ liệu cá nhân. Cùng với đó, các cơ quan chức năng tăng cường hoạt động giám sát, cảnh báo sớm và ứng cứu sự cố, đồng thời khuyến khích doanh nghiệp đầu tư vào hệ thống phát hiện và ngăn chặn xâm nhập (IDS/IPS), phân tích hành vi người dùng, và các mô hình phát hiện bất thường dựa trên trí tuệ nhân tạo.

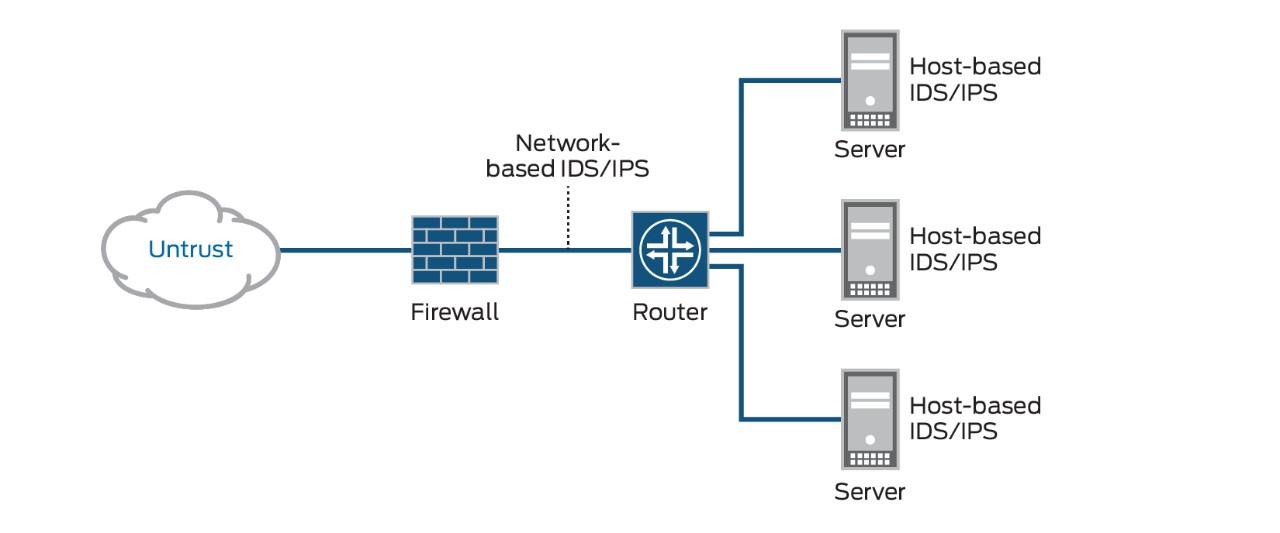
Cuộc cách mạng công nghiệp lần thứ tư đang mở ra nhiều cơ hội phát triển mới, song cũng đặt ra những yêu cầu cấp thiết trong việc bảo đảm an toàn, an ninh mạng cho toàn xã hội. Việc chủ động phòng ngừa, phát hiện và ứng phó kịp thời với các mối đe dọa trên không gian mạng chính là yếu tố then chốt để bảo vệ chủ quyền số quốc gia, duy trì ổn định an ninh, và bảo đảm sự phát triển bền vững trong thời kỳ chuyển đổi số hiện nay.

1. **TỔNG QUAN VỀ HỆ THỐNG PHÁT HIỆN VÀ PHÒNG CHỐNG XÂM NHẬP (IDS/IPS):**

***A. Giới thiệu về hệ thống phát hiện xâm nhập (Intrusion Detection System - IDS):***

* + 1. ***Khái niệm :***

Hệ thống phát hiện xâm nhập (Intrusion Detection System – IDS) là một phần mềm ứng dụng hoặc thiết bị được xây dựng để giám sát lưu lượng mạng và phát cảnh báo mỗi khi có các hành vi bất thường xâm nhập vào hệ thống.



*Hình 1. Mô hình phận loại của IDS bao gồm : NIDS (Network Intrusion Detection System) và HIDS (Host-based Intrusion Detection System).*

* + 1. ***Phân loại hệ thống phát hiện xâm nhập:***

Về mặt phân loại thì có 2 dạng chính :

* + - * **NIDS** (Network Intrusion Detection System)– Hệ thống phát hiện xâm nhập mạng, hệ thống sẽ tập hợp các gói tin để phân tích sâu bên trong nhằm xác định các mối đe dọa tiềm tàng mà không làm thay đổi cấu trúc của gói tin.
      * **HIDS** (Host-based Intrusion Detection System) – Hệ thống phát hiện xâm nhập dựa trên máy chủ, được cài đặt trực tiếp trên các máy tính cần theo dõi. HIDS giám sát lưu lượng đến và đi từ thiết bị để cảnh báo người dùng về những xâm nhập trái phép.

Các hệ thống IDS hiện đại được xây dựng để thu thập lưu lượng mạng từ mọi thiết bị thông qua cả NIDS và HIDS. Vì vậy có thể cải thiện đáng kể khả năng phát hiện xâm nhập trên hệ thống.

* + 1. ***Cách thức hoạt động của hệ thống phát hiện xâm nhập:***

Sau khi thu thập dữ liệu mạng, hệ thống IDS tiến hành phân tích và so sánh lưu lượng này với các mẫu hành vi hoặc dấu hiệu tấn công đã được lưu trong cơ sở dữ liệu trước đó. Quá trình này được gọi là tương quan mẫu (pattern correlation), giúp nhận diện các dấu hiệu bất thường trong lưu lượng mạng. Khi phát hiện hành vi khả nghi, IDS sẽ tự động gửi cảnh báo đến người quản trị hệ thống để họ kịp thời xử lý, cô lập mối đe dọa và ngăn chặn thiệt hại có thể xảy ra.

Hệ thống phát hiện xâm nhập thường hoạt động dựa trên hai cơ chế chính, bao gồm phát hiện dựa trên chữ ký và phát hiện dựa trên hành vi bất thường.

* + - * **Phát hiện dựa trên chữ ký (Signature-based Intrusion Detection):** Là phương pháp nhận dạng mối đe dọa thông qua việc đối chiếu lưu lượng mạng hoặc nhật ký hệ thống với các mẫu tấn công đã biết. Mỗi mẫu (signature) được xây dựng dựa trên đặc điểm của một loại tấn công cụ thể, có thể là chuỗi lệnh độc hại, đoạn mã hay mẫu byte đặc trưng. Phương pháp này giúp hệ thống nhanh chóng nhận diện và cảnh báo các cuộc tấn công đã được ghi nhận trước đó với độ chính xác cao.
      * **Phát hiện dựa trên sự bất thường (Anomaly-based intrusion detection):** được thiết kế để xác định các cuộc tấn công không xác định, chẳng hạn như phần mềm độc hại mới và thích ứng với chúng ngay lập tức bằng cách sử dụng máy học. Thông qua các kỹ thuật máy học cho phép Hệ thống phát hiện xâm nhập (IDS) tạo ra các đường cơ sở của mô hình tin cậy, sau đó so sánh hành vi mới với các mô hình tin cậy đã được xác minh. Cảnh báo giả có thể xảy ra khi sử dụng IDS dựa trên sự bất thường, vì lưu lượng mạng hợp pháp chưa từng được biết đến trước đây cũng có thể bị xác định sai là hoạt động độc hại.
    1. ***So sánh signature-based và anomaly-based detection:***

|  |  |  |
| --- | --- | --- |
| **Tiêu chí** | **Signature-based** | **Anomaly-based** |
| Nguyên lý | So khớp mẫu tấn công đã biết | So sánh hành vi với mô hình bình thường |
| Ưu điểm | Phát hiện nhanh, ít cảnh báo sai | Phát hiện tấn công mới, chưa có mẫu |
| Nhược điểm | Không phát hiện zero-day | Dễ cảnh báo nhầm (false positive) |
| Ứng dụng | IDS truyền thống, antivirus | IDS hiện đại, Machine Learning |

***B. Giới thiệu về hệ thống ngăn ngừa xâm nhập (Intrusion Prevention Systems - IPS):***

* + 1. ***Khái niệm :***

Hệ thống ngăn ngừa xâm nhập (Intrusion Prevention System – IPS) là một giải pháp bảo mật có chức năng giám sát và chủ động ngăn chặn các hoạt động tấn công hoặc xâm nhập trái phép vào hệ thống mạng. Không chỉ dừng lại ở việc phát hiện như IDS, IPS còn có khả năng phản ứng tự động, ngăn chặn các hành vi nguy hại ngay khi chúng xảy ra.

Cụ thể, IPS liên tục theo dõi lưu lượng dữ liệu đi qua hệ thống, phân tích và nhận diện những hoạt động đáng ngờ. Khi phát hiện dấu hiệu tấn công, hệ thống sẽ tự động kích hoạt biện pháp phòng vệ như chặn gói tin, ngắt kết nối hoặc cập nhật quy tắc trong tường lửa (firewall) để ngăn chặn nguồn tấn công. Đồng thời, IPS cũng lưu lại thông tin chi tiết về sự kiện và tạo báo cáo giám sát an ninh giúp quản trị viên dễ dàng đánh giá và xử lý.

Về bản chất, IPS được xem như phiên bản mở rộng của IDS, bởi ngoài chức năng giám sát và phát hiện, nó còn thực hiện hành động phản hồi dựa trên dữ liệu mà IDS phát hiện được. Các quy tắc (rules) sử dụng trong IPS thường tương tự IDS nhưng được thiết kế để tự động thực thi hành động ngăn chặn thay vì chỉ cảnh báo.

* + 1. ***Phân loại hệ thống ngăn ngừa xâm nhập :***

Hệ thống IPS có thể được triển khai dưới hai hình thức chính, tùy thuộc vào phạm vi giám sát và đối tượng bảo vệ:

* + - * **Hệ thống ngăn ngừa xâm nhập mạng** (NIPS – Network-based Intrusion Prevention): NIPS được đặt tại vị trí trung gian trên mạng, thường là trước hoặc sau tường lửa (firewall), để giám sát toàn bộ lưu lượng dữ liệu ra vào hệ thống. Khi triển khai trước firewall, NIPS có thể bảo vệ toàn bộ hạ tầng bên trong, bao gồm cả vùng DMZ và thiết bị tường lửa. Ngược lại, khi triển khai sau firewall, NIPS giúp phát hiện và ngăn chặn các cuộc tấn công đến từ bên trong mạng, chẳng hạn như các thiết bị di động kết nối qua VPN bị khai thác lỗ hổng.
      * **Hệ thống ngăn ngừa xâm nhập host** (HIPS – Host-based Intrusion Prevention):HIPS được cài đặt trực tiếp trên máy chủ hoặc máy trạm để theo dõi và bảo vệ riêng từng thiết bị. Hệ thống này hoạt động tương tự phần mềm chống virus, có khả năng phát hiện các hành vi bất thường, ngăn chặn mã độc, và cảnh báo khi có sự thay đổi bất hợp lệ trong tệp tin cấu hình hoặc hệ thống. Nhờ đặc điểm hoạt động độc lập trên từng host, HIPS đặc biệt hiệu quả trong việc bảo vệ các máy chủ có tính quan trọng cao hoặc chứa dữ liệu nhạy cảm.
    1. ***Các ưu điểm và hạn chế của hệ thống ngăn ngừa xâm nhập - IPS:***

**Ưu điểm:**

* + - * Cung cấp giải pháp bảo vệ toàn diện hơn đối với tài nguyên hệ thống.
      * Ngăn chặn kịp thời các tấn công đã biết hoặc chưa được biết.

**Hạn chế:**

* + - * Có thể gây ra tình trạng phát hiện nhầm (faulse positives), có thể không cho phép các truy cập hợp lệ tới hệ thống.

1. **KỸ THUẬT PHÁT HIỆN DỰA TRÊN BẤT THƯỜNG (ANOMALY-BASED INTRUSION DETECTION):**

Anomaly detection là gì? Đây là một thuật ngữ kỹ thuật và công nghệ có nghĩa là phát hiện bất thường (Đây là nghĩa tiếng Việt của Anomaly Detection - một thuật ngữ thuộc nhóm Technology Terms - Công nghệ thông tin). Anomaly detection (Phát hiện bất thường) là việc xác định các điểm dữ liệu, quan sát, các mục hoặc các sự kiện không phù hợp với các mô hình dự kiến của một nhóm nhất định. Những bất thường xảy ra thường xuyên nhưng có thể biểu hiện một mối đe dọa lớn và giống như sự xâm nhập mạng, gian lận.Phát hiện bất thường chủ yếu là quá trình khai thác dữ liệu và sử dụng để xác định các loại dị thường xảy ra trong một file dữ liệu nhất định và để xác định chi tiết về sự xuất hiện dị thường của chúng.

***2.3.1 Tầm quan trọng của việc phương pháp phát hiện dựa trên bất thường:***

Phát hiện bất thường đóng vai trò đặc biệt quan trọng đối với các ngành như tài chính, bán lẻ và an ninh mạng, tuy nhiên, mọi doanh nghiệp cần cân nhắc giải pháp phát hiện bất thường. Giải pháp này cung cấp một phương tiện tự động phát hiện các giá trị ngoại lai có hại và bảo vệ dữ liệu của bạn.

Phát hiện bất thường xác định hoạt động khả nghi khác với khuôn mẫu hành vi bình thường đã thiết lập của bạn. Một giải pháp bảo vệ hệ thống của bạn theo thời gian thực khỏi những phiên bản có thể dẫn đến thiệt hại đáng kể về tài chính, vi phạm dữ liệu hoặc các sự kiện có hại khác.

***2.3.2 Những ưu điểm và nhược điểm của phát hiện dựa trên bất thường:***

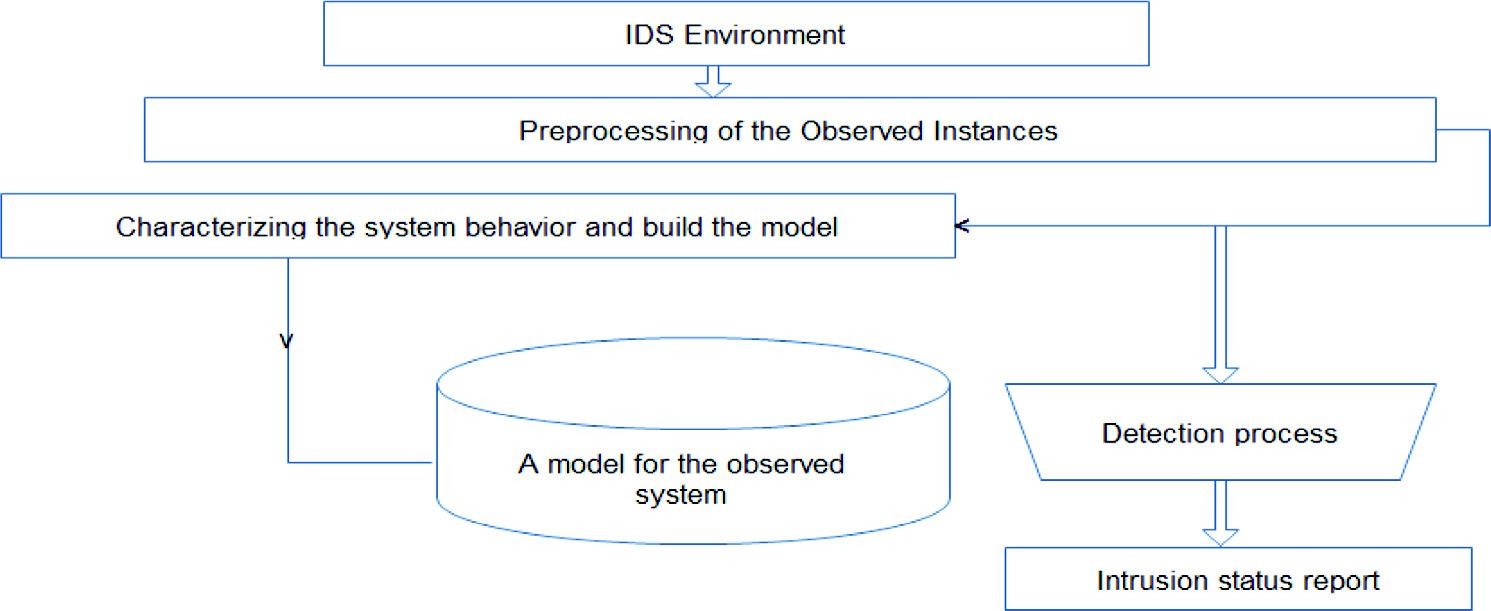
Những ưu điểm khi có thể kể đến như :

* Do phát hiện dựa trên bất thường nên có thể phát hiện nhiều cuộc tấn công mới mà không cần biết chi tiết về cuộc tấn công hoặc những dấu hiệu riêng (signature) của hình thức tấn công.
* Thông tin kết quả của kĩ thuật phát hiện bất thường có thể sử dụng làm mẫu trong kĩ thuật phát hiện sự lạm dụng hay dùng để xây dựng cơ sở dữ liệu chứa dấu hiệu riêng.

Nhược điểm của phát hiện xâm nhập dựa trên bất thường có thể kể đến như :

* Trong quá trình sử dụng thường đưa ra những cảnh báo sai đối với những hành vi mang tính bình thường của hệ thống hoặc của người dùng mà không thể dự đoán được.
* Tính chất phức tạp đòi hỏi phải mở rộng các bản ghi sự kiện hành vi được coi là bình thường của hệ thống.

***2.3.3 Mô hình của một IDS dạng Anomaly-Based Intrusion Detection:***

******

*Hình 2. Mô hình cơ bản của một IDS dạng Anomallly-Based Detection.*

Mô hình hoạt động dựa trên 3 giai đoạn chính :

* Formation Of Attributes (giai đoạn tiền xử lý các thuộc tính, dữ liệu) : các dữ liệu sẽ được sàng lọc và phân loại theo như cầu cần được xử lý.
* Observation Stage (giai đoạn quan sát) : dựa trên hình thức tự động hóa hoặc thông qua các mô hình (model).
* Functional Stage (giai đoạn chức năng) : giai đoạn thực hiện các chức năng phát hiện tùy theo model được dựng sẵn.

***2.3.4 Tổng quan cơ bản về Threshold (ngưỡng) và Profile (hồ sơ):***

+ Threshold (ngưỡng): là một giá trị giới hạn được thiết lập để xác định ranh giới giữa trạng thái bình thường và trạng thái bất thường. Khi một tham số giám sát (chẳng hạn như lưu lượng mạng, số lượng truy cập, hoặc nhiệt độ hệ thống) vượt quá giá trị ngưỡng này, hệ thống sẽ coi đó là dấu hiệu bất thường và phát cảnh báo. Ngưỡng có thể được thiết lập thủ công dựa trên kinh nghiệm quản trị hoặc tự động tính toán thông qua các thuật toán học máy hoặc thống kê.

+ Profile (hồ sơ): là mô hình mô tả hành vi bình thường của người dùng hoặc hệ thống, được xây dựng dựa trên dữ liệu lịch sử hoặc quan sát trong điều kiện vận hành ổn định. Khi một hành vi hoặc giá trị đo được lệch đáng kể so với hồ sơ đã học, hệ thống sẽ coi đó là dấu hiệu bất thường. Phương pháp này thường được sử dụng trong các kỹ thuật anomaly-based detection, nơi hành vi được theo dõi liên tục để phát hiện sai lệch theo thời gian.

+ Khác biệt lớn nhất: threshold dựa trên một số liệu cố định và profile dựa trên mô hình về hành vi bình thường.

+ Ví dụ về trường hợp liên quan đến Threshold:

* Giả sử bạn đang theo dõi nhiệt độ trong phòng máy chủ. Dữ liệu trung bình được ghi nhận ổn định ở mức 22°C. Một ngày, cảm biến báo nhiệt độ tăng lên 40°C do hỏng thiết bị làm mát. Khi đó, nếu hệ thống thiết lập ngưỡng cảnh báo ở mức 25°C, nó sẽ ngay lập tức phát hiện sự bất thường này và gửi cảnh báo đến quản trị viên.

+ Ví dụ về trường hợp liên quan đến Profile (hồ sơ):

* Xem xét tài khoản thẻ tín dụng của một người dùng tên John, 35 tuổi, sống tại New York, làm việc trong lĩnh vực tiếp thị. Lịch sử chi tiêu của John cho thấy thu nhập trung bình 2.000 USD/tháng và các khoản thanh toán thường xuyên gồm siêu thị, ăn uống, và giải trí. Tuy nhiên, gần đây John phát sinh giao dịch trị giá 10.000 USD tại cửa hàng trang sức cao cấp và thực hiện nhiều giao dịch nhỏ tại sòng bạc ở Las Vegas. Các hành vi này khác biệt rõ rệt so với hồ sơ chi tiêu thông thường, nên hệ thống phát hiện gian lận sẽ gắn cờ cảnh báo vì có khả năng tài khoản bị xâm phạm.

1. **CÁC KỸ THUẬT XÁC ĐỊNH THRESHOLD VÀ PROFILES :**

***2.4.1 Kỹ thuật dựa trên thống kê (Statistical-based techniques):***

**- Khái niệm:** Sử dụng các kĩ thuật thống kê như giá trị trung bình và phương sai để xây dựng một profile bình thường.

**- Hoạt động:** Hệ thống tiến hành thu thập dữ liệu hoạt động trong giai đoạn bình thường và tính toán các thông số thống kê tương ứng cho từng biến (chẳng hạn như số lượng gói tin, tần suất truy cập hoặc thời gian phản hồi). Sau đó, khi hệ thống hoạt động, mỗi sự kiện hoặc hành vi mới sẽ được so sánh với giá trị trung bình đã xác định. Nếu vượt quá ngưỡng thống kê cho phép (ví dụ vượt quá trung bình ± 3 lần độ lệch chuẩn), IDS sẽ phát cảnh báo.

**- Ưu điểm:**

* Không cần biết về signature của cuộc tấn công (có thể phát hiện được zero day attack).
* Việc ko cần phụ thuộc vào signature nên update không cần thường xuyên.
* Các cuộc tấn công xảy ra trong thời gian dài có thể xác định chính xác như dos.

**- Nhược điểm:**

* Các bản thống kê phân phối cần tính chính xác cao.
* Quá trình tạo profile dựa trên thống kê cần thời gian dài để chính xác và hiệu quả.

***2.4.2 Kỹ thuật dựa trên nhận thức (Cognitive-based or knowledge-based techniques):***

**- Khái niệm**: Nói gọn là kinh nghiệm rút từ những cuộc tấn công và các lỗ hổng đã được tìm thấy. Có thể phân loại thành : phân tích chuyển đổi trạng thái hệ thống và phân tích chữ ký.

**- Hoạt động**: Tổng hợp các thông tin về cuộc tấn công hay lỗ hổng đã được tìm thấy. Từ đó xây dựng những biện pháp để phát hiện xâm nhập hoặc tấn công. Khi phát hiện phù hợp sẽ tạo báo động.

**- Ưu điểm**: Độ chính xác cao và sai số thấp và giúp nhà phân tích dễ dàng bảo mật tốt hơn.

**- Nhược điểm**: Do dựa trên tri thức (nghiên cứu) từ những cuộc tấn công nên tốn nhiều thời gian. Tìm được đủ kiến thức cũng là một vấn đề.

***2.4.3 Kỹ thuật dựa trên khai thác dữ liệu (Data mining-based techniques):***

**- Khái niệm**: quá trình phân loại, sắp xếp các tập hợp dữ liệu lớn để xác định các mẫu và thiết lập các mối liên hệ nhằm giải quyết các vấn đề nhờ phân tích dữ liệu.

**- Các bước thực hiện:**

* Làm sạch dữ liệu : làm sạch tạp hoặc bất thường.
* Tích hợp dữ liệu : kết hợp nhiều nguồn.
* Chuyển đổi dữ liệu : để phân tích và tổng hợp.
* Khai phá dữ liệu : trích xuất từ hiện có.
* Đánh giá mẫu : thỏa yêu cầu.
* Trình bày thông tin : dạng cây, bảng, biểu đồ và ma trận.

**- Ưu điểm:**

* Xử lý được dữ liệu lớn.
* Xử lý nhanh hơn vì sử dụng những mô hình tính toán được thiết kế phù hợp trong giai đoạn đào tạo.
* Có thể tạo mẫu ở chế độ không giám sát.

**- Nhược điểm:**

* Là sản phẩm phụ của quá trình phân cụm nên không được tối ưu để phát hiện bất thường.
* Yêu cầu lưu trữ cao và phân loại chậm do kích thước.

1. **CÁC VẤN ĐỀ VÀ THÁCH THỨC ĐỐI VỚI PHÁT HIỆN BẤT THƯỜNG TRONG HỆ THỐNG IDS/IPS:**

Các thách thức có thể kể đến như :

* Có thể giảm tỷ lệ cảnh báo sai nhưng không giảm thiểu báo động sai → nhiệm vụ thách thức.
* Phát triển một phương pháp chung hay lấy dữ liệu ban đầu cũng là một thách thức.
* Mẫu mới được cập nhập trên cơ sở dữ liệu mà không cần hiệu suất cao là một thách thức.
* Các thuật toán còn độ phức tạp cao trong việc tiền xử lý dữ liệu trong giai đoạn đào tạo và triển khai.
* Xây dựng phương pháp phù hợp cho từng loại tấn công.

1. **MACHINE LEARNING LÀ GÌ VÀ ỨNG DỤNG MACHINE LEARNING TRONG PHÁT HIỆN VÀ NGĂN CHẶN XÂM NHẬP (IDS/IPS):**

Machine Learning (học máy) là một lĩnh vực thuộc Trí tuệ nhân tạo (Artificial Intelligence – AI), cho phép hệ thống máy tính có khả năng tự học từ dữ liệu và cải thiện hiệu suất theo thời gian mà không cần lập trình trực tiếp. Thay vì dựa vào các quy tắc cố định do con người thiết lập, các mô hình học máy phân tích dữ liệu quá khứ để nhận biết mẫu (pattern), mối quan hệ và hành vi, từ đó dự đoán hoặc đưa ra quyết định đối với dữ liệu mới.

Về cơ bản, Machine Learning được chia thành ba nhóm chính:

* **Học có giám sát (Supervised Learning):** mô hình được huấn luyện trên tập dữ liệu đã được gán nhãn (labelled data), ví dụ phân loại email “bình thường” hay “spam”.
* **Học không giám sát (Unsupervised Learning):** mô hình tự tìm kiếm cấu trúc ẩn trong dữ liệu chưa được gán nhãn, thường dùng để phát hiện bất thường hoặc gom cụm dữ liệu tương đồng.
* **Học tăng cường (Reinforcement Learning):** hệ thống học thông qua tương tác với môi trường và phản hồi (reward/punishment) nhằm tối ưu hành vi theo thời gian.

Trong lĩnh vực an ninh mạng, đặc biệt là các hệ thống phát hiện và ngăn chặn xâm nhập (IDS/IPS), Machine Learning đóng vai trò quan trọng trong việc tự động hóa quá trình nhận dạng và phân tích hành vi tấn công. Thay vì chỉ dựa vào mẫu tấn công có sẵn như phương pháp signature-based, các mô hình học máy có thể học và phát hiện các hành vi bất thường chưa từng xuất hiện (zero-day attacks), giúp nâng cao độ chính xác và khả năng thích ứng của hệ thống.

Một hướng nghiên cứu phổ biến là kết hợp học có giám sát và học không giám sát để cải thiện hiệu quả phát hiện. Trong đó, unsupervised learning được sử dụng để phát hiện các mẫu dữ liệu lạ hoặc bất thường, còn supervised learning được huấn luyện để phân loại, gán nhãn và đánh giá mức độ rủi ro của từng sự kiện. Cách tiếp cận này cho phép hệ thống tự động sinh và ưu tiên tập luật phát hiện (rules) dựa trên các đặc trưng mạng như địa chỉ IP, mẫu payload, giao thức và cổng truy cập (port patterns).

Nhờ đó, hệ thống IDS/IPS không chỉ dừng lại ở mức phát hiện và cảnh báo mà còn có khả năng tự học (self-learning) và tự tối ưu (self-optimization) theo thời gian. Việc kết hợp dữ liệu lớn (Big Data) và học máy giúp giảm đáng kể tỷ lệ cảnh báo giả (false positive), đồng thời tăng cường khả năng nhận diện tấn công phức tạp và tấn công mới chưa từng được biết đến. Đây là hướng phát triển tất yếu trong quá trình xây dựng hệ thống phát hiện và phòng chống xâm nhập thông minh, đáp ứng yêu cầu bảo mật của các mạng hiện đại trong kỷ nguyên chuyển đổi số.