|  |
| --- |
| BAN CƠ YẾU CHÍNH PHỦ  **HỌC VIỆN KỸ THUẬT MẬT MÃ**  ¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯  Logo HvKTMM  CHUYÊN ĐỀ CHUYÊN NGÀNH CHUYÊN SÂU  **NGHIÊN CỨU THU THẬP DỮ LIỆU**  **TRONG GIÁM SÁT AN TOÀN THÔNG TIN**  Ngành: An toàn thông tin  *Sinh viên thực hiện:*  **Đỗ Hoài Nam - MSSV: AT170636**  *Người hướng dẫn:*  **TS. Phạm Duy Trung**  Khoa An toàn thông tin – Học viện kỹ thuật mật mã  **Hà Nôi, 2025** |
| BAN CƠ YẾU CHÍNH PHỦ  **HỌC VIỆN KỸ THUẬT MẬT MÃ**  ¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯  Logo HvKTMM  CHUYÊN ĐỀ CHUYÊN NGÀNH CHUYÊN SÂU  **NGHIÊN CỨU THU THẬP DỮ LIỆU**  **TRONG GIÁM SÁT AN TOÀN THÔNG TIN**  Ngành: An toàn thông tin  *Sinh viên thực hiện:*  **Đỗ Hoài Nam - MSSV: AT170636**  *Người hướng dẫn:*  **TS. Phạm Duy Trung**  Khoa An toàn thông tin – Học viện kỹ thuật mật mã  **Hà Nôi, 2025** |

# MỤC LỤC

[MỤC LỤC 3](#_Toc207930021)

[DANH MỤC HÌNH ẢNH 5](#_Toc207930022)

[LỜI CẢM ƠN 6](#_Toc207930023)

[LỜI NÓI ĐẦU 7](#_Toc207930024)

[CHƯƠNG 1: TỔNG QUAN VỀ AN TOÀN THÔNG TIN VÀ GIÁM SÁT AN TOÀN THÔNG TIN 8](#_Toc207930025)

[1.1. Các khái niệm cơ bản trong an toàn thông tin 8](#_Toc207930026)

[1.2. Thực trạng an toàn thông tin tại Việt Nam 11](#_Toc207930027)

[1.3. Xu hướng sử dụng dịch vụ giám sát ATTT mạng trong bối cảnh nhiều rủi ro về bảo mật 12](#_Toc207930028)

[1.4. Các mối đe doạ và kỹ thuật tấn công mạng phổ biến hiện nay 13](#_Toc207930029)

[1.5. Các giải pháp phòng chống tấn công mạng 16](#_Toc207930030)

[1.5.1. Chống mã độc và thư rác (Antivirus, Antispam) 16](#_Toc207930037)

[1.5.2. Tường lửa (Firewall) 18](#_Toc207930038)

[1.5.3. Phát hiện và ngăn chặn xâm nhập (IDPS) 20](#_Toc207930039)

[1.5.4. Hệ thống bẫy (Honeypot) 21](#_Toc207930040)

[1.5.5. Mạng riêng ảo VPN 23](#_Toc207930041)

[1.5.6. Chống rò rỉ dữ liệu (DLP) 24](#_Toc207930042)

[1.6. Kết luận chương 1 26](#_Toc207930049)

[CHƯƠNG II: THU THẬP DỮ LIỆU TRONG GIÁM SÁT AN TOÀN THÔNG TIN 27](#_Toc207930050)

[2.1. Giới thiệu chung về thu thập dữ liệu trong giám sát 27](#_Toc207930052)

[2.1.1. Nhật ký hệ thống ( System logs collection) 27](#_Toc207930053)

[2.1.2. Nhật ký ứng dụng (Application Logs) 27](#_Toc207930054)

[2.1.3. Dữ liệu mạng (Network-based Logs / Network Flow & Packet) 28](#_Toc207930055)

[2.1.4. Nhật ký thiết bị / Iot ( Device/IoT Logs ) 28](#_Toc207930056)

[2.1.5. Dữ liệu bảo mật khác 28](#_Toc207930057)

[2.2. Phương pháp thu thập dữ liệu giám sát 29](#_Toc207930058)

[2.2.1. Thu thập trực tiếp (Direct Collection) 29](#_Toc207930059)

[2.2.2. Thu thập gián tiếp (Indirect Collection) 34](#_Toc207930060)

[2.2.3. Thu thập thủ công/định kỳ (Manual or Report-based Collection) 37](#_Toc207930061)

[2.2.4. Đánh giá các phương pháp thu thập 39](#_Toc207930062)

[2.3. Kết luận chương 2 40](#_Toc207930067)

[CHƯƠNG III: LƯU TRỮ VÀ PHÂN LOẠI DỮ LIỆU GIÁM SÁT 41](#_Toc207930068)

[3.1. Giới thiệu chung 41](#_Toc207930072)

[3.2. Lưu trữ dữ liệu giám sát 42](#_Toc207930073)

[3.2.1. Các mô hình lưu trữ log 42](#_Toc207930074)

[3.2.2. Chính sách lưu trữ 45](#_Toc207930075)

[3.3. Phân loại dữ liệu giám sát 46](#_Toc207930076)

[3.3.1. Theo nguồn gốc 46](#_Toc207930077)

[3.3.2. Theo mức độ quan trọng 46](#_Toc207930078)

[3.3.3. Theo mục đích giám sát 47](#_Toc207930079)

[3.4. Kết luận chương 3 48](#_Toc207930080)

[KẾT LUẬN 49](#_Toc207930081)

[TÀI LIỆU THAM KHẢO 50](#_Toc207930082)

[PHỤ LỤC 51](#_Toc207930083)

# DANH MỤC HÌNH ẢNH

[Hình 1. 1: Các biện pháp bảo vệ an toàn thông tin 9](#_Toc207929580)

[Hình 1. 2 Minh họa việc chặn gói tin tộc hại của tưởng lửa 17](#_Toc207929581)

[Hình 2. 1: Hiển thị thông tin log windows trong Event Viewer 29](#_Toc207929736)

[Hình 2. 2: Hiển thị bản ghi chi tiết với Event ID, User, TimeCreated, Message 30](#_Toc207929737)

[Hình 2. 3: Thống kê IP Flow Cache 32](#_Toc207929738)

[Hình 2. 4: Thu thập trực tiếp packet bằng tcpdump 33](#_Toc207929739)

[Hình 2. 5: Wazuh phát hiện trên máy Windows11 đã xảy ra nhiều sự kiện đăng nhập thất bại 35](#_Toc207929740)

# LỜI CẢM ƠN

Trong quá trình thực hiện đề tài chuyên đề này, em đã nhận được sự giúp đỡ tận tình của cán bộ hướng dẫn là TS. Phạm Duy Trung - Giảng viên Khoa An toàn thông tin - Học viện Kỹ thuật Mật mã đã tận tình hướng dẫn và hỗ trợ em trong việc lựa chọn đề tài, thực hiện triển khai và hoàn thiện báo cáo chuyên đề của mình.

Trong quá trình hoàn thiện đề tài, em đã cố gắng tìm những giải pháp thật tốt. Song với kiến thức và khả năng còn hạn chế, báo cáo của em không tránh khỏi những thiếu sót cần phải cải tiến thêm để hoàn thiện hơn. Em kính mong nhận được những góp ý của các thầy cô và những người quan tâm đến báo cáo này.

Em xin chân thành cảm ơn!.

**SINH VIÊN THỰC HIỆN CHUYÊN ĐỀ**

Đỗ Hoài Nam

# LỜI NÓI ĐẦU

Trong bối cảnh chuyển đổi số mạnh mẽ, an toàn thông tin đã và đang trở thành vấn đề được quan tâm hàng đầu tại Việt Nam. Các cuộc tấn công mạng ngày càng tinh vi, đặc biệt với sự hỗ trợ của trí tuệ nhân tạo (AI), gây ra những thách thức lớn cho công tác bảo vệ hệ thống thông tin quan trọng quốc gia cũng như các nền tảng số phục vụ người dân.

Theo đánh giá của Bộ Thông tin và Truyền thông, cùng với nhận định từ Hiệp hội An toàn thông tin Việt Nam (VNISA), bảo đảm an toàn thông tin cho hạ tầng dữ liệu và nền tảng số quốc gia chính là nền móng để bảo vệ không gian mạng, đồng thời thúc đẩy quá trình chuyển đổi số bền vững. Do đó, việc giám sát an toàn thông tin, đặc biệt là thu thập dữ liệu phục vụ giám sát, cần được đặt lên hàng đầu trong mọi chiến lược phát triển hạ tầng số.

Xuất phát từ thực tiễn trên, đề tài “**Nghiên cứu thu thập dữ liệu trong giám sát an toàn thông tin**” được thực hiện nhằm phân tích thực trạng, làm rõ vai trò, và đề xuất các giải pháp nâng cao hiệu quả thu thập dữ liệu, qua đó góp phần tăng cường năng lực bảo đảm an toàn thông tin tại Việt Nam.

# CHƯƠNG 1: TỔNG QUAN VỀ AN TOÀN THÔNG TIN VÀ GIÁM SÁT AN TOÀN THÔNG TIN

## Các khái niệm cơ bản trong an toàn thông tin

Trong suốt chiều dài lịch sử tồn tại và phát triển của mình, con người sử dụng thông tin như một công cụ để trao đổi và truyền đạt kiến thức. Trong kỉ nguyên bùng nổ thông tin hiện nay, thông tin dã trở thành một trong những nhu cầu, phương tiện sản xuất sống còn của con người. Tất cả các cá nhân hay doanh nghiệp đều sử dụng thông tin – Ví dụ, thông tin cá nhân, thông tin khách hàng,…Nếu thông tin đó bị xâm phạm một cách trái phép bằng một cách nào đấy có thể khiến cho hoạt động của doanh nghiệp bị ảnh hưởng, thập chí sụp đổ. Việc bảo vệ những thông tin ấy khỏi những sự truy cập bất hợp pháp được gọi là An toàn thông tin (Information Security).

**An toàn thông tin – Information Security**, được định nghĩa là việc bảo vệ thông tin và các hệ thống thông tin tránh khỏi việc bị truy nhập, sử dụng, tiết lộ, gián đoạn, sửa đổi hoặc phá hoại trái phép. Hiểu một cách nôm na an toàn thông tin là quá trình phát hiện và ngăn chặn mọi hành vi sử dụng trái phép máy tính máy tính/ điện thoại hoặc thiết bị chứa thông tin. Nó liên quan đến quá trình bảo vệ chống lại những kẻ xâm phạm sử dụng tài nguyên máy tính cá nhân hoặc văn phòng của bạn với mục đích xấu hoặc vì lợi ích riêng bất hợp pháp, hoặc thậm chí do vô tình. [1]

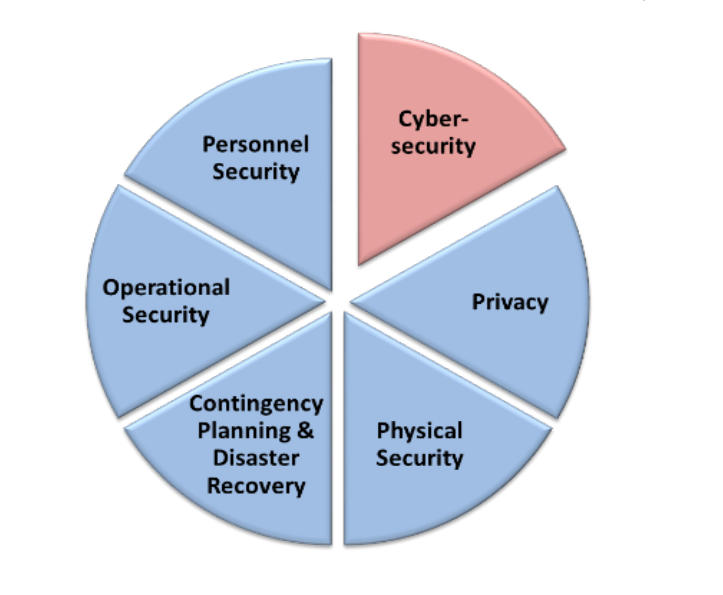
Tính toàn vẹn, tính bí mật và tính khả dụng là 3 tính chất cơ sở cốt lõi của an toàn thông tin. Trong đó:

* **Tính bí mật - Confidentiality**: là tính chất đảm bảo thông tin chỉ cung cấp cho những người có thẩm quyền. Điều này bao gồm việc đảm bảo những người không có thẩm quyền sẽ không có khả năng tiếp cận với các dữ liệu cá nhân hoặc bí mật (tính bí mật) và người sở hữu dữ liệu riêng tư có toàn quyền lưu trữ, sử dụng, và cấp phép cho người khác được tiếp cận với dữ liệu đó (tính riêng tư hoặc liên quan đến vấn đề bản quyền). Dữ liệu hoặc hệ thống bị mất tính bí mật khi nó bị tiếp cận hoặc tiết lộ trái phép.
* **Tính toàn vẹn - Integrity** là tính chất đảm bảo thông tin không bị thay đổi một cách trái phép hoặc thay đổi không như ý muốn. Bảo vệ thông tin chống lại việc sửa đổi hoặc phá hủy trái phép hoặc vô ý, bao gồm cả việc đảm bảo tính xác thực và chống chối bỏ của thông tin. Mất tính toàn vẹn xảy ra khi thông tin bị sửa đổi hoặc phá hủy trái phép hoặc sai lệch do lỗi đường truyền. Điều này bao gồm việc đảm bảo dữ liệu không bị thay đổi và đảm bảo cả về nguồn gốc của thông tin. Tính toàn vẹn cũng bao gồm việc đảm bảo rằng một hệ thống thực hiện đúng và đầy đủ các chức năng được thiết kế mà không có sự cố ý hoặc vô tình thao túng trái phép hệ thống.
* **Tính khả dụng - Availability** đảm bảo khả năng truy cập thông tin, tính năng của hệ thống thông tin mỗi khi người dùng hợp lệ có nhu cầu.

Ba tính chất bí mật, toàn vẹn và khả dụng của thông tin được gọi là tam giác CIA (Confidentiality - Integrity - Availability) và là 3 tính chất cốt lõi quan trọng nhất của an toàn thông tin. Mọi vấn đề liên quan đến việc đảm bảo an toàn thông tin đều xoay quanh việc đảm bảo 3 tính chất này của thông tin. Với việc càng ngày càng nhiều thông tin được số hóa, lưu trữ, xử lý và trảo đổi trực tuyến, khái niệm **An ninh mạng** (Cybersecurity) đã ra đời và trở thành một yếu tố chủ chốt của an toàn thông tin.

**An ninh mạng** – **Cybersecurity:** được định nghĩa là phương pháp bảo vệ an toàn cho máy tính, mạng, ứng dụng phần mềm, hệ thống quan trọng và dữ liệu khỏi các mối đe dọa kỹ thuật số tiềm ẩn. Các tổ chức chịu trách nhiệm bảo mật dữ liệu để duy trì lòng tin của khách hàng cũng như đáp ứng việc tuân thủ quy định. Họ sử dụng các biện pháp và công cụ an ninh mạng để bảo vệ dữ liệu nhạy cảm khỏi bị truy cập trái phép cũng như ngăn chặn gián đoạn trong hoạt động kinh doanh gây ra bởi hoạt động mạng ngoài ý muốn. Các tổ chức triển khai an ninh mạng bằng cách hợp lý hóa công tác phòng vệ kỹ thuật số giữa con người, quy trình và công nghệ. [2]

Là một phần của an toàn thông tin, an ninh mạng hoạt động kết hợp với nhiều kiểu bảo mật khác, một trong số đó được biểu diễn trong hình 1.1. Tổng thể, những thành phần giúp đảm bảo an toàn thông tin này cung cấp phương án phòng thủ trước nhiều mối đe dọa tới dữ liệu thông tin của tổ chức.



Hình 1. 1: Các biện pháp bảo vệ an toàn thông tin

• **An ninh vật lý (Physical Security)** – Biện pháp bảo vệ vật lý của tài sản như ổ khóa, hàng rào,….

• **An ninh nhân sự (Personnel Security)** – Việc đánh giá hành vi, tính chính trực, phán đoán, lòng trung thành, độ tin cậy và sự ổn định của các cá nhân đối với các nhiệm vụ và trách nhiệm đòi hỏi sự đáng tin cậy.

• **Phương án dự phòng và phục hồi sau thảm họa (Contingency Planning and Disaster Recovery)** – phương án khôi phục hoạt động bình thường của doanh nghiệp sau khi sự cố hay còn được biết đến là kế hoạch kinh doanh liên tục (Business Continuity Planning)

• **An ninh vận hành (Operational Security)** – Bảo vệ các kế hoạch kinh doanh, quy trình vận hành hằng ngày.

• **Riêng tư (Privacy)** – Bảo vệ các thông tin cá nhân

Chỉ cần thiếu một trong những thành phần này sẽ làm giảm độ hiệu quả các thành phần còn lại. Ví dụ, phương án bảo vệ an ninh vật tốt nhưng cũng không có ý nghĩa nếu nhân sự mà bạn tuyển có ý định làm hại đến công việc của tổ chức (an ninh nhân sự kém).

## Thực trạng an toàn thông tin tại Việt Nam

Trong bối cảnh Việt Nam đẩy mạnh chuyển đổi số trên mọi lĩnh vực, an ninh mạng đã trở thành một trụ cột chiến lược, giữ vai trò thiết yếu trong bảo đảm phát triển bền vững, ổn định của đất nước, bảo vệ chủ quyền quốc gia trên không gian mạng và quyền lợi hợp pháp của mọi tổ chức, cá nhân.

Tuy nhiên, thực tế hiện nay đang đặt ra những thách thức nghiêm trọng. Tình trạng lừa đảo chiếm đoạt tài sản qua mạng, xuyên tạc thông tin, bôi nhọ danh dự cá nhân và tổ chức trên không gian mạng diễn ra phổ biến với mức độ ngày càng nghiêm trọng, đe dọa trực tiếp an ninh trật tự quốc gia.

**Thiệt hại do tôi phạm mạng gây ra:** Số liệu thống kê chính thức cho thấy các cuộc tấn công mạng đang gia tăng đáng báo động. Năm 2023, Việt Nam ghi nhận khoảng 13.900 vụ tấn công vào các hệ thống thông tin, trung bình 1.160 vụ mỗi tháng, tăng 9,5% so với năm 2022.

**Mục tiêu tấn công nhằm vào dữ liệu cá nhân và hệ thống chính phủ:** Đã có 554 website của các cơ quan chính phủ và giáo dục (tên miền gov.vn, edu.vn) bị tin tặc xâm nhập, chèn mã độc quảng cáo cờ bạc. Ngoài ra, hơn 83.000 máy tính, máy chủ tại Việt Nam bị nhiễm mã độc ransomware mã hóa đòi tiền chuộc trong năm 2023, tăng 8,4% so với năm 2022. Thiệt hại ở cấp độ cá nhân cũng vô cùng nghiêm trọng. Tại TP.HCM, chỉ trong 9 tháng đầu năm 2024, Công an Thành phố đã tiếp nhận và thụ lý 461 vụ lừa đảo trên không gian mạng với tổng số tiền thiệt hại khoảng 982 tỷ đồng. Trung bình mỗi vụ, nạn nhân mất khoảng 5 tỷ đồng - một con số gây sửng sốt, cho thấy mức độ tàn phá tài chính đối với từng cá nhân, gia đình là rất lớn.

Hậu quả kéo theo không chỉ là mất tiền của, mà còn là những sang chấn tâm lý, mâu thuẫn gia đình, thậm chí có trường hợp cùng quẫn dẫn đến tự tử. Những con số và hệ lụy nêu trên cho thấy tội phạm mạng đã và đang trở thành vấn nạn nhức nhối, ảnh hưởng sâu sắc đến kinh tế - xã hội.

## Xu hướng sử dụng dịch vụ giám sát ATTT mạng trong bối cảnh nhiều rủi ro về bảo mật

**Chuyển dịch từ SOC truyền thống sang SOC-as-a-Service (SOC-dịch vụ):**

Truyền thống, SOC được xây dựng nội bộ, với hạ tầng, nhân lực và quy trình quản lý hoàn toàn thuộc quyền kiểm soát của doanh nghiệp. Tuy nhiên, mô hình này tỏ ra tốn kém, đòi hỏi nhân lực cao, và mất nhiều thời gian triển khai, đặc biệt đối với các doanh nghiệp vừa và nhỏ. Để khắc phục hạn chế trên, mô hình SOC-as-a-Service (SOC dưới dạng dịch vụ thuê ngoài) trở nên phổ biến. Thay vì xây dựng SOC tại chỗ, doanh nghiệp thuê dịch vụ từ nhà cung cấp chuyên nghiệp, nhận giám sát liên tục, cập nhật công nghệ mới và hỗ trợ phản ứng sự cố kịp thời ( Giúp tiết kiệm thời gian từ 12-18 tháng so với việc xây dựng SOC truyền thống ).

**Giám sát 24/7 với sự hỗ trợ của AI và Machine Learning:**

Các hệ thống giám sát an toàn thông tin hiện đại đang dịch chuyển sang mô hình hoạt động liên tục 24/7/365, đảm bảo khả năng phát hiện và phản ứng kịp thời trước các mối đe dọa mạng. Tuy nhiên, khối lượng dữ liệu log, cảnh báo và tín hiệu từ nhiều nguồn khác nhau (hệ thống mạng, endpoint, cloud, IoT) là rất lớn, vượt quá khả năng xử lý thủ công của con người.

Để giải quyết thách thức này, các doanh nghiệp tại Việt Nam bắt đầu ứng dụng trí tuệ nhân tạo (AI) và học máy (Machine Learning) vào quy trình giám sát. Các thuật toán học máy giúp:

* Phát hiện bất thường (anomaly detection) trong lưu lượng mạng.
* Phân loại và ưu tiên cảnh báo dựa trên mức độ rủi ro.
* Dự báo xu hướng tấn công dựa trên hành vi lịch sử và nguồn tình báo mạng (Cyber Threat Intelligence).

**Phát triển sản phẩm và giải pháp an toàn thông tin nội địa:**

Một xu hướng nổi bật khác là phát triển sản phẩm và giải pháp an toàn thông tin nội địa. Nhiều doanh nghiệp Việt Nam đã tập trung xây dựng:

* Hệ thống phòng chống DDoS, tấn công mạng.
* Nền tảng giám sát SOC nội địa.
* Các công cụ bảo mật hạ tầng và cloud.

Ưu điểm của sản phẩm nội địa gồm: tùy biến phù hợp hạ tầng Việt Nam, giảm phụ thuộc vào giải pháp nước ngoài, và góp phần nâng cao tự chủ công nghệ trong an ninh mạng. Dự báo thị trường sản phẩm ATTT nội địa đạt trên 5.000 tỷ đồng trong năm 2025, phản ánh xu hướng đầu tư mạnh mẽ vào giải pháp bảo mật nội địa.

## Các mối đe doạ và kỹ thuật tấn công mạng phổ biến hiện nay

Các cuộc tấn công mạng có thể nhắm mục tiêu vào nhiều nạn nhân từ người dung cá nhân đến doanh nghiệp hoặc thậm chí là chính phủ. Khi nhắm mục tiêu vào các doanh nghiệp hoặc tổ chức khác, mục tiêu của tin tặc thường là truy cập vào các tài nguyên nhạy cảm và có giá trị của công ty, chẳng hạn như dữ liệu khách hang hoặc chi tiết thanh toán. Dưới đây là 10 kỹ thuật tấn công mạng phổ biến hiện nay:

* **Mã độc – Malware:** Mã độc là những chương trình được tạo ra với mục đích gây hại cho máy tính, mạng hoặc là server. Mã độc là loại hình tấn công phổ biến nhất trong tấn công mạng, phần lớn bởi vì thuật ngữ này bao gồm nhiều tập hợp con như ransomware, trojan, spyware, viruses, worms, keyloggers, bots, cryptojacking, và bất kỳ loại tấn công phần mềm độc hại nào khác tận dụng phần mềm theo cách độc hại.
* **Tấn công từ chối dịch vụ -** **DoS:** Tấn công từ chối dịch vụ (DoS) là một cuộc tấn công độc hại, có chủ đích, làm tràn ngập mạng với các yêu cầu sai nhằm làm gián đoạn hoạt động của hệ thống. Trong một cuộc tấn công DoS, người dung không thể thực hiện các tác vụ thông thường và cần thiết, chẳng hạn như truy cập email, trang web, tài khoản trực tuyến hoặc các tài nguyên khác do máy tính hawojc mạng bị xâm nhập vận hành. Mặc dù hầu hết các cuộc tấn công DoS không dẫn đến mất dữ liệu nhưng chúng khiến tổ chức tốn thời gian, tiền bạc và các tài nguyên khác để khôi phục các hoạt động kinh doanh quan trọng.
* **Lừa đảo –** **Phishing:** Phishing là hình thức tấn công mạng sử dụng email, SMS, điện thoại, mạng xã hội và những kỹ thuật kỹ nghệ xã hội để dụ dỗ nạn nhân chia sẻ thông tin nhạy cảm của họ ví dụ như mật khẩu hoặc số tài khoản – hoặc là tải phần mềm độc hại, thứ sẽ cài đặt viruses vào thiết bị điện tử của họ.
* **Giả mạo** – **Spoofing:** Giả mạo là một kỹ thuật mà qua đó tội phạm mạng cải trang thành một nguồn đã biết hoặc đáng tin cậy. Khi làm như vậy, tin tặc có thể tương tác với mục tiêu và truy cập vào hệ thống hoặc thiết bị của họ với mục tiêu cuối cùng là đánh cắp thông tin, tống tiền hoặc cài đặt phần mềm độc hại hoặc phần mềm có hại khác trên thiết bị.
* **Tấn công dựa trên danh tính** – **Identity-Based Attacks:** Các cuộc tấn công dựa trên danh tính cực kỳ khó phát hiện. Khi thông tin đăng nhập hợp lệ của người dùng bị xâm phạm và kẻ thù đang giả mạo người dùng đó, thường rất khó phân biệt giữa hành vi điển hình của người dùng và hành vi của tin tặc bằng cách sử dụng các công cụ và biện pháp bảo mật truyền thống.
* **Tiêm mã** – **Code Injection Attacks:** Các cuộc tấn công tiêm mã bao gồm kẻ tấn công tiêm mã độc vào máy tính hoặc mạng dễ bị tấn công để thay đổi hành động của nó.
* **Tấn công chuỗi cung ứng** – **Supply Chain Attacks:** Tấn công chuỗi cung ứng là một loại tấn công mạng nhắm vào nhà cung cấp bên thứ ba đáng tin cậy cung cấp dịch vụ hoặc phần mềm quan trọng cho chuỗi cung ứng. Các cuộc tấn công chuỗi cung ứng phần mềm tiêm mã độc vào một ứng dụng để lây nhiễm cho tất cả người dùng ứng dụng, trong khi các cuộc tấn công chuỗi cung ứng phần cứng xâm phạm các thành phần vật lý cho cùng mục đích. Chuỗi cung ứng phần mềm đặc biệt dễ bị tổn thương vì phần mềm hiện đại không được viết từ đầu: thay vào đó, nó bao gồm nhiều thành phần sẵn có, chẳng hạn như API của bên thứ ba, mã nguồn mở và mã độc quyền từ các nhà cung cấp phần mềm.
* **Mối đe dọa nội bộ** - **Insider Threats:** Các tác nhân nội bộ gây ra mối đe dọa cho một tổ chức có xu hướng độc hại về bản chất. Một số động cơ thúc đẩy bao gồm lợi ích tài chính để đổi lấy việc bán thông tin bí mật trên web đen và/hoặc ép buộc tinh thần bằng cách sử dụng các chiến thuật lừa đảo xã hội, chẳng hạn như các cuộc tấn công lấy cớ hoặc xâm phạm email doanh nghiệp (BEC). Mặt khác, một số tác nhân đe dọa nội bộ không có bản chất độc hại mà thay vào đó lại có bản chất cẩu thả, dẫn đến việc thông tin nội bộ bị xâm phạm từ bên ngoài.
* **Đường hầm DNS** – **DNS Tunneling:** Đường hầm DNS là một loại tấn công mạng tận dụng các truy vấn và phản hồi của hệ thống tên miền (DNS) để vượt qua các biện pháp bảo mật truyền thống cũng như truyền dữ liệu và mã trong mạng. Sau khi bị nhiễm, hacker có thể tự do tham gia vào các hoạt động ra lệnh và kiểm soát. Đường hầm này cung cấp cho tin tặc một lộ trình để giải phóng phần mềm độc hại và/hoặc trích xuất dữ liệu, IP hoặc thông tin nhạy cảm khác bằng cách mã hóa từng chút một trong một loạt phản hồi DNS.
* **Tấn công dựa trên IoT –** **IoT-Based Attacks:** Cuộc tấn công IoT là bất kỳ cuộc tấn công mạng nào nhắm vào thiết bị hoặc mạng Internet of Things (IoT). Sau khi bị xâm nhập, tin tặc có thể chiếm quyền kiểm soát thiết bị, đánh cắp dữ liệu hoặc tham gia vào một nhóm thiết bị bị nhiễm để tạo mạng botnet nhằm khởi động các cuộc tấn công DoS hoặc DDoS. [3]

## Các giải pháp phòng chống tấn công mạng

Các giải pháp công nghệ mà có thể giúp bảo vệ doanh nghiệp trước các mối đe dọa an toàn thông tin có thể kể đến là:

* Chống mã độc và thư rác (Antivirus, Antispam)
* Tường lửa (Firewall)
* Phát hiện và ngăn chặn xâm nhập (IDPS)
* Hệ thống bẫy (Honeypot, Honeynet)
* Mạng riêng ảo (VPN)
* Chống rò rỉ dữ liệu (DLP)



### **Chống mã độc và thư rác (Antivirus, Antispam)**

Antivirus là một phần mềm được thiết kế để phát hiện, ngăn chặn và loại bỏ các phần mềm độc hại (malware) trong các thiết bị đầu cuối. Đây là thành phần cơ bản trong bảo vệ an toàn các thiết bị trong mạng máy tính.

Chức năng chính của phần mềm antivirus là quét các file, các chương trình và tổng thể hệ thống để tìm ra các mẫu, các đặc trưng của malware. Những đặc trưng này cơ bản là đặc điểm riêng biệt hoặc đoạn code mà có liên quan đến các kiểu malware. Khi phần mềm antivirus phát hiện một tệp tin hoặc chương trình có dấu hiệu phù hợp, phần mềm antivirus sẽ cảnh báo và thực hiện hành động để cách ly, xóa hoặc làm sạch tệp tin bị nhiễm.

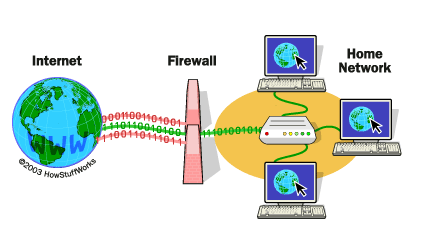
Dưới đây là một số tính năng và khả năng chính thường được tìm thấy trong phần mềm antivirus:

* Quét thời gian thực: Phần mềm antivirus liên tục giám sát các tệp tin và hoạt động trong thời gian thực để phát hiện và chặn malware ngay khi nó xuất hiện. Quét này có thể xảy ra trong quá trình tải xuống tệp tin, cài đặt, tệp đính kèm email và duyệt web.
* Phát hiện dựa trên dấu hiệu: Các chương trình antivirus duy trì một cơ sở dữ liệu chứa các dấu hiệu đã biết về malware. Trong quá trình quét, chúng so sánh các tệp tin và chương trình với cơ sở dữ liệu này để xác định và loại bỏ các mối đe dọa đã biết.
* Phân tích theo cách tiếp cận heuristic: Phần mềm antivirus sử dụng các kỹ thuật heuristic để xác định các hành vi có thể đáng ngờ hoặc độc hại. Nó phân tích mã và hành vi của các tệp tin và chương trình để phát hiện các mối đe dọa chưa biết hoặc zero-day mà không có các dấu hiệu tồn tại.
* Giám sát hành vi: Các chương trình antivirus giám sát hành vi của các tệp tin, quá trình và hoạt động hệ thống để xác định các hành động đáng ngờ có thể chỉ ra sự tồn tại của malware. Điều này bao gồm giám sát các sự thay đổi hệ thống không được ủy quyền, kết nối mạng không mong muốn và các hành vi bất thường khác.
* Cập nhật tự động: Phần mềm antivirus thường cập nhật cơ sở dữ liệu các dấu hiệu malware định kỳ để đối phó với các mối đe dọa mới và đang phát triển. Các cập nhật này cũng bao gồm các bản vá và sửa lỗi để nâng cao hiệu quả của phần mềm và khắc phục các lỗ hổng.
* Cách ly và khắc phục: Khi một tệp tin hoặc chương trình được phát hiện là bị nhiễm, phần mềm antivirus có thể cách ly nó, cô lập khỏi phần còn lại của hệ thống để ngăn chặn thiệt hại tiếp theo. Người dùng sau đó có thể chọn xóa hoặc cố gắng làm sạch tệp tin.
* Quét định kỳ: Phần mềm antivirus cho phép người dùng lên lịch quét hệ thống định kỳ để kiểm tra proactive các mối đe dọa malware. Các quét này có thể được thiết lập để xảy ra vào các thời điểm hoặc khoảng thời gian cụ thể, đảm bảo bảo vệ liên tục.
* Bảo vệ email và web: Nhiều chương trình antivirus bao gồm tính năng quét tệp đính kèm email, tải xuống web và liên kết trang web để phát hiện và chặn malware được lan truyền qua các kênh này.

Trên thị trường hiện đang có đa dạng các nhà cung cấp phần mềm antivirus cả phần mềm miễn phí và phần mềm thương mại. Một số phần mềm antivirus có thể kể đến: Microsoft Defender for Endpoint, ESET PROTECT Advanced, Kaspersky Antivirus, SentinelOne Singularity, Symantec End-user Endpoint Security, McAfee Antivirus.

### **Tường lửa (Firewall)**

Tường lửa (Firewall) là một thiết bị mạng có nhiệm vụ thực thi các chính sách an toàn thông tin dành cho lưu lượng mạng trong hệ thống mạng máy tính. Tường lửa tạo ra một rào chắn giữa các mạng máy tính khác nhau bằng cách đóng vai trò như một trạm kiểm soát mà lưu lượng mạng phải đi qua trước khi lưu lượng mạng đó có thể đi đến mạng máy tính. Tường lửa giúp hạn chế các lưu lượng mạng độc hại tới các thiết bị trong mạng máy tính, ví dụ như một attacker cố gắng thực hiện khai thác các lỗ hổng trong các phần mềm ứng dụng từ Internet.



Hình 1. 2 Minh họa việc chặn gói tin tộc hại của tưởng lửa

Tường lửa có thể là thiết bị phần cứng hoặc phần mềm. Tường lửa phần cứng là các thiết bị vật lý thường được cài đặt giữa mạng nội bộ và mạng ngoại vi. Chúng cung cấp một lớp bảo vệ bổ sung bằng cách lọc lưu lượng mạng dựa trên các yếu tố như địa chỉ IP, cổng và giao thức. Trong khi đó, tường lửa phần mềm là các chương trình được cài đặt trên các thiết bị hoặc server cá nhân. Chúng cung cấp bảo vệ ở mức thiết bị, giám sát và kiểm soát lưu lượng mạng cụ thể cho từng thiết bị.

Tường lửa hoạt động bằng cách kiểm tra các gói dữ liệu di chuyển qua mạng. Chúng so sánh thông tin trong các gói dữ liệu này với một tập hợp quy tắc hoặc chính sách đã được định trước. Nếu gói dữ liệu đáp ứng các tiêu chí xác định, nó được phép đi qua tường lửa và đến điểm đích. Nếu vi phạm bất kỳ quy tắc nào, tường lửa sẽ chặn hoặc loại bỏ gói dữ liệu, ngăn chúng tiếp cận mạng.

Tường lửa cung cấp một số chức năng quan trọng bao gồm:

* Kiểm soát truy cập: Tường lửa có thể hạn chế truy cập vào mạng bằng cách cho phép hoặc chặn địa chỉ IP, cổng hoặc giao thức cụ thể.
* Lọc gói dữ liệu: Chúng kiểm tra từng gói dữ liệu và đưa ra quyết định dựa trên thông tin như địa chỉ IP nguồn và đích, cổng và giao thức.
* Chuyển đổi địa chỉ mạng (NAT): Tường lửa có thể chuyển đổi địa chỉ IP riêng tư thành địa chỉ IP công cộng, cho phép nhiều thiết bị trong mạng riêng tư chia sẻ cùng một địa chỉ IP công cộng.
* Hỗ trợ VPN: Tường lửa thường bao gồm hỗ trợ Mạng Riêng Ảo (VPN), cho phép người dùng từ xa kết nối an toàn với mạng.
* Ngăn chặn xâm nhập: Một số tường lửa tiên tiến có khả năng ngăn chặn xâm nhập, phát hiện và chặn các hoạt động mạng đáng ngờ hoặc độc hại.

Một số tưởng lửa tiêu biểu Cisco ASA, Checkpoint Firewall, Paolo Alto Firewall, Iptable, UFW, Windows Firewall,..

### **Phát hiện và ngăn chặn xâm nhập (IDPS)**

IDPS là viết tắt của Intrusion Detection and Prevention System (tạm dịch: Hệ thống phát hiện và ngăn chặn xâm nhập). Đây là một công nghệ bảo mật mạng được thiết kế để phát hiện và ngăn chặn việc truy cập trái phép, hoạt động độc hại và các mối đe dọa mạng tiềm năng trong một mạng máy tính. IDPS kết hợp cả hệ thống phát hiện xâm nhập (IDS – Intrusion Detection System) và hệ thống ngăn chặn xâm nhập (IPS – Intrusion Prevention System) để cung cấp một cơ chế phòng thủ toàn diện.

Hệ thống Phát hiện xâm nhập (IDS) giám sát lưu lượng mạng, sự kiện hệ thống và hoạt động người ung để phát hiện bất kỳ dấu hiệu hành vi độc hại hoặc hoạt động đáng ngờ. Nó phân tích gói tin mạng, nhật ký và các nguồn dữ liệu khác nhau để phát hiện các mẫu hoặc chữ ký liên quan đến các cuộc tấn công đã biết hoặc các sự bất thường có thể chỉ ra một cuộc tấn công đang diễn ra. Khi một hệ thống IDS phát hiện một xâm nhập tiềm ẩn hoặc vi phạm bảo mật, IDS tạo ra cảnh báo hoặc thông báo để thông báo cho quản trị mạng hoặc nhân viên bảo mật.

Trong khi đó, hệ thống Ngăn chặn xâm nhập (IPS) hoạt động một cách chủ động để chặn hoặc ngăn chặn các mối đe dọa được xác định để không gian mạng không bị xâm nhập. Nó vượt xa giai đoạn phát hiện và thực hiện các hành động để ngăn chặn xâm nhập một cách tích cực. IPS có thể tự động loại bỏ, cách ly hoặc chặn lưu lượng mạng hoặc kết nối liên quan đến các mối đe dọa đã biết hoặc các hoạt động độc hại. IPS cũng có thể thực hiện các chính sách bảo mật, phát hiện và ngăn chặn các cuộc tấn công trái phép và bảo vệ khỏi nhiều loại cuộc tấn công, chẳng hạn như tấn công từ chối dịch vụ (DoS), quét cổng hoặc tấn công SQLi,..

Các chức năng chính của một IDPS bao gồm:

* Giám sát và Phân tích: IDPS liên tục giám sát lưu lượng mạng, nhật ký hệ thống và dữ liệu sự kiện để phát hiện các mối đe dọa tiềm ẩn và các sự cố bảo mật.
* Phát hiện Mối đe dọa: Phát hiện và phân tích các mẫu, chữ ký và các sự bất thường trong lưu lượng mạng và hành vi hệ thống để xác định các cuộc tấn công tiềm ẩn hoặc vi phạm bảo mật.
* Cảnh báo và Thông báo: IDPS tạo ra cảnh báo, thông báo hoặc báo động khi phát hiện các hoạt động đáng ngờ, cho phép quản trị mạng thực hiện các biện pháp phù hợp.
* Ngăn chặn và phản ứng: Một IDPS có thể ngăn chặn các mối đe dọa đã biết hoặc lưu lượng mạng đáng ngờ khỏi xâm nhập vào hoặc rời khỏi mạng bằng cách chặn các lưu lượng mạng đó, cung cấp phòng thủ tích cực chống lại các cuộc tấn công.
* Phản ứng sự cố: IDPS đóng vai trò quan trọng trong phản ứng sự cố bằng cách cung cấp thông tin quan trọng về các sự cố bảo mật, giúp đội ngũ bảo mật điều tra và giảm thiểu tác động của một cuộc xâm nhập.
* Phân tích Nhật ký và Báo cáo: IDPS ghi nhật ký và ghi lại các sự kiện bảo mật, cung cấp nguồn thông tin quý giá cho việc phân tích sau sự cố, kiểm tra tuân thủ quy định và mục đích báo cáo.

### **Hệ thống bẫy (Honeypot)**

Honeypot là một kỹ thuật hoặc công cụ bảo mật mạng được thiết kế để thu hút và đánh lừa các kẻ tấn công tiềm năng. Đó là một hệ thống hoặc tài nguyên mạng giả mạo có vẻ như là một mục tiêu hợp lệ, nhưng thực tế, hệ thống honeypot đã được cô lập và theo dõi chuyên sâu bởi các chuyên gia bảo mật. Mục đích của một honeypot là thu thập thông tin về các kẻ tấn công, phương pháp và động cơ của kẻ tấn công, nhằm hiểu rõ hơn và giảm thiểu các mối đe dọa tiềm năng.

Honeypot thường được triển khai như một môi trường được kiểm soát, mô phỏng các hệ thống hoặc dịch vụ thực tế. Hệ thống honeypot được thiết kế có ý đồ để bị tấn công, thu hút kẻ tấn công tương tác với chúng. Honeypot có thể mô phỏng nhiều loại tài nguyên, chẳng hạn như server, ứng dụng hoặc thiết bị mạng, phụ thuộc vào mục tiêu nghiên cứu bảo mật hoặc chiến lược phòng vệ mong muốn.

Khi kẻ tấn công tương tác với một Honeypot, hành vi và hoạt động của họ được ghi lại và phân tích. Điều này bao gồm việc ghi nhận lưu lượng mạng, lưu log lệnh nhập và giám sát bất kỳ cố gắng khai thác lỗ hổng nào. Bằng cách quan sát cẩn thận các hoạt động của kẻ tấn công, các chuyên gia bảo mật có thể thu thập thông tin quý giá về các kỹ thuật, công cụ và ý đồ của họ.

Hệ thống Honeypot có thể phân loại như sau:

* Production Honeypot: Đây là các hệ thống hoặc mạng thực sự được tạo ra mục đích để thu hút kẻ tấn công. Thông thường, chúng được triển khai song song với các hệ thống sản xuất thực tế để phát hiện và đáp ứng các mối đe dọa một cách tích cực.
* Research Honeypot: Các honeypot này được thiết kế cho mục đích nghiên cứu cụ thể. Chúng có thể tập trung vào nghiên cứu các mô hình tấn công cụ thể, phân tích phần mềm độc hại hoặc hiểu rõ ý đồ của kẻ tấn công.
* High-Interaction Honeypot: Những honeypot này cung cấp chức năng và khả năng tương tác rộng, thường mô phỏng các hệ điều hành hoặc dịch vụ đầy đủ. Mục tiêu của chúng là giữ kẻ tấn công tương tác trong thời gian dài, thu thập thông tin chi tiết hơn về hoạt động của họ.
* Low-Interaction Honeypot: Những honeypot này mô phỏng chức năng và dịch vụ hạn chế, tạo ra bề mặt tấn công nhỏ hơn. Chúng dễ dàng triển khai và bảo trì, nhưng có thể thu hút ít kẻ tấn công kỹ thuật hơn.

Honeypot mang lại nhiều lợi ích trong bảo mật mạng:

* Thông tin về mối đe dọa: Honeypot cung cấp thông tin quý giá về các chiến thuật, kỹ thuật và công cụ mà kẻ tấn công sử dụng. Thông tin này có thể nâng cao thông tin về mối đe dọa và giúp tổ chức cải thiện tổng thể bảo mật.
* Cảnh báo sớm: Honeypot có thể phát hiện và cảnh báo tổ chức về các cuộc tấn công tiềm ẩn ở giai đoạn đầu, cho phép triển khai biện pháp phòng vệ chủ động.
* Đánh lạc hướng: Bằng cách lạc hướng sự chú ý của kẻ tấn công đến honeypot, tổ chức có thể bảo vệ các hệ thống và mạng thực tế khỏi bị tấn công.
* Xác định lỗ hổng: Honeypot có thể giúp xác định các lỗ hổng mới hoặc chưa biết bằng cách thu hút và ghi lại các cuộc tấn công.

### **Mạng riêng ảo VPN**

VPN, viết tắt của Virtual Private Network (tạm dịch: Mạng riêng ảo), là một công nghệ cho phép người hem tạo ra một kết nối an toàn và được mã hóa trên mạng công cộng như internet. VPN cho phép người hem thiết lập một kết nối mạng riêng từ xa, như khi người hem trực tiếp kết nối đến một mạng riêng, ngay cả khi người hem truy cập internet thông qua một mạng công cộng hoặc mạng Wi-Fi không an toàn.

Mục đích chính của VPN là nâng cao bảo mật và sự riêng tư bằng cách mã hóa dữ liệu được truyền giữa thiết bị của người hem và server hoặc trang web đích. Khi người hem kết nối vào VPN, lưu lượng internet của họ được định tuyến qua một đường hầm an toàn, mã hóa dữ liệu và bảo vệ nó khỏi việc truy cập trái phép hoặc chặn bởi các hacker, giám sát chính phủ hoặc các thực thể độc hại khác.

Dưới đây là một số tính năng và lợi ích chính của việc sử dụng VPN:

* Truyền dữ liệu an toàn: VPN mã hóa dữ liệu truyền giữa thiết bị của người hem và server VPN, đảm bảo rằng nó không thể dễ dàng bị người khác chặn hoặc truy cập trái phép. Điều này đặc biệt quan trọng khi sử dụng mạng Wi-Fi công cộng, mà dễ bị nghe trộm và đánh cắp dữ liệu.
* Bảo vệ sự riêng tư: Bằng cách ẩn địa chỉ IP của người hem và mã hóa lưu lượng internet của họ, VPN cung cấp một mức độ bảo mật cao hơn. Nó ngăn người cung cấp dịch vụ internet (ISP), các trang web hoặc các thực thể khác từ việc theo dõi và giám sát hoạt động trực tuyến của người hem, nâng cao tính nặc danh.
* Vượt qua hạn chế địa lý: VPN cho phép người hem vượt qua các hạn chế địa lý được áp đặt bởi các trang web hoặc nền tảng phát trực tuyến. Bằng cách kết nối đến server VPN ở một quốc gia khác, người hem có thể truy cập nội dung và dịch vụ mà có thể bị hạn chế hoặc không khả dụng ở vị trí của mình.
* Tăng cường bảo mật cho việc truy cập từ xa: VPN cung cấp một phương pháp an toàn cho người lao động từ xa truy cập vào tài nguyên mạng nội bộ của tổ chức. Bằng cách kết nối vào VPN của tổ chức, nhân viên có thể truyền dữ liệu nhạy cảm và truy cập các nguồn tài nguyên của công ty một cách an toàn, ngay cả khi làm việc từ xa.
* Chia sẻ tập tin P2P: VPN có thể được sử dụng cho việc chia sẻ tập tin peer-to-peer (P2P) an toàn và riêng tư. Bằng cách mã hóa lưu lượng P2P, VPN ngăn ISP hoặc các bên khác theo dõi hoặc giới hạn kết nối dựa trên loại nội dung được chia sẻ.
* Ẩn danh và bảo vệ danh tính: VPN có thể giúp bảo vệ danh tính và hoạt động trực tuyến của người hem khỏi việc bị truy tìm lại. Bằng cách ẩn địa chỉ IP và mã hóa dữ liệu của người hem, VPN hem một lớp bảo vệ và nâng cao mức độ nặc danh.

### **Chống rò rỉ dữ liệu (DLP)**

DLP là viết tắt của Data Loss Prevention (DLP). DLP đề cập đến một tập hợp các công nghệ và phương pháp được thiết kế để ngăn chặn việc tiết lộ hoặc mất mát dữ liệu nhạy cảm trong một tổ chức. Các giải pháp DLP được sử dụng để xác định, giám sát và bảo vệ thông tin nhạy cảm để đảm bảo rằng các thông tin trên không rời khỏi giới hạn của tổ chức.

Mục tiêu chính của DLP là ngăn chặn việc vi phạm dữ liệu, bất kể có ý định hay vô tình, bằng cách giám sát và kiểm soát luồng dữ liệu nhạy cảm qua các kênh và điểm cuối khác nhau. Điều này bao gồm dữ liệu ở trạng thái yên ngủ (dữ liệu đã được lưu trữ), dữ liệu trong quá trình truyền (dữ liệu đang được truyền) và dữ liệu đang sử dụng (dữ liệu đang được truy cập hoặc xử lý).

Hệ thống DLP thường áp dụng sự kết hợp giữa chính sách, quy tắc và kiểm tra nội dung để xác định và phân loại dữ liệu nhạy cảm. Các hệ thống này có thể phát hiện thông tin nhạy cảm như thông tin cá nhân có thể xác định (PII), dữ liệu tài chính, sở hữu trí tuệ, bí mật thương mại hoặc thông tin bí mật khác dựa trên tiêu chí được xác định trước. Một số phương pháp thông thường được sử dụng trong DLP bao gồm khớp từ khóa, biểu thức chính quy, dấu vân tay dữ liệu và các thuật toán học máy.

Sau khi xác định dữ liệu nhạy cảm, các giải pháp DLP có thể thực hiện các biện pháp bảo mật khác nhau để ngăn chặn mất dữ liệu. Các biện pháp này có thể bao gồm mã hóa, kiểm soát truy cập, chặn hoặc cách ly dữ liệu, thông báo cho quản trị viên hoặc người dùng, hoặc áp dụng kỹ thuật xóa thông tin hoặc che giấu thông tin nhạy cảm.

Các giải pháp DLP có thể được triển khai ở các mức độ khác nhau trong hạ tầng của một tổ chức, bao gồm mức độ mạng, mức độ điểm cuối và mức độ lưu trữ dữ liệu. DLP mức độ mạng giám sát lưu lượng mạng để phát hiện và ngăn chặn rò rỉ dữ liệu qua giao thức mạng, email, ứng dụng web hoặc truyền tệp. DLP mức độ điểm cuối tập trung vào bảo vệ các thiết bị cá nhân như máy tính xách tay, máy tính để bàn hoặc thiết bị di động, đảm bảo rằng dữ liệu được bảo vệ ngay cả khi không kết nối với mạng của tổ chức. DLP mức độ lưu trữ dữ liệu liên quan đến bảo vệ các kho dữ liệu và cơ sở dữ liệu để ngăn chặn việc truy cập trái phép hoặc rò rỉ dữ liệu.



## Kết luận chương 1

Chương 1 cung cấp cái nhìn tổng quan về an toàn thông tin, giúp người đọc nắm được các khái niệm cơ bản như tính bảo mật, toàn vẹn, sẵn sàng và xác thực dữ liệu; nhận thức được thực trạng và mức độ rủi ro ATTT tại Việt Nam; đồng thời hiểu được các mối đe dọa phổ biến và các giải pháp bảo vệ cơ bản. Qua đó, chương này xây dựng nền tảng kiến thức cần thiết để triển khai giám sát và quản lý dữ liệu an toàn thông tin hiệu quả trong các chương tiếp theo

# CHƯƠNG II: THU THẬP DỮ LIỆU TRONG GIÁM SÁT AN TOÀN THÔNG TIN



## Giới thiệu chung về thu thập dữ liệu trong giám sát

Thu thập dữ liệu giám sát là bước nền tảng trong quản lý an toàn thông tin, nhằm phát hiện, phân tích và ứng phó kịp thời với các mối đe dọa. Dữ liệu giám sát được sinh ra từ nhiều nguồn khác nhau, bao gồm hệ thống, ứng dụng, mạng, thiết bị/IoT, và dữ liệu bảo mật bổ sung từ các nguồn bên ngoài.

Việc xác định, phân loại và lựa chọn phương pháp thu thập dữ liệu hợp lý giúp tổ chức triển khai giải pháp giám sát hiệu quả, đảm bảo bảo mật thông tin và tối ưu chi phí lưu trữ. Theo NIST SP 800-92 – Guide to Computer Security Log Management và NIST SP 800-137 – Information Security Continuous Monitoring (ISCM), dữ liệu giám sát có thể được phân loại thành các nhóm chính sau:

* Nhật ký hệ thống (System/Host-based Logs)
* Nhật ký ứng dụng (Application Logs)
* Nhật ký mạng (Network-based Logs / Network Flow & Packet)
* Nhật ký thiết bị/IoT (Device/IoT Logs)
* Dữ liệu bảo mật bổ sung (Supplemental Security Data)

### **Nhật ký hệ thống ( System logs collection)**

Nhật ký hệ thống (System/Host-based Logs) là nền tảng cơ bản của việc giám sát. Chúng bao gồm các log sinh ra từ hệ điều hành và các thành phần cơ bản của endpoint, ví dụ Windows Event Logs hay Syslog trên Linux/Unix. Dữ liệu này ghi nhận các sự kiện khởi động/tắt máy, quản lý tiến trình, driver, xác thực cơ bản (login/logout) và thay đổi cấu hình. Thông qua phân tích nhật ký hệ thống, tổ chức có thể phát hiện bất thường và tái hiện dòng sự kiện khi điều tra sự cố.

### **Nhật ký ứng dụng (Application Logs)**

Nhật ký ứng dụng (Application Logs) cung cấp thông tin chi tiết về hành vi phần mềm và người dùng. Log từ các ứng dụng doanh nghiệp, web server, cơ sở dữ liệu hay các công cụ bảo mật như antivirus, EDR giúp phát hiện lỗi, hành vi bất thường hoặc malware. Khi kết hợp với nhật ký hệ thống, nhật ký ứng dụng giúp cải thiện khả năng phát hiện và phân tích các mối đe dọa.

### **Dữ liệu mạng (Network-based Logs / Network Flow & Packet)**

Nếu như log hệ thống và log ứng dụng phản ánh hành vi cục bộ, thì dữ liệu mạng mang đến góc nhìn toàn cục, dữ liệu mạng (Network-based Logs / Network Flow & Packet) ghi nhận các luồng thông tin giữa các thiết bị trong tổ chức. Thông qua Network Flow, có thể thống kê kết nối, dung lượng dữ liệu và thời gian trao đổi, trong khi Packet Capture (PCAP) cung cấp chi tiết từng gói tin, phục vụ điều tra tấn công tinh vi. Việc kết hợp dữ liệu mạng với nhật ký hệ thống và ứng dụng giúp phát hiện các cuộc tấn công quy mô lớn, rò rỉ dữ liệu hay giao tiếp command-and-control.

### **Nhật ký thiết bị / Iot ( Device/IoT Logs )**

Ngoài ra, với sự phát triển của các thiết bị thông minh và IoT, nhật ký từ thiết bị/IoT (Device/IoT Logs) ngày càng quan trọng. Khác với flow/packet, log này không phải dữ liệu mạng, mà là nhật ký sự kiện thiết bị tự ghi lại. Log từ camera giám sát, cảm biến môi trường hay thiết bị điều khiển công nghiệp cung cấp thông tin về trạng thái thiết bị, truy cập vật lý và cảnh báo bất thường, bổ sung cho dữ liệu mạng và hệ thống để giám sát toàn diện.

### **Dữ liệu bảo mật khác**

Bên cạnh dữ liệu nội bộ, các tổ chức ngày càng dựa nhiều hơn vào dữ liệu bảo mật bổ sung từ các nguồn bên ngoài hoặc hệ thống chuyên biệt. Threat intelligence cung cấp các chỉ báo xâm phạm (Indicators of Compromise – IOC) như địa chỉ IP, tên miền hoặc mã băm file độc hại. Danh sách đen (blacklist) hoặc chữ ký (signatures) do cộng đồng hoặc nhà cung cấp phát hành cũng là nguồn dữ liệu hữu ích để phát hiện tấn công. Ngoài ra, các hệ thống bảo mật như IDS/IPS, tường lửa ứng dụng web (WAF), hay dịch vụ bảo mật đám mây đều tạo ra cảnh báo hỗ trợ giám sát.

Theo NIST SP 800-150 (2016), việc tích hợp threat intelligence vào quy trình giám sát không chỉ nâng cao hiệu quả phát hiện mối đe dọa mà còn cải thiện khả năng phản ứng kịp thời của tổ chức.

## Phương pháp thu thập dữ liệu giám sát

Theo hướng dẫn trong NIST SP 800-137 – Information Security Continuous Monitoring (ISCM), dữ liệu giám sát trong tổ chức có thể được thu thập bằng ba phương pháp chính: **thu thập trực tiếp, thu thập gián tiếp và thu thập thủ công/định kỳ**. Mỗi phương pháp có cơ chế triển khai, ưu điểm và hạn chế riêng, thường được kết hợp linh hoạt trong thực tế để đảm bảo tính toàn diện của hoạt động giám sát.

### **Thu thập trực tiếp (Direct Collection)**

Thu thập trực tiếp là phương pháp mà dữ liệu được lấy ngay tại hệ thống hoặc thiết bị nguồn. Người quản trị có thể truy cập trực tiếp vào file log, sử dụng lệnh dòng (command line), hoặc xây dựng các script để tự động hóa quá trình sao chép dữ liệu.

1. **Thu thập nhật ký trên Windows (Event Logs)**

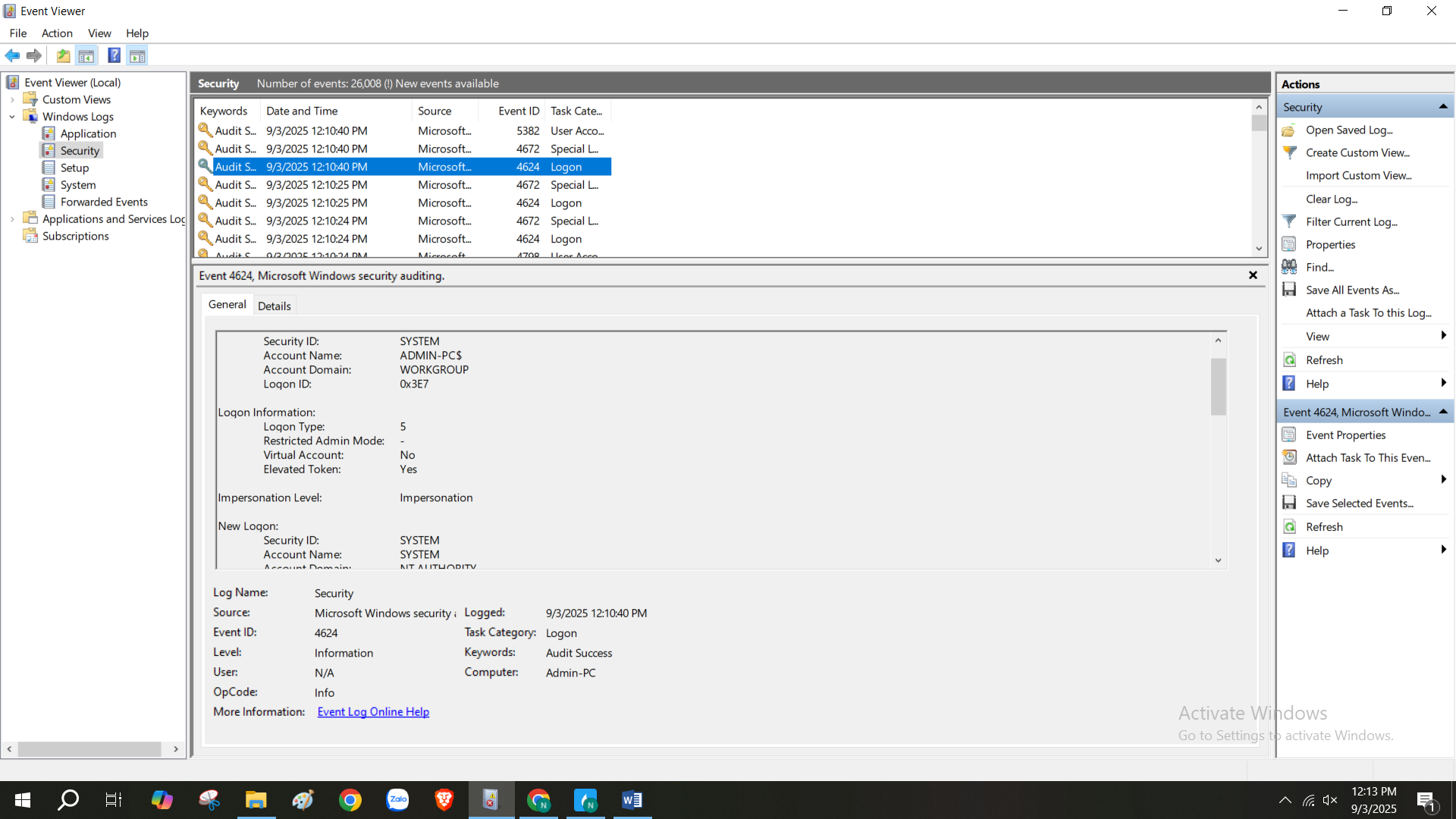
* **Giới thiệu:** Windows lưu trữ các sự kiện quan trọng (đăng nhập, cài đặt, lỗi hệ thống) trong Event Logs. Việc truy cập trực tiếp Event Logs giúp quản trị viên phát hiện hành vi bất thường hoặc sự cố bảo mật.

Các kênh chính:

* **System Log:** liên quan đến phần cứng và hệ thống (driver, khởi động, tắt máy).
* **Application Log:** ghi lại sự kiện do ứng dụng tạo ra.
* **Security Log:** chứa đăng nhập, đăng xuất, quyền truy cập và các vi phạm bảo mật.
* **Mục đích:**
  + - Thu thập log bảo mật từ hệ điều hành Windows.
    - Quan sát sự kiện đăng nhập và các cảnh báo bảo mật.
* **Cách thực hiện**

**Cách 1: Dùng Event Viewer**

* Mở **Event Viewer** (Win + R → eventvwr).
* Chọn **Windows Logs → Security**.
* Quan sát các sự kiện đăng nhập (Event ID 4624: đăng nhập thành công, 4625: đăng nhập thất bại).



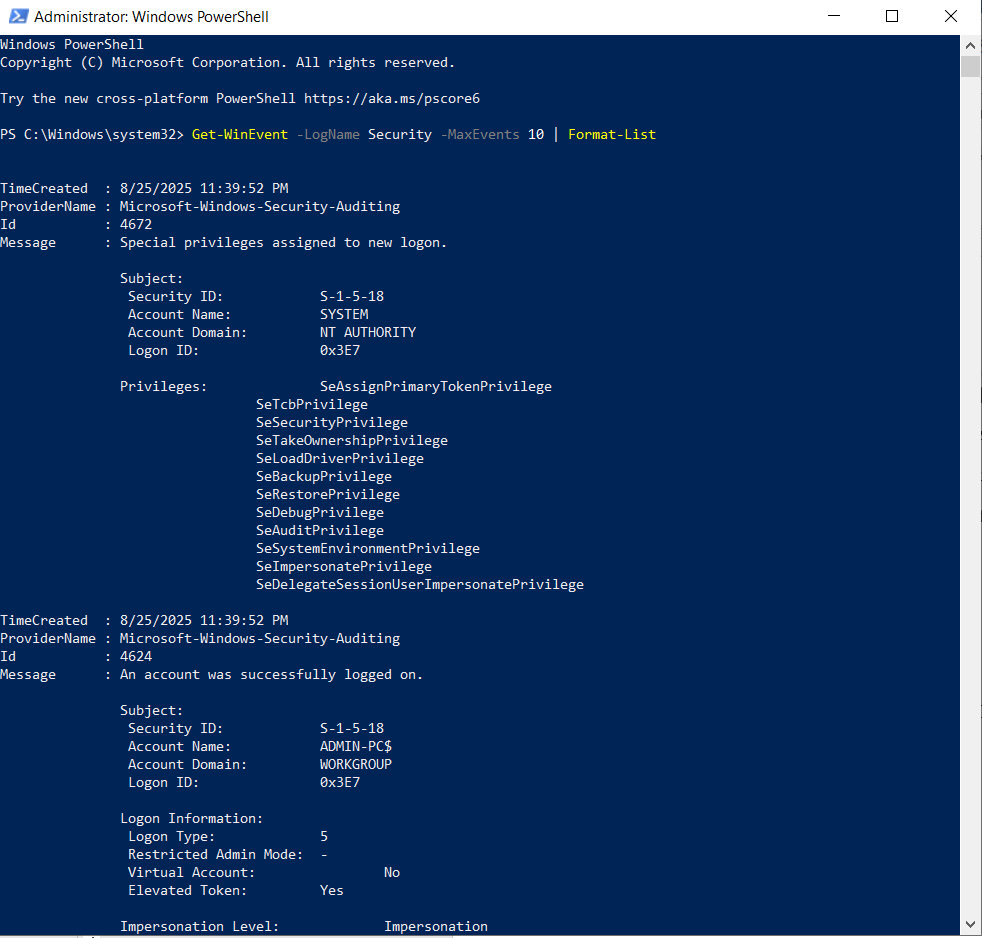
Hình 2. 1: Hiển thị thông tin log windows trong Event Viewer

Minh họa việc **hệ thống ghi nhận một phiên đăng nhập thành công (Event ID 4624)**, cụ thể là dạng đăng nhập của dịch vụ/tài khoản hệ thống chứ không phải người dùng trực tiếp nhập mật khẩu.

**Cách 2: Dùng PowerShell**

Chạy lệnh: *Get-WinEvent -LogName Security -MaxEvents 10 | Format-List*

* *LogName Security:* lấy log từ kênh Security.
* *MaxEvents 10*: hiển thị 10 sự kiện mới nhất.
* *Format-List:* hiển thị chi tiết từng sự kiện*.*



Hình 2. 2: Hiển thị bản ghi chi tiết với Event ID, User, TimeCreated, Message

Tài khoản SYSTEM đã đăng nhập vào hệ thống Windows. Khi đăng nhập, nó được cấp các đặc quyền quản trị cao nhất để có thể thực hiện các thao tác quan trọng trên hệ điều hành (backup, restore, debug, shutdown, impersonation…).

* **Nhân xét**
* Cả hai đều là thu thập trực tiếp, lấy dữ liệu ngay từ Windows Event Logs.
* Có thể dùng để điều tra sự cố đăng nhập, giám sát an toàn hệ thống

1. **Thu thập nhật ký Thu thập trực tiếp dữ liệu mạng (flow và packet)**

* **Giới thiệu:**

Bên cạnh nhật ký hệ thống và ứng dụng, việc thu thập dữ liệu mạng đóng vai trò quan trọng trong giám sát an toàn thông tin. Dữ liệu mạng có thể được thu thập dưới hai dạng phổ biến: **Network Flow** và **Packet (PCAP).**

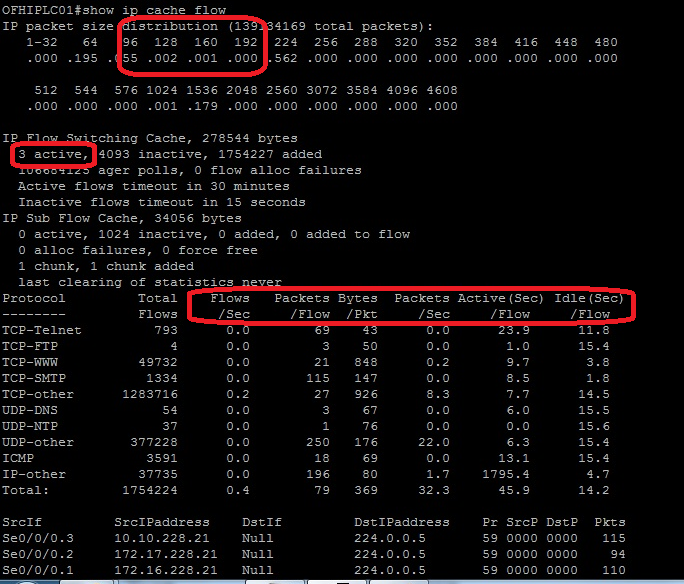
**Network Flow (luồng mạng):** Đây là dạng bản ghi tóm tắt thông tin về các kết nối mạng, ví dụ như địa chỉ IP nguồn, đích, cổng, giao thức, và dung lượng truyền tải. Một số chuẩn flow phổ biến là NetFlow, sFlow và IPFIX. Nếu thiết bị mạng (router hoặc switch) có hỗ trợ, quản trị viên có thể đăng nhập trực tiếp vào CLI để xem.

**Packet (gói tin – PCAP):** Khác với flow chỉ tóm tắt, packet capture (bắt gói tin) cho phép thu thập toàn bộ nội dung gói tin. Công cụ phổ biến như Wireshark có thể được sử dụng để bắt trực tiếp gói tin từ card mạng.

* **Mục đích:**
* Quan sát trực tiếp lưu lượng mạng.
* Phát hiện hành vi bất thường như quét cổng, DoS, truy vấn DNS bất thường.
* **Cách thực hiện**

**Thu thập Flow (Cisco router):**

Trên thiết bị Cisco, lệnh: show ip cache flow => lệnh này cho phép hiển thị các thông tin về flow đang được ghi nhận.



Hình 2. 3: Thống kê IP Flow Cache

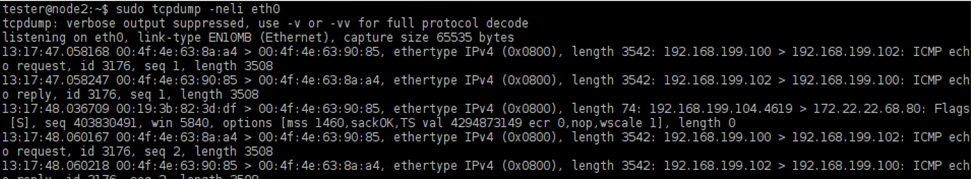
Hình trên cho ta biết, hệ thống đang xử lý 34169 gói tin với phân bố chủ yếu ở kích thước 64–160 bytes. Có 3 luồng lưu lượng đang hoạt động, còn lại là luồng cũ đã kết thúc. Các giao thức chính được ghi nhận: TCP (Telnet), UDP (DNS, NTP), ICMP. Lưu lượng Telnet chiếm nhiều nhất (793 flows) → có thể do quản trị viên hoặc người dùng kết nối thiết bị.

Thống kê này giúp quản trị viên nắm được:

* Tình trạng sử dụng băng thông.
* Giao thức nào đang được dùng nhiều.
* Phát hiện bất thường trong lưu lượng (nếu có).

**Thu thập Packet (Linux):**

Trên Linux, có thể sử dụng lệnh: sudo tcpdump -neli eth0



Hình 2. 4: Thu thập trực tiếp packet bằng tcpdump

Kết quả tcpdump cho biết:

* Có giao tiếp ICMP (ping) giữa 2 máy: 192.168.199.100 và 192.168.199.102.
* ICMP request/reply được thực hiện thành công (ping OK).
* Đồng thời có một kết nối TCP SYN từ 192.168.199.104 đến 172.22.22.68 cổng 80 (web).
* **Nhân xét:**
* Phương pháp này giúp quản trị viên thấy ngay hoạt động mạng của thiết bị.
* Packet cung cấp thông tin chi tiết để điều tra sự cố.
* Flow phù hợp để giám sát tổng thể trên router/switch.

### **Thu thập gián tiếp (Indirect Collection)**

Khác với thu thập trực tiếp, thu thập gián tiếp dựa vào các công cụ hoặc hệ thống trung gian. Thay vì quản trị viên phải vào từng máy, dữ liệu sẽ được agent hoặc forwarder trên hệ thống gửi về một máy chủ tập trung (ví dụ: SIEM). Trong phần thu thập này sẽ sử dụng Wazuh - giải pháp SIEM & XDR.

**Thu Thập gián tiếp bằng wazuh**

* **Giới thiệu:**

Wazuh là một nền tảng SIEM và EDR mã nguồn mở, có khả năng giám sát, phát hiện và phản ứng sự cố bảo mật. Việc thu thập log thông qua Wazuh là thu thập gián tiếp, vì dữ liệu không được lấy thủ công ngay tại hệ điều hành hay ứng dụng, mà được agent cài đặt trên endpoint thu thập rồi gửi về Wazuh Manager để phân tích và lưu trữ.

**Giải pháp SIEM** – Quản lý và phân tích sự kiện an ninh: SIEM cho phép các đơn vị, cơ quan có được cái nhìn toàn cảnh về các sự kiện an ninh. SIEM có thể phân tích một lượng lớn dữ liệu để phát hiện các cuộc tấn công ẩn dấu đằng sau chúng. Tuy nhiện SIEM có những hạn chế về khả năng mở rộng, độ phức tạp, tích hợp dữ liệu, khả năng phát hiện, nhận thức theo ngữ cảnh, phản hồi theo thời gian thực và cách tiếp cận lấy tuân thủ làm trung tâm. Các tổ chức cần nhận thức được những hạn chế này và xem xét các giải pháp bảo mật hiện đại tận dụng các công nghệ tiên tiến, như học máy, AI và tự động hóa, để khắc phục những hạn chế này và nâng cao tình trạng bảo mật của họ.

**Giải pháp EDR** – Phát hiện và phản hồi điểm cuối: cung cấp cho tổ chức khả năng giám sát các điểm cuối để tìm hành vi đáng ngờ và ghi lại mọi hoạt động và sự kiện đơn lẻ. Sau đó, nó tương quan thông tin để cung cấp ngữ cảnh quan trọng nhằm phát hiện các mối đe dọa nâng cao và cuối cùng chạy hoạt động ứng phó tự động, chẳng hạn như cách ly một điểm cuối bị nhiễm khỏi mạng trong thời gian gần thực tế. các phương thức xếp hạng cảnh báo và trực quan hóa dữ liệu giúp người quản trị nhanh chóng xác định được mối đe dọa và lên phương án phản ứng.

* **Mục đích:**
* Minh họa phương pháp thu thập gián tiếp trong giám sát an toàn thông tin.
* Tập trung log từ nhiều endpoint về một hệ thống duy nhất (Wazuh server).
* Kích hoạt cảnh báo tự động khi có sự kiện bất thường.
* **Cách thực hiện:**

**Bước 1:** Cài đặt Wazuh Server

**Bước 2:** Cài đặt Wazuh Agent trên endpoint (ví dụ: Windows/Linux)

(Bước 1 và 2 có hướng dẫn **Phụ lục 1**)

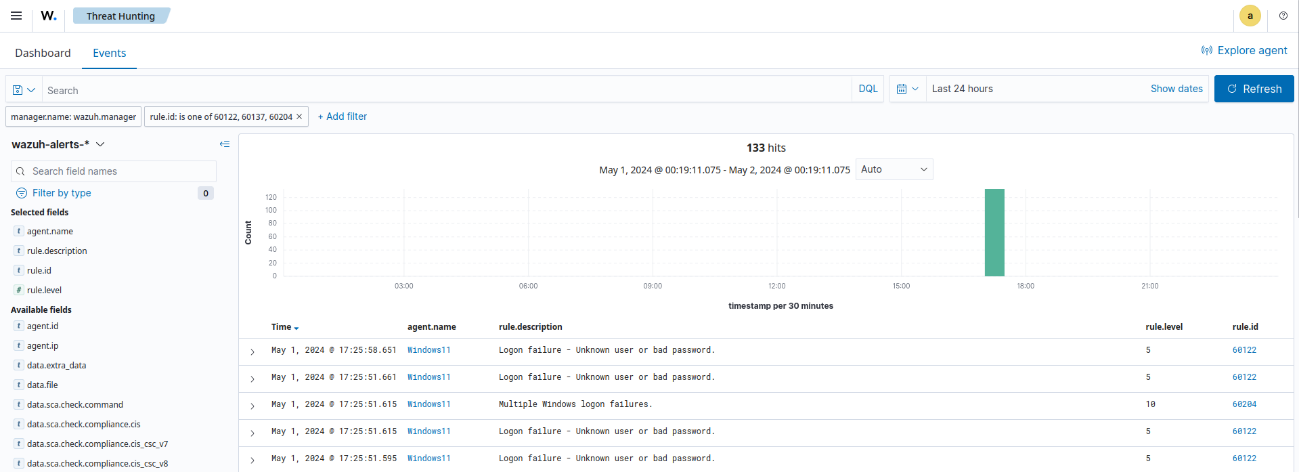
**Bước 3:** Thu thập log

* Agent giám sát file hệ thống (/var/log/auth.log, Windows Security Event Logs, log ứng dụng...).
* Agent gửi log về Wazuh Manager qua kết nối bảo mật (TCP/UDP 1514).

**Bước 4:** Quan sát trên Dashboard

Truy cập giao diện Kibana tích hợp với Wazuh, quản trị viên có thể xem:

* Các sự kiện đăng nhập thất bại (brute force).
* Hoạt động truy cập file hệ thống.
* Cảnh báo IDS/IPS (nếu tích hợp Suricata/Snort).



Hình 2. 5: Wazuh phát hiện trên máy Windows11 đã xảy ra nhiều sự kiện đăng nhập thất bại

Đây là cảnh báo về đăng nhập thất bại (failed login) trên Windows 11, cụ thể:

* Người dùng quên mật khẩu và nhập sai nhiều lần.
* Hoặc có dấu hiệu tấn công brute force (kẻ tấn công thử nhiều mật khẩu để xâm nhập).
* **Nhận xét:**
* Đây là thu thập gián tiếp, vì log không lấy trực tiếp trên thiết bị, mà thông qua agent gửi về server.
* Ưu điểm: tập trung log, có phân tích tự động, phát hiện sớm mối đe dọa.
* Nhược điểm: cần hạ tầng (server, agent, dashboard), triển khai phức tạp hơn so với cách thủ công.

Phương pháp gián tiếp đặc biệt phù hợp với hạ tầng lớn, vì giúp quản lý tập trung và giảm tải cho quản trị viên. Tuy nhiên, nhược điểm là phụ thuộc vào độ tin cậy của công cụ trung gian, đồng thời có thể phát sinh độ trễ khi truyền dữ liệu.

### **Thu thập thủ công/định kỳ (Manual or Report-based Collection)**

Thu thập thủ công hoặc định kỳ là phương pháp mà quản trị viên **chủ động thực hiện thao tác trích xuất dữ liệu/log theo yêu cầu** hoặc theo một lịch trình nhất định.

* Thường áp dụng khi cần báo cáo sự cố bảo mật, phục vụ kiểm tra tuân thủ (compliance) hoặc điều tra số liệu trong một giai đoạn cụ thể.
* Không liên tục như thu thập trực tiếp/gián tiếp, mà phụ thuộc vào thời điểm thực hiện hoặc theo lịch định kỳ (ví dụ: hàng ngày, hàng tuần).

**Thu thập log hệ thống Linux bằng Bash script**

* **Giới thiệu:**

Trên Linux, các log quan trọng của hệ thống thường được lưu tại thư mục /var/log/, ví dụ như:

* + syslog: nhật ký hệ thống.
  + auth.log: nhật ký đăng nhập và xác thực.

Nếu mỗi ngày quản trị viên đều phải sao chép thủ công log ra thư mục lưu trữ thì sẽ tốn thời gian và dễ bỏ sót. Vì vậy, một cách tiếp cận hiệu quả hơn là viết script Bash để tự động thực hiện việc thu thập..

* **Mục đích:**
* Minh họa phương pháp thu thập theo báo cáo/định kỳ.
* Thực hiện lưu trữ nhật ký hệ thống để phục vụ kiểm toán hoặc điều tra sự cố.
* Không yêu cầu hạ tầng SIEM phức tạp, nhưng vẫn đảm bảo có bản sao dữ liệu theo từng mốc thời gian.
* **Cách thực hiện:**

Bước 1: Tạo script collect\_logs.sh ( sử dụng lệnh : nano collect\_logs.sh dùng để tạo hoặc mở file collect\_logs.sh nếu đã tồn tại )

|  |
| --- |
| #!/bin/bash  # Tạo folder theo ngày  TIMESTAMP=$(date +"%Y%m%d")  mkdir -p /home/logs/$TIMESTAMP  # Sao chép log hệ thống  cp /var/log/syslog /home/logs/$TIMESTAMP/syslog.log  cp /var/log/auth.log /home/logs/$TIMESTAMP/auth.log |

Bước 2: chạy thử script bằng lệnh *bash collect\_logs.sh*

Trong thư mục /home/logs/20250903/ sẽ có file syslog.log và auth.log

Bước 3: Tự động hóa (cronjob)

|  |
| --- |
| crontab -e   1. 0 \* \* \* /home/admin/scripts/collect\_logs.sh |

Lệnh này sẽ chạy script vào mỗi 0:0 giờ hàng ngày và tự động lưu log

* **Nhận xét:**
* Đây là thu thập thủ công/định kỳ, vì log chỉ được lấy vào thời điểm xác định (không theo thời gian thực).
* Ưu điểm: dễ triển khai, không cần hạ tầng lớn, phù hợp cho kiểm toán, điều tra.
* Nhược điểm: dữ liệu không tức thời → có thể bỏ lỡ sự kiện quan trọng xảy ra giữa các lần thu thập

### **Đánh giá các phương pháp thu thập**

|  |  |  |  |
| --- | --- | --- | --- |
| **Phương pháp** | **Thu thập trực tiếp** | **Thu thập thủ công/định kỳ** | **Thu thập tập trung** |
| **Nguyên tắc** | Truy cập trực tiếp vào endpoint để xuất log hoặc copy file log. | Sử dụng script hoặc command-line tool để thu thập log theo lịch hoặc trigger sự kiện | Sử dụng agent hoặc công cụ quản lý log để gửi đồng thời log từ nhiều endpoint về server trung tâm |
| **Ưu điểm** | Nhanh chóng, trực quan, không cần hạ tầng phức tạp.  Thích hợp khi xử lý sự cố tại chỗ. | Đảm bảo có dữ liệu log lưu trữ định kỳ, thuận lợi cho báo cáo/khiểm toán.  Có thể triển khai dễ dàng bằng script, không tốn chi phí lớn. | Tự động, liên tục, có cảnh báo real-time.  Gom dữ liệu từ nhiều nguồn (Windows, Linux, Firewall, IoT…).  Hỗ trợ phân tích chuyên sâu, truy vấn nhanh, đáp ứng chuẩn an toàn thông tin. |
| **Nhược điểm** | Không lưu trữ lâu dài, dễ bỏ sót sự kiện.  Tốn công sức nếu phải kiểm tra nhiều hệ thống. | Không có khả năng cảnh báo theo thời gian thực.  Nếu script gặp lỗi, log có thể bị bỏ lỡ mà không phát hiện ngay. | Cần hạ tầng phức tạp (máy chủ SIEM, agent, network config).  Chi phí triển khai và quản lý cao. |
| **Phạm vị áp dụng** | Môi trường nhỏ, kiểm tra nhanh, thử nghiệm | Tổ chức vừa và nhỏ, khi chưa có hệ thống tập trung, nhưng cần lưu trữ log phục vụ kiểm tra định kỳ. | Tổ chức lớn, môi trường cần giám sát 24/7, đáp ứng tiêu chuẩn ISO 27001, PCI-DSS. |

**Thu thập trực tiếp** phù hợp với kiểm tra đột xuất và môi trường nhỏ, nhưng không đáp ứng nhu cầu giám sát liên tục.

**Thu thập thủ công / định kỳ** nâng cao tự động hóa, giảm lỗi con người và phù hợp với môi trường vừa, nhưng vẫn hạn chế mở rộng.

**Thu thập tập trung** là chuẩn công nghiệp cho tổ chức lớn, cho phép giám sát liên tục và phân tích nâng cao, nhưng đòi hỏi hạ tầng và quản lý phức tạp

Tùy thuộc vào quy mô tổ chức, số lượng endpoint và mục tiêu giám sát, có thể kết hợp các phương pháp. Ví dụ: script tự động tại endpoint + tập trung thu thập vào SIEM để cân bằng chi phí và hiệu quả.



## Kết luận chương 2

Chương 2 nhấn mạnh rằng thu thập dữ liệu là bước nền tảng trong giám sát an toàn thông tin. Việc hiểu rõ các loại dữ liệu và chọn phương pháp thu thập phù hợp giúp đảm bảo tính toàn vẹn, sẵn sàng và bảo mật, đồng thời tạo tiền đề cho việc lưu trữ, phân loại và phân tích dữ liệu hiệu quả trong các chương tiếp theo.

# CHƯƠNG III: LƯU TRỮ VÀ PHÂN LOẠI DỮ LIỆU GIÁM SÁT



## Giới thiệu chung

Trong hệ thống giám sát an toàn thông tin, việc thu thập dữ liệu mới chỉ là bước khởi đầu. Để dữ liệu có thể phục vụ hiệu quả cho quá trình phân tích, phát hiện sự cố và điều tra số, chúng cần được lưu trữ và phân loại một cách khoa học.

Theo khuyến nghị của **NIST SP 800-92** – Guide to Computer Security Log Management (2006), việc lưu trữ dữ liệu giám sát phải đáp ứng một số mục tiêu cơ bản sau:

1. **Tính toàn vẹn (Integrity):** Dữ liệu log, network flow và cảnh báo phải được bảo đảm nguyên vẹn trong suốt vòng đời, không bị chỉnh sửa, giả mạo hay mất mát.
2. **Tính sẵn sàng (Availability):** Người quản trị có thể truy xuất dữ liệu nhanh chóng khi cần, đặc biệt trong các tình huống ứng phó sự cố và điều tra pháp lý.
3. **Tính bảo mật (Confidentiality):** Dữ liệu giám sát thường chứa thông tin nhạy cảm, do đó cần áp dụng cơ chế kiểm soát truy cập và mã hóa để ngăn chặn truy cập trái phép.
4. **Tính xác thực (Authenticity):** Hệ thống phải đảm bảo dữ liệu được sinh ra từ nguồn hợp lệ, có thể chứng minh tính đúng đắn và tránh nguy cơ chối bỏ trách nhiệm.
5. **Khả năng mở rộng (Scalability):** Hệ thống lưu trữ phải đáp ứng sự gia tăng liên tục của dữ liệu do số lượng người dùng, thiết bị và ứng dụng ngày càng nhiều.

Ngoài ra, NIST cũng nhấn mạnh đến việc xây dựng **chính sách lưu giữ dữ liệu** (log retention policy). Chính sách này xác định rõ loại dữ liệu nào cần được bảo quản dài hạn nhằm phục vụ yêu cầu tuân thủ pháp lý và điều tra, và loại dữ liệu nào chỉ nên lưu ngắn hạn để tối ưu chi phí và tài nguyên.

Song song với lưu trữ, việc **phân loại dữ liệu giám sát** giúp tổ chức nhận biết mức độ quan trọng và giá trị khai thác của từng loại dữ liệu. Nhờ đó, tổ chức có thể áp dụng các phương thức lưu trữ khác nhau, tối ưu hóa hạ tầng và nâng cao hiệu quả phân tích trong công tác đảm bảo an toàn thông tin.

## Lưu trữ dữ liệu giám sát

Việc lưu trữ dữ liệu giám sát là bước quan trọng nhằm đảm bảo các thông tin thu thập được có thể được sử dụng hiệu quả trong quá trình giám sát, phân tích và điều tra sự cố an toàn thông tin. Theo hướng dẫn của NIST (SP 800-92), có ba mô hình lưu trữ phổ biến: lưu trữ tại chỗ (local storage), lưu trữ tập trung (centralized storage) và lưu trữ trên đám mây (cloud storage).

### **Các mô hình lưu trữ log**

1. **Lưu trữ tại chỗ (Local Storage)**

**Mô tả:** Trong mô hình này, dữ liệu giám sát được lưu trực tiếp trên thiết bị sinh ra dữ liệu, chẳng hạn như máy chủ, tường lửa, IDS/IPS, thiết bị mạng hoặc endpoint.

**Ưu điểm:**

* Cách triển khai đơn giản, không đòi hỏi hạ tầng phức tạp.
* Cho phép truy xuất dữ liệu nhanh ngay tại thiết bị.

**Hạn chế:**

* Quản lý kém hiệu quả khi số lượng thiết bị tăng cao.
* Nguy cơ mất mát dữ liệu nếu thiết bị gặp sự cố (hỏng phần cứng, tấn công xóa log).
* Giới hạn về dung lượng lưu trữ.

**Ví dụ:**

* Windows: nhật ký hệ thống được lưu trong Event Viewer (Security, Application, System).
* Linux: log lưu tại thư mục /var/log/ (syslog, auth.log, dmesg).
* Firewall/Router: lưu log truy cập, NAT, kết nối VPN trên bộ nhớ trong hoặc thẻ nhớ.
* Thiết bị IoT (camera, sensor): log sự kiện (motion detection, đăng nhập trái phép) ghi trực tiếp trên thẻ nhớ.

1. **Lưu trữ tập trung (Centralized Storage)**

**Mô tả:** Dữ liệu từ nhiều nguồn được thu thập và chuyển về một máy chủ trung tâm (log server, SIEM hoặc hệ thống quản lý log như ELK, Wazuh).

**Ưu điểm:**

* Thuận lợi trong việc phân tích, đối chiếu dữ liệu từ nhiều nguồn khác nhau.
* Hỗ trợ triển khai chính sách bảo mật và kiểm soát truy cập tập trung.
* Giảm thiểu rủi ro mất dữ liệu so với mô hình lưu trữ phân tán.

**Hạn chế:**

* Yêu cầu đầu tư hạ tầng máy chủ và giải pháp lưu trữ tập trung.
* Có thể trở thành điểm lỗi đơn (single point of failure) nếu không được dự phòng hợp lý.

**Ví dụ:**

* Wazuh SIEM: tập trung log từ Windows Event Viewer, Linux syslog và firewall về một máy chủ để phân tích an ninh.

1. **Lưu trữ đám mây (Cloud Storage)**

**Mô tả:** Dữ liệu giám sát được gửi và lưu trữ trên nền tảng đám mây như AWS S3, Azure Blob Storage, Google Cloud Storage, hoặc dịch vụ SIEM cloud-based như Splunk Cloud, ELK Cloud.

**Ưu điểm:**

* Khả năng mở rộng linh hoạt theo nhu cầu.
* Giảm gánh nặng vận hành và hạ tầng nội bộ.
* Hỗ trợ tính sẵn sàng cao (High Availability) và phục hồi sau thảm họa (Disaster Recovery).

**Hạn chế:**

* Phụ thuộc vào nhà cung cấp dịch vụ.
* Nguy cơ về bảo mật và tuân thủ (compliance) nếu không triển khai cơ chế kiểm soát thích hợp.

**Ví dụ:**

* Splunk Cloud: lưu trữ và phân tích log bảo mật từ toàn bộ hạ tầng CNTT mà không cần máy chủ nội bộ.
* Google Chronicle: nền tảng log cloud với khả năng phân tích bảo mật nâng cao bằng trí tuệ nhân tạo.

1. **So sánh các mô hình lưu trữ**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tiêu chí** | **Lưu trữ tại chỗ (Local)** | **Lưu trữ tập trung (Centralized)** | **Lưu trữ trên đám mây (Cloud)** |
| **Triển khai** | Đơn giản, trực tiếp | Cần hạ tầng máy chủ | Phụ thuộc nhà cung cấp |
| **Quản lý** | Phân tán, khó tập trung | Tập trung, dễ quản lý | Tập trung, dễ quản lý |
| **Khả năng mở rộng** | Hạn chế | Tương đối linh hoạt | Rất linh hoạt |
| **Chi phí đầu tư ban đầu** | Thấp | Trung bình đến cao | Thấp |
| **Rủi ro mất dữ liệu** | Cao nếu thiết bị hỏng | Thấp hơn nhờ tập trung | Thấp |
| **Tuân thủ & bảo mật** | Khó kiểm soát khi phân tán | Dễ triển khai chính sách bảo mật | Phụ thuộc SLA của nhà cung cấp |

### **Chính sách lưu trữ**

Để đảm bảo dữ liệu giám sát phục vụ hiệu quả cho công tác phân tích, điều tra sự cố và tuân thủ pháp lý, cần xây dựng các chính sách lưu trữ cụ thể.

1. **Thời gian lưu trữ (Retention Policy)**

Theo NIST SP 800-92: log hệ thống cần được lưu ít nhất 1 năm.

Ví dụ thực tế: Trong môi trường doanh nghiệp, log hệ thống Windows và Linux thường được lưu 90 ngày trên máy chủ nội bộ, sau đó nén và chuyển sang kho lưu trữ dài hạn trên NAS hoặc Cloud trong 1 năm

1. **Sao lưu và dự phòng (Backup & Redundancy)**

Cần có cơ chế sao lưu định kỳ để phòng ngừa sự cố.

Ví dụ thực tế: Log được Wazuh Agent thu thập từ nhiều endpoint và gửi về server trung tâm (SIEM/ELK). Hàng tuần, log tại server được tự động sao lưu sang ổ cứng rời hoặc dịch vụ cloud (AWS S3) để đảm bảo không mất dữ liệu nếu máy chủ chính gặp sự cố.

1. **Phân quyền truy cập (Access Control)**

Cần áp dụng mô hình **3-2-1**: giữ ít nhất 3 bản sao log, trên 2 loại phương tiện lưu trữ khác nhau, trong đó 1 bản ở ngoài site chính (off-site).

Ví dụ: Log tập trung tại SIEM → sao lưu sang NAS nội bộ → lưu trữ dài hạn trên cloud (AWS S3 Glacier).

1. **Bảo mật dữ liệu lưu trữ (Security of Stored Data)**

Mọi dữ liệu phải được mã hóa khi truyền tải và lưu trữ (TLS/SSL, AES-256). Hệ thống cần giám sát, ghi nhật ký truy cập và bảo vệ vật lý hạ tầng lưu trữ để chống rò rỉ, giả mạo hoặc truy cập trái phép.

Ví dụ thực tế: Khi triển khai ELK, dữ liệu log được truyền qua TLS/SSL từ các agent về server. Ngoài ra, các file log lưu trong ổ cứng được mã hóa bằng AES-256.

1. **Tuân thủ và hủy dữ liệu (Compliance & Sanitization)**

Khi log hết hạn lưu trữ, cần xóa hoặc hủy an toàn theo chuẩn NIST SP 800-88 để tránh rủi ro rò rỉ dữ liệu.

Ví dụ thực tế: Log của người dùng cuối (có chứa thông tin cá nhân) được lưu 6 tháng để phục vụ điều tra sự cố. Sau 6 tháng, log sẽ được hủy bằng lệnh shred trên Linux theo hướng dẫn của NIST SP 800-88.

## Phân loại dữ liệu giám sát

Phân loại dữ liệu giám sát là bước quan trọng trong quản lý thông tin an toàn. Nhờ việc phân loại, tổ chức có thể xác định phương thức lưu trữ, chính sách bảo mật và mức độ ưu tiên truy xuất dữ liệu, từ đó tối ưu hóa hiệu quả giám sát và phân tích.

### **Theo nguồn gốc**

Nguồn gốc dữ liệu cho biết dữ liệu được sinh ra từ đâu, từ đó quyết định cách thu thập, chuẩn hóa và quản lý. Dữ liệu giám sát có thể xuất phát từ **log hệ thống** như Windows Event hoặc Linux syslog, từ **dữ liệu mạng** bao gồm Netflow, firewall log và cảnh báo IDS/IPS, hoặc **từ các thiết bị đầu cuối (endpoint)** như thao tác người dùng, sensor, camera. Việc phân loại theo nguồn gốc không chỉ giúp áp dụng cơ chế bảo mật phù hợp mà còn hỗ trợ lựa chọn phương pháp lưu trữ và phân tích dữ liệu hiệu quả.

### **Theo mức độ quan trọng**

Theo hướng dẫn của NIST SP 800-92, dữ liệu giám sát thường được phân loại thành 3 cấp độ quan trọng để quản lý và bảo vệ hiệu quả:

* **Cao (High / Critical):**
* Bao gồm các dữ liệu nhạy cảm, cảnh báo bảo mật nghiêm trọng hoặc thông tin cá nhân.
* Những dữ liệu này yêu cầu bảo vệ chặt chẽ, quyền truy cập hạn chế, sao lưu đầy đủ và lưu trữ dài hạn.
* **Trung bình (Medium / Moderate):**
* Bao gồm log vận hành và nhật ký sự kiện thông thường.
* Dữ liệu ở mức này cần giám sát định kỳ và lưu trữ vừa phải, áp dụng các biện pháp bảo mật tiêu chuẩn.
* **Thấp (Low / Minimal):**
* Bao gồm log tạm thời, debug hoặc dữ liệu kiểm thử.
* Dữ liệu này có thể lưu trữ ngắn hạn, bảo mật cơ bản, và chủ yếu phục vụ tham khảo hoặc kiểm thử hệ thống.

Việc phân loại dữ liệu theo 3 cấp độ này giúp tổ chức ưu tiên nguồn lực, áp dụng chính sách lưu trữ và bảo mật hợp lý, tránh lãng phí và đảm bảo các dữ liệu quan trọng luôn được bảo vệ và truy xuất khi cần.

### **Theo mục đích giám sát**

Mục đích giám sát xác định cách dữ liệu được sử dụng trong tổ chức:

* Giám sát an ninh: hỗ trợ phát hiện và ứng phó sự cố, ví dụ log firewall, IDS/IPS hoặc cảnh báo nghiêm trọng.
* Giám sát vận hành: giúp tối ưu hóa hiệu suất hệ thống, như log CPU, memory, truy cập ứng dụng.
* Tuân thủ pháp lý và kiểm toán: lưu giữ để chứng minh các hoạt động và đáp ứng yêu cầu pháp lý.

Phân loại theo mục đích giúp tổ chức quản lý, lưu trữ và bảo vệ dữ liệu hợp lý, đồng thời nâng cao hiệu quả tổng thể của hệ thống giám sát. Nhờ kết hợp cả nguồn gốc, mức độ quan trọng và mục đích giám sát, tổ chức có thể quản lý dữ liệu một cách hiệu quả, tối ưu hóa phân tích rủi ro và đảm bảo tuân thủ các yêu cầu pháp lý.

## Kết luận chương 3

Chương 3 trình bày đầy đủ về lưu trữ và phân loại dữ liệu giám sát, nhấn mạnh tầm quan trọng của việc lưu trữ hợp lý để đảm bảo tính toàn vẹn, sẵn sàng và bảo mật theo hướng dẫn của NIST. Các mô hình lưu trữ, chính sách lưu giữ, sao lưu, phân quyền truy cập và tiêu chí hủy dữ liệu được phân tích, cùng với việc phân loại dữ liệu theo mức độ quan trọng (High, Medium, Low) giúp tổ chức tối ưu hóa quản lý, phân bổ nguồn lực và bảo vệ dữ liệu quan trọng. Khi kết hợp lưu trữ hợp lý và phân loại hiệu quả, tổ chức có thể truy xuất nhanh, phân tích dữ liệu hiệu quả, phát hiện rủi ro sớm, phản ứng kịp thời với sự cố, đồng thời đáp ứng các yêu cầu pháp lý và kiểm toán.

# KẾT LUẬN

Qua ba chương nghiên cứu, đồ án đã triển khai mục tiêu nghiên cứu về giám sát an toàn thông tin theo từng nội dung cụ thể:

Chương 1 cung cấp nền tảng nhận thức về an toàn thông tin, giúp người đọc nắm vững các khái niệm cơ bản (toàn vẹn, bảo mật, sẵn sàng, xác thực dữ liệu), thực trạng ATTT tại Việt Nam, xu hướng sử dụng dịch vụ giám sát, các mối đe dọa và kỹ thuật tấn công phổ biến, cùng các giải pháp bảo vệ cơ bản như Antivirus, Antispam, Firewall, IDPS, Honeypot, VPN và DLP. Đây là cơ sở lý thuyết quan trọng để triển khai các bước tiếp theo của giám sát ATTT.

Chương 2 tập trung vào thu thập dữ liệu giám sát, phân tích các loại dữ liệu và phương pháp thu thập (trực tiếp, gián tiếp, thủ công). Đặc biệt, thu thập tập trung (centralized collection) được đề cập nhưng chưa khai thác và làm rõ hết vai trò trong việc tổng hợp, chuẩn hóa, phân tích và phản ứng sự cố.

Chương 3 trình bày việc lưu trữ và phân loại dữ liệu theo hướng dẫn của NIST, bao gồm các mô hình lưu trữ, chính sách lưu giữ, phân quyền truy cập, phân loại theo mức độ quan trọng và hủy dữ liệu khi hết hạn. Việc lưu trữ và phân loại hợp lý giúp tối ưu hóa quản lý, truy xuất, phân tích dữ liệu, đồng thời nâng cao khả năng phát hiện và ứng phó sự cố, đảm bảo tuân thủ pháp lý và kiểm toán.

Mặc dù đồ án đã hoàn thành các mục tiêu nghiên cứu cơ bản, thời gian thực hiện ngắn, một số nội dung như thu thập dữ liệu tập trung, đánh giá hiệu quả trên hệ thống thực tế và thử nghiệm các tình huống sự cố vẫn còn hạn chế. Đây là những điểm cần được bổ sung, hoàn thiện trong các nghiên cứu tiếp theo để xây dựng một hệ thống giám sát an toàn thông tin toàn diện và hiệu quả hơn.

# TÀI LIỆU THAM KHẢO

[1] An toàn thông tin phải được đặt lên hàng đầu: <https://baochinhphu.vn/an-toan-thong-tin-phai-duoc-dat-len-hang-dau-102241121133114553.htm>

[2] An ninh mạng là gì, giải thích về an ninh mạng:

<https://aws.amazon.com/vi/what-is/cybersecurity/>

[3] 10 loại tấn công mạng phổ biến: <https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/common-cyberattacks/>

[4] **NIST SP 800-92:** <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-92.pdf>

**[5] NIST SP 800-137:** <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-137.pdf>

[6] Nhật ký linux: <https://phoenixnap.com/kb/how-to-view-read-linux-log-files>

[7] Phân tích và xoá nhât ký linux bằng Shell Script : <https://www.fosstechnix.com/parse-and-delete-linux-logs-using-shell-script/>

[8] Wazuh - giải pháp SIEM & XDR: <https://documentation.wazuh.com/current/getting-started/index.html>

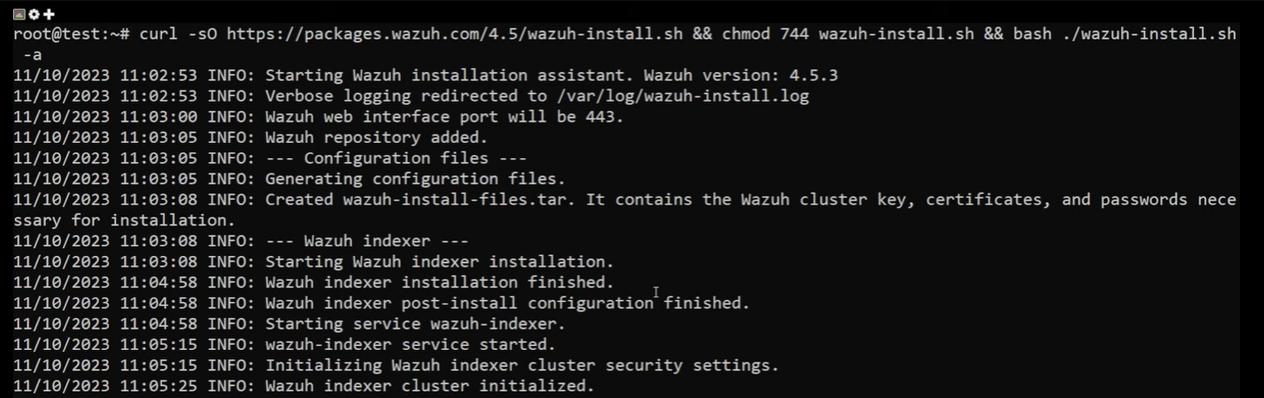
[9] Install wazuh server and agent: <https://www.youtube.com/watch?v=pCKApbHhQiY>

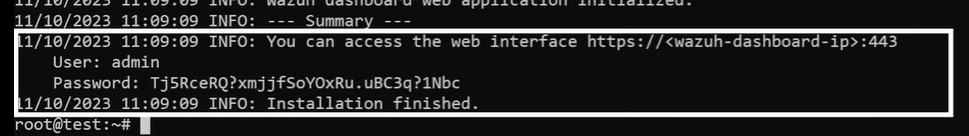
# PHỤ LỤC

**Phụ lục 1: Wazuh - giải pháp SIEM & XDR**

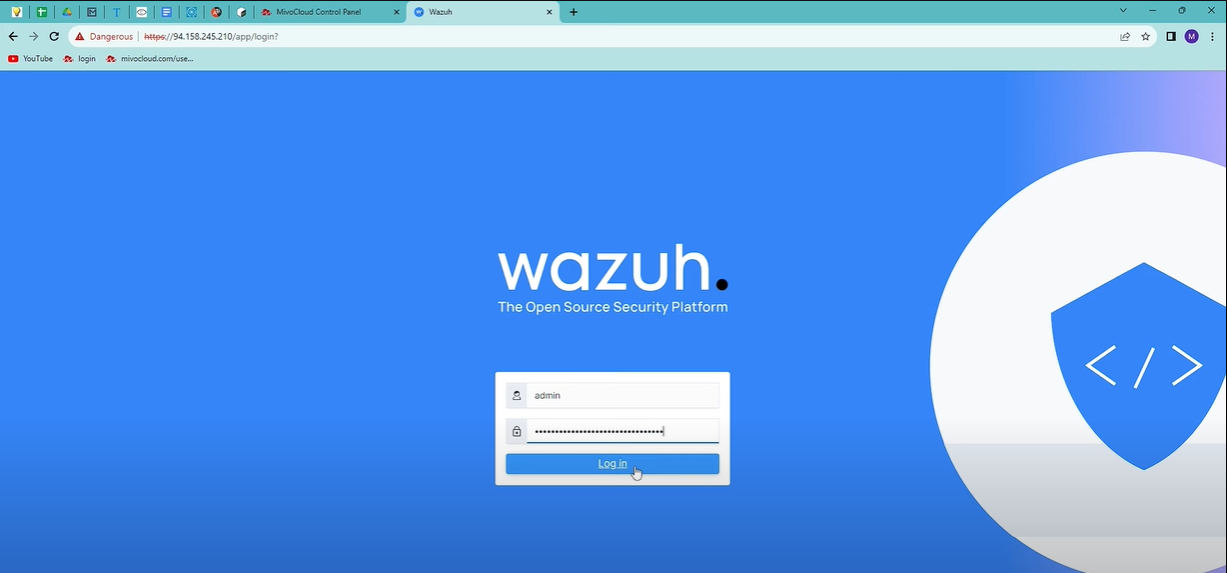
1. **Cài đặt Wazuh server**

|  |
| --- |
| *curl -s https://packages.wazuh.com/key/GPG-KEY-WAZUH | gpg --no-default-keyring --keyring gnupg-ring:/usr/share/keyrings/wazuh.gpg --import && chmod 644 /usr/share/keyrings/wazuh.gpg* |





Sau khi cài đặt thành công sẽ ta sẽ có 1 url để truy cập vào bảng điều khiển cũng như tên đăng nhập



1. **Cài đặt wazuh agents trên window**

Sử dụng PowerShell: .\wazuh-agent-4.12.0-1.msi /q WAZUH\_MANAGER="Địa chỉ IP Wazuh server "