### CHƯƠNG 2. PHÁT HIỆN TẤN CÔNG MẠNG VỚI CÔNG NGHỆ SIEM

#### **2.1. Giới thiệu về công nghệ Siem**

Hệ thống mạng ngày càng phát triển, một số nhu cầu đặt ra như: Làm sao để quản lý rất nhiều các thiết bị, việc quản lý thông tin và sự kiện an ninh trong hệ thống vận hành tốt và an toàn. Đó là câu hỏi mà các nhà nghiên cứu hay các quản trị viên an ninh mạng đang đi tìm câu trả lời sao cho tối ưu nhất. Đã có rất nhiều các thiết bị IDS, IPS ra đời nhằm trả lời điều đó. Nhưng hầu hết các thiết bị này đều hoạt động riêng lẻ, quản lý tập chung là không có, sự tương quan liên kết các sự kiện an ninh cũng không có, bên cạnh đó là việc đảm bảo tuân thủ an ninh còn hạn chế và có thể là rất mệt mỏi để hoàn thành các báo cáo về tình trạng an ninh của hệ thống. Một số giải pháp trước khi công nghệ Siem ra đời:

* Giải pháp quản lý an ninh thông tin - Security information management (SIM)

SIM là giải pháp công nghệ đầu tiên trong các giải pháp giám sát an ninh mạng. Ban đầu giải pháp SIM chỉ có khả năng lưu trữ các nhật ký sự kiện an ninh cho các hệ thống mạng (đây cũng là một trong các chức năng chính của giải pháp giám sát an ninh hiện nay). Hạn chế của giải pháp này là không có khả năng phân tích các sự kiện an ninh mà chỉ thực hiện việc lưu trữ chúng. Các nhật ký này chủ yếu là từ: Hệ Thống máy chủ, các ứng dụng, thiết bị network và từ các thiết bị chuyên về Security. Các thành phần chính của SIM bao gồm: Thành phần thu thập nhật ký, thành phần lưu trữ.

* Giải pháp quản lý các sự kiện an ninh - Security event management (SEM).SEM thực hiện việc xử lý log và các sự kiện an ninh từ các thiết bị gửi về bao gồm: Các thiết bị mạng (network devices), các máy chủ (Server), các ứng dụng theo thời gian thực nhằm thực việc việc theo dõi các sự kiện an ninh xẩy ra trong hệ thống, phân tích tính tương quan và thực hiện các hành động nhằm đảm bảo an toàn cho hệ thống. Các thành phần chính của hệ thống bao gồm: Thành phần thu thập nhật ký, thành phần phân tích nhật ký. Ngoài ra còn có các thành phần khác như: thành phần quản lý phản ứng trước các mối đe dọa (Threats Response Management), các module tạo báo cáo (Complaince Report).
* Giải pháp quản lý đăng nhập - Log Management System (LMS)

LMS là một hệ thống thu thập và lưu trữ tập tin đăng nhập (từ hệ điều hành, ứng dụng, ...vv). Thông tin tập trung được thu thập từ nhiều nguồn. Người quản trị thay vì phải kiểm tra từng hệ thống riêng lẻ thì quản trị tập trung tại một điểm duy nhất.

* Tương quan sự kiện an ninh - Security Event Correlation (SEC)

SEC là giải pháp tương quan các sự kiện an ninh thu thập được theo các quy tắc đã cài đặt nhằm tăng hoặc giảm mức cảnh báo đối với một sự kiện an ninh. - Giải pháp quản lý thông tin và sự kiện an ninh - Security information and event management (SIEM)

Giải pháp quản lý thông tin và sự kiện an ninh là một giải pháp bảo mật an ninh cung cấp cái nhìn tổng thể về hệ thống công nghệ thông tin của một tổ chức. SIEM là sự kết hợp của các giải pháp nêu trên.

Log từ các thiết bị tạo ra ngày càng nhiều (Có thể hàng trăm triệu bản ghi log trong một ngày) thì các quản trị viên hay các công cụ trước đó khó có thể phân tích một cách nhanh chóng, chính xác được. SIEM cung cấp việc tích hợp dữ liệu quản lý file log từ nhiều nguồn, bao gồm cả mạng, máy chủ, cơ sở dữ liệu, ứng dụng, cung cấp khả năng hợp nhất dữ liệu để tránh mất các sự kiện quan trọng. [9]

Giải pháp Quản lý và phân tích sự kiện an toàn thông tin là giải pháp toàn diện và hoàn chỉnh, cho phép các cơ quan, tổ chức thực hiện việc giám sát các sự kiện an toàn thông tin cho một hệ thống.

Đưa ra

cảnh báo

Tương quan sự kiện, Phát hiện

bất thường

Thiết bị bảo mật

Máy chủ

Thiết bị mạng

Cơ sở dữ liệu

Ứng dụng

Thông tin cấu

hình

Thông tin lỗ hổng

Hoạt động của

người dùng

# Hình 2.1: Hệ thống phát hiện tấn công mạng

Các thành phần chính của SIEM bao gồm: thu thập nhật ký, phân tích và lưu trữ, quản trị tập trung.

Nguyên lý cơ bản của SIEM là thu các dữ liệu về các sự kiện an ninh từ nhiều thiết bị khác nhau ở các vị trí khác nhau trong hệ thống và chúng ta có thể dễ dàng phân tích, theo dõi tất cả các dữ liệu ở tại một vị trí duy nhất để phát hiện xu hướng và theo dõi các dấu hiệu bất thường [9].

SIEM thu thập Log và các tài liệu liên quan đến an ninh khác để phân tích, tương quan liên kết. SIEM làm việc thu thập Log và các sự kiện an ninh thông qua các

Agent. Từ người dùng đầu cuối, các máy chủ, các thiết bị mạng và thậm chí là các thiết bị an ninh chuyên nghiệp như Firewall, AntiVirus hoặc các hệ thống phòng chống xâm nhập. Các thiết bị thu thập thông tin chuyển tiếp thông tin tới trung tâm nhằm chuẩn hóa, quản lý tập trung, phân tích, tương quan các sự kiện an ninh. Tiếp sau đó có thể xác định các sự kiện bất thường và thông báo tới quản trị viên.

SIEM được sử dụng nhằm theo dõi, xác định, quản lý hệ thống tài sản và ứng phó với các sự cố an ninh. Một số sự kiện an ninh như tấn công từ chối dịch vụ (DoS), tấn công có chủ ý, tấn công mã độc hại và phát tán virus. SIEM cũng có thể xác định mà không dễ phát hiện bằng các thiết bị khác. Nhiều sự kiện khó phát hiện hoặc bị che khuất bởi hàng ngàn các sự kiện an ninh khác trong mỗi giây. Bên cạnh đó SIEM có thể phát hiện những sự kiện an ninh khó phát hiện hơn như các hành vi vi phạm chính sách, cố gắng truy cập trái phép và phương thức tấn công của những kẻ tấn công có trình độ cao xâm nhập vào hệ thống.

Một mục tiêu quan trọng cho các nhà phân tích an ninh sử dụng SIEM là giảm số lượng cảnh báo giả. Hệ thống an ninh được cho là yếu kém, chẳng hạn như hệ thống phát hiện xâm nhập (IDS) thường có những cảnh báo về nhiều sự kiện giả. Nhiều cảnh báo này gây lãng phí thời gian, công sức của các nhà phân tích an ninh và thường tập trung chú ý vào các cảnh báo đó. Điều đó làm cho các nhà phân tích lại bỏ qua các cảnh báo quan trọng hơn. Với hệ thống SIEM, việc giảm các cảnh báo giả được thực hiện một cách cẩn thận bởi các bộ lọc và các quy tắc tương quan liên kết giữa các thông tin sự kiện an ninh với nhau. Điều đó xác định và cảnh báo chính xác khi có sự kiện an ninh bất chấp số lượng lớn những sự kiện an ninh.

SIEM cung cấp các dịch vụ sau:

* Quản lý nhật ký sự kiện an ninh (Log management).
* Tuân thủ các quy định về CNTT (IT regulatory compliance).
* Tương quan liên kết các sự kiện an ninh (Event correlation).
* Cung cấp các hoạt động ứng phó (Active response).
* Đảm bảo an ninh thiết bị đầu cuối (Endpoint security).

##### 2.1.1. Quản lý nhật ký sự kiện an ninh

SIEM quản lý Log từ các thiết bị trong hệ thống. Bắt đầu với việc cấu hình các vị trí quan trọng trong hệ thống để gửi các sự kiện an ninh vào một cơ sở dữ liệu tập trung. SIEM sẽ chuẩn hóa các Log này về một định dạng duy nhất để phân tích, tương quan liên kết. Sau đó, SIEM lưu trữ các file Log, tổ chức, tìm kiếm và các dịch vụ khác để đáp ứng nhu cầu quản lý mà các tổ chức yêu cầu. Phần quản lý dữ liệu này cũng sử dụng để phân tích về thời gian thực, tình trạng khai thác dữ liệu và an ninh của toàn bộ hệ thống.

##### 2.1.2. Tuân thủ các quy định về CNTT

Tất cả các sự kiện từ các hệ thống quan trọng đang được sử dụng truy nhập, chúng ta có thể xây dựng các bộ lọc hoặc các thiết lập các luật và tính toán thời gian để kiểm tra và xác thực việc tuân thủ của họ hoặc để xác định hành vi vi phạm các yêu cầu tuân thủ đã đặt ra của tổ chức. Các luật đó được đối chiếu với log được đưa vào hệ thống. Có thể giám sát số lần thay đổi mật khẩu, xác định hệ điều hành hoặc các bản vá lỗi ứng dụng, kiểm tra chống virus, phần mềm gián điệp và cập nhật. Chúng ta có thể xây dựng tập luật riêng của mình cho các bộ lọc hoặc các luật để hỗ trợ trong việc tuân thủ các quy định đã đề ra. Nhiều nhà cung cấp SIEM có các tập đóng gói sẵn các quy tắc được thiết kế đặc biệt để đáp ứng các yêu cầu về pháp luật và các quy định khác nhau mà các doanh nghiệp cần phải tuân thủ. Chúng được đóng gói và cung cấp bởi các nhà cung cấp một cách miễn phí hoặc mất một khoản chi phí.

##### 2.1.3. Tương quan liên kết các sự kiện an ninh

Sự tương quan liên kết giữa các sự kiện an ninh mang đem lại thông báo tốt hơn cho hệ thống. Chúng ta không chỉ qua một sự kiện duy nhất để quyết định cách ứng phó hay không ứng phó với nó. Với tương quan liên kết giữa các sự kiện an ninh, chúng ta xem xét điều kiện khác nhau trước khi kích hoạt báo động. Ví dụ, một máy chủ có CPU sử dụng 100% có thể được gây ra bởi nhiều nguyên nhân khác nhau. Nó có thể do một vấn đề xảy ra hoặc có thể không. Cũng có thể là một dấu hiệu cho thấy hệ thống bị quá tải với các hoạt động và yêu cầu một hoặc nhiều dịch vụ hoặc các ứng dụng cần được chia sẻ trên các máy chủ khác. Và cũng có thể là máy chủ đạt đến hết công suất do bị tấn công từ chối dịch vụ (DoS) vào hệ thống. Hoặc nó có thể là ngừng trệ tạm thời một cách tự nhiên của máy chủ.

Các công cụ tương quan trên một SIEM có thể kiểm tra và xem xét (tương quan) các sự kiện khác không phải liên quan đến việc sử dụng CPU. Có thể cung cấp một bức tranh đầy đủ hơn về tình trạng của máy chủ để loại trừ giả thuyết về nguyên nhân của vấn đề. Ví dụ, trong trường hợp sử dụng CPU 100%, SIEM có thể được cấu hình để xem xét một số nguyên nhân sau đây:

* Phần mềm chống virus xác định có phần mềm độc hại trên máy chủ hay không. - Bất kỳ máy chủ nào có CPU sử dụng 100% thì cần xem xét có hay không sự tồn tại của virus.
* Một ứng dụng hoặc nhiều ứng dụng, dịch vụ ngừng hoạt động.
* Sự gia tăng lưu lượng mạng do nhu cầu chính đáng của người dùng nhưng vượt quá sự cung cấp dịch vụ của máy chủ.
* Sự gia tăng lưu lượng mạng nhưng không do nhu cầu chính đáng của người dùng vượt quá sự cung cấp dịch vụ của máy chủ như một cuộc tấn công DoS. Đó là sự tương quan các sự kiện an ninh. Cảnh báo của SIEM giúp chúng ta đưa ra cách ứng phó tùy thuộc vào các điều kiện.

##### 2.1.4. Cung cấp các hoạt động ứng phó

Tất cả các thiết bị cung cấp đầu vào cho SIEM, các quy tắc và bộ lọc sẽ xác định và phân tích mối quan hệ giữa các thông tin đầu vào đó. Chúng ta có thể cấu hình các hành động và thực hiện các phản ứng ứng phó cho tất cả các sự kiện an ninh hoặc có thể cấu hình riêng biệt cho từng loại sự kiện khác nhau.

Lợi ích việc thực hiện các hoạt động ứng phó là rất tốt, nhưng bên cạnh đó nó cũng có điều bất lợi. Nếu chúng ta không cấu hình cẩn thận và chính xác thì nó có thể đưa ra các hành động ứng phó không cần thiết.

##### 2.1.5. Đảm bảo an ninh thiết bị đầu cuối

Hầu hết các hệ thống SIEM có thể giám sát an ninh cho các thiết bị đầu cuối để thông báo sự an toàn của hệ thống. SIEM cung cấp việc quản lý cũng như đánh giá tài sản các thiết bị. Bên cạnh là việc dò quét lỗ hổng và cập nhật các bản vá. Nhiều hệ thống SIEM có thể theo dõi các thiết bị như PC, server, Firewall. Một số hệ thống SIEM thậm chí có thể quản lý an ninh cho thiết bị đầu cuối, có sự điều chỉnh và hoàn thiện hơn đối với thiết bị an ninh đó trên hệ thống. Như cấu hình Firewall, cập nhật và theo dõi Anti-Virus, chống spyware, chống spam email.

##### 2.2. Thành phần và hoạt động của Siem

SIEM bao gồm nhiều phần, mỗi phần làm một công việc riêng biệt. Mỗi thành phần trong hệ thống này có thể hoạt động độc lập với các thành phần khác nhưng nếu tất cả không cùng hoạt động một lúc thì chúng ta sẽ không có một hệ thống SIEM hiệu quả. Tùy thuộc vào hệ thống đang sử dụng nhưng mỗi SIEM sẽ luôn luôn có các thành phần cơ bản cơ bản được mô tả trong phần này. Bằng sự hiểu biết từng phần của SIEM và cách thức hoạt động, chúng ta có thể quản lý một cách hiệu quả và khắc phục sự cố các vấn đề khi phát sinh.

Việc quan trọng khi thực hiện triển khai SIEM là cần phải hiểu nó làm việc như thế nào. Đối với mỗi nhà cung cấp khác nhau sẽ có đôi chút khác nhau nhưng chúng đều dựa trên những khái niệm cốt lõi. Thành phần cơ bản vẫn là thu thập thông tin, phân tích và lưu trữ. Các bản ghi Log được thu thập từ các thiết bị khác nhau và chúng có thể có những định dạng theo từng loại thiết bị. Chúng ta cần thu thập và chuyển nó về một định dạng chung. Quá trình này gọi là chuẩn hóa dữ liệu. Sau đó sẽ tiến hành phân tích từ các dữ liệu này và thực hiện tương quan sự kiện an ninh để đưa tới kết luận có một cuộc tấn công hay không. Các thông tin về môi trường mạng và các mối đe dọa phổ biến rất có ích trong giai đoạn này. Việc đưa ra cảnh báo và các báo cáo sẽ được tạo ra như một kết quả của việc phân tích. Các bản ghi Log được lưu trữ trực tiếp trên SIEM ít nhất vài giờ đồng hồ sau đó chuyển tới nơi lưu trữ lâu dài để phục vụ cho quá trình điều tra hoặc sử dụng sau này.

###### 2.2.1. Thiết bị Nguồn

Thành phần đầu tiên của Siem là các thiết bị đầu vào cung cấp dữ liệu cho Siem. Thiết bị nguồn có thể là một thiết bị thực tế trong hệ thống mạng như Router, Switch hoặc một số loại máy chủ và cũng có thể là các bản ghi log từ một ứng dụng hoặc chỉ là dữ liệu bất kỳ. Việc biết về những gì mình có trong hệ thống là rất quan trọng trong việc triển khai SIEM. Hiểu rõ những nguồn mà chúng ta muốn lấy các bản ghi log trong giai đoạn đầu sẽ giúp chúng ta tiết kiệm được công sức, số tiền đáng kể và giảm sự phức tạp trong triển khai.

Hệ điều hành: Microsoft Windows và các biến thể của Linux và UNIX, AIX, Mac OS là những hệ điều hành thường hay được sử dụng. Hầu hết các hệ điều hành vê cơ bản công nghệ khác nhau và thực hiện một nhiệm vụ nào đó nhưng một trong những điều mà tất cả đều có điểm chung là chúng tạo ra các bản ghi log. Các bản ghi log sẽ cho thấy hệ thống của bạn đã làm gì: Ai là người đăng nhập, làm những gì trên hệ thống... Các bản ghi log được tạo ra bởi một hệ điều hành về hệ thống và người sử dụng hoạt động sẽ rất hữu ích khi tiến hành ứng phó sự cố an ninh hoặc chẩn đoán vấn đề hay chỉ là việc cấu hình sai.

Thiết bị: Hầu hết các thiết bị là các hộp đen, các quản trị hệ thống không có quyền truy cập trực tiếp vào hệ thống để thực hiện một số việc quản lý cơ bản. Nhưng có thể quản lý các thiết bị thông qua một giao diện. Giao diện này có thể dựa trên web, dòng lệnh hoặc chạy qua một ứng dụng được tải về máy trạm của quản trị viên. Hệ điều hành các thiết bị mạng chạy có thể là một hệ điều hành thông thường, chẳng hạn như Microsoft Windows hoặc phiên bản của Linux, nhưng nó cũng có thể là một hệ điều hành riêng biệt. Ví dụ như một router hoặc switch. Nó phụ thuộc vào nhà cung cấp, chúng ta không bao giờ có thể truy cập trực tiếp vào hệ thống điều hành cơ bản của nó mà chỉ có thẻ truy cập vào thông qua dòng lệnh hoặc giao diện web được sử dụng để quản lý. Các thiết bị lưu trữ các bản ghi log của chúng trên hệ thống hoặc thường có thể được cấu hình để gửi các bản ghi ra thông qua syslog hoặc FTP.

Ứng dụng: Chạy trên các hệ điều hành là những ứng dụng được sử dụng cho một loạt các chức năng. Trong một hệ thống chúng ta có thể có hệ thống tên miền (DNS), dịch vụ cấp phát địa chỉ động (DHCP), máy chủ web, hệ thống thư điện tử và vô số các ứng dụng khác. Các bản ghi ứng dụng chứa thông tin chi tiết về tình trạng của ứng dụng, ví dụ như thống kê, sai sót, hoặc thông tin tin nhắn. Một số ứng dụng sinh ra bản ghi log sẽ có ích cho chúng ta.

Xác định bản ghi log cần thiết: Sau khi xác định các thiết bị nguồn trong hệ thống, chúng ta cần xem xét việc thu thập các bản ghi log từ các thiết bị nào là cần thiết và quan trọng cho SIEM. Một số điểm cần chú ý trong việc thu thập các bản ghi log như sau:

* Thiết bị nguồn nào được ưu tiên. Dữ liệu nào là quan trọng mà chúng ta cần phải thu thập.
* Kích thước các bản ghi log sinh ra trong khoảng thời gian nhất định là bao nhiêu. Những thông tin này dùng để xác định SIEM cần bao nhiêu tài nguyên cho chúng, đặc biệt là không gian lưu trữ.
* Tốc độ các thiết bị nguồn này sinh ra các bản ghi log là bao lâu. Thông tin này cùng với kích thước bản ghi log để lựa chọn việc sử dụng đường truyền mạng khi thu thập các bản ghi.
* Cách thức liên kết giữa các thiết bị nguồn với SIEM.
* Có cần các bản ghi log theo thời gian thực hay thiết lập quá trình thực hiện tại một thời điểm cụ thể trong ngày.

Các thông tin trên rất có ích trong việc xác định nguồn thiết bị cần thiết cho SIEM của chúng ta. Chúng có quá nhiều nhưng nó là cần thiết để xác định chính xác hơn điều gì là cần thiết cho SIEM. Số lượng người sử dụng, lịch bảo trì hệ thống và nhiều yếu tố khác có thể tác động đáng kể đến số lượng các bản ghi log được tạo ra mỗi ngày.

###### 2.2.2. Thu thập Log

SIEM thu thập các bản ghi Log từ rất nhiều các thiết bị khác nhau, việc truyền các bản ghi log từ các thiết bị nguồn tới SIEM cần được giữ bí mật, xác thực và tin cậy bằng việc sử dụng syslog hoặc các giao thức SNMP, OPSEC, SFTP, IDXP. Sau đó các bản ghi log chuẩn hóa đưa về cùng một định dạng. Nếu các thiết bị không hỗ trợ syslog hay các giao thức này chúng ta cần phải sử dụng các Agent. Đó là một điều cần thực hiện để thực hiện việc lấy các bản ghi log có định dạng mà SIEM có thể hiểu được. Việc cài đặt các Agent có thể kéo dài quá trình triển khai SIEM nhưng chúng ta sẽ có những bản ghi log theo dạng chuẩn mong muốn.

Khi các sự kiện an ninh được gửi đến máy chủ, mức độ ưu tiên sẽ được định dạng theo chuẩn từ 0 đến 5. Người quản trị có thể điều chỉnh các giá trị mặc định thông qua một bảng tiêu chuẩn và chính sách ưu tiên.

Chính sách thu thập thông tin: Có thể thiết lập một chính sách ưu tiên và thu thập ở các bộ cảm biến để lọc và củng cố các thông tin sự kiện an ninh trước khi gửi chúng đến máy chủ. Kỹ thuật này cho phép người quản trị để điều tiết sự kiện an ninh và quản lý những thông tin, nếu không sẽ rất nhiều các sự kiện an ninh trong hệ thống mạng làm cho chúng ta lúng túng không biết bắt đầu từ đâu.

Cơ chế thu thập các bản ghi log phụ thuộc vào từng thiết bị và có các phương thức thu thập như sau:

Push log: Các bản ghi log sẽ được các thiết bị nguồn gửi về SIEM. Phương pháp này có lợi ích: Dễ dàng cài đặt và cấu hình. Thông thường, chúng ta chỉ cần thiết lập một bộ tiếp nhận và sau đó kết nối thiết bị nguồn đến bộ phận tiếp nhận này. Ví dụ như syslog. Khi cấu hình thiết bị nguồn sử dụng syslog, chúng ta có thể thiết lập địa chỉ IP hoặc DNS tên của một máy chủ syslog trên mạng và thiết bị sẽ tự động gửi các bản ghi của nó thông qua syslog. Tuy nhiên phương pháp nay cũng còn một số nhược điểm. Ví dụ, sử dụng syslog trong môi trường UDP. Bản chất vốn của việc sử dụng syslog trong môi trường UDP có nghĩa là không bao giờ có thể đảm bảo rằng các gói tin đến đích, vì UDP là một giao thức không hướng kết nối. Nếu một tình huống xảy ra trên mạng chẳng hạn như khi một loại virus mạnh trên mạng, chúng ta có thể không nhận được gói tin syslog. Một vấn đề có thể phát sinh là nếu không đặt quyền điều khiển truy cập thích hợp trên máy thu nhận các bản ghi log thì khi cấu hình sai hoặc có phần mềm độc hại có thể làm tràn ngập các thông tin sai lệch. Điều đó làm cho các sự kiện an ninh khó được phát hiện. Nếu là một cuộc tấn công có chủ ý nhằm chống lại SIEM thì một kẻ xấu có thể làm sai lệch các thông tin và và thêm các dữ liệu rác vào SIEM. Do vậy sự hiểu biết về các thiết bị gửi các bản ghi log cho SIEM là điều rất quan trọng.

Pull log: Các bản ghi log sẽ được SIEM đi tới và lấy về. Không giống như phương pháp Push log, trong đó thiết bị nguồn gửi các bản ghi log cho SIEM mà không cần bất kỳ sự tương tác từ SIEM. Pull log đòi hỏi SIEM bắt đầu kết nối với các thiết bị nguồn và chủ động lấy các bản ghi từ các thiết bị nguồn đó. Một ví dụ nếu các bản ghi log được lưu trữ trong tập tin văn bản chia sẻ trên một mạng. SIEM sẽ thiết lập một kết nối lấy các thông tin được lưu trữ và đọc các file bản ghi từ các thiết bị nguồn. Đối với phương pháp Push Log, các bản ghi log của thiết bị nguồn thường gửi các bản ghi đến SIEM ngay sau khi nó được tạo ra. Nhưng với phương pháp Pull Log thì một kết nối sẽ được tạo ra để SIEM tiếp cận với các thiết bị nguồn và kéo các bản ghi log từ các thiết bị nguồn về. Chu kỳ của việc kết nối để lấy các bản ghi log của Pull Log có thể là vài giây hoặc theo giờ. Khoảng thời gian này chúng ta có thể cấu hình theo tùy chọn hoặc để cấu hình mặc định cho SIEM.

Prebuilt Log collection: Tùy thuộc vào SIEM, thường có các phương pháp được xây dựng sẵn có sẵn để lấy được các bản ghi từ các thiết bị hoặc các ứng dụng. Ví dụ, chúng ta có thể trỏ một máy chủ SIEM chạy cơ sở dữ liệu Oracle cung cấp cho các thông tin cơ sở dữ liệu SIEM. SIEM sẽ có những phương pháp xác thực và quy tắc (Rules) được xây dựng để lấy thông tin từ cơ sở dữ liệu Oracle. Ví dụ này làm cho việc lấy các bản ghi từ các thiết bị nguồn được dễ dàng hơn. Nhưng đối với một ứng dụng nào đó mà chúng ta muốn lấy những bản ghi log nhưng không có phương pháp hay quy tắc nào được xác định trước đó thì hơi khó. Trong trường hợp này, cần thay đổi các bản ghi từ các định dạng file gốc thành một dạng mà SIEM có thể hiểu được. Một ví dụ là nếu chúng ta đang chạy một ứng dụng trên một máy chủ và ứng dụng lưu trữ các bản ghi của nó trong một định dạng tập tin trên máy chủ. Chúng ta có thể sử dụng một ứng dụng khác để đọc tập tin này và gửi các bản ghi thông qua syslog.

Trong trường hợp máy chủ Windows, để làm việc với các bản ghi không chuẩn thì sử dụng Windows Event Log và đưa Windows Event Log vào SIEM.

###### 2.2.3. Chuẩn hóa và tổng hợp sự kiện an ninh

Vô số các bản ghi log được gửi từ các thiết bị và ứng dụng trong môi trường đến SIEM. Tại thời điểm này, tất cả các bản ghi đang ở định dạng gốc ban đầu, do đó chúng ta không thực hiện được bất cứ điều gì ngoại trừ lưu nó vào một nơi nào đó. Nhưng để các bản ghi log hữu ích trong SIEM chúng ta cần định dạng lại chúng sang một định dạng chuẩn duy nhất. Việc thay đổi tất cả các loại bản ghi log khác nhau thành các bản ghi có cùng một định dạng duy nhất được gọi là chuẩn hóa. Việc chuẩn hóa các bản ghi log giúp cho SIEM có thể thống nhất các bản ghi log, nhanh chóng phân tích cũng như tương quan sự kiện an ninh sau nay.

Diagram

Description automatically generated

# Hình 2.2: Mô tả chuẩn hóa sự kiện

Sau quá trình chuẩn hóa các bản ghi log thì quá trình tổng hợp sự kiện an ninh diễn ra. Mục đích của quá trình này là tổng hợp các sự kiện an ninh thuộc cùng một kiểu để thấy được sự tổng thể của hệ thống. Điều này có vẻ tương tự như với quá trình tương quan sự kiện an ninh nhưng thực sự không phải vậy. Tương quan sự kiện an ninh thì sẽ tổng hợp nhiều sự kiện an ninh khác nhau để đưa ra kết luận có hay không về một cuộc tấn công.

##### 2.2.4. Tương quan sự kiện an ninh

Quá trình tương quan sự kiện an ninh là từ các bản ghi sự kiện an ninh khác nhau được liên kết lại với nhau nhằm đưa ra kết luận có hay không một tấn công vào hệ thống. Quá trình đòi hỏi việc xử lý tập trung và chuyên sâu vì chúng phải hiểu được một tấn công diễn ra như thế nào. Mà thông thường sẽ sử dụng các thông tin dữ liệu trong cơ sở dữ liệu sẵn có và liên kết với các thông tin về bối cảnh trong môi trường mạng của hệ thống. Các thông tin này có thể như các thư mục người dùng, các thiết bị và vị trí của chúng.

Như chúng ta đã biết, có nhiều thiết bị IDS và IPS là những thiết bị phát hiện và ngăn chặn tấn công phổ biến đang được dùng trong hiện tại. Nhưng chúng đều là hoạt động một cách độc lập hay riêng lẻ. Một câu hỏi đặt ra trong nhu cầu tất yếu của sự phát triển liệu có một giải pháp nào giải quyết được những hạn chế trên. Nếu duy nhất một tập tin có thể chứa tất cả các thông tin cần thiết để phân tích an ninh, chúng ta sẽ không cần phải bận tâm với việc thu thập và liên kết các sự kiện từ nhiều nguồn khác nhau. Trong đó mỗi bản ghi log hoặc sự kiện có chứa một mẩu thông tin. Để đưa ra quyết định có chính xác về những gì đang xảy ra với các ứng dụng và trong hệ thống, chúng ta cần phải kết hợp các sự kiện từ nhiều nguồn khác nhau. Các sự kiện chúng ta muốn và phần nào của dữ liệu từ những sự kiện chúng ta cần là những gì mà SIEM mang đến cho chúng ta.

Tương quan là một tập hợp các quy tắc. Tương quan sự kiện an ninh giúp liên kết các sự kiện an ninh từ các nguồn khác nhau thành một sự kiện an ninh chính xác. Tương quan các sự kiện an ninh được thực hiện nhằm đơn giản hóa các thủ tục ứng phó sự cố cho cho hệ thống, bằng việc thể hiện một sự cố duy nhất được liên hệ từ nhiều sự kiện an ninh đến từ các thiết bị nguồn khác nhau. Nếu nhìn vào bảng các sự kiện an ninh theo thời gian thực của SIEM, nó cho thấy nhiều sự kiện an ninh đăng nhập thất bại trong khoảng thời gian 10 giây từ những địa chỉ khác nhau đến một số địa chỉ đích. Có thể thấy một một địa chỉ nguồn duy nhất đăng nhập thất bại vào địa chỉ đích nhiều lần. Điều này có thể là một cuộc tấn công brute-force với máy chủ. Nhưng trừ khi chúng ta có một trí nhớ rất tốt mới nhận ra điều đó. Thường là chúng ta có thể đã quên các sự kiện an ninh đầu tiên xảy ra.

Table

Description automatically generated

# Hình 2.3: Sự kiện an ninh theo thời gian thực

Mở rộng theo ví dụ này và thay vì chỉ có vài sự kiện an ninh trong một khoảng thời gian 10 giây, chúng ta có 1000 sự kiện an ninh trong 10 giây. Hãy chọn ra những sự kiện an ninh từ tất cả sự kiện trong hệ thống để đánh giá là sự kiện an ninh nguy hại là điều cực kỳ khó khăn. Chúng ta cần một cách để loại bỏ tất cả các thông tin sự kiện an ninh không liên quan trong các bản ghi log và chỉ cần theo dõi các thông tin sự kiện an ninh có thể chỉ ra một nguy hại qua nhiều sự kiện an ninh.

Vậy tương quan là việc liên kết nhiều sự kiện lại với nhau để phát hiện hành vi lạ. Nó là sự kết hợp của các sự kiện khác nhau nhưng liên quan đến một sự cố duy nhất trong hệ thống. Thông thường có hai kiểu tương quan. Một là dựa trên các quy tắc kiến thức đã biết (Rule - based) và hai là dựa trên phương pháp thống kê

(statistical-based).

Rule - based: Đây là phương pháp dựa trên hệ chuyên gia. Tương quan sự kiện dựa trên các quy tắc và kiến thức đã biết về các cuộc tấn công. Các kiến thức đã biết về các cuộc tấn công được sử dụng để liên kết các sự kiện lại với nhau và phân tích chúng trong một bối cảnh chung.

Các quy tắc được xây dựng vào các mẫu xác định và do các nhà cung cấp phát triển hoặc chúng ta có thể tự xây dựng, phát triển và bổ sung vào hệ thống theo thời gian và kinh nghiệm tích lũy.

Ví dụ: Nếu một quản trị viên dùng các quy tắc theo dõi giám sát xem có hành động quét các cổng trên các thiết bị. Nếu thấy rằng có hành động quét cổng để cố gắng telnet vào đó, các quy tắc sau đó tiếp tục theo dõi xem khoảng thời gian trước đó đã xảy ra hay chưa. Nếu có một kết nối telnet được xác định từ một nguồn IP không rõ thì hệ thống tương quan sự kiện này sẽ gửi cảnh báo qua giao diện điều khiển hoặc có thể nhắn tin, gửi mail đến cho các quản trị viên được biết.

Statistical - based: Phương thức tương quan không sử dụng bất kỳ kiến thức của các hoạt động được cho là nguy hiểm đã biết trước đó. Nhưng thay vì dựa vào những kiến thức của các hoạt động bình thường đã được công nhận và tích lũy theo thời gian. Các sự kiện đang diễn ra được đánh giá bởi một thuật toán và có thể được so sánh với mẫu bình thường để phân biệt hành vi bình thường và hành vi bất thường. Hệ thống phân tích các sự kiện an ninh trong một khoảng thời gian và sử dụng trọng số để đánh giá tài sản, hệ thống. Các giá trị này sau đó được phân tích để xác định nguy cơ kiểu tấn công này xảy ra. Các hệ thống này cũng thiết lập mức độ hoạt động mạng bình thường và tìm kiếm sai lệch so với những mẫu có hành vi bình thường có thể chỉ ra một cuộc tấn công.

##### 2.2.5. Lưu trữ Log

Khi phân tích thì các dữ liệu được lưu trữ trực tuyến và khi không còn cần thiết thì chúng sẽ được chuyển tới nơi khác để lưu trữ dài hạn. Dữ liệu có thể được lưu trữ dưới dạng đã chuẩn hóa nhằm đẩy nhanh tốc độ tìm kiếm sử dụng sau này. Bên cạnh đó chúng cũng được lưu trữ dưới dạng gốc ban đầu nhằm phục vụ nhu cầu bằng chứng và điều tra số sau này. Thông thường chúng được lưu trữ dưới dạng nén và có thể được mã hóa. SIEM cung cấp khả năng lưu trữ đến hàng trăm triệu sự kiện an ninh khác nhau.

Có ba cách mà có thể lưu trữ các bản ghi trong SIEM là: Dùng một cơ sở dữ liệu, file Text và dưới dạng file nhị phân.

* Lưu trữ dưới dạng cơ sở dữ liệu

Lưu trữ các bản ghi log trong cơ sở dữ liệu là cách lưu trữ các bản ghi log hay được dùng nhất trong SIEM. Cơ sở dữ liệu thường là một nền tảng cơ sở dữ liệu chuẩn như Oracle, MySQL, Microsoft SQL hoặc một trong các ứng dụng cơ sở dữ liệu lớn khác đang được sử dụng trong doanh nghiệp. Phương pháp này cho phép tương tác khá dễ dàng với dữ liệu vì các truy vấn cơ sở dữ liệu là một phần của ứng dụng cơ sở dữ liệu. Hiệu suất cũng khá tốt khi truy cập vào các bản ghi log trong cơ sở dữ liệu, phụ thuộc vào phần cứng cơ sở dữ liệu đang chạy, nhưng các ứng dụng cơ sở dữ liệu phải được tối ưu hóa để chạy với SIEM. Sử dụng cơ sở dữ liệu là một giải pháp tốt cho việc lưu trữ nhật ký.

* Lưu trữ dưới dạng file Text

Một tập tin văn bản chuẩn để lưu trữ các thông tin trong một định dạng có thể đọc được. Các thông tin cần phải có một ranh giới phân cách có thể là dấu phẩy, tab hoặc một số kí hiệu khác. Vì vậy thông tin có thể được phân tích và đọc đúng. Phương pháp lưu trữ này không được sử dụng thường xuyên. Hành động viết và đọc từ tập tin văn bản dường như chậm hơn so với các phương pháp khác. Thật sự không có nhiều lợi ích khi sử dụng một tập tin text để lưu trữ dữ liệu, nhưng nó dễ dàng cho các ứng dụng bên ngoài để truy cập dữ liệu này. Nếu các bản ghi log được lưu trữ trong một tập tin văn bản, thì sẽ không khó khăn khi một viết mã của riêng để mở các tập tin và lấy thông tin để cung cấp cho cho một ứng dụng khác. Một lợi ích khác là khi tập tin văn bản con người có thể đọc được và dễ dàng để nhà phân tích tìm kiếm và hiểu nó. Chúng ta có thể mở các tập tin và sử dụng lệnh grep hoặc một số công cụ tìm kiếm tập tin văn bản khác để tìm ra thông tin tìm kiếm mà không cần mở một giao diện điều khiển.

* Lưu trữ dưới dạng file nhị phân

Định dạng tập tin nhị phân là sử dụng một tập tin với định dạng tùy chỉnh để lưu trữ thông tin duới dạng nhị phân. SIEM biết làm thế nào để đọc và ghi vào những file này.

##### 2.2.6. Giám sát và cảnh báo

Khi đã có tất cả các bản ghi log trong SIEM và các sự kiện an ninh đã được xử lý, điều cần làm tiếp theo như thế nào để sử dụng hữu ích với các thông tin từ các bản ghi log khác nhau. SIEM có một giao diện điều khiển dựa trên web hoặc ứng dụng tải về máy trạm. Cả hai giao diện sẽ cho phép tương tác với các dữ liệu được lưu trữ trong SIEM. Giao diện điều khiển này cũng được sử dụng để quản lý SIEM. Giao diện ứng dụng này cho phép xử lý sự cố hoặc cung cấp cái nhìn tổng quan về môi trường của chúng ta. Bình thường khi muốn xem các thông tin hoặc xử lý sự cố các kỹ sư sẽ phải đi đến các thiết bị khác nhau và xem các bản ghi log trong định dạng gốc của nó. Nhưng với SIEM sẽ đơn giản và tiện lợi hơn nhiều. Nó có thể xử lý tại một nơi duy nhất, phân tích tất cả các bản ghi log khác nhau dễ dàng bởi vì SIEM đã chuẩn hóa các thông tin dữ liệu đó. Trong quản lý và giám sát giao diện điều khiển của SIEM, chúng ta có thể phát triển nội dung và quy định được sử dụng để tìm ra thông tin từ các sự kiện an ninh được xử lý. Giao diện điều khiển này là một cách để giao tiếp với các dữ liệu được lưu trữ trong SIEM.

SIEM cung cấp ba cách để thông báo tới các quản trị viên khi có một cuộc tấn công hay một hành vi bất thường đang xảy ra. Thứ nhất, SIEM có thể đưa ra một cảnh báo ngay khi chúng nhận ra rằng có điều gì bất thường. Thứ hai, SIEM sẽ gửi một thông báo vào một thời điểm được xác định trước của cuộc tấn công và thứ ba là các quản trị viên theo dõi giám sát SIEM theo thời gian thực thông qua một giao diện web. Các IDS thông thường đưa ra nhiều cảnh báo giả nhưng với SIEM nó tạo ra một tỷ lệ nhỏ các thông báo giả như vậy. Tuy nhiên tất cả những thông báo có thể là cần thiết để thực hiện một hành động hay đơn giản là bỏ qua nó còn tùy thuộc vào mức độ của sự kiện an ninh.

Một số sản phẩm của SIEM có thể thực hiện các hành động ứng phó như xóa các phần mềm độc hại, đóng một số cổng nào đó thông qua việc kết nối tới các thiết bị đó. Báo cáo được lập lịch để đưa ra các báo cáo thường xuyên. Các báo này được thể hiện theo chuẩn quốc tế và có thể thể hiện qua những biểu đồ trực quan về những số liệu. Những báo cáo này cung cấp nhanh chóng cái nhìn tổng quan cho các quản trị viên và nhà quản lý.

##### 2.3. Một số hệ thống triển khai Siem

###### 2.3.1. MARS

Hệ thống theo dõi, giám sát và ứng phó (MARS - Monitoring Analysis and

Response System) là một sản phẩm triển khai SIEM được sản xuất bởi Cisco. MARS là một sản phẩm thương mại rất được ưa chuộng trong việc triển khai SIEM. Do công ty chiếm lĩnh đa phần thị trường thiết bị mạng nên việc áp dụng rộng dãi và tác động của nó cũng trở lên dễ dàng hơn. Hiện nay trên một số sản phẩm phần cứng của hãng đã tích hợp công cụ này.

Khi triển khai một cách chính xác thì MARS có thể:

* Xác định một cuộc tấn công nào được tiến hành.
* Hiện thị các thông tin chi tiết và mạng hay đường dẫn liên quan đến sự việc.
* Xác định các thiết bị có thể sử dụng để ngăn chặn các cuộc tấn công.
* Trong nhiều trường hợp nó có thể cung cấp các lệnh cụ thể áp dụng cho các thiết bị để ngăn chặn các cuộc tấn công.

Giống như nhiều sản phẩm SIEM khác, MARS cũng cung cấp chức năng phục vụ điều tra số và báo cáo. Kiểm toán các chính sách về hệ thống mạng, báo cáo hiện trạng mạng, việc sử dụng các thiết bị và xác định định hướng quy hoạch phạm vi của SIEM.

MARS vẫn cung cấp khả năng thu thập thông tin sự kiện an ninh từ các thiết bị và ứng dụng qua Syslog hoặc giao thức SNMP.

###### 2.3.2. IBM Qradar

Hệ thống này bao gồm tất cả mọi thứ cần thiết để có được những thông tin về mức độ bảo mật của và các dịch vụ quản lý sự kiện an ninh. Hệ thống QRadar đáp ứng yêu cầu tuân thủ để lưu trữ sự kiện an ninh, giám sát, báo cáo và bao gồm các chức năng sau đây để thực hiện yêu cầu an ninh bảo mật của tổ chức:

* Khả năng hiển thị theo thời gian thực
* Giảm cảnh báo giả vào ưu tiên các cảnh báo chính xác
* Quản lý các mối đe dọa một cách hiệu quả
* Cung cấp thông tin bảo mật trong môi trường điện toán đám mây
* Báo cáo các hoạt động và truy cập dữ liệu chi tiết
* Cung cấp giao diện theo dõi, quản trị đầy đủ

Các dòng sản phẩm gồm: QRadar 21xx có giá vừa phải dành cho các doanh nghiệp vừa và nhỏ. Dòng sản phẩm QRadar 31xx dành cho các doanh nghiệp quy mô lớn hơn. Trong triển khai các thiết bị giám sát, lưu trữ trên toàn mạng tạo thành một hệ thống phân cấp và cung cấp nguồn dữ liệu vào Management Console. QRadar là một trong những sản phẩm đơn giản nhất triển khai SIEM. Nó dễ dàng được thu nhỏ và có nhiều tùy chỉnh mở rộng nhằm giảm thiểu các thông báo giả, lọc các luồng và phân tích sự kiện an ninh [14].

###### 2.3.3. Splunk

Splunk là một phần mềm giám sát mạng dựa trên sức mạnh của việc phân tích Log.

Splunk thực hiện các công việc tìm kiếm, giám sát và phân tích dữ liệu logs lớn được sinh ra từ các ứng dụng, các hệ thống và các thiết bị hạ tầng mạng. Nó có thể thao tác tốt với nhiều loại dịnh dạng dữ liệu khác nhau (Syslog, csv, apache-log, access\_combine …). Splunk được xây dựng dựa trên nền tảng Lucene and MongoDB với một giao diện web rất trực quan [13].

Graphical user interface, chart, bar chart

Description automatically generated

# Hình 2.4: Giao diện Web của Splunk

##### 2.3.4. AlienVault OSSIM

OSSIM (Open Source Security Information Management) là một giải pháp SIEM và OSSIM là hệ thống giám sát an ninh mạng dựa trên nền tảng mã nguồn mở.

OSSIM thực hiện việc thu thập các sự kiện từ rất nhiều các thiết bị khác nhau như Firewall, IDS/IPS, các máy chủ, máy trạm,…trong hệ thống giám sát, từ các sự kiện có thể đưa ra các cảnh báo (alert) khác nhau.

Các sự kiện thu thập được từ nhiều nguồn khác nhau sau đó sẽ được OSSIM phân tích để tìm ra mối liên hệ giữa các sự kiện khác nhau và đưa ra được những thông tin tổng hợp nhất có liên quan đến an ninh trong hệ thống.

Những thông tin OSSIM sau khi đã được phân tích, tổng hợp sẽ được đưa vào báo cáo cụ thể, nó là kết quả đánh giá về mức độ an ninh trong hệ thống được giám sát [11].

Chart

Description automatically generated with low confidence

# Hình 2.5: Báo cáo của AlienVault OSSIM