

HỌC VIỆN KỸ THUẬT MẬT MÃ
KHOA AN TOÀN THÔNG TIN



BÁO CÁO CHUYÊN ĐỀ CƠ SỞ
NGHIÊN CỨU VÀ TRIỂN KHAI GIẢI PHÁP SIEM
DỰA TRÊN SPLUNK

Nhóm sinh viên thực hiện:

Lê Minh Châu – AT180207 – L02

Bùi Ngọc Mai – AT180232 – L02

Trần Minh Tuấn – AT180249 – L02

Người hướng dẫn :

ThS. Thái Thị Thanh Vân

Khoa Công nghệ thông tin – Học viện Kỹ thuật mật mã

Hà Nội, 2024

LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em xin chân thành gửi lời cảm ơn đến cô Thái Thị Thanh Vân – giảng viên khoa Công nghệ thông tin trường Học viện Kỹ thuật Mật Mã Trong thời gian tham gia lớp học phần, cô đã tạo điều kiện học tập, cung cấp đầy đủ các kiến thức cũng như học liệu của môn học, giúp chúng em nắm vững được nền tảng cơ bản và mục tiêu của học phần. Đồng thời, cô cũng đã tận tình hướng dẫn, giúp đỡ, đưa ra những lời khuyên cần thiết để chúng em có thể hoàn thành đề tài này.

Chúng em xin chân thành cảm ơn!

Hà Nội, tháng 6 năm 2024

Nhóm sinh viên thực hiện

Lê Minh Châu

Bùi Ngọc Mai

Trần Minh Tuấn

MỤC LỤC

MỤC LỤC	i
DANH MỤC CHỮ VIẾT TẮT	iii
DANH MỤC BẢNG.....	v
DANH MỤC HÌNH ẢNH	vi
LỜI NÓI ĐẦU	viii
1.1. Tính cấp thiết của đề tài	viii
1.2. Mục tiêu thực hiện đề tài	viii
CHƯƠNG 1. TỔNG QUAN VỀ SIEM	1
1.1. Tổng quan về tình hình an ninh mạng	1
1.1.1. Xu hướng tình hình an ninh mạng	1
1.1.2. Nhu cầu xây dựng hệ thống giám sát an toàn mạng	3
1.2. Giới thiệu về hệ thống SIEM.....	4
1.2.1. Khái niệm về hệ thống SIEM.....	4
1.2.2. Các chức năng cơ bản của hệ thống SIEM	5
1.2.3. Mô hình kiến trúc của hệ thống SIEM.....	8
1.2.4. Lợi ích của hệ thống SIEM	15
1.3. Một số giải pháp SIEM	16
1.3.1. Splunk	16
1.3.2. ELK Stack	17
1.3.3. Wazuh	17
1.3.4. IBM QRadar.....	18
1.4. Kết luận Chương 1	19
CHƯƠNG 2. TÌM HIỂU CÔNG CỤ GIÁM SÁT AN NINH MẠNG SPLUNK	20
2.1. Tổng quan về Splunk.	20
2.2. Ứng dụng của Splunk	23
2.2.1. Quản lý ứng dụng.....	23
2.2.2. Quản lý hoạt động IT	24
2.2.3. An ninh trong lĩnh vực IT	25
2.3. Các thành phần của Splunk	26
2.3.1. Splunk Forwarder.....	26
2.3.2. Splunk Indexer	29
2.3.3. Splunk Search Head	35
2.4. Kiến trúc Splunk	42
2.5. Kết luận Chương 2	43
CHƯƠNG 3. TRIỂN KHAI THỰC NGHIỆM	44
3.1. Mục tiêu thực nghiệm	44
3.2. Xây dựng môi trường.....	44
3.2.1. Mô hình thực nghiệm.....	44

3.2.2. Giới thiệu công nghệ hỗ trợ	44
3.3. Triển khai kịch bản thử nghiệm và đánh giá	48
3.3.1. Phát hiện tấn công brute force	48
3.3.2. Phát hiện scan port	48
3.3.3. Phát hiện malware	48
3.3.4. Đánh giá	48
3.4. Kết luận Chương 3	49
KẾT LUẬN.....	50
TÀI LIỆU THAM KHẢO	51
PHỤ LỤC.....	52
BẢNG PHÂN CÔNG.....	62

DANH MỤC CHỮ VIẾT TẮT

Chữ viết tắt	Giải nghĩa
API	Application Programming Interface
BGP	Border Gateway Protocol
CLI	Command Line Interface
CPU	Central Processing Unit
DBA	Database Administrator
DDoS	Distributed Denial-of-Service
DHCP	Dynamic Host Configuration Protocol
DLL	Dynamic Link Library
DNS	Domain Name System
DoS	Denial of Service
EDR	Endpoint Detection and Response
ELK	Elasticsearch, Logstash and Kibana
FTP	File Transfer Protocol
GUI	Graphical User Interface
GUID	Globally Unique Identifier
HF	Heavy forwarder
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
OSPF	Open Shortest Path First
PPTP	Point-to-Point Tunneling Protocol
SEM	Security Event Management

SIEM	Security Information and Event Management
SIM	Security Information Management
SPL	Search Processing Language
UDP	User Datagram Protocol
UF	Universal Forwarder
WAF	Web Application Firewall

DANH MỤC BẢNG

Bảng 2.1. Một số lệnh Splunk Search Language	39
Bảng 2.2. Cảnh báo theo lịch trình và theo thời gian thực trong Splunk.....	40
Bảng 3.1. Cấu hình địa chỉ IP cho từng máy	Error! Bookmark not defined.

DANH MỤC HÌNH ẢNH

Hình 1.1. Chi phí ước tính của tội phạm mạng trên toàn thế giới	2
Hình 1.2. Mô hình kiến trúc SIEM	8
Hình 1.3. Chuẩn hóa dữ liệu	12
Hình 2.1. Mô tả Heavy forwarder trong Splunk	27
Hình 2.2. Mô hình thu thập log tập trung.....	28
Hình 2.3. Mô hình thu thập log cân bằng tải.....	28
Hình 2.4. Mô hình thu thập log định tuyến và lọc	29
Hình 2.5. Sơ đồ minh họa quy trình hoạt động của lập chỉ mục	30
Hình 2.6. Index bucket trong Splunk	35
Hình 2.7. Minh họa cú pháp câu lệnh search	39
Hình 2.8. Sơ đồ minh họa kiến trúc Splunk.....	42
Hình 3.1. Mô hình thực nghiệm	44
Hình 3.2. Vị trí lưu trữ logs Sysmon.....	46
Hình 3.3. Minh họa cấu trúc và thành phần cơ bản của file cấu hình Sysmon...	47
Hình 3.4. Giao diện Search của Splunk	Error! Bookmark not defined.
Hình 3.5. Giao diện cấu hình alert	Error! Bookmark not defined.
Hình 3.6. Cấu hình thông báo qua Splunk	Error! Bookmark not defined.
Hình 3.7. Cấu hình run a script	Error! Bookmark not defined.
Hình 3.8. Giao diện chức năng Visualization	Error! Bookmark not defined.
Hình 3.9. Cấu hình tạo dashboard.....	Error! Bookmark not defined.
Hình 3.10. Brute force với công cụ crowbar.....	Error! Bookmark not defined.
Hình 3.11. Kết quả được thông báo qua ứng dụng telegram	Error! Bookmark not defined.
Hình 3.12. Giao diện màn hình chính dashboard	Error! Bookmark not defined.
Hình 3.13. Một rule mới đã được thêm.....	Error! Bookmark not defined.
Hình 3.14. Kiểm tra kết nối tới Win 10	Error! Bookmark not defined.
Hình 3.15. Giao diện setting System logs.....	Error! Bookmark not defined.
Hình 3.16. Cấu hình nhận logs Pfsense	Error! Bookmark not defined.
Hình 3.17. Cấu hình nhận logs Pfsense	Error! Bookmark not defined.

Hình 3.18. Giao diện cấu hình alert	Error! Bookmark not defined.
Hình 3.19. Giao diện cấu hình gửi thông báo tới Telegram	Error! Bookmark not defined.
Hình 3.20. Giao diện cấu hình run a script	Error! Bookmark not defined.
Hình 3.21. Giao diện chức năng Visualization ...	Error! Bookmark not defined.
Hình 3.22. Giao diện cấu hình tạo dashboard.....	Error! Bookmark not defined.
Hình 3.23. Quá trình thực hiện quét cổng bằng nmap	Error! Bookmark not defined.
Hình 3.24. Kết quả thông báo qua Telegram	Error! Bookmark not defined.
Hình 3.25. Giao diện màn hình chính dashboards	Error! Bookmark not defined.
Hình 3.26. Rule mới đã được thêm	Error! Bookmark not defined.
Hình 3.27. Kết quả kiểm tra kết nối	Error! Bookmark not defined.
Hình 3.28. Giao diện công cụ Shellter	Error! Bookmark not defined.
Hình 3.29. Quá trình tiêm mã độc vào file.....	Error! Bookmark not defined.
Hình 3.30. Thông báo hoàn thành quá trình tiêm mã độc	Error! Bookmark not defined.
Hình 3.31. Tích hợp công cụ Virustotal vào splunk	Error! Bookmark not defined.
Hình 3.32. Kết quả của việc tích hợp Virustotal.	Error! Bookmark not defined.
Hình 3.33. Kết quả kiểm tra tại trang web	Error! Bookmark not defined.
Hình 3.34. Giao diện cấu hình alert	Error! Bookmark not defined.
Hình 3.35. Giao diện cấu hình thông báo tới Telegram	Error! Bookmark not defined.
Hình 3.36. Giao diện công cụ Metasploit	Error! Bookmark not defined.
Hình 3.37. Cấu hình phiên khai thác.....	Error! Bookmark not defined.
Hình 3.38. Máy mục tiêu chạy file malware.....	Error! Bookmark not defined.
Hình 3.39. Kết quả khi máy mục tiêu chạy file malware	Error! Bookmark not defined.
Hình 3.40. Kết quả thông báo qua Telegram	Error! Bookmark not defined.
Hình 3.41. Giao diện màn hình chính Vt4Splunk	Error! Bookmark not defined.

LỜI NÓI ĐẦU

1.1. Tính cấp thiết của đề tài

Với sự phát triển mạnh mẽ của Internet và World Wide Web đã đặt ra nhiệm vụ đảm bảo an toàn thông tin cho các hệ thống mạng của các cơ quan, tổ chức nhằm tránh khỏi những hiểm họa mất an toàn thông tin trước những tấn công mạng có thể xảy ra. Khi đó cần có một hệ thống giám sát an ninh mạng đủ mạnh nhằm kiểm soát, thu thập toàn bộ lưu lượng dữ liệu vào ra cho cả một hệ thống mạng và đưa ra những cảnh báo chính xác tới người quản trị hệ thống khi có tấn công xảy ra. Việc giám sát an ninh mạng hiện nay đã được các quốc gia trên thế giới vô cùng quan tâm và nó có vai trò sống còn cho an ninh quốc gia. Tại Việt Nam, trong những năm gần đây giám sát an ninh mạng cũng được xem là một nhiệm vụ trọng yếu được các cơ quan cấp bộ, ban, ngành vô cùng quan tâm và thực hiện công việc này một cách tích cực. Một hệ thống giám sát an ninh mạng tốt cần phải thu thập được tất cả các nhật ký vào ra của hệ thống, sau đó thực hiện phân tích những dữ liệu này, và dựa trên những dấu hiệu hoặc tập luật sẵn có để đưa ra cảnh báo tới người quản trị hệ thống.

Trên thực tế có nhiều giải pháp giám sát an ninh mạng hiện nay được sử dụng. Nhiều sản phẩm thương mại xuất hiện trên thị trường, mỗi sản phẩm có các ưu điểm và yếu điểm khác nhau. Nhưng nhìn chung, các sản phẩm hiện nay được sử dụng dựa trên công nghệ SIEM là chủ yếu. Vì vậy, chúng em quyết định chọn đề tài **“NGHIÊN CỨU VÀ TRIỂN KHAI GIẢI PHÁP SIEM DỰA TRÊN SPLUNK”**.

1.2. Mục tiêu thực hiện đề tài

Các mục tiêu bao gồm:

- Tìm hiểu về cơ chế hoạt động, chức năng cơ bản và ứng dụng của hệ thống giám sát an toàn mạng và quản lý sự kiện SIEM.
- Tìm hiểu về công cụ Splunk dùng để triển khai SIEM.
- Thực nghiệm triển khai hệ thống SIEM với mô hình mạng nhỏ, thực hiện một số tính năng nổi bật của SIEM.

CHƯƠNG 1. TỔNG QUAN VỀ SIEM

1.1. Tổng quan về tình hình an ninh mạng

1.1.1. Xu hướng tình hình an ninh mạng

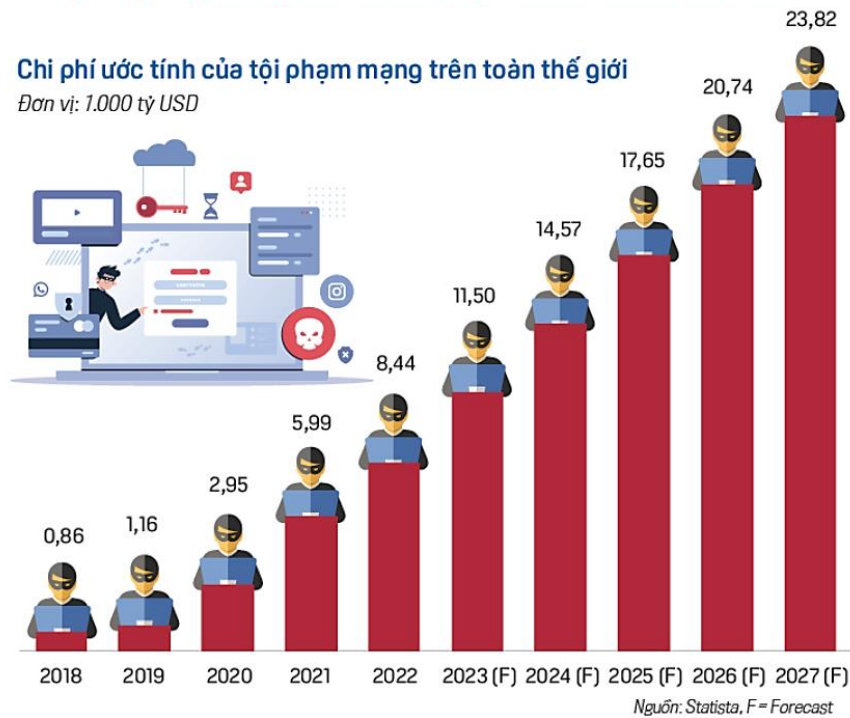
Mọi công ty đều cần có an ninh mạng nhằm đảm bảo và nâng cao sự thành công của quá trình chuyển đổi kỹ thuật số. Được thúc đẩy bởi sự gia tăng nhận thức về các mối đe dọa và rủi ro từ dữ liệu, thị trường an ninh mạng toàn cầu đã phát triển vô cùng mạnh mẽ trong những năm vừa qua. Nhu cầu duy trì an ninh mạng dự kiến sẽ tăng lên cùng với sự thâm nhập ngày càng sâu của Internet ở khắp nơi trên thế giới. An ninh mạng đã ngày càng trở thành một phần chi phối trong hoạch định chiến lược cấp cao nhất của các quốc gia.

Càng có nhiều người tham gia hoạt động trực tuyến, cho dù vì công việc hay cuộc sống cá nhân, thì tội phạm mạng càng có nhiều cơ hội để khai thác. Đồng thời, các kỹ thuật của kẻ tấn công trên môi trường online cũng trở nên tiên tiến hơn với sự ra đời của nhiều công cụ để trợ giúp những kẻ lừa đảo. Những năm gần đây, nhiều tổ chức phải đối mặt với số lượng các cuộc tấn công mạng tăng lên đáng kể do lỗ hổng bảo mật của công việc từ xa cũng như việc chuyển sang môi trường số hóa, chẳng hạn như cơ sở hạ tầng, dữ liệu và mạng lưới của điện toán đám mây. Theo ước tính của Statista (công ty cung cấp báo cáo thị trường uy tín của Đức), chi phí toàn cầu cho loại hình tội phạm mạng dự kiến sẽ tăng nhanh trong 5 năm tới, từ 8,44 nghìn tỷ USD vào năm 2022 lên 23,82 nghìn tỷ USD vào năm 2027. Bên cạnh đó, thị trường an ninh mạng toàn cầu được định giá gần 222 tỷ USD vào năm 2022. Đến năm 2030, thị trường được dự báo sẽ vượt 657 tỷ USD.

TỘI PHẠM MẠNG DỰ KIẾN SẼ TĂNG VỌT TRONG NHỮNG NĂM TỚI

Chi phí ước tính của tội phạm mạng trên toàn thế giới

Đơn vị: 1.000 tỷ USD



Hình 1.1. Chi phí ước tính của tội phạm mạng trên toàn thế giới

Sự ra đời của AI tạo ra những thách thức và cơ hội mới trong cả việc tấn công và bảo vệ tài sản. Khi AI bị những kẻ tấn công khai thác, các cuộc tấn công sẽ trở nên nhanh hơn, chính xác hơn và mở rộng hơn. Nghiên cứu X-Force năm 2024 cho thấy AI cũng trở thành công cụ hữu hiệu của tội phạm mạng, được sử dụng ngày càng nhiều trong các cuộc tấn công.

Tại Việt Nam, theo Công ty Công nghệ An ninh mạng Quốc Gia Việt Nam NCS cho biết: 13.900 vụ tấn công an ninh mạng vào các hệ thống năm 2023. Trong đó có tới 554 website của các cơ quan, tổ chức chính phủ và giáo dục có tên miền .gov.vn, .edu.vn bị xâm nhập, chèn mã quảng cáo cờ bạc, cá độ. Hơn 83.000 máy tính, máy chủ bị mã độc mã hóa dữ liệu tổng tiền tấn công. Tình trạng lộ dữ liệu cá nhân ở mức báo động, kèm theo hàng loạt hình thức lừa đảo trực tuyến liên tục xảy ra. Trong đó, tấn công mã hóa dữ liệu ransomware gây hậu quả nghiêm trọng. Không chỉ mã hóa dữ liệu nhằm đòi nạn nhân trả tiền chuộc, tin tặc có thể bán dữ liệu cho bên thứ ba để tối đa số tiền thu được. Theo tổng hợp của NCS, tỷ lệ máy tính tại Việt Nam bị mã độc tấn công trong năm 2023 là

43,6%. Trong 3 tháng đầu năm 2024, số sự cố tấn công mạng vào các hệ thống thông tin tại Việt Nam là 2.323 cuộc.

Sự phát triển nhanh chóng của AI không chỉ mang lại những lợi ích rõ ràng mà còn tạo ra những nguy cơ cho an ninh mạng. Thách thức lớn nhất đối diện với công nghệ AI ngày nay là lừa đảo và tấn công có chủ đích APT, với mức độ ngày càng phức tạp của các kịch bản lừa đảo, đặc biệt khi kết hợp giữa Deepfake và GPT. Bảo vệ dữ liệu cá nhân theo yêu cầu của Nghị định 13 là nhiệm vụ hàng đầu của các tổ chức. Chiến lược phòng thủ An ninh mạng sẽ có nhiều thay đổi, bên cạnh các kiến trúc bảo vệ nhiều lớp, ngăn chặn dựa trên các tập luật và mẫu nhận diện, các tổ chức sẽ tăng cường đầu tư vào giám sát an ninh mạng, săn tìm chủ động các nguy cơ, chấp nhận xác suất bị tấn công nhưng phát hiện sớm để khắc phục, giảm thiểu thiệt hại. Công nghệ tạo lập bẫy với dữ liệu giả (deception) để thu hút sự chú ý của tin tặc nhằm bảo vệ dữ liệu thật cũng sẽ được phổ biến trong thời gian tới.

1.1.2. Nhu cầu xây dựng hệ thống giám sát an toàn mạng

Trong an ninh mạng, phòng thủ khó khăn hơn nhiều so với việc tấn công. Trong khi các công ty cần phải tìm ra mọi lỗ hổng bảo mật để phòng tránh rủi ro, thì tội phạm mạng lại chỉ cần một lỗ hổng để tấn công hệ thống. Hơn nữa, chi phí để thực hiện một cuộc tấn công mạng đang tiếp tục giảm, vì tin tặc có khả năng tiếp cận với các công nghệ tinh vi hơn bao giờ hết. Các mối đe dọa an ninh mạng tiếp tục tăng cao và khi chúng trở nên tinh vi hơn, việc ngăn chặn chúng sẽ càng khó khăn hơn. Theo báo cáo Dự báo định hướng công nghệ 2024 - 2026 của FPT, các loại rủi ro mới nhất, đe dọa lớn nhất trong tương lai gần là các cuộc tấn công được thực hiện bởi trí tuệ nhân tạo (AI), các cuộc tấn công liên quan đến điện toán lượng tử. Với các rủi ro mới này, các khoản đầu tư cho các giải pháp an ninh mạng được dự báo tăng cao trong các năm tới. Đối với các cơ sở hạ tầng mạng, hệ thống giám sát an toàn mạng là vô cùng quan trọng, nhất là đối với các tổ chức, tập đoàn, doanh nghiệp.

Hệ thống giám sát an toàn mạng SIEM viết tắt của “Security Information and Event Management” là hệ thống được thiết kế nhằm thu thập thông tin nhật ký các sự kiện an ninh từ các thiết bị đầu cuối và lưu trữ dữ liệu một cách tập trung. Theo đó, các sản phẩm SIEM cho phép phân tích tập trung và báo cáo về các sự kiện an toàn mạng của tổ chức. Kết quả phân tích này có thể được dùng để phát hiện ra các cuộc tấn công mà không thể phát hiện được theo phương pháp thông thường. Một số sản phẩm SIEM còn có khả năng ngăn chặn các cuộc tấn công mà chúng phát hiện được.

Giám sát an toàn mạng mang lại nhiều lợi ích quan trọng, giúp các tổ chức bảo vệ hệ thống và dữ liệu của mình khỏi các mối đe dọa mạng:

- Quản lý tập trung.
- Giám sát an toàn mạng.
- Cải thiện hiệu quả trong hoạt động xử lý sự cố.

1.2. Giới thiệu về hệ thống SIEM

1.2.1. Khái niệm về hệ thống SIEM

SIEM viết tắt của Security Information and Event Management, một giải pháp bảo mật hỗ trợ trực tiếp cho tổ chức, doanh nghiệp có thể phát hiện và đưa ra những phản ứng kịp thời nhất với các mối đe dọa an ninh mạng. SIEM là sự kết hợp hài hòa giữa quản lý thông tin bảo mật (SIM - Security Information Management) và quản lý sự kiện bảo mật (SEM - Security Event Management) trong một hệ thống đồng bộ.

- SIM: thu thập các tệp nhật ký từ nhiều nơi và lưu trữ chúng trong kho lưu trữ trung tâm để dùng cho quá trình phân tích sau này. Do đó, SIM dùng để quản lý nhật ký.
- SEM: thiên về xử lý dữ liệu hơn trong thời gian thực để theo dõi mối tương quan và thông báo các sự kiện bảo mật. Nó xử lý trên những dữ liệu đã được thu thập trong SIM.

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 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. 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, Anti Virus hoặc các hệ thống phòng chống xâm nhập. Các thiết bị thu thập thông tin chuyên 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.

Việc sử dụng SIEM nhằm theo dõi, xác định, quản lý hệ thống tài sản và ứng phó với các sự cố 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 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 CNTT.

1.2.2. Các chức năng cơ bản của hệ thống SIEM

Hầu hết các giải pháp SIEM đều thực hiện một số chức năng tổng hợp, hợp nhất và sắp xếp dữ liệu ở một mức độ nào đó để xác định các mối đe dọa và tuân thủ theo các yêu cầu tuân thủ dữ liệu. Mặc dù một số giải pháp có khả năng khác nhau, song chúng đều cung cấp cùng một bộ chức năng cốt lõi:

❖ Quản lý nhật ký (Log management):

Quản lý nhật ký trong hệ thống SIEM bắt đầu với việc cấu hình các nút trong hệ thống CNTT, đặc biệt là các nút quan trọng hoặc có tính chất then chốt, để gửi các sự kiện hệ thống và ứng dụng liên quan (nhật ký) đến một cơ sở dữ liệu tập trung được quản lý bởi ứng dụng SIEM. Log này được chuẩn hóa về một định dạng duy nhất để thực hiện bước tiếp theo. Ứng dụng cơ sở dữ liệu SIEM này đầu tiên sẽ phân tích và chuẩn hóa dữ liệu được gửi bởi nhiều loại nút rất khác nhau trong hệ thống CNTT. Sau đó, SIEM thường cung cấp các dịch vụ lưu trữ, tổ chức, truy xuất và lưu trữ nhật ký để đáp ứng các yêu cầu quản lý nhật ký mà

doanh nghiệp có thể cần. Luồng dữ liệu vào thành phần quản lý nhật ký của hệ thống SIEM cũng cho phép sử dụng thêm phân tích gần như thời gian thực và khai thác dữ liệu về tình trạng hoạt động và an ninh của tất cả các hệ thống CNTT đang gửi dữ liệu vào hệ thống SIEM.

❖ **Tuân thủ các quy định trong CTTT (IT Regulatory Compliance):**

Tất cả các sự kiện từ các hệ thống quan trọng đang được sử dụng truy nhập, 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. Hoặc 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.

❖ **Tương quan liên kết các sự kiện an ninh (Event Correlation):**

Sự tương quan liên kết giữa các sự kiện an ninh mang đem lại thông báo tốt hơn cho hệ thống. 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, 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 một 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 cái nhìn 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%? 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? Từ các nguồn khác nhau? Có thể là một từ chối dịch vụ phân tán (DDoS)?

Đó là sự tương quan các sự kiện an ninh. Cảnh báo của SIEM giúp người giám sát đưa ra cách ứng phó tùy thuộc vào các điều kiện.

❖ **Phản hồi tự động:**

Bây giờ quản trị viên đã có tất cả các hệ thống phù hợp gửi sự kiện vào SIEM, các quy tắc và bộ lọc đã được xác định, và các quy tắc tương quan đã được thiết lập. SIEM có thể được cấu hình để tự động phản ứng với các sự kiện tương quan cụ thể. SIEM có thể tự động thực hiện các hành động khắc phục đối với các mối đe dọa hoặc cấu hình sai, giúp giảm bớt khối lượng công việc.

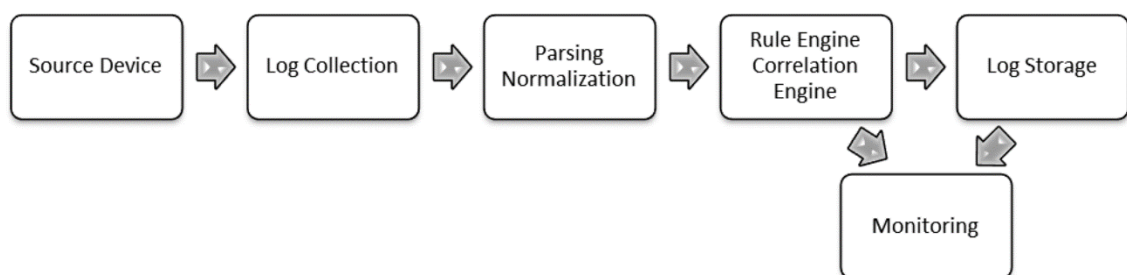
Lợi ích việc thực hiện các hoạt động ứng phó là rất tốt, nhưng bên cạnh đó nó cũng có điều bất lợi. Nếu không cấu hình cẩn thận và chính xác thì nó có thể đưa ra các hành động ứng phó không cần thiết.

❖ **Đảm bảo an ninh thiết bị đầu cuối:**

Hầu hết các hệ 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.

1.2.3. Mô hình kiến trúc của hệ thống 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 cùng một lúc thì sẽ không có một 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 như sau.



Hình 1.2. Mô hình kiến trúc SIEM

❖ 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 bất kỳ dữ liệu nào khác. Việc biết về những gì có trong hệ thống là rất quan trọng trong việc triển khai SIEM. Hiểu rõ những nguồn mà muốn lấy các bản ghi log trong giai đoạn đầu sẽ 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 về cơ bản công nghệ khác nhau và thực hiện chuyên 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 người dùng đã làm gì: Ai là người đăng nhập, làm những gì trên hệ thống?...

- **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 được cấu hình theo cách mà hệ điều hành thông thường. Một ví dụ như một router hoặc switch. Nó không phụ thuộc vào nhà cung cấp, người dùng không bao giờ có thể truy cập trực tiếp vào hệ thống điều hành cơ bản của nó mà chỉ có thể 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, 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, như thống kê, sai sót, hoặc thông tin tin nhắn.
- **Xác định bản ghi log cần thiết:** Sau khi xác định các thiết bị nguồn trong hệ thống, 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 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 đều có ích trong việc xác định nguồn thiết bị cần thiết cho SIEM.

❖ Thu thập Log:

Bước tiếp theo trong sơ đồ là làm thế nào để thu thập các bản ghi log từ các thiết bị khác nhau. Cơ chế thu thập các bản ghi log phụ thuộc vào từng thiết bị nhưng cơ bản nhất có hai phương thức như sau: Pull log và Push log.

- 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, 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, 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 này 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, người giám sát 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 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 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 có thể cấu hình theo tùy chọn hoặc để cấu hình mặc định cho SIEM.

❖ Phân tích, chuẩn hóa Log:

Tại thời điểm này, tất cả các bản ghi đang ở định dạng gốc ban đầu, do đó 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 cần định dạng lại chúng sang một định dạng chuẩn duy nhất.

- Parsing – phân tích cú pháp

Thành phần phần mềm có thể từ định dạng Log cụ thể chuyển nó về dữ liệu có cấu trúc. Ví dụ: log ssh

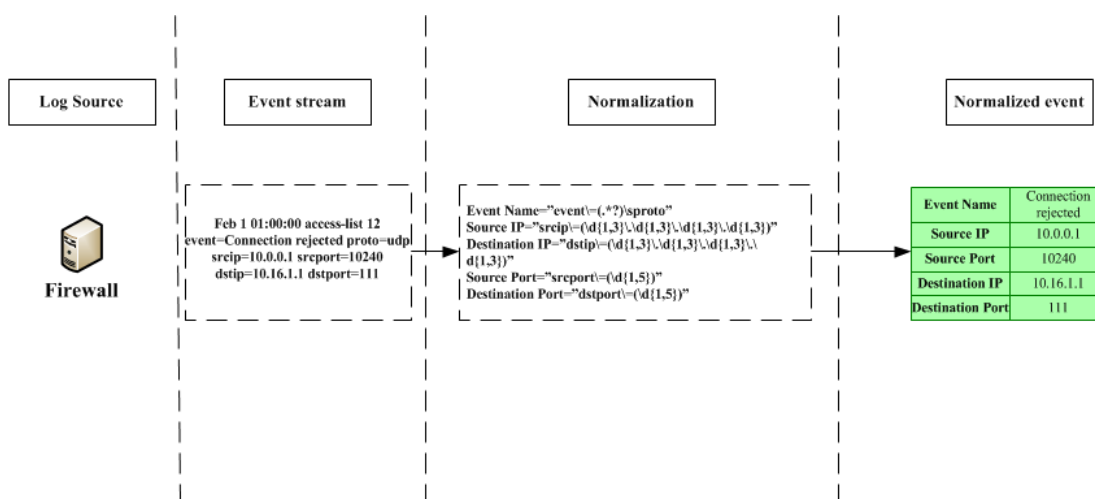
Sep 28 16:39:03 app_server sshd[8677]: Failed password for invalid user icecast2 from 10.72.109.227 port 57238 ssh2

Phân tích và chuẩn hóa:

```
host = app_server
process = sshd
source_user = icecast2
source_ip = 10.72.109.227
source_port = 57238
```

- Normalization - chuẩn hóa:

Hợp nhất các dữ liệu khác nhau thành một định dạng rút gọn chứa các thuộc tính sự kiện phổ biến. Tuân theo tiêu chuẩn để giảm các bản ghi thành các thuộc tính sự kiện chung, tên và giá trị trường chung.



Hình 1.3. Chuẩn hóa dữ liệu

❖ Correlation rules & alerts - quy tắc tương quan và cảnh báo

Là thành phần quan trọng nhất trong SIEM. 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 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. Thực hiện phân tích, kiểm tra các biểu thức logic

đã thêm. Chạy trên tất cả các nguồn nhật ký được chuyển và chuẩn hóa. Nói cách khác, quy tắc tương quan là một điều kiện có chức năng kích hoạt cảnh báo

Ví dụ:

Nếu server có người thực hiện thao tác đăng nhập. Trong khi nhiều lần đăng nhập thất bại và sau đó đột nhiên đăng nhập thành công, đó là dấu hiệu của tấn công brute force.

Khi này, quy tắc sẽ được đặt ra là: nếu đăng nhập 10 lần thất bại trong vòng 5 phút, sau đó đăng nhập thành công thì tiến hành gửi thông báo.

❖ Lưu trữ Log:

Với các bản ghi log gửi tới SIEM cần một cách để lưu trữ chúng để phục vụ các mục đích lưu giữ và truy vấn sau này. 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 khá dễ dàng tương tác 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, nhưng một số vấn đề có thể phát sinh tùy thuộc vào cách SIEM triển khai cơ sở dữ liệu tương ứng với nó. Nếu SIEM là một thiết bị thường không có nhiều sự tương tác với cơ sở dữ liệu, do đó việc cung cấp và bảo trì thường không phải là một vấn đề. Nhưng nếu SIEM đang chạy trên phần cứng riêng thì việc quản lý cơ sở dữ liệu cho mình cũng là vấn đề lớn. Điều này có thể là khó khăn nếu không có một DBA.

- 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. 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 dưới dạng nhị phân. SIEM biết làm thế nào để đọc và ghi vào những file này.

❖ Theo dõi và giám sát

Khi đã có tất cả các bản ghi log trong SIEM và các sự kiện an ninh đã được xử lý, cần một giao diện GUI để tiếp cận, tương tác 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 giám sát. 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 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, 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.

1.2.4. Lợi ích của hệ thống SIEM

- **Phát hiện mối đe dọa:** Bất kỳ mối đe dọa hoặc sự cố nào cũng có thể được tìm thấy với SIEM. Giám sát an ninh là một trong nhiều dịch vụ và chức năng được cung cấp bởi SIEM, bao gồm cả việc thu thập log tiêu chuẩn, chuẩn hóa, tương quan và phân tích. Khi ai đó vi phạm chính sách, dù là bên trong hay bên ngoài công ty, quản trị viên có thể cấu hình SIEM để thông báo cho các nhà phân tích an ninh. Nó cũng có thể cảnh báo khi phát hiện ra mối đe dọa.
- **Nâng cao hiệu suất:** SIEM hiệu quả hơn so với các giải pháp trước đó. Nhân viên có thể kiểm tra log sự kiện từ các thiết bị khác nhau trên mạng được thu thập bởi các hệ thống SIEM để phát hiện các vấn đề tiềm ẩn. Điều này cũng làm cho việc kiểm tra hoạt động trở nên dễ dàng hơn và đẩy nhanh quá trình phân tích file, cho phép nhân viên hoàn thành nhiệm vụ một cách dễ dàng và dành nhiều thời gian hơn cho các khía cạnh khác của công việc. Theo cách này, các hệ thống SIEM có thể nâng cao thủ tục báo cáo trong toàn tổ chức.
- **Quản lý dịch vụ bảo mật:** Xử lý tốt hơn các vi phạm và sự kiện bảo mật là một lợi thế khác của việc sử dụng SIEM. Bằng cách cung cấp phản ứng nhanh chóng với bất kỳ sự kiện bảo mật nào được quan sát, phần mềm có thể giảm đáng kể tác động tiêu cực của một vi phạm bảo mật đối với tổ chức. Phản ứng nhanh chóng từ phần mềm SIEM và các chuyên gia IT có thể giảm đáng kể tác động tài chính của một vi phạm cũng như mức độ thiệt hại đối với tổ chức và bất kỳ hệ thống IT hiện có nào. Phát hiện sớm một vi phạm hoặc phát hiện sự kiện bảo mật trước khi nó xảy ra cũng có thể ngăn chặn bất kỳ thiệt hại nào.

- **Báo cáo toàn diện:** Các tổ chức cũng sẽ hưởng lợi từ báo cáo toàn diện khi sử dụng SIEM. Có thể khó để có được các báo cáo đầy đủ về mức độ bảo mật trên toàn mạng vì có rất nhiều công cụ phần mềm được sử dụng để bảo mật các khía cạnh khác nhau của mạng. Điều này là vì mỗi sản phẩm phần mềm tạo ra các báo cáo dựa trên nhiệm vụ mà nó được thiết kế để thực hiện. Ví dụ, log hệ thống ngăn chặn xâm nhập mạng khác với log bảo mật của tường lửa. SIEM tạo ra các báo cáo chi tiết mô tả trạng thái của toàn bộ mạng, không chỉ một phần của nó, bằng cách thu thập tập trung.

1.3. Một số giải pháp SIEM

1.3.1. Splunk

Splunk được biết đến là một hệ thống xây dựng thu thập log tập trung, dùng để phân tích, xử lý, phản ứng, phát hiện kịp thời các hành vi bất thường trong hệ thống từ đó đưa ra các cảnh báo đến người quản trị để kịp thời xử lý về bảo mật, hệ thống.

Các tính năng chính của Splunk:

- Thu thập dữ liệu từ nhiều nguồn khác nhau và lập chỉ mục để dễ dàng tìm kiếm. Ngoài ra, công cụ Splunk Universal Forwarder có thể được cài đặt trên các máy chủ và thiết bị để chuyển dữ liệu đến Splunk indexers một cách hiệu quả và bảo mật.
- Search Processing Language (SPL) là ngôn ngữ tìm kiếm mạnh mẽ của Splunk, cho phép người dùng thực hiện các truy vấn phức tạp để tìm kiếm, phân tích và trực quan hóa dữ liệu.
- Giám sát dữ liệu theo thời gian thực, phát hiện sự cố ngay khi chúng xảy ra.
- Cho phép thiết lập cảnh báo dựa trên các điều kiện cụ thể. Cảnh báo có thể được gửi qua email, tin nhắn hoặc tích hợp với các hệ thống cảnh báo khác
- Tạo biểu đồ, đồ thị và bảng điều khiển để trực quan hóa dữ liệu.
- Splunk có khả năng tích hợp và mở rộng với nhiều ứng dụng và Add-ons phong phú.

1.3.2. ELK Stack

ELK Stack là một giải pháp an ninh mạng được xây dựng trên cơ sở của bộ công cụ ELK (Elasticsearch, Logstash và Kibana). ELK Stack được phát triển bởi Elastic, một công ty chuyên về các công cụ phân tích dữ liệu nguồn mở. Ngăn xếp này dựa trên dự án Apache Lucene, dự án này cung cấp các khả năng công cụ tìm kiếm cốt lõi cho Elasticsearch. Elasticsearch ban đầu được phát hành vào năm 2010, tiếp theo là Logstash vào năm 2012 và Kibana vào năm 2013.

- Elasticsearch: được coi là thành phần chính của ELK stack. Nó là một công cụ tìm kiếm và phân tích dữ liệu, có khả năng lưu trữ và xuất các tài liệu dưới dạng JSON.
- Logstash: là một quy trình xử lý dữ liệu có thể được sử dụng để thu thập, phân tích và lọc dữ liệu từ nhiều nguồn khác nhau trước khi gửi đến Elasticsearch để lưu trữ và phân tích.
- Kibana: là một công cụ trực quan có thể được sử dụng để tạo bảng điều khiển, biểu đồ và truy vấn giúp người dùng hiểu rõ hơn về dữ liệu của họ. Kibana rất dễ sử dụng và cung cấp nhiều loại trực quan tích hợp.

ELK Stack là một lựa chọn phổ biến cho việc xử lý và phân tích dữ liệu với tính linh hoạt, khả năng mở rộng và hiệu suất cao.

1.3.3. Wazuh

Wazuh là một nền tảng mã nguồn mở dùng cho việc giám sát an ninh. Được xây dựng từ các thành phần: OSSEC HIDS, OpenSCAP và Elastic Stack. Wazuh cung cấp một loạt các tính năng giám sát an ninh, bao gồm giám sát tệp, giám sát sự kiện, giám sát mạng và giám sát máy chủ. Nó cung cấp một số công cụ quan trọng để phát hiện các mối đe dọa an ninh và hỗ trợ việc phân tích bảo mật.

Wazuh có một loạt các tính năng quan trọng như:

- Hệ thống phát hiện xâm nhập (IDS) và hệ thống ngăn chặn xâm nhập (IPS)
- Giám sát tệp để phát hiện các hành động đáng ngờ như thay đổi tệp hoặc thực thi tệp lạ
- Giám sát sự kiện để phát hiện các hành vi đáng ngờ trên hệ thống

- Phân tích bảo mật để xác định các mối đe dọa mới và cung cấp thông tin về các mối đe dọa hiện tại
- Cảnh báo và báo cáo để thông báo cho người dùng về các mối đe dọa an ninh
- Tích hợp với các công cụ quản lý bảo mật khác để tăng tính toàn vẹn của hệ thống an ninh.

1.3.4. IBM QRadar

IBM QRadar là một nền tảng an ninh mạng và giám sát sự cố được phát triển bởi IBM. QRadar là viết tắt của "Quarantine Radar", tượng trưng cho khả năng phát hiện và phản ứng nhanh chóng đối với các mối đe dọa an ninh mạng. Là một hệ thống SIEM mạnh mẽ, QRadar giúp tổ chức và doanh nghiệp giám sát, phân tích và báo cáo về các sự kiện an ninh từ nhiều nguồn khác nhau trên mạng.

QRadar có khả năng thu thập và xử lý dữ liệu từ nhiều nguồn khác nhau như logs hệ thống, logs ứng dụng, dữ liệu từ thiết bị mạng và dữ liệu từ các nguồn bên ngoài như các công cụ bảo mật mạng và dịch vụ thông tin an ninh. Sau đó, nó phân tích dữ liệu này để phát hiện các mẫu, xu hướng và sự cố an ninh, đồng thời cung cấp các cảnh báo và báo cáo để giúp tổ chức phản ứng kịp thời và hiệu quả đối với các mối đe dọa.

Các tính năng chính của Qradar SIEM:

- Giám sát và phân tích các sự kiện an ninh để phát hiện các mối đe dọa an ninh mới và hiện tại.
- Tích hợp với các công cụ quản lý bảo mật khác để tăng tính toàn vẹn của hệ thống an ninh như hệ thống antivirus, IDS/IPS, tường lửa...
- QRadar thu thập logs từ các hệ thống và ứng dụng (máy chủ, máy cá nhân, ứng dụng), logs thiết bị mạng (tường lửa, router, switch), hoặc thu thập dữ liệu và các nguồn bên ngoài từ các dịch vụ được tích hợp.
- Thu thập dữ liệu và chuẩn hóa thành định dạng chuẩn, phân tích dữ liệu trong thời gian thực.

- Tạo ra các báo cáo và thống kê về hoạt động an ninh mạng, bao gồm các cảnh báo, xu hướng, và biểu đồ thống kê. Qradar đưa ra các thông báo khi phát hiện các sự kiện an ninh mạng hoặc hành vi đáng ngờ, cung cấp thông tin chi tiết về các mối đe dọa và sự cố.
- Thực hiện các biện pháp phản ứng tự động, như cắt kết nối hoặc cấm IP, đóng port để ngăn chặn các mối đe dọa an ninh mạng.

1.4. Kết luận Chương 1

Chương 1 đã cung cấp một cái nhìn tổng quan về mô hình kiến trúc, chức năng của hệ thống Quản lý Thông tin và Sự kiện Bảo mật (SIEM), và một số công cụ triển khai giải pháp SIEM. Cho thấy tầm quan trọng của nó trong việc bảo vệ an ninh mạng cho các tổ chức. Trong bối cảnh số lượng các cuộc tấn công mạng ngày càng gia tăng và trở nên phức tạp, SIEM là một công cụ thiết yếu giúp các tổ chức phát hiện và phản ứng kịp thời với các mối đe dọa ngày nay.

CHƯƠNG 2. TÌM HIỂU CÔNG CỤ GIÁM SÁT AN NINH MẠNG SPLUNK

2.1. Tổng quan về Splunk.

Splunk là hệ thống có thể captures, trích ra các dữ liệu thời gian thực có liên quan tới nhau từ đó nó có thể tạo ra các đồ thị, các báo cáo, các cảnh báo và các biểu đồ. Splunk hỗ trợ hầu hết các loại log trên hệ thống như phần mềm, Firewall, IDS/IPS,... của máy trạm.

Mục đích của Splunk là giúp cho việc xác định mô hình dữ liệu và thu thập dữ liệu máy trên toàn hệ thống dễ dàng hơn. Splunk có thể tìm kiếm các sự kiện đã và đang xảy ra, đồng thời cũng có thể báo cáo và phân tích thống kê các kết quả tìm được. Nó có thể nhập các dữ liệu của máy dưới dạng có cấu trúc hoặc không cấu trúc. Hoạt động tìm kiếm và phân tích sử dụng SPL (Search Processing Language), được tạo để quản lý Big Data. Do được phát triển từ Unix Piping và SQL nên Splunk có khả năng tìm kiếm dữ liệu, lọc, sửa đổi, chèn và xóa dữ liệu.

Cốt lõi của Splunk là một nền tảng dữ liệu cho mọi thứ, cho phép người dùng thu thập, lập chỉ mục, tìm kiếm, phân tích và trực quan hóa dữ liệu theo thời gian thực. Nó cung cấp một giao diện đa năng và thân thiện với người dùng để truy vấn và khám phá dữ liệu, làm cho nó trở nên dễ tiếp cận ngay cả với những người dùng có ít kiến thức kỹ thuật. Sự linh hoạt của Splunk cho phép nó tiếp nhận nhiều định dạng dữ liệu khác nhau, khiến nó phù hợp với các tổ chức có nhiều nguồn dữ liệu đa dạng.

Các chức năng cơ bản của Splunk:

❖ Index data (Chỉ mục dữ liệu)

Là thành phần quan trọng của hệ thống Splunk. Nó thu thập và xử lý các data đầu vào từ bất kỳ nguồn nào. Có thể xem Indexer là một nhà máy và data là những nhiên liệu thô cần phải xử lý. Khi data được chuyển vào nhà máy, Indexer đóng vai thanh tra nhìn vào các data đó mà đưa ra quyết định xử lý chúng. Lúc này data sẽ được gắn nhãn sourcetype để phân loại. Dựa vào nhãn sourcetype này

data sẽ được cắt thành các single event và gắn nhãn timestamp. Sau đó chúng được lưu ở Splunk index nơi mà người dùng có thể tìm kiếm.

Splunk có thể index cho rất nhiều kiểu dữ liệu. Các nguồn dữ liệu thông thường:

- Các dữ liệu có cấu trúc: Các tập tin CSV, JSON hay XML
- Các dịch vụ Web: Apache, IIS
- Các phần mềm vận hành IT: Nagios, NetApp, Cisco USC
- Dịch vụ cơ sở dữ liệu: Oracle, MySQL, Microsoft SQL Server
- Các nền tảng ảo hóa: VMWare, Xen Desktop, XenApp, Hyper-V
- Các dịch vụ ứng dụng: JMX & JMS, WebLogic, WebSphere
- Nền tảng của Microsoft: Exchange, Active Directory, Sharepoint

❖ **Investigational searching (Tìm kiếm điều tra)**

Thực hành tìm kiếm điều tra thường được xem như các quy trình kiểm tra môi trường, cơ sở hạ tầng hoặc một lượng lớn dữ liệu để tìm kiếm sự xuất hiện của các sự kiện, lỗi hoặc sự cố cụ thể. Ngoài ra, quá trình này có thể bao gồm việc tìm kiếm thông tin cho thấy tiềm năng xảy ra một sự kiện, lỗi hoặc sự cố.

Như đã đề cập, Splunk lập chỉ mục và cho phép tìm kiếm và điều hướng qua dữ liệu và các nguồn dữ liệu từ bất kỳ ứng dụng, máy chủ, hoặc thiết bị mạng nào trong thời gian thực. Điều này bao gồm các nhật ký, cấu hình, tin nhắn, báo và cảnh báo, kịch bản, và hầu như bất kỳ loại chỉ số nào, ở hầu như bất kỳ vị trí nào. Chức năng tìm kiếm mạnh mẽ của Splunk có thể được truy cập thông qua ứng dụng Tìm kiếm & Báo cáo của nó (đây cũng là giao diện mà người dùng sử dụng để tạo và chỉnh sửa báo cáo). Ứng dụng Tìm kiếm & Báo cáo cung cấp cho người dùng một thanh tìm kiếm, trình chọn phạm vi thời gian và bản tóm tắt dữ liệu đã được đọc và lập chỉ mục bởi Splunk. Ngoài ra, còn có một bảng điều khiển thông tin bao gồm các biểu tượng hành động nhanh, bộ chọn chế độ, trạng thái sự kiện và một số tab để hiển thị các kết quả sự kiện khác nhau.

Chức năng tìm kiếm của Splunk có thể cung cấp cho người dùng các tính năng sau:

- Xác định sự tồn tại của hầu hết mọi thứ (không chỉ là một danh sách ngăn các trường được xác định trước).
- Tạo các tìm kiếm kết hợp thời gian và thuật ngữ.
- Tìm lỗi vượt qua nhiều tầng của cơ sở hạ tầng (và thậm chí truy cập các môi trường dựa trên đám mây).
- Tìm và theo dõi các thay đổi cấu hình.

Một tính năng nâng cao hơn của Splunk là khả năng tạo và chạy các tìm kiếm tự động thông qua giao diện dòng lệnh (CLI), nâng cao hơn đó là thông qua API REST của Splunk. Các tìm kiếm Splunk được khởi tạo bằng các tính năng nâng cao này không đi qua Splunk Web. Do đó, chúng hiệu quả hơn nhiều vì khi này Splunk không tính toán hoặc tạo ra dòng thời gian sự kiện, giúp tiết kiệm thời gian xử lý.

❖ **Monitor and alert (giám sát và cảnh báo)**

Giám sát nhiều ứng dụng và môi trường là một yêu cầu điển hình của bất kỳ trung tâm dữ liệu hoặc hỗ trợ nào của tổ chức. Khả năng giám sát bất kỳ hạ tầng nào theo thời gian thực là rất quan trọng để nhận diện các vấn đề, sự cố và tấn công trước khi chúng có thể ảnh hưởng đến khách hàng, dịch vụ và cuối cùng là lợi nhuận. Với khả năng giám sát của Splunk, các dấu hiệu cụ thể sẽ được thiết lập dưới dạng các sự kiện để Splunk theo dõi, giúp cho các cá nhân không cần phải làm điều đó.

Splunk có khả năng kích hoạt thông báo theo thời gian thực để có thể thực hiện các biện pháp phù hợp như là theo dõi, tránh một sự kiện cũng như tránh thời gian chết và chi phí có thể gây ra bởi sự kiện đó. Splunk có khả năng thực hiện các hành động dựa trên các sự kiện hoặc điều kiện cụ thể.

Ngoài việc tìm kiếm và giám sát dữ liệu lớn, Splunk có thể được cấu hình để thông báo cho bất kỳ ai trong tổ chức khi một sự kiện xảy ra hoặc khi kết quả tìm kiếm đáp ứng các điều kiện cụ thể. Người dùng có thể tự động chạy tìm kiếm theo thời gian thực và lịch sử của mình theo một lịch trình đều đặn cho một loạt các kịch bản thông báo khác nhau.

❖ Report and analyze (Báo cáo và phân tích)

Các cảnh báo tạo ra các bản ghi khi chúng được kích hoạt. Splunk tổng hợp các báo cáo và sơ đồ hóa thành các biểu đồ. Trình quản lý cảnh báo của Splunk có thể được sử dụng để lọc các bản ghi kích hoạt (kết quả cảnh báo) theo ứng dụng, mức độ nghiêm trọng của cảnh báo và loại cảnh báo. Người dùng cũng có thể tìm kiếm các từ khóa cụ thể trong kết quả cảnh báo. Các bản ghi cảnh báo/kích hoạt có thể được thiết lập để tự động hết hạn, hoặc có thể sử dụng trình quản lý cảnh báo để xóa các bản ghi cảnh báo cụ thể một cách thủ công theo ý muốn. Các báo cáo có thể được chia sẻ lẫn nhau để mọi người có cái nhìn tổng quan về hệ thống.

2.2. Ứng dụng của Splunk

2.2.1. Quản lý ứng dụng

- Troubleshoot vấn đề một cách nhanh chóng, giảm chi phí và giảm thời gian để điều tra và khắc phục sự cố tới 70%.
- Giảm sự phức tạp bằng cách cung cấp cho các nhà phát triển được truy cập vào log của ứng dụng thông qua một vị trí trung tâm mà không cần quyền truy cập vào hệ thống đó.
- Giám sát toàn bộ môi trường ứng dụng trong thời gian thực để ngăn chặn các vấn đề ảnh hưởng tới người dùng, giữ lại log từ các sự kiện định kỳ để ngăn ngừa mất mát.
- Truy vết và giám sát các giao dịch của ứng dụng thông qua các tầng của kiến trúc phân tán và từ nhiều nguồn dữ liệu.
- Phát hiện các bất thường hoặc các vấn đề trong hoạt động, thời gian đáp ứng và chủ động giải quyết chúng trước khi nó ảnh hưởng tới người dùng ứng dụng.
- Theo dõi số liệu hoạt động quan trọng như thời gian đáp ứng end-to-end, độ dài thông điệp hàng đợi và đếm số lần giao dịch thất bại để đảm bảo ứng dụng đáp ứng được nhu cầu cần thiết.

- Nắm được toàn bộ hoạt động của ứng dụng trong thời gian thực trên toàn bộ cơ sở hạ tầng ứng dụng.
- Làm phong phú hệ thống bằng cách thêm các nguồn phi CNTT như giá cả cơ sở dữ liệu, thông tin khách hàng và thông tin vị trí.

Không giống các công cụ quản lý truyền thống, Splunk có thể index, phân tích, khai thác dữ liệu từ bất kỳ tầng ứng dụng nào. Nó cung cấp một góc nhìn trung tâm về toàn bộ hệ thống cơ sở hạ tầng. Ngôn ngữ tìm kiếm trong Splunk giúp người sử dụng so sánh các sự kiện, các giao dịch và chỉ số hoạt động quan trọng khác. Quyền điều khiển được trao cho nhiều nhóm trong một tổ chức. Những hiểu biết về dữ liệu ứng dụng có thể kết hợp với thông tin có cấu trúc như thông tin user hoặc giá cả thông tin để doanh nghiệp quyết định tốt hơn. Nhà sản xuất quản lý hoạt động ứng dụng AppDynamics và ExtraHop đã phát triển ứng dụng Splunk để giúp khách hàng quản lý tốt hơn các dữ liệu ứng dụng như log, các sự kiện, hoạt động của cơ sở hạ tầng.

2.2.2. Quản lý hoạt động IT

Trung tâm IT dữ liệu trên toàn thế giới đang trở nên cực kỳ phức tạp, với hàng trăm công nghệ khác nhau và thiết bị ở nhiều lớp. Ảo hóa và điện toán đám mây cũng đang trở nên phức tạp, đặc biệt là các vấn đề liên quan đến hiệu suất hoạt động. Đội ngũ quản trị và quản lý CNTT lãng phí nhiều thời gian trong việc di chuyển từ một giao diện điều khiển tới giao diện điều khiển khác, cố gắng theo dõi các dữ liệu cần thiết để đảm bảo hiệu suất và tính sẵn sàng cao.

Splunk cung cấp một cách tiếp cận tốt hơn mà không cần phải phân tích cú pháp hay tùy chỉnh nó. Splunk thu thập và lập các chỉ mục chứa tất cả dữ liệu được tạo ra bởi hệ thống IT (hệ thống mạng, server, OS, ảo hóa, v.v.v. Nó hoạt động với bất kỳ dữ liệu mà máy tạo ra, bao gồm log, file cấu hình, số liệu hiệu suất, SNMP trap và các ứng dụng log tùy chỉnh.

- Giải quyết vấn đề nhanh hơn, giảm thời gian Downtime
- Giúp nắm bắt được hoạt động ảo hóa, hệ thống cloud private và public từ 1 giao diện trung tâm.

- Giúp tìm được nguồn gốc của vấn đề nhanh hơn 70% mà không cần phải tìm kiếm trong hệ thống, server hay máy ảo.
- Quản lý hệ thống trong thời gian thực, ngăn ngừa vấn đề xảy ra trước khi nó ảnh hưởng tới người dùng, thêm kinh nghiệm xử lý các sự kiện xảy ra định kỳ để tránh mất mát.
- Chỉ cần một người quản lý có quyền truy cập trực tiếp, đảm bảo an toàn cho dữ liệu, giúp tránh leo thang đặc quyền.
- Tương quan các sự kiện ở tất cả các tầng layer của hệ thống
- Tìm các liên kết giữa người sử dụng, hiệu suất các sự kiện liên quan tới cơ sở hạ tầng được cung cấp bởi splunk.
- Kết hợp phân tích dữ liệu thời gian thực tương quan, so sánh với hàng triệu terabytes dữ liệu lịch sử.
- Phân tích phát hiện thành phần khả nghi có thể giúp dự đoán và ngăn ngừa mất mát hoặc vấn đề về hiệu năng.
- Chứa dữ liệu trên mỗi tầng của trung tâm dữ liệu. Quản lý môi trường để nhận biết được sự thay đổi, so sánh ngay lập tức để biết độ thiếu hụt hiệu năng của hệ thống, những vấn đề có sẵn hoặc vấn đề bảo mật, an ninh.
- Giảm chi phí cung cấp dịch vụ CNTT.

2.2.3. An ninh trong lĩnh vực IT

Với ứng dụng an ninh của Splunk chúng ta có thể sử dụng số liệu thống kê trên bất kỳ dữ liệu nào để tìm kiếm các mối đe dọa tiềm ẩn, trong khi vẫn có thể giám sát liên tục các mối đe dọa bị phát hiện bởi những sản phẩm an ninh truyền thống.

Ứng dụng an ninh Splunk chạy ở phía trên Splunk Enterprise và cung cấp công cụ để giám sát, cảnh báo và phân tích cần thiết để xác định và giải quyết các mối đe dọa đã biết và chưa biết. Nó phù hợp với đội ngũ an ninh nhỏ hoặc một trung tâm hoạt động bảo mật.

Bảng điều khiển an ninh cung cấp một cách xem hoàn toàn tùy biến với các từ khóa bảo mật quan trọng trong lĩnh vực an ninh domain. Ứng dụng an ninh

Splunk chứa một thư viện dựng sẵn các số liệu an ninh để hỗ trợ người dùng nhận diện được các tình huống và giám sát liên tục các nguy cơ bảo mật trên domain. Và tất cả thông tin đó đều được thể hiện rõ trên bảng điều khiển Dashboard.

Tính năng xem xét lại các sự kiện đã xảy ra: Cung cấp chi tiết quy trình công việc phân tích cần thiết để các ưu tiên của vụ việc, bối cảnh của sự cố, loại của nó và các máy chủ có liên quan.

Tính năng bảo vệ tài sản và điều tra nhận dạng mỗi nguy hiểm cung cấp cho nhà phân tích an ninh khả năng xem xét các mối đe dọa dựa trên một loạt các sự kiện an ninh. Đơn giản chỉ cần chọn một khung thời gian sự kiện hoặc nhiều sự kiện đại diện cho những hoạt động đáng ngờ và Splunk sẽ tự động hiển thị một bản tóm tắt mô hình an ninh. Dữ liệu được đặt ra theo thứ tự thời gian, đưa ra một cái nhìn trực tiếp cho người sử dụng hoặc tạo ra một tìm kiếm mới để xem các sự kiện đã xuất hiện này có tiếp tục xuất hiện hay không.

Splunk cung cấp dịch vụ out-of-the-box hỗ trợ cho 18 mã nguồn mở đe dọa tới dữ liệu nhằm tăng thêm tính bảo mật cho hệ thống. Splunk cho phép thêm mã nguồn mở riêng và nguồn cung cấp dữ liệu thanh toán mà không cần cam kết dịch vụ.

2.3. Các thành phần của Splunk

Có 3 thành phần chính trong Splunk:

- Splunk Forwarder: dùng để chuyển tiếp dữ liệu.
- Splunk Indexer: được sử dụng để phân tích cú pháp và lập chỉ mục dữ liệu
- Search Head: là GUI được sử dụng để tìm kiếm, phân tích và báo cáo.

2.3.1. Splunk Forwarder

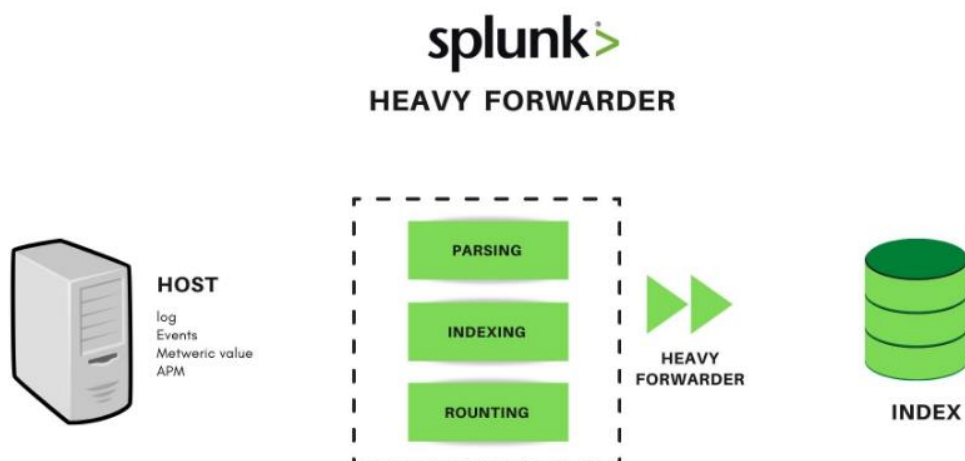
2.3.1.1. Các loại Forwarder

Trong Splunk, thành phần chính để thu thập log là Splunk Forwarder, có nhiệm vụ thu thập và chuyển tiếp dữ liệu từ các nguồn khác đến Splunk Indexer để phân tích và lưu trữ. Có hai loại Splunk Forwarder chính:

- Universal forwarder (UF): là một ứng dụng nhẹ và hiệu quả, được thiết kế để thu thập và chuyển tiếp dữ liệu từ các máy chủ và thiết bị khác nhau đến

Splunk Indexer. Nó chỉ thu thập và chuyển tiếp dữ liệu mà không thực hiện bất kỳ xử lý hoặc phân tích nào trên dữ liệu. UF thường được sử dụng để triển khai rộng rãi trên các hệ thống khác nhau vì nó tiêu thụ ít tài nguyên hệ thống. UF được cấu hình để theo dõi và thu thập dữ liệu từ các nguồn như tệp log, sự kiện hệ thống, dữ liệu mạng, và nhiều loại dữ liệu khác.

- Heavy forwarder (HF): là một ứng dụng mạnh mẽ hơn, có khả năng thu thập dữ liệu và thực hiện các tác vụ xử lý và phân tích trước khi gửi dữ liệu đến Splunk Indexer. HF có thể lọc, chuyển đổi và thực hiện các tìm kiếm phức tạp trên dữ liệu trước khi chuyển tiếp. Điều này có thể giúp giảm tải cho Indexer và tối ưu hóa quá trình phân tích. HF thường được cấu hình và sử dụng trong các kịch bản cần xử lý dữ liệu nhiều hoặc nơi cần thực hiện các thao tác đặc biệt trên dữ liệu trước khi gửi đến Indexer.

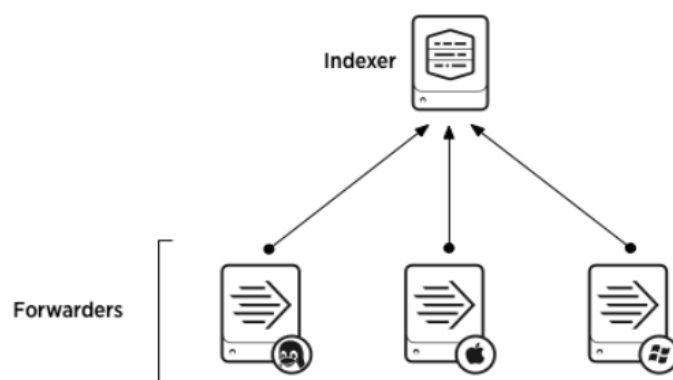


Hình 2.1. Mô tả Heavy forwarder trong Splunk

2.3.1.2. Các mô hình dữ liệu

❖ Mô hình chuyển tiếp dữ liệu tập trung (Data consolidation):

Là một trong những mô hình phổ biến với nhiều thiết bị forwarder từ các nguồn khác nhau gửi đến một splunk server. Mô hình này thường là các universal forwarder chuyển tiếp dữ liệu chưa phân tích từ máy trạm hoặc các thiết bị không phải là splunk server tới máy chủ splunk trung tâm để tổng hợp và đánh chỉ mục cho dữ liệu.



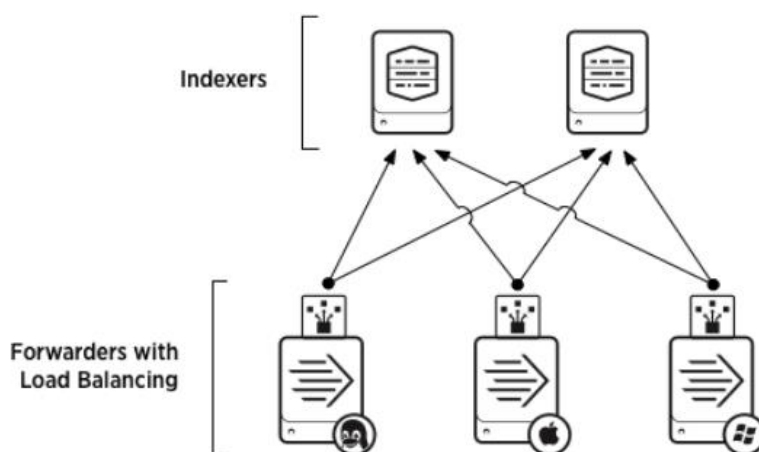
Hình 2.2. Mô hình thu thập log tập trung

Mô hình dữ liệu tập trung chỉ phù hợp với một hệ thống nhỏ còn đối với một hệ thống máy chủ lớn với lượng log gửi về nhiều khả năng chịu tải của máy chủ sẽ phải rất lớn do đó người ta nghĩ đến giải pháp cân bằng tải để tăng khả năng lưu trữ, cũng như giảm rủi ro và nâng cao hiệu năng tìm kiếm.

❖ Mô hình cân bằng tải (Load balancing):

Cân bằng tải là quá trình đơn giản hóa phân phối dữ liệu trên nhiều bộ chỉ mục để xử lý các yêu cầu như khối lượng dữ liệu lớn, mở rộng theo chiều ngang để nâng cao hiệu suất tìm kiếm và khả năng chịu lỗi. Trong quá trình cân bằng tải, forwarder định tuyến lần lượt đến các chỉ mục cài đặt khác nhau theo các khoảng thời gian xác định.

Forwarder thực hiện cân bằng tải tự động, trong đó forwarder chuyển đổi receiver theo các khoảng thời gian cố định. Nếu parsing được bật (dành cho heavy forwarder), việc chuyển đổi sẽ xảy ra ở ranh giới sự kiện.

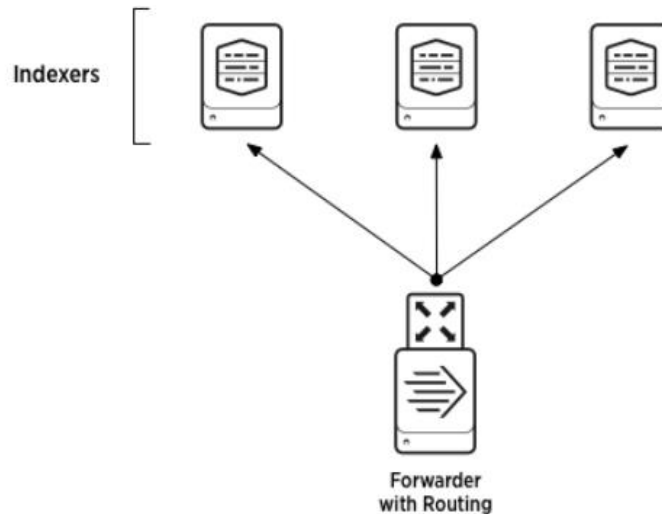


Hình 2.3. Mô hình thu thập log cân bằng tải

❖ Mô hình định tuyến và lọc (Routing and filtering)

Trong định tuyến dữ liệu, forwarder định tuyến các sự kiện đến các máy chủ cụ thể, dựa trên các tiêu chí như nguồn, loại nguồn hoặc mẫu trong chính sự kiện đó. Định tuyến ở cấp độ sự kiện yêu cầu một heavy forwarder.

Forwarder cũng có thể lọc và định tuyến các sự kiện đến các hàng đợi cụ thể hoặc loại bỏ chúng hoàn toàn bằng cách định tuyến đến hàng đợi rỗng.



Hình 2.4. Mô hình thu thập log định tuyến và lọc

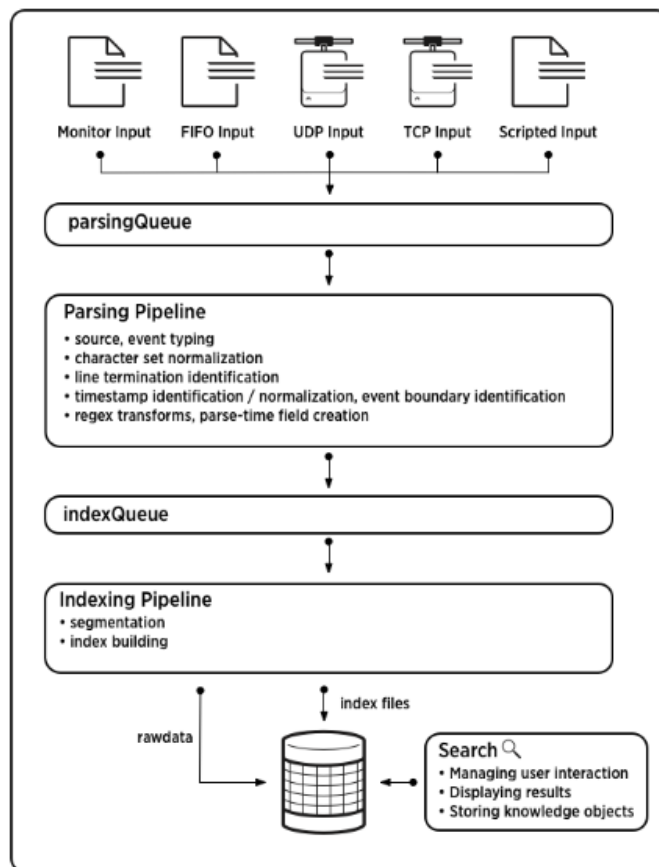
2.3.2. Splunk Indexer

Với một lượng dữ liệu lớn được truyền từ các máy chủ về máy chủ tập trung thì việc lưu trữ và tìm kiếm sẽ rất khó khăn. Bởi vậy việc đầu tiên sau khi thu thập được log về sẽ phải lập chỉ mục cho dữ liệu và lưu trữ chúng phục vụ cho việc tìm kiếm.

Indexer là một thành phần trong Splunk, nó nhận dữ liệu đến từ Splunk forwarder, chuyển đổi dữ liệu phi cấu trúc thành định dạng có cấu trúc (dạng event) và lưu trữ dữ liệu đó trong các chỉ mục để thực hiện các hoạt động tìm kiếm một cách nhanh chóng và hiệu quả.

2.3.2.1. Cách hoạt động của trình lập chỉ mục

Sơ đồ minh họa của trình hoạt động của trình lập chỉ mục:



Hình 2.5. Sơ đồ minh họa quy trình hoạt động của lập chỉ mục

Dữ liệu đi vào bộ chỉ mục và tiến hành thông qua một đường ống (pipeline) nơi diễn ra quá trình xử lý sự kiện. Cuối cùng, dữ liệu đã xử lý được ghi vào đĩa. Đường ống này bao gồm một số đường ống ngắn hơn được nối với nhau. Một thực thể duy nhất của đường dẫn dữ liệu đầu cuối này được gọi là bộ đường ống.

Quá trình xử lý sự kiện xảy ra trong hai giai đoạn chính, đó là phân tích cú pháp và lập chỉ mục. Tất cả dữ liệu đi vào Splunk đều đi qua đường dẫn phân tích cú pháp dưới dạng các khối lớn (10.000 byte). Trong quá trình phân tích cú pháp, Splunk Enterprise chia các khối này thành các sự kiện mà nó chuyển đến đường dẫn lập chỉ mục, nơi diễn ra quá trình xử lý cuối cùng.

❖ Phân tích cú pháp

Nếu đang nhận dữ liệu từ một Universal Forwarder thì trước tiên, trình lập chỉ mục sẽ phân tích dữ liệu rồi lập chỉ mục cho dữ liệu đó. Phân tích dữ liệu được thực hiện để loại bỏ dữ liệu không mong muốn. Tuy nhiên, nếu đang nhận dữ liệu từ Heavy Forwarder, thì trình lập chỉ mục sẽ chỉ lập chỉ mục dữ liệu.

Trong khi phân tích cú pháp, Splunk Enterprise thực hiện một số hành động, bao gồm:

- Trích xuất một tập hợp các trường mặc định cho từng sự kiện, bao gồm host, source và sourcetype.
- Cấu hình bộ ký tự chuyên hóa.
- Xác định điểm cuối dòng bằng cách sử dụng quy tắc ngắt dòng. Trong khi nhiều sự kiện ngắn và chỉ chiếm một hoặc hai dòng, những sự kiện khác có thể dài hơn.
- Xác định dấu thời gian (timestamp) hoặc tạo chúng nếu chúng không tồn tại. Đồng thời qua việc xử lý dấu thời gian, Splunk xác định ranh giới sự kiện.
- Splunk có thể được thiết lập để che giấu dữ liệu nhạy cảm (chẳng hạn như số thẻ tín dụng hoặc số căn cước công dân) ở giai đoạn này. Nó cũng có thể được cấu hình để áp dụng siêu dữ liệu tùy chỉnh cho các sự kiện sắp tới.

❖ **Lập chỉ mục**

Trong quy trình lập chỉ mục, Splunk Enterprise thực hiện xử lý bổ sung, bao gồm:

- Chia tất cả các sự kiện thành các phân đoạn để sau đó có thể tìm kiếm được. Người dùng có thể xác định mức độ phân đoạn, điều này ảnh hưởng đến tốc độ lập chỉ mục và tìm kiếm, khả năng tìm kiếm.
- Xây dựng cấu trúc dữ liệu chỉ mục.
- Ghi dữ liệu thô và các tệp chỉ mục vào đĩa, nơi diễn ra quá trình nén lập chỉ mục sau.

Sau khi các sự kiện được phân tích cú pháp và xử lý, Splunk lưu trữ tất cả dữ liệu mà nó xử lý dưới dạng các index. Chỉ mục là một tập hợp các cơ sở dữ liệu với các thư mục con nằm ở `SPLUNK_HOME/var/lib/splunk`. Các index chứa hai file là dữ liệu thô và file index. Splunk có thể có nhiều chỉ mục, mỗi chỉ mục dành riêng cho các loại dữ liệu hoặc trường hợp sử dụng cụ thể. Có thể chỉ định

các chỉ mục tùy chỉnh hoặc sử dụng chỉ mục chính mặc định khi định cấu hình Splunk.

2.3.2.2. *Các loại chỉ mục mặc định*

Khi cài đặt Splunk, có ba chỉ mục được cấu hình tự động:

- `main`: tất cả dữ liệu xử lý được lưu trữ tại đây trừ nếu chúng không áp dụng các quy tắc khác.
- `_internal`: lưu trữ log của Splunk và các số liệu xử lý.
- `_audit`: chứa các sự kiện liên quan đến sự giám sát thay đổi hệ thống, kiểm toán, và tất cả các lịch sử tìm kiếm của người dùng.

Quản trị viên có quyền tạo một index, sửa đổi hay xóa bỏ hoặc thay thế một index đã tồn tại. Việc quản lý index được thực hiện qua Web, CLI, và file cấu hình như `index.conf`.

2.3.2.3. *Quản lý chỉ mục*

Sau khi hoàn tất quá trình xử lý sự kiện, Splunk lưu trữ tất cả dữ liệu mà nó xử lý dưới dạng các index. Một index là một tập hợp các cơ sở dữ liệu với các thư mục con nằm ở `SPLUNK_HOME/var/lib/splunk`. Các index chứa hai file là dữ liệu thô và file index.

❖ **Tự định nghĩa các loại index**

Khi dữ liệu được thêm vào indexer xử lý và lưu trữ chúng dưới dạng index. Theo mặc định những dữ liệu đó được lưu trữ trong main index. Tuy nhiên người quản trị có thể cấu hình các index riêng cho các loại dữ liệu khác nhau. Indexer cũng có một vài index để sử dụng cho sự hoạt động trong bản thân hệ thống, cũng như cho các hoạt động khác như lập chỉ mục hay kiểm toán các sự kiện.

Có thể tạo các index không hạn chế trong Splunk. Tất cả các sự kiện mà không thuộc một index nào do người dùng định nghĩa thì sẽ được đẩy vào index main và kết quả tìm kiếm của ta nếu không chỉ ra tên index cụ thể đặt ra thì sẽ hiển thị các sự kiện trong main index. Việc tạo ra nhiều loại index giúp kiểm soát được người dùng truy cập: khi phân quyền cho người dùng theo các role, hạn chế người dùng tìm kiếm các thông tin nhạy cảm.

Nếu có các chính sách khác nhau cho các dữ liệu khác nhau, có thể chuyển dữ liệu từ index này sang index khác tùy theo nhu cầu sử dụng của quản trị viên. Một lí do khác để tạo ra nhiều index là nó sẽ rất hữu ích cho việc tìm kiếm. Giả sử ta có nhiều nguồn dữ liệu khác nhau như các sự kiện gửi từ windows và các sự kiện của một web server trên hệ điều hành linux. Tất cả dữ liệu này đều được lưu trữ trong cùng một index thì khi tìm kiếm các sự kiện trên windows phải tìm qua tất cả dữ liệu của hai nguồn, lúc này tốc độ chắc chắn sẽ chậm hơn rất nhiều.

❖ **Tìm kiếm index được định nghĩa**

Trong quá trình tìm kiếm mặc định các sự kiện cần tìm sẽ nằm trong main index, nếu muốn tìm kiếm trên các index riêng phải chỉ rõ tên index muốn tìm.

Ví dụ lệnh tìm kiếm sau để tìm kiếm dữ liệu có trong index tên là win10:
`index="win10" userid=henry.gale`

❖ **Xóa các index**

Xóa bỏ các index và dữ liệu trên index Có nhiều cách để xóa bỏ các dữ liệu index hoặc thậm chí cả một index hoàn chỉnh từ bộ indexer. Chẳng hạn xóa dữ liệu cũ trên các chính sách đã quá hạn.

Khi dữ liệu trong một index đã đạt tới một thời gian nhất định hoặc khi kích thước index phát triển tới mức giới hạn, nó sẽ được đưa vào trạng thái đóng băng. Nơi mà các indexer sẽ xóa nó từ các index mà nó được lưu trữ. Trước khi xóa dữ liệu indexer có thể di chuyển nó đến một nơi lưu trữ. Tất cả những việc trên đều phụ thuộc vào cách người dùng định nghĩa chính sách hết hạn.

2.3.2.4. Lưu trữ dữ liệu

Khi Splunk lập chỉ mục dữ liệu, nó sẽ tạo ra một số tệp. Các tệp này chứa một trong các tệp dưới đây:

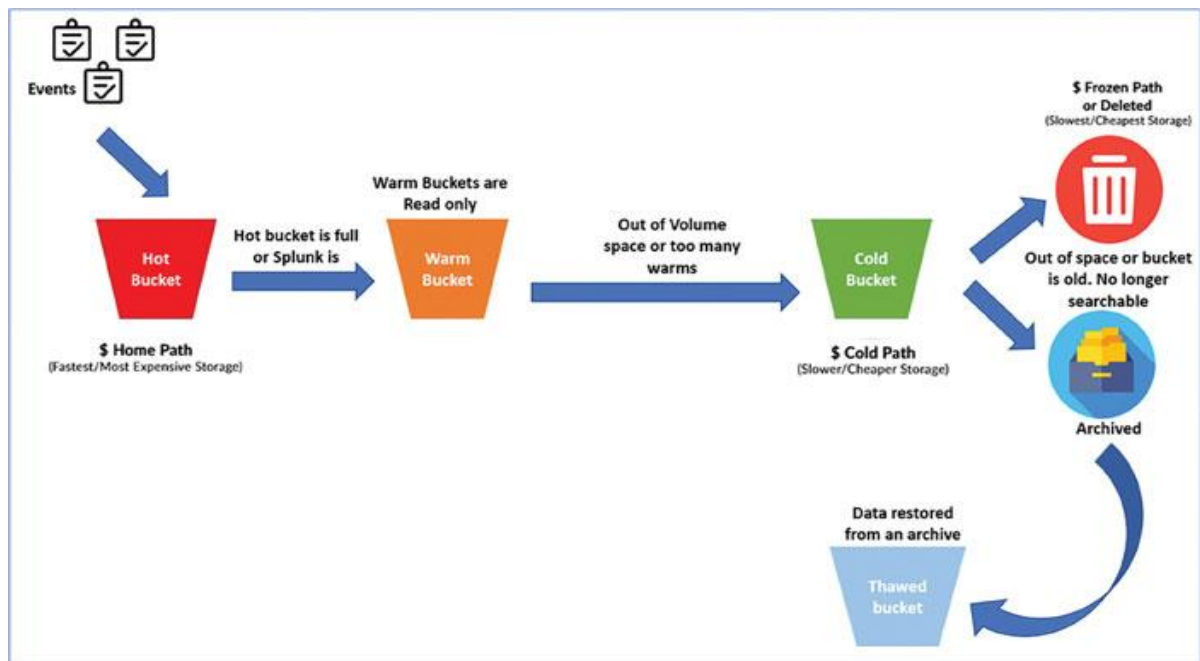
- Dữ liệu thô ở dạng nén (raw data)
- Các chỉ mục trỏ đến dữ liệu thô (index file, còn được gọi là tệp tsidx)
- Một số tệp siêu dữ liệu (metadata file)

Mỗi một thư mục index được xem như là một bucket (bộ chứa). Bucket có 4 giai đoạn:

- **Hot:** Trong giai đoạn này dữ liệu mới nhất được ghi vào và có thể tìm kiếm. Một index có thể có một vài hot bucket mở trong cùng một thời gian.
- **Warm:** Khi một điều kiện nhất định xảy ra (ví dụ như hot bucket đạt đến một kích thước giới hạn hay Splunk được khởi động lại) thì hot bucket sẽ chuyển sang giai đoạn warm bucket và một hot bucket sẽ được tạo ra tại vị trí của nó. Warm bucket sẵn sàng cho việc tìm kiếm nhưng không cho phép ghi tiếp dữ liệu vào. Trong một index thì có rất nhiều warm bucket.
- **Cold:** Khi một điều kiện tiếp tục được thỏa mãn (như index đạt đến số lượng tối đa các warm bucket). Indexer bắt đầu cuộn từ giai đoạn warm bucket sang cold, dựa trên tuổi của chúng. Nó luôn luôn lựa chọn những warm bucket lâu nhất để chuyển sang giai đoạn cold.
- **Frozen:** Sau một thời gian quy định, các cold bucket sẽ chuyển sang trạng thái frozen. Tại thời điểm này chúng sẽ được lưu trữ hoặc được xóa đi. Để định nghĩa các chính sách quá hạn ta cần chỉnh sửa các thuộc tính trong file inputs.conf. Nếu như dữ liệu frozen được lưu trữ nó có thể được khôi phục lại, đó là giai đoạn thawed. Giai đoạn này cho phép dữ liệu được phép tìm kiếm.

Khi có yêu cầu tìm kiếm, dữ liệu sẽ được trích xuất theo thứ tự Hot → Warm → Cold (tốc độ xuất ra kết quả ở hot bucket là nhanh nhất). Hot/Warm bucket có thể lưu ở một phân vùng và Cold bucket hoặc Frozen bucket ở hai phân vùng khác. Nếu muốn tìm kiếm từ Frozen bucket sẽ phải qua một bước khôi phục lại (thawed) từ archive và lập index lại.

Các giai đoạn của bucket được tóm tắt qua hình sau:



Hình 2.6. Index bucket trong Splunk

2.3.3. Splunk Search Head

Search head là thành phần được sử dụng để tương tác với Splunk. Nó cung cấp giao diện người dùng đồ họa cho người dùng để thực hiện các hoạt động khác nhau. Search Head cho phép người dùng sử dụng ngôn ngữ tìm kiếm (được gọi là Splunk Search Processing Language) để tìm kiếm dữ liệu đã được lập chỉ mục. Nó xử lý các yêu cầu tìm kiếm của người dùng và phân phối chúng đến các Indexer. Sau khi nhận được kết quả từ các Indexer, Search Head sẽ thực hiện hợp nhất các kết quả này trước khi phản hồi cho người dùng.

Ngoài ra, Search Head cũng cung cấp cho người dùng các công cụ hỗ trợ như cảnh báo, dashboard, báo cáo (report), và các công cụ trực quan hóa (visualizations). Những công cụ này giúp người dùng dễ dàng theo dõi, phân tích và trình bày dữ liệu một cách trực quan và hiệu quả.

2.3.3.1. Các loại tìm kiếm

Khi tìm kiếm, người dùng sẽ nhận ra các mẫu và xác định thêm thông tin hữu ích như các trường (field) để tìm kiếm. Người dùng có thể cấu hình Splunk để nhận dạng các trường mới này khi lập chỉ mục mới, hoặc tạo các trường mới khi tìm kiếm. Thông qua việc tìm kiếm này, người dùng sẽ có thêm kiến thức về việc sử dụng, thêm và chỉnh sửa các trường, sự kiện và giao dịch đối với dữ liệu

sự kiện của người dùng. Việc nắm bắt kiến thức này giúp người dùng xây dựng các tìm kiếm hiệu quả hơn và xây dựng các báo cáo chi tiết hơn. Thông thường, sau khi đưa dữ liệu vào trong quá trình triển khai Splunk, người dùng muốn:

- Điều tra để tìm hiểu thêm về dữ liệu vừa lập chỉ mục hoặc để tìm nguyên nhân cốt lõi của sự cố.
- Tóm tắt kết quả tìm kiếm thành một báo cáo bằng dạng bảng hoặc định dạng trực quan hóa khác.

Do đó, có hai loại tìm kiếm như sau:

- **Tìm kiếm sự kiện thô (Raw event searches):** Tìm kiếm sự kiện thô là các tìm kiếm chỉ truy xuất các sự kiện từ một chỉ mục hoặc chỉ mục và thường được sử dụng khi phân tích một vấn đề. Ví dụ: kiểm tra mã lỗi, các sự kiện tương quan, điều tra các vấn đề bảo mật và phân tích lỗi. Những tìm kiếm này thường không bao gồm các lệnh tìm kiếm (ngoại trừ chính nó) và kết quả thường là danh sách các sự kiện thô.
- **Tìm kiếm chuyển đổi (Transforming searches):** Tìm kiếm chuyển đổi là các tìm kiếm thực hiện một số loại tính toán thống kê so với một tập hợp kết quả. Đây là những tìm kiếm mà trước tiên phải truy xuất các sự kiện từ một chỉ mục và sau đó chuyển các sự kiện vào một hoặc nhiều lệnh tìm kiếm. Những tìm kiếm này sẽ luôn yêu cầu các trường và ít nhất một trong một tập hợp các lệnh thống kê. Ví dụ: đếm số sự kiện lỗi hàng ngày, đếm số lần một người dùng cụ thể đã đăng nhập.

Ngoài ra, tìm kiếm còn được phân loại theo mật độ thông tin như sau:

- **Tìm kiếm thưa thớt (sparse searches):** Tìm kiếm thưa thớt là các tìm kiếm tìm một sự kiện đơn lẻ hoặc một sự kiện xảy ra không thường xuyên trong một tập hợp dữ liệu lớn. Ví dụ: tìm kiếm địa chỉ IP hoặc mã lỗi cụ thể và duy nhất.
- **Tìm kiếm dày đặc (dense searches):** Tìm kiếm dày đặc là những tìm kiếm quét qua và báo cáo về nhiều sự kiện. Ví dụ: đếm số lỗi đã xảy ra hoặc tìm tất cả các sự kiện từ một máy chủ cụ thể.

2.3.3.2. Sử dụng trường dữ liệu (field) để tìm kiếm

Các trường tồn tại trong dữ liệu máy dưới nhiều dạng. Thông thường, một trường là một giá trị có vị trí cố định, được phân tách trên một dòng hoặc một cặp tên và giá trị, trong đó có một giá trị duy nhất cho mỗi tên trường. Một trường có thể có nhiều giá trị, nghĩa là một trường trong một sự kiện có thể có nhiều giá trị. Ví dụ: các trường sử dụng từ khóa `clientip` cho địa chỉ IP truy cập máy chủ Web của người dùng, `_time` dấu thời gian của một sự kiện và `host` tên miền của máy chủ. Một ví dụ phổ biến về trường đa giá trị là trường địa chỉ `email`. Mặc dù trường `From` sẽ chỉ chứa một địa chỉ email duy nhất, nhưng các trường `To` và `Cc` có một hoặc nhiều địa chỉ email được gắn với chúng.

Các trường là các cặp tên và giá trị có thể tìm kiếm để phân biệt sự kiện này với sự kiện khác. Không phải tất cả các sự kiện đều có trường và giá trị trường giống nhau. Sử dụng các trường để viết các tìm kiếm phù hợp hơn nhằm truy xuất các sự kiện cụ thể mà người dùng muốn.

Phần mềm Splunk trích xuất các trường từ dữ liệu sự kiện tại thời điểm lập chỉ mục và tại thời điểm tìm kiếm:

- **Thời gian lập chỉ mục:** Khoảng thời gian từ khi phần mềm Splunk nhận dữ liệu mới cho đến khi dữ liệu được ghi vào chỉ mục. Trong thời gian lập chỉ mục, dữ liệu được phân tích thành các phân đoạn và sự kiện. Các trường và dấu thời gian mặc định được trích xuất và áp dụng các phép biến đổi.
- **Thời gian tìm kiếm:** Khoảng thời gian bắt đầu khi tìm kiếm được bắt đầu và kết thúc khi tìm kiếm kết thúc. Trong thời gian tìm kiếm, một số loại xử lý sự kiện nhất định diễn ra, chẳng hạn như trích xuất trường thời gian tìm kiếm, đặt tên trường, đổi tên loại nguồn, khớp loại sự kiện, v.v.

Các trường mặc định và các trường được lập chỉ mục khác được trích xuất cho mỗi sự kiện khi dữ liệu của người dùng được lập chỉ mục.

Khi tìm kiếm các trường, sử dụng cú pháp `field_name = field_value`. Ngoài ra, tên trường có phân biệt chữ hoa chữ thường, nhưng giá

trị trường thì không. Dấu ngoặc kép là bắt buộc khi giá trị trường bao gồm dấu cách.

2.3.3.3. Ngôn ngữ xử lý tìm kiếm - SPL (*Search Processing Language*)

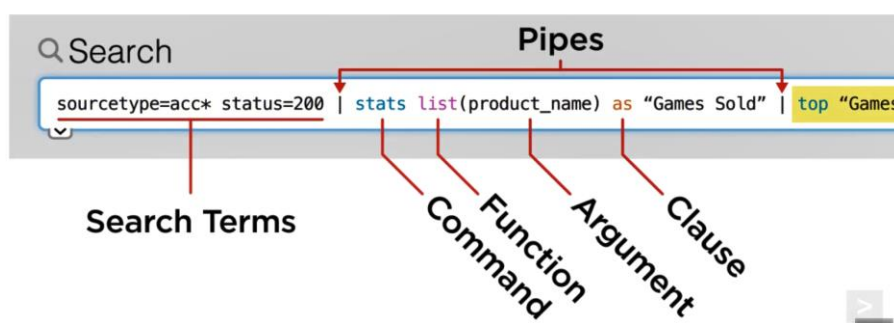
Ngôn ngữ xử lý tìm kiếm của Splunk (SPL) bao gồm tất cả các lệnh tìm kiếm cũng như các chức năng, đối số và mệnh đề của chúng. Các lệnh tìm kiếm sẽ cho phần mềm Splunk biết phải làm gì với các sự kiện người dùng đã truy xuất từ các chỉ mục. Ví dụ: người dùng cần sử dụng lệnh để lọc thông tin không mong muốn, trích xuất thêm thông tin, đánh giá các trường mới, tính toán số liệu thống kê, sắp xếp lại kết quả hoặc tạo biểu đồ.

Một số lệnh tìm kiếm có các hàm và đối số liên quan đến chúng. Sử dụng các hàm này và các đối số của chúng để chỉ định cách các lệnh tác động lên kết quả của người dùng và chúng tác động lên trường nào. Ví dụ: người dùng có thể sử dụng các hàm để định dạng dữ liệu trong biểu đồ, mô tả loại thống kê cần tính toán và chỉ định trường nào cần đánh giá. Một số lệnh cũng sử dụng các mệnh đề để chỉ định cách nhóm các kết quả tìm kiếm của người dùng.

Cú pháp câu lệnh search bao gồm:

- The Search Terms (cụm từ tìm kiếm): các từ khóa để tìm kiếm kết quả mong muốn. Ví dụ như: keywords, phrases (cụm từ), các toán tử kết hợp.
- Commands: Các lệnh cho Splunk biết sẽ làm gì với kết quả tìm kiếm, ví dụ như: thống kê tính toán, tạo biểu đồ.
- Functions: là các hàm được splunk hỗ trợ để thực hiện các chức năng như: tính toán và đánh giá kết quả.
- Arguments (đối số, tham số): các giá trị muốn truyền vào cho hàm.
- Clauses (mệnh đề): dùng để xác định hoặc nhóm các kết quả dưới dạng dữ liệu mong muốn.

Splunk Search Language Example



Hình 2.7. Minh họa cú pháp câu lệnh search

Thông thường, các lệnh được thực thi trong một đường dẫn, trong đó đầu ra của một lệnh sẽ trở thành đầu vào cho lệnh tiếp theo. Tính năng đường dẫn này cho phép người dùng xây dựng các yêu cầu phức tạp bằng cách xâu chuỗi nhiều lệnh và phép biến đổi. Dưới đây là một số lệnh SPL phổ biến và cách sử dụng chúng trong truy vấn tìm kiếm.

Bảng 2.1. Một số lệnh Splunk Search Language

Lệnh	Mô tả
stats	Tạo số liệu thống kê tóm tắt cho các sự kiện được truy vấn tìm kiếm trả về
timechart	Tạo biểu đồ và hình ảnh hóa dựa trên thời gian.
top	Hiển thị các giá trị thường xuyên nhất của một trường
table	Tạo bảng với các trường được chỉ định
eval	Tạo hoặc sửa đổi các trường bằng cách sử dụng biểu thức
join	Kết hợp các kết quả của hai hoặc nhiều truy vấn tìm kiếm dựa trên một trường chung.
fields	Cho phép chọn hoặc loại trừ các fields-trường cụ thể từ kết quả tìm kiếm
rename	Đổi tên các trường
sort	Sắp xếp sẽ cho phép hiển thị kết quả của mình theo thứ tự tăng dần hoặc giảm dần
deup	Dùng để loại bỏ các sự kiện trùng lặp dựa trên một hoặc nhiều trường

2.3.3.4. Cảnh báo

Chức năng tìm kiếm của Splunk rất quan trọng vì thông qua chức năng này, người dùng có thể tạo ra các cảnh báo tấn công dựa trên các kết quả của tìm kiếm. Khi tạo một cảnh báo, cần định nghĩa một điều kiện để kích hoạt cảnh báo đó.

Splunk hỗ trợ đặt cảnh báo Splunk qua:

- Một log event
- Chạy một script
- Gửi email
- Gửi một POST HTTP
- Các dịch vụ bên thứ ba như Slack, PagerDuty, Telegram, ...

Với cùng một điều kiện cảnh báo có thể đưa chúng vào nhiều lựa chọn khác nhau như vừa gửi mail vừa chạy script (đóng port, chặn IP, ...) để tạm thời khắc phục sự cố. Để tránh việc gửi cảnh báo quá thường xuyên, người quản trị cũng có thể giới hạn điều kiện cho một cảnh báo.

Splunk định nghĩa 2 loại cảnh báo như sau:

Bảng 2.2. Cảnh báo theo lịch trình và theo thời gian thực trong Splunk

Loại cảnh báo	Cách thức tìm kiếm	Điều kiện kích hoạt	Tùy chọn khác
Theo lịch trình (Scheduled)	Tìm kiếm theo lịch trình. Chọn từ các tùy chọn thời gian có sẵn hoặc sử dụng biểu thức cron để lên lịch tìm kiếm.	Chỉ định các điều kiện để kích hoạt cảnh báo dựa trên kết quả hoặc số lượng trường kết quả. Khi một tập hợp kết quả tìm kiếm đáp ứng các điều kiện kích hoạt, cảnh báo có thể kích hoạt một lần hoặc một lần cho mỗi kết quả.	Chỉ định một khoảng thời gian để kết thúc cảnh báo.

Theo thời gian thực (Real-time)	Tìm kiếm liên tục.	Per-result alert: bất cứ khi nào việc tìm kiếm trả về một kết quả.	Chỉ định khoảng thời gian và các giá trị trường tùy chọn để kết thúc cảnh báo.
		Rolling-window alert: tập hợp các kết quả phù hợp của việc tìm kiếm trong một khung thời gian quy định.	Chỉ định một khoảng thời gian để kết thúc cảnh báo.

Một số kịch bản của mỗi loại báo cáo:

❖ **Per-result alert:**

- Kích hoạt cảnh báo cho mỗi lần đăng nhập lỗi.
- Kích hoạt cảnh báo khi xảy ra những loại lỗi lựa chọn cho bất kỳ host nào.
- Cảnh báo xảy ra khi CPU trên host lên đến giá trị 100% trong một khoảng thời gian dài.

❖ **Scheduled alert:**

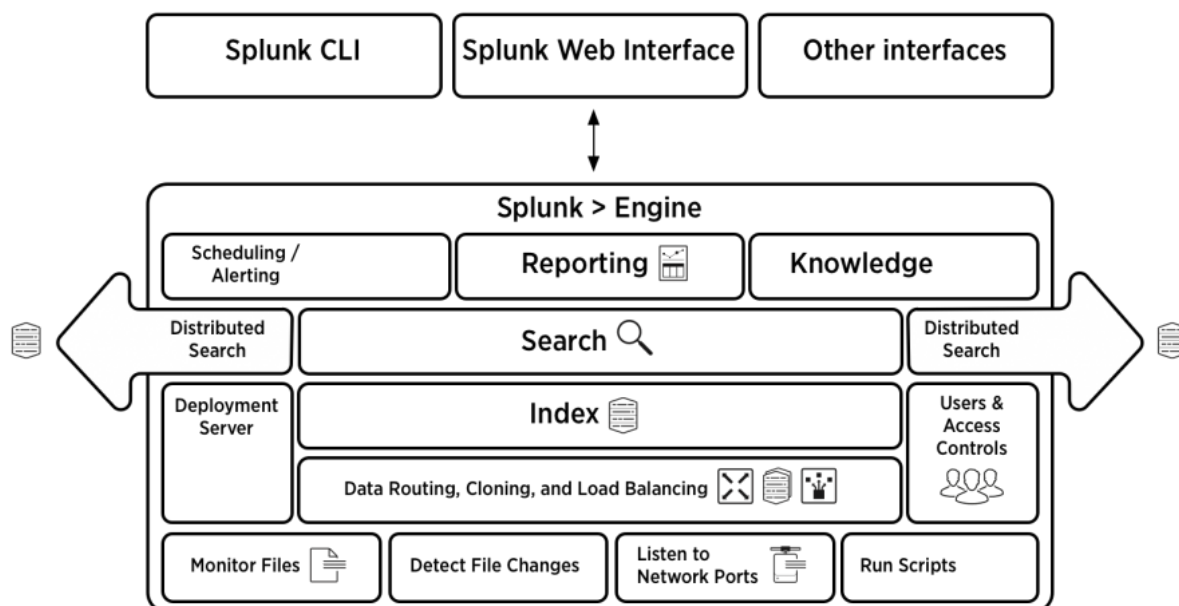
- Giả sử cần theo dõi số lượng lỗi đăng nhập thất bại trong ngày. Nếu số lượng lỗi vượt quá ngưỡng 100 lần, cảnh báo sẽ được kích hoạt.
- Quản trị viên muốn theo dõi tần suất người dùng truy cập trang web gặp lỗi HTTP 404. Tạo cảnh báo theo lịch trình để tìm kiếm lỗi 404 mỗi giờ và kích hoạt nếu có hơn 100 kết quả.

❖ **Rolling - window alert:**

- Một hành động cảnh báo sẽ xảy ra khi người dùng đăng nhập lỗi 3 lần trong vòng 10 phút. Có thể thiết lập điều chỉnh điều kiện để giới hạn việc gửi cảnh báo chỉ một lần trong vòng một giờ.

2.4. Kiến trúc Splunk

Sơ đồ sau đây minh họa toàn bộ kiến trúc Splunk.



Hình 2.8. Sơ đồ minh họa kiến trúc Splunk

- Ở mức thấp nhất, Splunk thu thập nhật ký bằng cách theo dõi tệp, phát hiện các thay đổi của tệp, nghe trên cổng hoặc chạy tập lệnh để thu thập dữ liệu nhật ký – tất cả những việc này đều được thực hiện bởi trình Splunk forwarder.
- Cơ chế lập chỉ mục, bao gồm một hoặc nhiều indexer, xử lý dữ liệu hoặc có thể nhận dữ liệu được xử lý trước bởi forwarder.
 - Máy chủ triển khai được sử dụng để quản lý toàn bộ quá trình triển khai, cấu hình và chính sách
 - Quyền truy cập và kiểm soát của người dùng được áp dụng ở cấp độ indexer – mỗi trình lập chỉ mục có thể được sử dụng cho một kho dữ liệu khác nhau, có thể có các quyền của người dùng khác nhau.
- Search head được sử dụng để cung cấp chức năng tìm kiếm theo yêu cầu và cũng hỗ trợ các tìm kiếm theo lịch trình do báo cáo tự động khởi tạo.
- Người dùng có thể xác định các đối tượng Scheduling, Reporting and Knowledge để lên lịch tìm kiếm và tạo cảnh báo.

- Dữ liệu có thể được truy cập từ UI, Splunk CLI hoặc API tích hợp với nhiều hệ thống bên ngoài.

2.5. Kết luận Chương 2

Trong chương này đã giới thiệu về Splunk từ tổng quan đến các chức năng quan trọng và các thành phần của Splunk. Ngôn ngữ xử lý tìm kiếm SPL của Splunk cung cấp nhiều công cụ mạnh mẽ để quản lý và phân tích dữ liệu log. Các lệnh và hàm của SPL cho phép người dùng dễ dàng tìm kiếm, lọc, và tạo ra các báo cáo chi tiết từ dữ liệu. Điều này không chỉ giúp xác định nguyên nhân gốc rễ của các vấn đề một cách nhanh chóng mà còn hỗ trợ trong việc ngăn ngừa các sự cố trước khi chúng ảnh hưởng đến người dùng. Việc quản lý hệ thống trong thời gian thực và khả năng tạo cảnh báo tự động cũng là những ưu điểm nổi bật của Splunk, giúp đảm bảo an toàn và hiệu quả cho hệ thống dữ liệu. Ngoài việc giám sát, Splunk còn cung cấp một cơ chế tự động khắc phục với các vấn đề xảy ra bằng việc tự động chạy các file script mà người dùng tự tạo khi có các cảnh báo xảy ra.

CHƯƠNG 3. TRIỂN KHAI THỰC NGHIỆM

3.1. Mục tiêu thực nghiệm

Phát hiện các hành động bất thường trong dữ liệu log của Windows cũng như tường lửa Pfsense để cảnh báo về các mối đe dọa tiềm ẩn bên trong cũng như bên ngoài hệ thống, từ đó đề ra các biện pháp phòng tránh cũng như cách ly các thiết bị độc hại ra khỏi hệ thống.

3.2. Xây dựng môi trường

3.2.1. Mô hình thực nghiệm

Hình 3.1. Mô hình thực nghiệm

Mô hình thực nghiệm được xây dựng trên hệ thống Splunk Enterprise được đặt tại máy Ubuntu thu thập logs tập trung từ Splunk Universal Forwarder đẩy về từ Win 10 (logs windows, logs sysmon System monitor) và logs tường lửa Pfsense.

3.2.2. Giới thiệu công nghệ hỗ trợ

❖ Windows Event Logs

Event Log của Windows được lưu trữ ở thư mục mặc định tại đường dẫn `%SystemRoot%\System32\winevt\logs`, có thể truy cập vào trực tiếp đường dẫn hoặc xem qua trình Event Viewer, để bật trình Event Viewer có thể vào RUN gõ keyword "**eventvwr**".

Trong Windows, mỗi log được cấu trúc thành các thành phần cơ bản để ghi lại thông tin liên quan đến sự kiện cụ thể. Dưới đây là các thành phần chính của một log trong Windows:

- **Event ID (ID Sự kiện):** Mỗi sự kiện trong log được gán một mã số duy nhất được gọi là Event ID. Mã này xác định loại sự kiện và có thể được sử dụng để xác định và phân loại các loại sự kiện.
- **Thời gian (Date/Time):** Thời điểm xảy ra sự kiện được ghi lại trong log. Thông thường, thời gian được ghi dưới dạng ngày, giờ, phút và giây.

- **Level (Mức độ):** Mức độ của sự kiện, thường được biểu diễn bằng các loại như Information, Warning, Error, hoặc Critical. Mức độ này cho biết mức độ quan trọng của sự kiện.
- **Source (Nguồn):** Chương trình hoặc thành phần gây ra sự kiện. Source giúp xác định nơi mà sự kiện đã xảy ra.
- **Category (Danh mục):** Là loại danh mục được gán khi log sinh ra.

Windows Event Logs chia thành ba loại chính:

- **Application Logs:** Ghi lại các sự kiện từ các ứng dụng và chương trình mà người dùng và hệ thống chạy. Điều này bao gồm các thông báo lỗi, cảnh báo và thông tin về việc chạy của các ứng dụng.
- **Security Logs:** Ghi lại các sự kiện liên quan đến bảo mật như đăng nhập, cố gắng đăng nhập không hợp lệ, quyền truy cập tài nguyên, và các hoạt động bảo mật khác. Chỉ những tài khoản người dùng có quyền admin mới có thể xem, xóa hoặc trích xuất logs.
- **System Logs:** Ghi lại các sự kiện hệ thống như khởi động, tắt máy, lỗi hệ thống và các cảnh báo hệ thống khác.

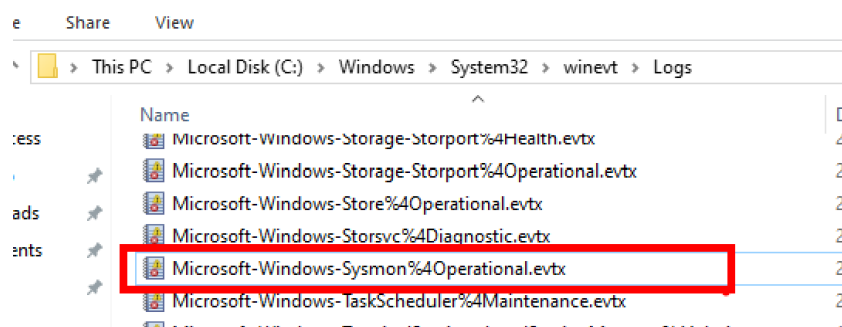
Ngoài ra còn có một số log khác:

- **Setup Logs:** Xác định các bản cập nhật bảo mật Windows, các bản vá lỗi và các hotfix đã được thêm vào hệ thống.
- **Forwarded Events Logs:** Nơi lưu nhật ký liên quan đến các máy tính khác trong mạng tác động tới máy tính user.

❖ Sysmon

Là bộ công cụ do Microsoft phát triển dùng để kiểm tra, giám sát và ghi lại những hoạt động của Windows. Công cụ cung cấp thông tin chi tiết về quá trình tạo lập tiến trình, kết nối mạng, thời gian chỉnh sửa của tệp tin. Bằng việc thu thập các sự kiện Sysmon tạo ra kết hợp với các công cụ quản lý sự kiện bảo mật khác (SIEM), sau đó phân tích chúng, quản trị viên có thể giám sát hoạt động độc hại hoặc bất thường và hiểu cách mà kẻ xâm nhập và phần mềm độc hại hoạt động trên mạng của mình.

Vị trí: Logs của Sysmon lưu trữ ở thư mục mặc định tại thư mục `%SystemRoot%\System32\winevt\Logs`



Hình 3.2. Vị trí lưu trữ logs Sysmon

Công dụng:

- Ghi lại sự kiện tạo tiến trình với đầy đủ dòng lệnh cho cả tiến trình hiện tại và tiến trình cha.
- Ghi lại giá trị băm của các tệp hình ảnh bằng SHA1(mặc định), MD5, SHA256 hoặc IMPHASH, hỗ trợ nhiều giá trị băm cùng một lúc.
- Gắn GUID vào các sự kiện tạo quá trình để liên kết các sự kiện ngay cả khi Windows tái sử dụng ID tiến trình hoặc trong cùng một phiên đăng nhập.
- Ghi lại việc tải các Drivers hoặc DLL với chữ ký và các giá trị băm của chúng.
- Ghi lại sự kiện mở quyền truy cập ổ đĩa và phân vùng.
- Tùy chọn ghi lại các kết nối mạng chi tiết.
- Nhận diện thay đổi thời gian tạo file để phát hiện tệp thực sự được tạo ra.
- Tự động tải lại cấu hình khi registry thay đổi.
- Lọc sự kiện linh hoạt bằng bộ lọc Rule.
- Tạo sự kiện từ quá trình khởi động để phát hiện phần mềm độc hại phức tạp.

File cấu hình trong Sysmon:

- File cấu hình Sysmon được viết dưới định dạng XML, được sử dụng nhằm giúp người dùng định nghĩa những sự kiện muốn lấy, tăng tính hiệu quả của công cụ.

```

<Sysmon schemaversion="4.82">
  <!-- Capture all hashes -->
  <HashAlgorithms>*</HashAlgorithms>
  <EventFiltering>
    <!-- Log all drivers except if the signature -->
    <!-- contains Microsoft or Windows -->
    <DriverLoad onmatch="exclude">
      <Signature condition="contains">microsoft</Signature>
      <Signature condition="contains">windows</Signature>
    </DriverLoad>
    <!-- Do not log process termination -->
    <ProcessTerminate onmatch="include" />
    <!-- Log network connection if the destination port equal 443 -->
    <!-- or 80, and process isn't InternetExplorer -->
    <NetworkConnect onmatch="include">
      <DestinationPort>443</DestinationPort>
      <DestinationPort>80</DestinationPort>
    </NetworkConnect>
    <NetworkConnect onmatch="exclude">
      <Image condition="end with">iexplore.exe</Image>
    </NetworkConnect>
  </EventFiltering>
</Sysmon>

```

Hình 3.3. Minh họa cấu trúc và thành phần cơ bản của file cấu hình Sysmon

- File được biểu diễn với hai thành phần chính:
 - Configuration Entries (Mục cấu hình): hình cho phép quản trị viên điều chỉnh cách Sysmon hoạt động theo cách tùy chỉnh và cho nhu cầu bảo mật cụ thể của tổ chức.
 - Event Filtering Entries (Mục lọc sự kiện): cho phép người dùng có thể lọc chọn lọc sự kiện theo ý của mình. Các bộ lọc được khởi tạo dưới mục EventFiltering và dán nhãn (tag) cho từng sự kiện.

❖ Virustotal

Là một dịch vụ trực tuyến miễn phí cung cấp bởi Google, cho phép người dùng phân tích các tệp tin và URL nghi ngờ để phát hiện virus, sâu máy tính, trojan, và các loại phần mềm độc hại khác. Công cụ này được sử dụng rộng rãi trong cộng đồng bảo mật để xác minh tính an toàn của các tệp tin và trang web

Công dụng:

- Phân tích đa công cụ: Sử dụng hơn 70 công cụ diệt virus để quét tệp tin và URL.
- Tích hợp dễ dàng: Cung cấp API để tích hợp vào các công cụ bảo mật khác.

- Chia sẻ và cộng tác: Kết quả phân tích có thể được chia sẻ công khai để hỗ trợ cộng đồng bảo mật.
- Báo cáo chi tiết: Cung cấp thông tin chi tiết về mối đe dọa, bao gồm loại mối đe dọa và thời gian phát hiện.
- Hỗ trợ nhiều định dạng tệp: Quét nhiều loại tệp tin khác nhau như tệp thực thi, tài liệu, và tệp nén.

❖ **Pfsense**

Là một nền tảng tường lửa và router mã nguồn mở, được xây dựng dựa trên hệ điều hành FreeBSD. Được thiết kế để cung cấp các tính năng bảo mật và quản lý mạng mạnh mẽ, pfSense là một lựa chọn phổ biến cho cả doanh nghiệp và người dùng cá nhân

Công dụng:

- Tường lửa mạnh mẽ: Lọc gói tin chi tiết, thiết lập quy tắc bảo mật phức tạp.
- Router hiệu quả: Hỗ trợ OSPF, BGP, RIP.
- VPN an toàn: Hỗ trợ IPsec, OpenVPN, PPTP.
- IDS/IPS: Tích hợp Snort, Suricata để phát hiện và ngăn chặn xâm nhập.
- Quản lý băng thông: Traffic Shaping, QoS tối ưu hiệu suất.
- Đa WAN: Cân bằng tải, dự phòng kết nối internet.
- Captive Portal: Quản lý truy cập mạng công cộng.
- Quản lý dễ dàng: Giao diện web trực quan, báo cáo chi tiết.

3.3. Triển khai kịch bản thử nghiệm và đánh giá

3.3.1. Phát hiện tấn công brute force

3.3.2. Phát hiện scan port

3.3.3. Phát hiện malware

3.3.4. Đánh giá

❖ Ưu điểm:

- Hệ thống sau khi xây dựng và triển khai đã kiểm tra và phát hiện sớm được đe dọa tiềm ẩn trong hệ thống.

- Giúp các tổ chức giám sát và đảm bảo an toàn thông tin cho hệ thống.
- Giúp đội giám sát sớm phát hiện các cuộc tấn công một cách chủ động hơn.

❖ **Nhược điểm:**

- Quá trình thực nghiệm chưa phân tích đa dạng các loại logs, đầy đủ các cuộc tấn công.
- Quá trình cảnh báo tấn công còn chưa tối ưu.
- Công cụ tấn công đang còn ở mức mô phỏng chưa đạt hiệu suất so với thực tế.

3.4. Kết luận Chương 3

Mô hình thực nghiệm được trình bày trong chương này cho thấy được quá trình phát hiện các cuộc tấn công và đưa ra cảnh báo diễn ra một cách nhanh chóng. Triển khai thử nghiệm hệ thống cùng với một số kịch bản tấn công để từ đó thấy các lợi ích, hiệu quả khi triển khai hệ thống Splunk và phương thức cảnh báo sớm. Việc này giúp cho người vận hành hệ thống sớm phát hiện và đưa ra những xử lý kịp thời nhằm giảm thiểu thiệt hại và đảm bảo sự vận hành của hệ thống.

KẾT LUẬN

Qua ba chương của đề tài đã thể hiện được việc triển khai một hệ thống giám sát tập trung SIEM là một nhu cầu thực tế đang được quan tâm tại các tổ chức hiện nay. Cụ thể:

Ở Chương 1 đã chỉ ra tầm quan trọng của hệ thống quản lý thông tin và sự kiện bảo mật trong việc giám sát và bảo vệ an ninh mạng, có cái nhìn tổng quan về hệ thống SIEM. Tiếp theo, thực hiện tìm hiểu về một số giải pháp SIEM được sử dụng phổ biến hiện nay.

Chương 2 đề cập đến Splunk - một giải pháp SIEM mạnh mẽ và phổ biến với khả năng thu thập, phân tích log và cảnh báo sự kiện an ninh một cách hiệu quả. Splunk thực hiện các công việc tìm kiếm, giám sát và phân tích các dữ liệu lớn được sinh ra từ các ứng dụng, các hệ thống và các thiết bị hạ tầng mạng đáp ứng đúng yêu cầu của một hệ thống SIEM.

Chương 3 đưa ra mô hình triển khai thực nghiệm, xây dựng một số kịch bản bao gồm một chuỗi hành vi tấn công để đánh giá hiệu quả của hệ thống, đồng thời đưa ra các hành động ngăn chặn giúp giải quyết sự cố tạm thời và giảm hậu quả.

Hướng phát triển:

Trong thời gian tới, các kỹ thuật tấn công sẽ trở nên ngày càng tinh vi và phức tạp. Do đó, sau khi hoàn thiện đề tài này, em sẽ tiếp tục phát triển một số hướng nghiên cứu sau:

- Nghiên cứu và phát triển các kỹ thuật tương quan sự kiện: Phát triển các kỹ thuật này sẽ giúp xây dựng các bộ quy tắc nhằm phát hiện cụ thể các cuộc tấn công thay vì chỉ nhận diện các hành vi đáng ngờ, nâng cao độ chính xác và hiệu quả của hệ thống SIEM.
- Nghiên cứu tích hợp các công nghệ và thiết bị an ninh: Việc tích hợp các công nghệ và thiết bị an ninh như IDS/IPS, Endpoint Detection and Response (EDR), Web Application Firewall (WAF),... sẽ tạo nên một hệ sinh thái bảo mật toàn diện hơn, tăng cường khả năng phòng thủ.

TÀI LIỆU THAM KHẢO

- [1] NCS Group, "Tổng kết an ninh mạng Việt Nam năm 2023 và dự báo 2024", 2023, <https://ncsgroup.vn/tong-ket-an-ninh-mang-viet-nam-nam-2023-va-du-bao-2024/>.
- [2] Báo Chính Phủ, "Nhiều vấn đề nổi cộm về an ninh mạng và dự báo năm 2024", 2024, <https://baochinhphu.vn/nhieu-van-de-noi-com-ve-an-ninh-mang-va-du-bao-nam-2024-102240118155026211.htm>.
- [3] Splunk Documentation, <https://docs.splunk.com/Documentation/Splunk>.
- [4] David R. Miller , Shon Harris, Allen Harper , Stephen VanDyke , Chris Blask, "Security Information and Event Management (SIEM) Implementation", McGraw Hill, pp. 78-92, 2010.
- [5] Jit Sinha, "Ultimate Splunk for Cybersecurity: Practical Strategies for SIEM Using Splunk's Enterprise Security (ES) for Threat Detection, Forensic Investigation, and Cloud Security", Orange Education Pvt Ltd, chapter 2, 2024.
- [6] James Miller , "Mastering Splunk", Packt Pub Ltd, chapter 6, 2014.
- [7] David Carasso, "Exploring Splunk", Cito Research, 2012.

PHỤ LỤC

Cài đặt Splunk Server

Tải file .deb dành cho Linux tại

https://www.splunk.com/en_us/download/splunk-enterprise.html

```
l1chow@splunk:~/Downloads$ ls
splunk-9.2.1-78803f08aabb-linux-2.6-amd64.deb
l1chow@splunk:~/Downloads$
```

Cài thành phần phụ thuộc cho Splunk: `sudo apt install curl`

Chạy lệnh sau để cài Splunk:

```
sudo dpkg -i splunk-9.2.1-78803f08aabb-linux-2.6-amd64.deb
```

Sau khi cài đặt hoàn tất, sử dụng lệnh sau để khởi chạy Splunk:

```
sudo /opt/splunk/bin/splunk start --accept-license --
answer-yes
```

Nhập username và password để đăng nhập vào Splunk

```
l1chow@splunk:~/Downloads$ sudo /opt/splunk/bin/splunk start --accept-license --answer-yes
This appears to be your first time running this version of Splunk.

Splunk software must create an administrator account during startup. Otherwise, you cannot log in.
Create credentials for the administrator account.
Characters do not appear on the screen when you type in credentials.

Please enter an administrator username: admin
Password must contain at least:
 * 8 total printable ASCII character(s).
Please enter a new password:
Please confirm new password:
Copying '/opt/splunk/etc/openldap/ldap.conf.default' to '/opt/splunk/etc/openldap/ldap.conf'.
Generating RSA private key, 2048 bit long modulus
```

Sau khi thiết lập hoàn tất, có thể thấy Splunk đang chạy trên <http://127.0.0.1:8000>

```
Waiting for web server at http://127.0.0.1:8000 to be available.....
..... Done

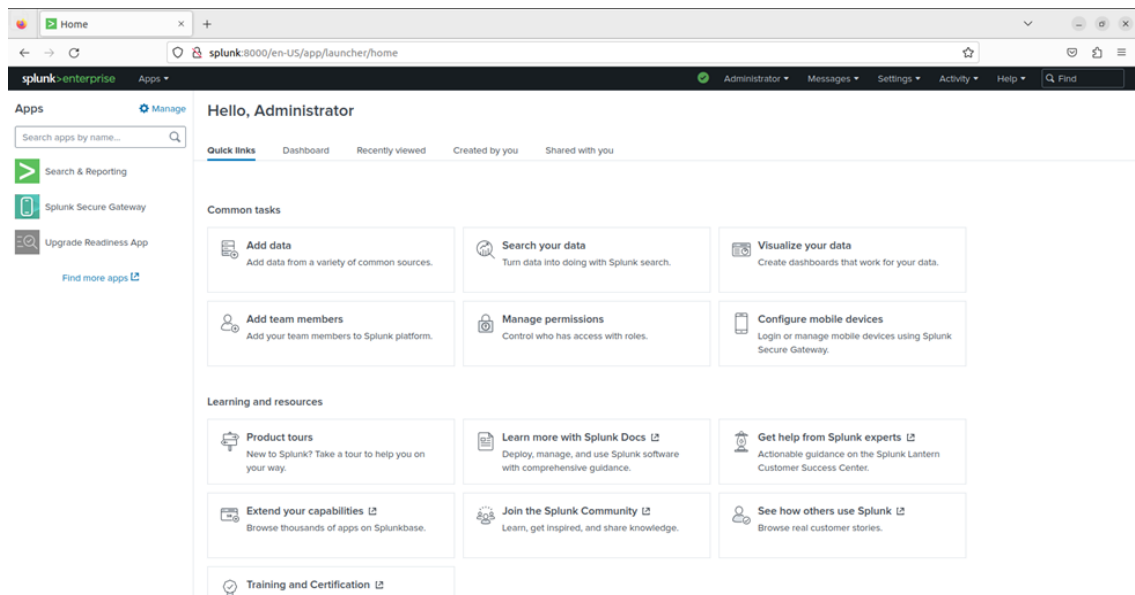
If you get stuck, we're here to help.
Look for answers here: http://docs.splunk.com

The Splunk web interface is at http://splunk:8000
```

Sử dụng lệnh sau để cho phép Splunk tự động khởi động khi hệ thống khởi động:

```
sudo /opt/splunk/bin/splunk enable boot-start
```

Truy cập vào Splunk bằng <http://splunk:8000> hoặc <http://127.0.0.1:8000>



Cấu hình để thu log từ Universal Forwarder

Administrator Messages **Settings** Activity Help Find

Add Data

KNOWLEDGE

Searches, reports, and alerts

Data models

Event types

Tags

Fields

Lookups

DATA

Data inputs

Forwarding and receiving

Indexes

Report acceleration summaries

Virtual indexes

Source types

Forwarding and receiving

Forward data
Set up forwarding between two or more Splunk instances.

Type	Actions
Forwarding defaults	
Configure forwarding	+ Add new

Receive data
Configure this instance to receive data forwarded from other instances.

Type	Actions
Configure receiving	+ Add new

Configure receiving
Set up this Splunk Instance to receive data from forwarder(s).

Listen on this port *

For example, 9997 will receive data on TCP port 9997.

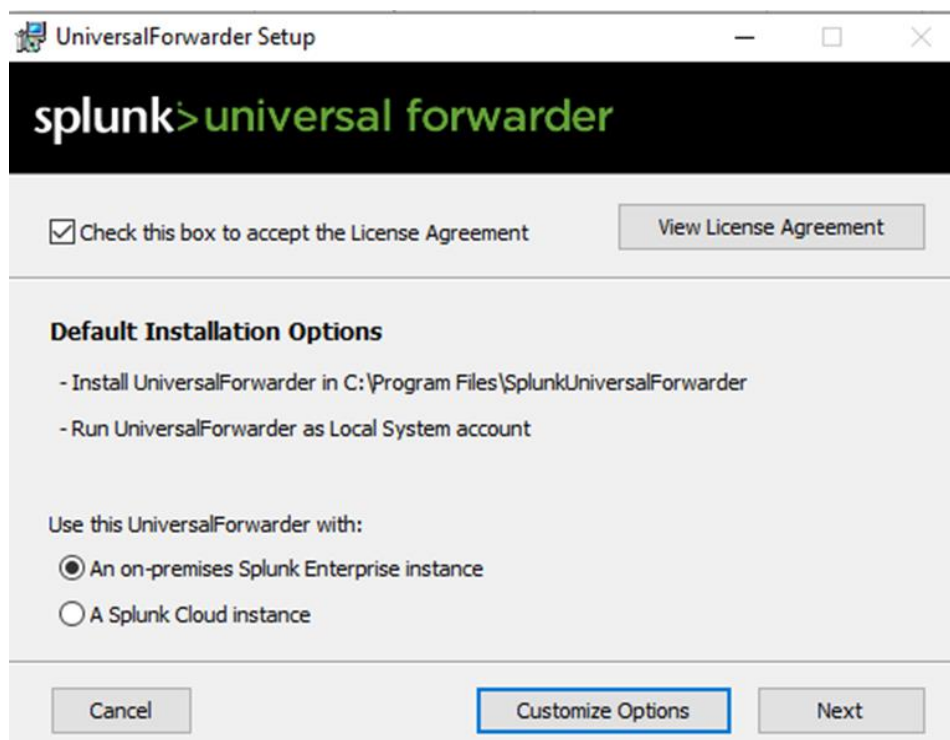
Cancel Save

Cài đặt Splunk Universal Forwarder

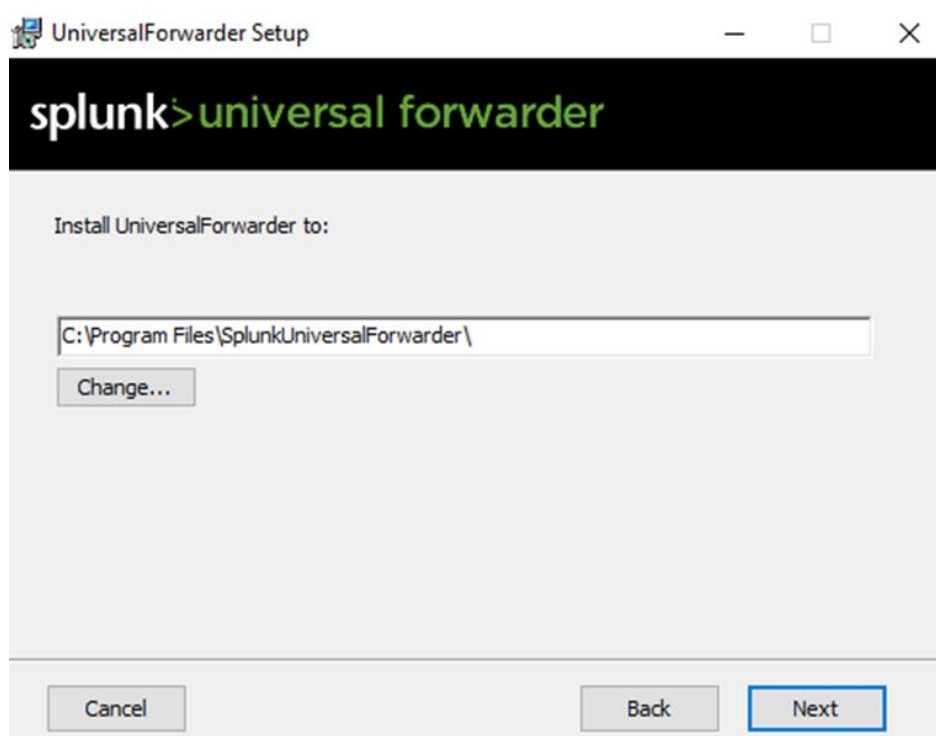
Tải về phiên bản phù hợp với máy tại

https://www.splunk.com/en_us/download/universal-forwarder.html

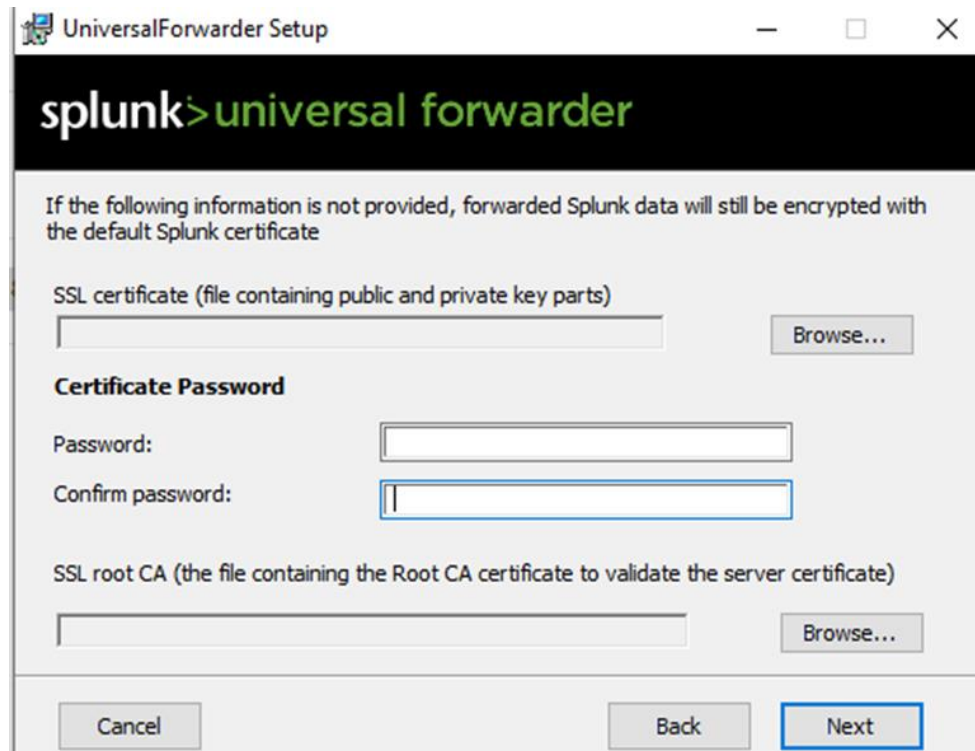
Chạy file .msi và tích vào ô **Check this box to accept the License Agreement**



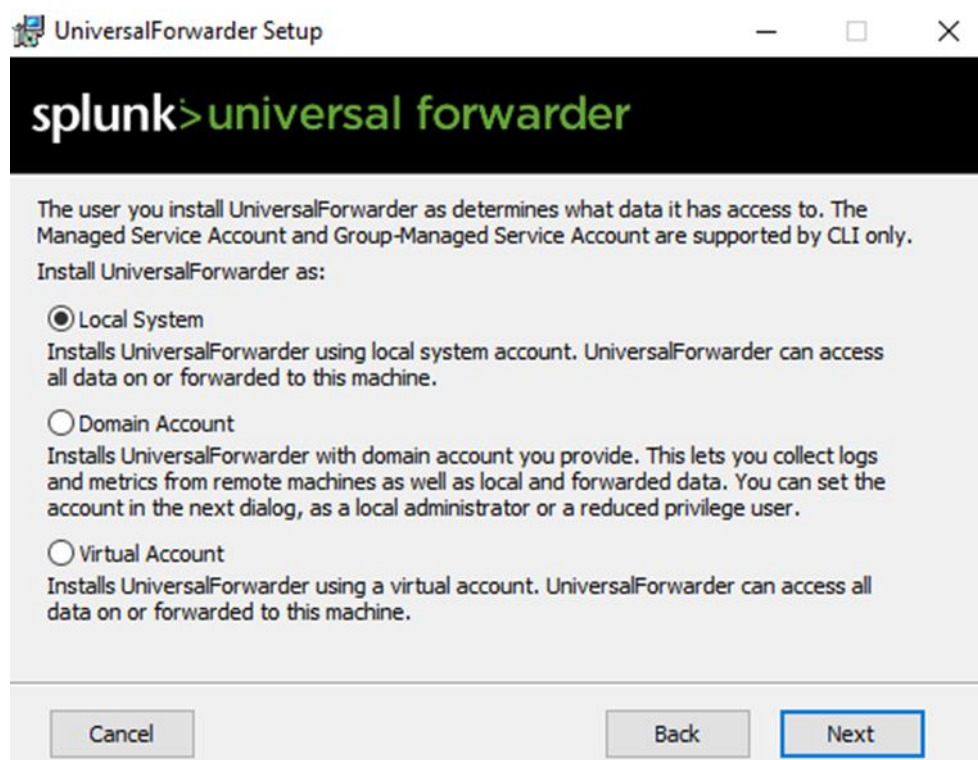
Chọn đường dẫn để cài đặt và **Next**



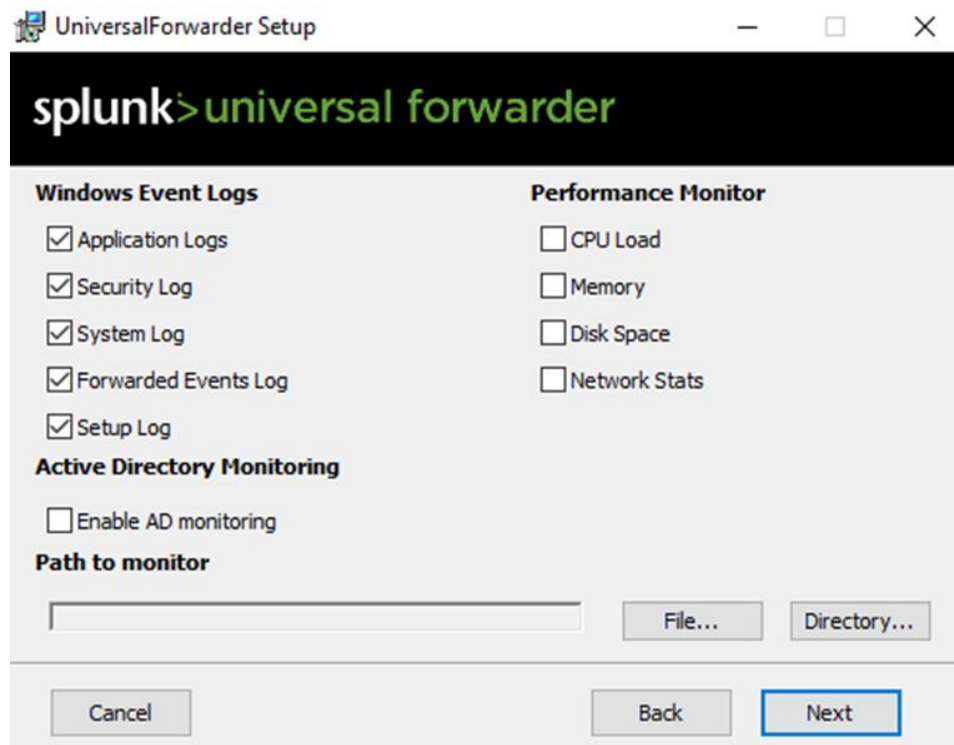
Chọn **Next** để bỏ qua phần cài đặt chứng chỉ SSL



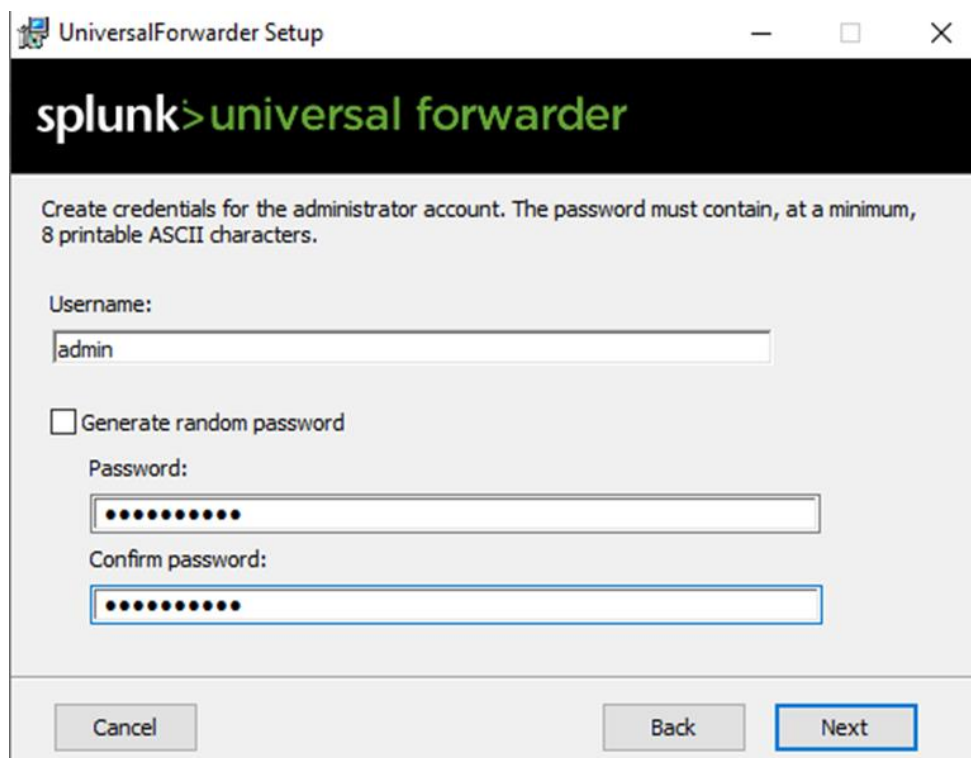
Chọn **Local System** và **Next**



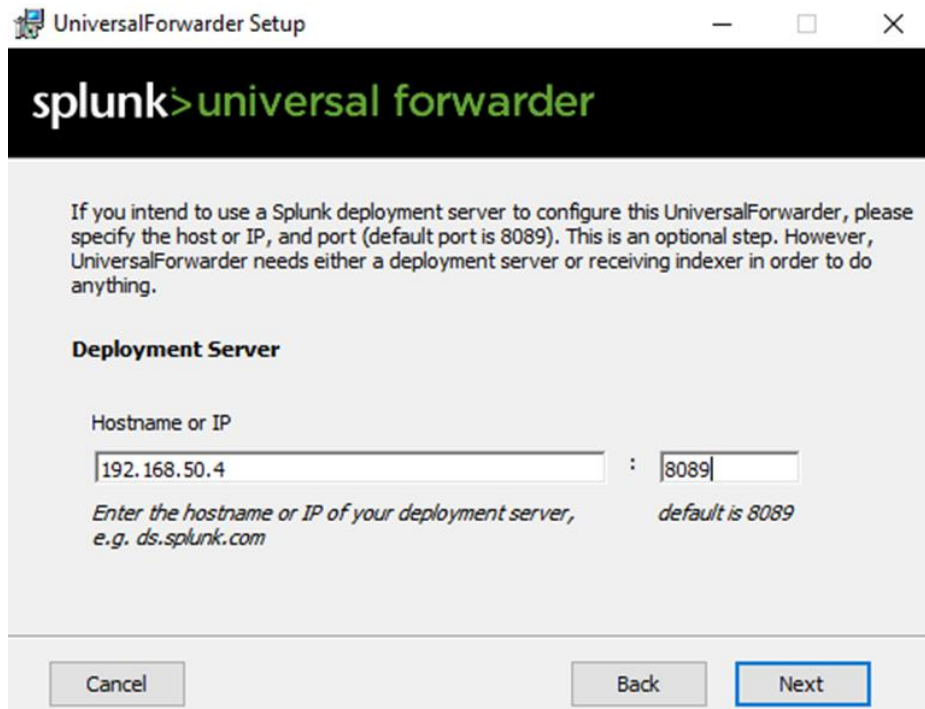
Tích chọn vào các loại logs muốn giám sát



Nhập username và password vừa cấu hình ở Splunk Server

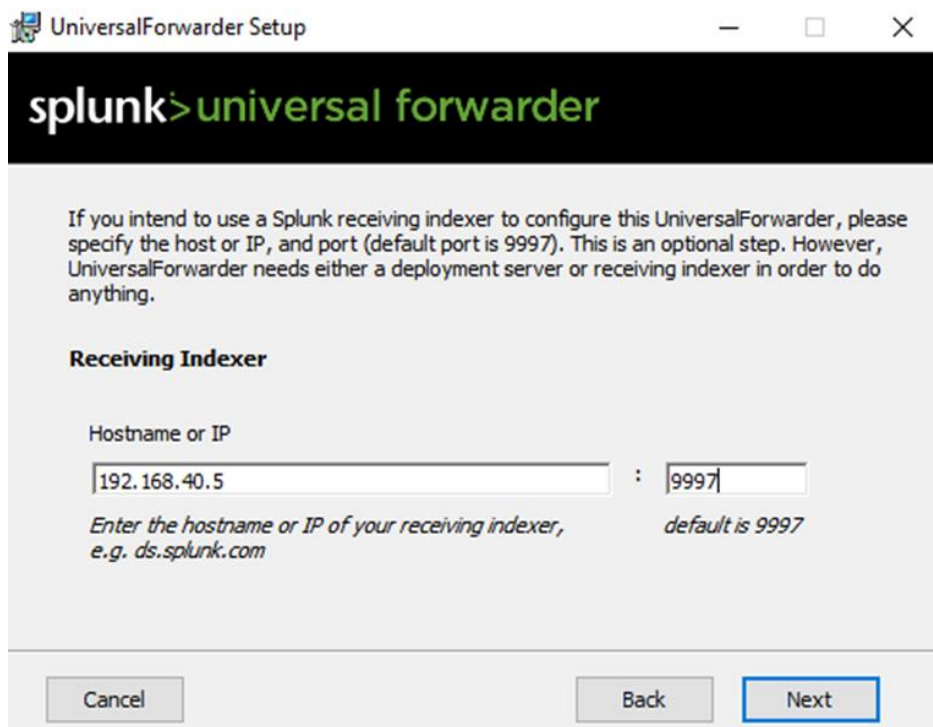


Nhập địa chỉ IP của Splunk server và port là 8089



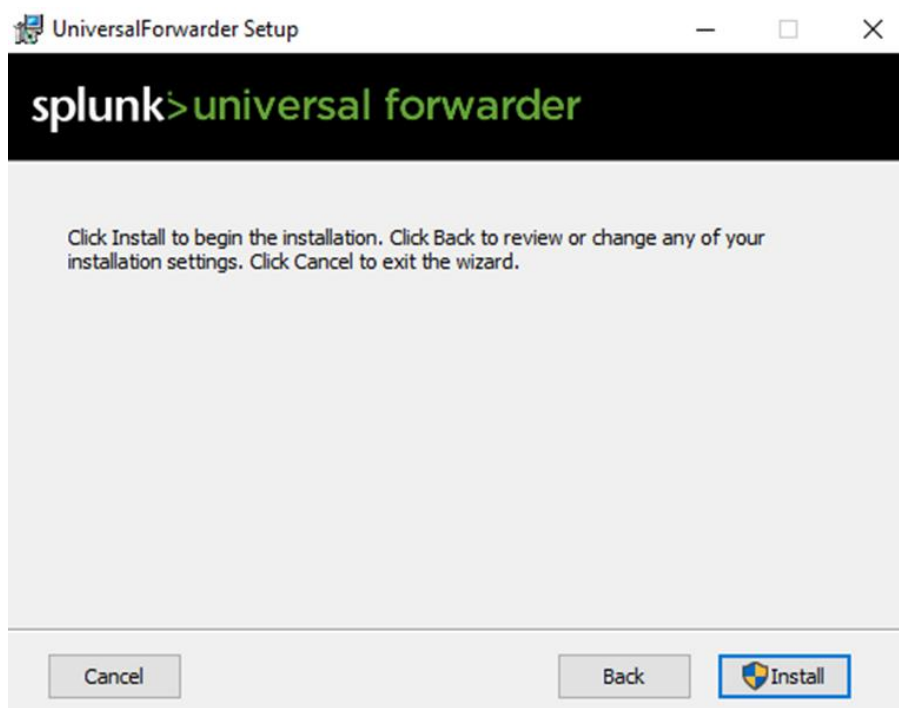
The screenshot shows the 'UniversalForwarder Setup' window. At the top, there's a black header with the 'splunk' logo and 'universal forwarder' in green. Below the header, a text block explains that if a Splunk deployment server is used, the host or IP and port (default 8089) must be specified. The 'Deployment Server' section has a 'Hostname or IP' label. Below it, there are two input fields: the first contains '192.168.50.4' and the second contains '8089', separated by a colon. A note below the fields says 'Enter the hostname or IP of your deployment server, e.g. ds.splunk.com' and 'default is 8089'. At the bottom, there are three buttons: 'Cancel', 'Back', and 'Next' (which is highlighted with a blue border).

Nhập lại địa chỉ IP của Splunk server và port là 9997



The screenshot shows the 'UniversalForwarder Setup' window. At the top, there's a black header with the 'splunk' logo and 'universal forwarder' in green. Below the header, a text block explains that if a Splunk receiving indexer is used, the host or IP and port (default 9997) must be specified. The 'Receiving Indexer' section has a 'Hostname or IP' label. Below it, there are two input fields: the first contains '192.168.40.5' and the second contains '9997', separated by a colon. A note below the fields says 'Enter the hostname or IP of your receiving indexer, e.g. ds.splunk.com' and 'default is 9997'. At the bottom, there are three buttons: 'Cancel', 'Back', and 'Next' (which is highlighted with a blue border).

Chọn **Install** để tiến hành cài đặt



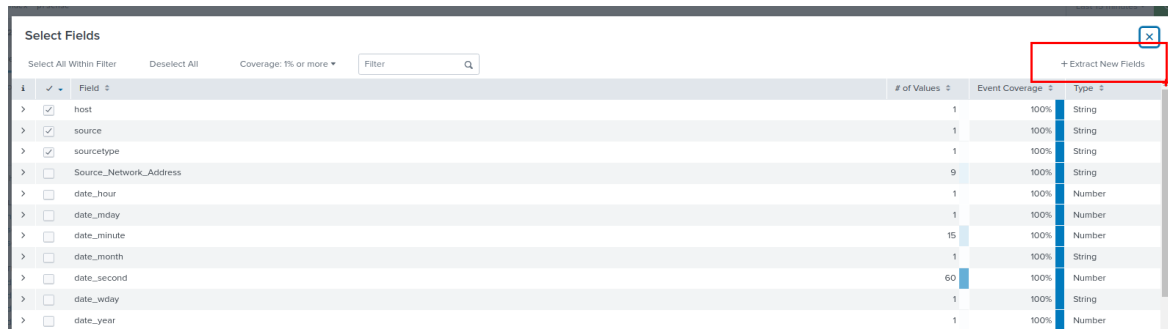
Chọn **Finish** sau khi quá trình cài đặt hoàn tất



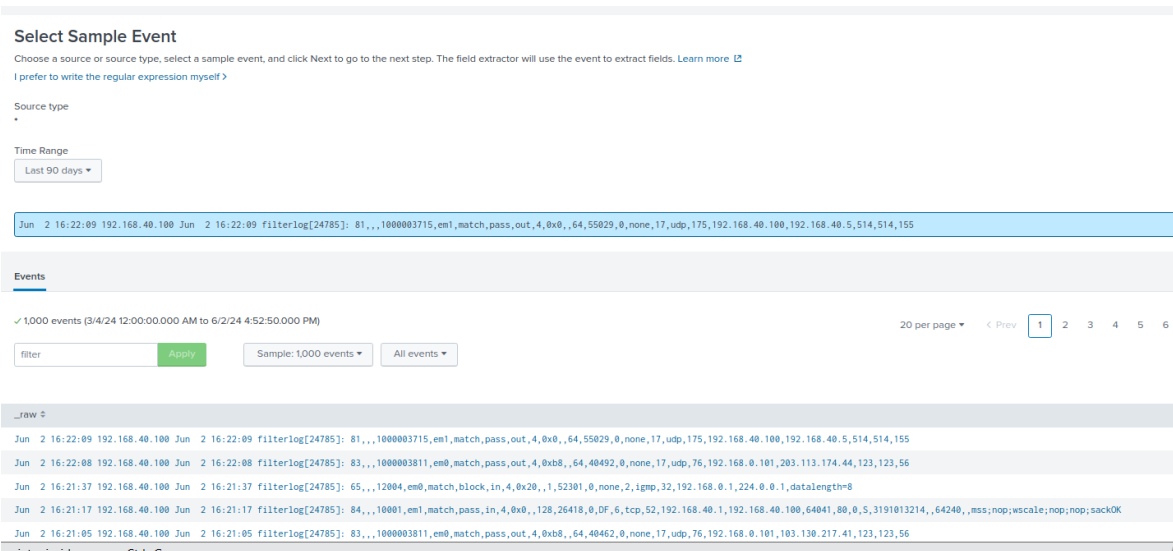
Tạo field mới cho logs tường lửa PfSense

Khi lấy logs lên Splunk Server không phải lúc nào các Field trong sự kiện cũng có sẵn để tiến hành tìm kiếm. Để tạo một Field mới trên các sự kiện có sẵn thì cần phải làm theo các bước sau đây.

Trong phần **Search and Reporting** chọn **Extract Field** sau khi đã tìm kiếm sự kiện chưa gắn Field:



Chọn mẫu sự kiện muốn tạo Field mới:



Chọn Method tạo Field là **Delimiters** rồi nhấn **Next**:

Extract Fields

Select Sample Select Method Select Fields Save

< Back Next >

Learn more [Learn more](#)

log[24785]: 81,,1000003715,em1,match,pass,out,4,0x0,,64,55029,0,none,17,udp,175,192.168.40.100,192.168.40.5,514,514,155

(.*?)
Regular Expression
Splunk Enterprise will extract fields using a Regular Expression.

x|y|z
Delimiters
Splunk Enterprise will extract fields using a delimiter (such as commas, spaces, or characters). Use this method for delimited data like comma separated values (CSV).

Sau đó chọn **Other** và nhập dấu phẩy để phân tách các field

Rename Fields

Select a delimiter. In the table that appears, rename fields by clicking on field names or values. [Learn more](#)

Delimiter

Space Comma Tab Pipe Other ,

field1 field2 field3 field4 field5 field6 field7 field8 field9 field10 field11 field12

Jun 2 16:22:09			1000003715	em1	match	pass	out	4	0x0		64
192.168.40.100											
Jun 2 16:22:09											
filterlog[24785]:											
81											

Chọn field muốn tạo và đổi tên field rồi ấn Next



Đặt tên cho extraction và đặt quyền cho nó

Save
Name the extraction and set permissions.

Extractions Name **REPORT-**

Owner **admin**

App **search**

Permissions

Thử tìm kiếm với field mới đã tạo:

Search > Analytics > Datasets > Reports > Alerts > Dashboards

Scan port Save Save As View Create Table View Close

Index="pfame" dst_ip="192.168.40.10"

100 of 995 events matched No saved sampling

200 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000

Events (995) Patterns Statistics Visualization

Format Timeline Zoom Out Zoom In Search Disabled Fields per column

1 2 3 4 5 6 7 8 9 Next

Time	Event
6/2/24 4:17:00 PM	Jun 2 16:17:11 192.168.40.100 Jun 2 16:17:11 Filtering[24785]: 81,...,1000007715,net_match_pass_out,4,840,...,63,4932,8,0F,8,top,68,192.168.140.58,192.168.40.10,58714,445,8,5,394792955,...,32120,...,ms
6/2/24 4:17:00 PM	Jun 2 16:17:11 192.168.40.100 Jun 2 16:17:11 Filtering[24785]: 81,...,1000007715,net_match_pass_out,4,840,...,63,43608,8,0F,8,top,68,192.168.140.58,192.168.40.10,58454,3389,8,5,166832016,...,32120,...,ms
6/2/24 4:17:00 PM	Jun 2 16:17:11 192.168.40.100 Jun 2 16:17:11 Filtering[24785]: 81,...,1000007715,net_match_pass_out,4,840,...,63,25301,8,0F,8,top,68,192.168.140.58,192.168.40.10,58474,3389,8,5,2096216189,...,32120,...,ms
6/2/24 4:17:00 PM	Jun 2 16:17:11 192.168.40.100 Jun 2 16:17:11 Filtering[24785]: 81,...,1000007715,net_match_pass_out,4,840,...,63,30981,8,0F,8,top,68,192.168.140.58,192.168.40.10,58726,445,8,5,4195516211,...,32120,...,ms
6/2/24 4:17:00 PM	Jun 2 16:17:11 192.168.40.100 Jun 2 16:17:11 Filtering[24785]: 81,...,1000007715,net_match_pass_out,4,840,...,63,8816,8,0F,8,top,68,192.168.140.58,192.168.40.10,58478,3389,8,5,718189540,...,32120,...,ms

SELECTED FIELDS
host 1
source 1
sourcetype 1

INTERESTING FIELDS
date_hour 1
date_minute 1
date_second 2
date_month 1
date_year 1
dst_ip 1
dst_ip 100-
field 10
field2 1

BẢNG PHÂN CÔNG

Nội dung		Châu	Mai	Tuấn
Chương 1. Tổng quan về SIEM	1.1. Tổng quan về tình hình an ninh mạng		x	
	1.2. Giới thiệu về hệ thống SIEM		x	
	1.3. Một số giải pháp SIEM	x		
Chương 2. Tìm hiểu công cụ giám sát an ninh mạng Splunk	2.1. Tổng quan về Splunk		x	
	2.2. Ứng dụng của Splunk		x	
	2.3. Các thành phần của Splunk	x		x
	2.4. Kiến trúc Splunk	x		
Chương 3. Triển khai thực nghiệm	3.1. Mục tiêu thực nghiệm			x
	3.2. Xây dựng môi trường			x
	3.3. Triển khai và đánh giá	x	x	x
Sửa báo cáo		x		
Làm Slide		x		x