HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

NGUYỄN CÔNG TÙNG

NGHIÊN CỨU GIẢI PHÁP AN TOÀN THÔNG TIN VÀ ỨNG DỤNG TẠI VIỆN KHCN SÁNG TẠO VIỆT NAM

LUẬN VĂN THẠC SĨ KỸ THUẬT

(***Theo định hướng ứng dụng***)

HÀ NỘI – NĂM 2020

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

NGUYỄN CÔNG TÙNG

NGHIÊN CỨU GIẢI PHÁP AN TOÀN THÔNG TIN VÀ ỨNG DỤNG TẠI VIỆN KHCN SÁNG TẠO VIỆT NAM

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

LUẬN VĂN THẠC SĨ KỸ THUẬT

(***Theo định hướng ứng dụng***)

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. NGUYỄN TẤT THẮNG

HÀ NỘI – NĂM 2020

LỜI CAM ĐOAN

Tôi cam đoan đề tài: ***"Nghiên cứu giải pháp an toàn thông tin và ứng dụng tại Viện KHCN Sáng tạo Việt Nam"*** là công trình nghiên cứu của riêng tôi dưới sự hướng dẫn của **TS. Nguyễn Tất Thắng.**

Những phân tích, kết luận, kết quả trong luận văn này đều là kết quả của tác giả, số liệu nêu ra là trung thực và chưa từng được công bố trong bất kỳ công trình nào khác.

Hà Nội, ngày 10 tháng 11 năm 2020

Tác giả

Nguyễn Công Tùng

LỜI CẢM ƠN

Lời đầu tiên cho tôi xin gửi lời cảm ơn chân thành đến các thầy, cô giáo của Học viện Công nghệ Bưu chính Viễn thông đã tận tình chỉ bảo, hướng dẫn, giúp đỡ tôi trong suốt quá trình thực hiện luận văn này.

Tôi xin gửi lời cảm ơn chân thành đặc biệt tới thầy hướng dẫn khoa học **TS. Nguyễn Tất Thắng**, tận tình chỉ bảo và hướng dẫn, đưa ra định hướng đúng đắn giúp em hoàn thành được luận văn này.

Xin trân trọng cảm ơn các cảm ơn tập thể lớp Cao học hệ thống thông tin khoá 2019-2021 đã đồng hành, khích lệ và chia sẻ trong suốt quá trình học tập và làm luận văn.

Trong quá trình thực hiện luận văn, mặc dù bản thân đã cố gắng, chủ động sưu tầm tài liệu, củng cố kiến thức… tuy nhiên khó có thể tránh khỏi những thiếu sót, hạn chế. Rất mong nhận được sự chỉ dạy, góp ý của các thầy, cô giáo và các bạn cùng lớp để luận văn được hoàn thiện hơn nữa và có tính ứng dụng cao hơn trong thực tiễn.

Xin trân trọng cảm ơn!

Hà Nội, ngày 10 tháng 11 năm 2020

Học viên

Nguyễn Công Tùng

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng Việt
AI	Artificial Intelligence	Trí tuệ nhân tạo
APT	Advanced Persistent Threat	Mối đe dọa liên tục nâng cao
ATTT	Safety information	An toàn thông tin
CNTT	Information Technology	Công nghệ thông tin
IDS	Intrusion Detection System	Hệ thống phát hiện xâm nhập
IPS	Intrusion Prevention System	Hệ thống ngăn chặn xâm nhập
KHCN	Science and technology	Khoa học công nghệ
LAN	Local Area Network	Mạng lưới khu vực địa phương
SIEM	Security Information and Event Management	Thông tin bảo mật và quản lý sự kiện
VPN	Virtual Private Network	Mạng riêng ảo

DANH SÁCH CÁC BẢNG

DANH MỤC CÁC HÌNH

MỞ ĐẦU

1. Lý do chọn đề tài

Trong những năm gần đây, công nghệ thông tin (CNTT) là một trong những lĩnh vực phát triển nhanh chóng, toàn diện và được ứng dụng rộng rãi trong tất cả các lĩnh vực đời sống, xã hội. Khi các giá trị từ hệ thống CNTT mang lại ngày càng lớn, các nguy cơ bị hacker tấn công ngày càng cao.

Hiện này đã có nhiều giải pháp bảo đảm an toàn thông tin cho hệ thống CNTT đã được quan tâm nghiên cứu và triển khai. Tuy nhiên, thực tế, vẫn thường xuyên có các hệ thống bị tấn công, bị đánh cắp thông tin, phá hoại gây ra những hậu quả vô cùng nghiêm trọng đối với nhiều doanh nghiệp, cơ quan nhà nước cũng như toàn xã hội.

Theo báo cáo thống kê của Microsoft, Việt Nam là những nước đứng đầu trong 05 nước toàn cầu về nguy cơ nhiễm mã độc. Khu vực Đông Nam Á có 02 nước là Việt Nam và Indonesia. Cả hai nước có tỷ lệ bị nhiễm mã độc rơi vào khoảng 46% ở quý II/2016, cao gấp đôi so với trung bình 21% toàn thế giới [4].

Tại Việt Nam, Cục An toàn thông tin – Bộ Thông tin & Truyền thông đã ghi nhận trong năm 2018 có 10.220 cuộc tấn công mạng vào các hệ thống thông tin tại Việt Nam. Trong 6 tháng đầu năm 2019 đã có tổng số 3.159 cuộc tấn công mạng vào các hệ thống thông tin tại Việt Nam [26].

Trước những thực trạng cấp thiết đó, học viên xin chọn đề tài ***"Nghiên cứu giải pháp an toàn thông tin và ứng dụng tại Viện KHCN Sáng tạo Việt Nam"*** làm đề tài luận văn nhằm nghiên cứu, đưa ra các giải pháp giám sát an toàn thông tin trong giai đoạn hiện nay.

2. Tổng quan về đề tài nghiên cứu

Luận văn nghiên cứu giải pháp giám sát an toàn thông tin dựa trên SIEM (Security Information and Event Management) là hệ thống được thiết kế nhằm thu thập và phân tích nhật ký, các sự kiện an toàn thông tin từ các thiết bị đầu cuối và được lưu trữ tập trung. Hệ thống SIEM cho phép phân tích tập trung và báo cáo về các sự kiện an toàn thông tin của tổ chức, phát hiện thông qua các bộ luật tương quan (correlation rule).

Hệ thống SIEM có thể phục vụ rất nhiều công việc như: Quản lý tập trung, giám sát an thông tin mạng, cải thiện hiệu quả trong phục sự cố. Trong Luận văn sẽ tập trung tìm hiểu, phân tích, nghiên cứu chủ đề chính là giám sát an toàn thông tin.

Giám sát an toàn thông tin là việc sử dụng một hệ thống để liên tục theo dõi một số thông tin, xem xét tình trạng hoạt động của các thiết bị, dịch vụ hệ thống đó, cảnh báo cho quản trị viên trường hợp mạng không hoạt động hoặc có các sự cố khác (tắc nghẽn, sập,...), hành vi tấn công (dựa trên tập luật đã được cấu hình), hành vi bất thường ....

Thông thường một hệ thống CNTT tối thiểu cần có máy chủ (server), đường truyền, các thiết bị kết nối (Repeater, Hub, Switch, Bridge,...), máy tính người dùng (client), card mạng (Network Interface Card – NIC) để kết nối các máy tính lại với nhau.

Một hệ thống giám sát gồm có nhiều thành phần:

- Log Source (Nguồn nhật ký/sự kiện)
- Event Collector: Thành phần thu thập nhật ký/sự kiện.
- Event Processor (Xử lý nhật ký/sự kiện)
- Magistrate (Thành phần lõi xuất ra các báo cáo, cảnh báo ATTT).

Các sự kiện diễn ra trong các thiết bị đều được ghi lại trong "log". Nhiệm vụ của hệ thống giám sát ATTT là sử dụng Event Collector thu thập log từ Log Source (thành phần có log) và gửi về cơ sở dữ liệu trung tâm. Event Processor phân tích các sự kiện được gửi về và báo cho quản trị viên để có các hành động ứng phó thích hợp.

Giải pháp giám sát an toàn thông tin có khả năng phân tích, cảnh báo thời gian thực các sự cố, nguy cơ mất ATTT đối với hệ thống. Với giải pháp này, hệ thống quản lý sẽ được bảo đảm ATTT ở mức cao hơn. Và để hiểu rõ giải pháp giám sát an toàn thông tin SIEM, cần phải nghiên cứu cả về mặt lý thuyết lẫn triển khai ứng dụng.

3. Mục tiêu nghiên cứu của đề tài

Mục tiêu nghiên cứu của luận văn là khảo sát các yêu cầu và giải pháp an toàn thông tin. Để đưa ra giải pháp an toàn thông tin cho Viện KHCN Sáng tạo Việt Nam có khả năng triển khai áp dụng trong thực tế.

4. Đối tượng và phạm vi nghiên cứu

- Đối tượng nghiên cứu: Luận văn nghiên cứu về giải pháp an toàn thông tin và các vấn đề liên quan tới giải pháp an toàn thông tin. Trong đó, Luận văn tập trung vào nghiên cứu giải pháp Splunk trong việc xây dựng hệ thống giám sát, đảm bảo an toàn thông tin. Cách thức chuẩn hóa sự kiện an toàn thông tin và đưa ra cảnh báo

- Phạm vi nghiên cứu: Luận văn nghiên cứu một cách tổng quan về giải pháp an toàn thông tin; đặc điểm, ưu điểm và nhược điểm của hệ thống. Nghiên cứu các giải pháp xây dựng hệ thống; các vấn đề an toàn thông tin tại Viện KHCN Sáng tạo Việt Nam và các giải pháp đảm bảo an toàn thông tin hiện nay.

5. Phương pháp nghiên cứu của đề tài

- ***Về mặt lý thuyết:*** Thu thập, khảo sát, phân tích các tài liệu liên quan đến giải pháp toàn thông tin.

- Về mặt thực nghiệm: Khảo sát hệ thống CNTT của Viện KHCN Sáng tạo Việt Nam và ứng dụng giải pháp an toàn thông tin tại Viện.

6. Bố cục luận văn

Luận văn được trình bày trong 3 chương:

Chương 1 của luận văn sẽ khảo sát tổng quan về tình hình an toàn thông tin và các mối đe dọa an toàn thông tin.

Chương 2 của luận văn tập trung nghiên cứu các giải pháp an toàn thông tin, từ đó sẽ đưa ra giải pháp an toàn thông tin

Chương 3 của luận văn tập trung nghiên cứu về hệ thống mạng Viện KHCN Sáng tạo và đề xuất ứng dụng giải pháp an toàn thông tin thông qua nghiên cứu từ chương 2 cho hệ thống CNTT của Viện KHCN Sáng tạo Việt Nam.

CHƯƠNG 1. TỔNG QUAN VỀ AN TOÀN THÔNG TIN

Chương 1 của luận văn sẽ khảo sát tổng quan về tình hình an toàn thông tin và các mối đe dọa an toàn thông tin, những yêu cầu, khái niệm cơ bản về giám sát hệ thống mạng, cách ứng dụng cũng như các yêu cầu chung khi triển khai một hệ thống giám sát an toàn thông tin. Nội dung chính của chương 1 bao gồm:

1.1 Tổng quan chung về tình hình an toàn thông tin

Ngày nay ở Việt Nam, các tổ chức, doanh nghiệp đều xây dựng, vận hành một hệ thống mạng của riêng mình. Hệ thống mạng giúp gia tăng khả năng làm việc giữa các nhân viên, các đơn vị với nhau, gia tăng hiệu suất và giúp cơ quan, tổ chức hoạt động một cách hiệu quả. Tuy nhiên, khi vận hành hệ thống mạng, có rất nhiều vấn đề có thể phát sinh làm ảnh hưởng đến khả năng hoạt động của hệ thống. Những vấn đề đó đến từ nhiều nguyên nhân khác nhau, có thể là hỏng hóc máy móc thiết bị hay lỗi do người dùng tạo ra. Hệ thống càng lớn, các hoạt động diễn ra bên trong hệ thống phức tạp, các vấn đề nảy sinh cũng càng tăng theo. Do đó, hệ thống mạng luôn cần có một hệ thống giám sát an toàn thông tin bao quát toàn bộ các hoạt động, các vấn đề, có thể túc trực, quản lý, dễ dàng phát hiện các sự cố xảy ra bên trong hệ thống, thông qua đó quản trị viên sẽ đưa ra các biện pháp ứng phó.

Từ tình hình trên, việc xây dựng hệ thống giám sát an toàn thông tin để quản lý hệ thống mạng đang ngày càng trở nên cấp thiết hơn bao giờ hết.

1.2. Các mối đe dọa an toàn thông tin và phương thức tấn công mạng

1.2.1 Các mối đe dọa an toàn thông tin

- Mối đe dọa không có cấu trúc:

Là những hành vi xâm nhập mạng trái phép một cách đơn lẻ, không có tổ chức. Kiểu tấn công này có rất nhiều công cụ trên internet có thể hack và rất nhiều script có sẵn. Chỉ cần ai muốn tìm hiểu có thể tải chúng về và sử dụng thử để nghiên cứu trên mạng nội bộ của công ty. Rất nhiều người lại thích với việc tấn công và xâm nhập vào máy tính và thử thách các hành động vượt tường lửa đi ra khỏi tầm bảo vệ. Đa phần tấn công không có cấu trúc đều được gây ra bởi Script Kiddies hay những người có trình độ thấp hoặc vừa phải vừa phải. Những cuộc tấn công đó có thể do sở thích cá nhân, nhưng đôi khi có nhiều cuộc tấn công có ý đồ xấu để lấy cắp thông tin. Các trường hợp đó sẽ có ảnh hưởng nghiêm trọng đến hệ thống và các chủ thể sở hữu mạng. Thậm chí, có một đoạn mã độc là có thể phá hủy chức năng của mạng nội bộ.

- Mối đe dọa có cấu trúc:

Là những cách thức tấn công hoặc xâm nhập hệ thống mạng trái phép, có động cơ và kỹ thuật cao. Kiểu này hoạt động độc lập hoặc theo từng nhóm, chúng thường có kỹ năng phát triển ứng dụng và sử dụng các kỹ thuật phức tạp nhằm xâm nhập vào mục tiêu có chủ đích. Mục đích của các hình thức tấn công này có thể vì tiền hoặc hoạt động chính trị, đôi khi là tức giận hay báo thù. Các nhóm tội phạm, các đối tác, đối thủ cạnh tranh hay các tổ chức sắc tộc thuê các hacker để thực hiện các cuộc tấn công, kiểm soát dạng structured threat. Những cuộc tấn công vào hệ thống thường có nhiều mục đích từ trước, qua đó có thể lấy được mã nguồn của những đối thủ cạnh tranh với nhau.

Các cuộc tấn công như vậy rất có thể gây hậu quả nghiêm trọng, có thể gây nên sự phá hủy cho toàn hệ thống mạng của doanh nghiệp hoặc các tổ chức.

- Mối đe dọa từ bên ngoài:

Là những cuộc tấn công được tạo ra khi Hacker không có một quyền nào kiểm soát trong hệ thống. Người dùng có thể bị tấn công trên toàn thế giới thông qua mạng Internet. Những mối đe dọa từ bên ngoài này thường là mối đe dọa nguy hiểm, các chủ doanh nghiệp sở hữu mạng LAN thường phải bỏ rất nhiều tiền và thời gian để bảo vệ hệ thống.

- Mối đe dọa từ bên trong hệ thống:

Là kiểu tấn công được thực hiện từ một cá nhân hoặc một tổ chức có một số quyền truy cập vào hệ thống mạng nội bộ của công ty. Những cách tấn công này thường từ bên trong, được thực hiện từ một vị trí tin cậy trong mạng nội bộ, rất khó phòng chống bởi đôi khi chính là các nhân viên truy cập mạng rồi tấn công. Nhưng nếu có hệ thống giám sát và phân tích sẽ rất dễ bắt được các đối tượng này.

1.2.2 Những cách thức tấn công hệ thống mạng máy tính

- Cách thức lấy cắp thông tin bằng kiểu tấn công Packet Sniffers:

Chương trình ứng dụng này tạo ra dùng để bắt giữ các các gói tin lưu chuyển trên hệ thống mạng hoặc trên một miền mạng riêng. Kiểu Sniffer thường được dùng phân tích lưu lượng (traffic). Nếu một số ứng dụng không mã hóa mà gửi dữ liệu dưới dạng clear text (telnet, POP3, FTP, SMTP,...) thì phần mềm sniffer cũng là một công cụ giúp cho hacker bắt được các thông tin nhạy cảm như là username, password, từ đó có thể đăng nhập vào các hệ thống máy chủ.

- Cách thức lấy cắp mật khẩu bằng Password attack:

Hacker thường tấn công lấy cắp mật khẩu bằng các phương pháp như: kiểu brute-force attack, chương trình Trojan Horse, hoặc IP spoofing và packet sniffer. Đối với kiểu dùng packet sniffer hoặc IP spoofing có thể lấy được tài khoản và mật khẩu (user account và password), tuy nhiên các Hacker lại thường sử dụng kiểu brute-force để lấy tài khoản và mật khẩu. Cách thức tấn công brute-force được thực hiện bằng phương pháp dùng một chương trình chạy trên hệ thống mạng, sau đó cố gắng login vào các phần chia sẻ tài nguyên trên máy chủ.

- Cách thức tấn công bằng Mail Relay:

Phương pháp này rất phổ biến hiện nay. Nếu máy chủ chạy dịch vụ Email không cấu hình theo chuẩn hoặc tài khoản và mật khẩu của người dùng sử dụng mail bị lộ. Các Hacker thường lợi dụng máy chủ Email để gửi rất nhiều mail cùng lúc gây ngập băng thông mạng và phá hoại các hệ thống email khác. Đặc biệt kiểu gắn thêm những đoạn script trong mail, các hacker có thể gây ra các cuộc tấn công Spam đồng thời với khả năng tấn công gián tiếp đến các máy chủ chứa Database nội bộ của công ty hoặc các cuộc tấn công DoS vào một mục tiêu nào đó có chủ đích.

- Cách thức tấn công tầng ứng dụng:

Hacker tấn công vào tầng ứng dụng được thực hiện bằng rất nhiều cách khác nhau. Những cách thông dụng nhất thường là tấn công vào các điểm yếu của các phần mềm như HTTP hoặc FTP. Các nguyên nhân chủ yếu của các cuộc tấn công tầng ứng dụng là chúng sử dụng các cổng được mở bởi tưởng lửa của hệ thống. Ví dụ Hacker thường tấn công dịch vụ Web server bằng cách sử dụng một số phần mềm quét port 80 sau đó tấn công hoặc dịch vụ mail server qua port 25.

- Cách thức tấn công bằng Virus và phần mềm Trojan Horse:

Những nguy hiểm của các máy workstation và người dùng đầu cuối là những tấn công virus và Trojan (thường gọi là Trojan horse). Phần mềm Virus thường là có hại, chúng được đính kèm vào các chương trình thực thi để thực hiện một cách thức phá hại nào đó. Còn phần mềm Trojan horse thì hoạt động theo kiểu gián điệp, nghe lén và lấy cắp thông tin.

1.3 Giới thiệu tổng quan về hệ thống SIEM

1.3.1 Tổng quan về SIEM

SIEM là viết tắt của cụm từ Securit Information and Event Management được hiểu đơn giản là giải pháp quản lý và phân tích sự kiện an toàn thông tin. Là một hệ thống giám sát an ninh mạng tân tiến nhất hiện nay. Nó tiến hành một loạt

hoạt động như thu thập, phân tích, đánh giá nhật ký từ mọi thiết bị trong hệ thống… Từ đó cho phép cho chúng ta phân tích một lượng lớn dữ liệu để phát hiện các cuộc tấn công ẩn dấu đằng sau để các đơn vị, cơ quan có được cái nhìn toàn cảnh về các sự kiện an ninh mạng.

Hệ thống SIEM là giải pháp kết nối giữa hai giải pháp SIM và SEM. Trong đó SIM thực hiện việc thu thập nhật ký và phân tích đồng thời đưa ra cảnh báo. Nhật ký này được lấy từ máy chủ, ứng dụng, thiết bị mạng, thiết bị chuyên về bảo mật. Nó hỗ trợ việc theo dõi, giám sát hành động người dùng

SIEM thực hiện việc xử lý nhật ký và các sự kiện an ninh được gửi về từ các thiết bị các thiết bị mạng, các máy chủ (Server), các ứng dụng. SIEM giúp theo dõi sự kiện anh ninh của hệ thống và thực hiện các hành động bảo vệ an toàn hệ thống. Nó gồm 2 thành phần chính thu thập nhật ký, thành phần phân tích nhật ký.

Giải pháp SIEM được xem là cách toàn diện, hoàn chỉnh và hiệu quả giúp các cơ quan tổ chức thực hiện việc giám sát an toàn thông tin cho hệ thống. Đây là giải pháp được ngày càng nhiều doanh nghiệp tổ chức áp dụng nhằm đảm bảo an toàn tuyệt đối và nhất quán, linh hoạt trong việc lắp đặt và sử dụng thiết bị cho hệ thống an ninh mạng công nghệ thông tin.

1.3.2. Chức năng chính của SIEM

- Quản lý tập trung: SIEM giúp tập hợp các dữ liệu thông qua giải pháp nhật ký tập trung. Thiết bị đầu cuối của hệ thống thường ghi lại và truyền dữ liệu nhật ký về máy chủ SIEM. Máy chủ SIEM nhận nhật ký từ nhiều máy và tiến hành thống kê, phân tích và tạo ra một báo cáo duy nhất. Nhờ có hệ thống này mà giúp tiết kiệm công sức trong việc tập trung dữ liệu và báo cáo an ninh định kỳ.

- Giám sát an toàn mạng: Đây chính là chức năng chính của SIEM, hệ thống sẽ phát hiện được các sự cố mà các thiết bị thông thường không phát hiện được. Cùng với đó nó có thể cho thấy sự tương quan giữa các thiết bị với nhau. Hệ thống SIEM sẽ thấy được những phần khác nhau của các cuộc tấn công bởi Hacker thông qua các thiết bị khác nhau. SIEM sẽ tiến hành kiểm tra và cách ly máy chủ mục tiêu của cuộc tấn công.

- Giúp ích cho việc xử lý sự cố: SIEM có giao diện đơn giản để có thể xem tất cả nhật ký từ nhiều thiết bị một cách thuận tiện để khắc phục sự cố một cách dễ dàng và hiệu quả.

1.3.3. Các thành phần của hệ thống

Việc xây dựng hệ thống SIEM có thể tiến hành theo nhiều cách, thường gồm 3 thành phần chính như sau.

- Thu thập nhật ký ATTT: Phần thu thập ATTT gồm các giao diện có chức năng thu thập nhật ký từ mọi thiết bị. Sau khi tập hợp nó sẽ gửi toàn bộ nhật ký về thành phần phân tích.

- Phân tích và lưu trữ Log: các Log được tập trung về và tiến hành phân tích so sánh. Sau khi thực hiện thuật toán phân tích hệ thống sẽ đưa ra các cảnh báo cần thiết. Thậm chí còn có thể phân tích dữ liệu trong quá khứ.

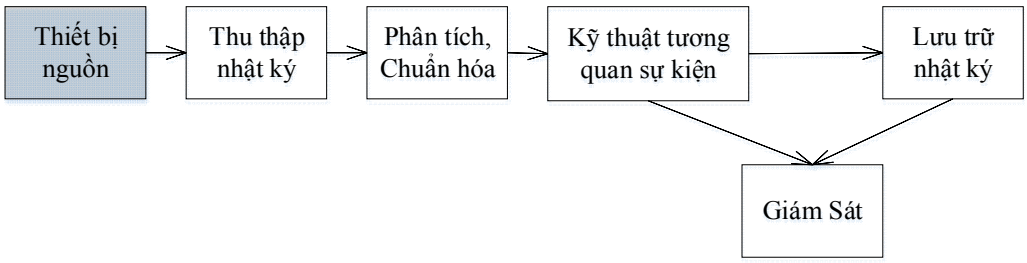
- Quản trị tập trung: cung cấp giao diện quản lý tập trung cho toàn bộ hệ thống giám sát an ninh. Hệ thống có sẵn hàng ngàn mẫu báo cáo để có thể sử dụng ngay.

1.3.4 Kiến trúc và cách thức hoạt động của hệ thống SIEM

Kiến trúc của hệ thống SIEM bao gồm: Thiết bị nguồn; Thu thập log; Phân tích và chuẩn hóa log; Kỹ thuật tương quan sự kiện; Lưu trữ log (nhật ký); Giám sát. Cụ thể như sau:

1.3.4.1 Thiết bị nguồn trong kiến trúc SIEM

Thiết bị nguồn: là các thiết bị đầu vào cung cấp các dữ liệu thu thập cho hệ thống SIEM, nó nằm ở vị trí đầu tiên trong kiến trúc của hệ thống SIEM được biểu thị trên Hình 1-1.



Hình 1.1: Bộ phận thiết bị nguồn

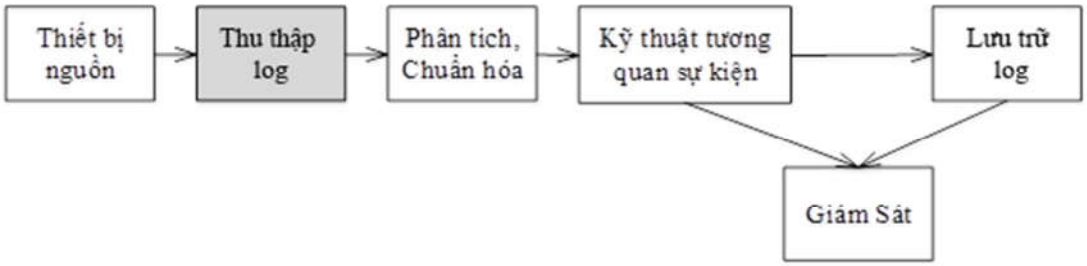
Các thiết bị nguồn có thể là Firewall, Router, Switch hoặc có thể là những bản ghi nhật ký từ một dịch vụ đang hoạt động. Đặc biệt người quản trị cần phải hiểu rõ những nguồn log mà mình muốn lấy sẽ tiết kiệm được rất nhiều công sức, thời gian và giảm sự phức tạp trong triển khai.

1.3.4.2 Bộ phận thu thập log

Thành phần thu thập log nằm ở bước 2 trong cấu trúc của SIEM (Hình 1-2). Cơ chế thu thập các bản ghi log sẽ phụ thuộc vào từng thiết bị, dịch vụ dữ liệu đầu vào nhưng cơ bản sẽ có hai phương thức: Phương thức đẩy nhật ký (Pull log) và phương thức tự lấy nhật ký (Push log), ngoài ra còn có tính năng xây dựng nhật ký để thu thập (Prebuilt Log collection) và tự tạo nhật ký thu thập log (Custom Log Collection), đồng thời còn kết hợp nhiều cách thu thập log khác (Mixed Environments).

- Đẩy nhật ký (Pull log)

Những bản ghi log sẽ được đẩy về từ các thiết bị nguồn. Phương pháp này rất dễ dàng cấu hình và cài đặt, người quản trị sẽ thiết lập một bộ tiếp nhận log, sau đó kết nối log từ thiết bị nguồn đến bộ phận tiếp nhận log.



Hình 1.2: Bộ phận thu thập log

Phương án này đôi khi bị thất lạc gói tin hoặc gửi tin không tới đích, dẫn tới có thể hệ thống không đủ gói tin để phân tích và đưa ra thông tin sai lệch.

Nếu Hacker sử dụng một cuộc tấn công vào hệ thống có chủ ý nhằm chống lại SIEM thì hacker có thể làm sai lệch các thông tin và thêm các dữ liệu rác vào SIEM. Qua đó người quản trị phải rất hiểu biết về các thiết bị gửi các bản ghi log cho SIEM là điều rất quan trọng, quyết định tới thành công của hệ thống.

- Lấy nhật ký (Push log)

Những bản ghi log sẽ được hệ thống SIEM gửi bản tin yêu cầu tới thiết bị nguồn và lấy bản ghi log về. Phương pháp Pull log đòi hỏi SIEM tạo nối tới các thiết bị nguồn đồng thời chủ động lấy những bản ghi từ những thiết bị nguồn đó.

Việc kết nối để lấy những bản ghi log của Pull Log có thể là nhiều giờ hoặc một số phút, đôi khi vài giây, nó phụ thuộc vào lượng log của thiết bị nguồn ít hoặc nhiều. Lúc này người quản trị phải hiểu rõ khối lượng log ở thiết bị nguồn để cấu hình thời gian để gửi log về hoặc để mặc định cho SIEM.

- Xây dựng hệ thống để thu thập nguồn log

Với phương pháp này sẽ xây dựng sẵn hệ thống, tạo ra phương pháp xác thực và các quy tắc, giao thức để tập hợp log. Với phương pháp này thì việc lấy các bản ghi log sẽ rất đơn giản. Nhưng điểm khó của nó là những ứng dụng có những bản ghi không tuân theo quy tắc hoặc giao thức sẽ gặp khó khăn trong việc thu thập log.

- Xây dựng theo phương pháp tự thu thập log

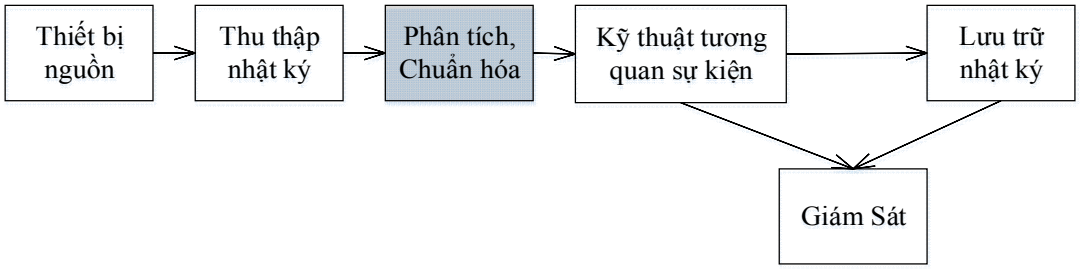
Trong hệ thống mạng sẽ có rất nhiều trang thiết bị và các dòng khác nhau, lúc này nguồn log cũng khác nhau. Người quản trị sẽ xây dựng một phương pháp riêng để lấy bản ghi log cung cấp cho SIEM. Qua đó sẽ kiểm soát được tất cả các nguồn log và cả quá trình phân tích, tìm kiếm log.

- Phối hợp nhiều phương pháp để thu thập log

Khi hệ thống mạng có nhiều thiết bị, nhiều nguồn log đổ về, người quản trị sẽ cần nhiều phương pháp thu thập log. Quản trị viên có thể lấy log qua Syslog, hoặc cơ sở dữ liệu MySQL sẽ lưu các bản ghi log trong một tệp tin trên máy chủ, còn Windows Server sẽ lưu các log trên ổ C của hệ điều hành. Qua các phương pháp đó sẽ cần các giao thức khác nhau để lấy log về.

1.3.4.3 Bộ phận phân tích và chuẩn hóa log

Thông qua bộ phận thu thập log, rất nhiều kiểu bản ghi khác nhau với định dạng nguồn khác nhau sẽ được chuyển về SIEM, để các bản ghi này hoạt động hiệu quả thì hệ thống cần phải đưa về một định dạng chuẩn duy nhất. Việc chuyển đổi các bản ghi này về dạng chuẩn được thực hiện bởi bộ phận chuẩn hóa log (Hình 1-3). Khi chuẩn hóa các bản ghi, hệ thống sẽ tổng hợp và phân tích nhanh hơn.

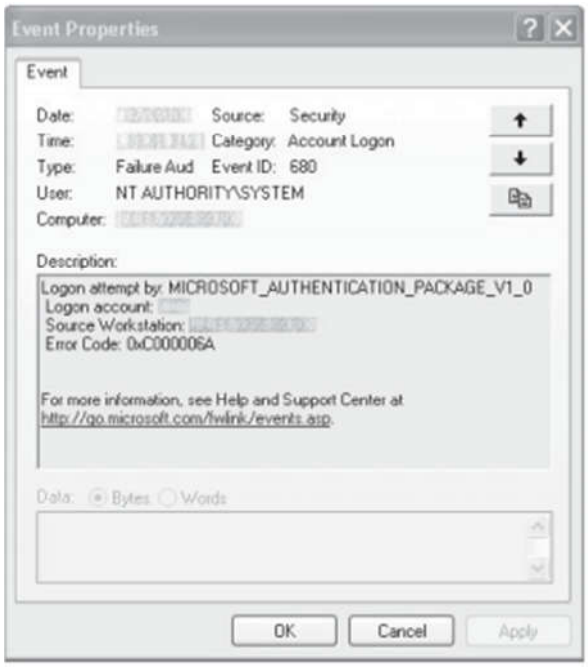


Hình 1.3: Bộ phận phân tích, chuẩn hóa log

Mỗi một hệ thống khác nhau thì nguồn vào các bản ghi cũng khác nhau, như Windows Server Event Log ở (Hình 1-4) và log đăng nhập trên Firewall ASA ở (Hình 1-5), cả hai nguồn cho thấy đều là log người dùng đăng nhập vào thiết bị. Các cách đăng nhập vào hệ thống là khác nhau nhưng những hành động đó là tương tự nhau.

Tuy nhiên nguồn log ban đầu của chúng lại là hai định dạng từ 2 nguồn thiết bị khác nhau.

Kết quả các bản ghi log đều được chuẩn hóa và tổng hợp cuối cùng được đẩy lên SIEM để phân tích.



Hình 1.4: Log đăng nhập trên hệ thống máy chủ windows

Priority	Hostname	Message
Local4.Info	192.168.1.1	:%ASA-sys-6-605005: Login permitted from 192.168.1.18/42925 to INSIDE:192.168.1.1/ssh for user "aiel"

Hình 1.5: Hệ thống firewall ASA hiển thị Log đăng nhập

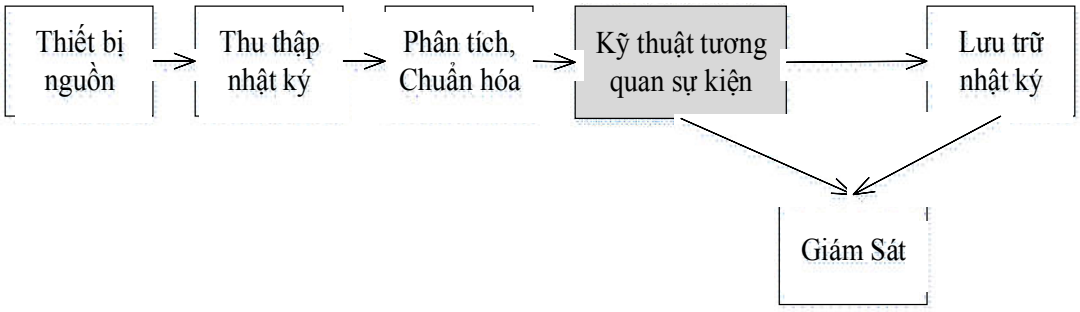
Time	Date	Source Device IP Address	Event Message	Event ID
22:54:53 CST	17-Oct-20	192.168.1.1	User login	ASA-sys-6-605005
22:54:53 CST	17-Oct-20	192.168.1.18	User login	Security: 680

Hình 1.6: Log được chuẩn hóa

1.3.4.4 Bộ phận kỹ thuật tương quan sự kiện

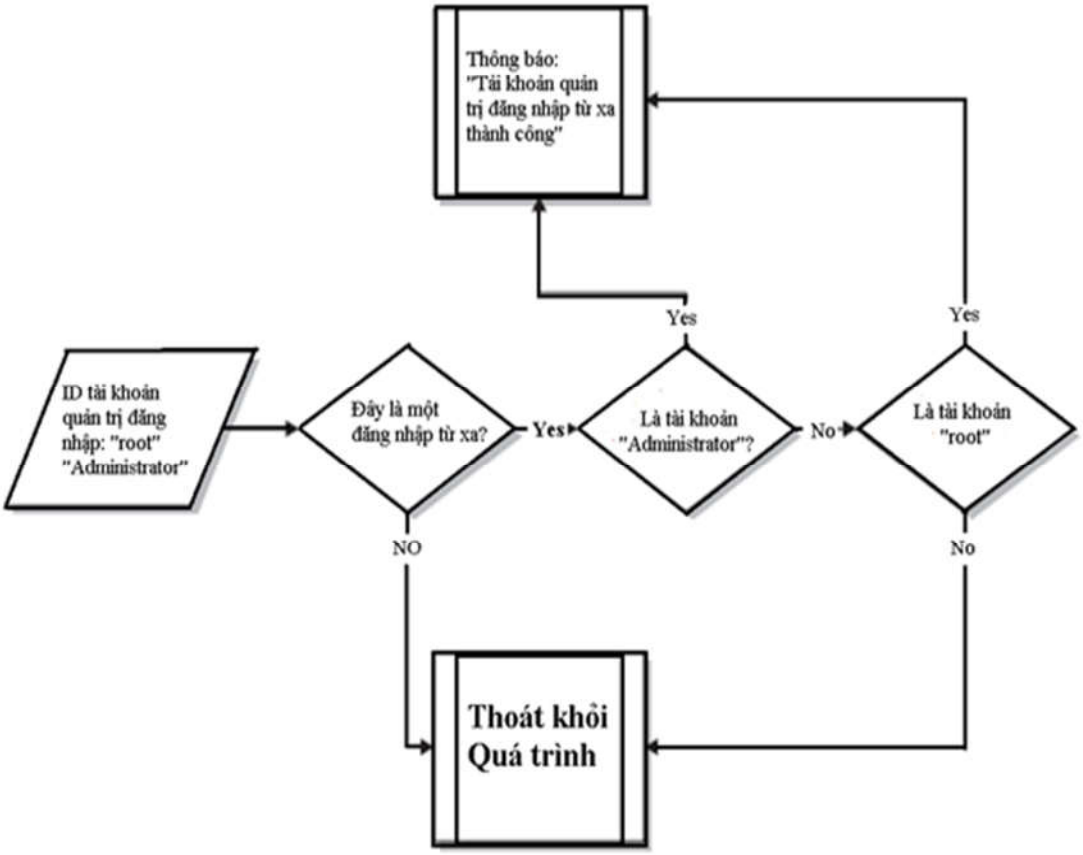
Đây là bộ phận tập hợp các tập giao thức, tập luật, dùng để so sánh, tổng hợp và đưa ra cảnh báo. Hệ thống phân tích chuyên nghiệp hay không sẽ phụ thuộc vào

người quản trị đặt các tập luật. Đôi khi người quản trị sẽ tự viết ra các quy tắc phục vụ phù hợp với hệ thống hiện tại của doanh nghiệp.



Hình 1.7: Bộ phận tương quan sự kiện

Chu trình về kiểm soát đăng nhập:



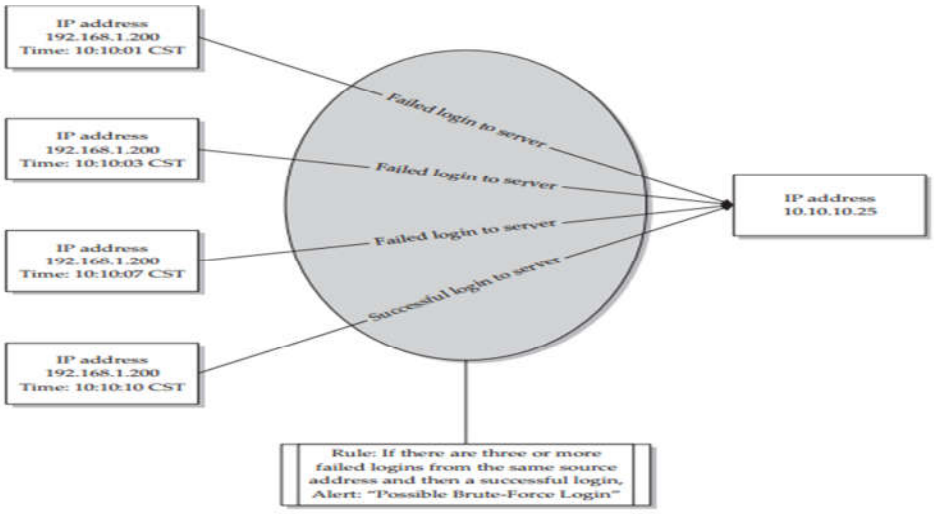
Hình 1.8: Kiểm soát quá trình đăng nhập tài khoản (nguồn:internet)

Bộ phận tương quan sự kiện an ninh giúp liên kết và kết nối các sự kiện an ninh từ nhiều nguồn khác nhau thành một sự kiện an ninh nhanh chóng, chính xác. Thông qua tương quan các sự kiện an ninh được thực hiện chuyên nghiệp sẽ đơn giản hóa các thủ tục ứng phó với các sự cố hệ thống.

Time	Event Number	Source	Destination	Event
10:20:01 CST	1035	192.168.1.200	10.10.10.25	Failed login to server
10:20:02 CST	1036	192.168.1.90	10.10.10.21	Successful login to server
10:20:03 CST	1037	192.168.1.200	10.10.10.25	Failed login to server
10:20:04 CST	1038	192.168.1.91	10.10.10.35	Failed login to server
10:20:05 CST	1039	192.168.1.10	10.10.10.2	Successful login to server
10:20:06 CST	1040	192.168.1.10	10.10.10.3	Successful login to server
10:20:07 CST	1041	192.168.1.200	10.10.10.25	Failed login to server
10:20:08 CST	1042	10.10.10.54	192.168.1.201	Failed login to server
10:20:09 CST	1043	10.10.10.34	192.168.1.10	Failed login to server
10:20:10 CST	1045	192.168.1.200	10.10.10.25	Successful login to server

Hình 1.9: Hệ thống SIEM hiển thị sự kiện cơ bản

Qua bảng sự kiện đăng nhập của hệ thống, ta thấy trong một thời gian ngắn nhưng có liên tục các lần đăng nhập. Điều này có thể do người dùng quên mật khẩu hoặc hệ thống đang bị tấn công brute-force tới máy chủ. Lúc này hệ thống sẽ đưa ra sự cảnh báo và gửi tới các bộ phận liên quan.



Hình 1.10: Sơ đồ minh họa về tương quan sự kiện (nguồn:internet)

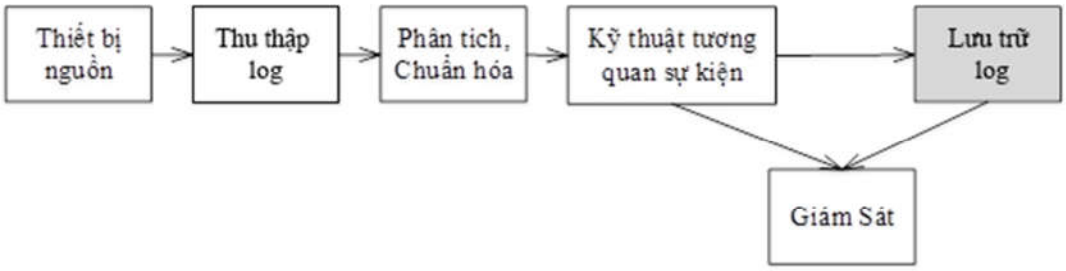
Khi hệ thống bị tấn công liên tục, cảnh báo gửi tới người quản trị, lúc này quản trị viên sẽ kịp thời theo dõi và đưa ra phương án xử lý nhanh chóng, phù hợp.

Thông qua sự kiện trong Hình 1.10, người quản trị có thể tự viết một đoạn code để thực hiện quá trình kiểm tra đó.

If [(failed logins >= 3) and then (Successful Login)] from the same source within 20 seconds = Possible Brute Force Compromise

1.3.4.5 Bộ phận lưu trữ log

Hệ thống thiết bị nguồn nhiều lượng log đổ về lớn. Lúc này trong SIEM có một bộ phận lưu lại các log đó để phục vụ phân tích.



Hình 1.11: Bộ phận lưu trữ log

- Sử dụng hệ quản trị cơ sở dữ liệu lưu trữ log

Các bản ghi khi đưa về sau khi chuẩn hóa sẽ được lưu vào trong hệ cơ sở dữ liệu như Oracle, MySQL, SQL hoặc các ứng dụng cơ sở dữ liệu lớn khác như Hadoop. Nếu xét về hiệu suất sẽ rất tốt, nhưng những ứng dụng cơ sở dữ liệu phải được tối ưu hóa để chạy với hệ thống SIEM.

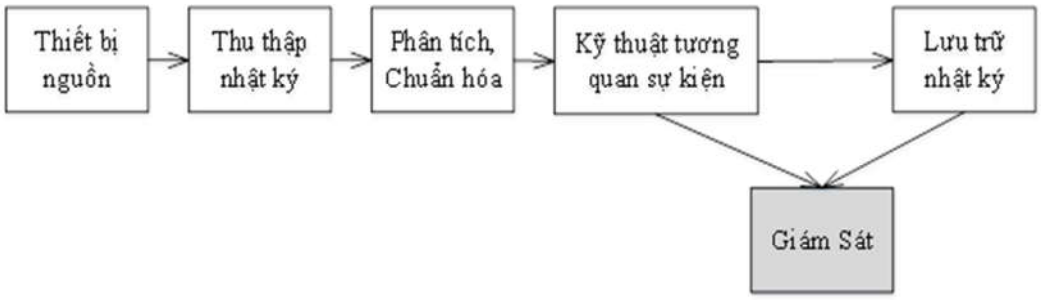
- Sử dụng định dạng tập tin văn bản lưu trữ log

Đối với cách thức lưu trữ tệp tin văn bản thì các thông tin cần phải có một ranh giới phân cách bằng dấu phẩy, tab hoặc kí hiệu khác. Từ đó nguồn thông tin đổ về có thể được phân tích và đọc đúng dễ giàng. Phương pháp này dễ dàng cho các ứng dụng bên ngoài để truy cập dữ liệu để phân tích. Ngoài ra con người có thể đọc được và thuận tiện phân tích tìm kiếm và hiểu nó.

- Lưu trữ log dưới dạng tập tin nhị phân

Phương pháp lưu trữ log dưới định dạng tập tin nhị phân là cách sử dụng một tập tin với định dạng tùy chỉnh để lưu trữ thông tin.

1.3.4.3 Bộ phận giám sát



Hình 1.12: Bộ phận giám sát

Trong hệ thống SIEM, các thông tin từ các bản ghi nhật ký được thể hiện trên nền giao diện Web để quản lý, giám sát các sự kiện. Đây chính là Bộ phận giám sát.

Thông qua giao diện ứng dụng này cho phép quản trị viên xử lý sự cố hoặc cung cấp cái nhìn tổng quan về môi trường hoạt động của hệ thống. Trước kia khi muốn xử lý sự cố quản trị viên thường phải đọc log của thiết bị nguồn. Nhưng với SIEM sẽ tập trung log tại một nơi duy nhất, hệ thống sẽ phân tích các bản ghi log khác nhau một cách dễ dàng bởi vì các bản ghi đó đã được chuẩn hóa các thông tin dữ liệu.

Trong quá trình quản lý và giám sát giao diện điều khiển của hệ thống SIEM, người quản trị có thể tiếp tục phát triển nội dung, các tập luật để tìm ra thông tin từ các sự kiện an ninh được xử lý. Vai trò của giao diện web sẽ hỗ trợ điều khiển và giao tiếp với các dữ liệu được lưu trữ bên trong hệ thống SIEM.

1.4. Kết luận chung chương một

Trong chương 1, luận văn đã nghiên cứu tổng quan chung về an ninh mạng, giám sát tập trung và các yêu cầu giám sát hệ thống mạng, cũng như các vấn đề liên quan đến hệ thống SIEM. Qua đó ta thấy cấu trúc hoạt động rất phức tạp bởi có nhiều bộ phận hoạt động chuyên biệt. Nhưng cũng tạo ra được sức mạnh tổng hợp và phân tích log rất tốt, linh hoạt, hỗ trợ tối đa cho việc quản trị hệ thống.

Chương tiếp theo sẽ trình bày các giải pháp và cách thức áp dụng SIEM một cách hiệu quả.

CHƯƠNG 2. NGHIÊN CỨU GIẢI PHÁP AN TOÀN THÔNG TIN

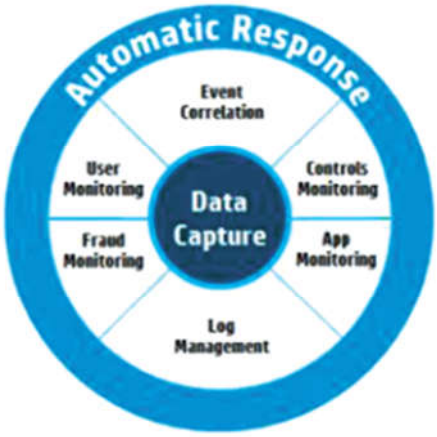
Chương 2 của luận văn tập trung nghiên cứu các giải pháp giám sát an toàn thông tin, phân tích một vài công cụ giám sát thành phần chuyên biệt, từ đó so sánh và đưa ra công cụ sử dụng giải pháp giám sát tập trung từ đó sẽ đưa ra giải pháp an toàn thông tin.

2.1 Các giải pháp giám sát an toàn thông tin hiện nay

2.1.1 Giải pháp HP ArcSight ESM

HP ArcSight Enterprise Security Manager là một sản phẩm trong bộ sản phẩm ArcSight của HP. Hệ thống rất hiệu quả về trong việc quản lý và vận hành phân tích log và có thể xử lý lượng log lớn đổ về cùng hỗ trợ nhiều định dạng lấy log khác nhau. Trong hệ thống cho phép phản hồi nhanh chóng và nhận dạng rất nhanh các cuộc tấn công từ bên ngoài hệ thống mạng hoặc bên trong.

Hệ thống cung cấp một giải pháp tổng thể về các hoạt động như: phân tích các mối tấn công từ bên ngoài mà hackers gây ra hoặc các mối tấn công từ bên trong. Ngoài ra còn phân tích, phát hiện các rủi ro từ các điểm yếu trên các phần mềm hoặc các dịch vụ hoạt động.



Hình 2.1: HP ArcSight Enterprise Security Manager (nguồn:internet)

Ưu điểm:

- Phân tích dữ liệu toàn diện.
- Cảnh báo tấn công hoặc lỗi theo thời gian thực.
- Có thể tìm kiếm và tạo báo cáo tổng hợp.

Nhược điểm:

- Hiệu năng xử lý phụ thuộc vào thiết bị, muốn nâng cấp phải mua thiết bị mới.

- Chi phí đầu tư tốn kém, đắt đỏ.

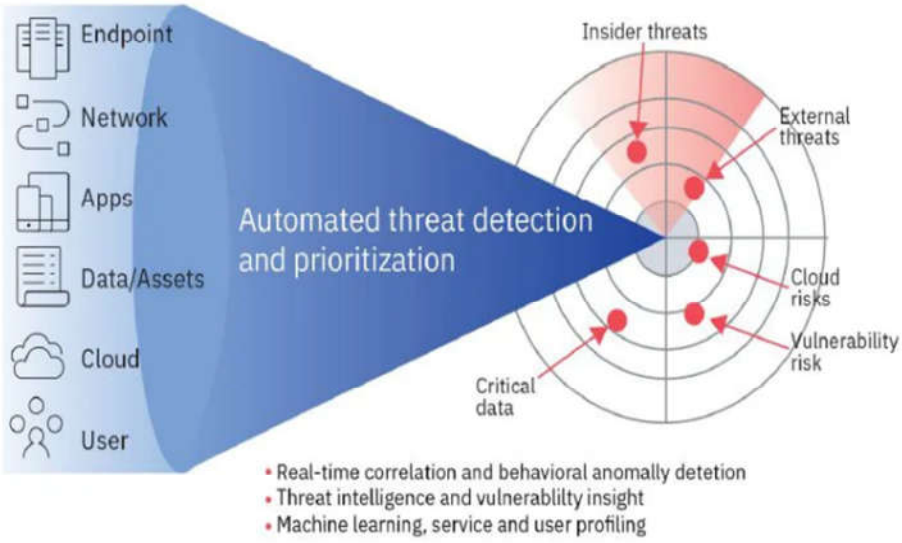
2.1.2 Giải pháp IBM Security Qradar

IBM QRadar SIEM (Security Information and Event Management – Quản lý sự kiện và bảo mật thông tin) được thiết kế để cung cấp cho các nhóm bảo mật khả năng hiển thị tập trung vào dữ liệu bảo mật toàn doanh nghiệp và hiểu biết sâu sắc về các mối đe dọa ưu tiên cao nhất.

Bước đầu tiên, giải pháp sử dụng một lượng lớn dữ liệu trong toàn doanh nghiệp để cung cấp cái nhìn toàn diện về hoạt động trên khắp các môi trường tại chỗ và trên nền tảng đám mây. Khi dữ liệu được sử dụng, QRadar áp dụng trí thông minh bảo mật tự động, thời gian thực để phát hiện và ưu tiên nhanh chóng và chính xác các mối đe dọa. Cảnh báo có thể thực hiện cung cấp bối cảnh lớn hơn vào các sự cố tiềm ẩn, cho phép các nhà phân tích bảo mật phản ứng nhanh chóng để hạn chế tác động của kẻ tấn công. Không giống như các giải pháp khác, chỉ QRadar được xây dựng có mục đích để giải quyết các trường hợp sử dụng bảo mật và được thiết kế có chủ ý để dễ dàng mở rộng quy mô với sự tùy chỉnh giới hạn cần có.

QRadar hỗ trợ nhiều công nghệ, ứng dụng và dịch vụ đám mây để giúp khách hàng có được tầm nhìn toàn diện vào hoạt động toàn doanh nghiệp. Khi dữ liệu này được tập trung, nó có thể được phân tích tự động để xác định các mối đe dọa đã biết, sự bất thường có thể chỉ ra các mối đe dọa chưa biết và rủi ro quan trọng có thể khiến dữ liệu nhạy cảm bị lộ.

- Tự động hóa thông tin bảo mật để nhanh chóng phát hiện các mối đe dọa



Hình 2.2: Tự động hóa thông tin bảo mật phát hiện các mối đe dọa (nguồn:internet)

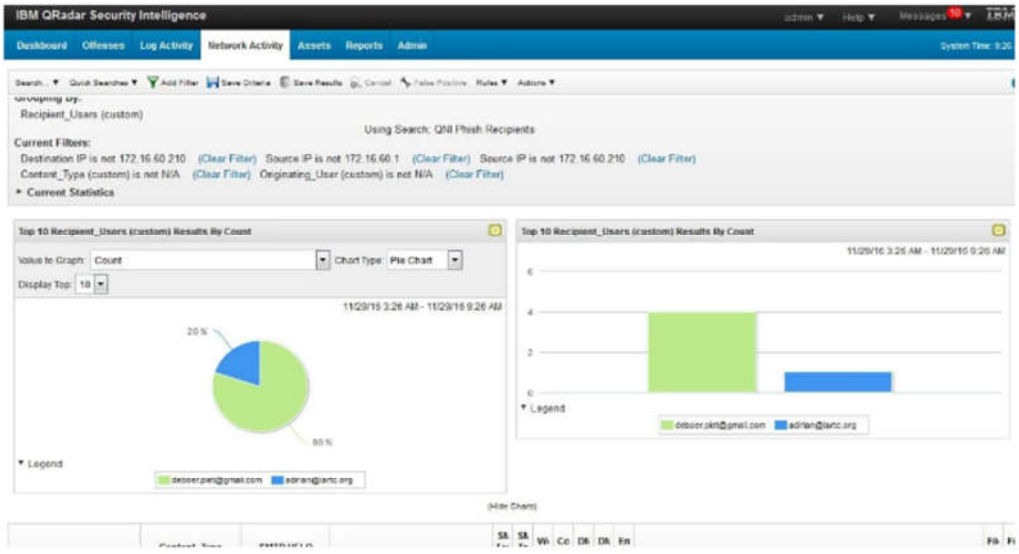
QRadar SIEM được thiết kế để tự động phân tích và tương quan hoạt động trên nhiều nguồn dữ liệu bao gồm nhật ký, sự kiện, luồng mạng, hoạt động của người dùng, thông tin về lỗ hổng và thông tin về mối đe dọa để xác định các mối đe dọa đã biết và chưa biết.

- Phát hiện hoạt động của mạng, người dùng và ứng dụng bất thường

Khi những kẻ tấn công trở nên tinh vi hơn trong các kỹ thuật của chúng, việc phát hiện mối đe dọa cần phải có khả năng phát hiện những thay đổi nhỏ trong hành vi của mạng, người dùng hoặc hệ thống có thể chỉ ra các mối đe dọa như thông tin bị xâm phạm hoặc fileless malware. QRadar chứa nhiều khả năng phát hiện bất thường để xác định những thay đổi trong hành vi có thể là chỉ số của một mối đe dọa chưa biết. Và khả năng độc đáo của QRadar để giám sát và phân tích lưu lượng ứng dụng lớp 7 cho phép nó xác định chính xác hơn các bất thường mà các giải pháp khác có thể bỏ lỡ.

Bằng cách tùy chọn sử dụng QRadar Network Insights như một phần của việc triển khai SIEM, các tổ chức có thể hiểu rõ hơn về các hệ thống giao tiếp với nhau, ứng dụng nào có liên quan và thông tin nào được trao đổi trong các gói. Bằng cách tương quan thông tin này với hoạt động mạng, nhật ký và người dùng khác, các nhà phân tích bảo mật có thể phát hiện ra hoạt động mạng bất thường có thể là dấu hiệu của máy chủ bị xâm nhập, người dùng bị xâm phạm hoặc nỗ lực đánh cắp dữ liệu.

- Quản lý việc tuân thủ với quy tắc, nội dung và báo cáo được xây dựng trước



Hình 2.3: Phát hiện hoạt động của mạng, người dùng và ứng dụng bất thường

QRadar cung cấp tính minh bạch và khả năng đo lường đối với việc đáp ứng các quy định và báo cáo về việc tuân thủ quy định. Nó có hàng trăm báo cáo được xây dựng trước và các mẫu quy tắc có thể giúp các tổ chức dễ dàng giải quyết các yêu cầu tuân thủ của ngành hơn.

Dễ dàng thay đổi, mở rộng quy mô

Kiến trúc linh hoạt, có thể mở rộng của QRadar được thiết kế để hỗ trợ cả các tổ chức lớn và nhỏ với nhiều nhu cầu khác nhau.

Giải pháp QRadar SIEM bao gồm các thành phần sau: bộ thu sự kiện, bộ xử lý sự kiện, bộ thu lưu lượng, bộ xử lý luồng, nút dữ liệu (để lưu trữ chi phí thấp và tăng hiệu suất) và bảng điều khiển trung tâm. Tất cả các thành phần có sẵn như phần cứng, phần mềm hoặc thiết bị ảo.

2.1.3 Giải pháp Mcafee ESM

Intel Security's McAfee Enterprise Security Manager (ESM) là một bộ sản phẩm quản lý sự kiện và thông tin bảo mật của hãng Intel Security's McAfee. Nó được thiết kế dưới dạng máy ảo hoặc thiết bị và hỗ trợ số lượng lớn các sản phẩm để tạo ra thông tin hữu ích cho các nhà quản trị an ninh.

Mcafee ESM thu thập các bản ghi trên một số lượng lớn thiết bị và tích hợp với danh bạ hoạt động. Người nhận lấy logs và đều dễ dàng có hiệu lực cho người dùng. Tính năng kiểm soát truy cập và tài khoản người dùng dễ dàng giới hạn người dùng chỉ có quyền truy cập những gì họ thực sự cần xem.

Điểm đặc biệt của sản phẩm là tính năng phân tích tương quan nâng cao. ESM tạo ra mối tương quan giữa các nguồn khác nhau và tìm ra yếu tố ngoại lai nhanh nhất có thể cùng với các sự kiện khả nghi trong hệ thống mạng. ESM cực kỳ hữu ích cho việc theo dõi nhanh chóng, dễ dàng các sự kiện bảo mật và vi phạm chính sách. Tính năng đó cho phép chúng ta nhìn thấy các bản ghi chính xác gây ra cảnh báo hoặc sự kiện, cho phép quản trị viên hệ thống quyết định sự kiện có đáng gỡ xuống.

McAfee Enterprise Security Manager là một sản phẩm hữu ích tuy nhiên giá thành tương đối cao (đắt gần gấp 10 lần giá của sản phẩm SIEM) nhưng nếu có điều kiện mua thì đây là sản phẩm đáng đầu tư.

2.1.4 Giải pháp MARS của Cisco

Cisco Security Manager là một doanh nghiệp cung cấp cái nhìn sâu sắc và kiểm soát an ninh của Cisco và các thiết bị mạng. Cisco Security Manager cung cấp

quản lý an ninh toàn diện (cấu hình và quản lý sự kiện) trên một loạt các thiết bị bảo mật của Cisco, trong đó có Cisco ASA Adaptive Security Appliances, IPS Sensor gia dụng, định tuyến dịch vụ tích hợp, Firewall Services Modules, và Cisco Catalyst 6000 Series Switches. Cisco Security Manager cho phép bạn quản lý hiệu quả các mạng, từ các mạng nhỏ, mạng lưới rộng lớn bao gồm hàng trăm các thiết bị.

Cisco Security MARS giám sát và phân tích các sự kiện từ nhiều nguồn, trong đó có 5.500 Series Cisco ASA và Cisco Catalyst 6500 Series FWSM. Cisco Security MARS tích hợp với Cisco Security Manager để tương quan các sự kiện an ninh với các quy tắc cấu hình tường lửa mà có thể ảnh hưởng đến sự kiện bảo mật.

Nhược điểm của dòng sản phẩm này là đi theo thiết bị phần cứng, không tùy chỉnh và cấu hình theo yêu cầu của doanh nghiệp được mà phải liên lạc với hãng để chỉnh trên dòng thiết bị cứng.

2.1.5 Giải pháp AlienVault OSSIM

OSSIM là một sản phẩm SIEM mã nguồn mở của AlienVault. OSSIM đã được tích hợp một số công cụ bảo mật mạnh mẽ như Snort, ntop, OpenVAS, P0f, PADs, arpwatch, OSSEC,Osiris, Nagios, OCS và Kismet.

Các log được thu thập và đưa vào chuẩn hóa sau đó gửi đến một máy chủ trung tâm. Tại máy chủ trung tâm sẽ đánh giá các rủi ro, những mối tương quan và lưu trữ vào trong cơ sở dữ liệu. Tiếp theo sẽ tiến hành xử lý khi có kết quả sẽ đưa ra thông báo hoặc có thể gửi email cảnh báo tới admin.

Trong OSSIM việc liên kết SE là một trong những tính năng cốt lõi để phân phân biệt nó với hệ thống IDS/IPS. Nó giúp giảm các cảnh báo giả bằng cách tương quan liên kết nhiều SE khác nhau (Hệ thống sẽ phân tích nhiều SE cùng một lúc và đối chiếu với nhau để đưa ra kết quả chính xác nhất) và báo động cho các quản trị viên biết và chú ý đến các SE.

2.1.6 Giải pháp Splunk

Splunk là một hệ thống giám dựa trên việc phân tích Log, nó thực hiện các công việc tìm kiếm log, phân tích và giám sát lượng dữ liệu lớn của log sinh ra từ những dịch vụ đang chạy, hạ tầng mạng. Hệ thống Splunk được tạo ra dựa trên nền tảng Lucene và MongoDB, ngoài ra có thể quản lý trên nền giao diện web trực quan. Splunk có những tính năng sau:

- Splunk tích hợp và hỗ trợ được gần như hết các loại log của hệ thống.
- Cho phép thực hiện thu thập log từ nhiều nguồn khác nhau đổ về.

- Có thể cập nhật dữ liệu liên tục theo thời gian thực và đưa ra cảnh báo theo thời gian thực.

- Tốc độ đánh chỉ mục rất nhanh cho khối lượng dữ liệu lớn đổ về.

- Có thể cập nhật liên tục lượng dữ liệu lớn đổ về và làm việc rất tốt.

- Cung cấp thông tin cho người quản trị nhanh chóng dựa trên cơ chế cảnh báo và cách tìm kiếm do người quản trị đặt ra.

*** Lợi thế của Splunk so với các giải pháp SIEM khác**

- Linh hoạt mềm dẻo khi sử dụng:

Linh hoạt, khả năng mở rộng và đủ linh hoạt từ bất kỳ nguồn dữ liệu, các ứng dụng tùy chỉnh và cơ sở dữ liệu. Splunk tự động cung cấp một cái nhìn chi tiết theo dòng thời gian của tất cả các dữ liệu thu thập được.

- Điều tra theo thời gian thực:

Splunk cho phép bạn xem thông tin thời gian thực từ an ninh và thiết bị mạng, hệ điều hành, cơ sở dữ liệu và các ứng dụng, trên một thời gian cho phép các đội an ninh để nhanh chóng phát hiện và hiểu được ý nghĩa end-to -end của một sự kiện an ninh.

Splunk sẽ giải quyết được khó khăn của các hệ thống bảo mật hiện tại khi tìm kiếm và phát hiện các hành vi nguy hiểm đang hoạt động trong hệ thống. Với khả năng phát hiện từng hành vi bất hợp phát nhỏ nhất, Splunk sẽ giúp phát hiện những cuộc tấn công tinh vi nhằm vào hệ thống một cách nhanh chóng và hiệu quả nhất.

- Liên kết thông tin theo thời gian thực và cảnh báo:

Tương quan của thông tin từ bộ dữ liệu khác nhau có thể cung cấp cái nhìn sâu sắc thêm và bối cảnh. Splunk có thể liên kết với tất cả các thông tin dữ liệu từ mọi nguồn trên hệ thống một cách nhanh chóng và chính xác theo thời gian thực.

- Splunk là phần mềm mã nguồn mở, có bản không tính phí nên không tốn kém khi triển khai.

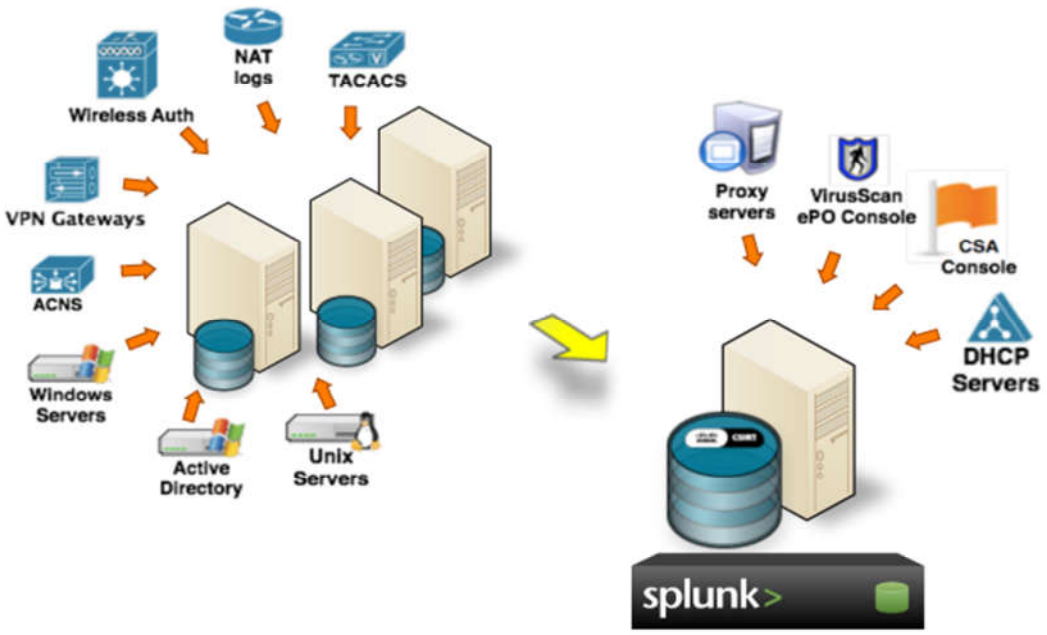
- Giải quyết được hầu hết các bài toán trong giám sát hệ thống mạng: giám sát hạ tầng, giám sát dịch vụ, giám sát an ninh, giám sát người dùng... Đây là đặc điểm chính giúp cho Splunk trong tương lai sẽ được các tổ chức, doanh nghiệp sử dụng để triển khai hệ thống giám sát tập trung bên trong hệ thống mạng của họ.

2.2. Lựa chọn giải pháp Splunk

2.2.1. Giới thiệu tổng quan về giải pháp Splunk

Splunk là phần mềm cho phép tìm kiếm và duyệt logs và các dữ liệu trong thời gian thực. Người dùng có thể ngay lập tức phát hiện ra sự cố ở bất cứ ứng dụng nào, hoặc ở các máy chủ và thiết bị; cảnh báo các nguy cơ tiềm ẩn và báo cáo các hoạt động của các dịch vụ và thành phần khác nhau trong mạng.

Splunk là một công cụ dữ liệu rất linh hoạt và khả năng mở rộng cho các dữ liệu máy tính được tạo ra bởi cơ sở hạ tầng CNTT. Nó thu thập, lập chỉ mục và khai thác những dữ liệu được tạo ra từ bất cứ nguồn nào, định dạng hoặc vị trí bao gồm cả đóng gói và các ứng dụng tùy chỉnh, máy chủ ứng dụng, máy chủ web, cơ sở dữ liệu, mạng, máy ảo, hypervisors, hệ điều hành và nhiều hơn nữa mà không cần phải phân tích cú pháp tùy chỉnh, bộ điều hợp hoặc một cơ sở dữ liệu trên các phụ trợ.



Hình 2.4: Mô hình hoạt động của Splunk (nguồn:internet)

Splunk có những điểm nổi bật như sau:

- Giải quyết tốt vấn đề hạ tầng mạng

Splunk cho phép bạn tìm kiếm, cảnh báo và báo cáo trên mạng trong thời gian thực trên các sự kiện mạng và các giao dịch hoàn chỉnh. Bắt đầu từ hiện tượng bất thường rồi tìm ra nguyên nhân một cách nhanh chóng với các thông tin từ syslog, SNMP traps, các cấu hình và dữ liệu netflow. Việc tích hợp Splunk với hệ thống giám sát cho phép phát hiện sớm các vấn đề và đi sâu vào tìm hiểu cặn kẽ nguyên nhân của vấn đề đó.

- Giải quyết tốt vấn đề ảo hóa

Ảo hóa mang đến nhiều lợi ích không thể phủ nhận, nhưng nó cũng khiến cho hệ thống càng trở nên phức tạp hơn. Các vấn đề liên quan đến tài nguyên trên các máy chủ vật lý, máy áo, năng suất hệ thống trở nên quá phức tạp, khiến cho các hệ thống cũ không thể theo kịp.

Splunk cung cấp một cái nhìn rõ ràng xuyên suốt các hệ thống ảo hóa. Thu thập tất cả các dữ liệu từ các hệ thống ảo hóa, hệ thống vật lý, các phiên giao dịch rồi tổng hợp, liên kết chúng lại với nhau để phục vụ việc phân tích, tìm kiếm, tối ưu hệ thống.

- Giám sát hệ thống cloud (đám mây)

Sự phát triển của các cơ sở hạ tầng lai đã tạo ra một thách thức đối với các bộ phận CNTT phải làm thế nào để các đối tượng không nằm trong quyền kiểm soát trực tiếp của họ. Với khả năng phân tích, lập chỉ mục cho mọi loại dữ liệu, Splunk cung cấp một cái nhìn chi tiết, chính xác về môi trường đám mây của bất kì nhà cung cấp nào.

Ngoài việc cung cấp khả năng hiển thị hoạt động vào các môi trường điện toán đám mây, Splunk có thể cung cấp những thông tin như tình hình sử dụng tài nguyên của khách hàng, sử dụng tài sản, quản lý tài khoản và kế hoạch hoạt động.

- Giám sát hệ thống Email

Hệ thống tin nhắn là cực kì phức tạp. Từ truy tìm các tin nhắn email để quản lý tuân thủ và phân tích thư rác và tấn công lừa đảo, cơ sở hạ tầng khá phức tạp và chuyên sâu. Splunk giúp bạn tìm kiếm các giao dịch tin nhắn trong thời gian thực trên cơ sở hạ tầng của bạn.

- Giám sát máy chủ

Chi phí quản lý máy chủ đang bị đẩy lên rất cao. Quản lý tập trung các máy chủ là một vấn đề khó khăn, phải tốn quá nhiều các agents để lấy thông tin từ cùng một máy chủ. Nhận dạng và chuẩn đoán các vấn đề của máy chủ lại liên quan đến việc truy cập trực tiếp và gây ảnh hưởng đến hệ thống đang chạy. Splunk tích hợp logs, cấu hình, các traps, metrics vào một nơi. Tìm kiếm, phân tích, báo cáo xuyên suốt hệ thống chỉ trong vài giây, giải quyết sự cố một cách nhanh chóng.

- Giám sát ứng dụng

Hệ thống phân bố ứng dụng phức tạp có thể ẩn chứa rất nhiều lỗi. Nhưng để tìm và sửa các lỗi này thì lại là một vấn đề không hề đơn giản, tiêu tốn nhiều thời

gian cũng như tiền bác. Những nhóm phát triển cũng như người quản trị không thể truy cập vào các dữ liệu mà họ cần để làm việc. Splunk cho phép điều tra phân tích các vấn đề một cách nhanh chóng từ khu vực quản lý trung tâm. Splunk cũng cho phép nhóm phát triển truy cập các dữ liệu họ cần để nhanh chóng giải quyết và khắc phục vấn đề.

2.2.2 Tính năng của giải pháp Splunk

2.2.2.1 Quản lý ứng dụng của Splunk

- Khắc phục sự cố nhanh hơn:

Splunk giúp 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, khắc phục sự cố vấn đề một cách nhanh chóng, giảm chi phí và giảm thời gian để điều tra và khắc phục sự cố tới 70%. Đồng thời, 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.

- Nắm được hoạt động của toàn bộ ứng dụng:

+ Cho phép 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.

+ Đạt được cái nhìn toàn diện về cách mà người dùng sử dụng dịch vụ, từ đó có thể cung cấp dịch vụ tốt hơn.

+ 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 1 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.

2.2.2.2 Quản lý các hoạt động công nghệ thông tin

Splunk cung cấp một cách tiếp cận tốt hơn, nó thu thập và lập indexes chứa tất cả dữ liệu được tạo ra bởi hệ thống CNTT (hệ thống mạng, server, OS, ảo hóa, v.v.). Splunk 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.

- Giúp nắm bắt được hoạt động ảo hóa, hệ thống cloud private và public từ một giao diện trung tâm, dễ dàng 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 và có thêm kinh nghiệm xử lý các sự kiện xảy ra định kỳ để tránh mất mát.

- 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.

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

Sử dụng sức mạnh và khả năng mở rộng của Splunk không chỉ cho hoạt động quản lý CNTT mà còn dùng để hỗ trợ kiểm toán, an ninh. Giảm số lượng các công cụ và kỹ năng cần thiết để duy trì quản lý cơ sở hạ tầng phức tạp.

- Phân tích hoạt động

+ Splunk phân tích hoạt động toàn diện theo nhiều tầng giúp cho định hướng của doanh nghiệp tốt hơn tùy theo từng trường hợp cụ thể.

+ Chủ động trong việc nhận diện và khắc phục lỗi dịch vụ để đảm bảo sự hài lòng của khách hàng và giúp tăng số lượng khách hàng sử dụng.

+ Nắm bắt được những nguy hiểm tiềm tàng trong quá trình hoạt động kinh doanh, giúp đạt được các mục tiêu kinh doanh bằng cách cung cấp tầm nhìn toàn diện trên toàn hệ thống công nghệ không đồng nhất, các dịch vụ, cách quản lý, lên kế hoạch về dung lượng, phân tích mức sử dụng của người dùng và nhiều hơn nữa.

- *Giám sát cơ sở hạ tầng*

+ Máy chủ: Splunk cho phép chúng ta có thể chủ động giám sát các máy chủ và hiểu biết sâu hơn về hiệu suất, cấu hình, truy cập và các lỗi phát sinh. Tương quan hiệu suất máy chủ, các lỗi và dữ liệu sự kiện với người dùng, ảo hóa và ứng dụng thành phần để ngăn ngừa và khắc phục lỗi. Phân tích và tối ưu hóa chi phí cho việc theo dõi dung lượng máy chủ, báo cáo an ninh trong thời gian thực.

+ Hệ thống lưu trữ: Splunk có thể cho ta tương quan log, số liệu hiệu suất và các sự kiện từ hệ thống lưu trữ với máy chủ, mạng và dữ liệu từ các ứng dụng để giải quyết các vấn đề và làm tăng sự hài lòng của khách hàng. Sử dụng công cụ phân tích mạnh mẽ để khắc phục sự cố trong thời gian thực và phân tích hiệu suất hệ thống lưu trữ. Giảm thời gian phát triển và cắt giảm chi phí bằng việc dễ dàng tích hợp với các nhà cung cấp dịch vụ lưu trữ, như NetApp và EMC.

+ Hệ thống mạng: Splunk cho phép ta có thể giám sát và theo dõi dữ liệu mạng từ các thiết bị không dây, switch, router, firewall và trên những thiết bị khác bằng cách sử dụng SNMP, Netflow, syslog, PCAP,…

Chủ động nhận diện các vấn đề an ninh mạng và thực hiện phân tích vấn đề. Tương quan dữ liệu mạng với các ứng dụng, hệ thống lưu trữ và phân tích máy chủ để giữ cho mạng an toàn và hoạt động mọi lúc.

- *Splunk cho hệ điều hành*

Splunk và ứng dụng của splunk có thể giúp ta:

+ Tương quan số liệu hệ thống và dữ liệu sự kiện với các dữ liệu ở các tầng công nghệ khác một cách dễ dàng. Tìm liên kết giữa vấn đề hiệu suất ứng dụng và hệ điều hành, ảo hóa, hệ thống lưu trữ, mạng, và cơ sở hạ tầng máy chủ.

+ Nắm được toàn bộ hoạt động hệ thống bằng cách cung cấp bảng điều khiển trung tâm môi trường không đồng bộ.

+ Theo dõi những thay đổi và đảm bảo an ninh cho môi trường bằng cách giám sát môi trường để phát hiện những hoạt động bất ngờ, thay đổi vai trò của người sử dụng, truy cập trái phép,…

- *Quản lý ảo hóa*

+ Cơ sở hạ tầng ảo hóa tạo ra môi trường năng động, nơi mà tài nguyên máy tính như máy chủ, storage, phần cứng mang được ảo hóa từ các ứng dụng, hệ điều hành và người sử dụng. Môi trường ảo hóa phức tạp đòi hỏi cách tiếp cận mới với các dịch vụ CNTT truyền thống như xử lý sự cố hiệu suất, quản lý và phân tích rủi ro.

+ Ứng dụng ảo hóa của Splunk kết hợp sức mạnh và tính năng của Splunk Enterprise được thiết kế dành riêng cho công nghệ ảo hóa. Kết hợp dữ liệu hạ tầng ảo hóa với dữ liệu tầng công nghệ khác sẽ cho một góc nhìn bao quát hơn về hệ thống trung tâm dữ liệu.

+ Splunk App cho ảo hóa có thể tương thích và thu thập dữ liệu ảo hóa từ các công nghệ ảo hóa như WMware vSphere, Citrix XenServer và Microsoft Hyper-V và công nghệ ảo hóa máy tính bàn như Citrix XenApp và Citrix XenDesktop.

+ Nó tạo các báo cáo đa dạng, đồng nhất về các công nghệ ảo hóa từ tất cả các lớp ứng dụng và cơ sở hạ tầng.

+ Giúp chủ động ngăn chặn, quản lý vấn đề hiệu suất, tắc nghẽn cổ chai, những sự kiện bất ngờ, những thay đổi và lỗi an ninh bảo mật nguy hiểm. Nó phân tích và báo cáo chính xác giúp cho người dùng có trải nghiệm tối ưu.

+ Tương quan dữ liệu ảo hóa, giúp việc tìm ra các sự kiện có liên quan một cách dễ dàng hơn, tương quan các vấn đề về hiệu năng, mạng và kiến trúc hệ thống máy chủ.

+ Giữ lại số liệu về hiệu suất hoạt động của máy để theo dõi và phân tích. Thu thập dữ liệu có chiều sâu từ máy chủ, máy ảo, hệ thống máy tính. Cung cấp khả năng hiển thị hoạt động và phân tích hoàn chỉnh bằng cách xác định khả năng của máy chủ, các máy ảo nhàn rỗi, các máy chủ sử dụng đúng mức, sức chứa dữ liệu, theo dõi thống kê hiệu suất để tìm mô hình sử dụng và tránh khả năng tắc nghẽn có thể.

+ Theo dõi chi tiết sự thay đổi mà người dùng thực hiện, tự động hóa các tác vụ của vSphere cũng như báo cáo tình trạng các thành phần ảo. Cải thiện an ninh bằng cách giám sát môi trường để tìm các hoạt động đáng ngờ, vai trò của người sử dụng bị thay đổi, truy cập trái phép và nhiều hơn nữa.

2.2.2.3 An ninh trong lĩnh vực CNTT

- Quản lý log

Phần mềm Splunk giúp khách hàng cải thiện vấn đề phân tích dữ liệu log để quản lý việc kinh doanh của họ tốt hơn. Splunk tự động index dữ liệu, bất kể có cấu trúc hay không cấu trúc, cho phép ta nhanh chóng tìm kiếm, báo cáo, chẩn đoán các hoạt động và các vấn đề an ninh một cách ít tốn kém hơn. Với Spunk-việc quản lý log sẽ dễ hơn bao giờ hết.

- Ứng dụng Splunk dành cho an ninh

Với ứng dụng an ninh của Splunk 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 đã 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. Chỉ một click chuột là ta có thể thấy được các dữ liệu thô mà ứng dụng an ninh splunk lưu trữ và điều tra nhận dạng mối 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. Với 1 cú click chuột, ta có thể xem tất cả các dữ liệu thô được đặt ra theo thứ tự thời gian, đưa ra 1 cái nhìn trực tiếp cho đồng nghiệp hoặc tạo ra một tìm kiếm mới để xem các sự kiện đã xuất hiện này có tiếp tục xuất hiện hay không.

- Phân tích và dự đoán

Nhấp vào điểm đó sẽ hiện các giải pháp để biết được hướng đi tương lai của điểm đó và dự báo giá trị dựa trên mô hình dữ liệu. Chỉ cần chọn kiểu dữ liệu, bất kỳ đối tượng chứa kiểu dữ liệu đó, kiểu hàm trình diễn, thuộc tính và chu kỳ phân tích mà ta muốn tạo. Danh sách các mối đe dọa: 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 ta thêm mã nguồn mở của riêng mình và nguồn cung cấp dữ liệu thanh toán chỉ với vài click chuột mà không cần một cam kết dịch vụ. Splunk còn cộng tác với trung tâm bảo mật Norse Security, một trung tâm bảo mật uy tín toàn cầu.

2.2.3 Thành phần của Splunk

2.2.3.1 Thành phần thu thập Log của Splunk

Đối với bộ công cụ splunk thì thành phần thu thập log được chia làm ba loại là: universal forwarder, heavy forwarder và light forwarder.

Universal forwarder là một streamlined

Nó là phiên bản chuyên dụng của Splunk mà chỉ chứa các thành phần thiết yếu để chuyển dữ liệu từ máy trạm đến máy server. Đây là phiên bản khá gọn nhẹ để tích hợp trên các nguồn sinh log chính vì thế mà nó không bao gồm các tính năng như lập chỉ mục cho dữ liệu và tìm kiếm dữ liệu.

Heavy forwarder

Có kích thước nhỏ hơn một Splunk indexer nhưng vẫn giữ được hầu hết các tính năng ngoại trừ việc tìm kiếm các kết quả phân phối. Và một số thành phần ví dụ Web splunk nếu cần thiết có thể bị vô hiệu hóa để giảm bớt kích thước. Một heavy forwarder phân tích dữ liệu trước khi chuyển và có thể định tuyến dữ liệu dựa trên các đặc điểm như nguồn và các loại sự kiện. Nó cũng có thể đánh chỉ mục cho dữ liệu trên máy cục bộ cũng như chuyển dữ liệu đến một splunk đặc biệt khác.

Light forwarder

Những loại dữ liệu được chuyển đó là dữ liệu thô, dữ liệu chưa được phân tích và dữ liệu đã được phân tích. Mỗi công cụ chuyển tiếp dữ liệu cho phép chuyển các loại dữ liệu khác nhau. Với universal và light forwarder làm việc với dữ liệu thô và dữ liệu chưa được phân tích. Còn heavy forwarder làm việc với dữ liệu thô hoặc dữ liệu đã được phân tích.

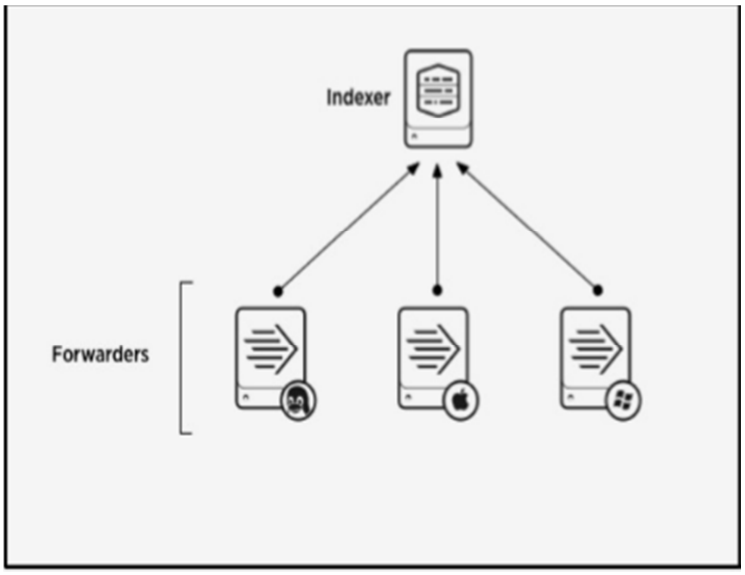
Với dữ liệu thô thì luồng dữ liệu được chuyển tiếp như dữ liệu TCP đơn thuần. Dữ liệu không được chuyển đổi sang định dạng của splunk. Thiết bị chuyển tiếp chỉ lựa chọn dữ liệu và đẩy chúng đi. Với việc chuyển dữ liệu như thế này hữu ích cho việc chuyển dữ liệu sang hệ thống không phải là splunk.

Với dữ liệu không được phân tích thì universal forwarder thực hiện các tiến trình tối thiểu. Mặc dù dữ liệu không được phân tích nhưng nó vẫn được định dạng các thẻ để xác định nguồn, loại nguồn, host. Nó cũng chia các luồng dữ liệu thành các block có kích thước 64K. Thực hiện việc gắn nhãn thời gian lên luồng dữ liệu để bộ phận tiếp nhận có thể phân biệt được dữ liệu khi mà nó không có một mốc thời gian cụ thể. Universal forwarder không xác định, kiểm tra, và gán thẻ lên các sự kiện cá nhân.

Đối với dữ liệu đã được phân tích công cụ heavy forwarder chuyển đổi dữ liệu thành các dạng dữ liệu riêng biệt. Nó sẽ gán thẻ và chuyển dữ liệu đến splunk indexer. Nó cũng có thể kiểm tra các sự kiện. Bởi vì dữ liệu đã được phân tích, sau đó bộ phận forwarder có thể thực hiện việc định tuyến dữ liệu dựa trên sự kiện dữ liệu, chẳng hạn như giá trị của các trường.

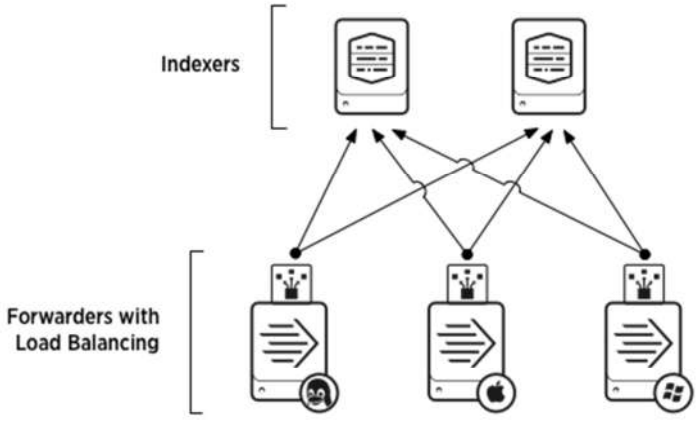
Các mô hình dữ liệu:

- Mô hình chuyển tiếp dữ liệu tập trung



Hình 2.5: Mô hình thu thập log tập trung (nguồn: internet)

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 2.6: Mô hình thu thập log cân bằng tải(nguồn:internet)

- Mô hình cân bằng tải

Mô hình định tuyến và lọc dữ liệu thì công cụ chuyển tiếp sẽ định tuyến dữ liệu đến một hệ thống splunk riêng hoặc một hệ thống thứ ba dựa trên thông tin về nguồn, loại nguồn hoặc các mẫu có sẵn trong bản thân các sự kiện. Tất nhiên hệ thống này yêu cầu bộ công cụ heavy forwarder do cần phải phân tích dữ liệu để lấy thông tin. Ta có thể lọc dữ liệu gửi đi theo yêu cầu ví dụ như chỉ gửi đi những log chứa xâu kí tự error.

2.2.3.2 Thành phần xử lý dữ liệu đầu vào

Những loại dữ liệu được đưa vào Splunk bao gồm: các sự kiện của mạng, nguồn dữ liệu từ hệ điều hành windows, các nguồn khác.

Các sự kiện của mạng

Splunk có thể đánh chỉ mục dữ liệu từ bất kì cổng mạng nào. Ví dụ, Splunk có thể đánh chỉ mục cho dữ liệu từ syslog-nd hoặc bất kì ứng dụng khác có chức năng chuyển dữ liệu theo định dạng TCP. Nó cũng có thể đánh chỉ mục cho dữ liệu ở dạng UDP nhưng ta nên sử dụng TCP để nâng cao độ tin cậy của dữ liệu.

Nó cũng có thể nhận và xử lý các sự kiện SNMP, các cảnh báo được đưa ra từ các thiết bị từ xa.

Nguồn dữ liệu từ windows

Các phiên bản splunk cho windows định nghĩa một loạt các input (đầu vào) riêng biệt cho windows. Nó cũng cung cấp việc đánh số trang trong hệ thống splunk để định nghĩa việc xác định các loại đầu vào riêng biệt đặc trưng cho windows như:

Dữ liệu Windows event log, dữ liệu Windows Registry, dữ liệu WMI, dữ liệu Active Directory, dữ liệu từ các tiện ích được thiết lập giám sát.

Nguồn dữ liệu khác

Splunk cũng hỗ trợ các loại nguồn dữ liệu như: Hàng đợi First-in, First-out, đầu vào từ các script: lấy dữ liệu từ các API và các giao diện từ các dữ liệu từ xa khác, các tin nhắn hàng đợi, module đầu vào: xác định khả năng đầu vào để mở rộng các khung Splunk.

Ngoài những dữ liệu được đưa vào từ các công cụ forwarder người dùng có thể cấu hình để thêm dữ liệu mà ta mong muốn.

Sau khi đã nhận được dữ liệu Splunk tiến hành xử lí các sự kiện được chuyển vào. Các sự kiện là các bản ghi của hành động được lưu trữ lại trong tập tin nhật ký, lưu trữ trong các index. Các sự kiện cung cấp thông tin về các hệ thống, tạo ra các file nhật ký. Dưới đây là một sự kiện của hành động đăng xuất được ghi lại:

172.26.34.223-[01/oct/2020:12:05:27-0700]"GET/trade/app?action=logout HTTP/1.1" 200 2953

Quá trình xử lí các sự kiện bao gồm

- Định dạng bộ kí tự cho dữ liệu đầu vào để phù hợp với định dạng mà Splunk có thể xử lý.
- Quá trình phân mảnh các sự kiện
- Gán nhãn thời gian cho các sự kiện.
- Trích xuất dữ liệu để tạo các trường đánh chỉ mục.

Định dạng bộ kí tự

Người quản trị có thể cấu hình ngôn ngữ cho nguồn dữ liệu. Splunk cung cấp một bộ kí tự để hỗ trợ cho việc chuẩn hóa quốc tế cho việc phát triển Splunk. Nó hỗ trợ nhiều ngôn ngữ bao gồm cả những định dạng không thuộc chuẩn UTF8.

Theo mặc định Splunk áp dụng bộ kí tự UTF-8 cho các nguồn. Nếu một nguồn dữ liệu không thuộc định dạng UTF-8 hoặc không phải tệp tin ASCII thì Splunk sẽ cố gắng chuyển đổi sang định dạng UTF-8 nếu không ta phải định nghĩa việc thiết lập kí tự bằng cách đặt từ khóa CHARSET trong file props.conf.

Các kí tự đã được hỗ trợ trong splunk bao gồm: UTF-8; UTF-16LE; Latin-1; BIG5; SHIFT-JIS

Xử lí các sự kiện nhiều hơn một dòng

Một vài sự kiện chứa nhiều hơn một dòng Splunk sẽ tự động xử lý các sự kiện này theo mặc định. Nếu ta có các sự kiện nhiều dòng mà splunk xử lí không đúng yêu cầu cần phải cấu hình để thay đổi hành vi phá vỡ dòng của Splunk.

Có hai cách để xử lý các sự kiện mà có nhiều dòng:

- Chia nhỏ luồng dữ liệu vào các dòng và gộp lại thành các sự kiện:

Cách này thường đơn giản vì nó thường có các thuộc tính có sẵn để ta có thể định nghĩa các quy tắc hợp nhất dữ liệu. Sử dụng thuộc tính LINE_BREAKER để chia dữ liệu thành nhiều hàng sau đó thiết lập SHOULD_LINEMERGE=true và thiết lập các quy tắc hợp nhất dữ liệu như BREA_ONLY_BEFORE để chỉ ra cách Splunk lắp ráp dữ liệu đó thành các sự kiện.

- Chỉ chia nhỏ các luồng dữ liệu:

Bằng cách sử dụng tính năng LINE_BREAKER và không cần hợp nhất chúng lại. Việc này giúp tăng tốc độ đánh chỉ mục trên dữ liệu nhưng khi làm việc với các dữ liệu này sẽ khó khăn hơn. Nó phù hợp với những người quản trị đã có kinh nghiệm với việc xử lí dữ liệu này và phải thu thập một lượng lớn các log. Với cách này chỉ cần sử dụng thuộc tính LINE_BREAKER và đặt SHOULD_LINEMERGE=false.

Ví dụ việc chỉ định chia nhỏ sự kiện

[my_custom_sourcetype]

BREAK_ONLY_BEFORE = ^\d+\s*$

Với cấu hình trên thì nó chỉ định cho Splunk phân chia các sự kiện bằng cách giả định bất kì dòng chứa các chỉ số bắt đầu một sự kiện mới. Nó sẽ áp dụng cho bất kì loại nguồn được đặt trong “my_custom_sourcetype”.

Cấu hình nhãn thời gian (timestamps)

Timestamps là rất quan trọng đối với Splunk. Nó sử dụng nhãn thời gian để tương quan các sự kiện theo thời gian, để tạo các biểu đồ thời gian trong Web splunk, và thiết lập các khoảng thời gian cho việc tìm kiếm.

Splunk gán các timestamp một cách tự động trong thời gian nó đánh chỉ mục sử dụng thông tin từ dữ liệu sự kiện thô. Nếu một sự kiện mà không chứa thời gian cụ thể, nó sẽ gán dựa theo các cách thức khác. Một vài dữ liệu có thể cần định nghĩa ra cách đánh timestamp.

Trích xuất các trường được đánh index trong Splunk

Khi Splunk đánh chỉ mục dữ liệu, nó phân tích dữ liệu trong một chuỗi các sự kiện. Một phần của tiến trình này, nó thêm một số trường vào sự kiện dữ liệu.

Các trường đó bao gồm những trường mặc định được tự động thêm vào và bất kì trường nào do người dùng định nghĩa.

Quá trình thêm các trường tới sự kiện được gọi là trích xuất trường (field extraction). Có hai loại field extraction là:

- *Indexed field extraction,* Splunk lưu trữ các trường này trong index và coi nó như là một phần của dữ liệu sự kiện.

- *Search-time field extraction,* các trường chỉ được thêm vào trong quá trình tìm kiếm mà không được lưu trữ trong index.

Bảng 2.1: Các trường trong Index

Loại trường	Danh sách các trường	Mô tả
Trường nội bộ	_raw,_time, _indextime _cd	Trường này chứa thông tin mà Splunk sử dụng cho các tiến trình nội bộ
Trường mặc định cơ bản	Host, index, linecount, punct, source, sourcetype, splunk_server, timestamp	Trường này cung cấp các thông tin cơ bản về một sự kiện, chẳng hạn như nơi sinh ra sự kiện, loại dữ liệu mà nó chứa, vị trí index, có bao nhiêu dòng mà nó chứa và nó được sinh ra khi nào
Các trường thời gian mặc định	Date_hour, date_mday, Date_minute, Date_mounth, Data_second, date_wday, date_year, date_zone	Các trường này cung cấp thêm thông tin tìm kiếm các sự kiện theo các timestamp.

Khi Splunk đánh chỉ mục dữ liệu, nó gán thẻ mỗi sự kiện với một số trường. Những trường này sẽ trở thành một phần của sự kiện đã được đánh chỉ mục. Các trường này được tự động thêm vào và được xem như các trường mặc định.

Các trường mặc định phục vụ cho một số mục đích. Ví dụ như, trường index định nghĩa index mà sự kiện được lưu trữ, trường linecount mô tả số dòng mà sự kiện đó chứa, và timestamp định nghĩa thời gian sự kiện đó xảy ra. Splunk sử dụng giá trị trong một vài trường đặc biệt là sourcetype, khi đánh chỉ mục dữ liệu giúp tạo ra các sự kiện đúng. Khi một dữ liệu đã được đánh chỉ mục, có thể sử dụng các trường mặc định đó để thực hiện việc tìm kiếm.

Định nghĩa về host, source, và sourcetype host

Một máy chủ của sự kiện thường là tên máy, địa chỉ IP, hoặc tên miền đầy đủ của các máy chủ mạng mà sự kiện đó sinh ra. Giá trị máy chủ cho phép dễ dàng xác định vị trí dữ liệu sinh ra từ các thiết bị cụ thể.

Source: nguồn của sự kiện là tên của tệp tin, luồng hoặc các đầu vào khác nơi mà các sự kiện được sinh ra. Các dữ liệu được giám sát từ các tệp tin và thư mục, giá trị của nguồn là các đường dẫn cụ thể như "/archive/server1/var/log/messages.0" hoặc "/var/log/". Giá trị của nguồn cho dữ liệu từ mạng là giao thức, cổng chẳng hạn như UDP:514.

Sourcetype: các loại nguồn của một sự kiện là định dạng của dữ liệu đầu vào mà dữ liệu sinh ra, chẳng hạn như "access_combined" hoặc "cisco_syslog". Loại nguồn xác định các Splunk định dạng dữ liệu.

So sánh giữa source và sourcetype

Ta không được nhầm lẫn giữa source và sourcetype, cả hai đều là các trường mặc định nhưng chúng có những điểm khác nhau quan trọng như:

- Source là tên của tệp tin, luồng và các đầu vào khác xuất phát từ nguồn gốc các sự kiện đặc biệt.
- Sourcetype chỉ rõ định dạng của sự kiện. Splunk sử dụng trường này để xác định làm thế nào để định dạng luồng dữ liệu đi vào thành các sự kiện khác nhau.

Các sự kiện với cùng một loại nguồn có thể đến từ các nguồn khác nhau. Ví dụ người quản trị đang giám sát "source=/var/log/messages" và nhận đầu vào syslog trực tiếp từ udp: 514. Nếu tìm kiếm "sourcetype=linux_syslog", Splunk sẽ trả lại các sự kiện từ cả hai nguồn này.

Phân đoạn các sự kiện (event segmentation)

Segmentation là cái mà Splunk sử dụng để phá vỡ các sự kiện thành các phân đoạn tìm kiếm tại các chỉ số thời gian và thời điểm tìm kiếm. Các phân đoạn có thể được chia ra thành major và minor. Trong đó các sự kiện ban đầu có thể được chia ra thành các phân đoạn lớn gọi là major và minor là các phân đoạn nhỏ hơn nữa được chia ra từ major. Ví dụ địa chỉ IP 192.168.2.223 là một phân đoạn major, và major này lại được phá vỡ thành các phân đoạn nhỏ hơn như 192 hay một nhóm các phân đoạn nhỏ như 192.168.2.

Người quản trị có thể định nghĩa cách phân đoạn các sự kiện. Điều này là rất quan trọng bởi vì index-time segmentation ảnh hưởng đến việc đánh chỉ mục, tốc độ tìm kiếm, kích thước lưu trữ và khả năng sử dụng tính năng typehead (Splunk cung cấp các mục mà phù hợp với văn bản nhập vào thanh tìm kiếm). Search-time

segmentation, mặt khác, ảnh hưởng đến tốc độ tìm kiếm và khả năng tạo các tìm kiếm bằng việc chọn các mục từ kết quả được hiển thị trong Web splunk.

Quá trình phân đoạn các sự kiện có ba loại chính sau:

Inner segmentation: Phá vỡ các sự kiện thành các phân đoạn nhỏ nhất có thể. Ví dụ như với địa chỉ 192.168.2.223 tiến hành thông qua inner segmentation nó sẽ được chia thành 192, 168, 2, 223. Loại phân đoạn này dẫn đầu trong việc đánh chỉ mục, tìm kiếm và giảm bộ nhớ sử dụng. Tuy nhiên nó hạn chế tính năng type ahead, vì vậy người dùng chỉ có thể tìm kiếm ở mức phân đoạn nhỏ.

Outer segmentation: Thì ngược lại với inner segmentation, với loại này thì Splunk chỉ đánh chỉ mục các phân đoạn lớn (major). Chẳng hạn như với địa chỉ 192.168.2.223 được đánh chỉ mục như 192.168.2.223 có nghĩa là không thể tìm kiếm trên từng cụm nhỏ. Ta vẫn có thể sử dụng các kí tự đại diện để tìm kiếm trên các điểm nhỏ. Ví dụ, Ta có thể tìm kiếm theo 192.168* kết quả trả về sẽ hiển thị bất kì sự kiện có địa chỉ bắt đầu là 192.168.

Full segmentation: Là sự phối hợp của hai loại phân đoạn trên. Khi đó địa chỉ IP được đánh chỉ mục ở cả phân đoạn lớn và nhỏ. Nó mang lại hiệu suất đánh chỉ mục nhỏ nhất nhưng cung cấp nhiều lựa chọn cho việc tìm kiếm.

Tệp tin "segmenters.conf" ở trong thư mục "$SPLUNK_HOME/etc/system /default" định nghĩa tất cả các loại phân đoạn sẵn có. Theo mặc định thì việc phân đoạn đánh chỉ mục được thiết lập cho loại indexing, được kết hợp cả inner và outer segmentation. Phân đoạn tìm kiếm được thiết lập là full segmentation.

2.2.3.3 Thành phần đánh chỉ mục và lưu trữ

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. Có nhiều công cụ thực hiện công việc này ví dụ như elastic trong bộ công cụ logstash, hay splunk indexer trong bộ công cụ splunk.

Đối với công cụ splunk thì việc đầu tiên là splunk sẽ phải cung cấp dữ liệu, khi đã nhận được dữ liệu nó sẽ đánh chỉ số và làm cho chúng sẵn sàng để tìm kiếm. Với universal indexer được tích hợp thì splunk sẽ biến đổi dữ liệu thành một loạt các sự kiện liên quan đến từng lĩnh vực tìm kiếm. Ta có thể xử lí dữ liệu trước và sau khi splunk đánh chỉ số cho nó, nhưng điều này thường là không cần thiết.

Sau khi đánh index có thể bắt đầu tìm kiếm dữ liệu, hoặc sử dụng nó để tạo báo cáo, biểu đồ, cảnh báo hoặc nhiều công việc khác.

Những loại dữ liệu mà splunk có thể đánh chỉ mục thường là bất kì một loại dữ liệu nào như windows event logs, webserver log, log từ các ứng dụng đang chạy, log từ hệ thống mạng, log giám sát, tin nhắn hàng đợi, tệp tin archive, hoặc bất kì nguồn nào có thể hữu ích.

Khi nguồn dữ liệu chuyển dữ liệu đầu vào splunk ngay lập tức đánh chỉ mục và đưa chúng đến các chuỗi dữ liệu để người dùng có thể tìm kiếm chúng ngay lập tức. Nếu như kết quả tìm kiếm không thỏa mãn yêu cầu, người quản trị có thể cấu hình lại cách đánh chỉ mục sao cho phù hợp.

Quá trình đánh index

Event processing xảy ra qua hai giai đoạn là phân tích và đánh chỉ mục. Tất cả dữ liệu khi được đưa đến splunk đều được đẩy vào đường ống phân tích như các khối lớn khoản 10,000 byte. Trong quá trình phân tích splunk chuyển đổi các khối này thành các sự kiện phù hợp, nơi mà kết thúc tiến trình đánh chỉ mục cho các sự kiện.

Trong quá trình phân tích, splunk thực hiện một vài hành động bao gồm: Trích xuất dữ liệu đặc trưng của mỗi dự kiện bao gồm host, source, sourcetype.

Cấu hình các kí tự để thiết lập mã.

Xác định việc chấm dứt dòng sử dụng quy tắc linebreaking. Có nhiều sự kiện ngắn chỉ một đến hai dòng nhưng cũng có những sự kiện chiếm rất nhiều dòng.

Xác định "tem" thời gian hoặc tạo chúng nếu chúng không tồn tại. Trong cùng thời gian xử lý "tem" thời gian, splunk xác định ranh giới các sự kiện.

Splunk có thể thiết lập để che dấu các sự kiện nhạy cảm như số thẻ tín dụng, số an sinh xã hội. Nó cũng có thể được cấu hình để áp dụng siêu dữ liệu trong các sự kiện sắp tới.

Trong quá trình đánh chỉ mục, splunk thực hiện các tiến trình sau:

- Chia nhỏ các sự kiện thành các phân đoạn từ đó phục vụ cho việc tìm kiếm. Ta có thể tự định nghĩa kích thước của các phân đoạn tùy theo nhu cầu về tốc độ đánh chỉ mục và tìm kiếm, cũng như chất lượng tìm kiếm và hiệu năng của đĩa.
- Xây dựng cấu trúc dữ liệu chỉ mục. Ghi dữ liệu thô và các file index ra đĩa.

Định nghĩa Index: Splunk lưu trữ tất cả dữ liệu mà nó xử lí dưới dạng các index. Một index là một tập hợp các cơ sở dữ liệu với các thư mục con nằm ở SLPUNK_HOME/var/lib/splunk. Các index chứa hai file là dữ liệu thô và file index.

Các loại index mặc định là:

- Main: tất cả dữ liệu xử lí 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 lí.

- _ 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.

Một người quản trị splunk có quyền tạo một index, sửa đổi hay xóa bỏ hoặc thay thế một index đã tồn tại. Việc quản lí index được làm qua Web, CLI, và file cấu hình như index.conf.

Quản lí index

Khi dữ liệu được thêm vào indexer xử lí 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.

Một index là một tập hợp các thư mục và tệp tin. Các thư mục index còn được gọi là các bucket.

Để xem danh sách các index trong giao diện quản trị web ta truy nhập vào settings sau đó chọn Indexes. Có 3 loại index sẵn có là main, _internal, _audit.

Tạo ra nhiều loại index khác nhau.

Theo mặc định thì index main sẽ chứa tất cả các sự kiện. 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 ghi lại các sự kiện.

Ta 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 có thể giúp:

Kiểm soát được người dùng truy cập: khi phân quyền cho người dùng theo các role, ta có thể hạn chế người dùng tìm kiếm các thông tin nhạy cảm.

Nếu có các chính sách khác nhau cho các dữ liệu khác nhau ta có thể chuyển dữ liệu từ index này sang index khác tùy theo nhu cầu sử dụng của ta. 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 ta 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 win:

index=win userid=henry.gale

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.

Xóa dữ liệu cũ dựa trên các chính sách 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 ta định nghĩa chính sách hết hạn của mình.

Quản lí việc lưu trữ index

Cách lưu trữ dữ liệu sau khi đánh index

Khi indexer lập chỉ mục cho dữ liệu, nó tạo ra một nhóm các file, các file này chứa hai loại dữ liệu sau:

- Dữ liệu thô ở dạng nén (rawdata)
- Các index (chỉ số) ánh xạ tới dữ liệu thô, cộng thêm một số file metadata. (file index).

Thông thường dữ liệu này được thiết lập lưu trữ trong các thư mục sắp xếp theo thời gian của dữ liệu. Một số thư mục chứa dữ liệu cũ, một số khác có thể chứa dữ liệu mới. Số lượng các thư mục có thể phát triển lên khá lớn, tùy thuộc vào cách ta giới hạn chúng.

Sau một thời gian dài, thông thường là vài năm, các indexer sẽ xóa dữ liệu trong hệ thống. Nếu cần xử lý một lượng lớn dữ liệu, trong đó chứa những dữ liệu quan trọng thì ta cần phải có các hành động để lưu trữ dữ liệu này để tránh bị xóa theo mặc định như sao lưu các dữ liệu đó.

Định nghĩa tuổi của dữ liệu:

Mỗi một thư mục index được xem như là một bucket (bộ chứa).

Một bucket được chuyển qua các giai đoạn như là các độ tuổi sau: Hot; Warm; Cold; Frozen; Thawed.

Một tuổi của bucket là việc chuyển từ giai đoạn hiện tại sang giai đoạn kết tiếp. Đối với dữ liệu vừa được đánh index thì nó sẽ đi tới giai đoạn hot bucket.

Trong giai đoạn này dữ liệu có thể được tìm kiếm và ghi vào. Một index có thể có một vài hot bucket mở trong cùng một thời gian.

Khi một điều kiện nhất định xảy ra (ví dụ như hot bucket đạt đến một kích thước giới hạn hay splunkd 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.

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. 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 flozen 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.

Vị trí lưu trữ các thư mục index:

Mỗi một index sẽ chiếm giữ một thư mục cho riêng chúng ở trong *"$/SPLUNK_HOME/var/lib/splunk"*. Tên của thư mục giống như tên của index. Trong mỗi thư mục này lại chứa một chuỗi các thư mục con được phân chia theo các giai đoạn của bucket (hot/warm, cold hoặc thawed).

Tất cả các thư mục trên đều phải có quyền ghi.

Sao lưu và lưu trữ index

- Đối với dữ liệu đã được đánh chỉ mục có hai cách cơ bản để backup là:
- Liên tục, sao lưu gia tăng các dữ liệu ở giai đoạn warm (incremental backup)
- Sao lưu tất cả dữ liệu, ví dụ trước khi nâng cấp indexer (full backup)
- Quá trình sao lưu index
- Sao lưu theo cách Incremental backup

Khuyến cáo đưa ra là nên backup tất cả các warm bucket mới, bằng cách sao lưu gia tăng theo sự lựa chọn của ta. Nếu ta đang chuyển tiếp các giai đoạn của dữ liệu một cách thường xuyên ta nên backup cả những thư mục dữ liệu ở giai đoạn cold để đảm bảo không bị mất mát dữ liệu khi chuyển giai đoạn từ warm sang cold

mà việc backup chưa kịp xảy ra. Nếu tên thư mục không thay đổi khi chuyển đổi giai đoạn có thể chỉ phải tìm kiếm các thư mục backup trong cold bằng tên.

Bảng sau mô tả vị trí của các thư mục:

Bảng 2.2: Vị trí lưu trữ các thư mục index

Giai đoạn	Vị trí mặc định	Chú ý
Hot	$SPLUNK_HOME/var/lib/splunk / defaultdb/db/*	Có thể có nhiều thư mục con cho mỗi hot bucket.
Warm	$SPLUNK_HOME/var/lib/splunk / defaultdb/db/*	Có các thư mục con riêng biệt cho warm bucket
Cold	$SPLUNK_HOME/var/lib/splunk / defaultdb/colddb/* 18001091	Có nhiều thư mục con cho cold bucket. Khi các warm bucket được chuyển qua cold, chúng thực hiện di chuyển các thư mục nhưng không đổi tên
Frozen	Dữ liệu của frozen sẽ bị xóa hoặc lưu trữ trong thư mục mà ta định nghĩa	Việc xóa là mặc định. Nếu ta cấu hình thì dữ liệu sẽ được lưu trữ
Thawed	$SPLUNK_HOME/var/lib/splunk / defaultdb/thaweddb/*	Vị trí của dữ liệu đã được lưu trữ và sau đó được khôi phục lại.

2.2.3.4 Thành phần cảnh báo

Một cảnh báo là một hành động được kích hoạt dựa trên các kết quả của tìm kiếm. Khi tạo một cảnh báo, cần định nghĩa một điều kiện mà kích hoạt cảnh báo đó. Hành động điển hình là gửi email dựa trên các kết quả tìm kiếm. Ngoài ra cũng có thể chọn các hành động khác như chạy một đoạn mã script hoặc đưa chúng vào trong danh sách các cảnh báo. Với cùng một điều kiện cảnh báo có thể đưa chúng vào nhiều lựa chọn khác nhau như vừa gửi mail vừa chạy script. Để tránh việc gửi cảnh báo quá thường xuyên, ta cũng có thể giới hạn điều kiện cho một cảnh báo. Splunk định nghĩa ba loại cảnh báo là:

- Per result alert: Dựa trên việc tìm kiếm thời gian thực. Điều kiện kích hoạt là bất cứ khi nào việc tìm kiếm trả về một kết quả.

- Scheduled alert. Chạy tìm kiếm theo lịch trình được chỉ định khi tạo cảnh báo. Ta định nghĩa các kết quả của việc tìm kiếm để kích hoạt cảnh báo đó.
- Rolling-window alert. Dựa trên việc tìm kiếm thời gian thực. Điều kiện kích hoạt là tập hợp các kết quả phù hợp của việc tìm kiếm trong một khung thời gian quy định.

Tiếp theo sẽ giới thiệu chi tiết các kịch bản của mỗi loại cảnh báo. Đối với per result alert: cảnh báo khi việc tìm kiếm thời gian thực trả về một kết quả phù hợp với điều kiện. Thông thường, ta định nghĩa một giới hạn điều kiện vì vậy cảnh báo được kích hoạt chỉ trong một khoảng thời gian quy định.

Các ví dụ về một kết quả trả về bao gồm các điều kiện dưới đây:

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

Cần chú ý khi triển khai per result alert trong một hệ thông yêu cầu độ sẵn sàng cao. Nếu một mạng ngang hàng không sẵn sàng, việc tìm kiếm thời gian thực có thể không được đảm bảo. Trong trường hợp này nên sử dụng scheduled alert.

Sử dụng một scheduled alert để đưa ra cảnh báo khi một lịch trình tìm kiếm trả về các kết quả phù hợp với điều kiện được định nghĩa. Một scheduled alert là hữu ích khi việc cảnh báo không cần thiết phải thực hiện luôn.

Một số ví dụ về scheduled alert:

- Kích hoạt cảnh báo chạy hàng ngày, cảnh báo xảy ra khi số lượng kết quả của ngày đó ít hơn 500.
- Kích hoạt cảnh báo theo giờ, giả sử như khi lỗi 404 trong mỗi giờ lớn hơn 100.

Rolling-window alert được sử dụng để giám sát các kết quả của việc tìm kiếm thời gian trong một khoảng thời gian được định nghĩa.

Ví dụ, giám sát các kết quả trong khoảng 10 phút hoặc mỗi giờ. Ví dụ như:

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

Kích hoạt cảnh báo khi một máy chủ không thể chuyển một tệp tin đến máy chủ khác trong vòng một giờ. Có thể thiết lập điều chỉnh điều kiện để thực hiện cảnh báo cho một giờ cho mỗi máy.

Sử dụng bộ điều chỉnh để giới hạn các cảnh báo

Một cảnh báo có thể được kích hoạt thường xuyên dựa trên các kết quả mà việc tìm kiếm trả về. Lịch trình chạy một cảnh báo cũng có thể kích hoạt các cảnh báo thường xuyên. Để giảm bớt hành động này theo yêu cầu của người dùng có hai cách sau:

- Giới hạn thời gian chạy cảnh báo.
- Xác định giá trị các trường mà kết quả tìm kiếm trả về.

Ví dụ, muốn tạo cảnh báo khi một lỗi hệ thống xảy ra, có khoảng 20 hoặc hơn 20 lỗi xảy ra mỗi phút nhưng người quản trị chỉ muốn gửi cảnh báo một lần mỗi giờ. Để giảm bớt số lần cảnh báo trong trường hợp này cần phải cấu hình bộ điều chỉnh cho cảnh báo này như sau:

Bước 1: Từ trang tìm kiếm nhập vào thông tin sau

index=_internal log_level=ERROR

Bước 2: Chọn Save As > Alert

Bước 3: Trong Result Type, chọn Real Time để cấu hình loại per result alert.

Bước 4: Chọn Next

Bước 5: Chọn các hành động muốn kích hoạt

Bước 6: Chọn Throttle

Bước 7: Chập log_level để giới hạn cảnh báo cho trường log_level. Ta có thể cấu hình bộ điều chỉnh để giới hạn nhiều hơn một trường.

Bước 8: Nhập 1 hour là thời gian giới hạn việc kích hoạt cảnh báo.

Bước 9: Nhấp vào Save Đối với việc tìm kiếm theo lịch trình được chạy thường xuyên, ta không muốn thông báo xảy ra cho mỗi lần chạy, có thể cấu hình bộ điều chỉnh để kiểm soát các cảnh báo trong một khung thời gian dài hơn.

Đối với việc tìm kiếm theo thời gian thực, nếu cấu hình chỉ cảnh báo một lần cho mỗi điều kiện kích hoạt, thì ta không cần phải cấu hình bộ điều chỉnh.

Khi ta cấu hình bộ điều chỉnh cho việc tìm kiếm theo thời gian thực, khi bắt đầu có thể đặt khoảng thời gian đưa ra cảnh báo phù hợp, sau đó có thể mở rộng khoảng thời gian nếu cần thiết. Việc này sẽ giúp ngăn chặn nhiều thông báo cho một hành động.

2.2.4 Cách thức hoạt động của Splunk

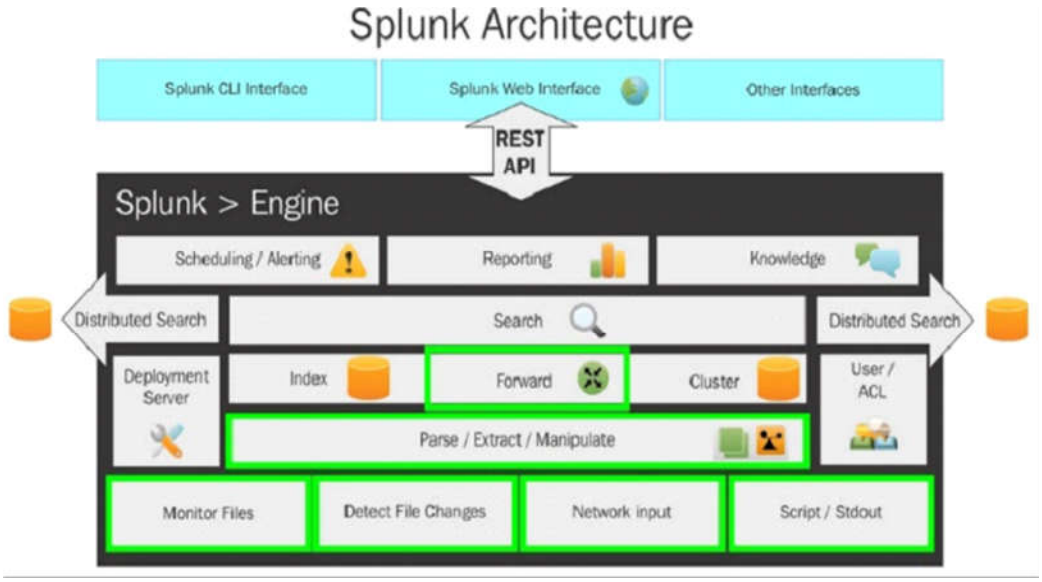
- Mức thấp nhất của kiến trúc Splunk mô tả các phương thức nhập liệu khác nhau được hỗ trợ bởi Splunk. Những phương thức nhập này có thể được cấu hình để gửi dữ liệu trên các bộ phân loại Splunk.

- Trước khi dữ liệu đến được các bộ phân loại Splunk, nó có thể được phân tích cú pháp hoặc thao tác, có nghĩa là làm sạch dữ liệu có thể được thực hiện nếu cần.

- Một khi dữ liệu được lập chỉ mục trên Splunk, nó sẽ tiến hành đi vào cụ thể để phân tích dữ liệu.

- Splunk hỗ trợ hai loại triển khai: triển khai độc lập và triển khai phân tán. Tùy thuộc vào loại triển khai, tìm kiếm tương ứng được thực hiện. Công cụ Splunk có các thành phần bổ sung khác của quản lý dữ liệu, báo cáo và lên kế hoạch và cảnh báo. Toàn bộ công cụ Splunk được tiếp xúc với người dùng thông qua Splunk CLI, Splunk Web Interface, và Splunk SDK, được hỗ trợ bởi hầu hết các ngôn ngữ.

- Splunk cài đặt một quy trình máy chủ phân tán trên máy chủ được gọi là splunkd. Quá trình này có trách nhiệm lập chỉ mục và xử lý một số lượng lớn dữ liệu thông qua các nguồn khác nhau. Splunkd có khả năng xử lý số lượng lớn dữ liệu phát trực tuyến và lập chỉ mục cho phân tích thời gian thực trên một hoặc nhiều đường ống (Pipeline)



Hình 2.7: Cơ chế hoạt động của Splunk (nguồn:internet)

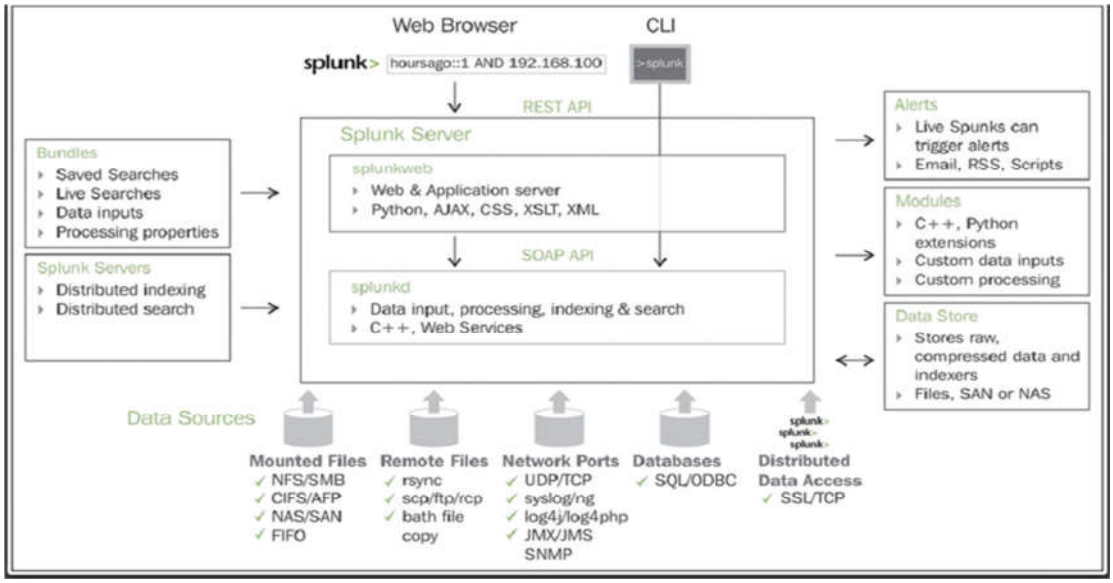
- Danh sách dưới đây là các khối kiến trúc splunk:

- **Pipeline:** Đây là một quá trình cấu hình đơn luồng duy nhất nằm trong splunk.
- **Bộ vi xử lý:** Chúng là những hàm số có thể tái sử dụng cá nhân hoạt động trên dữ liệu đi qua một đường ống. Đường ống trao đổi dữ liệu giữa họ thông qua một hàng đợi.

- Splunk cho phép người dùng tìm kiếm, điều hướng và quản lý dữ liệu trên Splunk Enterprise thông qua giao diện web được gọi là Splunk Web.

- Một trong những thành phần quan trọng của kiến trúc của Splunk là kho dữ liệu. Nó có trách nhiệm nén và lưu trữ dữ liệu ban đầu (nguyên vẹn). Dữ liệu được lưu trữ trong các tệp Time Series Index (T SIDX).

- Các triển khai của Splunk Enterprise có thể bao gồm từ việc triển khai các máy chủ đơn (có chỉ số vài gigabyte dữ liệu mỗi ngày và được truy cập bởi một vài người dùng đang tìm kiếm, phân tích và hình dung dữ liệu) tới các triển khai lớn của doanh nghiệp ở nhiều trung tâm dữ liệu, lập chỉ mục hàng trăm terabytes dữ liệu và tìm kiếm được thực hiện bởi hàng trăm người dùng. Splunk hỗ trợ giao thức truyền thông TCP chuyển tiếp dữ liệu từ một máy chủ Splunk sang một máy khác để lưu trữ dữ liệu và các yêu cầu phân phối và phân phối dữ liệu khác thông qua giao tiếp TCP Splunk-to-Splunk.



Hình 2.8: Sơ đồ hoạt động của Splunk (nguồn:internet)

- Bundles là các thành phần của kiến trúc Splunk lưu trữ cấu hình dữ liệu đầu vào, tài khoản người dùng, ứng dụng Splunk, tiện ích và môi trường khác.

- Các Modul là những thành phần của kiến trúc Splunk được sử dụng để thêm các tính năng mới bằng cách sửa đổi hoặc tạo bộ xử lý và đường ống (pipeline). Các Modul chỉ là các kịch bản tùy chỉnh và các phương pháp nhập dữ liệu hoặc phần mở rộng có thể thêm một tính năng mới hoặc sửa đổi các tính năng hiện có của Splunk.

2.3 Kết luận chương 2

Trong chương 2, luận văn đã nghiên cứu các giải pháp giám sát an toàn thông tin cần phải đáp ứng và thực hiện. Mỗi một giải pháp sẽ giám sát một thành phần nhất định trong hệ thống mạng. Từ các thành phần cần giám sát, luận văn đã đưa ra giải pháp áp dụng công cụ để giám sát tập trung trong thực tế.

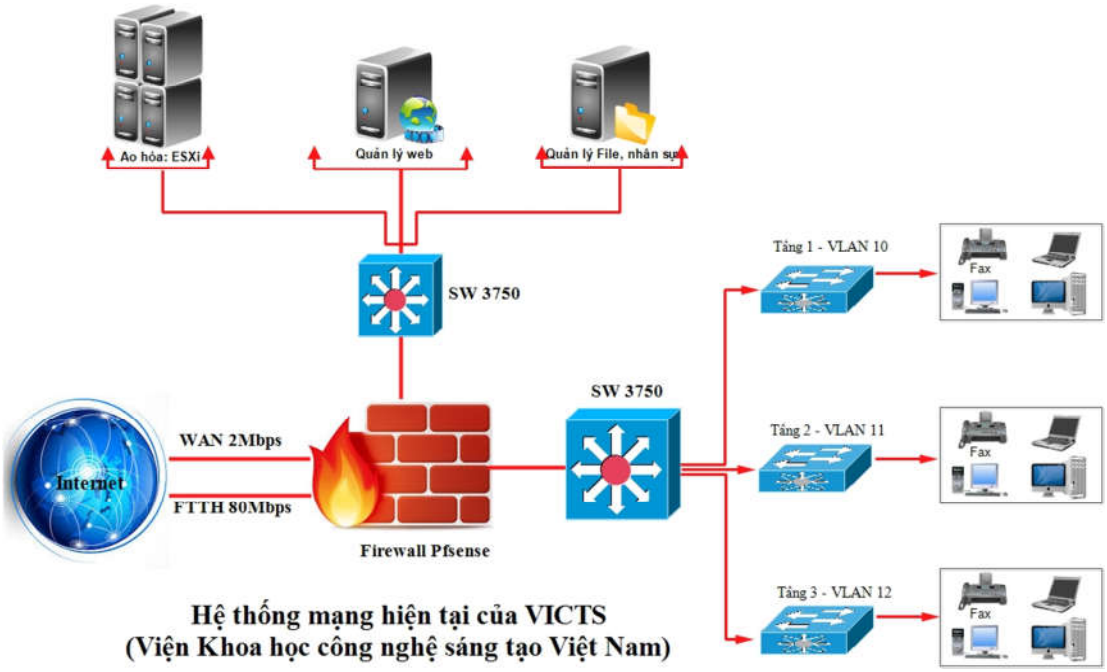
Trong chương 3, luận văn sẽ ứng dụng giải pháp giám sát an toàn thông tin bằng Splunk để xây dựng hệ thống giám sát cho hệ thống mạng Viện Khoa học công nghệ sáng tạo Việt Nam.

CHƯƠNG 3. XÂY DỰNG HỆ THỐNG GIÁM SÁT AN TOÀN THÔNG TIN CHO HỆ THỐNG MẠNG VIỆN KHCN SÁNG TẠI VIỆT NAM

Chương 3 của luận văn sẽ nghiên cứu đề xuất một số giải pháp an toàn thông tin cho hệ thống mạng của Viện Khoa học Công nghệ Sáng tạo Việt Nam. Luận văn cũng thực hiện một số thử nghiệm đánh giá hiệu quả các giải pháp an toàn thông tin. Kết quả thử nghiệm được trình bày trong phần phụ lục.

3.1 Khảo sát mạng nội bộ Viện KHCN Sáng tạo Việt Nam

3.1.1 Chức năng, trang thiết bị và mô hình hiện có của hệ thống mạng Viện KHCN Sáng tạo Việt Nam



Hình 3.1: Mô hình hoạt động của hệ thống mạng tại Viện KHCN Sáng tạo Việt Nam

Hệ thống mạng hiện tại đang sử dụng kiến trúc mô hình mạng Client - Server nhằm chia sẻ dữ liệu từ các máy chủ tới các máy con. Với kiến trúc mạng hình sao ở các tầng, ta sẽ đạt được tốc độ nhanh nhất có thể, kiểm soát tốt khi xảy ra lỗi cũng như mở rộng tùy ý muốn trong toàn hệ thống.

Hạ tầng mạng được phân cấp: máy tính của các phòng ban sẽ kết nối tới các Switch của các tầng, từ Switch các tầng kết nối tới Switch tổng của tòa nhà. Switch tổng kết nối tới Firewall rồi ra ngoài Internet.

Hệ thống máy chủ Web, Mail, File kết nối vào Core Switch. Hệ thống Core Switch đặt sau Firewall nên rất an toàn. Từ Firewall sẽ chia cổng ngoài Internet.

Về trang thiết bị và số lượng người sử dụng:

- Số lượng phòng ban và các đơn vị trực thuộc sử dụng máy tính là 8.
- Tổng số máy tính cho cán bộ nhân viên là 125.
- Số lượng máy chủ là 08 máy đặt tập trung: 02 máy quản lý File Server, 03 máy chủ chạy Website của Viện (https://www.victs.vn), 03 máy chủ chạy Tạp chí điện tử (donghanhviet.vn) và các phòng ban, 03 máy chủ chạy ảo hóa trên đó đặt dịch vụ: Email, DHCP và DNS.
- Số lượng Firewall là: 1 Server cài dịch vụ Firewall Pfsense có 4 Card mạng
- Số lượng Switch layer 3 là: 2 Switch 3750
- Số lượng Switch ở các tòa nhà và các tầng là 21 Access Switch và 3 Distribution Swich.
- Số lượng tổng đài nội bộ dùng IP là 1
- Số lượng Camera sử dụng 24 chiếc.
- Số lượng đường truyền là 2 ra ngoài Internet: Viettel (wan) và Fpt (FTTH)

3.1.2 Yêu cầu sử dụng

- Hệ thống phải luôn kết nối được Internet.
- Hệ thống Firewall phải bảo vệ hệ thống máy chủ và người dùng 24/7 .
- Các dịch vụ File, Mail, Web luôn phải ổn định để cán bộ nhân viên trong Viện và khách hàng có thể sử dụng. Luôn luôn kiểm soát được số lượng người truy cập dịch vụ.
- Dữ liệu tại các phòng ban phải được tập trung, không phân tán, dễ quản lý, được phân quyền phù hợp với chức trách.
- Khả năng cung ứng cao, đáp ứng được một lượng lớn kết nối vào trong hay ra ngoài mạng mà vẫn giữ được sự ổn định.
- Có khả năng mở rộng trong tương lai.

3.1.3 Hiện trạng các vấn đề liên quan trong quá trình vận hành, khai thác mạng máy tính tại Viện KHCN Sáng tạo Việt Nam

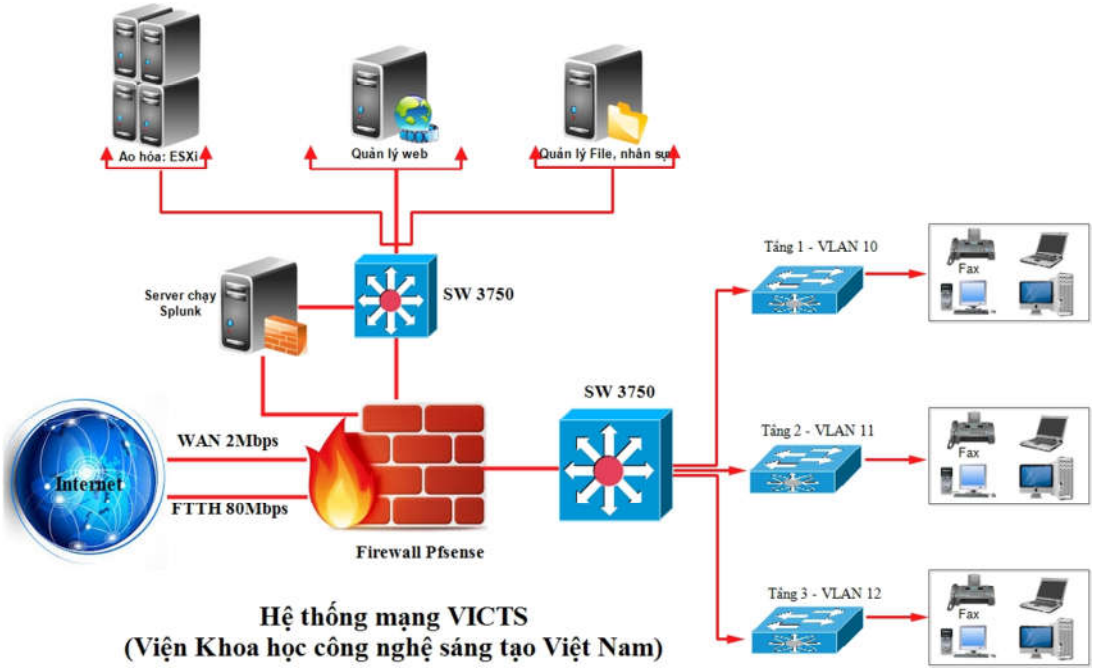
- Vào thời gian cao điểm từ 6h30 tới 8h30 sáng và 13h00 đến 15h00, số lượng người truy cập máy chủ là rất lớn do thời điểm này các độc giả truy cập đọc tin tức từ báo điện tử do Viện quản lý. Do không đo lường, kiểm soát được hiệu

năng của máy chủ dẫn tới không phân luồng kịp thời lưu lượng về máy chủ gây ra chậm hoặc nghẽn khi đông người truy cập một thời điểm.

- Website và phần mềm của Viện có nhiều dữ liệu quan trọng cần phải có giải pháp bảo mật tối ưu.

- Không theo dõi kịp thời Firewal Pfsense dẫn tới không kịp chặn hoặc điều luồng dữ liệu khi cần thiết.

3.2 Kiến nghị đề xuất giải pháp giám sát Splunk cho mạng máy tính tại Viện KHCN Sáng tạo Việt Nam



Hình 3.2: Hệ thống mạng của Viện KHCN Sáng tạo Việt Nam

Để giám sát tập trung, đồng thời kịp đánh giá các trạng thái hoạt động cho hệ thống mạng, tác giả đề xuất giải pháp cài đặt bộ công cụ giám sát tập trung Slpunk để xử lý các bài toán cần phải giám sát. Máy chủ cài đặt Slpunk cần xây dựng bên trong hệ thống máy chủ, cạnh các máy chủ dịch vụ khác như Web, Mail, File đồng thời vẫn giám sát được Firewall và lưu lượng vào ra của phía người dùng từ các phòng ban. Làm như vậy vừa để bảo vệ máy chủ giám sát do có Firewall bảo vệ, vừa thuận tiện cho việc giám sát các máy chủ, dịch vụ,... của cơ quan.

3.3 Cài đặt và vận hành hệ thống

3.3.1 Cài đặt máy chủ giám sát tập trung Splunk

- Truy cập vào file /tmp trên máy chủ: [root@server1 ~]# cd /tmp/

- Tại máy chủ ta tiến hành tạo 1 tài khoản cho Splunk.

[root@server1 tmp]# groupadd splunk

[root@server1 tmp]# useradd -d /opt/splunk -m -g splunk splunk

- Đăng nhập và chỉnh sửa thông tin tài khoản.

[root@server1 tmp]# su – splunk

[splunk@server1 tmp]# id uid=1001(splunk) gid=1001(splunk) group=1001(splunk)

[splunk@server1 tmp]# getconf LONG_BIT

Hình 3.3 Cấu hình thông tin tài khoản

- Sử dụng máy tính cá nhân, truy cập vào trang download Splunk enterprise tại https://www.splunk.com/en_us/download/splunk-enterprise.html .

- Đăng nhập vào tài khản. Nếu chưa có tài khoản, ta tạo 1 tài khoản mới với các thông tin điền vào form "Create Your Account".

- Tải phiên bản linux với đuôi .tgz về máy và sử dụng WinSCP để đưa file vừa tải vào địa chỉ /opt/splunk/ trên server.

[root@server1 tmp]# wget splunk-8.0.5-a1a6394cc5ae-Linux-x84_64.tgz

- Giải nén file vừa tạo và chuyển đến thư mục /opt/splunk trên server.

[root@server1 tmp]# tar -xvf splunk-8.0.5-a1a6394cc5ae-Linux-x84_64.tgz

Hình 3.4 Giải nén file sau khi tải về

[root@server1 tmp]# cp -rp splunk /* /opt/splunk/

[root@server1 tmp]# chown -R splunk: /opt/splunk/

- Đăng nhập vào tài khoản splunk và tiến hành cài đặt.

[root@server1 tmp]# su – splunk

[splunk@server1 ~]$ cd /opt/splunk/bin/

[splunk@server1 bin]$ ./splunk start --accept-license

Hình 3.5 Đăng nhập vào tài khoản splunk và tiến hành cài đặt

Thêm các thông tin tài khoản Splunk cần thiết

Hình 3.6 Bổ sung các thông tin cho tài khoản Splunk

- Mở các port cần thiết cho phép máy tính cá nhân có thể quản lý Splunk.

firewall-cmd --permanent --zone=public --add-port=8000/tcp --add-port=8089/tcp --add-port=8191/tcp --add-port=8065/tcp --add-port=9997/tcp

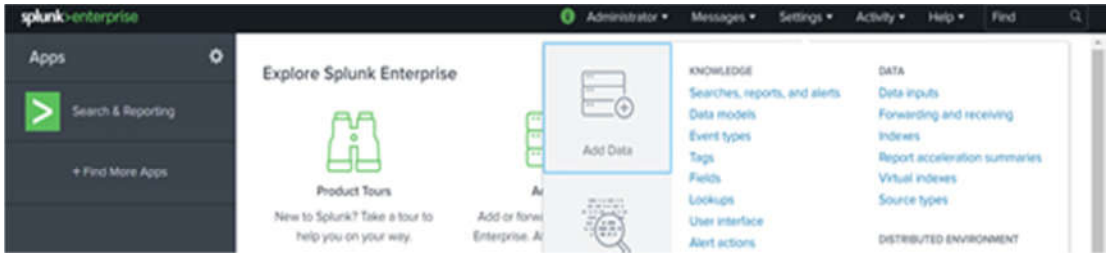
firewall-cmd --reload

- Sử dụng trình duyệt web trong máy tính cá nhân, tiến hành đăng nhập Splunk của web với: https://[địa chỉ server]:8000

- Thiết lập cài đặt tài khoản admin cho lần đầu tiên sử dụng Splunk.

- Sau khi hiển thị lên giao diện chính của Splunk thì ta đã hoàn tất cài đặt.

- Giao diện sau khi cài đăt xong

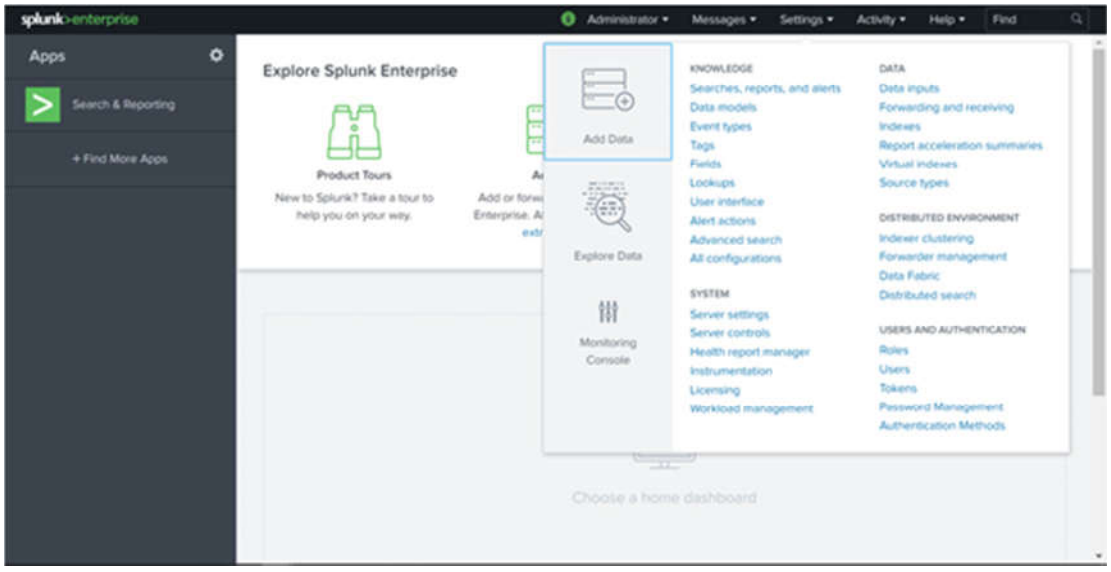


Hình 3.7 Giao diện Slunk sau khi cài đăt

3.3.2 Lấy log từ máy chủ Linux đưa vào Splunk để phân tích

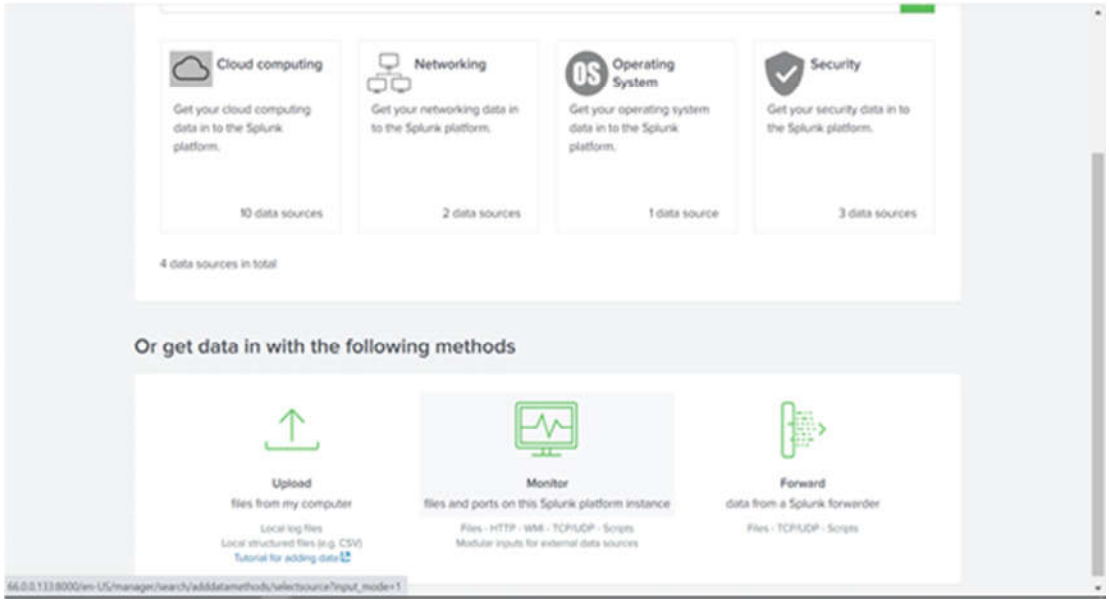
- Sử dụng WinSCP để kiểm tra lại vị trí các file log cần thiết trên server. (thường các file log của server nằm tại vị trí /var/log/).

- Tại trình duyệt web đang hiển thị Splunk ta vào lần lượt các bước: Chọn Settings và chọn Add Data.



Hình 3.8 Giao diện tùy chọn Add data của Splunk

Tại giao diện Add Data, ta chọn Monitor.



Hình 3.9 Giao diện tùy chọn Monitor của Splunk

Trong đó có các lựa chọn sau:

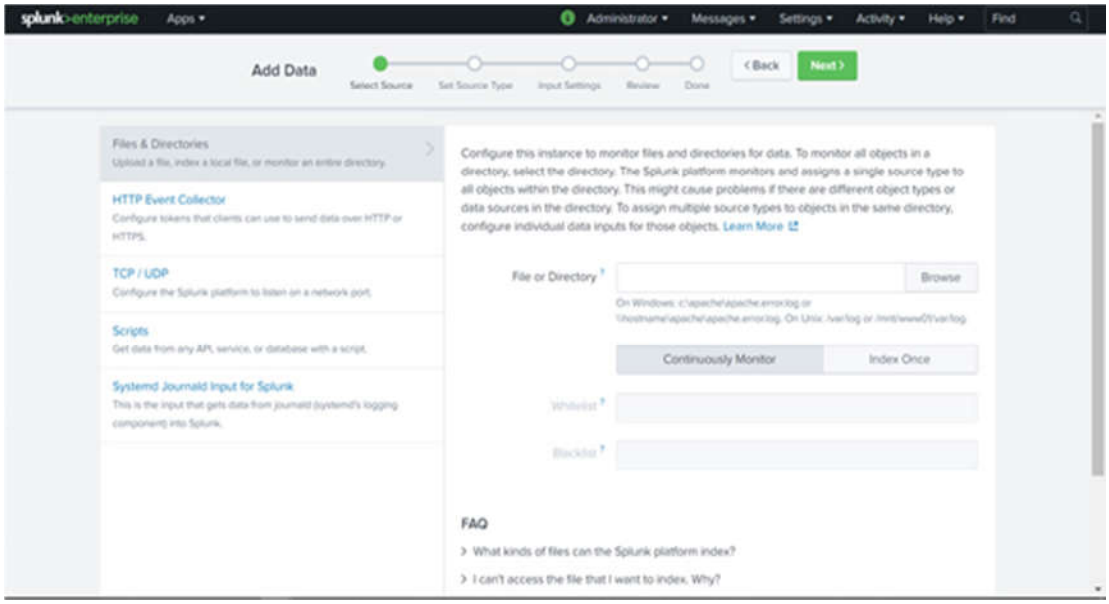
File & Directories : Giám sát files và folders.

HTTP Event Collection : Giám sát luồng dữ liệu khi đi qua HTTP.

TCP/UDP : Giám sát các cổng dịch vụ.

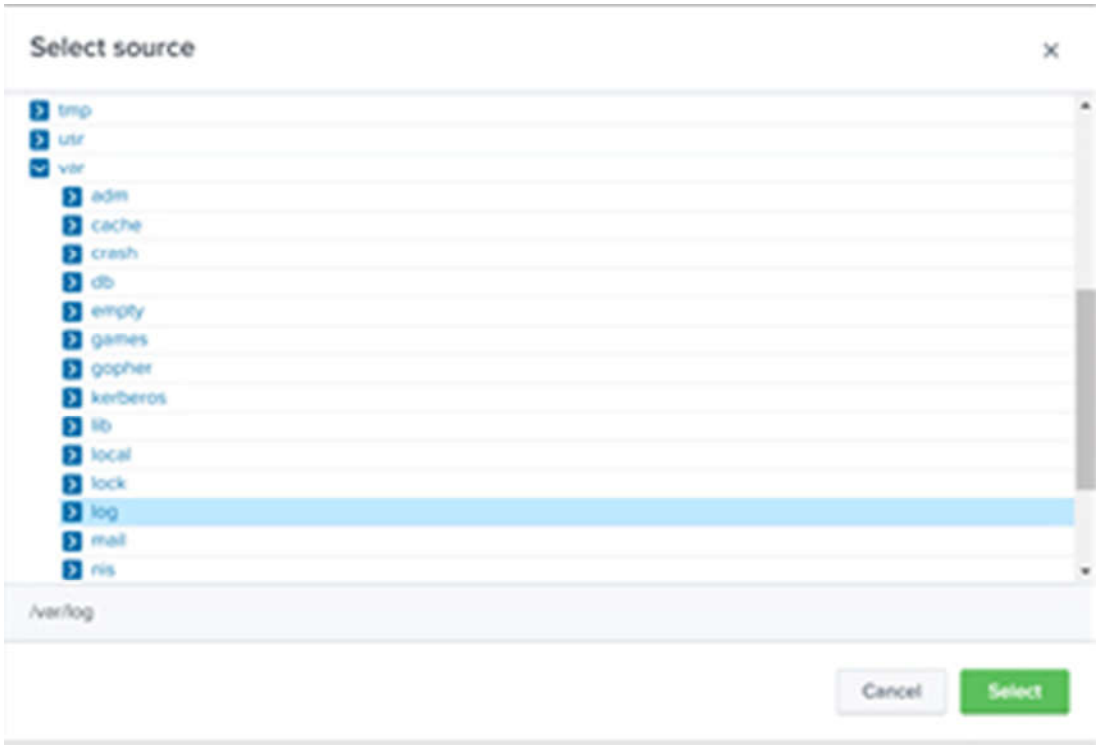
Scripts : Giám sát các script.

Ta lựa chọn vào File & Directories.



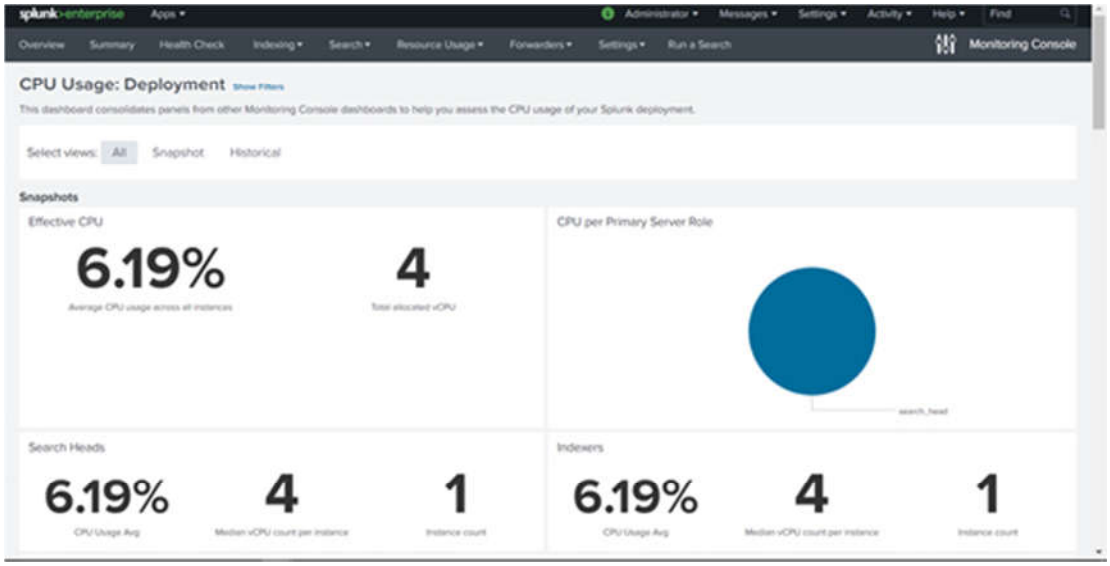
Hình 3.10 Lựa chọn File & Directories để lấy log

- Lựa chọn lần lượt đến các địa chỉ các file mà bạn muốn thu thập log (trong ví dụ là địa chỉ /var/log/) và lựa chọn Select.

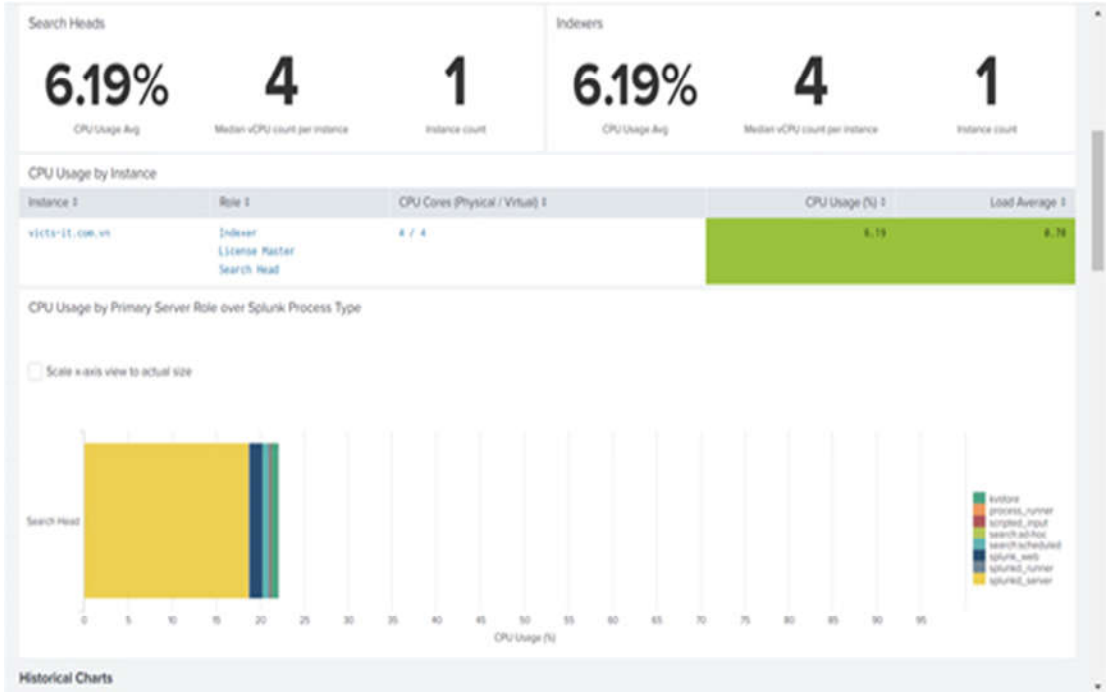


Hình 3.11 Lựa chọn các file để thu thập log

- Tại giao diện Add Data, ta đưa thêm các lựa chọn cần thiết và ấn vào Next.
- Lựa chọn hoặc thay đổi các thông tin cần thiết và nhấn Review.
- Thông tin chung hiển thị trên Splunk:

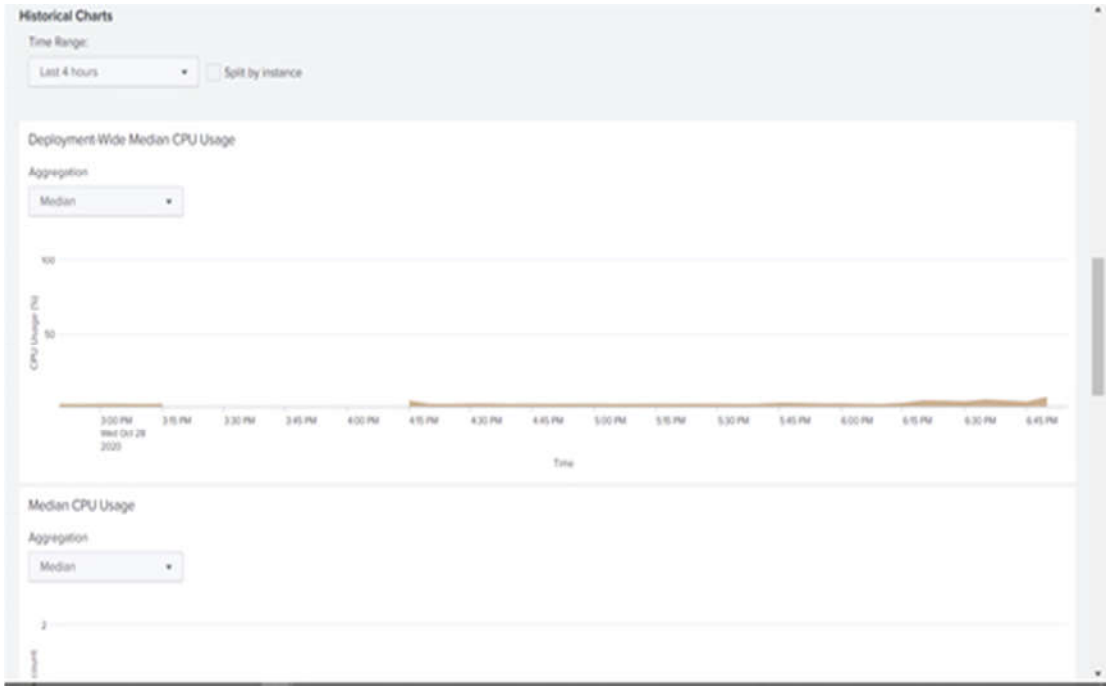


Hình 3.12 Hiển thị thông tin trên Splunk – giao diện 1

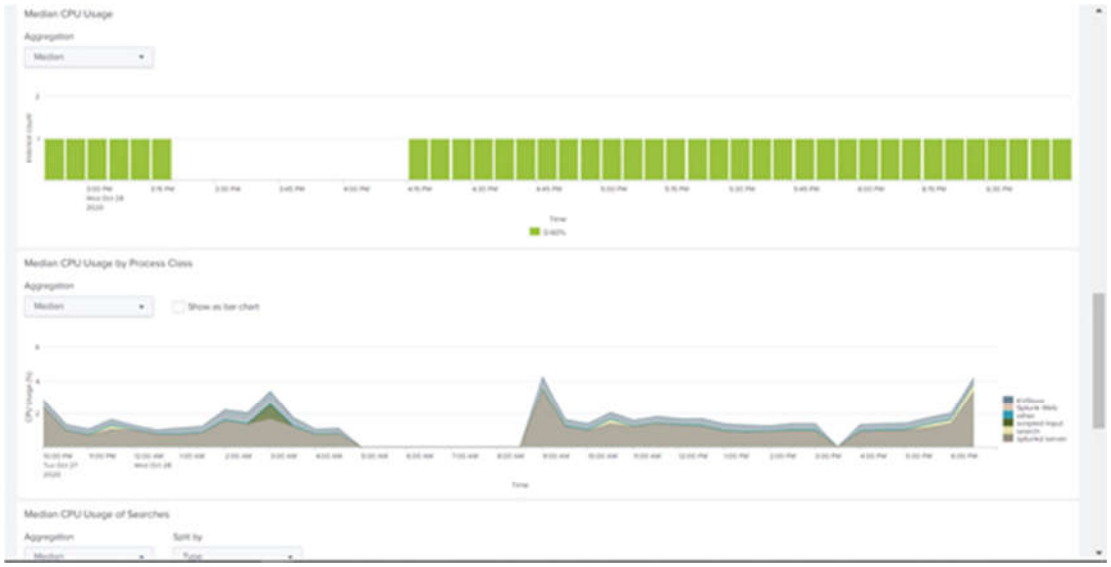


Hình 3.13 Hiển thị thông tin trên Splunk – giao diện 2

- Thông tin về sử dụng CPU:



Hình 3.14 Thông tin hiển thị về sử dụng CPU – giao diện 1



Hình 3.15 Thông tin hiển thị về sử dụng CPU – giao diện 2

Hình 3.16 Thông tin hiển thị về sử dụng CPU – Giao diện 3

- Địa chỉ các key log chính được hiện thị trong Splunk

  /var/log/message : Thông tin chung về hệ thống.

  /var/log/auth.log : Các log về xác thực.

  /var/log/kern.log : Các log về nhân của hệ điều hành.

  /var/log/maillog : Các log về máy chủ email.

  /var/log/httpd : Thư mục log truy cập và lỗi của dịch vụ web Apache.

  /var/log/boot.log : Các log của quá trình khởi động hệ thống.

  /var/log/mysqld.log : Các log của Mysql.

  /var/log/secure : Các log về xác thực.

  /var/log/utmp hoặc /var/log/wtmp : file lưu bản ghi đăng nhập.

  /var/log/yum.log : Các log về file cài đặt trên máy.

- Địa chỉ các key log phụ được hiện thị trong Splunk

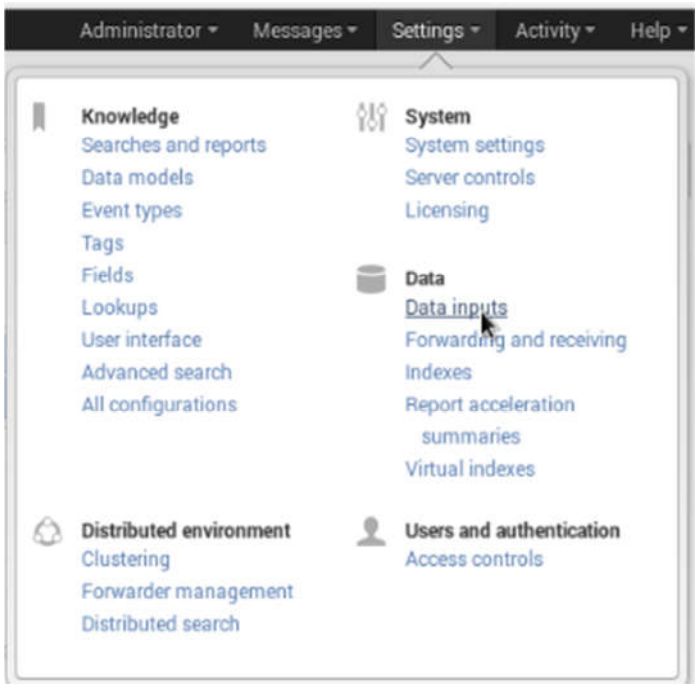
/var/log/cron.log : Các log về dịch vụ Crond (dịch vụ lập trình chạy tự động).

/var/log/qmail : Các log của phần mềm Qmail.

/var/log/lighttpd : Thư mục log truy cập và lỗi của phần mềm Lighttpd.

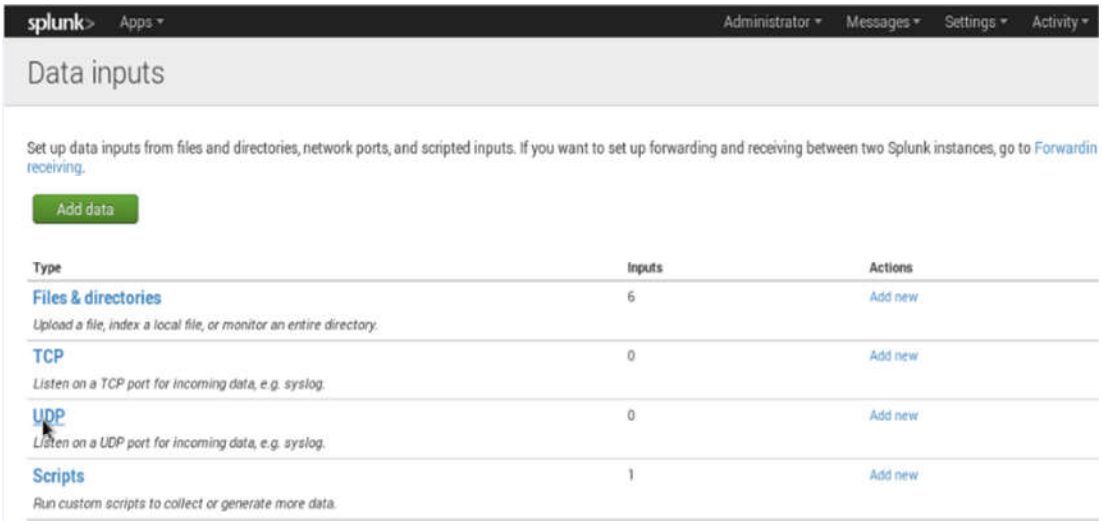
3.3.3 Lấy Log từ máy chủ Firewall Pfsense

- Trên máy Centos đã cài sẵn Splunk:



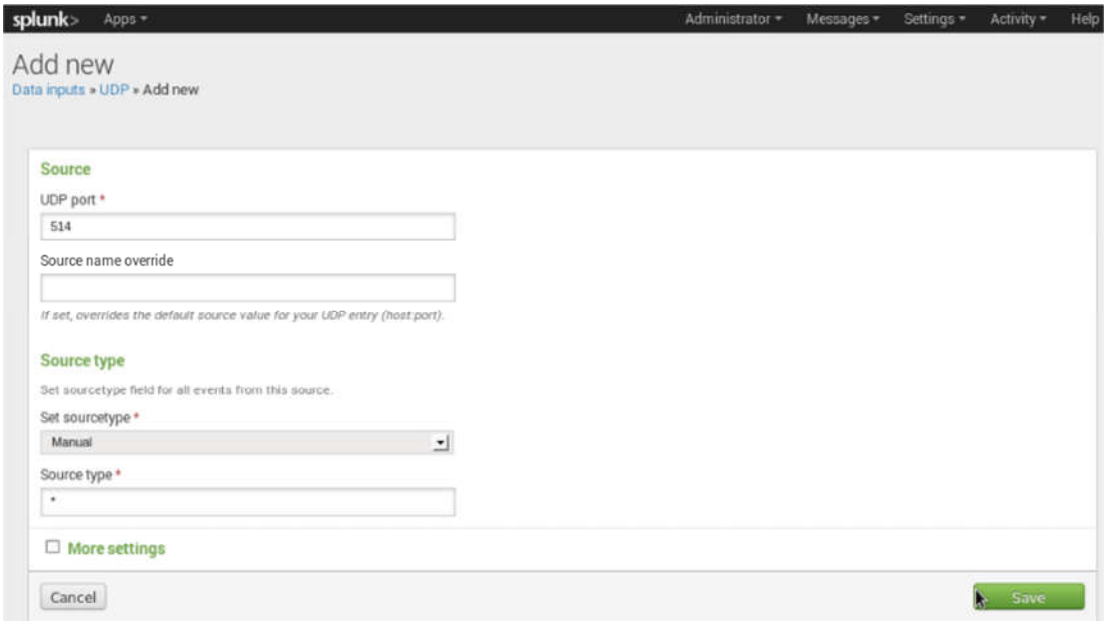
Hình 3.17 Giao diện cấu hình của Splunk

- Chọn Setting =>Data inputs



Hình 3.18 Lựa chọn Data inputs để lấy Log từ máy chủ Firewall Pfsense

- Chọn Add New UDP

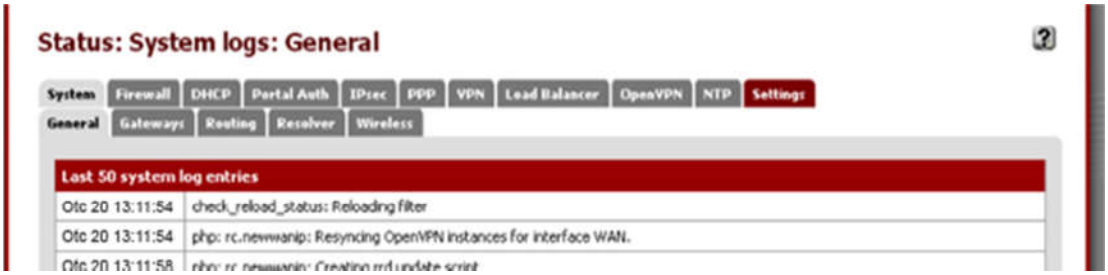


Hình 3.19 Lựa chọn Add New UDP để lấy Log từ máy chủ Firewall Pfsense

- Chọn port nhận các gói tin UDP từ client là 514, Đặt sourcetype là Manual, chọn Save

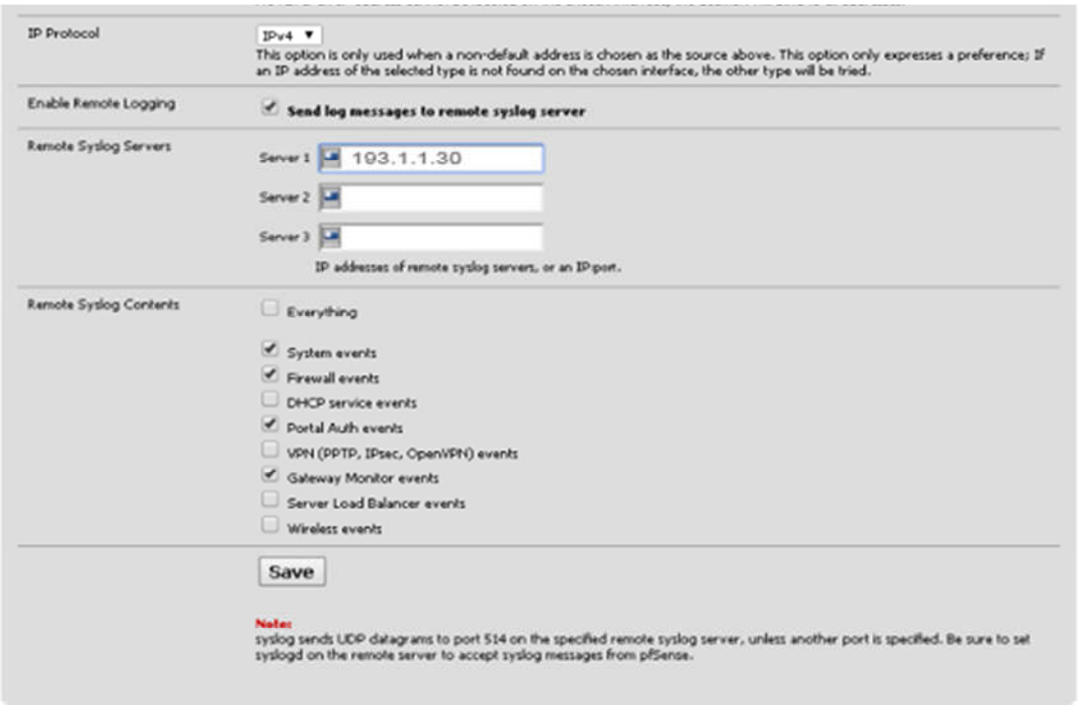
Hình 3.20 Giao diện hiển thị kết quả sau khi lưu port

- Splunk sẽ nhận các gói tin UDP từ port 514
- Trên máy Pfsense: Vào Status=>System Logs



Hình 3.21 Lựa chọn System Logs trên máy Pfsense

- Chọn tab Setting



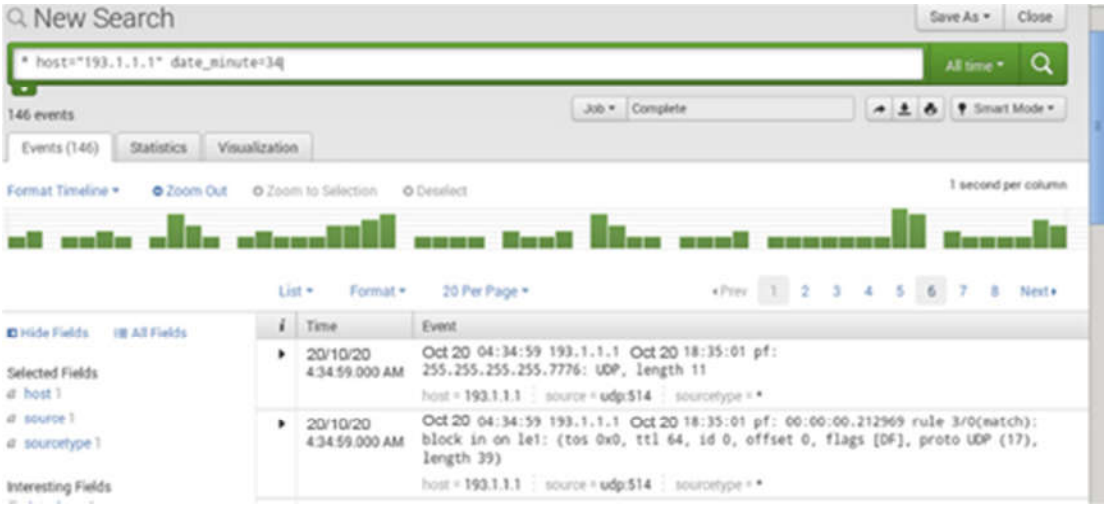
Hình 3.22 Lựa chọn Setting trên máy Pfsense

- Tích vào ô Send log Messages to remote syslog server, IP remote server nhận log là 193.1.1.30 (máy Splunk), tích tùy chọn các log muốn gửi qua Splunk, Chọn Save.

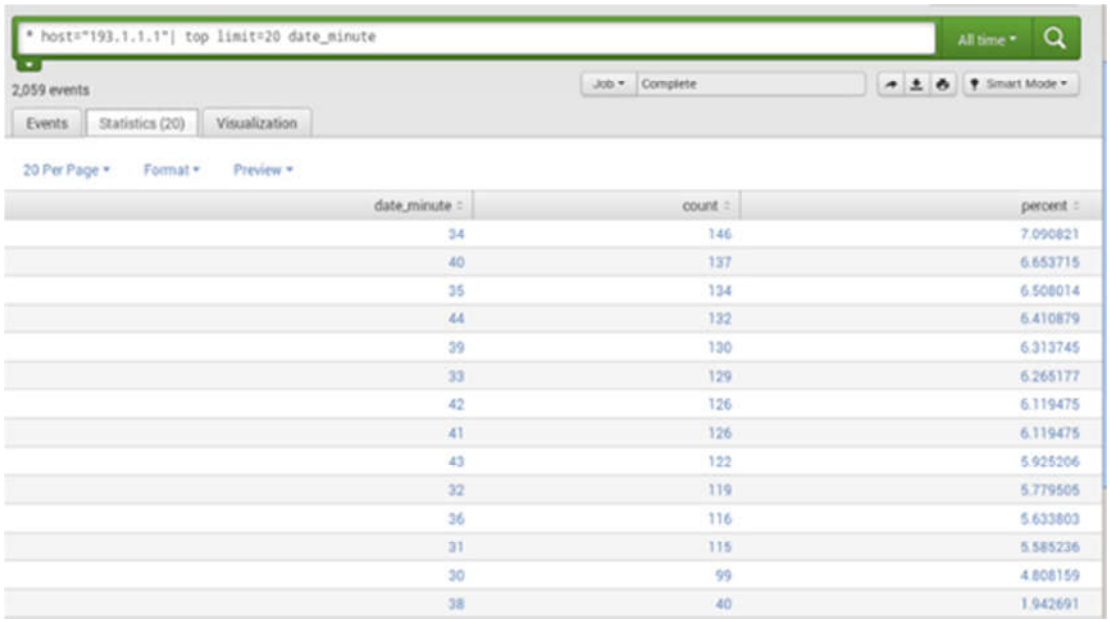
- Lưu ý là Pfsense chỉ gửi log bằng giao thức UDP.

- Kết quả: Splunk đã nhận được log từ Pfsense

- Nhấn thanh search tìm địa chỉ IP 193.1.1.1 của Pfsense



Hình 3.23 Tìm kiếm thành công máy chủ Pfsense trên Splunk - giao diện 1



date_minute	count	percent
34	146	7.090821
40	137	6.653715
35	134	6.508014
44	132	6.410879
39	130	6.313745
33	129	6.265177
42	126	6.119475
41	126	6.119475
43	122	5.925206
32	119	5.779505
36	116	5.633803
31	115	5.585236
30	99	4.808159
38	40	1.942691

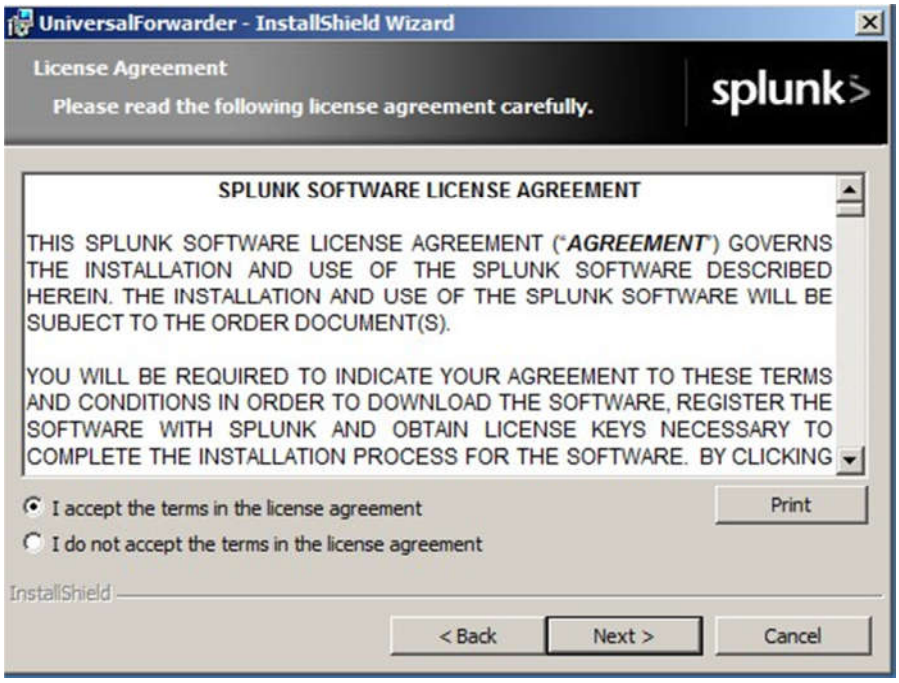
Hình 3.24 Tìm kiếm thành công máy chủ Pfsense trên Splunk - giao diện 2

3.3.4 Lấy Log từ máy chủ Windows Server

Trên máy Window Server 2012:

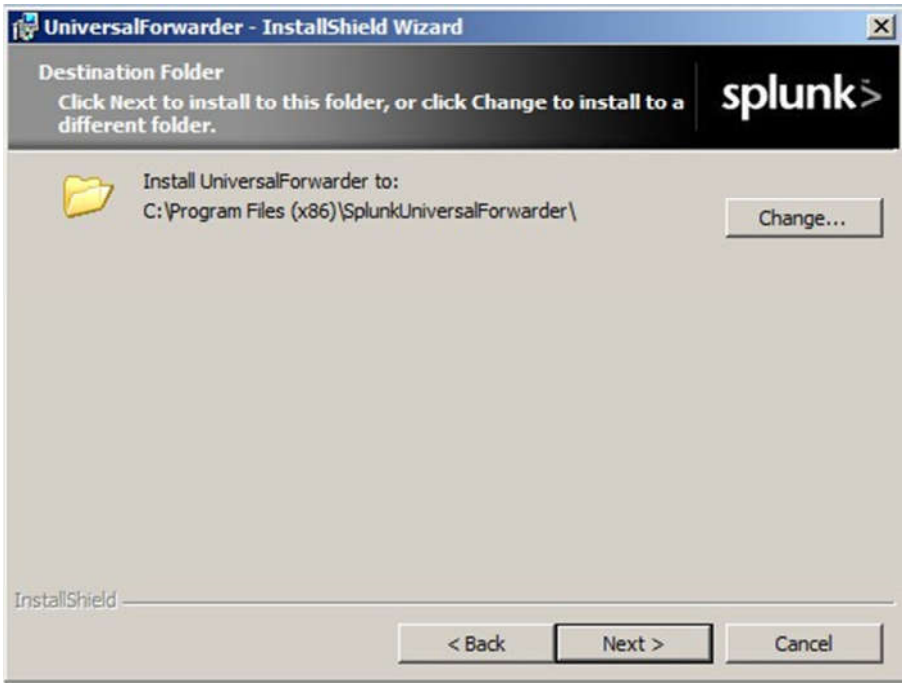
Cài đặt Splunk Forwarder

Chọn chấp nhận các điều khoản của splunk sau đó bấm Next



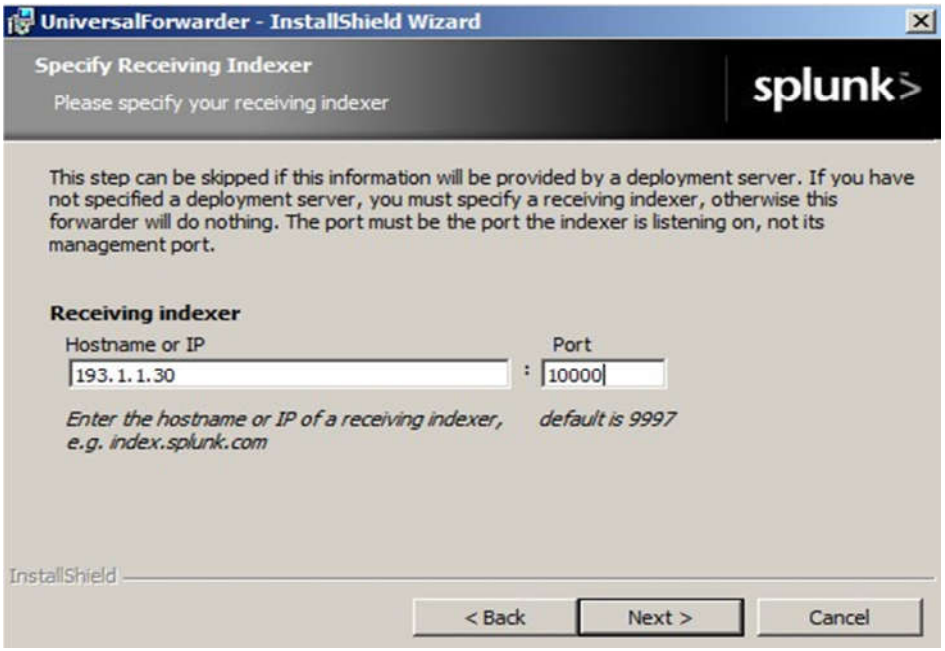
Hình 3.25 Cài đặt Splunk Forwarder để lấy Log từ máy chủ Windows Server

Tiếp tục chọn Next.



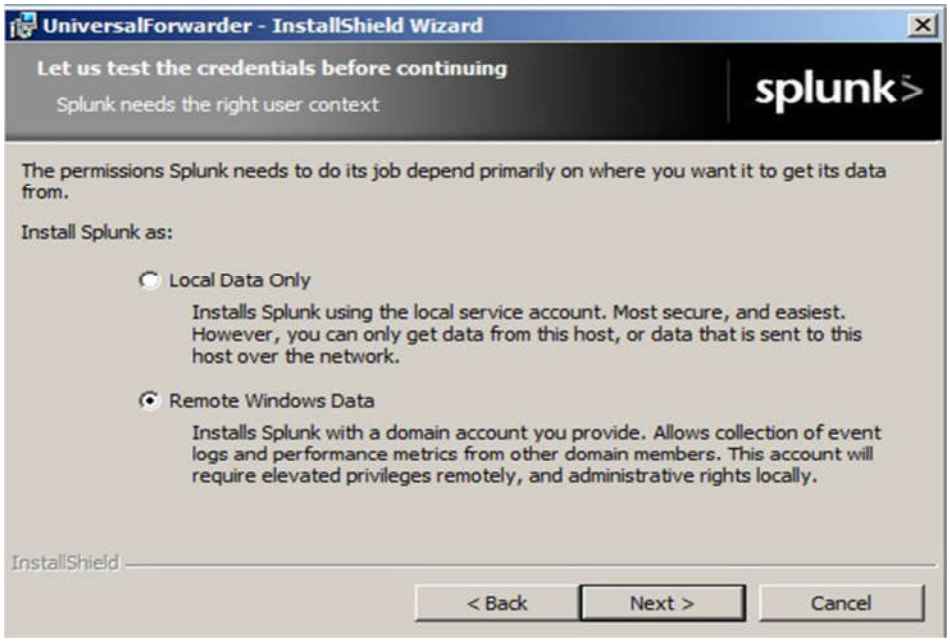
Hình 3.26 Chọn chấp nhận các điều khoản của splunk để cài đặt

Nhập vào địa chỉ của máy chủ splunk và port. Lưu ý: ta chọn port trùng với port sẽ cấu hình trên Splunk (Setting > Forwarding and Receive data) để dữ liệu có thể gửi qua Splunk.



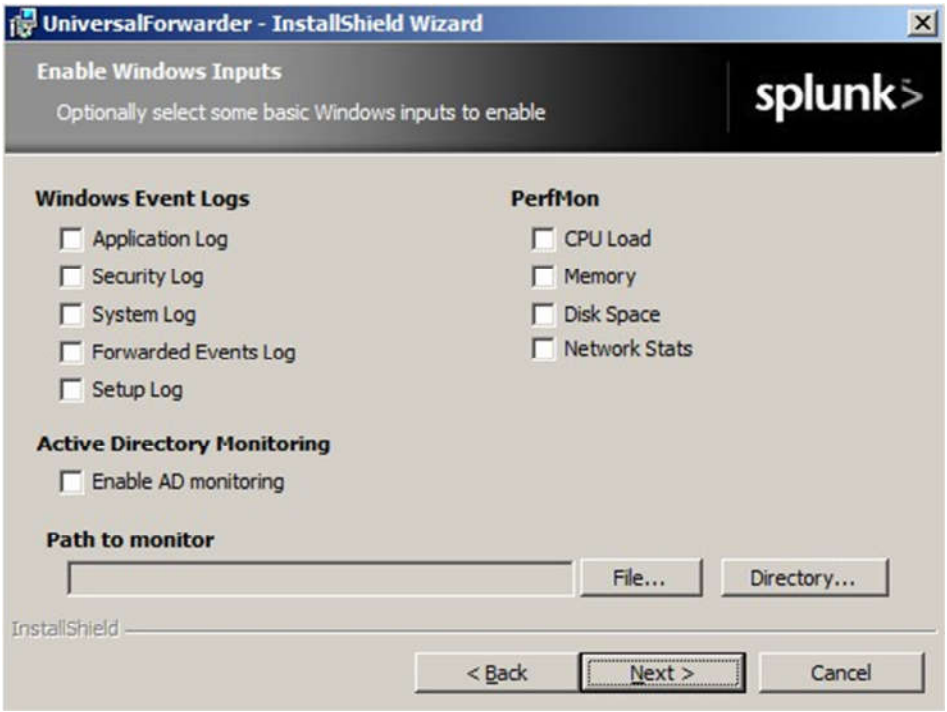
Hình 3.27 Giao diện nhập địa chỉ IP và cổng kết nối

Chọn mục Remote Windows Data để gửi thông tin các log, event, và performance của DomainController



Hình 3.28 Chọn Remote Windows Data

Chọn loại log mà ta cần giám sát, ta có thể tùy chỉnh lại ở file sau khi cài đặt.



Hình 3.29 Giao diện lựa chọn kiểu lấy log

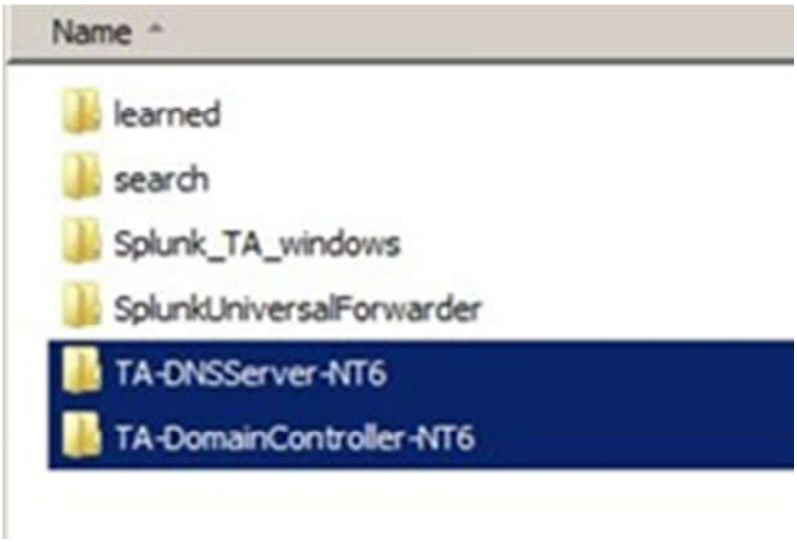
Ta chọn cài đặt Splunk Add-on for Windows.

Hình 3.30 Giao diện chọn Splunk Add-on for Windows

Chọn Finish để kết thúc quá trình cài đặt.

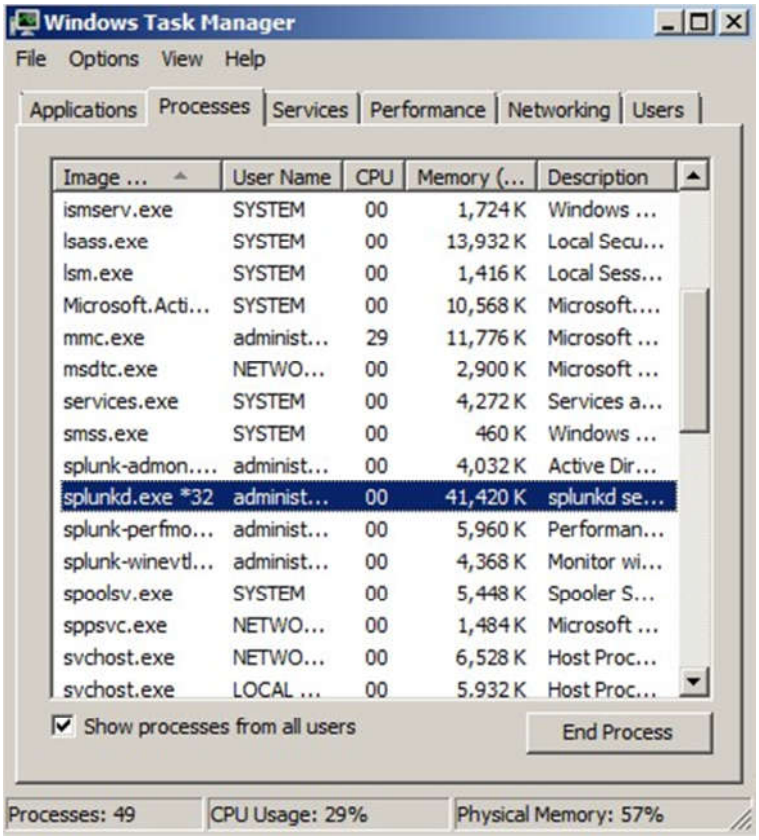
Hình 3.31 Giao diện lựa chọn kết thúc quá trình cài đặt

Copy 2 thư mục TA-DNSServer-NT6 và TA-DomainController-NT6 vào thư mục câu hình của Splunk để có thể gửi các thông tin tới. Sau đó ta restart server để splunk có thể hoạt động.



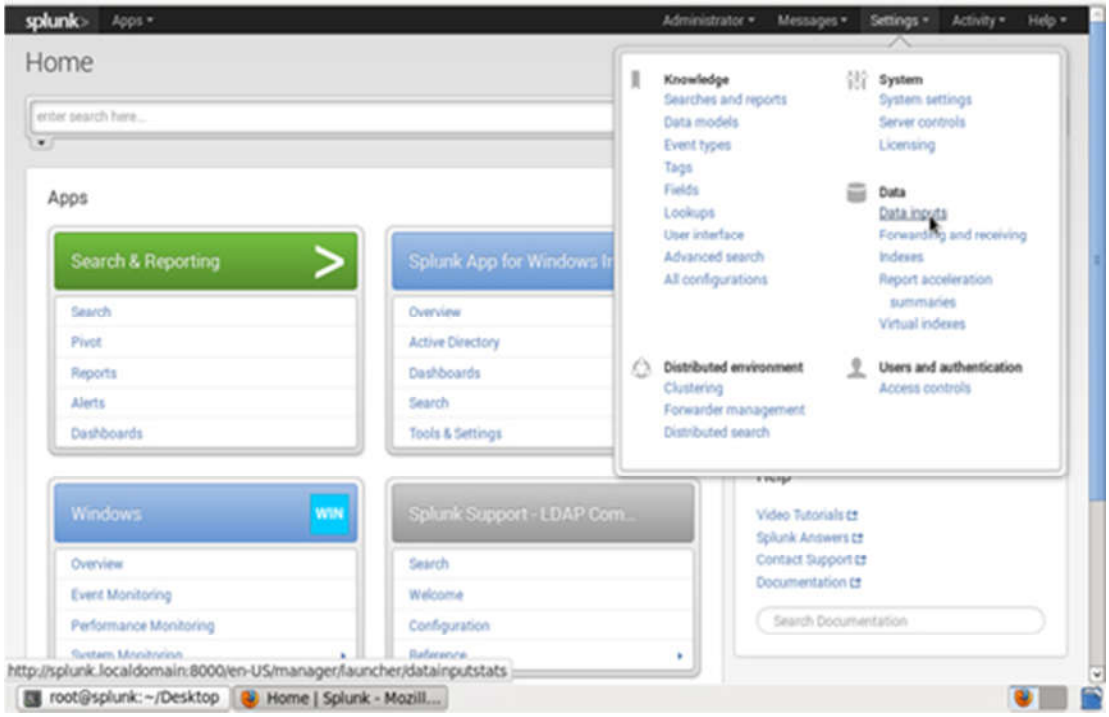
Hình 3.32 Giao diện lưu trữ 2 thư mục

Sau khi cài đặt xong, ta khởi động lại windows server và kiểm tra tiến trình đã thấy Splunk hoạt động



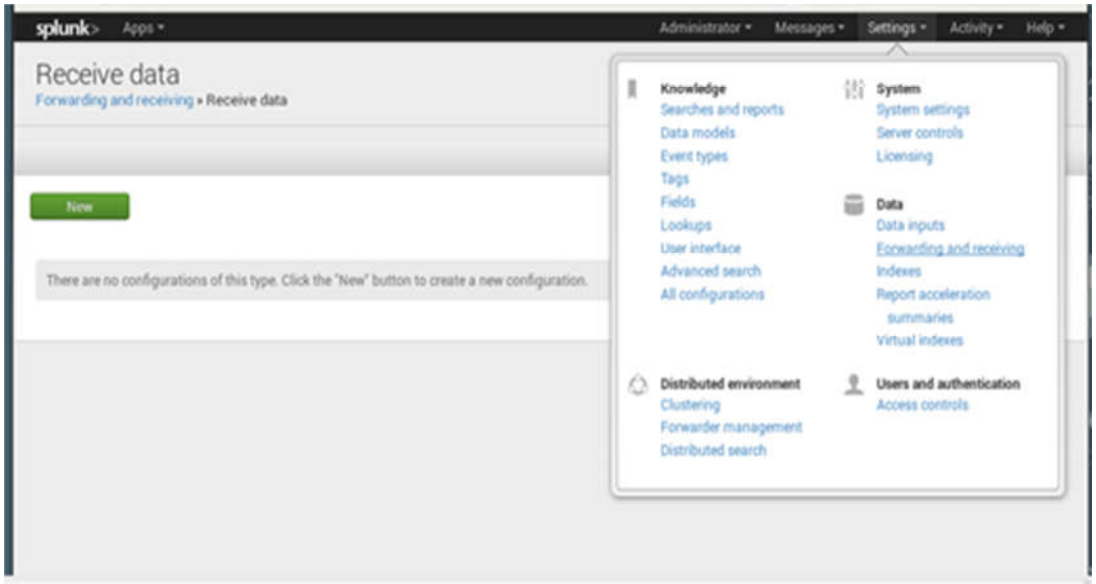
Hình 3.33 Splunk đã hoạt động trên Task Manager

Tại giao diện của Splunk,



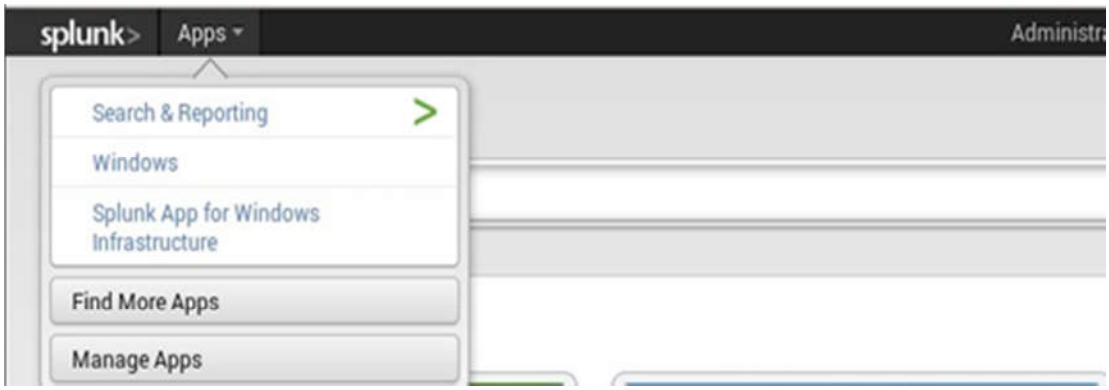
Hình 3.34 Giao diện cấu hình của Splunk

Vào Forwarding and Receiving add thêm port 10000 trùng với port lúc cài đặt ở Windows Server



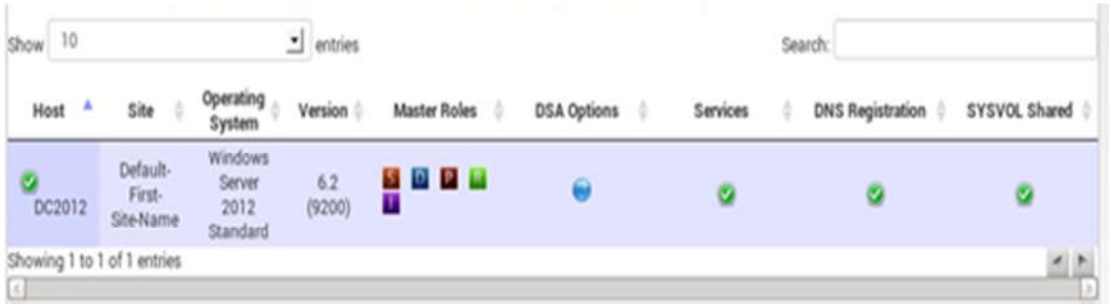
Hình 3.35 Cấu hình Receive data thành công

Sau khi sau khi gắn port ta, vào Search & Reporting kiểm tra dữ liệu đã được gửi qua cho Splunk



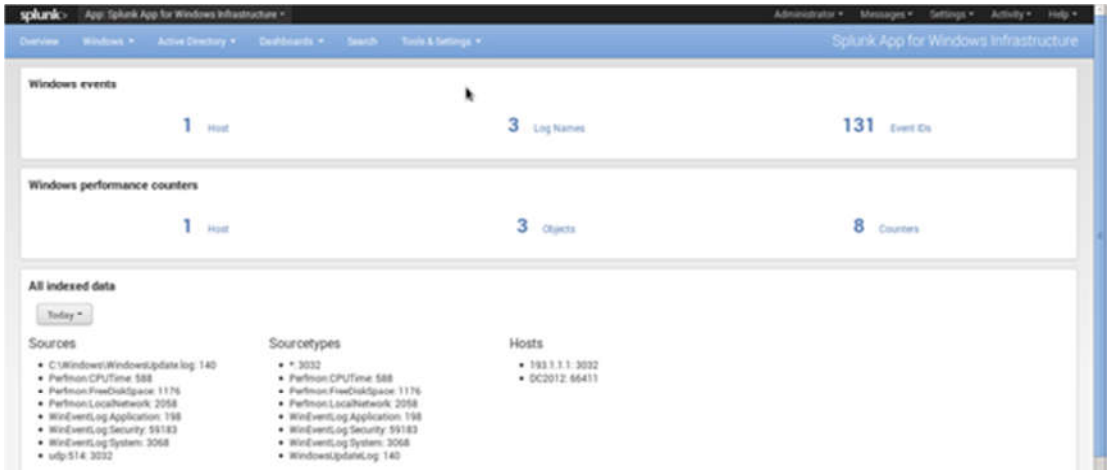
Hình 3.36 Giao diện tìm kiếm của Splunk

Kết quả trả ra sau khi tìm kiếm máy chủ Windows trên Splunk

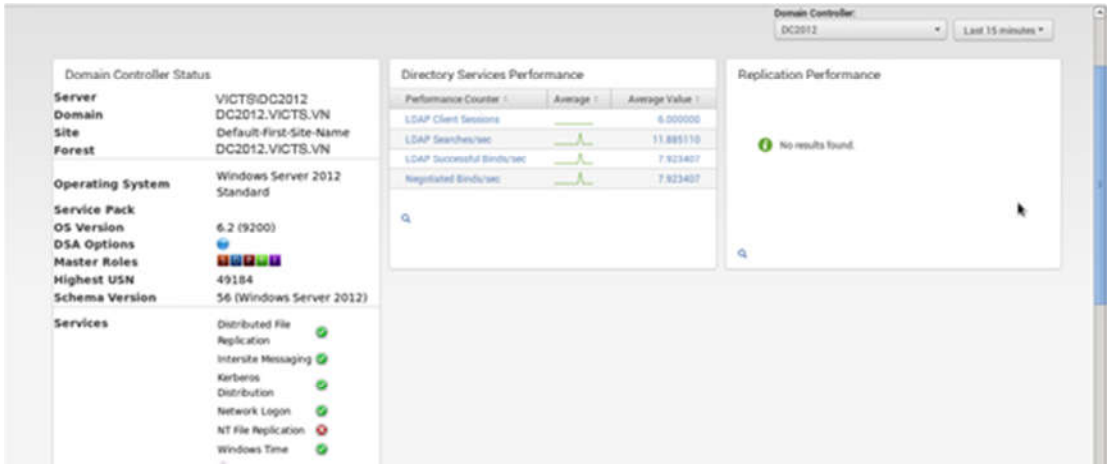


Hình 3.37. Giao diện hiển thị kết quả tìm kiếm thành công máy chủ Windows

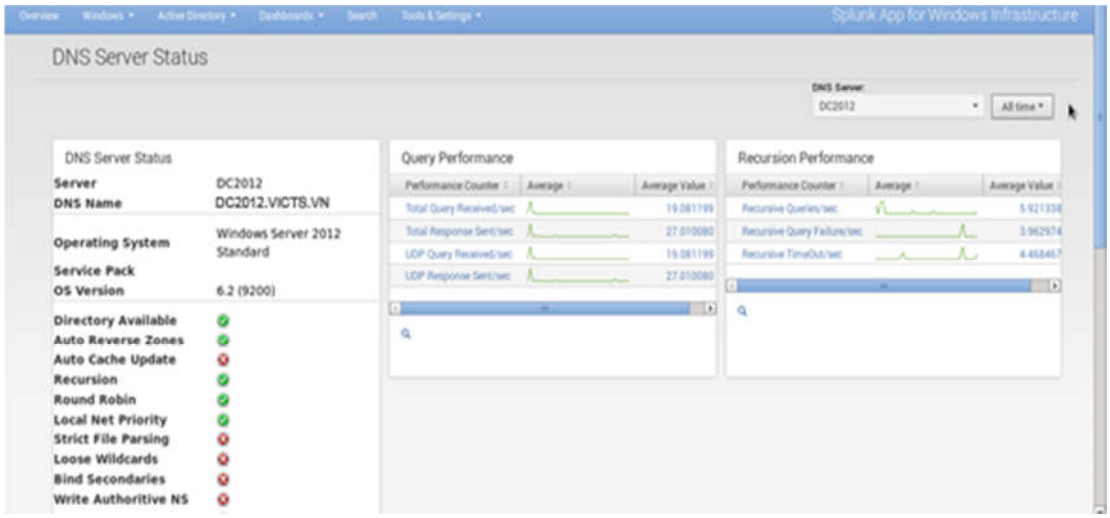
Hình 3.38 Giao diện hiển thị dịch vụ DNS chạy trên máy chủ Windows



Hình 3.39 Giao diện hiển thị log của máy chủ Windows trên Splunk



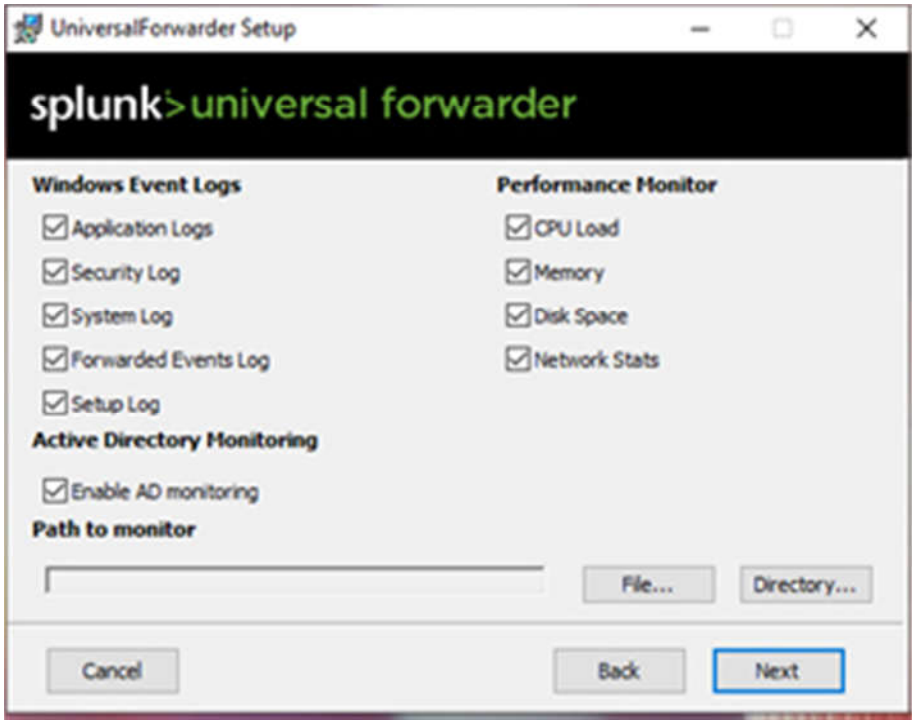
Hình 3.40 Giao diện hiển thị tổng quan các thông số của máy chủ Windows



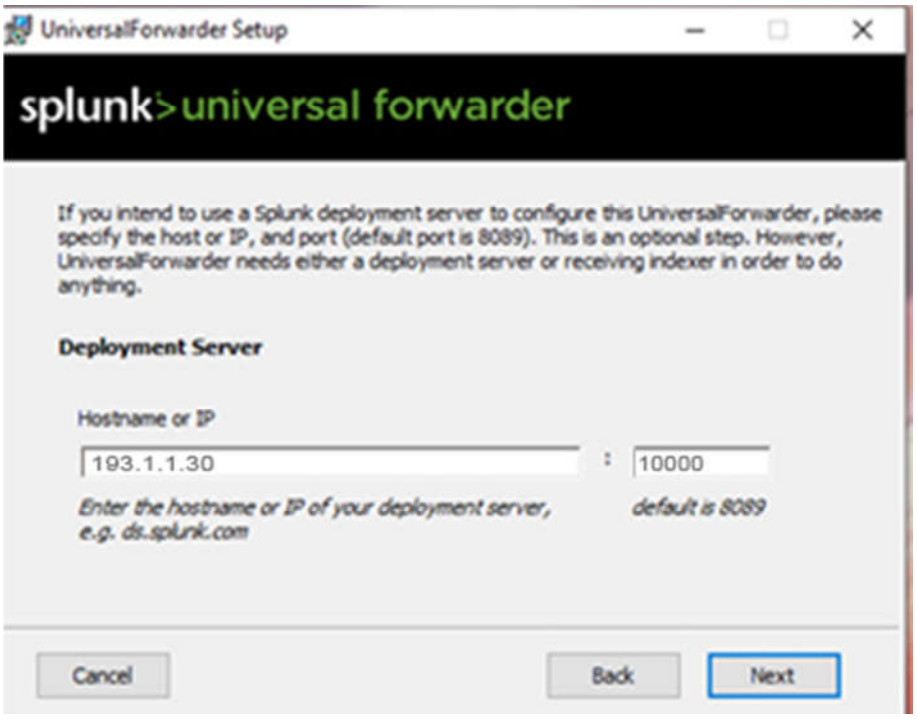
Hình 3.41 Giao diện tổng quan hiển thị dịch vụ DNS trên Splunk

3.3.5 Lấy Log từ máy chủ Windows 10 của người dùng về phân tích

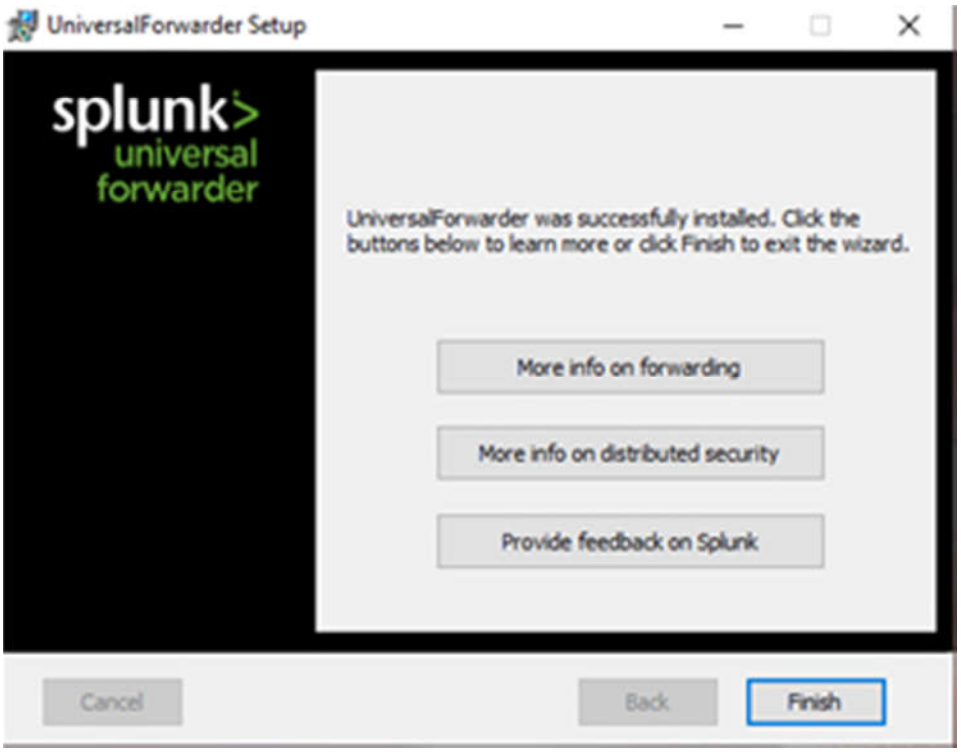
Cài đặt Splunk Forwarder trên Windows 10



Hình 3.42 Giao diện lựa chọn kiểu lấy log trên Windows 10



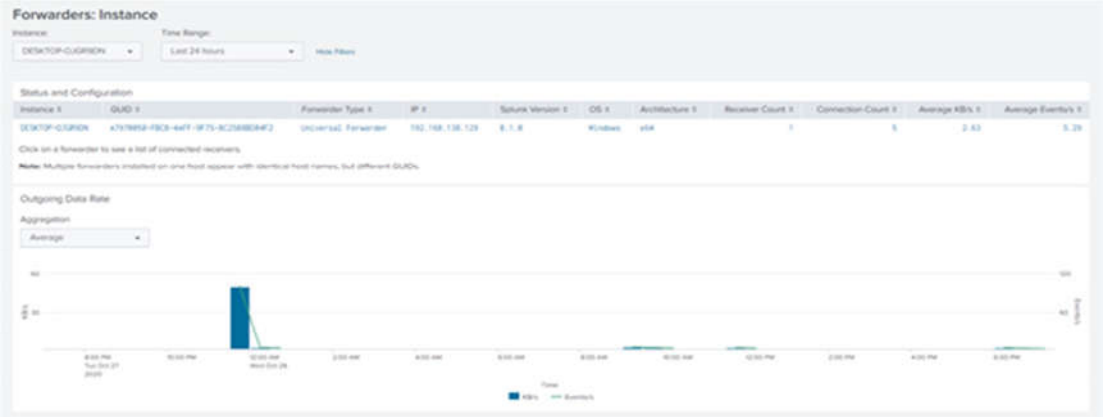
Hình 4.43 Giao diện nhập địa chỉ IP và cổng kết nối



Hình 3.44 Giao diện lựa chọn kết thúc quá trình cài đặt trên Windows 10

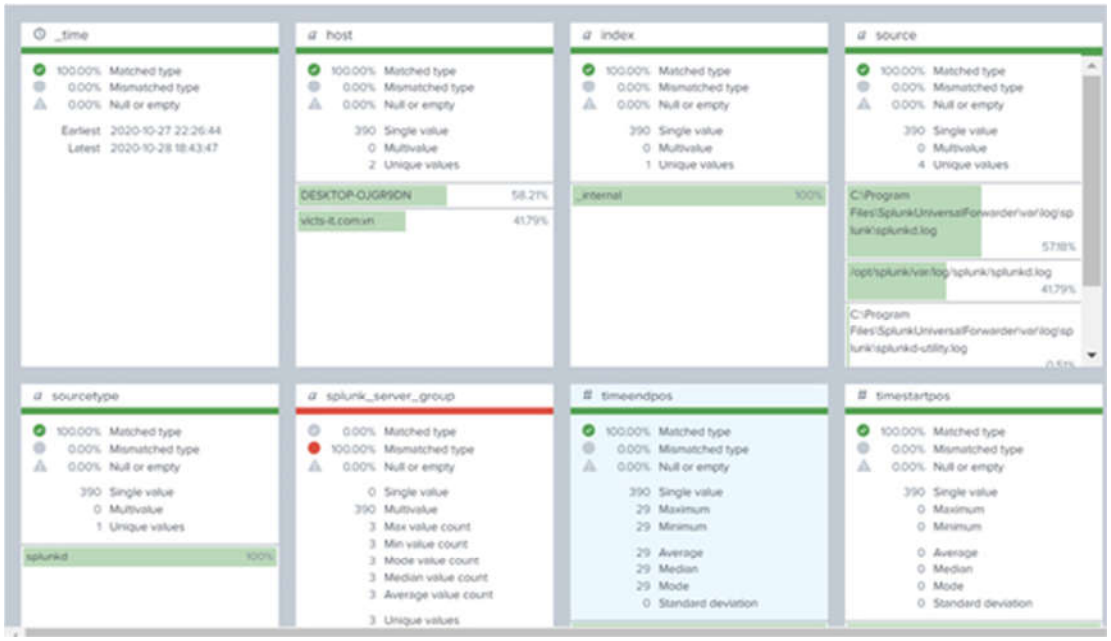
Kết quả:

Với hiển thị cụ thể cho các đăng nhập từ máy windows 10: host="desktop-ojgr9dn" "logname=system" source="WinEventLog:System" "sid=s-1-5"



Hình 3.45. Giao diện hiển thị thông tin từ forwarding

- Thông tin các error log



Hình 3.46 Giao diện hiển thị thông tin các error log

3.4 Thử nghiệm và đánh giá

3.4.1 Nội dung thử nghiệm

Luận văn thực hiện thử nghiệm một số nội dung sau đây

(1) Cài đặt máy chủ giám sát tập trung Splunk

Ta sẽ tiến hành cài đặt máy chủ Splunk, nơi sẽ thông tin đổ về từ các máy thu thập, lưu trữ và phân tích chúng. Ta cần phải đảm bảo đường truyền mạng ổn định, các kết nối tới các máy giám sát đều không gặp trục trặc.

(2) Lấy log từ máy chủ Linux cài đặt Splunk để phân tích

Ta sẽ tiến hành cấu hình máy chủ Splunk để lấy Log của máy chủ Linux đưa ra phân tích.

(3) Cài đặt giám sát lấy Log từ máy chủ Firewall Pfsense về phân tích

Ta sẽ cài đặt và cấu hình ở System Logs của Pfsense và mở cổng để lấy log từ máy chủ Firewall Pfsense về để phân tích, từ đó ta có thể theo dõi được các hoạt động vào ra của hệ thống mạng.

(4) Lấy Log từ máy chủ Windows Server

Ta sẽ tiến hành cài đặt Cài đặt Splunk Forwarder và cấu hình sử dụng công cụ Splunk để lấy Log từ máy chủ Windows Server vào phân tích.

(5) Lấy Log từ máy chủ Windows 10 của người dùng về phân tích

Trong phần này sẽ cài đặt và cấu hình để lấy Log từ Windows 10 của người dùng về để phân tích.

3.4.2 Kết quả thử nghiệm và đánh giá

Phần thử nghiệm sẽ đưa ra kết quả và trình bày trong phần phụ lục của Luận văn. Các kết quả thử nghiệm đều khả quan, vận hành tốt cũng như ổn định và đáp ứng được các nhu cầu giám sát an toàn thông tin.

Các giải pháp đã thử nghiệm có thể ứng dụng cho mạng nội bộ tại Viện Khoa học công nghệ sáng tạo Việt Nam.

3.5 Kết luận chương 3

Chương 3 của luận văn đã khảo sát mạng nội bộ tại Viện KHCN Sáng tạo Việt Nam, các vấn đề nảy sinh trong quá trình sử dụng và các yêu cầu trong giám sát hệ thống mạng nhằm đáp ứng nhu cầu đào tạo của Viện KHCN Sáng tạo Việt Nam.

Luận văn cũng đề xuất giải pháp giám sát an toàn thông tin cho hệ thống mạng của Viện KHCN Sáng tạo Việt Nam. Kết quả thử nghiệm phù hợp với các yêu cầu đề ra ban đầu.

KẾT LUẬN

Kết quả dự kiến đạt được của luận văn:

Với mục tiêu nghiên cứu, áp dụng giải pháp giám sát an toàn thông tin vào hệ thống CNTT của Viện KHCN Sáng tạo Việt Nam, luận văn đã đạt được một số kết quả sau đây:

- Tổng quan về giám sát an toàn thông tin
- Yêu cầu giám sát hệ thống an toàn thông tin
- Giải pháp giám sát an toàn thông tin Splunk SIEM
- Tăng cường bảo đảm ATTT cho hệ thống CNTT tại Viện KHCN Sáng tạo Việt Nam.

Hướng phát triển tiếp theo:

Học viên sẽ tiếp tục nghiên cứu, hoàn thiện, tối ưu giải pháp hơn nữa để có thể đảm bảo ATTT cho hệ thống CNTT của Viện KHCN Sáng tạo Việt Nam ở mức cao hơn như sử dụng công nghệ AI để phát hiện hành vi tấn công.

DANH MỤC TÀI LIỆU THAM KHẢO

Tài liệu trong nước

[1] Trần Công Cẩn (2019) – *Mô phỏng mạng máy tính Trường Đại học Khánh Hòa* – Trường Đại học Khánh Hòa

[2] Hoàng Xuân Dậu (2007) - *Bài giảng an toàn bảo mật hệ thống thông tin*. Học viện Công nghệ Bưu chính Viễn thông

[3] PGS.TSKH. Hoàng Đăng Hải (2018) - *Quản lý an toàn thông tin - Học viện Công nghệ Bưu chính Viễn thông.* Nhà xuất bản Khoa học Kỹ thuật

[4] Phương Minh Nam (2010) - *Nguy cơ mất an ninh, an toàn thông tin, dữ liệu và một số giải pháp khắc phục* – Bộ Công An

[5] Trần Văn Khá (2008) – *Firewall trong Linux bằng Iptables*

Tài liệu ngước ngoài

[6] Arne Mikalsen and Per Borgesen (2002) - *Local Area Network Management - Design n Security.* University College Norway

[7] Certified Ethical Hacker – Ec Council

[8] Certified Information System Security Professional – Microsoft

[9] Cisco Certified Network Associate – Cisco Academy

[10] David Miller et al (2010) Security Information and Event Management

[11] Eliud Ir. Eliud Aganze (2014) - *Design Implementation And Management Of* Secured LAN MSc. Jomokenyatta University of Agriculture And Technology.

[12] Gert De Laet, Gert Schauwers (2004) - *Network Security Fundamentals.* Publisher Cisco Press.

[13] Helling (2015) - *Home Network Security.* Eindhoven University of Techonogy.

[14] IETF RFC 1701: Generic Routing Encapsulation (GRE)

[15] IETF RFC 2637: Point - to - Point Tunneling Protocol (PPTP)

[16] IETF RFC 2661: Layer Two Tunnuling Protocol (L2TP)

[17] Jan Vykopal (2008) - *Security Analysis of a Computer Network.* Masaryk University Faculty of Informatics.

[18] Kaiyuan Yang (2011) - *Bachelor's Thesis.* Abstract Turku University of Applied Sciences.

[19] Kevin Wey Kaye Tham (2006) - *Dev Security Service For Network*

Architectures. PhD Quyeesland University of Technology.

[20] Overview of Virtual Private Networks and IPSec Technologies - Cisco System.

[21] R.C.Sreijl (2000) - Analysis of Managed Virtual Private Network

[22] Tamirat Atsemegiorgis (20130 - *Building a Secure Local Area Network*. Helsinki Metropolia University of Applied Sciences.

Tài liệu từ Internet:

[23] http://www.cisco.com/go/vpn Ngày 29/4/2020

[24] http://www.lpi.org/ Ngày 29/4/2020

[25] http://www.vjst.vn/vn/tin-tuc/2653/big-data-va-ung-dung-trong-bao-mat-thong-tin.aspx Ngày 29/4/2020

[26] https://ictnews.vietnamnet.vn/cntt/bao-mat/nua-dau-nam-2019-so-cuoc-tan-cong-mang-vao-cac-he-thong-thong-tin-viet-nam-tiep-tuc-giam-184932. Ngày 29/4/2020.

[27] https://securitydaily.net/splunk-cong-cu-toan-nang-cho-cac-chuyen-gia-giam-sat-an-ninh-mang/ Ngày 29/4/2020

[28] https://trendmicro.ctydtp.vn/10-vu-tan-cong-internet-lon-nhat-lich-su.html Ngày 29/4/2020.

[29] https://vnetwork.vn/vi/news/10-thong-ke-ve-an-ninh-mang-2019

[30] https://www.elastic.co/elk-stack/ Ngày 29/4/2020

[31] https://www.elastic.co/products Ngày 29/4/2020

[32] https://www.snort.org/ Ngày 29/4/2020