HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG KHOA AN TOÀN THÔNG TIN



BÁO CÁO BÀI TẬP LỚN

Kỹ thuật theo dõi và giám sát an toàn mạng Tìm hiểu về Splunk để phân tích và thống kê dữ liệu trong hệ thống mạng máy tính

Giảng viên hướng dẫn: Ninh Thị Thu Trang

Nhóm lớp: 04

Nhóm Bài tập lớn: 06

Họ và tên sinh viên: B20DCAT087 - Trần Trọng Huy

B20DCAT103 - Nguyễn Quốc Khánh

B20DCAT091 - Nguyễn Thái Hưng

HÀ NỘI, tháng 3 năm 2024

LÒI CẨM ƠN

Sau thời gian học tập và cố gắng tại lớp học phần **Kỹ thuật theo dõi và giám sát an toàn mạng**, bằng sự biết ơn và kính trọng, chúng em xin gửi lời cảm ơn chân thành đến **cô Ninh Thị Thu Trang** đã nhiệt tình hướng dẫn, giảng dạy và tạo mọi điều kiện thuận lợi giúp đỡ chúng em trong suốt quá trình học tập, nghiên cứu và hoàn thiện đề tài nghiên cứu "*Tìm hiểu về Splunk để phân tích và thống kê dữ liệu trong hệ thống mạng máy tính*".

Tuy nhiên điều kiện về năng lực bản thân còn hạn chế, bài tập nghiên cứu chắc chắn không tránh khỏi những thiếu sót. Kính mong nhận được sự đóng góp ý kiến của cô để bài nghiên cứu của nhóm chúng em được hoàn thiện hơn.

Em xin trân trọng cảm ơn!

Hà Nội, ngày 17 tháng 03 năm 2024

Nhóm 06

MỤC LỤC

LÒI	CÅM ON	2
MŲ(C LŲC	3
DAN	NH MỤC HÌNH ẢNH	4
I. 1	Γổng quan về NSM và phân tích dữ liệu trong NSM	5
1.1.	Tổng quan về NSM	5
1.2.	Tìm hiểu về phân tích dữ liệu trong NSM	
1.2	2.1. Phân tích gói tin	6
1.2	2.2. Mối đe dọa bảo mật và tài nguyên cần bảo vệ	
1.2	2.3. Quy trình phân tích	8
1.	2.4. Truy tìm các mối đe dọa	9
II.	Giới thiệu về Splunk	9
2.1 I	Định nghĩa và các ứng dụng phổ biến của Splunk	9
	1.1. Splunk là gì?	
2.	1.2. Úng dụng	10
2.2.	Kiến trúc hệ thống Splunk	11
III.	Thu thập dữ liệu với Splunk	12
3.1.	Nguyên tắc	12
3.2.	Loại dữ liệu phổ biến	13
3.3.	Cách Splunk xử lý dữ liệu từ nguồn khác nhau	13
IV.	Các tính năng và chức năng chính của Splunk	13
4.1.	Bảng điều khiển và báo cáo:	13
4.	1.1. Table	14
4.	1.2. Dashboard	14
4.2.	Thực hiện giám sát và cảnh báo	14
4.2	2.1. Giám sát	15
4.2	2.2. Cảnh báo	15
4.3.	Quản lý quyền truy cập và bảo mật	16
V.	Ngôn ngữ truy vấn Splunk (SPL)	18
VI.	Ứng dụng thực hành: Phân tích tấn công DoS với Splunk	20
6.1	Cấu hình rules	20

6.2.	Thử nghiệm (Demo tấn công DoS)	22
6.2.2.	Kịch bản	22
6.2.3.	Chuẩn bị	22
6.2.4.	Thử nghiệm	23
VII.	Nhận xét	24
7.1.	Ưu điểm	24
7.2.	Nhược điểm	
VIII.	Kết luận	25
IX. Tà	i liệu tham khảo	27
	DANH MỤC HÌNH ẢNH	
Hình 1. C	Chu trình NSM	6
Hình 2. (Gói in HTTP GET trong Wireshark	6
Hình 3. E	Oịnh dạng của tiêu dề gói tin TCP	7
Hình 4. C	Chu trình thu thập thông tin về các mỗi đe doạ NSM	8
Hình 5. k	Kiến trúc của Splunk	11
Hình 6. C	Các loại data, log mà Splunk index được	13
Hình 7. T	Table dạng số trong Splunk	14
Hình 8. D	Danh sách cảnh báo trong Splunk	16
Hình 9. S	plunk Search Language ví dụ	19
Hình 10.	The Search Pipeline	19
Hình 11.	Event được Splunk tạo ra khi phân tích log	20
Hình 12.	Splunk hỗ trợ cấu hình alert dựa trên event	21
Hình 13.	Mẫu cấu hình Alert cho tấn công DoS	22
Hình 14.	Sử dụng BurpSuite để tấn công DoS trang web	23
Hình 15.	Event hiển thị trong Splunk	23
Hình 16.	Request bất thường hiển thị trong Splunk	24
Hình 17.	Hệ thống hiển thị cảnh báo tấn công DoS	24

I. Tổng quan về NSM và phân tích dữ liệu trong NSM

1.1. Tổng quan về NSM

NSM (Network Security Monitoring) là phương pháp và hệ thống các kỹ thuật, công cụ được sử dụng để giám sát và bảo vệ mạng máy tính khỏi các mối đe dọa an ninh mạng. NSM tập trung vào việc thu thập, phân tích và phản ứng đối với các hoạt động không bình thường diễn ra trên mạng, từ đó cung cấp sự nhận biết sớm và bảo vệ hiệu quả đối với hệ thống của tổ chức.

Phần lớn các hệ thống NSM được dành riêng để phát hiện xâm nhập từ đó ứng phó sự cố tốt hơn, NSM được coi là mô hình mới cho lĩnh vực phát hiện xâm nhập và đã xây dựng được một tập các đặc tính khác biệt hoàn toàn so với phát hiện xâm nhập truyền thống như: Phòng chống đến cùng cho dù thất bại, Tập trung vào tập dữ liệu, Quy trình theo chu trình, Phòng thủ theo mối đe dọa.

Chu trình giám sát an toàn mạng của NSM bao gồm việc thu thập dữ liệu, phát hiện xâm nhập và phân tích dữ liệu an ninh mạng:

- Thu thập dữ liệu: NSM bắt đầu bằng việc thu thập dữ liệu từ các nguồn khác nhau trên mạng, bao gồm lưu lượng mạng, logs hệ thống, dữ liệu từ các thiết bị mạng như tường lửa, máy chủ, và các thiết bị giám sát khác sau đó được sắp xếp và lưu trữ dữ liệu cho việc phát hiện xâm nhập và phân tích dữ liệu trong hệ thống NSM.
- Phát hiện xâm nhập: Là quá trình mà qua đó dữ liệu thu thập được kiểm tra và cảnh báo sẽ được tạo ra dựa trên các sự kiện quan sát được và dữ liệu thu thập không được như mong đợi. Điều này thường được thực hiện thông qua một số hình thức chữ ký, sự bất thường, hoặc phát hiện dựa trên thống kê. Kết quả là tạo ra các dữ liệu cảnh báo.
- Phân tích dữ liệu: Dữ liệu thu thập được được phân tích để xác định các mẫu hành vi không bình thường hoặc có thể đe dọa. Phân tích này có thể bao gồm việc xác định các biểu hiện của malware, tấn công

mạng, các cuộc tấn công từ chối dịch vụ (DoS), hoặc các hành vi đăng nhập không hợp lệ.



Hình 1. Chu trình NSM.

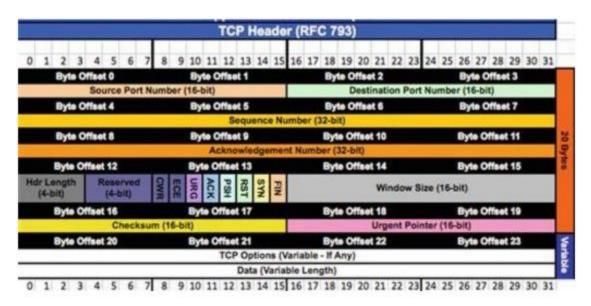
1.2. Tìm hiểu về phân tích dữ liệu trong NSM

1.2.1. Phân tích gói tin

Xâm nhập gói tin: là quá trình phân tích và kiểm tra các gói tin dữ liệu được truyền qua mạng từ một thiết bị đến thiết bị khác. Các gói tin là các đơn vị cơ bản nhất trong mạng, chúng được định dạng và truyền đi để thiết lập và duy trì kết nối giữa các máy tính và thiết bị mạng khác nhau. Xâm nhập gói tin giúp phát hiện và ngăn chặn các mối đe dọa mạng bằng cách kiểm tra nội dung của các gói tin, phân tích các giao thức mạng, và xác định các hoạt động bất thường hoặc độc hại trong mạng.

Hình 2. Gói tin HTTP GET trong Wireshark.

Phân tích chi tiết gói tin là quá trình nghiên cứu cẩn thận từng phần của gói tin mạng, tập trung vào các tiêu đề và dữ liệu của từng giao thức trong gói tin. Các bước cơ bản trong phân tích bao gồm xác định cấu trúc của tiêu đề IP, độ dài của tiêu đề TCP, và các tùy chọn được hỗ trợ trong tiêu đề TCP. Điều này giúp hiểu rõ hơn về cách mạng hoạt động và phát hiện các vấn đề bảo mật có thể tồn tại trong giao tiếp mạng.



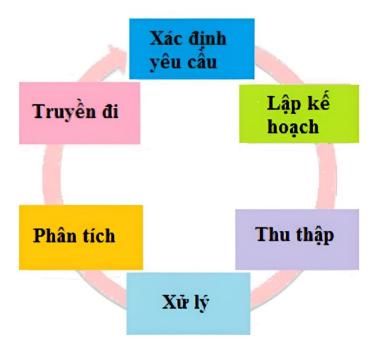
Hình 3. Định dạng của tiêu dề gói tin TCP.

1.2.2. Mối đe dọa bảo mật và tài nguyên cần bảo vệ

Mối đe dọa bảo mật và tài nguyên cần bảo vệ là hai khía cạnh quan trọng trong bảo mật mạng.

Thông tin về mối đe dọa bảo mật và tài nguyên cần được bảo vệ (Friendly and threat intelligence –TI): Bao gồm các thông tin như địa chỉ IP, URL, tên miền, địa chỉ email, tên file, đường dẫn, và thông tin liên quan đến máy chủ như tên file, hàm băm, khóa, thư viện. Những thông tin này giúp xác định mối đe dọa và hỗ trợ quyết định bảo mật.

Chu trình thu thập thông tin về mối đe dọa NSM: Thu thập thông tin từ hệ thống giám sát mạng (NSM), bao gồm phân tích dữ liệu gói tin, log, và các thông tin từ thiết bị mạng.



Hình 4. Chu trình thu thập thông tin về các mỗi đe doạ NSM.

Tạo thông tin về tài nguyên cần bảo vệ: Xác định và mô tả các tài nguyên quan trọng cần được bảo vệ trong hạ tầng mạng, như máy chủ, dịch vụ, và dữ liệu quan trọng.

Tạo thông tin về mối đe dọa bảo mật: Đánh giá và mô tả các mối đe dọa cụ thể đối với hệ thống mạng, cung cấp thông tin chi tiết về các loại mối đe dọa, phương pháp tấn công và tác động tiềm ẩn.

1.2.3. Quy trình phân tích

Các phương pháp phân tích

- Điều tra quan hệ: Phương pháp này thích hợp cho các tình huống phức tạp và có nhiều máy tính tham gia.
- Chẩn đoán khác biệt: Phương pháp này thường hiệu quả trong các tình huống ít máy tính liên quan và có thể gắn với một vài dấu hiệu khác biệt.

Các quy chuẩn thực tiễn cho phân tích

- Luôn đặt ra các giả định.
- Cần phải lưu ý về dữ liệu.
- Nên làm việc theo nhóm.
- Không bao giờ đánh động tin tặc.
- Gói tin vốn dĩ là vô hại.
- Wireshark chỉ là một công cụ phân tích.
- Cần thực hiện phân loại sự kiện rõ ràng.
- Tuân thủ quy tắc 10.

1.2.4. Truy tìm các mối đe dọa

Truy tìm mối đe dọa là quá trình chủ động tìm kiếm các mối đe dọa mạng tiềm ẩn dựa trên thông tin có sẵn. Nó giúp tối ưu hóa chi phí bảo mật, phát hiện sớm các hành vi bất thường và lỗ hổng trong hệ thống. Bằng cách này, nó giúp kiểm soát thiệt hại hiệu quả hơn và giảm thời gian dừng của hệ thống.

Các phương pháp truy tìm bao gồm xác định các đường cơ sở cho lưu lượng mạng, phân tích các mẫu luồng dữ liệu, và tận dụng sự hiểu biết về hoạt động bình thường của người dùng. Đồng thời, việc từ chối tường lửa ngoại vi, theo dõi các liên lạc ra ngoài, và phát hiện giao tiếp với các thiết bị bất thường cũng đóng vai trò quan trọng trong quá trình này. Ánh xạ khung làm việc MITRE ATT&CK cũng là một công cụ hữu ích để xác định và đối phó với các mối đe dọa..

II. Giới thiệu về Splunk

2.1 Định nghĩa và các ứng dụng phổ biến của Splunk

2.1.1. Splunk là gì?

Splunk là hệ thống có thể captures, trích ra các dữ liệu thời gian thực có liên quan tới nhau từ đó nó có thể tạo ra các đồ thị, các báo cáo, các cảnh báo và các biểu đồ.

Mục đích của Splunk là giúp cho việc xác định mô hình dữ liệu và thu thập dữ liệu máy trên toàn hệ thống dễ dàng hơn.Nó cung cấp số liệu, chẩn đoán các vấn đề xảy ra , phục vụ tốt cho hoạt động kinh doanh

Splunk có thể tìm kiếm các sự kiện đã và đang xảy ra, đồng thời cũng có thể báo cáo và phân tích thống kê các kết quả tìm được. Nó có thể nhập các dữ liệu của máy dưới dạng có cấu trúc hoặc không cấu trúc. Hoạt động tìm kiếm và phân tích sử dụng SPL(Search Processing Language), được tạo để quản lý Big Data. Do được phát triển từ Unix Piping và SQL nên Splunk có khả năng tìm kiếm dữ liệu, lọc, sửa đổi, chèn và xóa dữ liệu.

2.1.2. Úng dụng

Phân tích mối đe dọa an ninh: Splunk được sử dụng để phân tích mối đe dọa an ninh trong môi trường mạng. Với khả năng thu thập và chỉ mục dữ liệu mạng từ nhiều nguồn khác nhau mà không cần quan tâm đến định dạng hay kích thước, Splunk cho phép các nhà phân tích dễ dàng tìm kiếm, phân tích và giám sát các mẫu hoạt động bất thường, từ đó giúp phát hiện và phản ứng với các mối đe dọa an ninh một cách hiệu quả.

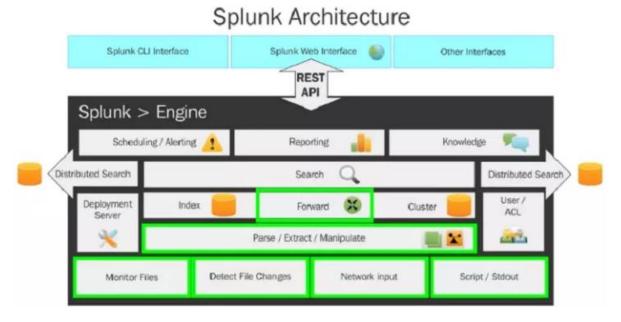
Quản lý log: Splunk được sử dụng để cải thiện quá trình quản lý và phân tích dữ liệu log. Bằng cách tự động index hóa dữ liệu log từ nhiều nguồn khác nhau và cho phép tìm kiếm, báo cáo và chẩn đoán dễ dàng, Splunk giúp tối ưu hóa việc quản lý log và phát hiện sớm các vấn đề hoặc mối đe dọa an ninh trong hệ thống.

Phát triển: Cung cấp các API hỗ trợ việc tạo các ứng dụng trên Splunk của người dùng. Một số bộ API điển hình như Splunk SDK (cung cấp các SDK trên nền tảng Python, Java, JS, PHP), Shep (Splunk Hadoop Intergration - đây là sự kết hợp giữa Splunk và Hadoop), Shuttl (là một sản phẩm hỗ trợ việc sao lưu dữ liệu trong Splunk), Splunkgit (Giúp bạn hình dung dữ liệu tốt hơn),

Splunk power shell resource Kit (Bộ công cụ hỗ trợ việc mở rộng và quản lý hệ thống).

Khắc phục sự cố: Splunk còn cung cập một cơ chế tự động khắc phục với các vấn đề xảy ra bằng việc tự động chạy các file Script mà người dùng tự tạo (Ví dụ như: Chặn IP, đòng Port ...) khi có các cảnh báo xảy ra.

2.2. Kiến trúc hệ thống Splunk



Hình 5. Kiến trúc 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.

Mỗi đường ống đơn bao gồm một loạt các bộ vi xử lý, dẫn đến xử lý dữ liệu nhanh hơn và hiệu quả hơn. 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.

III. Thu thập dữ liệu với Splunk

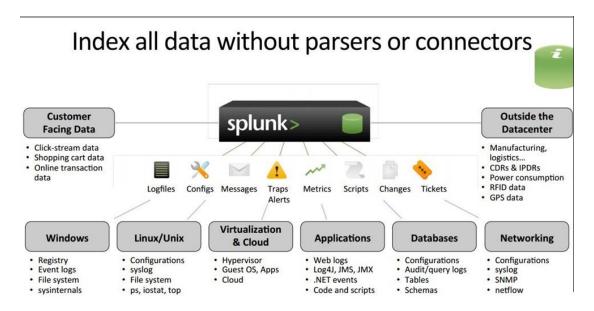
3.1. Nguyên tắc

Splunk thu thập dữ liệu từ nhiều nguồn khác nhau, bao gồm log hệ thống, log ứng dụng, dữ liệu hệ thống do máy móc tạo ra.

Dữ liệu hệ thống bao gồm nhiều hạng mục record của tất cả các hoạt động và hành vi của khách hàng, giao dịch của user, hành vi của hệ thống.

3.2. Loại dữ liệu phổ biến

Hỗ trợ hầu như tất cả các loại log của hệ thống, thiết bị hạ tầng mạng, phần mềm, Firewall, IDS/IPS, Log Event, Register của các máy trạm



Hình 6. Các loại data, log mà Splunk index được.

3.3. Cách Splunk xử lý dữ liệu từ nguồn khác nhau

Splunk có khả năng thu thập và lập chỉ mục dữ liệu từ một loạt các nguồn, bao gồm hệ thống máy chủ, hệ thống mạng và các ứng dụng. Điều này được thực hiện thông qua việc triển khai các agent hoặc forwarder Splunk trên các thiết bị để gửi dữ liệu về Splunk Enterprise để xử lý.

IV. Các tính năng và chức năng chính của Splunk

4.1. Bảng điều khiển và báo cáo:

Splunk cho phép người dùng tạo các bảng điều khiển và báo cáo tùy chỉnh để hiển thị thông tin theo cách tốt nhất cho nhu cầu của họ.

Các bảng điều khiển và báo cáo có thể được tạo ra từ các truy vấn dữ liệu và có thể được tùy chỉnh với nhiều biểu đồ, biểu đồ và dạng báo cáo khác nhau.

4.1.1. Table

Hàm pipe (|) trong splunk dùng để đưa kết quả output của 1 tiến trình thành input cho 1 tiến trình khác.

Một số hàm để tạo fields :eval, rex

Hàm lọc event: head, where

Hàm thay thế event với report : top, stats

logger \$	count \$	percent \$
BarClass	242	63.185379
FooClass	49	12.793734
AuthClass	47	12.271540
LogoutClas	s 45	11.749347

Hình 7. Table dạng số trong Splunk.

4.1.2. Dashboard

Dashboard là công cụ giúp chúng ta nắm bắt, nhóm và tùy chỉnh các bảng biểu đồ một cách hiệu quả. Nó chứa nhiều bảng thông tin, mỗi bảng chạy một truy vấn khác nhau. Mỗi dashboard có 1 link URL riêng biệt, dễ dàng trong việc chia sẻ. Dashboard có thể tùy biến, tùy chỉnh hiển thị các giá trị cần thiết, thanh tìm kiếm trong Dashboard được loại bỏ.

4.2. Thực hiện giám sát và cảnh báo

Splunk cho phép người dùng thực hiện giám sát liên tục trên hệ thống và ứng dụng bằng cách theo dõi các chỉ số quan trọng và sự kiện.

Người dùng có thể thiết lập các cảnh báo để được thông báo khi các điều kiện nhất định được đáp ứng, giúp họ phát hiện sớm và ứng phó với các vấn đề.

4.2.1. Giám sát

Máy chủ: 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: Tương quan log, số liệu hiệu suất và các sự kiện từ hệ thống lưu trữ của chúng ta 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 của chúng ta. Giảm thơi 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: Với Splunk, chúng 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,v.v. 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 của chúng ta an toàn và hoạt động mọi lúc. Đạt được chỉ số ROI tối đa bằng cách tối ưu hóa dung lượng mạng lưới của chúng ta, xác định độ trễ, quản lý bang thông, xác định top 10 tài nguyên mạng thường được sử dụng và mô hình sử dụng.

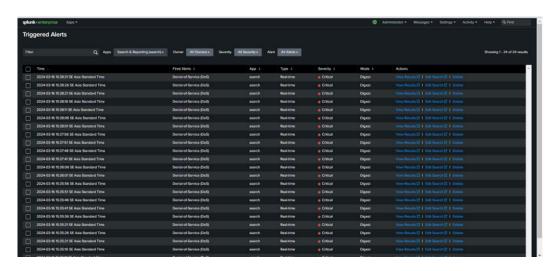
4.2.2. Cảnh báo

Splunk tự động list ra thời gian cụ thể của từng sự kiện xảy trong hệ thống mà nó đang giám sát.

Cảnh báo trong thời gian thực. Ta có thể chỉnh tùy chọn, định nghĩa các loại cảnh báo và có thể chỉ định ai nhận được cảnh báo đó.

Chủ động cảnh báo khi thiếu hụt dung lượng xảy ra. Lấy lại không gian lưu trữ không sử dụng để cho người dùng có trải nghiệm tối ưu. Sử dụng xu hướng theo thời gian và tối ưu hóa dựa trên tiêu thụ. Dự báo thông tin CPU, bộ nhớ, nhu cầu ổ cứng cần thiết và hiệu năng của từng máy chủ, máy ảo VMs thông qua lịch sử sử dụng tài nguyên.

Cấu hình cảnh báo dựa trên kịch bản có sẵn cho các vấn đề thường gặp như bộ nhớ, CPU, dung lượng ổ đĩa thấp.



Hình 8. Danh sách cảnh báo trong Splunk.

4.3. Quản lý quyền truy cập và bảo mật

Splunk cung cấp các tính năng quản lý quyền truy cập và bảo mật mạnh mẽ, giúp kiểm soát việc truy cập vào dữ liệu và chức năng của hệ thống.

Người quản trị có thể thiết lập các vai trò và phân quyền cho người dùng và nhóm người dùng, đồng thời áp dụng các biện pháp bảo mật như mã hóa và xác thực hai yếu tố để bảo vệ dữ liệu.

Kiểm soát truy cập dựa trên vai trò (RBAC) cung cấp các công cụ linh hoạt và hiệu quả mà bạn có thể sử dụng để bảo vệ dữ liệu trên nền tảng Splunk.

Nền tảng Splunk che giấu dữ liệu cho người dùng giống như cách cơ sở dữ liệu quan hệ quản lý RBAC. Trong một số trường hợp, tổng phân đoạn dữ liệu có thể là cần thiết.

Nguyên tắc cơ bản của việc định cấu hình quyền truy cập:

- Sử dụng vai trò để xác định quyền và khả năng của người dùng.
- Chỉ định các vai trò cho người dùng để kiểm soát phạm vi tác vụ, dữ liệu tìm kiếm và tài nguyên hệ thống.
- Người dùng có thể giữ nhiều vai trò, mỗi vai trò cung cấp quyền truy cập cụ thể vào tài nguyên hoặc chức năng.

Các vai trò được xác định trước:

Splunk đi kèm với các vai trò được xác định trước như admin,
 power, user, và nhiều vai trò khác để phù hợp với nhu cầu quản lý và sử dụng.

Đặt mức độ chi tiết về quyền với các vai trò tùy chỉnh:

• Có thể tạo và quản lý các vai trò tùy chỉnh, cho phépthực hiện điều chỉnh chi tiết về quyền truy cập của người dùng, bao gồm kế thừa vai trò, khả năng, chỉ mục được phép, và han chế tìm kiếm.

Kế thừa và kết hợp vai trò:

- Người dùng có thể kế thừa quyền và khả năng từ nhiều vai trò,
 với vai trò có quyền cao hơn thay thế các quyền của vai trò có quyền thấp hơn.
- Kết hợp các vai trò giúp người dùng nhận được quyền truy cập và khả năng mở rộng nhất có thể từ các vai trò khác nhau.

Giới hạn tìm kiếm và bộ lọc:

- Ngoài việc chỉ định quyền truy cập vào chỉ mục, Splunk có thể giới hạn kết quả tìm kiếm của người dùng thông qua việc áp dụng bộ lọc tìm kiếm.
- Bộ lọc tìm kiếm giúp xác định tập dữ liệu mà người dùng cuối cùng sẽ nhìn thấy từ kết quả tìm kiếm.

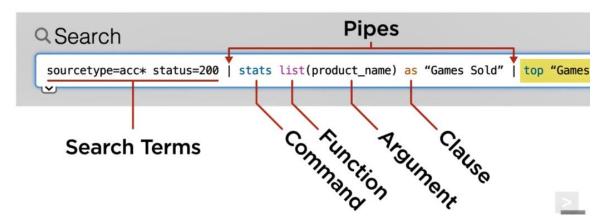
V. Ngôn ngữ truy vấn Splunk (SPL)

SPL là ngôn ngữ truy vấn của Splunk, cho phép người dùng tương tác với dữ liệu. SPL cung cấp cú pháp linh hoạt và mạnh mẽ để thực hiện các truy vấn, phân tích, và trích xuất thông tin từ dữ liệu.

Ngôn ngữ tìm kiếm được xây dựng từ năm thành phần:

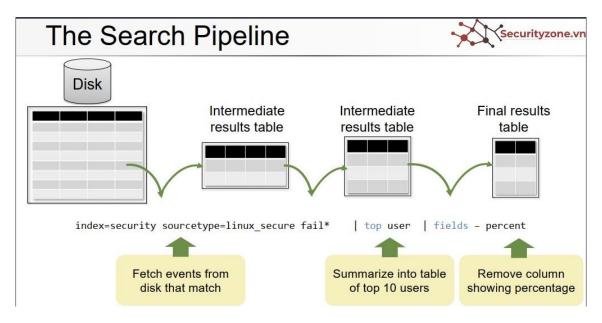
- The Search Terms(cụm từ tìm kiếm): các từ khóa để tìm kiếm kết quả mong muốn.Ví dụ như: keywords, phrases(cụm từ), các toán từ kết hợp.
- Commands: Các lệnh cho Splunk biết sẽ làm gì với kết quả tìm kiếm, ví dụ như:, thống kê tính toán, tạo biểu đồ.
- Functions: là các hàm được splunk hỗ trợ để thực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

Splunk Search Language Example



Hình 9. Splunk Search Language ví du.

Khi câu lệnh tìm kiếm Splunk thực hiện, Splunk sẽ đọc kết quả từ đĩa và tạo một bản sao trong bộ nhớ, xóa những sự kiện trong kết quả tìm kiếm mà không khớp với câu lệnh tìm kiếm. Kết quả cuối cùng giống như một bảng tính, chỉ chứa các trường và giá trị bạn đã tìm kiếm. Cùng với việc sử dụng bộ lọc thời gian thời gian cho các sự kiện, với các cụm từ tìm kiếm tối ưu, sẽ làm cho bảng tính nhỏ hơn và câu lệnh tìm kiếm nhanh hơn.



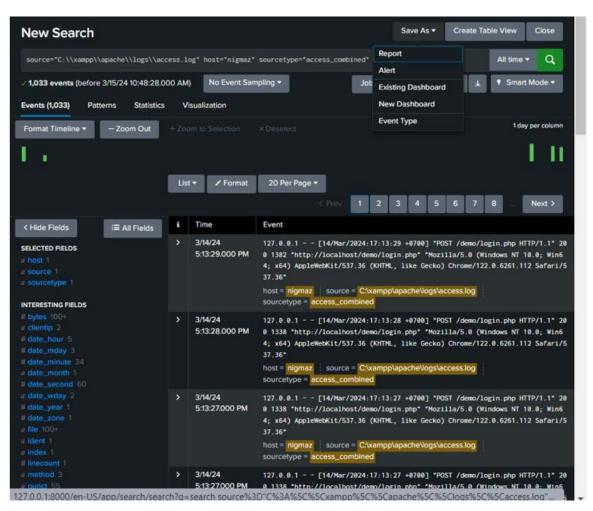
Hình 10. The Search Pipeline.

Cú pháp truy vấn cơ bản Splunk như Fields command, Table command, Rename command, Sort Command, Deup command và khi nhập các câu lệnh tìm kiếm, splunk sẽ cung cấp các tính năng tự động hoàn thành, các thành phần trong câu lệnh tìm kiếm có màu tự động.

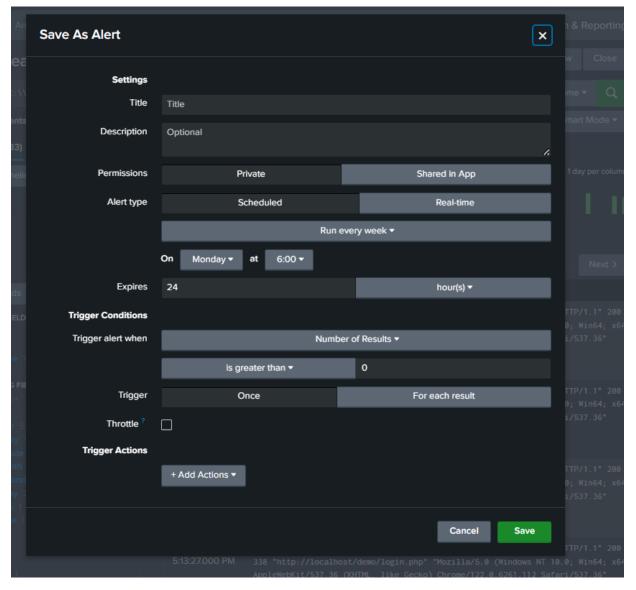
VI. Úng dụng thực hành: Phân tích tấn công DoS với Splunk.

6.1. Cấu hình rules

Để có thể phát hiện được tác động hoặc tấn công đến trang web, Splunk cần viết và cấu hình các rules để có thể alert ra thông báo nếu bị tác động hoặc tấn công DoS.

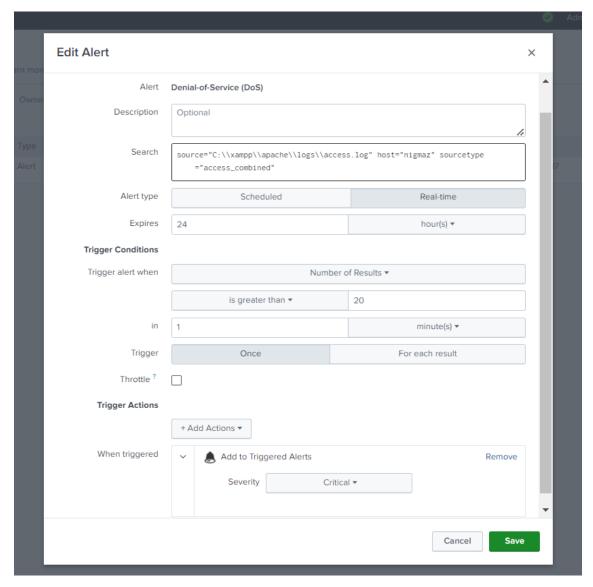


Hình 11. Event được Splunk tạo ra khi phân tích log.



Hình 12. Splunk hỗ trợ cấu hình alert dựa trên event.

- Thêm rule lên alert (Settings/Searchs, reports, and alerts/New Alert) và cấu hình các mục:
 - Tittle: Tên alert
 - Permisstions: Quyền của alert
 - Alert type: Real-time Chạy theo thời gian thực, khi tìm thấy event thì sẽ alert.
 - Trigger alert: Nhiều hơn 50 events/1 phút thì sẽ lên alert.



Hình 13. Mẫu cấu hình Alert cho tấn công DoS.

6.2. Thử nghiệm (Demo tấn công DoS)

6.2.2. Kịch bản

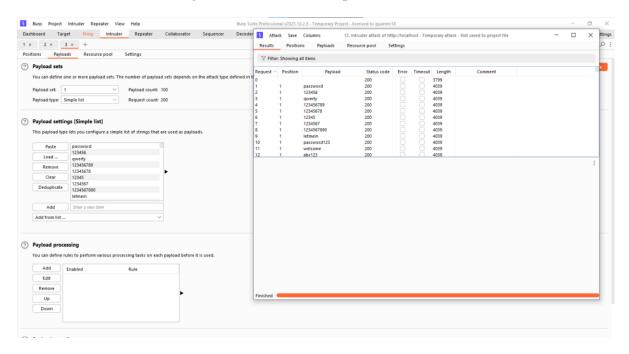
- Tạo trang đăng nhập, thực hiện tấn công Brute Force.
- Việc tấn công Brute Force với lượng request lớn sẽ khiến trang web bị DoS.

6.2.3. Chuẩn bị

- Môi trường: apache2, MySQL
- Công cụ tấn công: Burp Suite
- Công cụ giám sát: Splunk

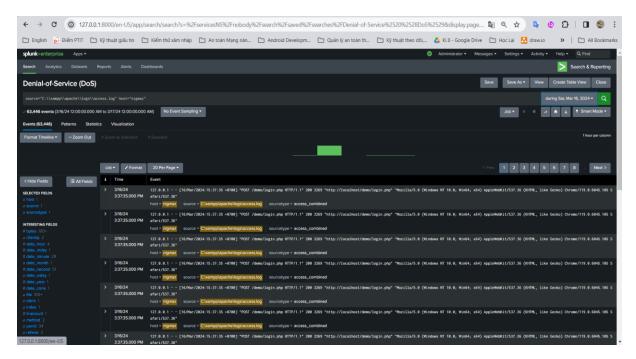
6.2.4. Thử nghiệm

- Thực hiện tấn công Brute Force với Burp Suite.



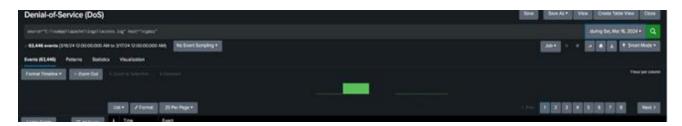
Hình 14. Sử dụng BurpSuite để tấn công DoS trang web.

- Tất cả các event được hiển thị trên Splunk.



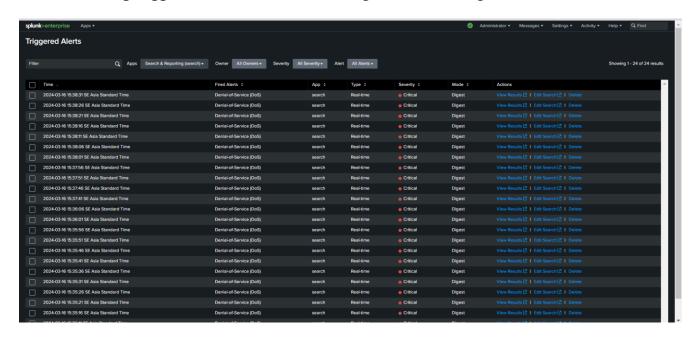
Hình 15. Event hiển thị trong Splunk.

- Biểu đồ lượng request nhiều bất thường trong một khoảng thời gian.



Hình 16. Request bất thường hiển thị trong Splunk.

- Hệ thống trigger alert đưa ra cảnh báo trang web bị tấn công DoS.



Hình 17. Hệ thống hiển thị cảnh báo tấn công DoS.

VII. Nhận xét

7.1. Ưu điểm

- Khả năng xử lý dữ liệu lớn: Splunk có khả năng xử lý và phân tích các lượng dữ liệu lớn từ nhiều nguồn khác nhau, bao gồm dữ liệu cấp độ doanh nghiệp.
- Tích hợp dễ dàng: Nó có khả năng tích hợp với nhiều nguồn dữ liệu khác nhau, bao gồm logs, metric, và dữ liệu từ các ứng dụng khác.

- Tìm kiếm và truy vấn linh hoạt: Splunk cung cấp một ngôn ngữ truy vấn mạnh mẽ (SPL) cho phép người dùng tìm kiếm, lọc, và phân tích dữ liệu một cách linh hoạt.
- Trực quan hóa dữ liệu: Nó cung cấp các công cụ trực quan hóa mạnh mẽ giúp người dùng hiểu và phân tích dữ liệu một cách dễ dàng thông qua biểu đồ, bảng, và các loại hình trực quan hóa khác.
- Bảo mật và phân quyền: Splunk cung cấp các tính năng bảo mật mạnh mẽ, bao gồm kiểm soát truy cập và phân quyền dựa trên vai trò người dùng.
- Thời gian phản hồi nhanh: Nhờ vào khả năng xử lý dữ liệu nhanh chóng, Splunk có thể cung cấp thời gian phản hồi nhanh cho các truy vấn và yêu cầu.

7.2. Nhược điểm

- Chi phí cao: Splunk là một giải pháp phần mềm có giá khá đắt, đặc biệt đối với các tổ chức lớn với nhu cầu lưu trữ và xử lý dữ liệu lớn.
- Học phức tạp: SPL, ngôn ngữ truy vấn của Splunk, có thể khó hiểu và học đối với người dùng mới.
- Yêu cầu cấu hình phức tạp: Đôi khi, việc cấu hình và triển khai Splunk có thể phức tạp, đặc biệt là trong môi trường doanh nghiệp lớn.
- Quản lý dữ liệu: Splunk có thể tạo ra các tệp nhật ký lớn và yêu cầu quản lý dữ liệu kỹ lưỡng để đảm bảo hiệu suất và khả năng mở rộng.

VIII. Kết luận

Như vậy, qua phần tìm hiểu chúng ta đã có cơ hội hiểu rõ hơn về Splunk và khả năng của nó trong việc phân tích và thống kê dữ liệu trong hệ thống mạng máy tính. Chúng ta đã làm quen với định nghĩa, khái niệm và tính chất

của công việc thu thập và phân tích dữ liệu, cũng như sự linh hoạt và tiện ích của việc sử dụng và cá nhân hoá các công cụ phân tích dữ liệu như Splunk.

Splunk không chỉ giúp chúng ta dễ dàng tiếp cận việc theo dõi và giám sát hệ thống một cách hiệu quả với các chế độ và tùy chọn cấu hình đơn giản, mà còn cho phép chúng ta tìm hiểu và thích nghi với các giải pháp cụ thể được cung cấp bởi Splunk để tối ưu hoá chi phí sử dụng.

Tuy nhiên Splunk cũng có những hạn chế nhất định. Điều này nhấn mạnh sự cần thiết từ kiến thức và nỗ lực liên tục từ các chuyên viên an ninh mạng khi sử dụng Splunk nhằm tạo ra một môi trường an toàn và tiện ích cho người sử dụng.

Về đề tài nghiên cứu "Tìm hiểu về Splunk để phân tích và thống kê dữ liệu trong hệ thống mạng máy tính", nhóm đã hoàn thành các phần Cơ sở lý thuyết và Thực hiện thử nghiệm một cách toàn diện. Dù vậy, để nghiên cứu được hoàn thiện hơn, chúng tôi rất mong nhận được ý kiến đóng góp từ cô để cải thiện và phát triển thêm bài nghiên cứu của chúng tôi.

IX. Tài liệu tham khảo

- [1] "Splunk Documentation" [Trực tuyến]. Available: https://docs.splunk.com/Documentation/Splunk/9.2.0
- [2] Welcome | Documentation | Splunk Developer Program
- [3] James Miller, Mastering Splunk
- [4] https://ffeathers.wordpress.com/2011/06/18/wiki-documentation-splunk-on-mediawiki/
- [5] https://infohub.delltechnologies.com/en-US/l/white-paper-cloud-native-splunk-enterprise-with-smartstore-predictive-maintenance-for-it-operations/splunk-documentation-2/