HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

VƯƠNG MINH VIỆT

NGHIÊN CỨU XÂY DỰNG HỆ THỐNG PHÂN TÍCH LOG TRUY NHẬP CHO PHÁT HIỆN BẤT THƯỜNG VÀ CÁC NGUY CƠ AN TOÀN THÔNG TIN

CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN

MÃ SỐ: 8.48.01.04

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI – 2019

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: **TS. HOÀNG XUÂN DẬU**

Phản biện 1: ……

Phản biện 2: ……

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: … giờ … ngày … tháng … năm ……

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

Log (còn gọi là nhật ký, hay vết) là các mục thông tin do hệ điều hành, hoặc các ứng dụng sinh ra trong quá trình hoạt động. Mỗi bản ghi log thường được sinh ra theo 1 hoạt động, hoặc sự kiện, nên còn được gọi là nhật ký sự kiện (event log). Các nguồn sinh log phổ biến bao gồm các thiết bị mạng (như router, firewall,…), hệ điều hành, các máy chủ dịch vụ (máy chủ web, máy chủ cơ sở dữ liệu, máy chủ DNS, email,…) và các chương trình ứng dụng. Mục đích của việc thu thập, xử lý và phân tích log bao gồm:

- Kiểm tra sự tuân thủ các chính sách an ninh;
- Kiểm tra sự tuân thủ vấn đề kiểm toán và luật pháp;
- Phục vụ điều tra số;
- Phục vụ phản ứng các sự cố mất an toàn thông tin ;
- Hiểu các hành vi của người dùng trực tuyến, trên cơ sở đó tối ưu hóa hệ thống Việc xử lý và phân tích log có nhiều ứng dụng, đặc biệt trong đảm bảo an toàn thông tin và cải thiện chất lượng hệ thống và các dịch vụ kèm theo, như quảng cáo trực tuyến. Hiện nay, trên thế giới đã có một số nền tảng và công cụ cho thu thập, xử lý và phân tích các dạng log phiên bản thương mại cũng như mã mở, như IBM Qradar SIEM, Splunk, Graylog và Logstash,... Tuy nhiên, việc nghiên cứu sâu các phương pháp xử lý và phân tích log và ứng dụng ở Việt Nam vẫn cần được tiếp tục thực hiện nhằm xây dựng các mô hình, hệ thống xử lý và phân tích log hiệu quả với chi phí hợp lý. Đây cũng là mục đích của đề tài luận văn *"Nghiên cứu xây dựng hệ thống phân tích log truy nhập cho phát hiện bất thường và các nguy cơ an toàn thông tin"*.

Luận văn bao gồm ba chương chính với nội dung như sau:

- Chương 1: Tổng quan về phân tích log truy nhập: khái niệm log truy nhập, các dạng log truy nhập, các phương pháp thu thập, xử lý và phân tích log, ứng dụng của phân tích log và giới thiệu một số nền tảng, công cụ phân tích log.

- Chương 2: Các kỹ thuật và mô hình xử lý, phân tích log truy nhập: Mô hình xử lý log; Thu thập và tiền xử lý; Các kỹ thuật phân tích log: Các kỹ thuật nhận dạng mẫu (Pattern Discovery), phân tích mẫu (Pattern Analysis), phân tích tương quan (Correlation Analysis).

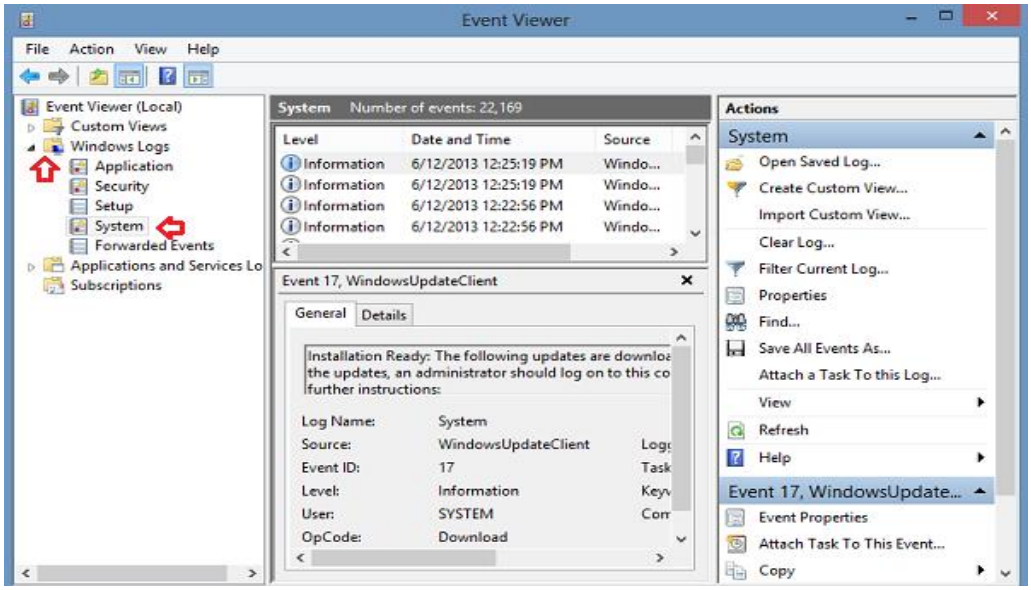
- Chương 3: Cài đặt, thử nghiệm và đánh giá: Giới thiệu môi trường và công cụ thử nghiệm; Cài đặt hệ thống: Cài đặt OSSEC, cài đặt ELK, kết hợp OSSEC và ELK; Nội dung thử nghiệm, kết quả và nhận xét.

CHƯƠNG 1. TỔNG QUAN VỀ PHÂN TÍCH LOG TRUY NHẬP

1.1. Khái quát về log truy nhập

1.1.1. Khái niệm log truy nhập

Log truy nhập hay nhật ký, hoặc vết truy nhập (gọi tắt là log) là một danh sách các bản ghi mà một hệ thống ghi lại khi xuất hiện các yêu cầu truy nhập các tài nguyên của hệ thống **Error! Reference source not found.**. Chẳng hạn, log truy nhập web (gọi tắt là web log) chứa tất cả các yêu cầu truy nhập các tài nguyên của một website. Các tài nguyên của một website, như các file ảnh, các mẫu định dạng và file mã JavaScript. Khi một người dùng thăm một trang web để tìm một sản phẩm, máy chủ web sẽ tải xuống thông tin và ảnh của sản phẩm và log truy nhập sẽ ghi lại các yêu cầu của người dùng đến các tài nguyên thông tin và ảnh của sản phẩm.



Hình 1.1. Xem Windows log sử dụng công cụ Event Viewer

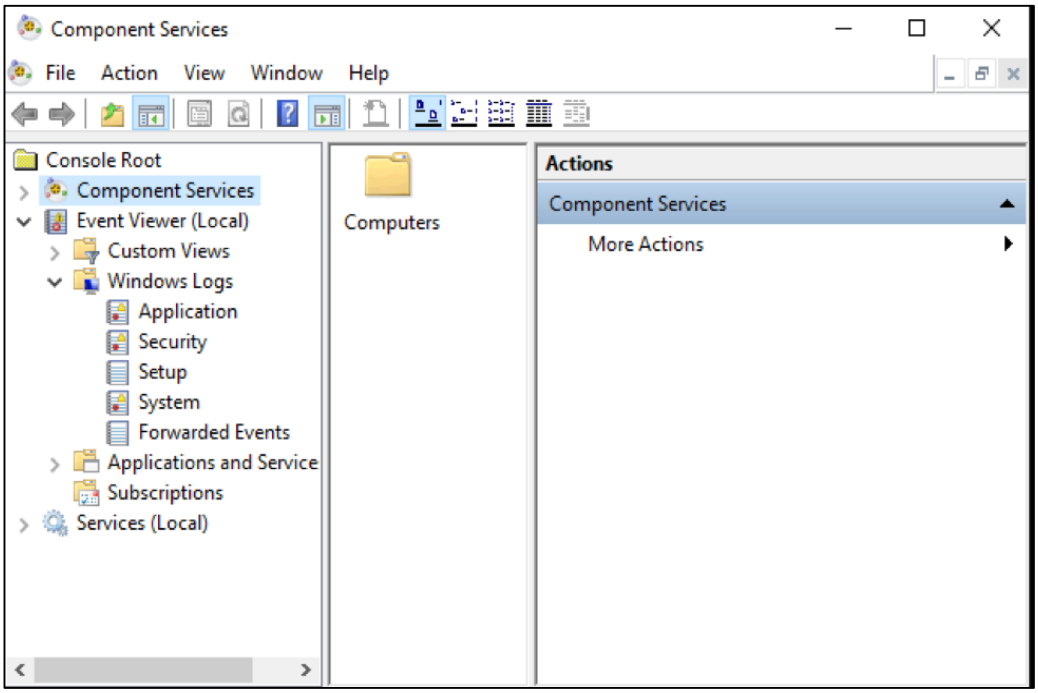
Hình 1.2. Các bản ghi log tạo bởi máy chủ e-mail.

1.1.2. Các dạng log truy nhập

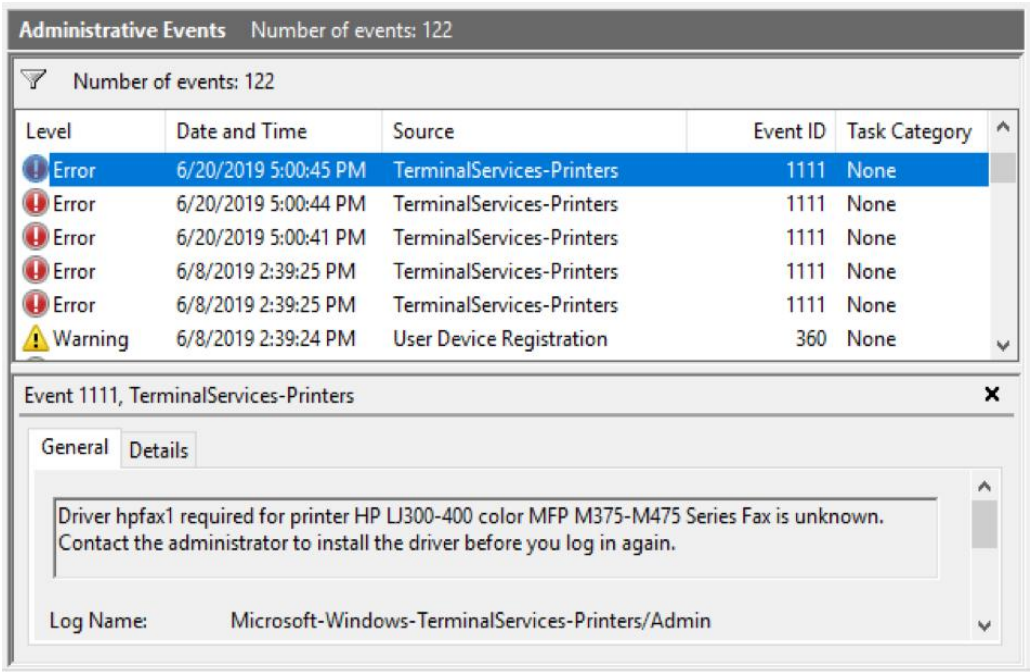
Như đã đề cập, có nhiều nguồn sinh log trong hệ thống, như log sinh bởi hệ điều hành, log sinh bởi các máy chủ dịch vụ mạng và log sinh bởi các thiết bị mạng và thiết bị đảm bảo an toàn thông tin. Mục này trình bày khái quát về các dạng log này.

1.1.2.1. Log sinh bởi hệ điều hành

a. Windows logs



Hình 1.3. Các thành phần của Windows Logs ***Error! Reference source not found.***



Hình 1.4. Một bản ghi Windows log mô tả lỗi dịch vụ ***Error! Reference source not found.***

b. Linux/Unix logs

```
*.err;kern.debug;auth.notice /dev/console
daemon,auth.notice           /var/log/messages
lpr.info                     /var/log/lpr.log
mail.*                       /var/log/mail.log
ftp.*                        /var/log/ftp.log
auth.*                       @prep.ai.mit.edu
auth.*                       root,amrood
netinfo.err                  /var/log/netinfo.log
install.*                    /var/log/install.log
*.emerg                      *
*.alert                      |program_name
mark.*                       /dev/console
```

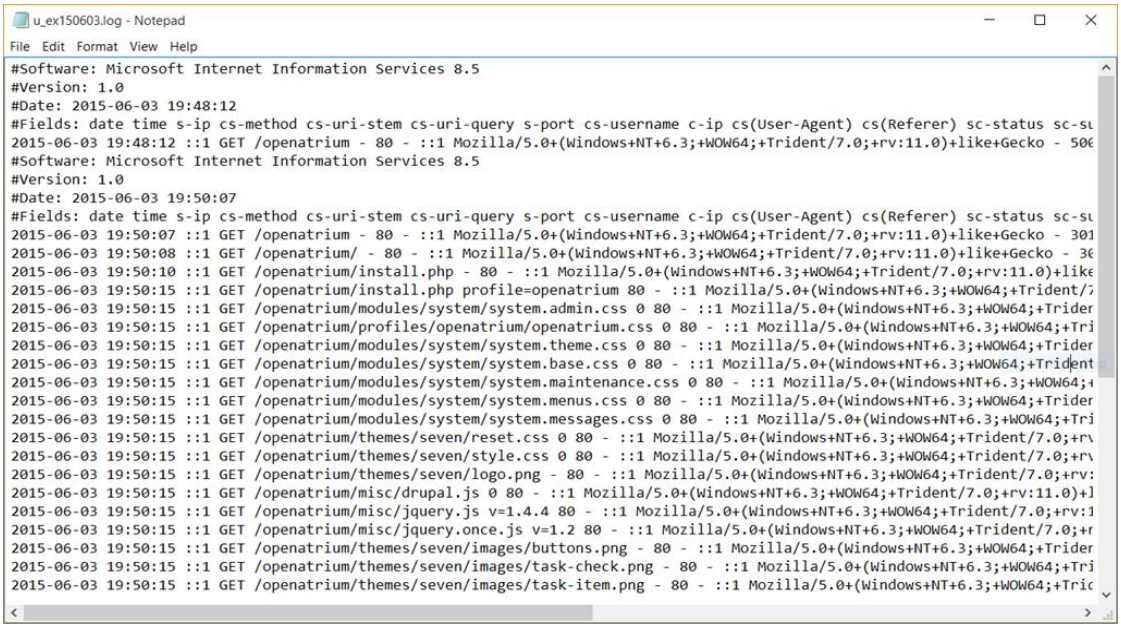
Hình 1.5. Một phần tập tin cấu hình syslog - syslog.conf

```
Jun  1 22:20:05 secserv kernel: Kernel logging (proc) stopped.
Jun  1 22:20:05 secserv kernel: Kernel log daemon terminating.
Jun  1 22:20:06 secserv exiting on signal 15
Nov 27 08:05:57 galileo kernel: Kernel logging (proc) stopped.
Nov 27 08:05:57 galileo kernel: Kernel log daemon terminating.
Nov 27 08:05:57 galileo exiting on signal 15
```

Hình 1.6. Một số bản ghi kern log của hệ điều hành Linux

1.1.2.2. Log sinh bởi các dịch vụ mạng

a. Web log



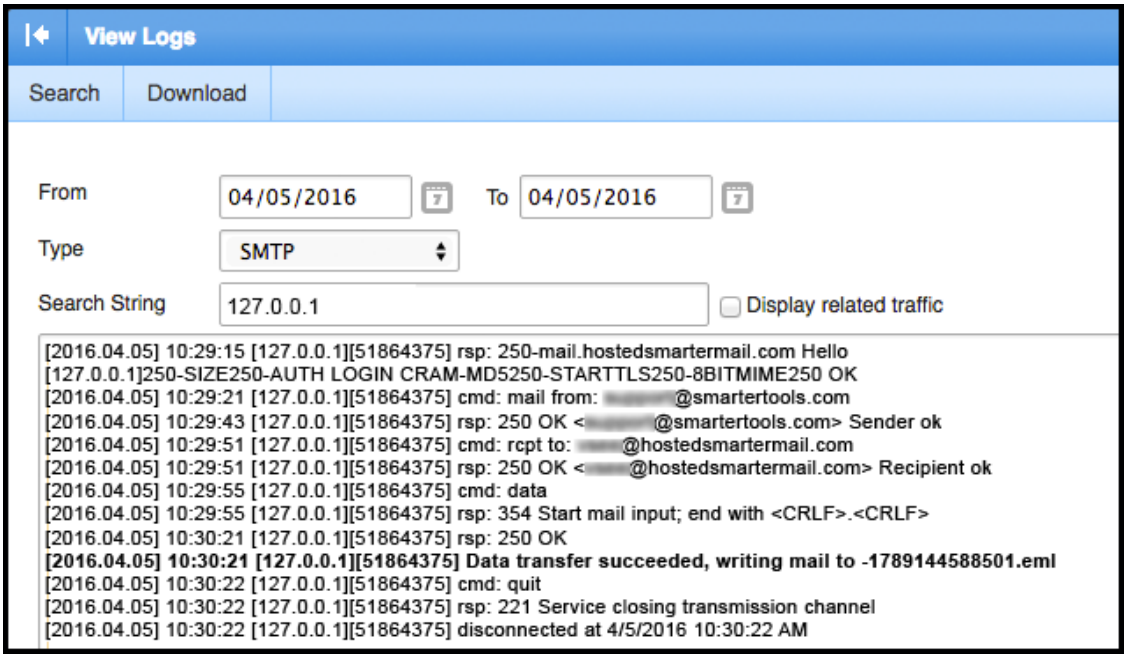
Hình 1.7. Một phần file log theo định dạng W3C Extended log file format

b. DNS log

No.	Time	Source	Destination	Protocol	Length	Info
803	4.416709000	192.168.2.52	192.168.2.4	DNS	85	Standard query 0x8040 A cooking.stackexchange.com
808	4.419638000	192.168.2.52	192.168.2.4	DNS	89	Standard query 0xcc0f A electronics.stackexchange.com
810	4.424106000	192.168.2.52	192.168.2.4	DNS	83	Standard query 0x967f A emacs.stackexchange.com
818	4.471499000	192.168.2.4	192.168.2.52	DNS	101	Standard query response 0x8040 A 198.252.206.16
819	4.472231000	192.168.2.52	192.168.2.4	DNS	85	Standard query 0xbb20 A gamedev.stackexchange.com
820	4.472263000	192.168.2.4	192.168.2.52	DNS	105	Standard query response 0xcc0f A 198.252.206.16
821	4.472725000	192.168.2.52	192.168.2.4	DNS	83	Standard query 0x8b07 A money.stackexchange.com
822	4.473017000	192.168.2.4	192.168.2.52	DNS	99	Standard query response 0x967f A 198.252.206.16
823	4.473380000	192.168.2.52	192.168.2.4	DNS	83	Standard query 0xe44a A music.stackexchange.com
832	4.518543000	192.168.2.4	192.168.2.52	DNS	99	Standard query response 0x8b07 A 198.252.206.16
833	4.519218000	192.168.2.52	192.168.2.4	DNS	86	Standard query 0x893d A outdoors.stackexchange.com
834	4.519302000	192.168.2.4	192.168.2.52	DNS	101	Standard query response 0xbb20 A 198.252.206.16
836	4.520024000	192.168.2.52	192.168.2.4	DNS	89	Standard query 0xefb3 A programmers.stackexchange.com
838	4.523652000	192.168.2.4	192.168.2.52	DNS	99	Standard query response 0xe44a A 198.252.206.16
839	4.524066000	192.168.2.52	192.168.2.4	DNS	86	Standard query 0x422b A puzzling.stackexchange.com
848	4.564736000	192.168.2.4	192.168.2.52	DNS	105	Standard query response 0xefb3 A 198.252.206.16
849	4.565487000	192.168.2.4	192.168.2.52	DNS	102	Standard query response 0x893d A 198.252.206.16
850	4.565516000	192.168.2.52	192.168.2.4	DNS	81	Standard query 0x6350 A rpg.stackexchange.com
851	4.566114000	192.168.2.52	192.168.2.4	DNS	84	Standard query 0x5bcd A travel.stackexchange.com
855	4.567001000	192.168.2.4	192.168.2.52	DNS	102	Standard query response 0x422b A 198.252.206.16
856	4.567692000	192.168.2.52	192.168.2.4	DNS	85	Standard query 0x0261 A tridion.stackexchange.com
867	4.606352000	192.168.2.4	192.168.2.52	DNS	101	Standard query response 0x0261 A 198.252.206.16
868	4.606361000	192.168.2.4	192.168.2.52	DNS	97	Standard query response 0x6350 A 198.252.206.16
869	4.607101000	192.168.2.4	192.168.2.52	DNS	100	Standard query response 0x5bcd A 198.252.206.16
870	4.607158000	192.168.2.52	192.168.2.4	DNS	84	Standard query 0x268d A area51.stackexchange.com
871	4.607175000	192.168.2.52	192.168.2.4	DNS	86	Standard query 0x234b A bicycles.stackexchange.com

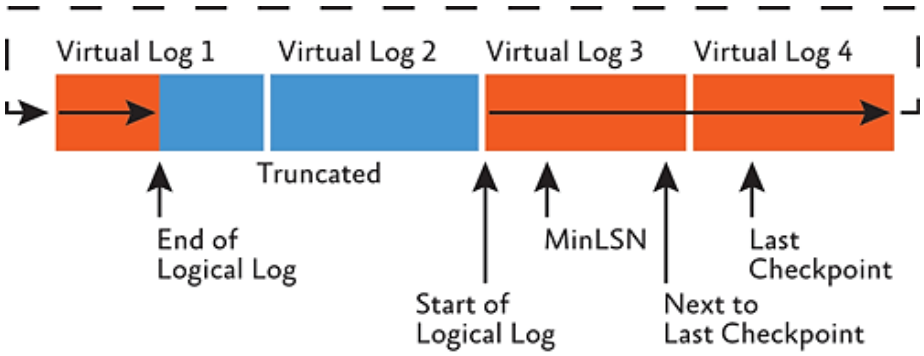
Hình 1.8. Trích xuất một số bản ghi DNS log

c. Mail log



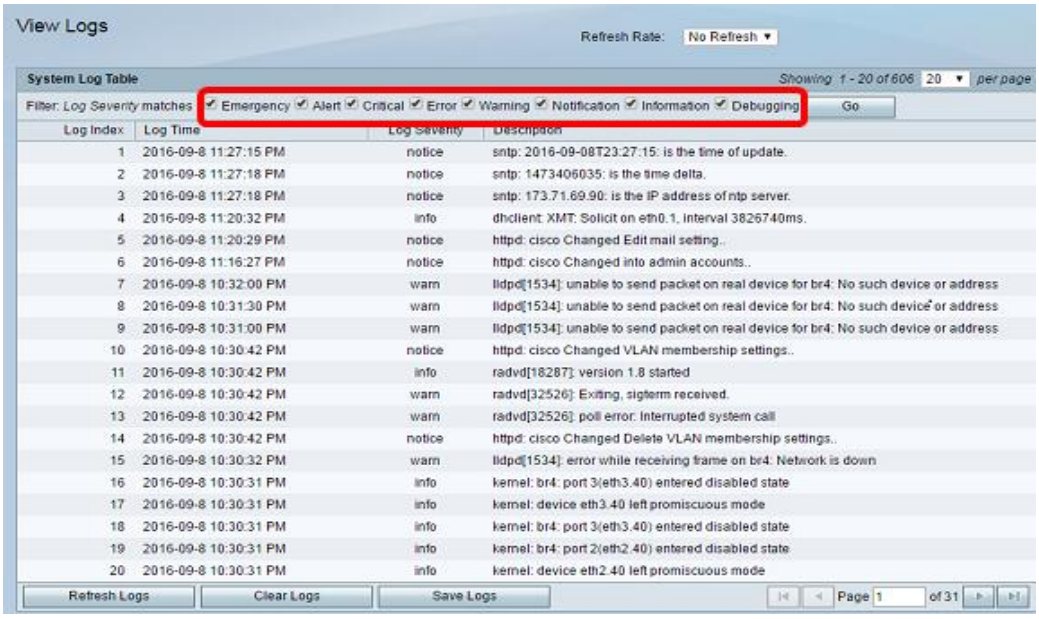
Hình 1.9. Một phần log truy nhập máy chủ email SMTP

d. Database log



Hình 1.10. Mô hình quản lý dữ liệu log của Microsoft SQL Server

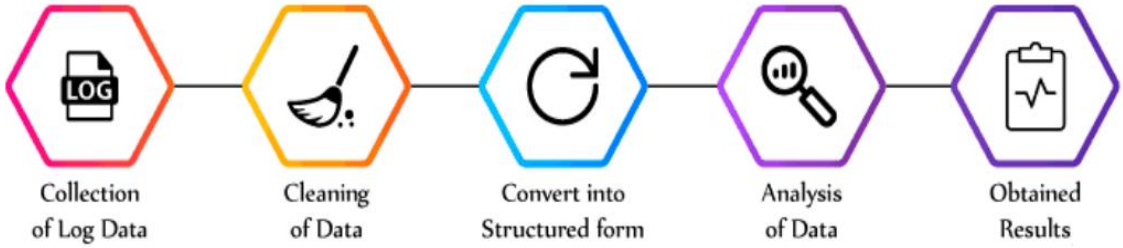
1.1.2.3. Log sinh bởi các thiết bị mạng và thiết bị đảm bảo ATTT



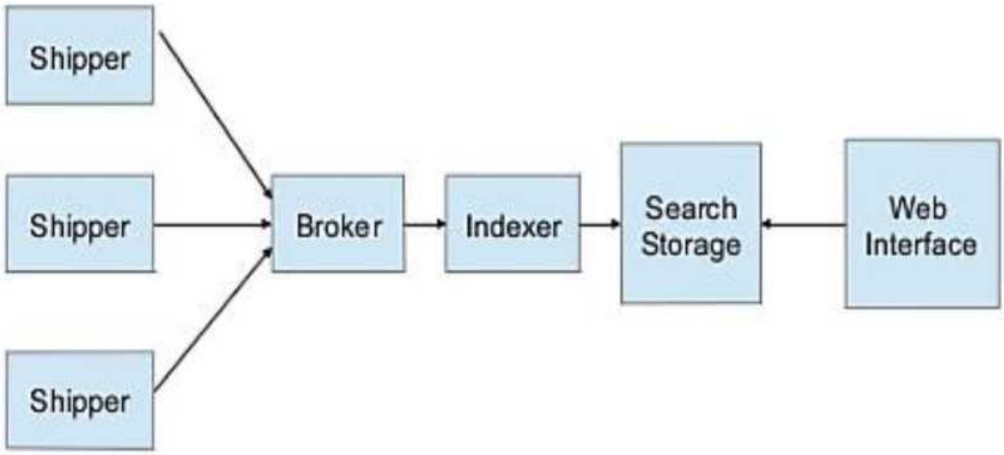
Hình 1.11. Một phần log của Cisco RV Series Router

1.2. Thu thập, xử lý và phân tích log truy nhập

Thu thập, xử lý và phân tích log là các khâu cơ bản của một hệ thống phân tích log. Hình 1.12 biểu diễn các khâu cụ thể của quá trình thu thập, xử lý và phân tích log thường được áp dụng trên thực tế. Theo đó, các khâu xử lý cụ thể gồm:



Hình 1.12. Các khâu của quá trình thu thập, xử lý và phân tích log



Hình 1.13. Kiến trúc điển hình của hệ thống thu thập, xử lý và phân tích log

1.3. Ứng dụng của phân tích log truy nhập

Việc phân tích log truy cập thường được thực hiện cho các mục đích **Error! Reference source not found.**:

- Đảm bảo an toàn thông tin cho hệ thống;
- Hỗ trợ khắc phục sự cố hệ thống;
- Hỗ trợ điều tra số;
- Hỗ trợ hiểu được hành vi người dùng trực tuyến.

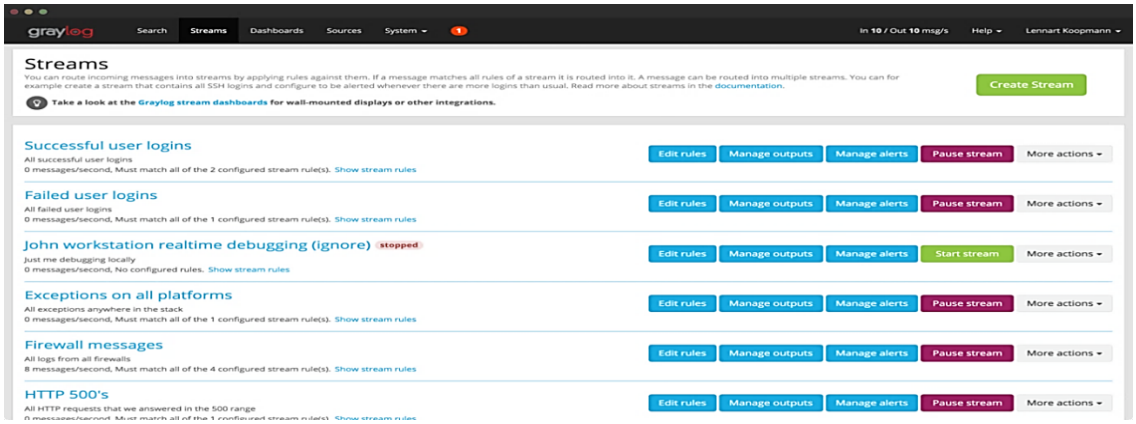
1.4. Một số nền tảng và công cụ xử lý, phân tích log

Có nhiều nền tảng và công cụ xử lý, phân tích log truy cập thương mại cũng như mã nguồn mở được cung cấp hiện nay như Splunk **Error! Reference source not found.**, Sumo Logic, VNCS Web Monitoring, ELK Stack **Error! Reference source not found.**, Graylog, Webzlizer, IBM QRadar SIEM và OSSEC... Mục này giới thiệu

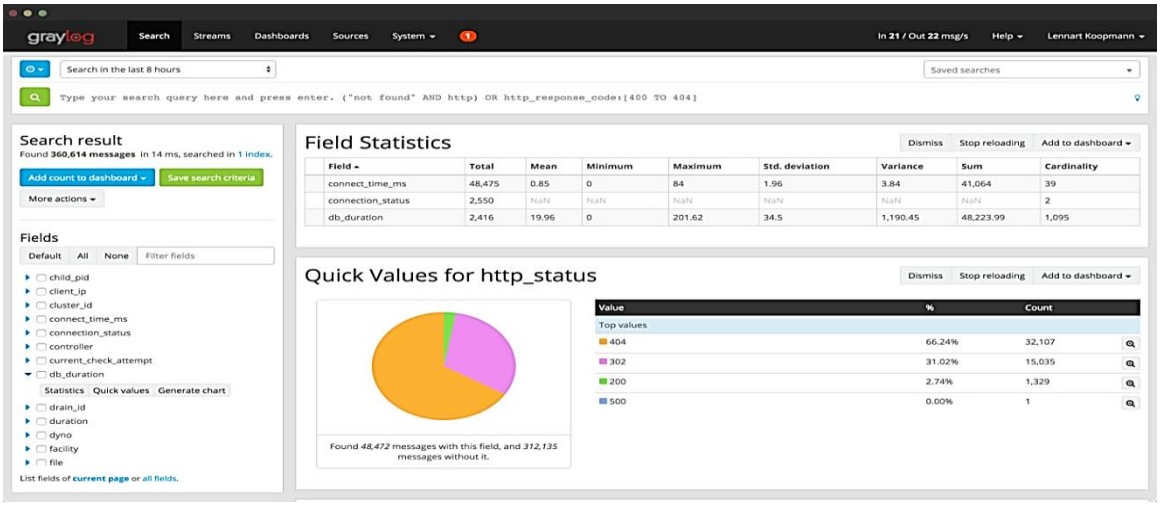
khái quát về tính năng và các ưu nhược điểm của một số công cụ phân tích log điển hình, bao gồm Graylog, Webzlizer, và ELK Stack, và một số công cụ thu thập và xử lý log cho đảm bảo ATTT, bao gồm IBM QRadar SIEM và OSSEC.

1.4.1. Các công cụ phân tích log điển hình

1.4.1.1. Graylog

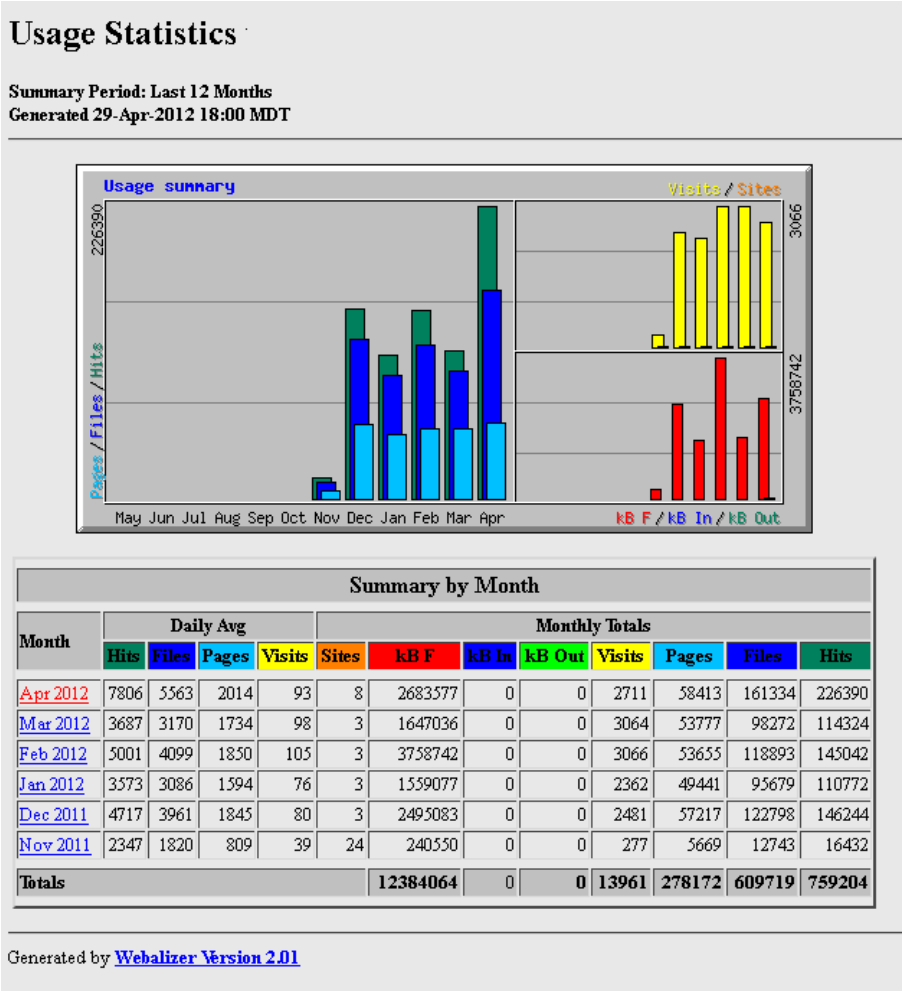


Hình 1.14. Màn hình quản lý các nguồn thu thập log của Graylog ***Error! Reference source not found.***



Hình 1.15. Màn hình báo cáo tổng hợp của Graylog ***Error! Reference source not found.***

1.4.1.2. Webalizer



Usage Statistics

Summary Period: Last 12 Months
Generated 29-Apr-2012 18:00 MDT

Summary by Month

Month	Daily Avg				Monthly Totals							
	Hits	Files	Pages	Visits	Sites	kB F	kB In	kB Out	Visits	Pages	Files	Hits
Apr 2012	7806	5563	2014	93	8	2683577	0	0	2711	58413	161334	226390
Mar 2012	3687	3170	1734	98	3	1647036	0	0	3064	53777	98272	114324
Feb 2012	5001	4099	1850	105	3	3758742	0	0	3066	53655	118893	145042
Jan 2012	3573	3086	1594	76	3	1559077	0	0	2362	49441	95679	110772
Dec 2011	4717	3961	1845	80	3	2495083	0	0	2481	57217	122798	146244
Nov 2011	2347	1820	809	39	24	240550	0	0	277	5669	12743	16432
Totals						**12384064**	0	**0**	**13961**	**278172**	**609719**	**759204**

Generated by Webalizer Version 2.01

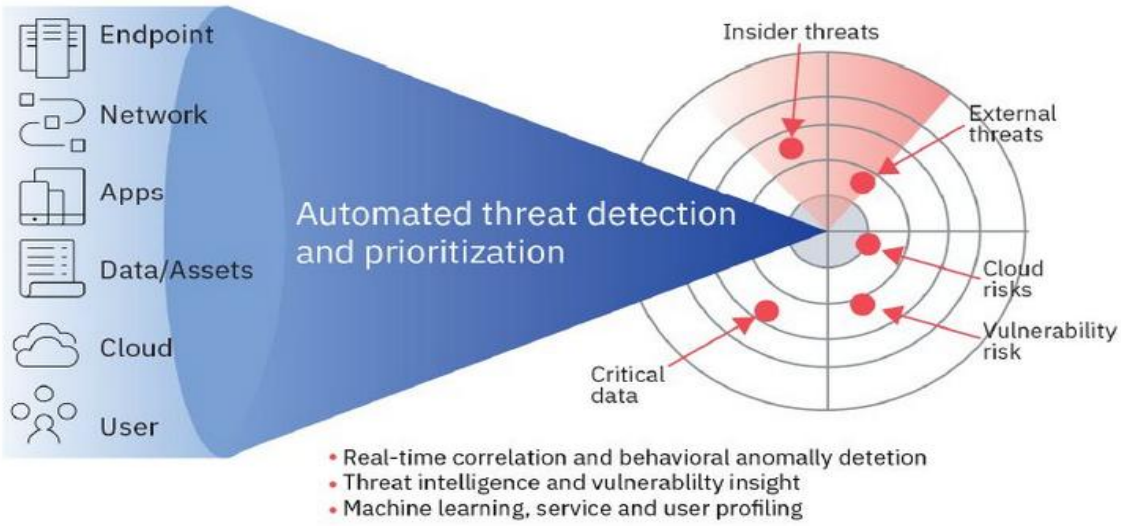
Hình 1.16. Một mẫu báo cáo của Webalizer ***Error! Reference source not found.***

1.4.1.3. ELK Stack

1.4.2. Các công cụ thu thập và xử lý log cho đảm bảo ATTT

1.4.2.1. IBM QRadar SIEM

QRadar SIEM (Security Information and Event Management) **Error! Reference source not found.** là hệ thống quản lý các thông tin và sự cố an ninh được phát triển và cung cấp bởi hãng IBM, Hoa Kỳ. QRadar SIEM cho phép phát hiện các bất thường, các nguy cơ an toàn thông tin với độ chính xác cao và tỷ lệ cảnh báo sai thấp thông qua việc xử lý, phân tích dữ liệu log và luồng mạng từ hàng ngàn thiết bị và ứng dụng phân tán trong mạng, như minh họa trên Hình 1.17.



Hình 1.17. Mô hình thu thập và xử lý dữ liệu của QRadar SIEM ***Error! Reference source not found.***

1.4.2.2. Hệ thống phát hiện xâm nhập OSSEC

1.4.3. Nhận xét

Bảng 1.1 cung cấp thông tin so sánh các ưu điểm và nhược điểm của các nền tảng, công cụ xử lý, phân tích log truy nhập đã đề cập ở trên.

Bảng 1.1. So sánh các công cụ xử lý log truy cập

Công cụ	**Ưu điểm**	**Nhược điểm**
Graylog	- Mã mở, miễn phí - Hỗ trợ phân tích log truy cập từ nhiều nguồn và phân tích hành vi người dùng dùng cho cho phát hiện và cảnh báo các truy cập bất thường cũng như trích xuất các mẫu hành vi truy cập phục vụ cho tối ưu hóa các trang web	- Không có khả năng phân tích chuyên sâu các nguy cơ mất an toàn thông tin, như dấu hiệu xuất hiện các dạng mã độc và các dạng tấn công lên các dịch vụ và tài nguyên mạng.
Webalizer	- Mã mở, miễn phí - Có khả năng phân tích nhiều dạng web log - Các báo cáo dưới dạng biểu đồ có tính biểu biễn cao.	- Chỉ có khả năng phân tích tình hình sử dụng các trang web - Ít có khả năng trích xuất các thông tin cho cảnh báo các nguy cơ mất an toàn thông tin.
ELK Stack	- Mã mở, miễn phí - Chi phí cài đặt, vận hành thấp	- Không có khả năng phân tích chuyên sâu các nguy cơ mất an

	- Hỗ trợ trích xuất các mẫu hành vi truy cập phục vụ cho tối ưu hóa các trang web - Giao diện hiển thị đa dạng, phong phú	toàn thông tin, như dấu hiệu xuất hiện các dạng mã độc và các dạng tấn công lên các dịch vụ và tài nguyên mạng.
IBM QRadar SIEM	- Hỗ trợ thu thập và xử lý nhiều loại log khác nhau với khối lượng lớn và dữ liệu từ luồng mạng - Hỗ trợ thu thập dữ liệu từ hàng ngàn thiết bị mạng - Hỗ trợ phát hiện các bất thường, các nguy cơ ATTT với độ chính xác cao và tỷ lệ cảnh báo sai thấp	- Chi phí cài đặt ban đầu và phí bản quyền khá lớn - Đòi hỏi thiết bị chuyên dụng - Khó khăn trong vận hành và bảo trì.
OSSEC	- Mã mở, miễn phí - Hỗ trợ thu thập và xử lý nhiều loại log khác nhau - Hỗ trợ phát hiện các bất thường, các nguy cơ ATTT - Hỗ trợ giám sát tính toàn vẹn của các file và tham số hệ thống	- Giao diện hiển thị và cảnh báo hạn chế - Khó quản trị, giám sát - Việc kết nối giám sát nhiều phân đoạn mạng gặp khó khăn.

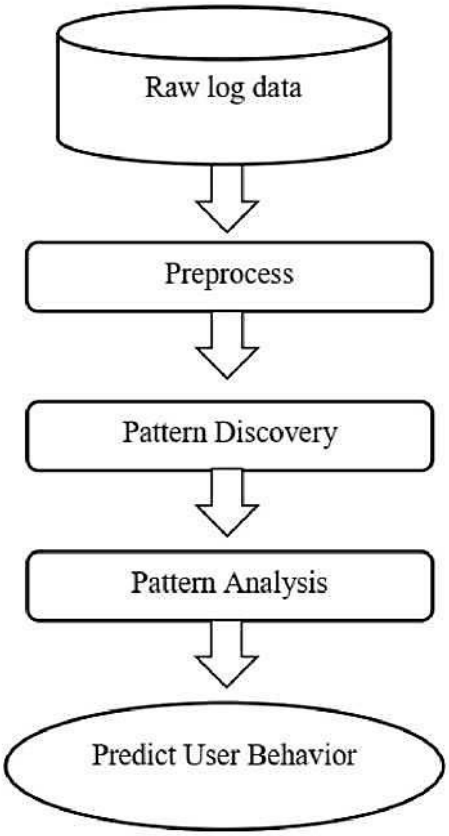
1.5. Kết luận chương

Chương này đã trình bày khái quát về log truy nhập, các nguồn sinh log, vấn đề thu thập, xử lý và phân tích log. Chương cũng giới thiệu chi tiết các dạng log truy nhập phổ biến, các khâu xử lý, phân tích log cũng như ứng dụng của phân tích log. Đồng thời, chương cũng khảo sát một số nền tảng và công cụ xử lý, phân tích log phổ biến hiện nay và rút ra nhận xét.

CHƯƠNG 2. CÁC KỸ THUẬT VÀ MÔ HÌNH XỬ LÝ, PHÂN TÍCH LOG TRUY NHẬP

2.1. Mô hình xử lý log

Hình 2.1 mô tả mô hình xử lý log truy nhập khái quát, mô hình gồm các pha chính: Pha tiền xử lý và chuẩn hóa - Preprocess; Pha nhận dạng mẫu - Pattern Discovery; Pha phân tích mẫu - Pattern Analysis; Pha dự đoán hành vi người dùng - Predict User Behavior.



Hình 2.1. Mô hình xử lý log truy nhập khái quát

- Tiền xử lý và chuẩn hóa - Preprocess:

2.2. Thu thập và tiền xử lý

2.2.1. Thu thập log

Log truy nhập có thể được sinh ra ở nhiều vị trí khác nhau trong mạng, do đó có nhiều cách để thu thập log. Log có thể được nhận từ nhiều nguồn khác nhau như: từ file, từ mạng internet hay từ đầu ra của các ứng dụng khác. Một số nguồn thu thập log cụ thể có thể kể ra như:

- Lấy các sự kiện từ file log.
- Nhận đầu ra của các công cụ dòng lệnh như là một sự kiện.

- Tạo các sự kiện dựa trên các bản tin SNMP.
- Đọc các bản tin syslog.
- Đọc sự kiện từ một TCP socket.
- Đọc sự kiện thông qua giao thức UDP.
- Nhận các sự kiện từ framework Elastic Beats.
- Đọc các kết quả truy vấn từ một cụm Elasticsearch.

2.2.2. Tiền xử lý và chuẩn hóa log

Quá trình tiền xử lý và chuẩn hóa thực hiện việc làm sạch, hợp nhất dữ liệu từ nhiều nguồn khác nhau và chuẩn hóa dữ liệu theo một định dạng thống nhất. Quá trình này cung cấp các dữ liệu tối ưu và thống nhất cho quá trình phân tích log.

2.2.2.1. Làm sạch và hợp nhất dữ liệu

2.2.2.2. Chuẩn hóa log

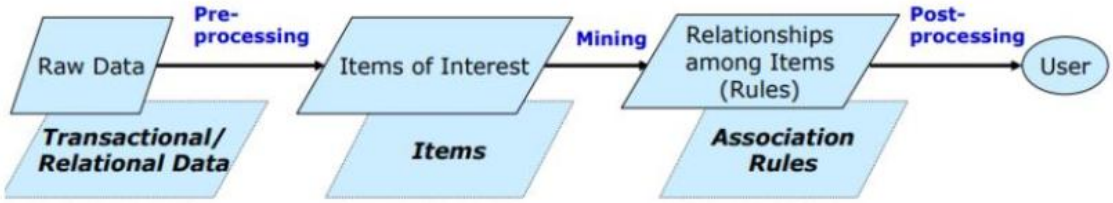
2.3. Các kỹ thuật phân tích log

2.3.1. Các kỹ thuật nhận dạng và phân tích mẫu

2.3.1.1. Phân tích thống kê

2.3.1.2. Luật kết hợp

Phương pháp này nhằm phát hiện ra các luật kết hợp giữa các thành phần dữ liệu trong CSDL. Mẫu đầu ra của giải thuật khai phá dữ liệu là tập luật kết hợp tìm được.

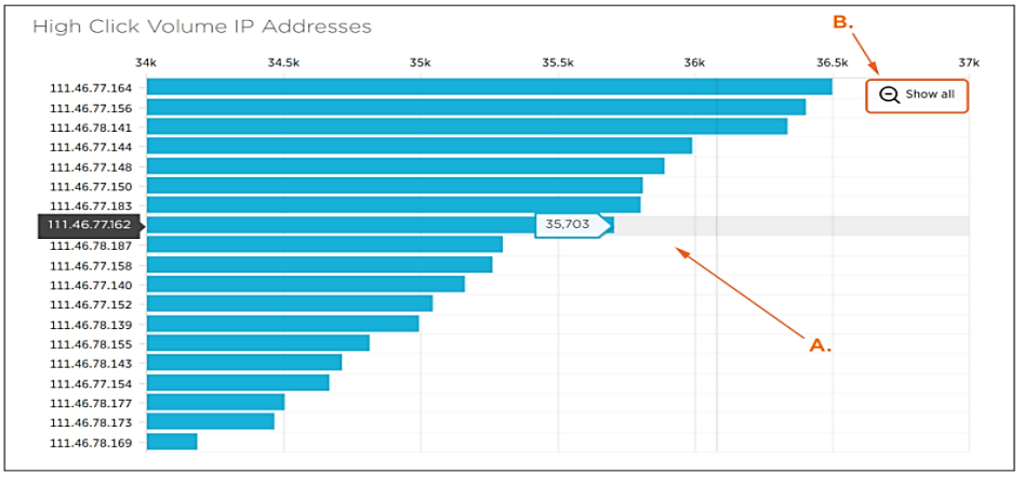


Hình 2.2. Quá trình sử dụng luật kết hợp

2.3.1.3. Phân lớp

2.3.1.4. Phân cụm

2.3.1.5. Phân tích mẫu



Hình 2.3. Phân tích mẫu sử dụng data visualization

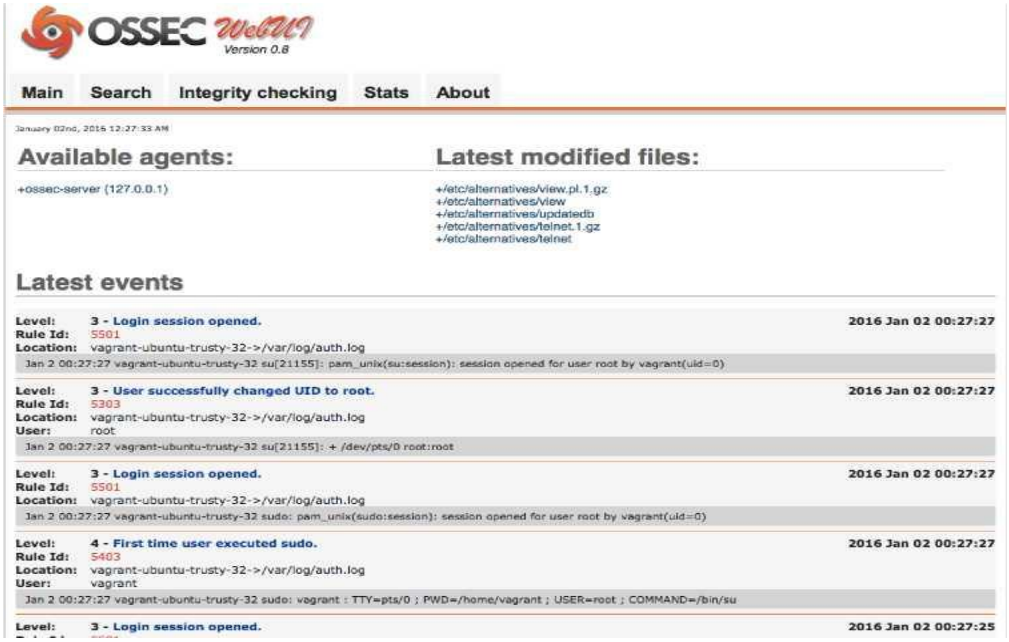
2.3.2. Phân tích tương quan

2.4. Xây dựng mô hình phân tích log dựa trên OSSEC kết hợp ELK Stack cho phát hiện bất thường và các nguy cơ ATTT

2.4.1. Hệ thống phát hiện xâm nhập OSSEC

2.4.1.1. Giới thiệu

OSSEC là hệ thống phát hiện xâm nhập dựa trên host (HIDS - Host-based Intrusion Detection) dựa trên log mã nguồn mở, miễn phí, đa nền tảng có thể mở rộng và có nhiều cơ chế bảo mật khác nhau. OSSEC có thể phát hiện xâm nhập bằng cả chữ ký hoặc dấu hiệu bất thường. OSSEC cung cấp kiến trúc đa nền tảng tập trung, cho phép quản lý bảo mật máy tính từ một vị trí trung tâm.



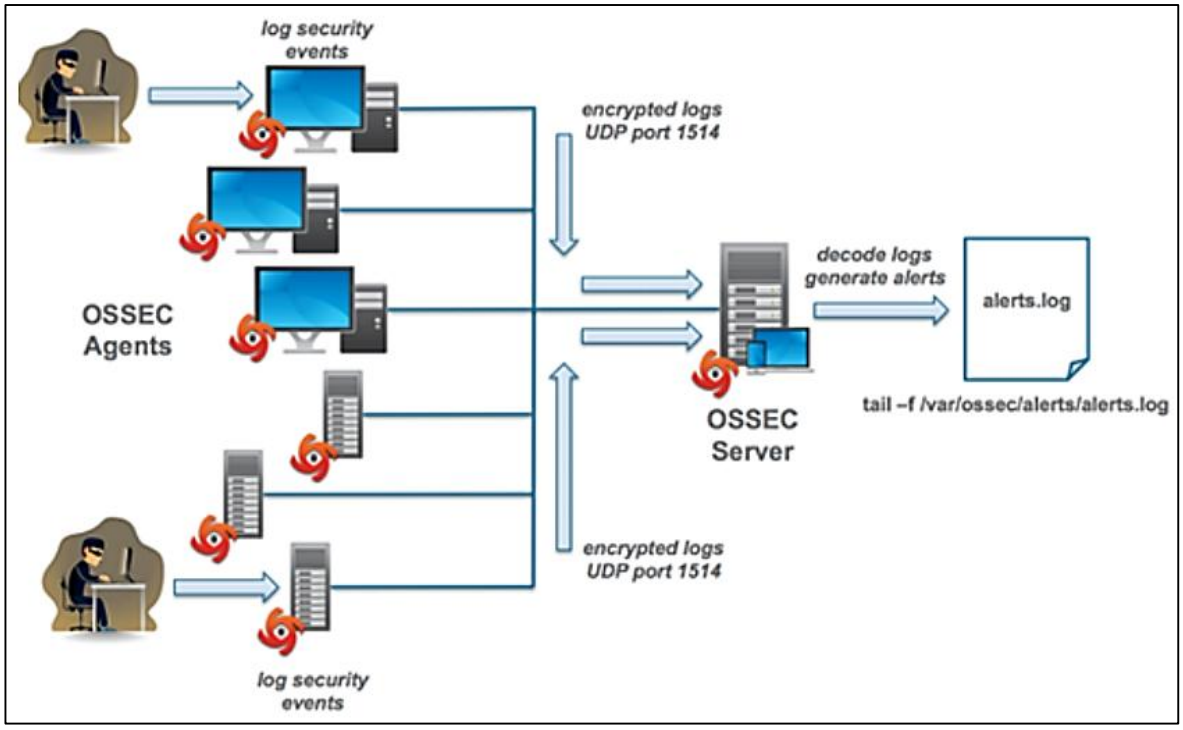
Hình 2.4. Giao diện người dùng của OSSEC

2.4.1.2. Các tính năng nổi bật của OSSEC

Các tính năng nổi bật của OSSEC bao gồm:

- *Theo dõi và phân tích các log:*
- *Kiểm tra tính toàn vẹn của file:*
- *Giám sát Registry:*
- *Phát hiện Rootkit:*
- *Phản ứng chủ động:*
- *Giám sát toàn vẹn tập tin:*

2.4.1.3. Kiến trúc và hoạt động của OSSEC

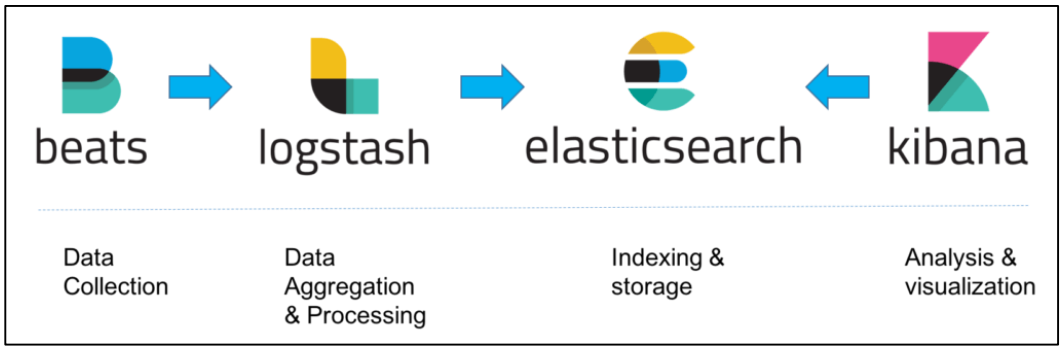


Hình 2.5. Luồng hoạt động của hệ thống phát hiện xâm nhập OSSEC ***Error! Reference source not found.Error! Reference source not found.***

2.4.2. Bộ công cụ xử lý và phân tích log ELK Stack

2.4.2.1. Giới thiệu

Hình 2.6 biểu diễn các thành phần chính của ELK Stack và tương tác giữa chúng. Theo đó, các thành phần ELK Stack gồm:



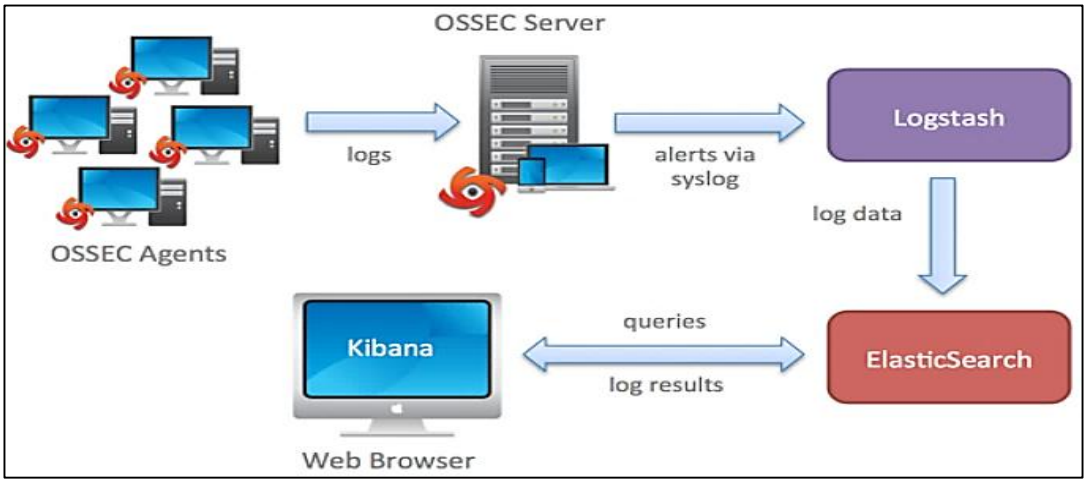
Hình 2.6. Các thành phần của bộ công cụ xử lý và phân tích log ELK ***Error! Reference source not found.***

2.4.2.2. Các ưu điểm khi sử dụng ELK Stack

2.4.3. Mô hình triển khai tích hợp OSSEC và ELK Stack

Đầu ra tiêu chuẩn của hệ thống phát hiện xâm nhập OSSEC là các cảnh báo dưới dạng các dòng log lưu vào file, như biểu diễn trên Hình 2.5. Gói phần mềm OSSEC cũng có thành phần hỗ trợ giao diện web, nhưng có tính năng khá hạn chế và không hỗ trợ phân tích chuyên sâu log kết quả **Error! Reference source not found.Error!**

Reference source not found.. Trong khi đó, ELK Stack là bộ công cụ cho phép thu thập, xử lý và phân tích log chuyên sâu với nhiều tính năng mạnh và khả năng hiển thị, trình bày phong phú. Do vậy, việc tích hợp ELK Stack với hệ thống phát hiện xâm nhập OSSEC cho phép khai thác hiệu quả các điểm mạnh của ELK Stack, bổ trợ hiệu quả cho OSSEC. Điều này giúp tạo thành một hệ thống xử lý và phân tích log cho phát hiện bất thường và nguy cơ an toàn thông tin với khả năng quản lý log với khối lượng lớn và các tính năng phân tích log chuyên sâu và khả năng hiển thị log cũng như kết quả xử lý đa dạng dưới nhiều hình thức khác nhau.



Hình 2.7. Mô hình tích hợp OSSEC và ELK ***Error! Reference source not found.***

Hình 2.7 biểu diễn mô hình tích hợp OSSEC và ELK **Error! Reference source not found.**. Theo đó, dữ liệu log và các cảnh báo (alert) xuất ra từ OSSEC được xử lý tiếp như sau:

- Dữ liệu log và các cảnh báo (gọi chung là log) được thu thập và xử lý bởi thành phần Logstash. Tại đây, log được làm sạch, chuẩn hóa và chuyển sang khâu tiếp theo.
- Dữ liệu log sau chuẩn hóa được chuyển đến ElasticSearch quản lý và lập chỉ số phục vụ phân tích, tìm kiếm.
- Kibana là thành phần cuối cùng trong hệ thống cho phép phân tích log chuyên sâu và biểu diễn log và kết quả xử lý dưới nhiều dạng khác nhau (báo cáo, đồ thị, biểu đồ,…).

2.5. Kết luận chương

Chương 2 đã trình bày về các kỹ thuật xử lý và phân tích log, bao gồm mô hình khái quát cho xử lý và phân tích log, vấn đề tiền xử lý, chuẩn hóa log, cũng như các kỹ thuật phân tích log. Phần cuối chương mô tả việc xây dựng mô hình phân tích log dựa

trên OSSEC kết hợp ELK Stack cho phát hiện bất thường và các nguy cơ ATTT làm cơ sở cho thử nghiệm tại chương 3.

CHƯƠNG 3. CÀI ĐẶT, THỬ NGHIỆM VÀ ĐÁNH GIÁ

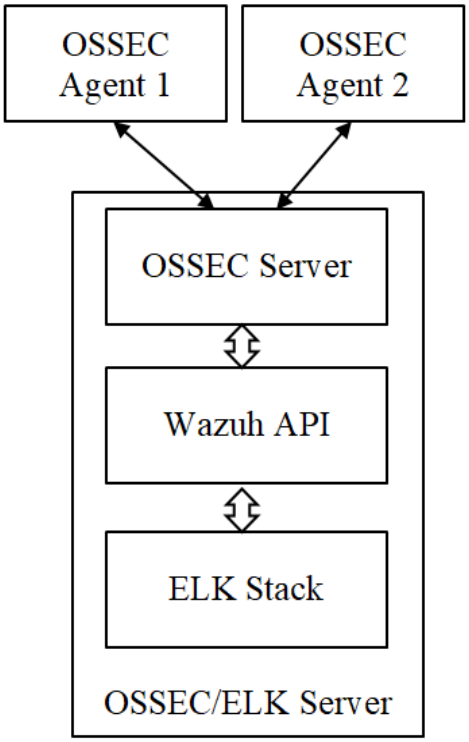
3.1. Môi trường thử nghiệm và mô hình triển khai cài đặt

3.1.1. Môi trường và công cụ thử nghiệm

Môi trường thử nghiệm sử dụng trong luận văn là hệ thống mạng mô phỏng dựa trên phần mềm ảo hóa VMWare Professional 15. Các phần mềm và công cụ thử nghiệm bao gồm:

3.1.2. Mô hình cài đặt hệ thống thử nghiệm

Hình 3.1 biểu diễn mô hình cài đặt hệ thống thử nghiệm. Theo đó, hệ thống thử nghiệm được triển khai cài đặt gồm 3 máy như sau:



Hình 3.1. Mô hình cài đặt hệ thống thử nghiệm

3.2. Triển khai cài đặt hệ thống thử nghiệm

Do bộ công cụ Wazuh đã tích hợp OSSEC server và các công cụ quản lý vào gói phần mềm Wazuh Manager, nên các thành phần thực tế cần cài đặt trên máy chủ OSSEC/ELK Server bao gồm: Wazuh Manager, Wazuh API, Filebeat và ELK Stack. Filebeat là công cụ cho phép thu thập log trên bản thân máy chủ OSSEC/ELK Server. Thành phần phải cài đặt trên các máy trạm/máy được giám sát là Wazuh agent. Bản chất của Wazuh agent là OSSEC agent được đóng gói trong gói phần mềm Wazuh.

3.2.1. Cài đặt Wazuh Manager, Wazuh API và Filebeat

3.2.1.1. Thêm thông tin gói phần mềm Wazuh vào thư viện quản lý của Ubuntu

3.2.1.2. Cài đặt Wazuh Manager

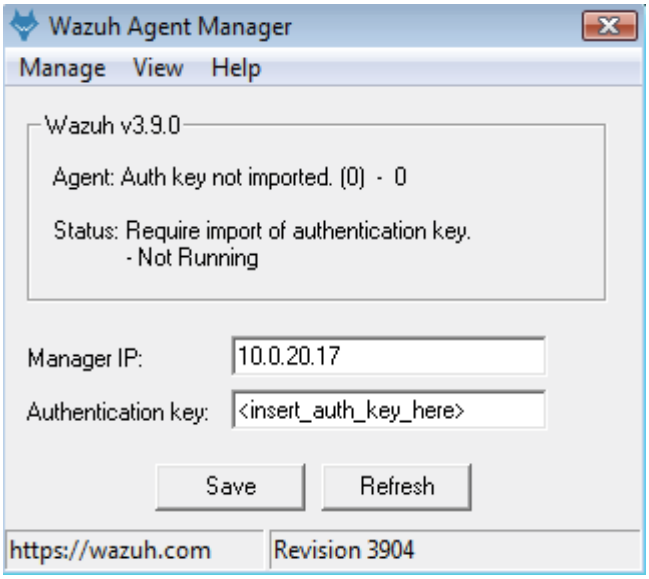
3.2.1.3. Cài đặt Wazuh API

3.2.1.4. Nạp thông tin gói ELK và khóa GPG vào thư viện

3.2.1.5. Cài đặt ElasticSearch

3.2.2. Cài đặt Wazuh agent trên các máy được giám sát

+ Thêm một agent (tên, địa chỉ IP)
+ Tạo chuỗi xác thực cho agent đó
+ Nạp địa chỉ IP của Wazuh Manager (192.168.186.130) và chuỗi xác thực vào các ô Manager IP và Authentication key, bấm Save và menu Manage / Restart để khởi động lại agent trong giao diện quản lý của agent, như biểu diễn trên Hình 3.2.



Hình 3.2. Giao diện quản lý, đăng ký Wazuh agent với Wazuh Manager

3.3. Thử nghiệm và kết quả

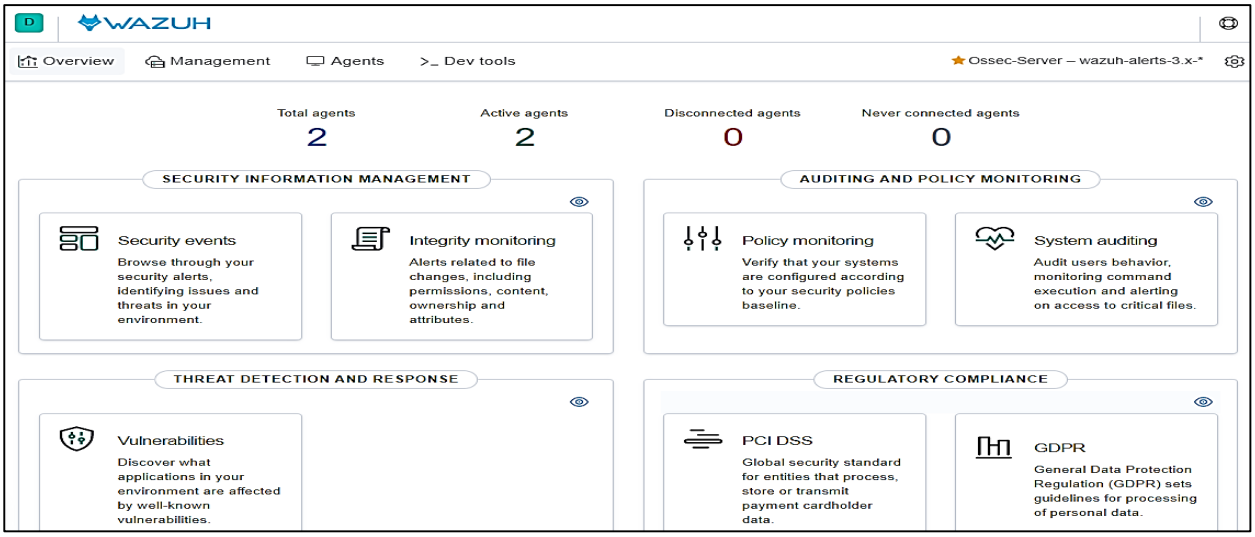
3.3.1. Nội dung thử nghiệm

Sau khi hoàn thành cài đặt và cấu hình xong các thành phần của hệ thống như mô tả trong Mục 3.2, mục này thử nghiệm một số tính năng thu thập, xử lý và phát hiện các bất thường trong hệ thống thử nghiệm. Cụ thể, các tính năng đã thử nghiệm bao gồm:

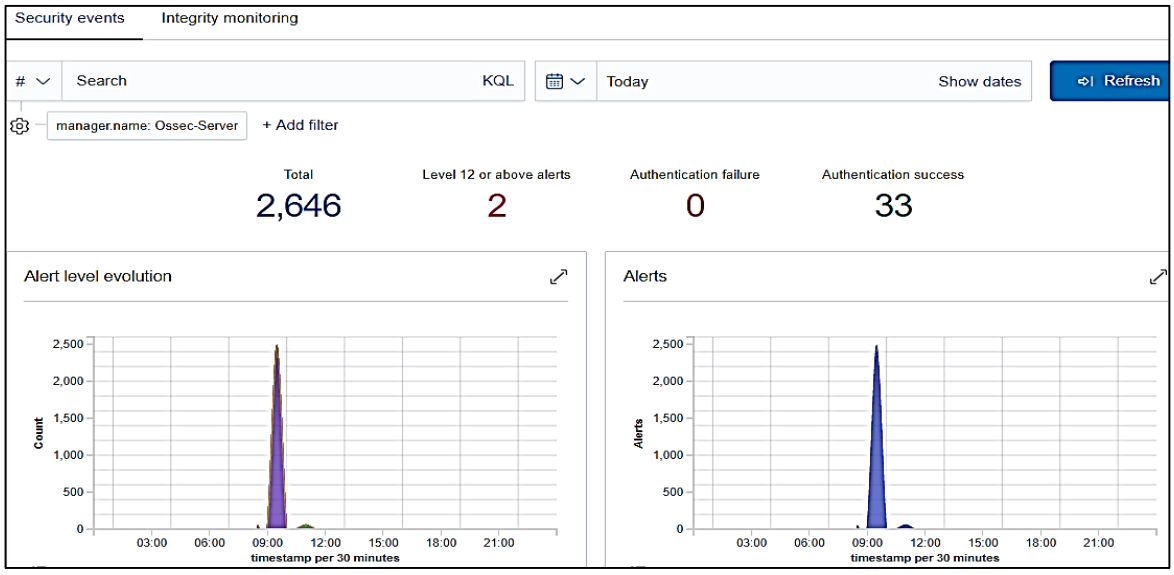
- Hiển thị màn hình tổng hợp các sự kiện an ninh và giám sát toàn vẹn file
- Quản lý hệ thống
- Quản lý và hiển thị thông tin từ các agent
- Công cụ cho nhà phát triển

3.3.2. Kết quả

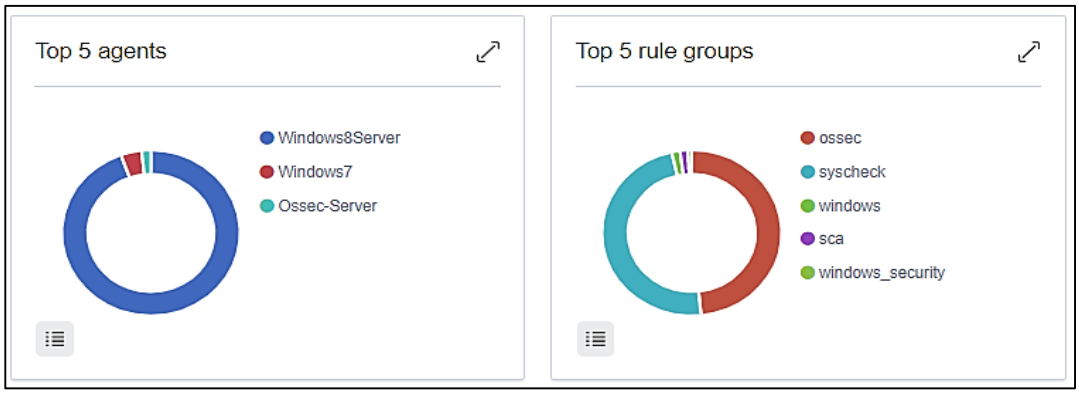
Mục này trình bày một số giao diện hệ thống là kết quả các thử nghiệm các nội dung đã trình bày ở Mục 3.3.1.



Hình 3.3. Giao diện tổng hợp của Wazuh OSSEC-ELK



Hình 3.4. Tổng hợp các sự kiện an ninh

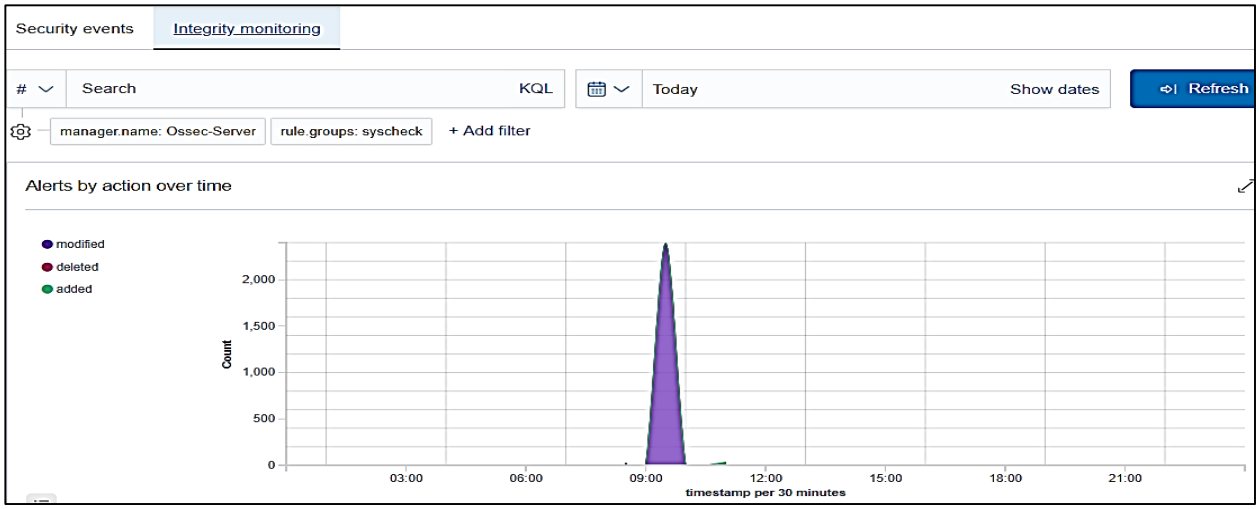


Hình 3.5. Các sự kiện an ninh thu thập từ top 5 agent và top 5 nhóm luật được kích hoạt

Alerts summary

Rule ID	Description	Level	Count
550	Integrity checksum changed.	7	2,428
554	File added to the system.	5	25
61104	Service startup type was changed	3	18
60106	Windows Logon Success	3	16
5501	PAM: Login session opened.	3	10
60137	Windows User Logoff	3	6
5502	PAM: Login session closed.	3	6
60775	SessionEnv was unavailable to handle a notification event	5	5
60118	Windows Workstation Logon Success	3	5
533	Listened ports status (netstat) changed (new port opened or closed).	7	5

Hình 3.6. Tổng hợp các cảnh báo an ninh



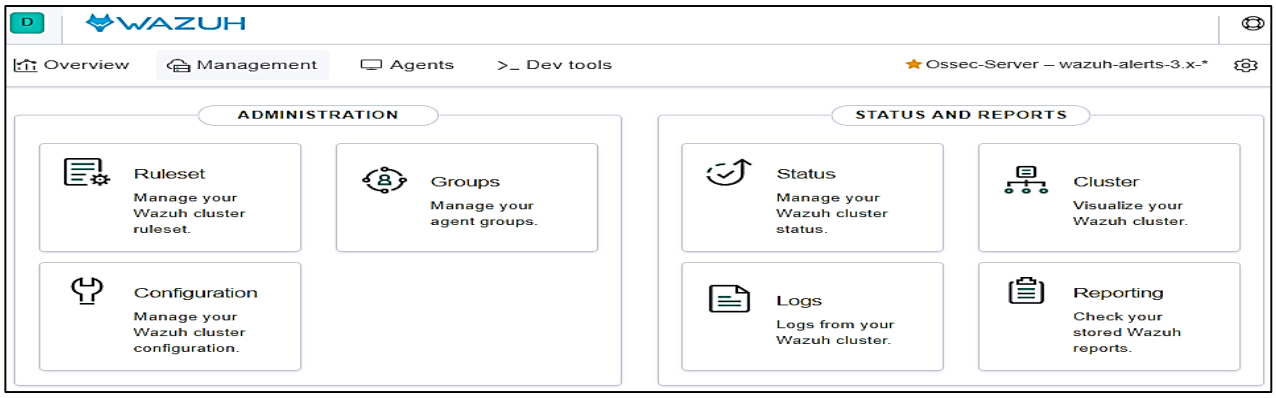
Hình 3.7. Tổng hợp giám sát tính toàn vẹn của file

Hình 3.8. Giám sát tính toàn vẹn của file chia theo agent

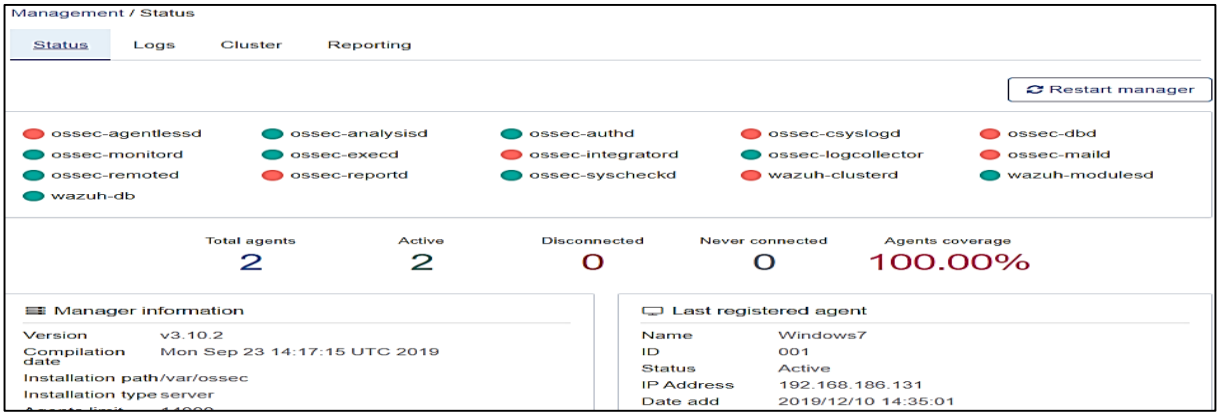
Alerts summary

Agent	Path	Action	Count
Ossec-Server	/etc/cups/subscriptions.conf.O	modified	3
Ossec-Server	/etc/cups/subscriptions.conf	modified	3
Windows7	HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\VSS\Diag\VolSnap	modified	3
Windows7	HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\LanmanServer\Parameters	modified	3
Ossec-Server	/boot/grub/grubenv	modified	2
Windows7	HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\mssmbios\Data	modified	2
Windows7	HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\VSS\Diag\SPP	modified	2
Windows7	HKEY_LOCAL_MACHINE\Security\SAM\Domains\Account	modified	2
Windows8Server	HKEY_LOCAL_MACHINE\Security\SAM\Domains\Account	modified	2
Windows7	HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\VSS\Diag\Lovelace	added	1

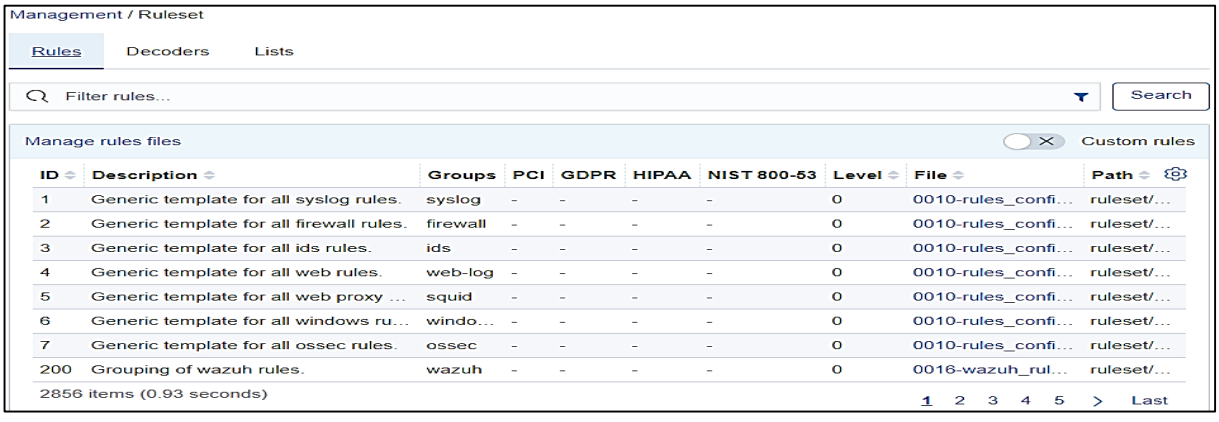
Hình 3.9. Tổng hợp các cảnh báo giám sát toàn vẹn file



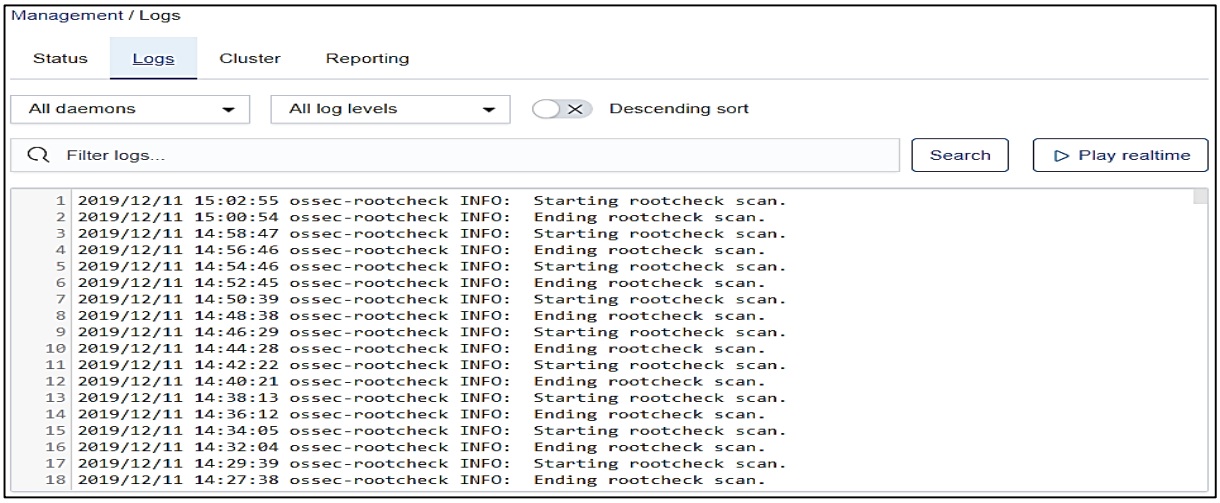
Hình 3.10. Màn hình quản lý hệ thống Management



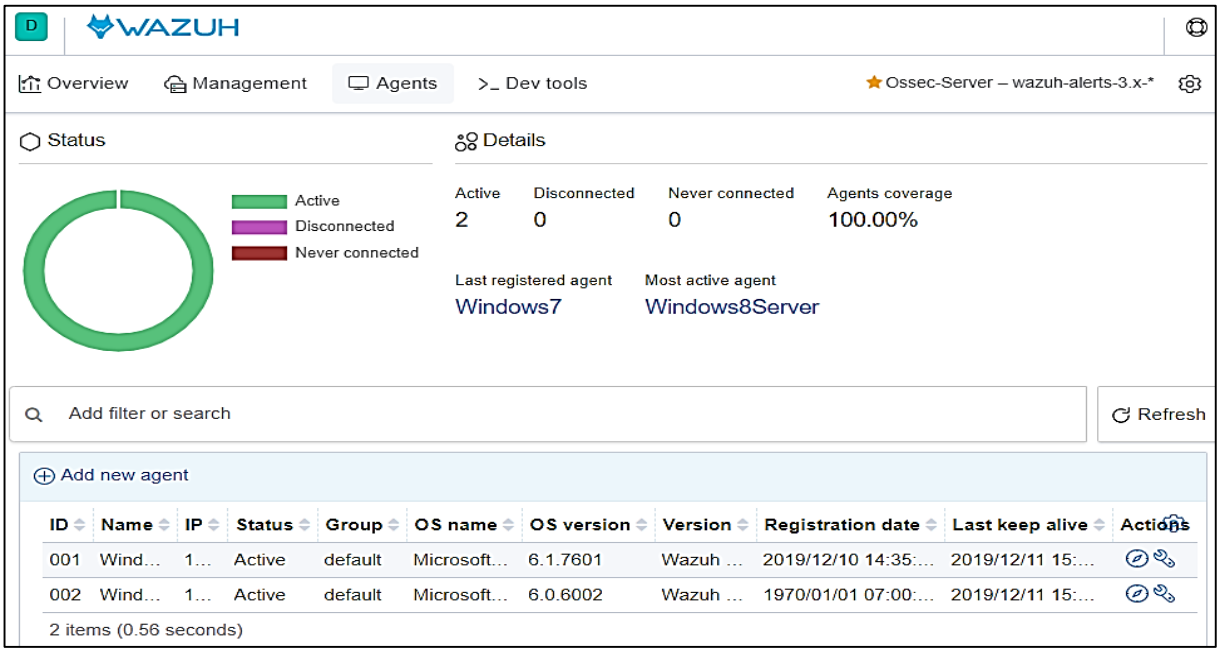
Hình 3.11. Trạng thái hệ thống



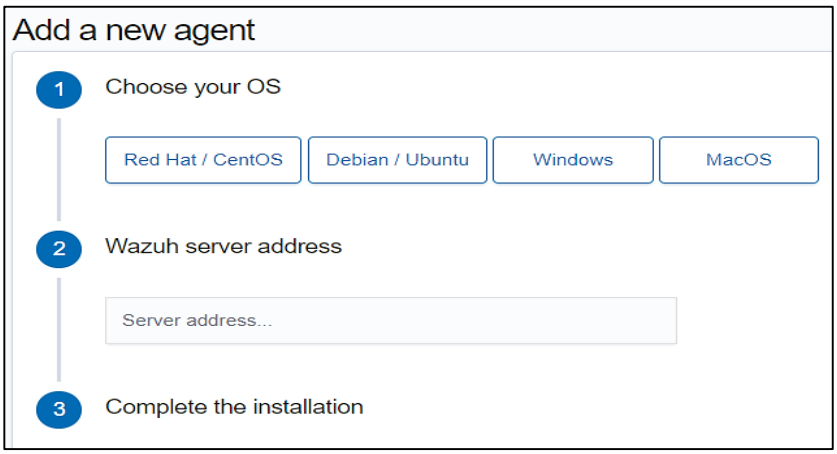
Hình 3.12. Tập luật dựng sẵn của OSSEC



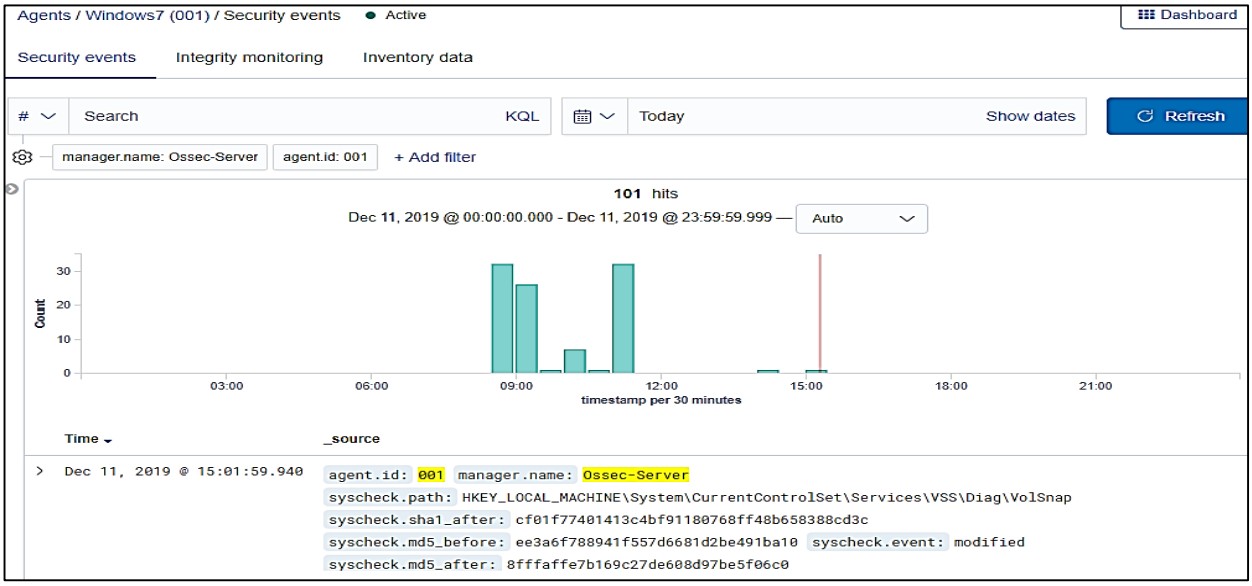
Hình 3.13. Hiển thị log thu thập hỗ trợ hiển thị theo thời gian thực



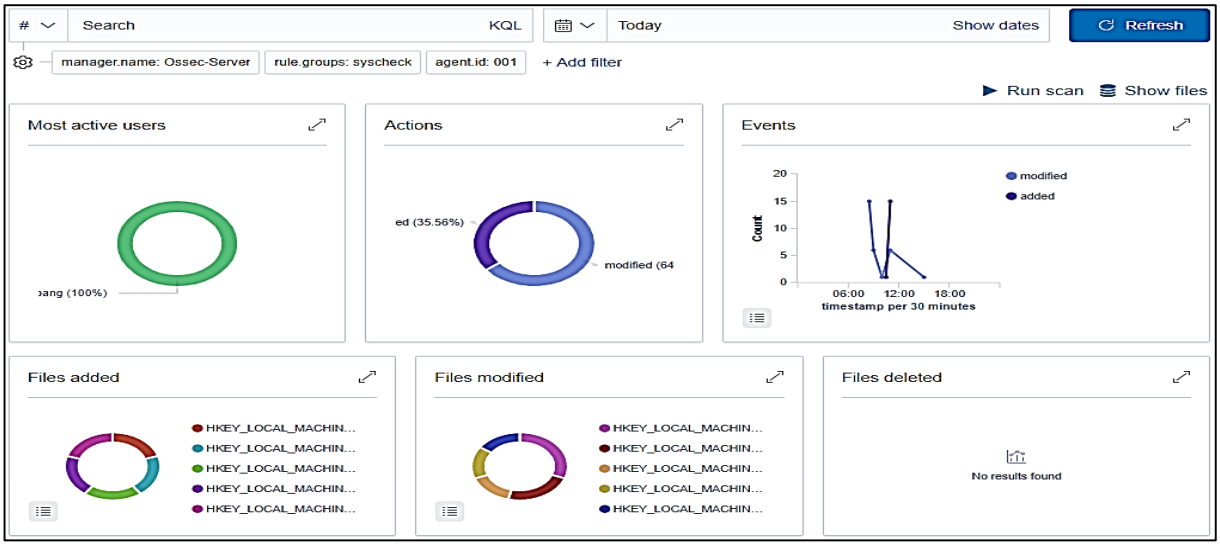
Hình 3.14. Giao diện hiển thị và quản lý các agent



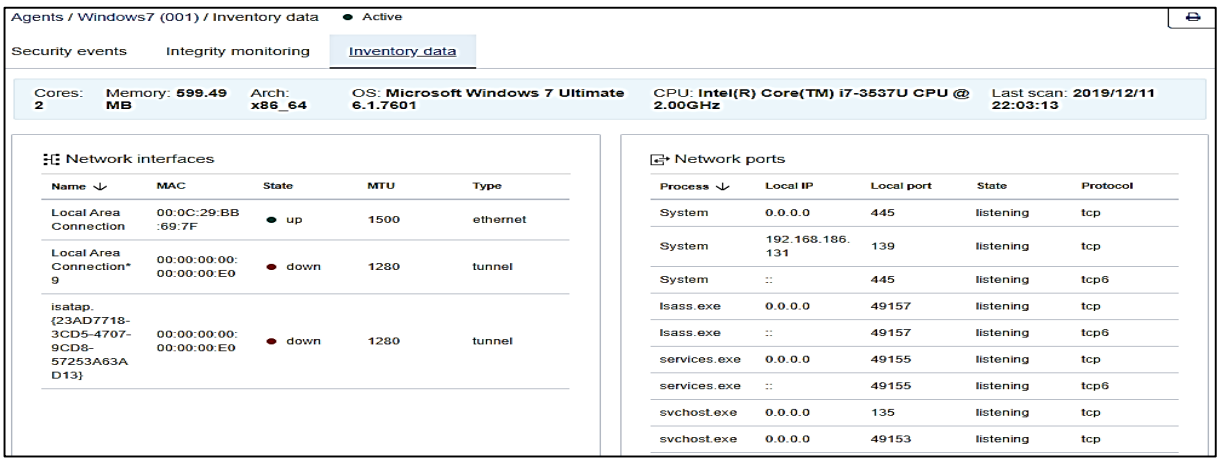
Hình 3.15. Hỗ trợ thêm agent



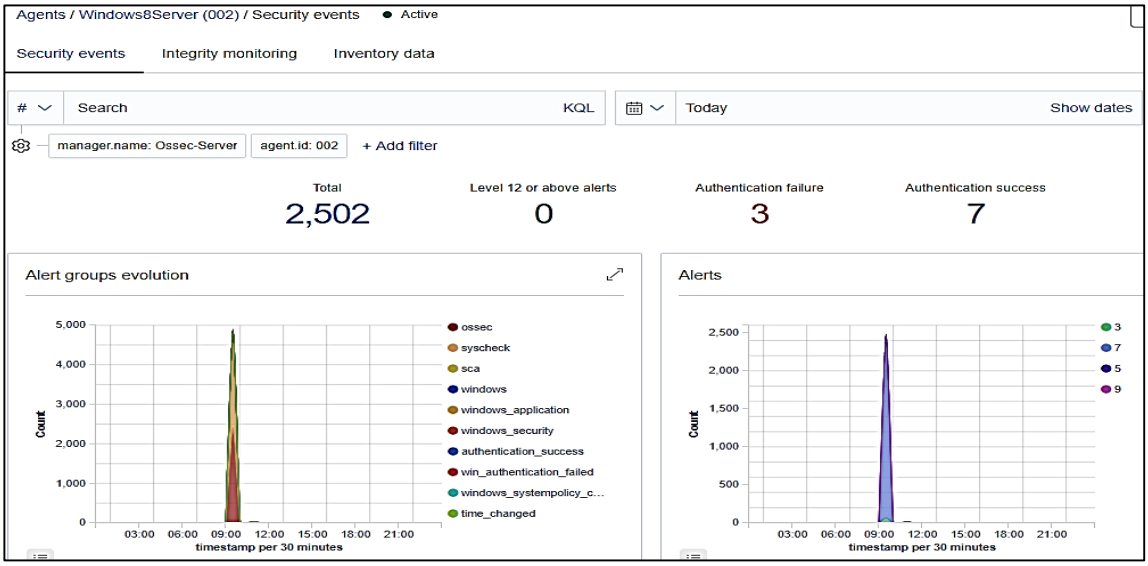
Hình 3.16. Các sự kiện an ninh từ agent số 001



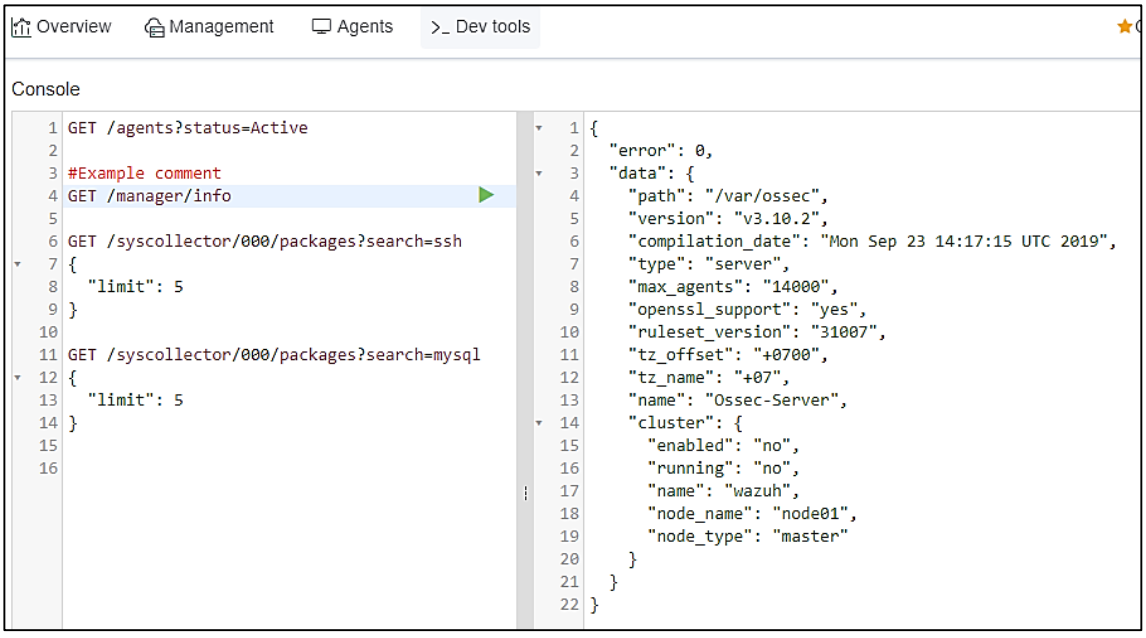
Hình 3.17. Giám sát tính toàn vẹn file từ agent 001



Hình 3.18. Giám sát sử dụng tài nguyên trên máy chạy agent 001



Hình 3.19. Giám sát tổng hợp từ máy chạy agent 002



Hình 3.20. Giao diện hỗ trợ phát triển – trực tiếp chạy các lệnh giám sát

3.3.3. Nhận xét

3.4. Kết luận chương

Chương 3 đã trình bày quá trình thử nghiệm triển khai cài đặt thử nghiệm hệ thống xử lý và phân tích log truy nhập cho phát hiện các bất thường và nguy cơ an toàn thông tin dựa trên việc tích hợp hệ thống phát hiện xâm nhập OSSEC và bộ công cụ xử lý phân tích log ELK. Các kết quả ban đầu cho thấy hệ thống tích hợp vận hành ổn định, có khả năng giám sát và phát hiện các bất thường và nguy cơ ATTT, cũng như hỗ trợ các tính năng quản trị đơn giản và hiển thị dữ liệu, kết quả đa dạng, có tính biểu diễn cao.

KẾT LUẬN

Các kết quả đạt được:

Luận văn này tập trung nghiên cứu về thu thập, xử lý, phân tích log truy cập, phục vụ phát hiện các hành vi bất thường và nguy cơ mất an toàn thông tin trong các hệ thống mạng. Các nội dung đã thực hiện trong luận văn bao gồm:

- Trình bày khái quát về log truy nhập, các dạng log truy nhập, vấn đề thu thập, xử lý và phân tích log truy nhập, cũng như ứng dụng của nó.
- Mô tả một số nền tảng và công cụ xử lý và phân tích log truy nhập, từ đó rút ra so sánh, đánh giá để tìm ra mô hình triển khai phù hợp.
- Trình bày mô hình xử lý và phân tích log khái quát, vấn đề tiền xử lý và chuẩn hóa log và các kỹ thuật phân tích log.
- Xây dựng, cài đặt và thử nghiệm thành công mô hình phân tích log dựa trên OSSEC kết hợp ELK Stack cho phát hiện bất thường và các nguy cơ ATTT trên hệ thống mạng mô phỏng.

Hướng phát triển:

Luận văn có thể được phát triển theo các hướng sau:

- Triển khai thử nghiệm mô hình phân tích log dựa trên OSSEC kết hợp ELK Stack cho phát hiện bất thường và các nguy cơ ATTT trên hệ thống mạng thực.
- Xây dựng và bổ sung thêm các luật giám sát, phát hiện bất thường và các nguy cơ ATTT, đảm bảo khả năng phát hiện kịp thời các bất thường và nguy cơ ATTT.