**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA AN TOÀN THÔNG TIN**



**ĐỒ ÁN TỐT NGHIỆP**

**ĐỀ TÀI: NGHIÊN CỨU, XÂY DỰNG HỆ THỐNG QUẢN LÝ LOG ĐẢM BẢO AN TOÀN SỬ DỤNG GRAYLOG**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn:** | **TS. Đinh Trường Duy** |
| **Sinh viên thực hiện:** | **Lê Việt Thọ** |
| **Mã sinh viên:** | **B18DCAT242** |
| **Lớp:** | **D18CQAT02-B** |
| **Khóa học:** | **2018-2023** |

**Hà Nội, 2023**

**LỜI CẢM ƠN**

Đầu tiên, em xin gửi lời cảm ơn chân thành và sâu sắc nhất tới Thầy giáo hướng dẫn đồ án tốt nghiệp của em – TS. Đinh Trường Duy. Trong suốt khoảng thời gian làm đồ án, thầy đã tận tình hướng dẫn và chỉ dạy, giúp em có những định hướng và phương pháp tiếp cận trong quá trình tìm hiểu, thực hiện cho tới những bước cuối cùng để đồ án được hoàn thành.

Ngoài ra, em xin gửi lời cảm ơn chân thành tới các Thầy, Cô trong Khoa Công Nghệ Thông Tin 1 và Khoa An Toàn Thông Tin và toàn thể các cán bộ của Học viện Công nghệ Bưu chính Viễn thông Hà Nội, đã truyền đạt cho em những kiến thức hữu ích và quan trọng trong suốt quá trình em học tập tại trường, để em có thể hoàn thành đồ án của mình một cách tốt nhất.

Cuối cùng, em xin cảm ơn gia đình, bạn bè và đồng nghiệp đã luôn ở bên em, quan tâm động viên, giúp em có điều kiện tốt nhất trong cả quá trình học tập cũng như khi thực hiện đồ án.

Với trình độ hiểu biết còn nhiều hạn chế của bản thân, và vốn kiến thức vẫn còn ít ỏi, nên trong đồ án của em không tránh khỏi những thiếu sót. Em rất mong nhận được sự góp ý của các Thầy, Cô để đồ án của em được hoàn thiện hơn. Em xin chân thành cảm ơn!

|  |  |
| --- | --- |
|  | Hà Nội, tháng 1 năm 2023 |
|  | Sinh viên thực hiện |
|  | **Lê Việt Thọ** |

**NHẬN XÉT, ĐÁNH GIÁ, CHO ĐIỂM**

**(Của Người hướng dẫn)**

……………………………………………………………………………………… .……………………………………………………………………………………… ……………………………………………………………………………………… ……………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ..…………………………………………………………………………………… . . .…………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ..…………………………………………………………………………………… ...…………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ……………………………………………………………………………………… ……………………………………………………………………………………… **Điểm:**..................................................(bằng chữ ………………………………)

**Đồng ý/Không đồng ý** cho sinh viên bảo vệ trước hội đồng chấm đồ án tốt nghiệp?

|  |  |
| --- | --- |
|  | Hà Nội, ngày.........tháng ........năm 2023 |
|  | CÁN BỘ - GIẢNG VIÊN HƯỚNG DẪN |
|  | (ký, họ tên) |

**NHẬN XÉT, ĐÁNH GIÁ, CHO ĐIỂM**

**(Của Người hướng dẫn)**

……………………………………………………………………………………… .……………………………………………………………………………………… ……………………………………………………………………………………… ……………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ..…………………………………………………………………………………… . . .…………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ..…………………………………………………………………………………… ...…………………………………………………………………………………… ……………………………………………………………………………………… .……………………………………………………………………………………… ……………………………………………………………………………………… ……………………………………………………………………………………… **Điểm:**..................................................(bằng chữ ………………………………)

**Đồng ý/Không đồng ý** cho sinh viên bảo vệ trước hội đồng chấm đồ án tốt nghiệp?

|  |  |
| --- | --- |
|  | Hà Nội, ngày.........tháng ........năm 2023 |
|  | CÁN BỘ - GIẢNG VIÊN PHẢN BIỆN |
|  | (ký, họ tên) |

# MỤC LỤC

[MỤC LỤC 4](#_Toc123635039)

[DANH MỤC BẢNG 6](#_Toc123635040)

[DANH MỤC HÌNH ẢNH 7](#_Toc123635041)

[DANH MỤC CÁC TỪ VÀ THUẬT NGỮ VIẾT TẮT 9](#_Toc123635042)

[LỜI MỞ ĐẦU 10](#_Toc123635043)

[CHƯƠNG 1. TỔNG QUAN VỀ GIÁM SÁT AN TOÀN THÔNG TIN 12](#_Toc123635044)

[1.1. Khái quát về giám sát an toàn thông tin 12](#_Toc123635045)

[*1.1.1.* *Khái niệm* 12](#_Toc123635046)

[*1.1.2.* *Cách thức hoạt động của giám sát an toàn thông tin* 12](#_Toc123635047)

[1.2. Các thành phần chính của hệ thống giám sát an toàn thông tin 14](#_Toc123635048)

[*1.2.1.* *Hệ thống phát hiện xâm nhập* 14](#_Toc123635049)

[*1.2.2.* *Hệ thống quản lý các sự kiện và thông tin bảo mật* 20](#_Toc123635050)

[*1.2.3.* *Cơ sở dữ liệu về thông tin tình báo về mối đe dọa* 24](#_Toc123635051)

[1.3. Kết chương 26](#_Toc123635052)

[CHƯƠNG 2. TỔNG QUAN VỀ QUẢN LÝ NHẬT KÝ 27](#_Toc123635053)

[2.1. Giới thiệu về log 27](#_Toc123635054)

[*2.1.1.* *Khái niệm về log* 27](#_Toc123635055)

[*2.1.2.* *Các loại log liên quan tới an toàn thông tin* 27](#_Toc123635056)

[2.2. Hệ thống quản lý nhật ký 31](#_Toc123635057)

[*2.2.1.* *Kiến* *trúc* 31](#_Toc123635058)

[*2.2.2.* *Chức năng* 32](#_Toc123635059)

[2.3. Kết chương 33](#_Toc123635060)

[CHƯƠNG 3. GIỚI THIỆU PHẦN MỀM QUẢN LÝ NHẬT KÝ GRAYLOG 35](#_Toc123635061)

[3.1. Khái niệm 35](#_Toc123635062)

[3.2. Các thành phần của graylog 35](#_Toc123635063)

[*3.2.1.* *Elasticsearch* 35](#_Toc123635064)

[*3.2.2.* *MongoDB* 38](#_Toc123635065)

[*3.2.3.* *Graylog server* 38](#_Toc123635066)

[*3.2.4.* *Graylog UI* 38](#_Toc123635067)

[3.3. Kiến trúc 38](#_Toc123635068)

[*3.3.1.* *Kiến trúc tối giản* 38](#_Toc123635069)

[*3.3.2.* *Kiến trúc nâng cao* 39](#_Toc123635070)

[3.4. Các tính năng chính của graylog 41](#_Toc123635071)

[*3.4.1.* *Graylog Sidecar* 41](#_Toc123635072)

[*3.4.2.* *Tìm kiếm* 41](#_Toc123635073)

[*3.4.3.* *Luồng (stream)* 42](#_Toc123635074)

[*3.4.4.* *Bảng điều khiển* 43](#_Toc123635075)

[*3.4.5.* *Cảnh báo và thông báo (Alert and notification)* 44](#_Toc123635076)

[*3.4.6.* *Pipelines* 48](#_Toc123635077)

[*3.4.7.* *Các bộ luật* 50](#_Toc123635078)

[3.5. Cấu hình graylog 51](#_Toc123635079)

[CHƯƠNG 4. Cài đặt và thử nghiệm 57](#_Toc123635080)

[4.1. Cài đặt 57](#_Toc123635081)

[*4.1.1.* *Mô hình cài đặt* 57](#_Toc123635082)

[*4.1.2.* *Giới thiệu các thành phần* 58](#_Toc123635083)

[*4.1.3.* *Cơ chế hoạt động của hệ thống* 59](#_Toc123635084)

[*4.1.4.* *Môi trường và các công cụ* 60](#_Toc123635085)

[4.2. Thử nghiệm 70](#_Toc123635086)

[KẾT LUẬN 78](#_Toc123635087)

[TÀI LIỆU THAM KHẢO 80](#_Toc123635088)

# DANH MỤC BẢNG

[Bảng 3.1. Thành phần của graylog 37](#_Toc123557612)

# DANH MỤC HÌNH ẢNH

[Hình 1.1. Mô hình cơ bản của hệ thống phát hiện xâm nhập 20](#_Toc123635140)

[Hình 1.2. Mô hình của hệ thống phát hiện xâm nhập dựa trên máy. 20](#_Toc123635141)

[Hình 1.3. Mô hình hệ thống phát hiện xâm nhập dựa trên mạng 21](#_Toc123635142)

[Hình 1.4. Mô tả chức năng của SIEM. 26](#_Toc123635143)

[Hình 1.5. Kiến trúc cơ bản của SIEM 26](#_Toc123635144)

[Hình 2.1. Nhật ký của suricata 33](#_Toc123635145)

[Hình 2.2. Nhật ký của hệ điều hành Window. 35](#_Toc123635146)

[Hình 3.1. Mô tả các shard được phân vào các nút elasticsearch 42](#_Toc123635147)

[Hình 3.2. Mô tả kiến trúc khi không có replicas shard 42](#_Toc123635148)

[Hình 3.3. Mô tả kiến trúc khi có replicas shard 43](#_Toc123635149)

[Hình 3.4. Kiến trúc tối giản của graylog 44](#_Toc123635150)

[Hình 3.5. Kiến trúc nâng cao của graylog 45](#_Toc123635151)

[Hình 3.6. Giao diện tìm kiếm của graylog 47](#_Toc123635152)

[Hình 3.7. Ví dụ minh họa về các bản tin vào luồng test sẽ được lưu trong index test 47](#_Toc123635153)

[Hình 3.8. Ví dụ minh họa về bộ luật để định tuyến bản tin vào luồng tương ứng 48](#_Toc123635154)

[Hình 3.9. Ví dụ về hiển thị nhiều đồ thị trên cùng 1 màn hình 49](#_Toc123635155)

[Hình 3.10. Ví dụ về xác định sự kiện bảo mật 49](#_Toc123635156)

[Hình 3.11. Hình ảnh minh họa đặt điều kiện cho sự kiện bằng bộ lọc 50](#_Toc123635157)

[Hình 3.12. Hình ảnh minh họa về sự tổng hợp (aggregation) 51](#_Toc123635158)

[Hình 3.13. Thông báo cảnh báo bằng email 51](#_Toc123635159)

[Hình 3.14. Nội dung của thông báo mail 52](#_Toc123635160)

[Hình 3.15. Cấu hình thông báo telegram 52](#_Toc123635161)

[Hình 3.16. Phần body gồm dữ liệu cảnh báo được gửi lên cùng request 53](#_Toc123635162)

[Hình 3.17. Cấu hình thông báo HTTP 53](#_Toc123635163)

[Hình 3.18.Cấu trúc của pipeline 54](#_Toc123635164)

[Hình 3.19. Xác định mức độ ưu tiên của giai đoạn và xác định điều kiện để thực hiện giai đoạn tiếp theo 55](#_Toc123635165)

[Hình 3.20. Trạng thái của mongodb 57](#_Toc123635166)

[Hình 3.21. Trạng thái của elasticsearch 58](#_Toc123635167)

[Hình 3.22. Nội dung file cấu hình của graylog 59](#_Toc123635168)

[Hình 3.23. Trạng thái của graylog-server 60](#_Toc123635169)

[Hình 3.24. Giao diện quản trị của graylog 60](#_Toc123635170)

[Hình 4.1. Mô hình của hệ thống 62](#_Toc123635171)

[Hình 4.2. Cấu trúc filebeat 64](#_Toc123635172)

[Hình 4.3. Trạng thái của suricata 66](#_Toc123635173)

[Hình 4.4. Xác định interface máy đang dùng 66](#_Toc123635174)

[Hình 4.5. Cấu hình interface cho suricata 67](#_Toc123635175)

[Hình 4.6. Các bộ luật được cung cấp bởi các nhà cung cấp 68](#_Toc123635176)

[Hình 4.7. Các bộ luật được bật 68](#_Toc123635177)

[Hình 4.8. Kết quả thử request 69](#_Toc123635178)

[Hình 4.9. Cảnh báo tương ứng với yêu cầu ở trên 69](#_Toc123635179)

[Hình 4.10. Cấu hình suricata ở chế độ chặn 70](#_Toc123635180)

[Hình 4.11. Luật tùy chỉnh để thử chế độ ngăn chặn 70](#_Toc123635181)

[Hình 4.12. Kết quả suricata chặn ip 8.8.8.8 70](#_Toc123635182)

[Hình 4.13. Nhật ký cảnh báo chặn ip 8.8.8.8 của suricata 71](#_Toc123635183)

[Hình 4.14. Trạng thái của máy chủ python 71](#_Toc123635184)

[Hình 4.15. Cấu hình của filebeat 72](#_Toc123635185)

[Hình 4.16. Cấu hình module suricata của filebeat 73](#_Toc123635186)

[Hình 4.17. Trạng thái của filebeat 73](#_Toc123635187)

[Hình 4.18. Tạo input beats 74](#_Toc123635188)

[Hình 4.19. Cấu hình thông báo về telegram 75](#_Toc123635189)

[Hình 4.20. Cấu hình bảng tra cứu 76](#_Toc123635190)

[Hình 4.21. Luật để tra cứu ip với bảng tra cứu 77](#_Toc123635191)

[Hình 4.22. Api check\_ip\_on\_virustotal 78](#_Toc123635192)

[Hình 4.23. Luồng hoạt động của api check\_ip\_on\_virustotal 78](#_Toc123635193)

[Hình 4.24. Api chặn ip trên suricata 78](#_Toc123635194)

[Hình 4.25. Luật để chặn ip trên suricata 79](#_Toc123635195)

[Hình 4.26. Trạng thái khi ping 8.8.8.8 lần đầu tiên 79](#_Toc123635196)

[Hình 4.27. Python server 192.168.10.170 nhận được request check ip với ip 8.8.8.8 80](#_Toc123635197)

[Hình 4.28. Python server 192.168.10.157 nhận được yêu cầu chặn ip 80](#_Toc123635198)

[Hình 4.29. Cảnh báo chặn ip 8.8.8.8 báo về telegram 80](#_Toc123635199)

[Hình 4.30. Ip 8.8.8.8 đã bị chặn 81](#_Toc123635200)

[Hình 4.31. Thêm whitelist ip 81](#_Toc123635201)

[Hình 4.32. Yêu cầu nhận được sau khi chặn ip 8.8.8.8 81](#_Toc123635202)

[Hình 4.33. Yêu cầu bỏ chặn ip 8.8.8.8 81](#_Toc123635203)

[Hình 4.34. Thông báo bỏ chặn ip 8.8.8.8 82](#_Toc123635204)

# DANH MỤC CÁC TỪ VÀ THUẬT NGỮ VIẾT TẮT

|  |  |  |
| --- | --- | --- |
| **Ký hiệu** | **Tên tiếng anh** | **Ý nghĩa tiếng việt** |
| IDS | Intrusion Detection System | Hệ thống phát hiện xâm nhập |
| IPS | Intrusion Prevention System | Hệ thống ngăn chặn xâm nhập |
| SIEM | Security Information And Event Management | Hệ thống quản lý sự kiện và thông tin bảo mật |
| HIDS | Host Intrusion Detection System | Hệ thống phát hiện xâm nhập dựa trên máy |
| NIDS | Network Intrusion Detection System | Hệ thống phát hiện xâm nhập dựa trên mạng |
| IP | Internet Protocol | Giao thức mạng |
| TLS | Transport Layer Security | Lớp truyền tải |
| OS | Operating System | Bảo mật tầng giao vận |
| SIM | Security Information Management | Hệ thống quản lý thông tin bảo mật |
| SEM | Security Event Management | Hệ thống quản lý sự kiện bảo mật |
| API | Application Protocol Interface | Giao diện lập trình ứng dụng |
| DoS | Denial-of-Service | Một loại tấn công từ chối dịch vụ |
| DDoS | Distributed Denial-of-Service | Một loại tấn công từ chối dịch vụ với nhiều máy |
| AV | Antivirus | Phần mềm chống vi-rút |
| APT | Advance Persistent Threat | Mối đe dọa nâng cao kéo dài |
| VPN | Virtual Private Network | Mạng riêng ảo |
| URL | Uniform Resource Locator | Định vị tài nguyên thống nhất |
| CPU | Central Processing Unit | Đơn vị xử lý trung tâm |
| IoC | Indicator of Compromise | Manh mối, bằng chứng của cuộc tấn công |

# LỜI MỞ ĐẦU

Trong thời đại mà công nghệ đang trong giai đoạn bùng nổ như hiện nay, mạng internet đã trở thành một thành phần không thể thiếu đối với mỗi người. Đi cùng với những lợi ích to lớn của internet là việc phải đối mặt với những nguy cơ làm mất an toàn trên không gian mạng. Hiện nay các cuộc tấn công vào các trang web, hệ thống với kỹ thuật cao diễn ra rất nhiều. Nếu không có các hệ thống phòng thủ, giám sát an toàn thông tin thì rất khó có thể tổ chức bị xâm nhập một thời gian dài mà cũng không hề hay biết. Do đó, để đối phó với những mối đe dọa, các cuộc tấn công kỹ thuật cao hiện nay thì hệ thống giám sát an toàn thông tin là một công cụ vô cùng cần thiết trong việc phát hiện sớm các cuộc tấn công. Hệ thống giám sát an toàn thông tin là tập hợp gồm nhiều các thành phần bảo mật như hệ thống phát hiện/ngăn chặn xâm nhập, hệ thống quản lý nhật ký, các công cụ hỗ trợ việc phân tích các dấu hiệu tấn công… Đồ án này tập trung và việc xây dựng hệ thống quản lý nhật ký graylog. Hệ thống này đóng vai trò lưu trữ, chuẩn hóa nhật ký, gọi đến các API để phân tích các trường trong nhật ký, đưa ra cảnh báo khi gặp điều kiện thích hợp. Ngoài ra để nâng cao khả năng phát hiện, ngăn chặn sớm các ip độc hại, một số cuộc tấn công như DoS, SSH brute-force thì trong đồ án này còn nghiên cứu tích hợp tự động hóa để phân tích ip và tự động ngăn chặn, bỏ ngăn chặn trên hệ thống ngăn chặn xâm nhập suricata. Việc phân tích có thể dựa vào dấu hiệu cuộc tấn công hoặc phân tích trên Virustotal. Đồ án bao gồm bốn chương với nội dung như sau:

Chương 1: Tổng quan về hệ thống giám sát an toàn thông tin. Chương 1 trình bày khái niệm về giám sát an toàn thông tin, cách thức hoạt động cũng như các lợi ích của giám sát an toàn thông tin, các thành phần chính của hệ thống giám sát an toàn thông tin

Chương 2: Tổng quan về quản lý nhật ký. Chương 2 trình bày về quản lý nhật ký bao gồm giới thiệu về nhật ký và hệ thống quản lý nhật ký. Chương này trình bày khái niệm về nhật ký, các loại nhật ký, kiến trúc và chức năng của hệ thống quản lý nhật ký

Chương 3: Giới thiệu phần mềm quản lý nhật ký graylog. Chương 3 trình bày về phần mềm quản lý nhật ký graylog bao gồm khái niệm, các thành phần, kiến trúc, tính năng của graylog. Ngoài ra chương này cũng trình bày về hướng dẫn cài đặt phần mềm graylog trên hệ điều hành ubuntu.

Chương 4: Cài đặt và thử nghiệm. Chương 4 mô tả về quá trình cài đặt hệ thống bao gồm thành phần: hệ thống quản lý log graylog, hệ thống phát hiện xâm nhập suricata, máy chủ API. Ngoài ra chương này có giới thiệu về các phần mềm, môi trường cho từng máy và cuối cùng chương này sẽ mô tả, cài đặt và trình bày về các kịch bản thử nghiệm liên quan tới hệ thống quản lý nhật ký graylog kết hợp với việc tự động hóa phát hiện ra các ip độc hại và tự động ngăn chặn ip độc hại đó trên hệ thống ngăn chặn xâm nhập suricata.

Do còn nhiều hạn chế về kiến thức chuyên sâu cũng như thời gian thực hiện nên đồ án không thể tránh khỏi những sai sót. Vì vậy em rất mong nhận được ý kiến đóng góp của thầy cô và các bạn sinh viên để em có thể nâng cao kiến thức phục vụ quá trình làm việc sau này.

# CHƯƠNG 1. TỔNG QUAN VỀ GIÁM SÁT AN TOÀN THÔNG TIN

*Chương 1 trình bày khái niệm về giám sát an toàn thông tin, cách thức hoạt động cũng như các lợi ích của giám sát an toàn thông tin, các thành phần chính của hệ thống giám sát an toàn thông tin*

## Khái quát về giám sát an toàn thông tin

### *Khái niệm*

Giám sát an toàn thông tin là quá trình thu thập và phân tích thông tin để phát hiện ra các hành vi nghi ngờ hoặc các hoạt động trái phép trên mạng, định nghĩa với loại dấu hiệu nào thì sẽ đưa ra cảnh báo và thực hiện các hành động để xử lý nếu cần thiết. Công việc giám sát an toàn thông tin được thực hiện thường xuyên sẽ cho chúng ta cơ hội phát hiện sự có ở giai đoạn sớm dẫn tới việc ứng phó sự có sẽ hiệu quả hơn. [1]

Giám sát an toàn thông tin không chỉ cho phép đặt cảnh báo cho các hành vi độc hại từ bên ngoài mạng, nó còn làm nổi bật các hoạt động bất thường của chính nhân viên trong nội bộ của tổ chức.

Với sự gia tăng của các mối đe dọa và các băng nhóm tội phạm tinh vi, khả năng phát hiện các sự cố trước khi chúng gây ra thiệt hại nghiêm trọng và việc hành động sớm rất là quan trọng.

### *Cách thức hoạt động của giám sát an toàn thông tin*

Trước tiên, các gói tin không xác định đi vào mạng của tổ chức, nhờ sự hỗ trợ của các giao thức, công cụ bảo mật, nó sẽ được lưu trữ trong cơ sở dữ liệu an toàn thông tin để các chuyên gia có thể phân tích các gói tin hoặc các gói tin cũng có thể được phân tích bởi các công cụ như IDS/IPS…Nếu các gói tin có hại thì các chuyên gia an toàn thông tin có thể sẽ xử lý thủ công và sẽ thực hiện các hành động tương ứng và gửi cảnh báo cho các nhóm chịu trách nhiệm tương ứng. Ngoài ra,có thể tự động hóa các quy trình để không phải can thiệp thủ công quá sâu vào từng bước trong quy trình giám sát an toàn thông tin. Để hiểu rõ hơn, có thể chia giám sát an toàn thành hai loại [2]:

* Giám sát các thiết bị đầu cuối (Endpoint monitoring)
* Giám sát mạng (Network monitoring)

#### *Giám sát các thiết bị đầu cuối*

Các thiết bị đầu cuối là những thiết bị có kết nối mạng như máy tính xách tay, máy tính để bàn, điện thoại thông minh, đồng hồ thông minh. Giám sát các thiết bị đầu cuối bao gồm việc phân tích hành vi của các thiết bị được kết nối với một mạng cụ thể. Nó sẽ giúp các nhóm gồm các chuyên gia bảo mật phát hiện ra các mối đe dọa và họ có thể thực hiện các biện pháp phòng ngừa khi có hành vi độc hại, bất thường hoặc đáng ngờ.

#### *Giám sát mạng*

Mạng là một sự kết nối giữa các thiết bị với nhau do đó chúng có thể liên lạc với nhau và chia sẻ thông tin và nội dung. Giám sát mạng đòi hỏi phải theo dõi và phân tích các gói tin, lưu lượng mạng mà từ đó nó sẽ phản hồi trên cơ sở kết quả giám sát mạng nhận được trong quá trình giám sát. Nếu các thành phần mạng không hoạt động bình thường như là thành phần nào đó bị quá tải, liên tục gặp sự cố, chậm,...Tất cả những điều đó có thể dẫn đến một số mối đe dọa mạng nhất định và khiến hệ thống dễ bị tấn công. Có nhiều công cụ chẩn đoán sẽ tiếp tục chẩn đoán các thành phần và lưu nhật ký kết quả và nếu có bất kỳ sự xáo trộn hoặc mối đe dọa nào, nó sẽ tự động thông báo ngay cho nhóm an bảo mật thông qua nhiều phương tiện. Từ đó, nhóm an toàn thông tin có thể khắc phục lỗi hoặc sự cố.

#### *Lợi ích của giám sát an toàn thông tin*

1. Giảm thiểu tối đa việc vi phạm dữ liệu

Giám sát an toàn thông tin liên tục sẽ giúp phát hiện bất kỳ mối đe dọa nào trước khi sự kiện xảy ra và tổ chức có thể tự ngăn mình khỏi các loại tấn công ảnh hưởng đến thông tin của người dùng và nhân viên của họ. Vì vậy, thực hiện giám sát an toàn thông tin liên tục sẽ hỗ trợ đắc lực cho việc giảm thiểu tối đa việc vi phạm dữ liệu của người dùng, nhân viên. [2]

1. Cải thiện thời gian phản hồi với các kiểu tấn công mạng

Hầu hết các tổ chức đều thực hiện các biện pháp bảo mật để ngăn chặn các mối đe dọa và tấn công mạng, nhưng nếu kẻ xấu bằng cách nào đó tấn công thành công tổ chức thì tổ chức phải sẵn sàng ứng phó với cuộc tấn công và khắc phục nó ngay khi được phát hiện. Bởi vì tài sản của tổ chức (hệ thống, phần mềm) phải có sẵn cho người dùng 24/7. [2]

1. Giảm thời gian chết của hệ thống

Một trong những lợi ích chính của dịch vụ giám sát an ninh mạng là giảm thời gian ngừng hoạt động cho các công ty. Một mạng đầy đủ chức năng là điều cần thiết cho các hoạt động kinh doanh hàng ngày và nhà cung cấp dịch vụ bảo mật sẽ đảm bảo rằng mọi thứ đang hoạt động ở mức tối ưu bằng cách liên tục kiểm tra mạng thường xuyên. Các dịch vụ giám sát này cũng sẽ sửa chữa và thực hiện bất kỳ cải tiến nào đối với mạng của tổ chức để giảm thiểu khả năng ngừng hoạt động có thể dẫn đến chi phí tài chính đáng kể cho tổ chức. [2]

1. Hạn chế các kiểu tấn công, mối đe dọa mới

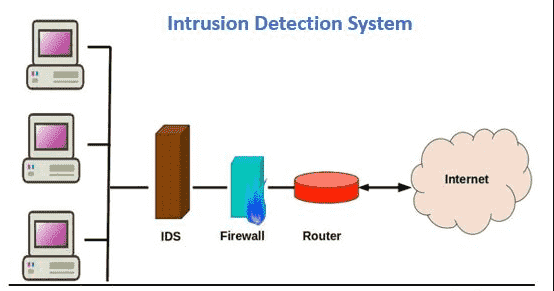
Tội phạm mạng ngày càng thông minh và sắc bén hơn. Họ luôn cố gắng vượt qua lớp phòng thủ mà bất kỳ tổ chức nào cũng thiết lập cho mạng lưới của họ. Bọn tội phạm mạng ngày càng đang đưa ra các cuộc tấn công, thủ thuật và chiến thuật mới để thực hiện hoạt động độc hại của chúng. Cách tốt nhất để giải quyết vấn đề này là liên tục giám sát an toàn thông tin bởi vì giám sát an toàn thông tin có thể giúp chúng ta phát hiện ra những dấu hiệu bất thường từ đó có những hành vi ngăn chặn kịp thời. [2]

## Các thành phần chính của hệ thống giám sát an toàn thông tin

### *Hệ thống phát hiện xâm nhập*

#### *Khái niệm*

IDS là một hệ thống giám sát để phát hiện ra các dấu hiệu, hành vi độc hại trên một hoặc nhiều các tài nguyên máy tính. Các hành vi độc hại sẽ được phát hiện bằng cách so khớp chữ ký (bao gồm 1 tập dấu hiệu được xác định trước) hoặc phát hiện dựa trên bất thường sử dụng các kĩ thuật như thống kê, học máy. IDS sẽ phản ứng với các hành vi độc hại bằng cách sinh ra cảnh báo. Các cảnh báo này có thể được báo cáo tới người quản trị hoặc được đưa lên hệ thống quản lý các sự kiện và thông tin bảo mật (SIEM). [3]

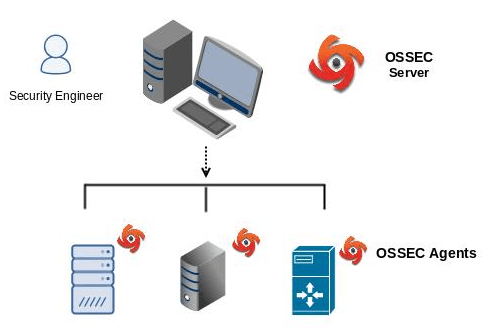


Hình 1.1. Mô hình cơ bản của hệ thống phát hiện xâm nhập

#### *Phân loại hệ thống phát hiện xâm nhập*

1. Hệ thống phát hiện xâm nhập dựa trên máy

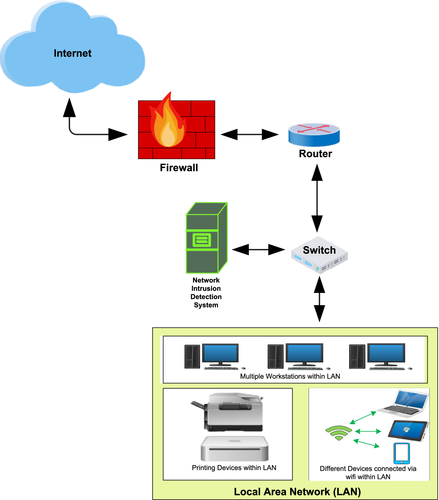
Hệ thống phát hiện xâm nhập dựa trên host (HIDS) sẽ giám sát và gửi cảnh bảo nếu phát hiện hành vi đáng nghi trên một máy đơn lẻ như máy tính cá nhân, máy chủ, điện thoại di động hoặc các thiết bị đầu cuối khác. Hầu hết các hệ thống phát hiện xâm nhập dựa trên host đều triển khai phần mềm được biết đến như là một “agent” sẽ giám sát và báo cáo về hoạt động đáng ngờ. HIDS có thể giám sát được việc truy cập tệp, sửa đổi tệp, thay đổi cấu hình, các tiến trình đang chạy, ứng dụng và các nhật ký của hệ thống. HIDS thường được cài đặt trên các máy quan trọng như máy chủ hoặc các máy có chứa dữ liệu nhạy cảm hoặc các máy ở ngoài internet. [4]



*Hình 1.2. Mô hình của hệ thống phát hiện xâm nhập dựa trên máy.*

1. Hệ thống phát hiện xâm nhập dựa trên mạng

Hệ thống phát hiện xâm nhập dựa trên mạng (NIDS) được thiết lập tại một điểm dự kiến trong mạng để kiểm tra lưu lượng từ tất cả các thiết bị trên mạng. Nó thực hiện quan sát lưu lượng truyền trên mạng và sẽ phân tích và so khớp nó với các chữ ký được thực hiện bởi các bộ luật. Khi một cuộc tấn công được xác định hoặc phát hiện ra các hành vi đáng ngờ thì nó sẽ gửi cảnh báo tới người quản trị hoặc gửi lên hệ thống quản lý sự kiện và thông tin bảo mật (SIEM). [4]



*Hình 1.3. Mô hình hệ thống phát hiện xâm nhập dựa trên mạng*

#### *Cơ chế phát hiện của hệ thống phát hiện xâm nhập*

1. Phát hiện xâm nhập dựa trên chữ ký

Phát hiện xâm nhập dựa trên chữ ký là một trong những kĩ thuật so sánh các chỉ số như địa chỉ ip,mã băm…,các dữ liệu đầu vào với chữ ký đã được xác định trước.

Phát hiện xâm nhập dựa trên chữ ký trước hết cần xây dựng cơ sở dữ liệu các chữ ký, hoặc các dấu hiệu của các loại tấn công, xâm nhập đã biết. Hầu hết các chữ ký, dấu hiệu được biểu diễn dưới dạng là các luật phát hiện (Detection rule). Bước tiếp theo là sử dụng cơ sở dữ liệu các chữ ký để giám sát hành vi của hệ thống hoặc mạng và cảnh báo nếu các chỉ số, dữ liệu đầu vào khớp với chữ ký, bộ luật. [5]

Phương pháp này cho tốc độ xử lý cao, đồng thời yêu cầu tài nguyên tính toán tương đối thấp, có khả năng phát hiện tấn công chính xác cao nếu các dấu hiệu tấn công đã tồn tại trong cơ sở dữ liệu. Tuy nhiên nhược điểm của phương pháp này chính là không có khả năng phát hiện tấn công, xâm nhập mới do các dấu hiệu tấn công chưa tồn tại trong cơ sở dữ liệu chứa các chữ ký. Hơn nữa, nó cũng đòi hỏi nhiều công sức xây dựng và cập nhật sở sở dữ liệu chữ ký, dấu hiệu của tấn công, xâm nhập.

1. Phát hiện xâm nhập dựa trên bất thường

Phát hiện xâm nhập dựa trên bất thường là kỹ thuật thực hiện việc so sánh các sự kiện bất thường với hành vi bình thường của đối tượng. Bất kể một hành động nào sai lệch so với hành vi bình thường đều được coi là xâm nhập. [5]

Phát hiện xâm nhập dựa trên bất thường dựa trên giả thiết: các hành vi tấn công, xâm nhập thường có quan hệ chặt chẽ với các hành vi bình thường. Quá trình xây dựng và triển khai một hệ thống phát hiện xâm nhập bất thường gồm 2 giai đoạn: (1) huấn luyện và (2) phát hiện. Trong giai đoạn huấn luyện, hồ sơ của đối tượng trong chế độ làm việc bình thường được xây dựng. Để thực hiện giai đoạn huấn luyện này, cần giám sát đối tượng trong một khoảng thời gian đủ dài để thu thập được đầy đủ dữ liệu mô tả các hành vi của đối tượng trong điều kiện bình thường làm dữ liệu huấn luyện. Tiếp theo, thực hiện huấn luyện dữ liệu để xây dựng mô hình phát hiện, hay hồ sơ của đối tượng. Trong giai đoạn phát hiện, thực hiện giám sát hành vi hiện tại của hệ thống và cảnh báo nếu có khác biệt rõ nét giữa hành vi hiện tại và hành vi lưu trong hồ sơ của đối tượng.

Ưu điểm của phát hiện xâm nhập dựa trên bất thường là có tiềm năng phát hiện các loại tấn công, xâm nhập mới mà không yêu cầu biết trước thông tin về chúng. Tuy nhiên, phương pháp này có tỉ lệ cảnh báo sai tương đối cao với phương pháp phát hiện dựa trên chữ ký bởi vì trong thực tế bất thường không nhất thiết là xâm nhập. Điều này làm giảm khả năng ứng dụng thực tế của phát hiện xâm nhập dựa trên bất thường. Ngoài ra, nó cũng tiêu tốn tài nguyên của hệ thống cho việc xây dựng hồ sơ của đối tượng và phân tích hành vi hiện tại.

#### *Một số phần mềm phát hiện xâm nhập*

1. Snort [6]

Snort thuộc sở hữu của Cisco Systems, là một dự án mã nguồn mở và miễn phí sử dụng. Đây là NIDS hàng đầu hiện nay. Snort có thể được cài đặt trên Windows, Linux và Unix.

Các tính năng chính:

* Miễn phí sử dụng
* Lắng nghe gói tin
* Ngôn ngữ luật phát hiện

Snort là một hệ thống phát hiện xâm nhập dựa trên mạng. Nó sẽ lắng gói tin, lưu lượng mạng để phân tích. Tuy nhiên, công cụ này có các chế độ khác và một trong số đó là phát hiện xâm nhập. Khi ở chế độ phát hiện xâm nhập, Snort áp dụng các bộ luật phát hiện của công cụ. Tuy nhiên chúng ta cần tinh chỉnh các bộ luật sao cho phù hợp với các hoạt động điển hình của mạng và giảm tỷ lệ cảnh báo sai. Chúng ta có thể viết các bộ luật của riêng mình, tuy nhiên chúng ta có thể tải xuống một gói gồm các bộ luật từ trang web.

Snort có một cộng đồng người dùng rất lớn và những người dùng đó giao tiếp với nhau thông qua một diễn đàn.

Ưu điểm:

* Hoàn toàn miễn phí và là mã nguồn mở
* Cộng đồng lớn chia sẻ các bộ quy tắc và cấu hình mới để quản trị viên hệ thống triển khai trong môi trường của họ
* Hỗ trợ lắng nghe gói tin để phân tích lưu lượng trực tiếp

Nhược điểm:

* Phần lớn khi muốn sự hỗ trợ cần phụ thuộc cá nhiều vào cộng đồng
* Có thể cần phải điều chỉnh chính sách nhiều hơn để loại bỏ các cảnh báo

1. Suricata [6]

Suricata cũng là một NIDS hoạt động ở lớp ứng dụng, mang lại khả năng hiển thị nhiều gói tin. Mặc dù các hệ thống phát hiện dựa trên chữ ký này hoạt động ở lớp ứng dụng, nhưng chúng vẫn có quyền truy cập vào các chi tiết của gói, cho phép chương trình xử lý lấy thông tin ở cấp giao thức từ các tiêu đề gói.

Các tính năng chính:

* Phân tích dữ liệu ở lớp ứng dụng
* Sử dụng miễn phí
* Tương thích với snort
* Snort cũng sử dụng các phương pháp phát hiện dựa trên sự bất thường. Ngoài dữ liệu gói, Suricata có thể kiểm tra chứng chỉ TLS, yêu cầu HTTP và giao dịch DNS
* Công cụ này cũng có thể trích xuất các phân đoạn từ các tệp ở cấp độ bit để phát hiện vi-rút

Ưu điểm:

* Thu thập dữ liệu ở các lớp ứng dụng, mang lại khả năng hiển thị duy nhất mà các sản phẩm như Snort không thể nhìn thấy
* Có thể giám sát nhiều giao thức và kiểm tra tính toàn vẹn của chứng chỉ trong TLS, HTTP.

Nhược điểm:

* Không có cộng đồng hỗ trợ rộng rãi như Snort và Zeek

1. OSSEC [7]

OSSEC là một HIDS mã nguồn mở miễn phí do Trend Micro sản xuất. Đây là một bộ xử lý dữ liệu tệp nhật ký rất hiệu quả, nhưng nó không đi kèm với giao diện người dùng. Hầu hết người dùng đặt Kibana hoặc Graylog để làm giao diện để quản lý dữ liệu,sự kiện từ OSSEC.

Các tính năng chính:

* Công cụ miễn phí
* Công cụ này sẽ tổ chức lưu trữ tệp nhật ký của bạn và bảo vệ tệp khỏi bị giả mạo.
* Các bộ quy tắc này có thể được mua miễn phí từ cộng đồng người dùng. Phần mềm OSSEC có thể được cài đặt trên Windows, Linux, Unix hoặc Mac OS. Nó theo dõi các bản ghi sự kiện của Windows và registry. Nó sẽ bảo vệ tài khoản root trên Linux, Unix và Mac OS.
* Hỗ trợ có sẵn miễn phí từ cộng đồng người dùng đang hoạt động hoặc có thể trả tiền cho Trend Micro để có gói hỗ trợ chuyên nghiệp.

Ưu điểm:

* Có thể được sử dụng trên nhiều hệ điều hành, Linux, Windows, Unix và Mac.
* Dễ dàng tùy chỉnh và có tính trực quan cao.

Nhược điểm:

* Yêu cầu các công cụ phụ như Graylog và Kibana để phân tích.
* Một số chức năng cao cấp hơn phải yêu cầu trả phí.

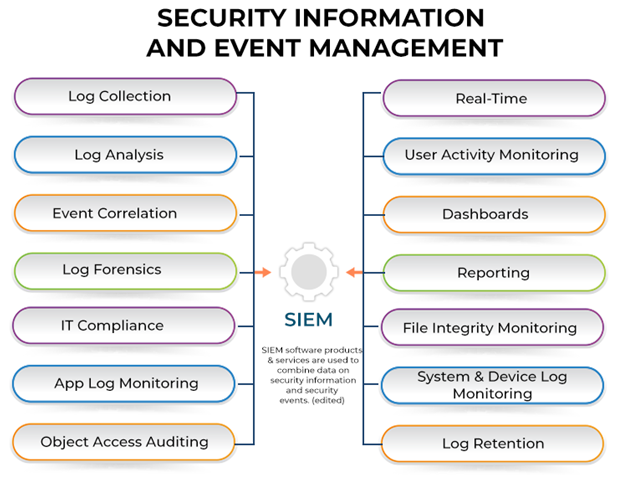
### *Hệ thống quản lý các sự kiện và thông tin bảo mật*

#### *Khái niệm*

Security information management (SIM) đề cập tới việc thu thập các tệp nhật ký và lưu trữ trong cơ sở dữ liệu tập trung. SIM được coi giống như việc quản lý nhật ký (log management). Các giải pháp SIM thường có agent chạy trên các máy được giám sát như máy tính cá nhân, máy chủ. Các agent này sẽ làm nhiệm vụ gửi nhật ký, các thông tin liên quan tới bảo mật lên máy chủ tập trung (SIM server). Tại đây, các nhà bảo mật, các nhà quản trị hệ thống có thể đăng nhập vào bảng điều khiển và xuất ra các báo cáo, đồ thị dựa trên các thông tin được gửi lên.

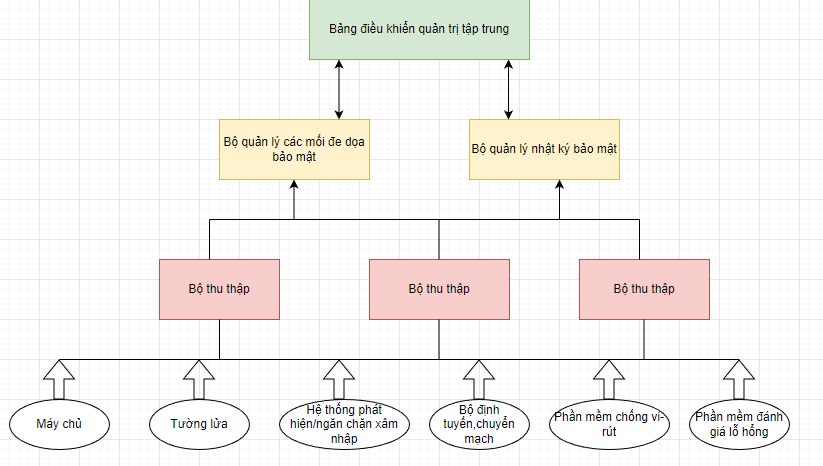
Security event management (SEM) đề cập tới việc giám sát thời gian thực, tạo ra các bộ luật để đưa ra cảnh báo, thực hiện việc cảnh báo tới các nhà bảo mật, nhà quản trị.

SIEM là sự kết hợp của SIM và SEM**.** SIEM là một giải pháp bảo mật giúp xác định các mối đe dọa và rủi ro bảo mật. Nó thu thập thông tin từ các thiết bị bảo mật khác nhau, giám sát và phân tích thông tin này, sau đó trình bày kết quả theo cách phù hợp với doanh nghiệp sử dụng nó. Nó cảnh báo các tổ chức về các cuộc tấn công tiềm ẩn, sự cố bảo mật thông tin hoặc thậm chí các vấn đề tuân thủ. Các giải pháp SIEM cung cấp khả năng giám sát và phân tích các sự kiện theo thời gian thực. Có thể kể đến một số chức năng của SIEM như theo dõi, ghi nhật ký, thu thập và quản lý dữ liệu bảo mật, báo cáo, tổng hợp dữ liệu, giám sát bảo mật, giám sát hoạt động của người dùng.



*Hình 1.4. Mô tả chức năng của SIEM*.

#### *Kiến trúc*



*Hình 1.5. Kiến trúc cơ bản của SIEM*

Bộ thu thập (collector) là thành phần để xử lý việc thu thập dữ liệu từ từ máy khách. Bộ thu thập có thể thu thập thông tin được đẩy lên bởi agent được cài trên các máy được giám sát với nhiều định dạng khác nhau, bộ thu thập sẽ chuẩn hóa những thông tin đó về dạng chuẩn và sau đó đẩy lên hệ thống tập trung. Không phải loại dữ liệu nào cũng có agent để đẩy lên bộ thu thập nên bộ thu thập cũng cần khả năng kéo dữ liệu từ client về.

Bộ để phân tích mối đe dọa (threat analysis engine) là thành phần dùng để xử lý và tương quan các sự kiện được truyền bởi bộ thu thập và nó sẽ báo cáo lên bảng điều khiển hay giao diện quản lý những mối đe dọa tìm được.

Bộ quản lý nhật ký (log manager) để lưu trữ dữ liệu, nó có thể lưu cả dữ liệu thô và dữ liệu đã được chuẩn hóa. Ngoài ra nó có khả năng nén lại dữ liệu cho việc lưu trữ lâu dài để phục vụ cho một số việc như điều tra sự cố.

Bảng điều khiển quản lý tập trung là giao diện web hiển thị các thông tin về sự kiện bảo mật cho các nhà bảo mật. Đây là một giao diện hiển thị các biểu đồ với các dữ liệu đã được trực quan hóa. Nó là giao diện mà các nhà bảo mật thao tác hàng ngày có chức năng như tìm kiếm, xem các dữ liệu nhật ký trong lịch sử, tạo bảng điều khiển (dashboard)…

#### *Quy trình hoạt động của SIEM*

1. Thu thập dữ liệu

SIEM thu thập nhật ký và sự kiện bảo mật từ rất nhiều nguồn như máy tính người dùng, máy chủ, các thiết bị mạng, hoặc các hệ thống bảo mật khác như tường lửa, phần mềm chống vi-rút, hệ thống phát hiện xâm nhập… Mỗi thiết bị sinh ra sự kiện mỗi khi có thứ gì đó xảy ra trên thiết bị. Nó sẽ thu thập các sự kiện đó và ghi vào một tệp nhật ký các sự kiện đó. SIEM có thể thu thập dữ liệu thông qua 2 cách chính:

* Thông qua một agent được cài trên máy cần được giám sát (đây là phương pháp phổ biến nhất)
* Trực tiếp kết nối với thiết bị sử dụng một giao thức mạng nào đấy hoặc sử dụng lời gọi API

Việc tiền xử lý dữ liệu có thể xảy ở bộ thu thập (collector) để đưa dữ liệu về dạng chuẩn và loại bỏ đi các thông tin dư thừa.

1. Lưu trữ dữ liệu

Các công cụ SIEM có khả năng thu thập một lượng dữ liệu khổng lồ. Dữ liệu này có thể được lưu trữ tại chỗ, trên đám mây hoặc cả hai. Quan trọng nhất, các vị trí lưu trữ cần được bảo mật chặt chẽ để tránh mất mát dữ liệu dưới bất kỳ hình thức nào. Nơi đặt dữ liệu phải dựa trên mức độ liên quan và tầm quan trọng của nó. Ví dụ, dữ liệu nóng (hot data) là những dữ liệu được sử dụng thường xuyên trong giám sát thì cần phải được lưu trong bộ nhớ có hiệu suất cao. Dữ liệu lạnh (cold data) là những dữ liệu ít được sử dụng, truy xuất nên thì không cần phải đặt trong bộ nhớ không cần hiệu suất cao như hot data mà có thể đặt ở những bộ nhớ vừa đủ có giá thành thấp hơn.

1. Tạo ra các chính sách và bộ luật

SIEM cho phép các nhân viên bảo mật có thể định nghĩa hồ sơ, trạng thái bình thường của hệ thống. Các nhân viên bảo mật có thể đặt các bộ luật và ngưỡng cảnh báo để xác định xem loại sự kiện bảo mật nào có thể được xem xét thành sự cố bảo mật. Hơn thế, SIEM có thể tận dụng học máy để tự động phát hiện ra các bất thường và tự động định nghĩa bộ luật từ dữ liệu đang có để tìm ra các sự kiện bảo mật cần được điều tra.

1. Hợp nhất và tương quan dữ liệu

SIEM cho phép tập hợp tất cả dữ liệu với nhau và cho phép tương quan giữa các nhật ký và sự kiện trên tất cả hệ thống của tổ chức. Thông báo lỗi trên máy chủ có thể liên quan đến kết nối bị chặn trên tường lửa hoặc xác thực bị sai. Nhiều điểm dữ liệu được kết hợp thành các sự kiện bảo mật có ý nghĩa và được chuyển đến các nhà phân tích bằng các thông báo hoặc bảng điều khiển (dashboard).

#### *Ứng dụng của SIEM trong an toàn thông tin*

1. Giám sát an toàn thông tin

SIEM giúp giám sát thời gian thực các hệ thống của tổ chức đối với các sự cố bảo mật. SIEM cung cấp quan điểm độc đáo về các sự cố bảo mật vì nó có quyền truy cập vào nhiều nguồn dữ liệu. Ví dụ: nó có thể kết hợp các cảnh báo từ hệ thống phát hiện xâm nhập (IDS) với thông tin từ sản phẩm chống vi-rút (AV) và nhật ký xác thực. Nó giúp các nhóm bảo mật xác định các sự cố bảo mật mà việc sử dụng từng công cụ riêng lẻ như chỉ sử dụng IDS,AV sẽ không xác định được điều đó.

1. Phát hiện các mối đe dọa nâng cao

SIEM có thể giúp phát hiện, giảm thiểu và ngăn chặn các mối đe dọa nâng cao, bao gồm:

* Lấy dữ liệu (dữ liệu nhạy cảm được chuyển bất hợp pháp ra bên ngoài tổ chức): SIEM có thể thu thập dữ liệu truyền bất thường về kích thước, tần suất hoặc trọng tải của chúng
* Các thực thể bên ngoài, bao gồm các nhóm APT: SIEM có thể phát hiện các dấu hiệu cảnh báo sớm cho thấy rằng một thực thể bên ngoài đang thực hiện một cuộc tấn công tập trung hoặc chiến dịch dài hạn chống lại tổ chức.

1. Điều tra và phản ứng sự cố

SIEM có thể giúp các nhà phân tích bảo mật xác định rằng một sự cố bảo mật đang diễn ra, phân loại sự kiện và xác định các bước để leo thang và khắc phục. Ngay cả khi sự cố đã được nhân viên an ninh biết, cần có thời gian thu thập dữ liệu để hiểu đầy đủ về cuộc tấn công và ngăn chặn nó. SIEM có thể tự động thu thập dữ liệu này và giảm đáng kể thời gian phản hồi. Khi nhân viên an ninh phát hiện ra vi phạm trong lịch sử hoặc sự cố bảo mật cần được điều tra, SIEM cung cấp dữ liệu pháp y phong phú để giúp khám phá chuỗi tiêu diệt (kill chain), các tác nhân đe dọa và giảm thiểu.

1. Báo cáo và kiểm toán tuân thủ

SIEM có thể giúp các tổ chức chứng minh với các kiểm toán viên và cơ quan quản lý rằng họ có các biện pháp bảo vệ thích hợp và các sự cố an ninh đã được biết và ngăn chặn.

### *Cơ sở dữ liệu về thông tin tình báo về mối đe dọa*

#### *Khái niệm*

Thông tin tình báo là những thông tin được thu thập từ nhiều nguồn về những cuộc tấn công, mối đe dọa hiện tại hoặc tiềm ẩn đối với tổ chức. Thông tin tình báo là những thông tin đã được phân tích, chuẩn hóa, tinh chỉnh và sau đó được sử dụng để giảm nhẹ,tối thiểu hóa các mối đe dọa an ninh mạng. Mục đích chính của thông tin tình báo về mối đe dọa là cho các tổ chức thấy các rủi ro khác nhau mà họ phải đối mặt từ các mối đe dọa bên ngoài, chẳng hạn như các mối đe dọa chưa được phát hiện (zero-day) hoặc là những chiến dịch tấn công có chủ đích sử dụng những kỹ thuật cao (APT). Thông tin tình báo về mối đe dọa bao gồm thông tin chuyên sâu và bối cảnh về các mối đe dọa cụ thể, chẳng hạn như ai đang tấn công, khả năng và động cơ của chúng cũng như là các bằng chứng, dấu hiệu của cuộc tấn công, mối đe dọa. Với thông tin này, các tổ chức có thể đưa ra quyết định đúng đắn về cách bảo vệ trước các cuộc tấn công tiềm ẩn và nguy hiểm.

Cơ sở dữ liệu về thông tin tình báo về mối đe dọa là kho lưu trữ thông tin cung cấp cho người dùng luồng dữ liệu thời gian thực liên quan đến các mối đe dọa mạng khác nhau. Dữ liệu này có thể bao gồm các các bằng chứng, dấu hiệu về sự xâm phạm như thông tin về các tác nhân đe dọa, dữ liệu về mối đe dọa, các cuộc tấn công mạng, tên miền độc hại, địa chỉ IP độc hại, mẫu phần mềm độc hại… Các nhóm bảo mật cũng có thể sử dụng thông tin này để hiểu rõ hơn về các chiến thuật, kỹ thuật và quy trình của tin tặc và cải thiện lớp phòng thủ của họ. [8]

Cơ sở dữ liệu thông tin tình báo về mối đe dọa giúp các tổ chức được thông báo về các mối đe dọa mới nhất có thể ảnh hưởng xấu đến cơ sở hạ tầng của họ. Bằng cách sử dụng cơ sở dữ liệu tình báo về mối đe dọa, các tổ chức có thể nhận được thông tin cập nhật kịp thời về các mối đe dọa mới và đang nổi lên. Cơ sở dữ liệu thông tin tình báo về mối đe dọa là các luồng dữ liệu chứa các thông tin về các mối đe dọa mạng tiềm ẩn. Thông tin này có thể được sử dụng để cải thiện tình hình bảo mật của tổ chức bằng cách giúp họ phát hiện, điều tra và đối phó với các mối đe dọa.

#### *Các nguồn phổ biến để thu thập thông tin*

Cơ sở dữ liệu thông tin tình báo về mối đe dọa thu thập thông tin từ nhiều nguồn khác nhau, bao gồm honeypots , diễn đàn các web đen, phương tiện truyền thông xã hội, phần mềm độc hại… [8]

Honeypots là các máy chủ được thiết kế đặc biệt để bị tấn công, cung cấp phương án để theo dõi và giám sát hoạt động độc hại của kẻ tấn công. Honeypots như là các hệ thống giả được sử dụng để thu hút những kẻ tấn công để thu thập các thông số liên quan tới cuộc tấn công và quy trình tấn công của chúng. Bằng cách nghiên cứu cách các hệ thống bị tấn công, các chuyên gia có thể có được cái nhìn sâu sắc về các phương pháp và công cụ được sử dụng bởi những kẻ tấn công.

Diễn đàn web đen cung cấp nơi cho tội phạm mạng mua bán thông tin nhạy cảm. Web đen là nơi có nhiều diễn đàn khác nhau, nơi người dùng có thể mua và bán dữ liệu, thông tin đăng nhập, khai thác, bộ dụng cụ khai thác, phần mềm độc hại cũng như các công cụ và dịch vụ độc hại khác. Rò rỉ dữ liệu là một mặt hàng đặc biệt phổ biến trên các diễn đàn web đen vì chúng có thể cung cấp nhiều thông tin xác thực nhạy cảm. Bộ dụng cụ khai thác và khai thác cũng thường được giao dịch trên các diễn đàn web đen, vì chúng cho phép những kẻ tấn công lợi dụng các lỗ hổng để cài đặt phần mềm độc hại hoặc truy cập vào dữ liệu nhạy cảm. DDoS as a service là một sản phẩm phổ biến khác trên các diễn đàn web tối, vì nó cho phép những kẻ tấn công thực hiện các cuộc tấn công nhằm vào mục tiêu của chúng mà không cần phải đầu tư vào cơ sở hạ tầng. Tất cả các sản phẩm và dịch vụ này đều có sẵn trên các diễn đàn web đen khiến chúng trở thành nguồn tài nguyên quý giá cho những kẻ tấn công. Do đó, bằng cách giám sát các diễn đàn này, các chuyên gia bảo mật có thể đón đầu các mối đe dọa mới nhất.

Phương tiện truyền thông xã hội là một nguồn dữ liệu quan trọng khác cho các nguồn cấp dữ liệu tình báo về mối đe dọa; bằng cách theo dõi các bài đăng và nhận xét, các chuyên gia bảo mật có thể xác định các chủ đề thịnh hành và các khu vực mục tiêu để điều tra thêm.

Các thông tin tình báo được chia sẻ bởi các tổ chức bảo mật.

## Kết chương

Chương 1 đã trình bày về tổng quan giám sát an toàn thông tin bao gồm khái niệm về giám sát an toàn thông tin, cơ chế hoạt động cũng như một số lợi ích của giám sát an toàn thông tin. Chương này giới thiệu về các thành phần cơ bản của hệ thống giám sát an toàn thông tin như hệ thống phát hiện xâm nhập, cơ sở dữ liệu chứa các IoC, SIEM.

# CHƯƠNG 2. TỔNG QUAN VỀ QUẢN LÝ NHẬT KÝ

*Chương 2 trình bày về quản lý nhật ký bao gồm giới thiệu về nhật ký và hệ thống quản lý nhật ký. Chương này trình bày khái niệm về nhật ký,các loại nhật ký,kiến trúc và chức năng của hệ thống quản lý nhật ký*

## Giới thiệu về log

### *Khái niệm về log*

Nhật ký là bản ghi các sự kiện xảy ra trong hệ thống và mạng của tổ chức. Nhật ký bao gồm các mục, mỗi mục chứa thông tin liên quan đến một sự kiện cụ thể đã xảy ra trong một hệ thống hoặc mạng. Ban đầu, nhật ký được sử dụng chủ yếu để khắc phục sự cố, nhưng hiện nay nhật ký làm nhiều chức năng trong các tổ chức, chẳng hạn như tối ưu hóa hệ thống và hiệu suất mạng, ghi lại hành động của người dùng và cung cấp dữ liệu hữu ích để điều tra các hoạt động độc hại. Nhật ký đã được tiến hóa theo thời gian để chứa thông tin liên quan đến nhiều loại sự kiện khác nhau xảy ra trong mạng và hệ thống của tổ chức. Trong một tổ chức, nhiều nhật ký chứa các bản ghi liên quan đến bảo mật, các ví dụ phổ biến về nhật ký bảo mật là nhật ký kiểm tra theo dõi các lần thử xác thực người dùng và nhật ký của thiết bị bảo mật ghi lại các cuộc tấn công có thể xảy ra. Ở trong chương này tập trung chủ yếu vào những nhật ký chứa thông tin liên quan đến các sự kiện bảo mật. Các nhật ký có thể được sinh ra liên tục hoặc cũng có thể được sinh ra theo định kỳ. [9]

### *Các loại log liên quan tới an toàn thông tin*

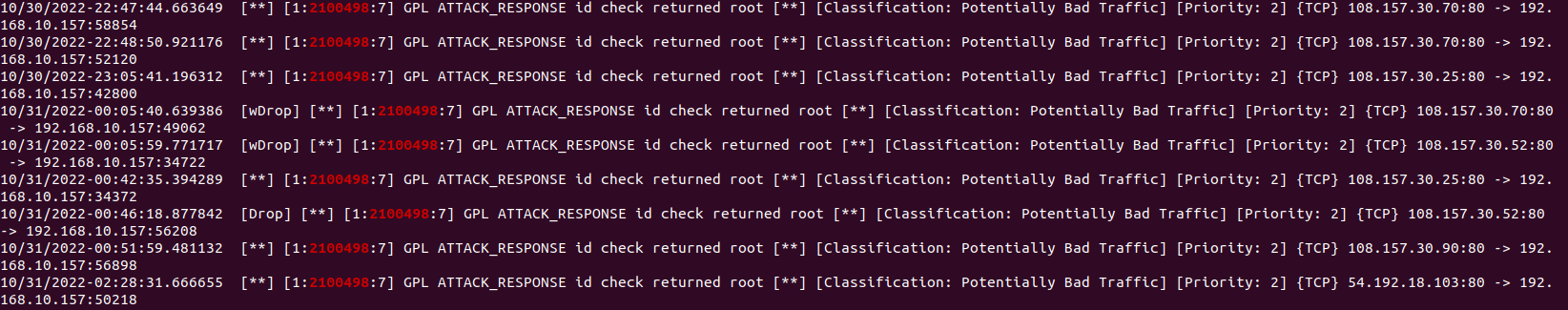
Trong phần này sẽ tập trung vào 3 loại nhật ký chính:

* Nhật ký của phần mềm bảo mật (antivirus,ids/ips…) chủ yếu chứa các thông tin liên quan tới bảo mật thông tin
* Nhật ký hệ điều hành
* Nhật ký ứng dụng

#### *Nhật ký của phần mềm bảo mật*

Hầu hết các tổ chức sử dụng một số loại phần mềm bảo mật dựa trên mạng và dựa trên máy chủ để phát hiện hoạt động độc hại, bảo vệ hệ thống và dữ liệu, đồng thời hỗ trợ việc ứng phó sự cố. Do đó, các phần mềm là một nguồn chính của dữ liệu nhật ký liên quan bảo mật. Các loại phần mềm bảo mật dựa trên mạng và máy chủ phổ biến bao gồm:

* Phần mềm chống phần mềm độc hại: Hình thức phổ biến nhất của phần mềm chống phần mềm độc hại là phần mềm chống vi-rút, thường ghi lại tất cả các mẫu phần mềm độc hại được phát hiện, những lần chống lây nhiễm tập tin, hệ thống và cách ly tập tin. Ngoài ra, phần mềm chống vi-rút cũng có thể ghi lại nhật ký khi phần mềm quét mã độc đã được chạy và khớp với chữ ký ở trong cơ sở dữ liệu hoặc khi có cập nhật phần mềm. Phần mềm chống gián điệp và các loại phần mềm chống phần mềm độc hại khác cũng là các nguồn phổ biến trong an toàn thông tin.
* Hệ thống phát hiện xâm nhập và ngăn chặn xâm nhập: Hệ thống phát hiện xâm nhập ngăn chặn xâm nhập ghi lại thông tin chi tiết về hành vi đáng ngờ và các cuộc tấn công được phát hiện, cũng như bất kỳ hành động nào mà hệ thống ngăn chặn xâm nhập được thực hiện để ngăn chặn hoạt động độc hại đang diễn ra. Một số hệ thống phát hiện xâm nhập, chẳng hạn như phần mềm kiểm tra tính toàn vẹn của tệp, chạy định kỳ thay vì liên tục.

****

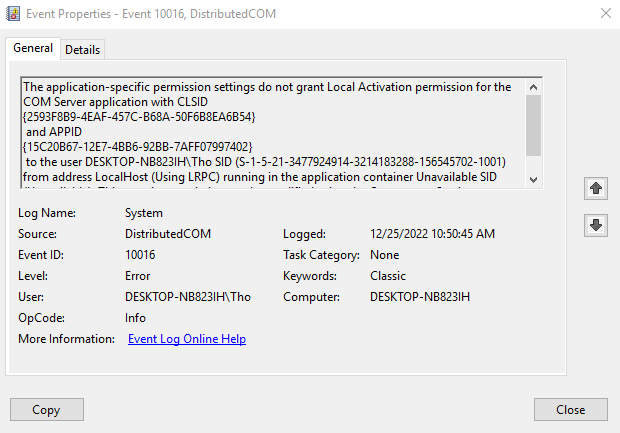
Hình 2.1. Nhật ký của suricata

* Phần mềm truy cập từ xa. Truy cập từ xa thường được bảo mật thông qua mạng riêng ảo (VPN). Các hệ thống VPN thường ghi nhật ký các lần đăng nhập thành công và thất bại, cũng như ngày và giờ mỗi người dùng kết nối và ngắt kết nối cũng như lượng dữ liệu được gửi và nhận trong mỗi phiên người dùng
* Web proxy: Web proxy là máy chủ trung gian thông qua đó các trang Web được truy cập. Web proxy thực hiện các yêu cầu trang Web thay mặt cho người dùng và chúng lưu trữ các bản sao của trang web đã truy xuất trước đó để gia tăng hiệu suất khi truy cập vào trang web. Web proxy cũng có thể được sử dụng để hạn chế các truy cập vào web và thêm một lớp bảo vệ giữa máy khách và máy chủ web. Web proxy thường lưu giữ các bản ghi của tất cả các URL được truy cập thông qua chúng
* Phần mềm quản lý lỗ hổng: phần mềm quản lý lỗ hổng, bao gồm phần mềm quản lý bản vá và phần mềm đánh giá lỗ hổng, thường ghi lại lịch sử cài đặt bản vá và trạng thái lỗ hổng của từng máy chủ, bao gồm các lỗ hổng đã biết và các bản cập nhật phần. Phần mềm quản lý lỗ hổng cũng có thể ghi lại thông tin bổ sung về cấu hình của máy chủ. Phần mềm quản lý lỗ hổng thường chạy định kỳ chứ không chạy liên tục và có khả năng tạo ra một lượng lớn log.
* Máy chủ xác thực. Máy chủ xác thực như máy chủ thư mục (active directory) thường ghi lại từng lần thử xác thực, bao gồm, tên người dùng, thành công hoặc thất bại, và ngày và thời gian.
* Bộ định tuyến: Bộ định tuyến có thể được cấu hình để cho phép hoặc chặn một số loại lưu lượng mạng nhất định dựa trên chính sách. Các bộ định tuyến chặn lưu lượng thường được cấu hình để chỉ ghi nhật ký cơ bản nhất các đặc điểm của hoạt động mà bị chặn
* Tường lửa: giống như bộ định tuyến, tường lửa cho phép hoặc chặn hoạt động dựa trên chính sách, tuy nhiên tường lửa sử dụng các phương pháp phức tạp hơn nhiều để kiểm tra lưu lượng mạng. Tường lửa cũng có thể theo dõi trạng thái lưu lượng mạng và thực hiện kiểm tra nội dung. Tường lửa có có chính sách phức tạp hơn,chi tiết hơn và tạo nhật ký hoạt động chi tiết hơn bộ định tuyến

#### *Nhật ký của hệ điều hành*

Hệ điều hành cho máy chủ, máy trạm và thiết bị mạng ghi lại nhiều thông tin liên quan đến bảo mật như là ghi lại các lần đăng nhập thành công thất bại vào thiết bị,… Nhật ký của hệ điều hành cũng có thể chứa thông tin từ phần mềm bảo mật và các ứng dụng khác chạy trên hệ thống. Nhật ký hệ điều hành rất có lợi cho việc xác định hoặc điều tra hoạt động đáng ngờ liên quan đến một máy chủ cụ thể. Sau khi hoạt động đáng ngờ được phần mềm bảo mật xác định, nhật ký hệ điều hành thường được tham khảo để biết thêm thông tin về hoạt động. Ví dụ: một thiết bị bảo mật mạng có thể phát hiện một cuộc tấn công chống lại một máy chủ cụ thể; nhật ký hệ điều hành của máy chủ đó có thể cho biết xem là có người dùng đã đăng nhập vào máy chủ tại thời điểm tấn công và xem cuộc tấn công đã thành công chưa. Các loại nhật ký có liên quan tới bảo mật của hệ điều hành thường được chia làm 2 loại:

* Sự kiện của hệ thống: các sự kiện hệ thống là các hành động vận hành được thực hiện bởi các thành phần của hệ điều hành, chẳng hạn như việc tắt hệ thống hoặc bắt đầu một dịch vụ. Thông thường, các sự kiện thất bại và thành công được ghi lại, một số hệ điều hành có thể cho phép quản trị viên chỉ định loại sự kiện nào sẽ được ghi lại. Các chi tiết được ghi lại cho mỗi sự kiện cũng rất khác nhau; mỗi sự kiện thường có thời gian xảy ra, tên sự kiện, trạng thái và mã lỗi, tên dịch vụ, và có thể tài khoản người dùng hoặc hệ thống được liên kết với một sự kiện
* Bản ghi kiểm toán: các bản ghi kiểm toán chứa thông tin sự kiện bảo mật như những lần thử xác thực thành công hoặc thất bại,các hành động truy cập tệp, thay đổi chính sách bảo mật, thay đổi tài khoản như là thêm,sửa,xóa tài khoản,gán đặc quyền cho tài khoản… Nhiều hệ điều hành cho phép chúng ta chỉ định các hành động được ghi vào bản ghi kiểm toán như các hành động xác thực,sửa đổi thông tin của tài khoản…



*Hình 2.2. Nhật ký của hệ điều hành Window.*

#### *Nhật ký ứng dụng*

Một số ứng dụng có thể tạo tệp nhật ký của riêng chúng, trong khi những ứng dụng khác sử dụng khả năng ghi nhật ký của hệ điều hành nơi mà chúng được cài đặt. Các ứng dụng sẽ có sự khác nhau đáng kể về loại thông tin mà chúng ghi lại. Một số loại thông tin được ghi lại phổ biến nhất:

* Yêu cầu của phía khách và phản hồi của máy chủ: có thể rất hữu ích trong việc xây dựng lại chuỗi các sự kiện và xác định kết quả rõ ràng của chúng. Nếu ứng dụng ghi lại sự xác thực thành công của người, có thể xác định người dùng nào đã thực hiện từng yêu cầu. Một số các ứng dụng có thể thực hiện ghi nhật ký rất chi tiết, chẳng hạn như máy chủ mail ghi lại người gửi, người nhận, tên chủ đề và tên tệp đính kèm cho mỗi mail, Máy chủ web ghi lại từng URL được yêu cầu và loại phản hồi được cung cấp bởi máy chủ. Thông tin này có thể được sử dụng để xác định hoặc điều tra các sự cố và giám sát việc sử dụng ứng dụng cho mục đích tuân thủ và kiểm toán.
* Thông tin tài khoản: Thông tin về những lần đăng nhập thành công, thất bại, những lần thay đổi tài khoản như tạo tài khoản, xóa tài khoản hoặc gán các đặc quyền cho tài khoản và thông tin về đặc quyền mà tài khoản đang sử dụng. Ngoài ra thông tin này còn có thể được sử dụng để xác định các sự kiện bảo mật như brute-force mật khẩu hoặc leo thang đặc quyền. Thông tin này còn có thể được sử dụng để xác định ai đang truy cập vào ứng dụng.
* Các thông tin liên quan tới các hành động: khởi động hoặc tắt ứng dụng,ứng dụng bị chạy lỗi hoặc là thay đổi cấu hình của ứng dụng. Nó có thể được sử dụng để xác định xem là ứng dụng đã bị xâm nhập chưa hoặc kiểm tra nguyên nhân bị lỗi ứng dụng.
* Thông tin liên quan tới giao dịch: Là những thông tin liên quan tới số giao dịch được thực hiện trong 1 khoảng thời gian,các thông tin liên quan tới kích thước trong mỗi lần giao dịch như thông tin liên quan tới kích thước của tệp,thư mục,email.

## Hệ thống quản lý nhật ký

### *Kiến* *trúc*

Một hệ thống quản lý nhật ký thường bao gồm 3 thành phần:

* Thành phần làm nhiệm vụ sinh ra nhật ký: Đây thường là các thiết bị đầu cuối như các máy trạm,máy chủ,…Hầu như các máy đều chạy ứng dụng ghi nhật ký ở phía máy khách và chạy ứng dụng để có thể chuyển nhật ký tới thành phần lưu trữ. Ngoài cách chạy ứng dụng phía khách để chuyển nhật ký tới thành phần lưu trữ thì còn có cách là máy khách cho phép máy chủ xác thực để truy xuất nhật ký hoặc bản sao chép của nhật ký phía máy khách
* Thành phần phân tích và lưu trữ nhật ký: Thành phần này bao gồm một hoặc nhiều máy chủ nhận dữ liệu nhật ký từ một hoặc nhiều máy gửi lên ở thành phần phía trên. Dữ liệu được vận chuyển tới máy chủ có thể là thời gian thực hoặc gần thời gian thực,đôi khi nó có thể được gửi lên định kỳ. Những máy chủ nhận nhật ký từ các nguồn còn được gọi là bộ thu thập hay bộ tổng hợp. Nhật ký có thể được lưu trữ trên một máy chủ hoặc nhiều máy chủ. Đây là thành phần có thể thay đổi cấu trúc. Cấu trúc đơn giản nhất là chỉ có 1 máy chủ riêng lẻ sẽ làm tất cả các chức năng như xử lý nhật ký và lưu trữ nhật ký. Cấu trúc phức tạp hơn là sẽ có nhiều máy chủ, mỗi máy chủ thực hiện một nhiệm vụ riêng biệt như một máy chủ sẽ làm nhiệm vụ thu thập, một máy chủ sẽ làm nhiệm vụ phân tích, một máy chủ sẽ làm nhiệm vụ lưu trữ nhật ký ngắn hạn, một máy chủ sẽ làm nhiệm vụ lưu trữ dài hạn.
* Thành phần giám sát nhật ký: Thành phần này chứa bảng điều khiển được sử dụng để giám sát và xem lại các dữ liệu nhật ký. Các dữ liệu này có thể được trực quan hóa dưới dạng bảng hoặc đồ thị. Ngoài ra bảng điều khiển này có thể được sử dụng để tạo ra các báo cáo. Trong một số hệ thống quản lý nhật ký, bảng điều khiển có thể được sử dụng để quản lý máy chủ nhật ký và máy khách. Ngoài ra bảng điều khiển cũng có thể được sử dụng để phân quyền cho người dùng. Nói ngắn gọn hơn thì có thể coi bảng điều khiển như một giao diện quản trị cho hệ thống quản lý nhật ký.

### *Chức năng*

Hệ thống quản lý nhật ký thường thực hiện một số chức năng hỗ trợ việc lưu trữ, phân tích và xử lý dữ liệu nhật ký

* Phân tích cú pháp nhật ký (log parsing): là việc trích xuất dữ liệu từ nhật ký sao cho các giá trị được phân tích cú pháp có thể được sử dụng làm đầu vào cho quá trình ghi nhật ký khác. Một ví dụ đơn giản về phân tích cú pháp nhật ký là đọc tệp nhật ký dưới dạng văn bản chứa 10 giá trị được phân tách bằng dấu phẩy trên mỗi dòng và trích xuất 10 giá trị từ mỗi dòng. Phân tích cú pháp nhật ký được thực hiện như một phần của nhiều chức năng ghi nhật ký khác, chẳng hạn như chuyển đổi nhật ký hoặc là trực quan hóa dữ liệu nhật ký.
* Xoay vòng nhật ký (log rotation): là việc đóng một tệp nhật ký và mở một tệp nhật ký mới khi tệp đầu tiên được coi là trọn vẹn. Xoay vòng nhật ký thường được thực hiện theo lịch trình (hàng giờ, hàng ngày, hàng tuần) hoặc khi tệp nhật ký đạt đến một kích thước nhất định. Những lợi ích chính của luân chuyển nhật ký là bảo quản các mục nhật ký và giữ cho kích thước của tệp nhật ký có thể quản lý được. Khi một tệp nhật ký được được xoay, tệp nhật ký được bảo quản có thể được nén để tiết kiệm dung lượng. Ngoài ra, trong quá trình luân chuyển nhật ký, tập lệnh thường chạy trên nhật ký lưu trữ. Ví dụ: một tập lệnh có thể phân tích nhật ký cũ để xác định hoạt động độc hại. Nhiều trình tạo nhật ký cung cấp khả năng luân chuyển nhật ký, nhiều tệp nhật ký cũng có thể được xoay vòng thông qua các tập lệnh đơn giản hoặc tiện ích của bên thứ ba trong một số trường hợp tính năng xoay vòng nhật không được cung cấp bởi trình tạo nhật ký.
* Lưu trữ nhật ký (log archival): là lưu giữ nhật ký trong một khoảng thời gian dài thường được lưu trên máy chủ hoặc thiết bị lưu trữ nhật ký chuyên dụng.
* Nén nhật ký (log compression): là việc lưu trữ tệp nhật ký theo cách giảm dung lượng lưu trữ cần thiết cho tệp mà không làm thay đổi ý nghĩa nội dung của nó. Nén nhật ký thường được được thực hiện khi nhật ký được luân chuyển hoặc lưu trữ.
* Chuẩn hóa nhật ký (log normalization): mỗi trường dữ liệu nhật ký được chuyển đổi thành một biểu diễn dữ liệu cụ thể và được phân loại một cách nhất quán. Việc chuẩn hóa dữ liệu giúp phân tích và báo cáo dễ dàng hơn khi có nhiều định dạng nhật ký đang được sử dụng. Tuy nhiên, việc chuẩn hóa có thể rất tốn tài nguyên, đặc biệt đối với các ứng dụng phức tạp.
* Báo cáo nhật ký (log reporting): hiển thị kết quả phân tích nhật ký. Báo cáo nhật ký thường được thực hiện để tóm tắt hoạt động quan trọng trong một khoảng thời gian cụ thể hoặc để ghi lại chi tiết thông tin liên quan đến một sự kiện cụ thể hoặc một loạt các sự kiện.
* Xóa nhật ký: xóa nhật thường được thực hiện để xóa dữ liệu nhật ký cũ không còn cần thiết trên hệ thống bởi vì nó không quan trọng hoặc nó đã được lưu trữ.

## Kết chương

Như vậy, chương 2 đã trình bày tổng quan về quản lý nhật bao gồm trình bày về nhật ký cũng như là hệ thống quản lý nhật ký. Đối với phần trình bày liên quan tới nhật ký, trong chương này đã trình bày về khái niệm, ba loại nhật ký cơ bản có thông tin liên quan tới an toàn thông tin là: nhật ký của phần mềm bảo mật, nhật ký của hệ điều hành, nhật ký của ứng dụng cũng như đã trình bày chi tiết về từng loại nhật ký. Đối với phần trình bày liên quan tới hệ thống quản lý nhật ký thì trong chương này đã trình bày về kiến trúc của hệ thống quản lý nhật ký, ba thành phần chính của hệ thống quản lý nhật ký bao gồm thành phần làm nhiệm vụ sinh ra nhật ký, thành phần phân tích và lưu trữ nhật ký, thành phần giám sát nhật ký. Ngoài ra đối với phần trình bày về hệ thống quản lý nhật ký này thì trong chương cũng đã trình bày về các chức năng của hệ thống quản lý nhật ký. Chương 2 đã trình bày tổng quan về quản lý nhật ký và đây là các khái niệm, vấn đề cơ bản liên quan tới quản lý log và là tiền đề cho chương 3 liên quan tới phần mềm quản lý nhật ký graylog.

# CHƯƠNG 3. GIỚI THIỆU PHẦN MỀM QUẢN LÝ NHẬT KÝ GRAYLOG

*Chương 3 trình bày về phần mềm quản lý nhật ký graylog bao gồm khái niệm,các thành phần,kiến trúc,tính năng của graylog. Ngoài ra chương này cũng trình bày về hướng dẫn cài đặt phần mềm graylog trên hệ điều hành ubuntu*

## Khái niệm

Graylog là một nền tảng quản lý nhật ký mã nguồn mở được tích hợp đầy đủ để thu thập, lập chỉ mục và phân tích cả dữ liệu có cấu trúc và không có cấu trúc từ hầu hết mọi nguồn. Nó dựa trên Elasticsearch, MongoDB và Java. Graylog có một máy chủ chính, máy chủ này nhận dữ liệu từ các máy khách và sẽ xử lý những dữ liệu đó và sẽ gửi những dữ liệu đã được xử lý lên elasticsearch để lưu trữ. Graylog có một giao diện web, hiển thị dữ liệu và cho phép làm việc với các bản ghi được tổng hợp bởi máy chủ chính. [10]

## Các thành phần của graylog

|  |  |
| --- | --- |
| Elasticsearch | - Dùng để lưu trữ các bản ghi,nhật ký  - Cung cấp khả năng tìm kiếm mạnh mẽ |
| MongoDB | - Dùng để lưu trữ các cấu hình,siêu dữ liệu |
| Graylog server | - Dùng để tiếp nhận và xử lý dữ liệu được gửi lên từ máy khách và giao tiếp với các thành phần khác như elasticsearch hay mongodb |
| Graylog UI | - Là một giao diện dùng để trực quan hóa  - Cho phép người dùng cấu hình dễ dàng hơn khi sử dụng giao diện  - Cung cấp một số chức năng như tìm kiếm, tạo bảng điều khiển, xem một số thông tin về indice… |

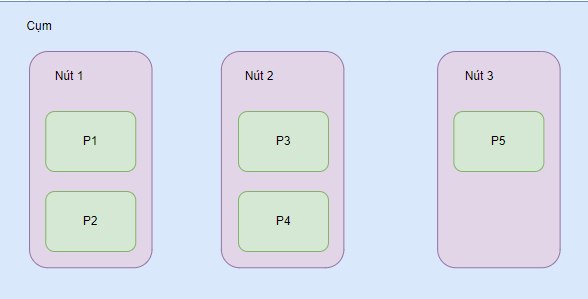
Bảng 3.1. Thành phần của graylog

### *Elasticsearch*

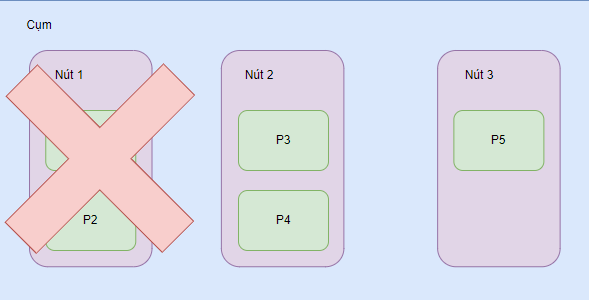
Elasticsearch là 1 thành phần quan trọng trong các thành phần của graylog vì nó được dùng để lưu trữ các dữ liệu chính. Elasticsearch là một công cụ lưu trữ, tìm kiếm và phân tích phân tán, thời gian thực có thể mở rộng theo chiều ngang và có thể giải quyết được nhiều bài toán. Elasticsearch có thể chạy trên 1 nút đơn và có thể mở rộng lên tới hàng trăm nút. Mở rộng theo chiều ngang là khả năng mở rộng hệ thống theo chiều ngang bằng cách khởi chạy nhiều thực thể cùng loại thay vì làm cho một thực thể ngày càng mạnh mẽ hơn. Mở rộng theo chiều dọc là việc nâng cấp một thực thể đơn lẻ bằng cách thêm sức mạnh xử lý (tăng số CPU, lõi CPU),bộ nhớ,dung lượng lưu trữ. Sẽ có giới hạn về khả năng mở rộng theo chiều dọc bởi vì giá thành và các yếu tố khác như là sự sẵn có của các phần cứng cao cấp hơn. Không giống như các cơ sở dữ liệu truyền thống chỉ cho phép mở rộng theo chiều dọc, Elasticsearch có thể mở rộng theo chiều ngang. Nó có thể chạy trên hàng trăm nút thay vì trên 1 nút mà giá rất đắt. Thêm một nút mới vào một cụm elasticsearch đã tồn tại dễ dàng như việc khởi chạy 1 nút trong cùng 1 mạng mà hầu như không cần phải cấu hình thêm.Với elasticsearch dữ liệu có sẵn trong vòng một giây sau khi được lập chỉ mục (lưu). Không phải tất cả hệ thống lưu trữ dữ liệu lớn đều có khả năng thời gian thực. Elasticsearch cho phép lưu trữ hàng nghìn tới hàng trăm nghìn tài liệu mỗi giây và làm nó sẵn dùng cho việc tìm kiếm gần như ngay lập tức.

Một số khái niệm cơ bản trong Elasticsearch:

* Chỉ mục (Index): giống như một thùng chứa để lưu trữ và quản lý các dữ liệu, tài liệu trong elasticsearch.
* Tài liệu (Documents): Một tài liệu bao gồm nhiều trường và là đơn vị cơ bản được lưu trữ trong elasticsearch. Ví dụ 1 tài liệu về một sản phẩm, 1 khách hàng. Mỗi trường trong tài liệu JSON là một kiểu cụ thể. Tên trường giống với tên cột trong cơ sở dữ liệu quan hệ. Giá trị của trường như là giá trị của 1 ô trong cơ sở dữ liệu quan hệ.
* Nút (Nodes): một Elasticsearch node là một máy chủ đơn lẻ của Elasticsearch cái mà có thể là 1 phần của cụm bao gồm nhiều node. Nó sẽ tham gia vào việc lưu trữ, tìm kiếm và thực hiện các hoạt động khác của elasticsearch.
* Cụm (Clusters): 1 cluster bao gồm nhiều node
* Shard: Một index bao gồm nhiều document. Shards giúp phân tán các document trong index vào các node trong cluster. Sẽ có giới hạn nếu lượng lớn dữ liệu chỉ lưu trong 1 node (bộ nhớ, dung lượng, khả năng xử lý). Nếu chúng ta chia nhỏ dữ liệu cho các node thì bộ nhớ cần sử dụng cho từng node cũng được giảm và khả năng xử lý cũng tăng lên. Tại thời điểm tạo index, chúng ta có thể chỉ định số lượng shard mà dữ liệu sẽ được chia cho index. Mỗi shard sẽ chứa 1/5 dữ liệu của index. Khi truy vấn,elasticsearch sẽ đi qua từng shard và hợp nhất kết quả.

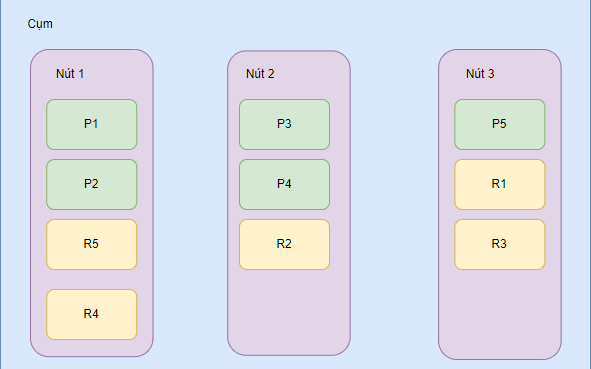
* Hình 3.1*. *Mô tả các shard được phân vào các nút elasticsearch*

* Replicas shard: Giả sử node 1 bị sập thì dữ liệu lưu ở shard 1 và shard 2 cũng sẽ bị mất. Mỗi shard trong một index có thể được cấu hình để có 0 hoặc nhiều replica shard. Replica shard là những bản sao của shard gốc và cung cấp tính sẵn dùng cao cho dữ liệu



*Hình 3.2. Mô tả kiến trúc khi không có replicas shard*

* Giả sử mỗi shard đều có 1 replica shard. Shard gốc màu xanh còn replica shard là màu xanh. Giả sử nút 1 bị sập thì vẫn có các bản sao chép ở nút 2 và nút 3.Replica shard có thể được thăng lên làm shard chính khi mà shard chính tương ứng bị sập.Các hoạt động như truy vấn,tìm kiếm đều có thể thực hiện ở replica shard



*Hình 3.3. Mô tả kiến trúc khi có replicas shard*

### *MongoDB*

MongoDB: MongoDB là cơ sở dữ liệu NoSQL hướng tài liệu được sử dụng để lưu trữ dữ liệu khối lượng lớn. Graylog sử dụng MongoDB để lưu trữ dữ liệu liên quan tới cấu hình, không phải dữ liệu nhật ký của bạn. Chỉ siêu dữ liệu được lưu trữ, chẳng hạn như thông tin người dùng hoặc cấu hình luồng (stream), đầu vào (inputs), các bộ trích xuất để chuẩn hóa (extractor), các bộ luật, …. Không có dữ liệu nhật ký nào được lưu trữ trong MongoDB. Đây là lý do tại sao MongoDB không có tác động lớn đến hệ thống và sẽ không phải lo lắng quá nhiều về việc mở rộng quy mô nó.

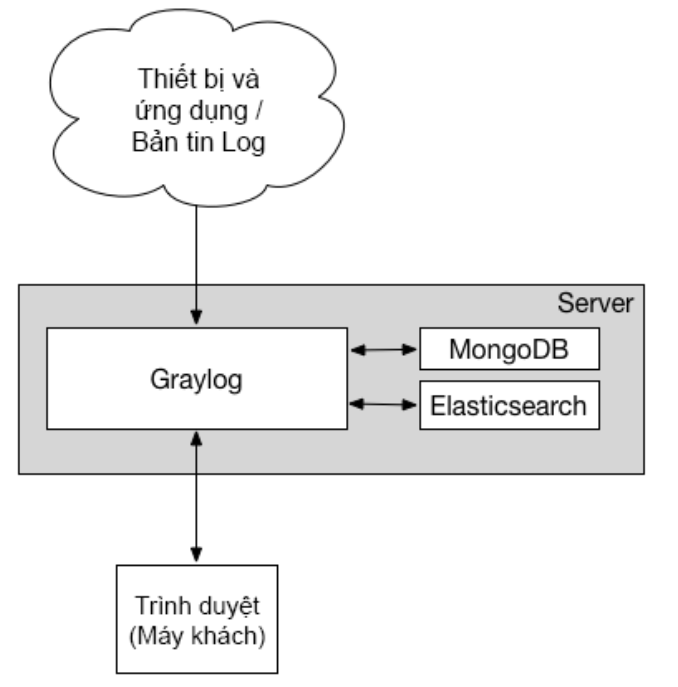
### *Graylog server*

Graylog server là một thành phần chịu trách nhiệm nhận dữ liệu được gửi lên và chuẩn hóa,xử lý nó. Ngoài ra graylog server còn có mục đích là tích hợp và giao tiếp với các thành phần khác như MongoDB hay Elasticsearch.

### *Graylog UI*

Graylog UI: cung cấp một giao diện trực quan. Nó cho phép người sử dụng có thể xem dữ liệu được đẩy lên, có thể cấu hình dễ dàng các đầu vào (input), pipeline, alert… Nó còn cho phép người sử dụng tìm kiếm, phân tích dữ liệu nhanh chóng và còn cho phép người dùng tạo các dashboard gồm các biểu đồ để dễ dàng cho việc quan sát, phân tích

1. **Kiến trúc**
2. ***Kiến trúc tối giản***

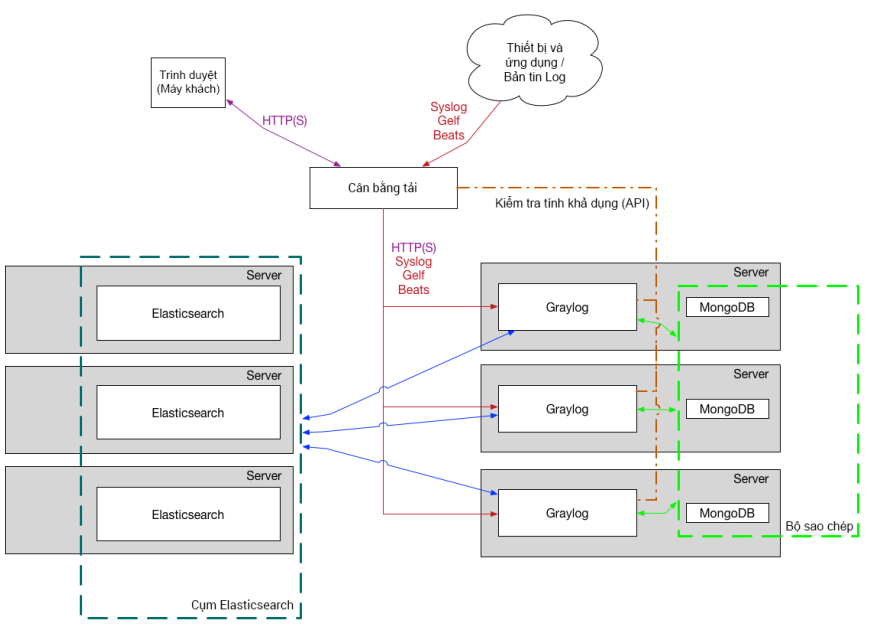


*Hình 3.4. Kiến trúc tối giản của graylog*

Đây là kiến trúc graylog tối thiểu có thể được sử dụng cho các hệ thống nhỏ, không quan trọng hoặc thử nghiệm. Không có thành phần nào là dư thừa

Khi nhật ký được chuyển đến thông qua đầu vào (input), Graylog sẽ sử dụng các cài đặt mà nó đã lưu trữ trong MongoDB (đầu vào, luồng, các bộ trích xuất, các bộ luật, đường ống…) để tìm ra cách xử lý, chuẩn hóa nhật ký, bản ghi. Sau khi xử lý xong các nhật ký, bản ghi (chuẩn hóa ra thành các trường dữ liệu và các tính toán,xử lý khác) thì nó sẽ gửi đến elasticsearch để lưu trữ dữ liệu. Việc tạo dashboard hay truy vấn dựa trên các dữ liệu nhật ký sẽ được gửi từ graylog tới elasticsearch và kết quả sẽ được trả về và hiển thị trên giao diện graylog.

1. ***Kiến trúc nâng cao***



*Hình 3.5. Kiến trúc nâng cao của graylog*

Trong kiến trúc tối giản, các thành phần như MongoDB, Graylog server hay Elasticsearch đều được đặt chung trong một server. Đối với môi trường lớn hơn, yêu cầu khả năng mở rộng, tính khả dụng cao, khả năng chịu lỗi thì graylog có thể được triển khai với mô hình multi-node phức tạp hơn. Khi dữ liệu trở nên lớn hơn thì việc có nhiều nút sẽ chia sẻ công việc cho các nút sẽ giúp giảm gánh nặng lưu trữ hay gánh nặng xử lý khi chỉ có 1 nút đơn lẻ. Ngoài ra khi có nhiều nút chạy song song thì tốc độ xử lý công việc như tìm kiếm, chuẩn xóa sẽ được tăng lên.

Kiến trúc nâng cao thường được áp dụng cho môi trường thực, doanh nghiệp lớn. Nó có nhiều nút Graylog đằng sau bộ cân bằng tải làm nhiệm vụ phân phối xử lý vào các nút. Bộ cân bằng tải có thể ping các nút Graylog qua HTTP trên Graylog REST API để kiểm tra xem chúng còn sống hay không và đưa các nút chết ra khỏi cụm.

Có 1 số quy tắc chung khi mở rộng tài nguyên cho Graylog:

* Các nút Graylog nên tập trung vào CPU.
* Các nút Elasticsearch nên có càng nhiều RAM càng tốt và chọn ổ cứng chạy nhanh nhất có thể. Mọi thứ phụ thuộc vào tốc độ I/O ở đây
* MongoDB chủ yếu lưu trữ thông tin về cấu hình và metadata nên không cần quá nhiều tài nguyên.

Các bản tin, nhật ký chỉ được lưu trữ vào trong elasticsearch. Nếu bị mất dữ liệu trong cụm elasticsearch thì dữ liệu gần như bị mất hoàn toàn trừ khi đã tạo bản sao lưu từ trước đó.

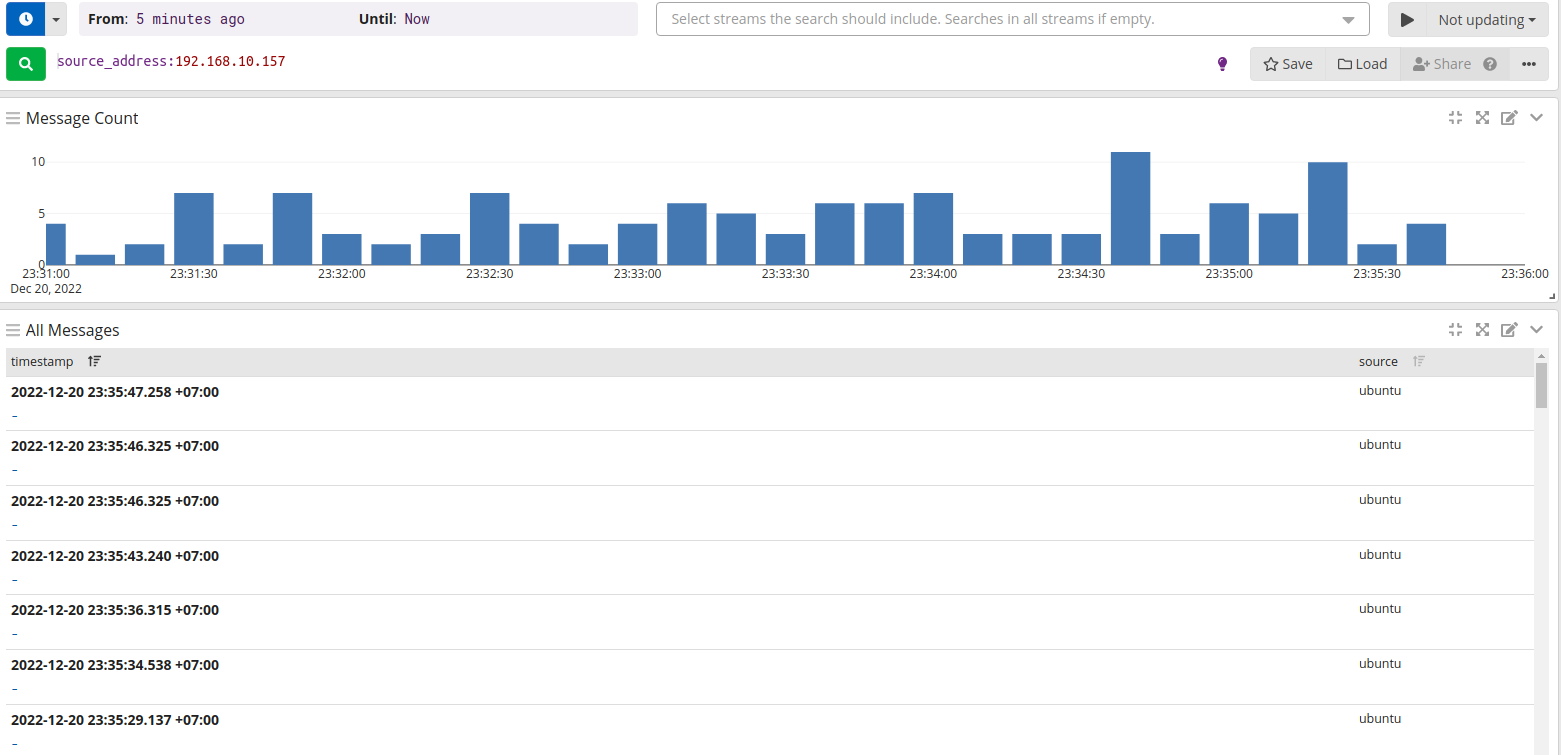
1. **Các tính năng chính của graylog**
2. ***Graylog Sidecar***

Graylog Sidecar là một hệ thống quản lý cấu hình cho các bộ thu thập nhật ký khác nhau. Các nút Graylog hoạt động như một trung tâm tập trung chứa các cấu hình của bộ thu thập nhật ký. Trên các thiết bị / máy chủ tạo ra bản ghi nhật ký được hỗ trợ, Sidecar có thể chạy như một dịch vụ trên máy Windows hoặc tiến trình nền (daemon) trên máy Linux.

Cấu hình bộ thu thập nhật ký được quản lý tập trung thông qua giao diện web Graylog. Theo định kỳ, tiến trình nền (daemon) Sidecar sẽ tìm nạp tất cả các cấu hình có liên quan cho mục tiêu, sử dụng REST API. Trong lần chạy đầu tiên hoặc khi phát hiện thay đổi cấu hình, Sidecar sẽ tạo các tệp cấu hình backend có liên quan. Sau đó, nó sẽ bắt đầu hoặc khởi động lại, các bộ thu thập nhật ký được cấu hình lại.

1. ***Tìm kiếm***

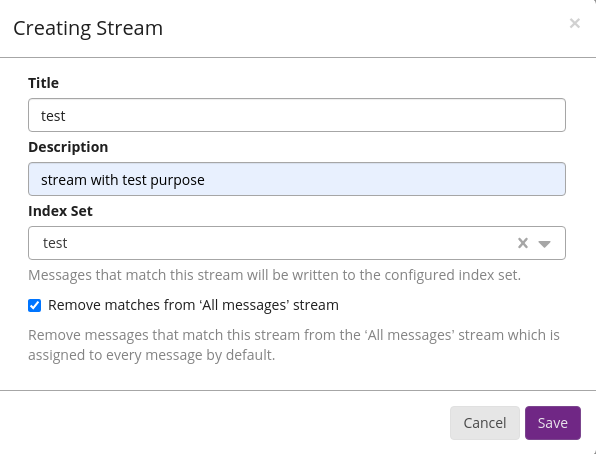
Trang Graylog Search là giao diện được sử dụng để tìm kiếm nhật ký trực tiếp. Graylog sử dụng cú pháp đơn giản, rất giống với Lucene. Phạm vi thời gian tương đối hoặc tuyệt đối có thể được chỉ định. Người dùng có thể cấu hình chế độ xem của riêng họ và có thể chọn xem dữ liệu tóm tắt hoặc dữ liệu đầy đủ từ các thông báo sự kiện. Chúng ta có thể thực hiện một tìm kiếm (truy vấn) và trực quan hóa kết quả với nhiều tiện ích con. Mọi tìm kiếm đều có thể được lưu hoặc xuất dưới dạng dashboard. Các tìm kiếm đã lưu cho phép dễ dàng sử dụng lại các cấu hình tìm kiếm cụ thể.

****

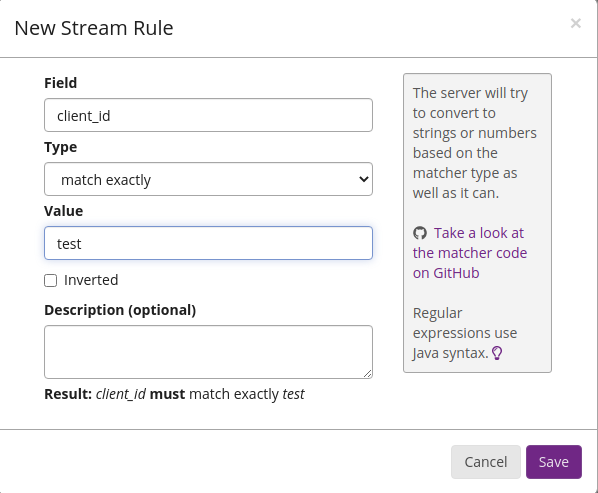
*Hình 3.6. Giao diện tìm kiếm của graylog*

1. ***Luồng (stream)***

Các luồng Graylog (streams) là một cơ chế định tuyến các nhật ký, bản ghi thành các danh mục trong thời gian thực trong khi chúng đang được xử lý.Các luồng được sử dụng để định tuyến dữ liệu lưu trữ vào chỉ mục (index) nhất định.

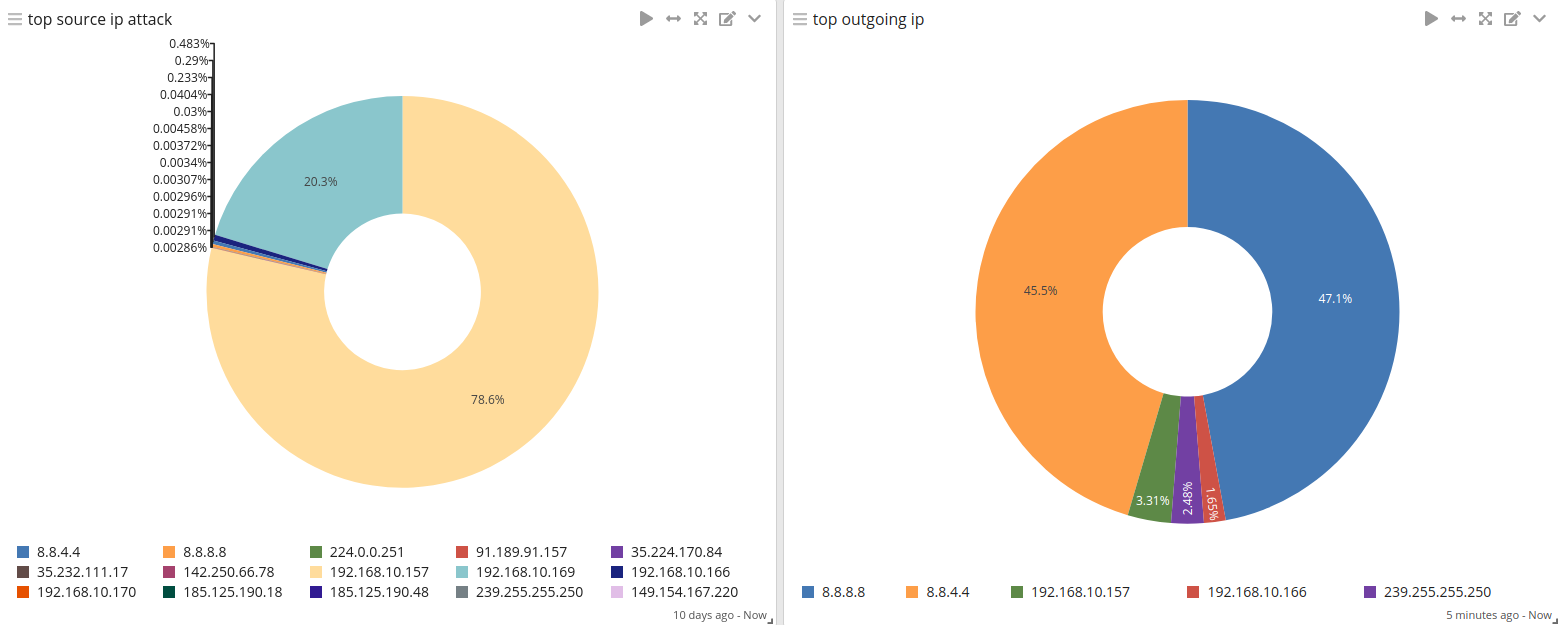
*Hình 3.7. Ví dụ minh họa về các bản tin vào luồng test sẽ được lưu trong index test*

Có thể xác định các quy tắc trong Graylog để định tuyến các thông báo vào các luồng nhất định. Giả sử có ba bản ghi được gửi tới graylog. Dưới đây là ví dụ định nghĩa quy tắc là nếu có trường client\_id = “test” trong nhật ký thì sẽ định tuyến bản tin vào luồng “test”

 *Hình 3.8. Ví dụ minh họa về bộ luật để định tuyến bản tin vào luồng tương ứng*

1. ***Bảng điều khiển***

Sử dụng bảng điều khiển (dashboard) cho phép tạo các bảng biểu đồ thị trực quan hóa cho các tìm kiếm được xác định trước trên dữ liệu, vì vậy khi cần xem thông tin quan trọng một cách trực quan chỉ cần một cú nhấp chuột. Bảng điều khiển cũng cho phép tạo nhiều tab cho các trường hợp sử dụng khác nhau, hiển thị kết quả ở chế độ toàn màn và chia sẻ kết quả với người khác. Ngoài ra chúng ta cũng có thể hiển thị nhiều đồ thị,bảng biểu của nhiều trường hợp khác nhau trên cùng 1 màn hình



*Hình 3.9. Ví dụ về hiển thị nhiều đồ thị trên cùng 1 màn hình*

1. ***Cảnh báo và thông báo (Alert and notification)***

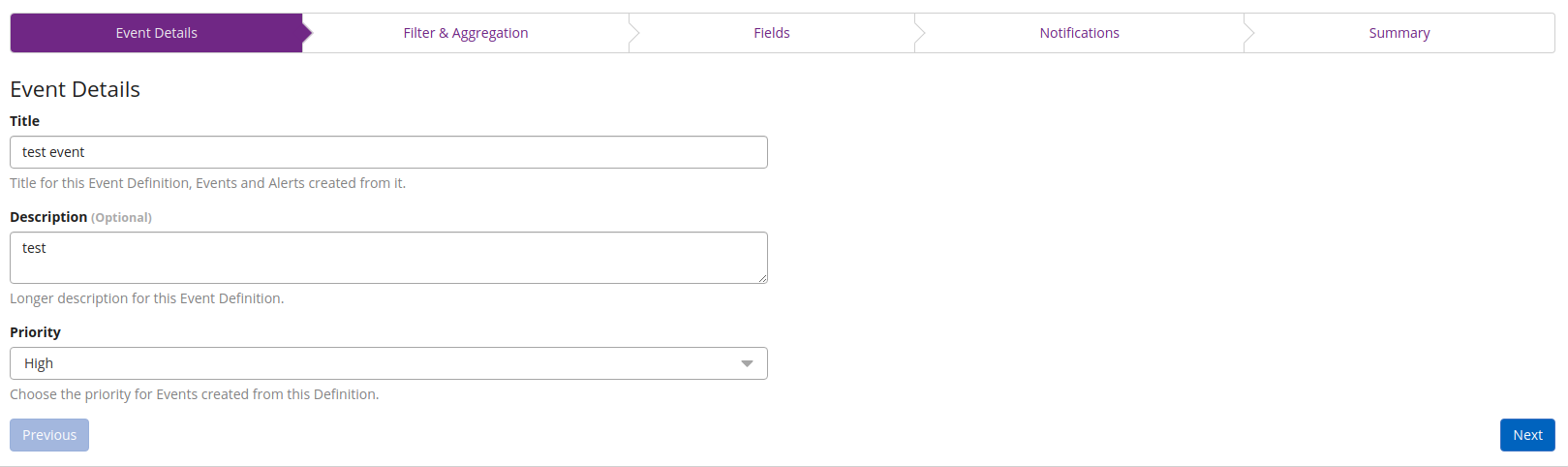
Cảnh báo được tạo ra bằng cách định nghĩa các sự kiện an toàn thông tin cái mà sẽ khớp với một số điều kiện. Khi mà một tập điều kiện được khớp, nó sẽ được lưu trữ dưới dạng sự kiện và có thể được sử dụng để kích hoạt thông báo

Quy trình để tạo ra cảnh báo trong graylog:

* Định nghĩa sự kiện
* Đặt điều kiện cho sự kiện
* Thêm thông báo (có thể có hoặc không)

1. *Định nghĩa sự kiện*

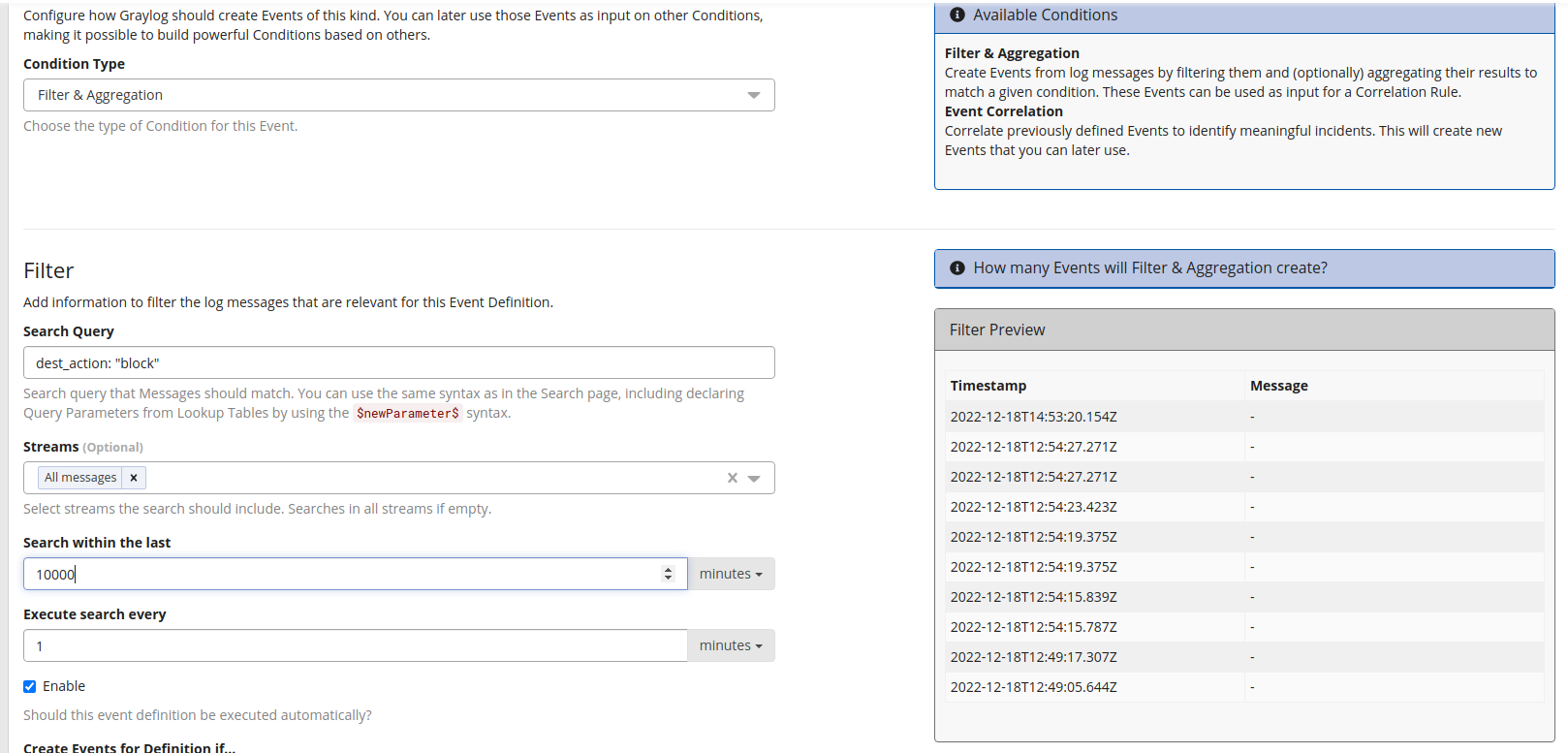
Trong bước này cần đặt tên sự kiện, thêm mô tả và chọn mức độ ưu tiên. Mức độ ưu tiên của một sự kiện là một phân loại cho mục đích của người dùng.



*Hình 3.10. Ví dụ về xác định sự kiện bảo mật*

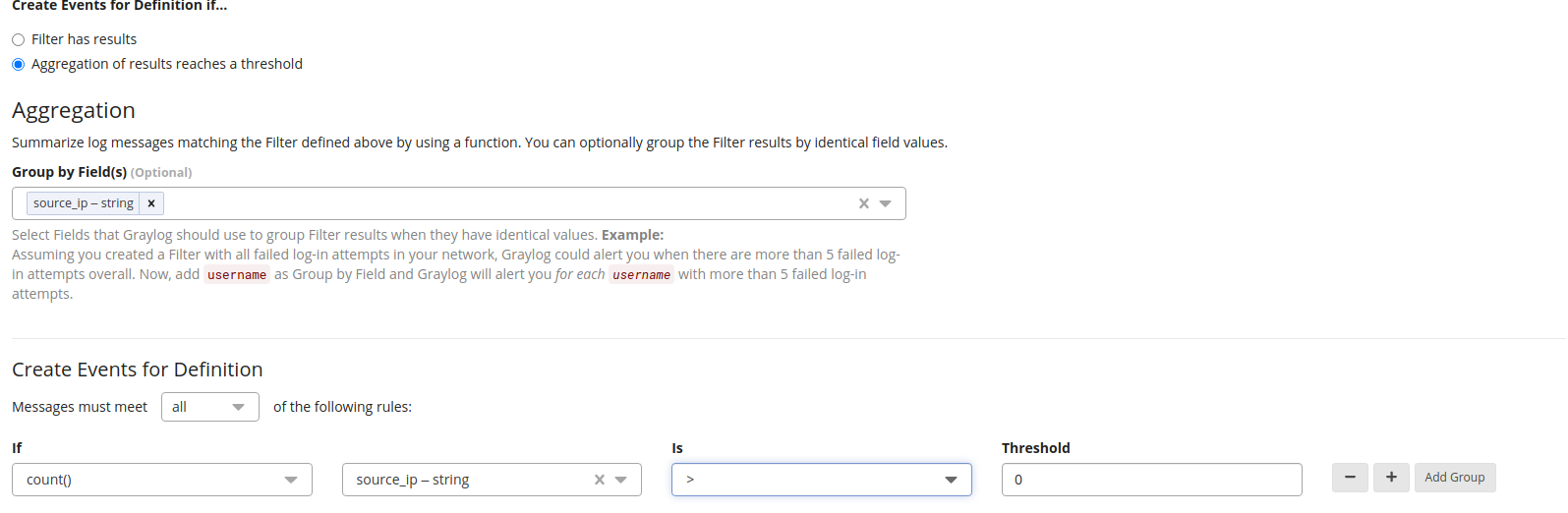
1. *Đặt điều kiện cho sự kiện*

Bằng cách kết hợp bộ lọc và sự tổng hợp, chúng ta có thể mô tả cụ thể các tiêu chí của sự kiện. Chúng ta có thể xác định bộ lọc bằng cách sử dụng truy vấn tìm kiếm theo cú pháp giống như trang tìm kiếm. Chọn một luồng để truy vấn các bản tin trên luồng đó. Xác định khoảng thời gian mà bộ lọc sẽ tìm kiếm ngược các bản tin trong khoảng thời gian đó để khớp với các thông báo. Việc tìm kiếm sẽ được thực hiện trong khoảng thời gian nhất định. Nếu dữ liệu vào khớp với bộ lọc, một sự kiện có thể được tạo. Nếu các bộ lọc khớp với các bản tin xuất hiện trong khoảng thời gian chúng ta truy vấn thì kết quả xem trước của bộ lọc sẽ được hiển thị ở bên tay phải



*Hình 3.11. Hình ảnh minh họa đặt điều kiện cho sự kiện bằng bộ lọc*

Sự tổng hợp (aggregation) có thể chạy một phép toán (tổng,trung bình cộng,lớn nhất, nhỏ nhất) trên trường giá trị số. Ngoài ra chúng ta có thể nhóm theo trường. Ví dụ chúng ta tạo bộ lọc với những cần cố gắng đăng nhập nhưng thất bại vào mạng của chúng ta, graylog có thể sẽ cảnh báo khi có 5 lần đăng nhập thất bại tuy nhiên 5 lần này có thể không phải là của một người dẫn tới kết quả cảnh báo sẽ không chính xác. Nếu chúng ta thêm trường username vào Group by Field thì graylog sẽ cảnh báo mỗi khi với một username xác định đăng nhập thất bại 5 lần.



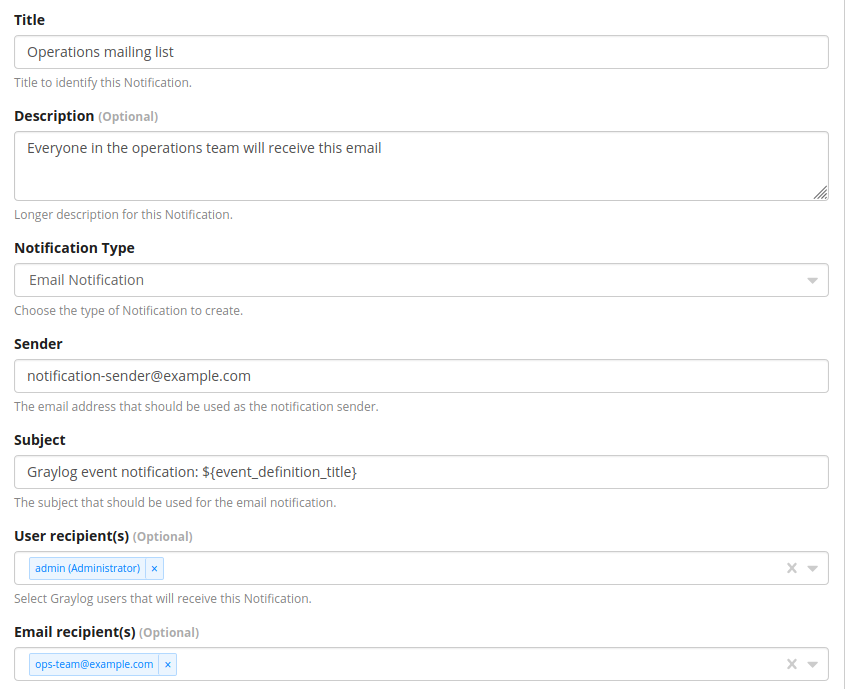
*Hình 3.12. Hình ảnh minh họa về sự tổng hợp (aggregation)*

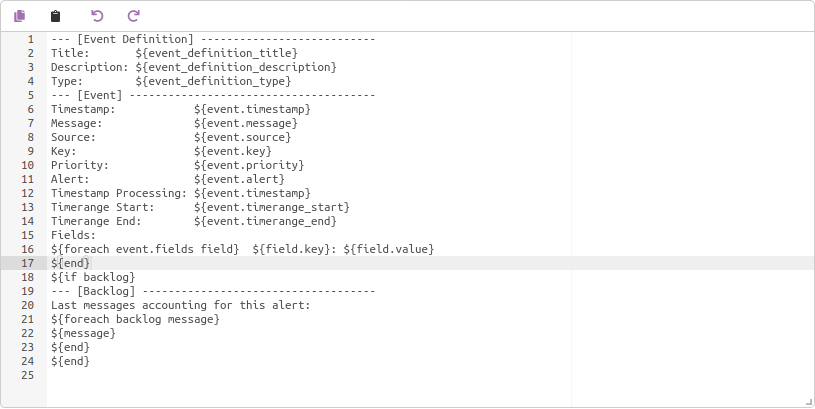
1. *Tạo thông báo*

Sau khi xác định các điều kiện để kích hoạt cảnh báo sau đó có thể đính kèm thông báo. Bằng cách đính kèm thông báo vào sự kiện hoặc nhóm sự kiện, chúng ta sẽ có thể xác định được mà cách thông tin từ graylog sẽ được chuyển đi có thể là qua các phương tiện truyền thông như email,telegram hoặc qua HTTP Request

Một số loại cảnh báo có thể tạo:

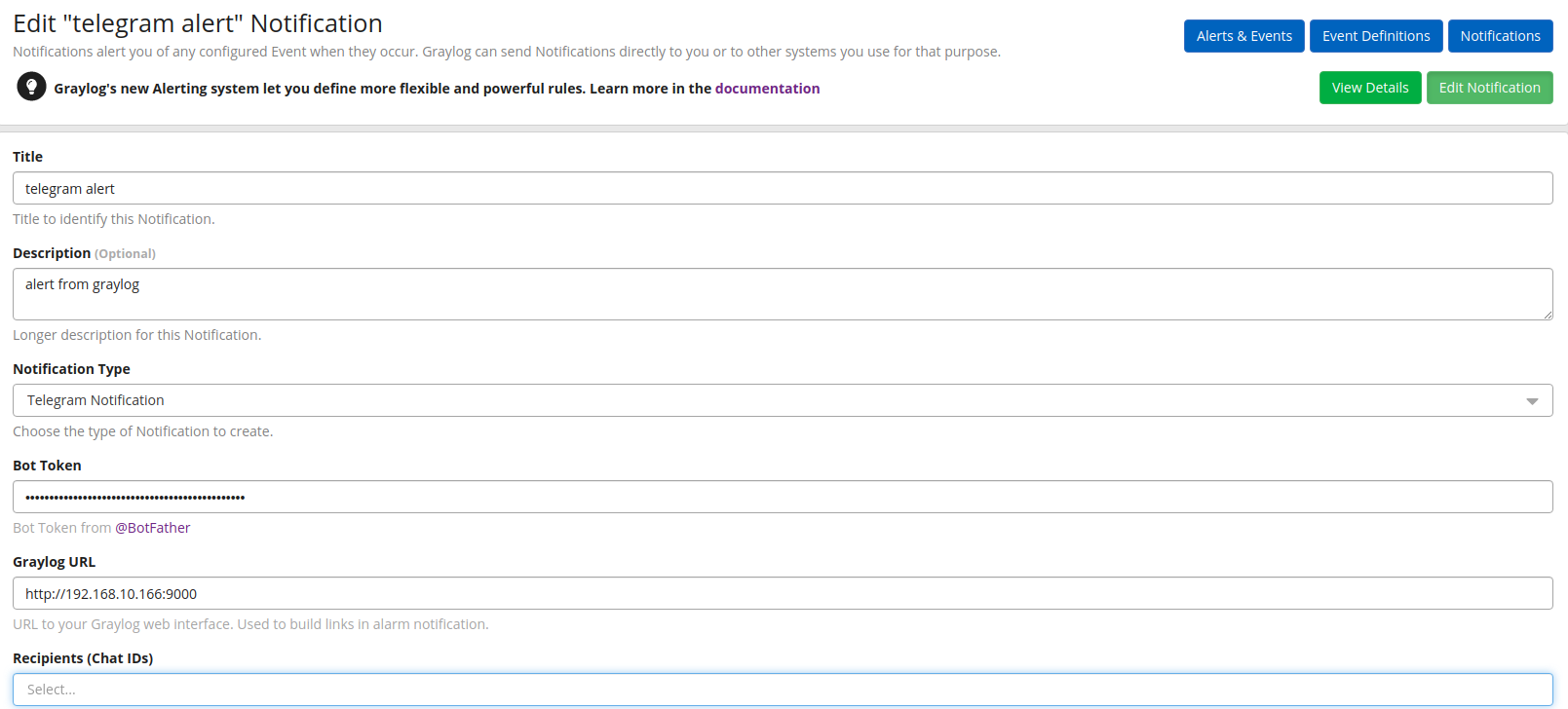
* Thông báo email: có thể được sử dụng để gửi thư tới người nhận cảnh báo khi mà một số điều kiện được kích hoạt

*Hình 3.13. Thông báo cảnh báo bằng email*



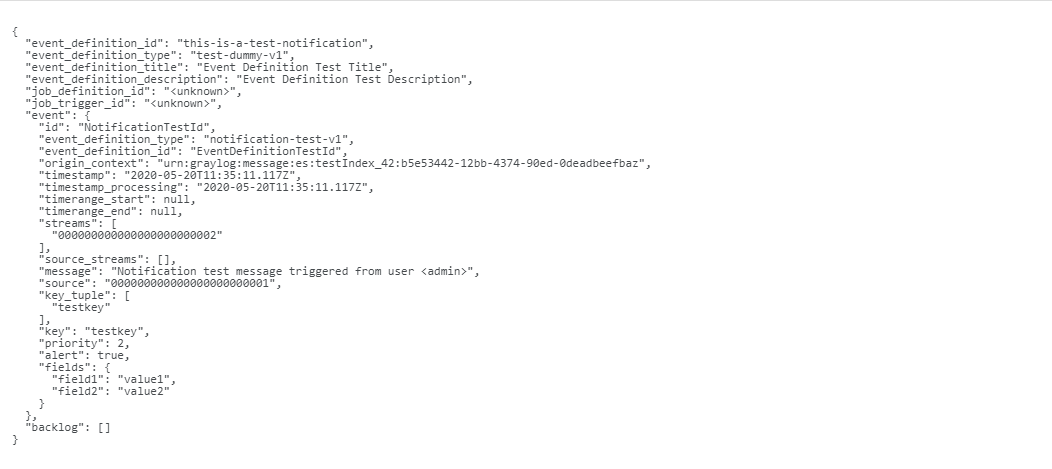
*Hình 3.14. Nội dung của thông báo mail*

* Thông báo telegram: Với thông báo telegram chưa có sẵn trong graylog mà phải tải plugin này về. Thông báo telegram cho phép gửi thông báo từ graylog về 1 group nào đó trên telegram với sự trợ giúp của con bot

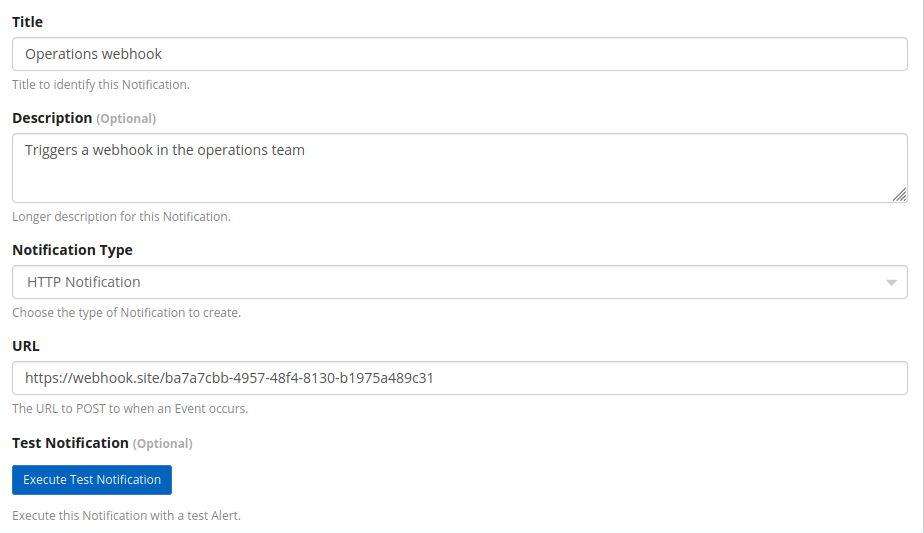


*Hình 3.15. Cấu hình thông báo telegram*

* Thông báo HTTP: Thông báo HTTP cho phép cấu hình một điểm cuối (URL) sẽ được gọi khi mà một cảnh báo trên graylog được kích hoạt. Graylog sẽ gửi yêu cầu POST cùng những thông tin về cảnh báo tới URL được cấu hình. Phần thân (body) của yêu cầu ở dưới dạng JSON như ảnh dưới



*Hình 3.16. Phần body gồm dữ liệu cảnh báo được gửi lên cùng request*



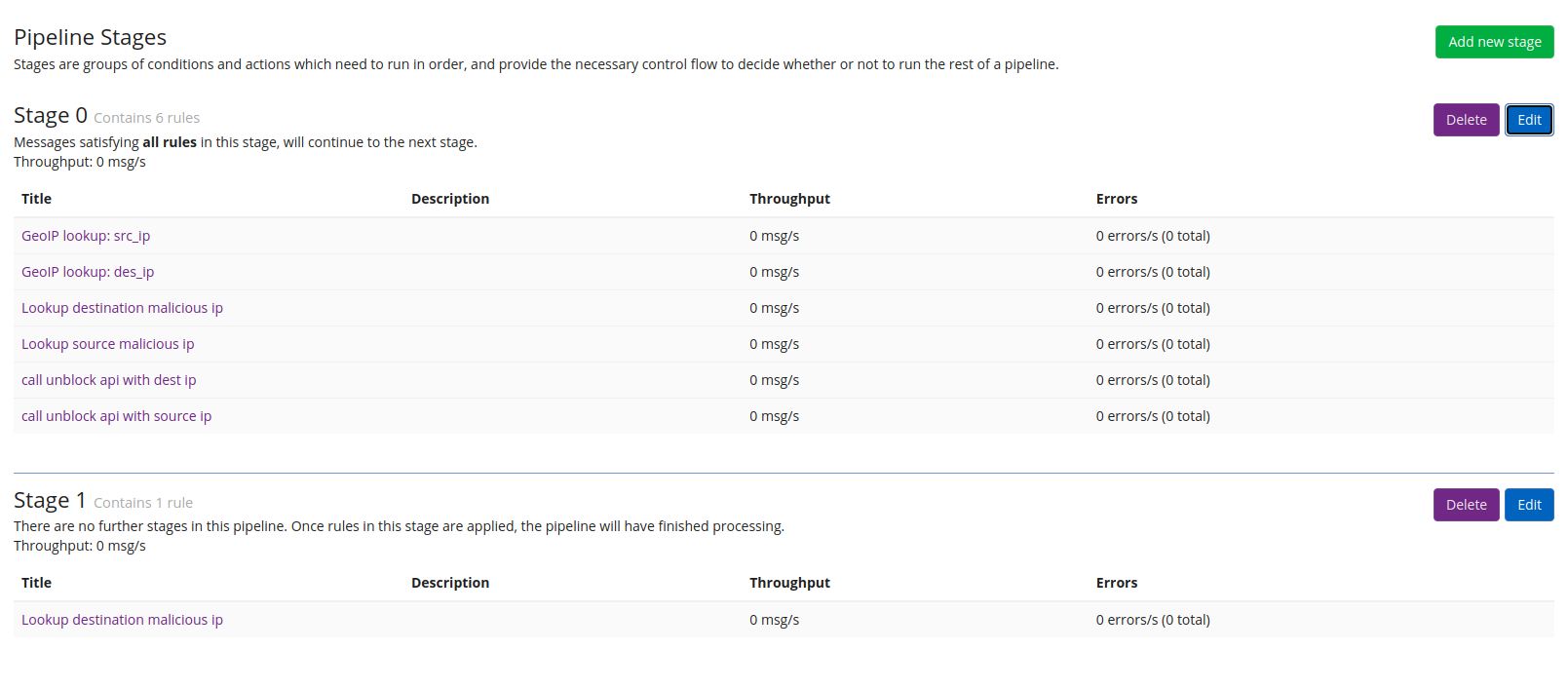
*Hình 3.17. Cấu hình thông báo HTTP*

1. ***Pipelines***
   * + 1. *Khái niệm*

Pipelines là một tập hợp các bước xử lý được áp dụng vào các bản tin. Pipelines chứa một tập bộ luật và có thể được kết nối tới một hay nhiều luồng cho phép kiểm soát chi tiết quá trình xử lý được áp dụng cho các bản tin. Các bộ luật xử lý chỉ đơn thuần là các điều kiện và theo sau đó là các hành động và bản thân chúng không có luồng kiểm soát. Vì thế sinh ra thêm khái niệm là giai đoạn (stage)

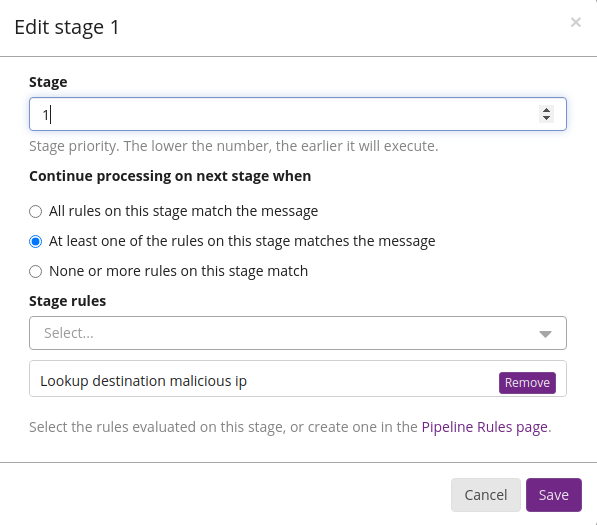
Có thể nghĩ đơn giản stage như là một tập các điều kiện và hành động được chạy theo thứ tự. Tất cả giai đoạn có độ ưu tiên giống nhau sẽ chạy đồng thời. Các giai đoạn (stage) cung cấp luồng điều khiển cần thiết để quyết định xem là có chạy các giai đoạn tiếp theo trong 1 pipeline.

* + - 1. *Cấu trúc của pipeline*



*Hình 3.18.Cấu trúc của pipeline*

Một pipeline có thể bao gồm nhiều giai đoạn được thực hiện theo mức độ ưu tiên được xác định theo những con số chứ không phụ thuộc vào thứ tự các giai đoạn được xác định. Ví dụ các bộ luật trong giai đoạn 0 sẽ được thực hiện trước các bộ luật ở trong giai đoạn 1 chứ không phụ thuộc vào giai đoạn nào được đặt trước.



*Hình 3.19. Xác định mức độ ưu tiên của giai đoạn và xác định điều kiện để thực hiện giai đoạn tiếp theo*

Trong mỗi giai đoạn có nhiều bộ luật và các bộ luật này sẽ được thực hiện đồng thời. Ở mỗi giai đoạn chúng ta có thể cấu hình điều kiện để có thể chuyển sang giai đoạn tiếp theo.Có một số điều kiện như, tất cả các bộ luật đều được thực hiện thì chuyển tới giai đoạn sau hoặc chỉ là một trong các bộ luật được thực hiện thì cũng sẽ được chuyển sang giai đoạn sau

1. ***Các bộ luật***
2. *Khái niệm*

Các bộ luật là nền tảng của pipeline. Chúng chứa logic về cách thay đổi, làm giàu, định tuyến hoặc loại bỏ bản tin. Graylog hỗ trợ một ngôn ngữ luật nhỏ để thực hiện xử lý logic. Graylog cũng hỗ trợ nhiều hàm được viết sẵn với nhiều chức năng như chuyển đổi dữ liệu, thao tác với chuỗi, truy xuất dữ liệu sử dụng bảng tra cứu và rất nhiều các hàm khác.

1. *Cấu trúc của các bộ luật*

rule “rule name”

when

condition

then

action

end

1. Điều kiện (Condition)

Trong các bộ luật của graylog, mệnh đề when là biểu thức đúng sai (boolean) sẽ được đánh giá đối với thông điệp được xử lý. Biểu thức hỗ trợ các toán hạng đúng sai chung như AND (&&), OR (||), NOT (!) và các toán hạng đúng sai (<; <=, >,>=,==,!=). Bất cứ hàm nào trả về một giá trị nào đó đều có thể được gọi trong mệnh đề when nhưng nó phải được đánh giá với một biểu thức đúng sai. Nếu trong biểu thức gọi một hàm không tồn tại thì nó sẽ tương ứng với giá trị false

1. Hành động (Action)

Mệnh đề then của luật chứa danh sách các hành động. Chúng sẽ được thực thi theo thứ tự mà chúng xuất hiện. Có hai loại hành động:

* Gọi hàm
* Gán biến

Lời gọi hàm cũng giống với lời gọi hàm ở trong mệnh đề when. Khi gọi hàm thì nó sẽ thực hiện chức năng tương ứng. Biến được sử dụng để tránh phải tính toán xử lý lại các phép toán. Nó sẽ giữ giá trị tạm thời và làm cho đoạn mã dễ đọc hơn. Biến cần phải được định nghĩa trước khi sử dụng. Danh sách các action có thể để trống

1. **Cấu hình graylog**

Trước tiên chúng ta cần cài java vì graylog được viết bằng java. Chúng ta cài bằng câu lệnh:

yum install -y java-1.8.0-openjdk-headless.x86\_64

Sau đó chúng ta cần cài mongodb để lưu trữ cấu hình,siêu dữ liệu. Đầu tiên chúng ta cần thêm mongodb repo:

vi /etc/yum.repos.d/mongodb-org.repo

[mongodb-org-4.4]  
name=MongoDB Repository  
baseurl=https://repo.mongodb.org/yum/redhat/7/mongodb-org/4.4/x86\_64/  
gpgcheck=1  
enabled=1  
gpgkey=https://www.mongodb.org/static/pgp/server-4.4.asc

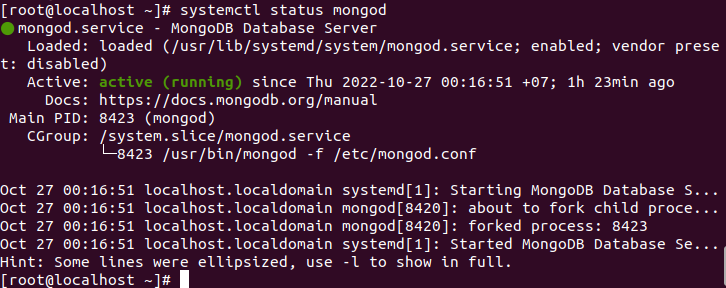
Sau đó cài đặt mongodb bằng câu lệnh:

yum install –y mongodb-org

Khởi chạy dịch vụ và cho dịch vụ khởi động cùng với máy:  
systemctl daemon-reload  
systemctl enable mongod.service  
systemctl start mongod.service

Sau đó kiểm tra status của mongodb:

systemctl status mongod



*Hình 3.20. Trạng thái của mongodb*

Cài đặt Elastic GPG key:

rpm --import <https://artifacts.elastic.co/GPG-KEY-elasticsearch>

Thêm elasticsearch repository:

vi /etc/yum.repos.d/elasticsearch.repo

[elasticsearch-7.x]  
name=Elasticsearch repository for7.x packages  
baseurl=https://artifacts.elastic.co/packages/oss-7.x/yum  
gpgcheck=1  
gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch  
enabled=1  
autorefresh=1  
type=rpm-md

Cài đặt elastic với bản phát hành mới nhất:

sudo yum install –y elasticsearch-oss

Sửa đổi file cấu hình của elasticsearch /etc/elasticsearch/elasticsearch.yml và đặt tên của cụm là graylog và bỏ comment action.auto\_create\_index: false

sudo tee -a /etc/elasticsearch/elasticsearch.yml > /dev/null <<EOT  
cluster.name: graylog  
action.auto\_create\_index: false  
EOT

Khởi chạy dịch vụ elasticsearch và cho khởi chạy cùng với lúc máy khởi động:

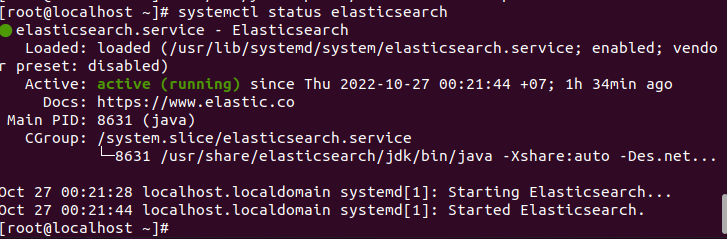
systemctl daemon-reload

systemctl enable elasticsearch

systemctl restart elasticsearch

Kiểm tra trạng thái của elasticsearch:

systemctl status elasticsearch



*Hình 3.21. Trạng thái của elasticsearch*

Cài đặt graylog-server:

rpm -Uvh  
<https://packages.graylog2.org/repo/packages/graylog-4.2-repository_latest.rpm>

yum install graylog-server

Cài đặt pwgen để tạo secret key cho graylog:

yum install epel-release

yum install pwgen

Tạo password\_secret:

password\_secret = pwgen –N 1 –s 96

Tạo root\_password\_sha2:

root\_password\_sha2 = echo –n password | sha256sum

Chỉnh sửa file cấu hình /etc/graylog/server/server.conf:

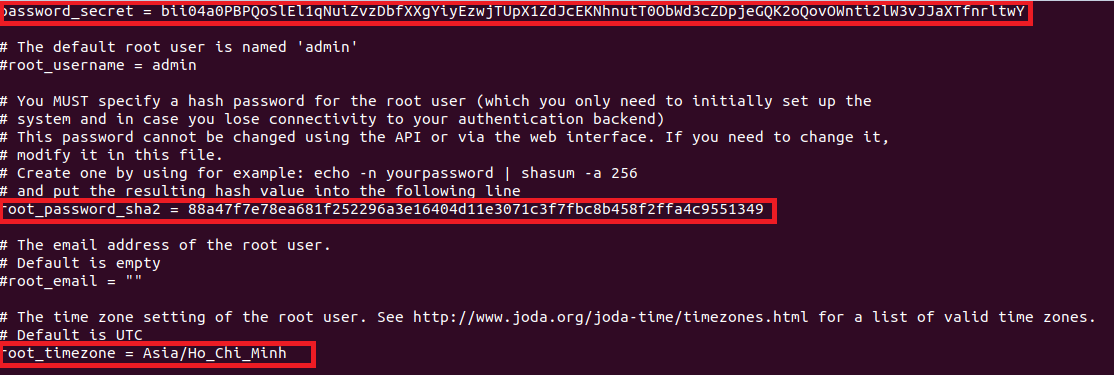
vi /etc/graylog/server/server.conf

password secret =

root\_password\_sha2 =

root\_timezone = Asia/Ho\_Chi\_Minh

http\_bind\_address = 0.0.0.0:9000



*Hình 3.22. Nội dung file cấu hình của graylog*

Cho graylog-server chạy cùng với lúc khởi động hệ thống và bắt đầu chạy dịch vụ:

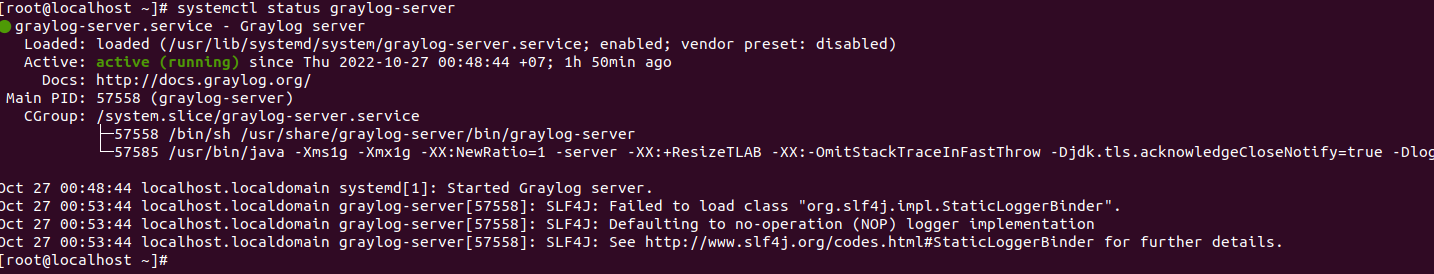
systemctl daemon-reload

systemctl enable graylog-server

systemctl start graylog-server

Kiểm tra status của graylog-server:

systemctl status graylog-server



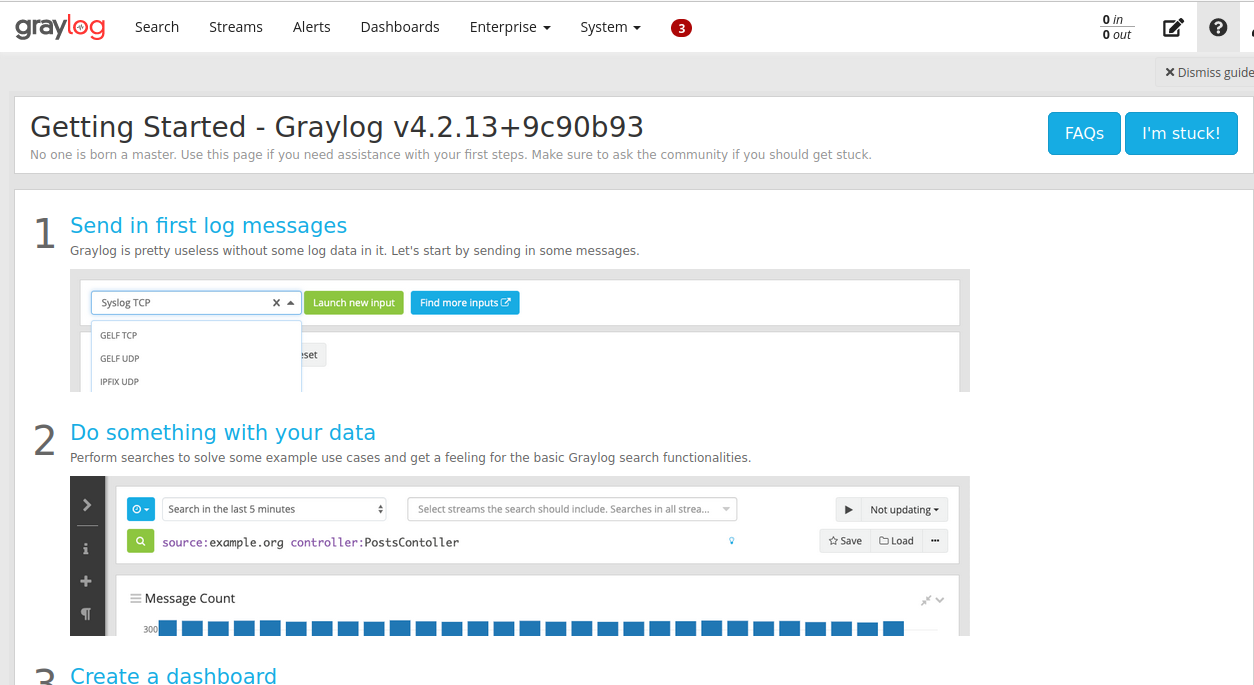
*Hình 3.23. Trạng thái của graylog-server*

Cho phép port 9000 trên firewall:

firewall-cmd --add-port=9000/tcp --permanent

firewall-cmd --reload

Sau khi hoàn thành các bước thì ta có thể vào giao diện quản trị của graylog



*Hình 3.24. Giao diện quản trị của graylog*

1. **Kết chương**

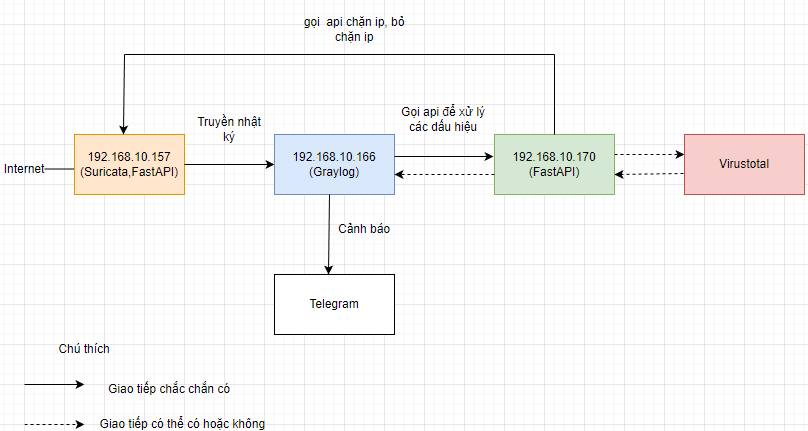
Chương 3 đã trình bày về phần mềm quản lý nhật ký graylog bao gồm khái niệm, các thành phần,chi tiết các thành phần,kiến trúc,tính năng cũng như là hướng dẫn cài đặt graylog. Graylog có rất nhiều tính năng nhưng một trong những tính năng đáng chú ý là pipeline và cho phép viết luật. Pipeline sẽ bao gồm nhiều luật cho phép chúng ta chuẩn hóa nhật ký và cho phép chúng ta có thể gọi API. Việc gọi được API là tiền đề để chúng ta có thể thực hiện được việc tự động hóa các quy trình như tự động tìm kiếm sự độc hại của các IoC trên virustotal và đây cũng là tiền đề để có thể thực hiện thử nghiệm trong chương sau.

# CHƯƠNG 4. Cài đặt và thử nghiệm

*Chương 4 mô tả về quá trình cài đặt hệ thống bao gồm thành phần: hệ thống quản lý log graylog, hệ thống phát hiện xâm nhập suricata, máy chủ API. Ngoài ra chương này có giới thiệu về các phần mềm, môi trường cho từng máy và cuối cùng chương này sẽ mô tả, cài đặt và trình bày về các kịch bản thử nghiệm liên quan tới hệ thống quản lý nhật ký graylog kết hợp với việc tự động hóa phát hiện ra các ip độc hại và tự động ngăn chặn ip độc hại đó trên hệ thống ngăn chặn xâm nhập suricata*

## 4.1. Cài đặt

1. ***Mô hình cài đặt***

****

*Hình 4.1. Mô hình của hệ thống*

Hệ thống gồm 5 thành phần chính, trong đó mỗi thành phần là 1 khối riêng biệt:

* Máy có ip 192.168.10.157: Đây là máy đóng vai trò như một IPS, trên máy này có cài suricata đóng vai trò là IPS,FastAPI (Python framework) để viết các api để ngăn chặn cũng như là bỏ ngăn chặn các ip,hash trên suricata. Ngoài ra trên máy này còn cài Filebeat để đẩy nhật ký ghi được ở suricata lên hệ thống Graylog có ip là 192.168.10.166
* Máy có ip 192.168.10.166: Đây là máy dùng để quản lý,lưu trữ log cũng như gửi các api đến 192.168.10.165 để kiểm tra ip,mã băm mỗi khi có thông điệp được gửi lên và cũng dùng để đưa ra cảnh báo tới telegram trong một số điều kiện nhất định.
* Máy có ip 192.168.10.165: Đây là máy cài đặt FastAPI (Python framework) dùng để viết các api để xử lý các IoC (ip, mã băm) được gửi tới. Cách api nó xử lý như nào thì phụ thuộc vào từng kịch bản.
* Virustotal: Dùng để kiểm tra sự độc hại của các IoC
* Telegram: Dùng để nhận cảnh báo từ graylog

1. ***Giới thiệu các thành phần***
   * + 1. *Suricata*

Đã được giới thiệu ở phần 1.2.1.4.b

* + - 1. *Graylog*

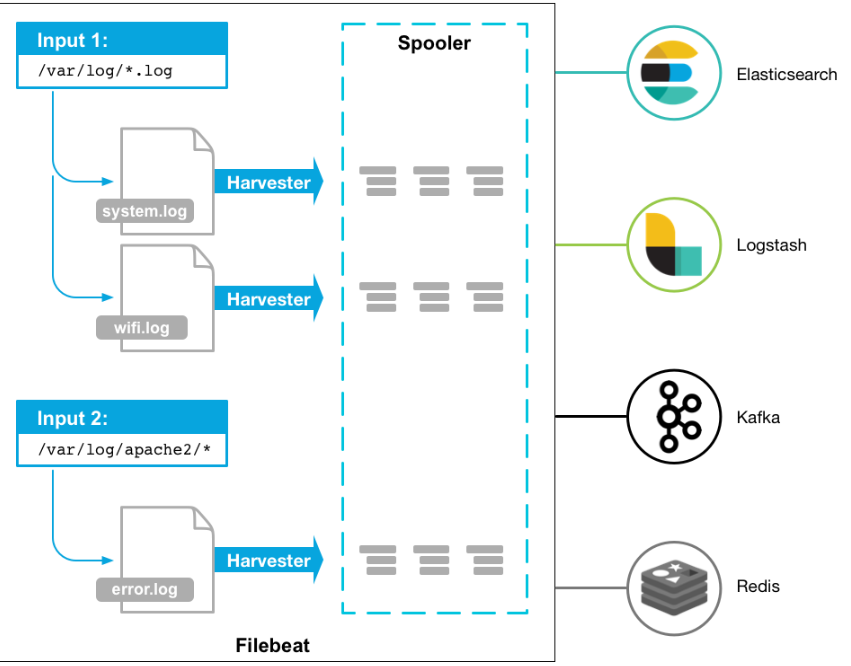
Đã được giới thiệu ở chương 2

* + - 1. *FastAPI*

FastAPI là một framework hiện đại,hiệu suất nhanh được sử dụng để xây dựng API với ngôn ngữ Python.Một số các tính năng chính của FastAPI: [11]

* Nhanh: hiệu suất rất cao ngang ngửa với NodeJS và Go
* Ít lỗi hơn: Giảm được khoảng 40% lỗi do con người gây ra
* Dễ dàng: Framework được thiết kế để cho người dùng dễ học và sử dụng
* Ngắn gọn: Giảm thiểu việc lặp lại mã
  + - 1. *Filebeat*

Filebeat là một công cụ vận chuyển nhật ký để chuyển tiếp và tập trung dữ liệu nhật ký trên hệ thống tập trung. Filebeat sẽ được cài đặt ở trên máy chủ, nó sẽ giám sát các tệp nhật ký tại nơi mà chúng ta chỉ định, nó sẽ thu thập các sự kiện nhật ký và chuyển tiếp chúng tới các hệ thống tập trung để lưu trữ như Elasticsearch hay Logstash. Khi chúng ta khởi động filebeat, nó sẽ bắt đầu tìm kiếm các vị trí các nhật ký mà chúng ta đã chỉ định trong đầu vào lúc cấu hình. Đối với mỗi nhật ký mà Filebeat định vị được, Filebeat sẽ khởi động trình thu hoạch (harvester). Mỗi trình thu hoạch đọc một nhật ký đơn lẻ cho nội dung mới và sẽ gửi dữ liệu nhật ký mới tới libbeat cái sẽ tổng hợp sự kiện và gửi dữ liệu tổng hợp mà chúng ta đã cấu hình cho Filebeat. [12]



*Hình 4.2. Cấu trúc filebeat*

* + - 1. *Virustotal*

Virustotal là một trang web được tạo ra bởi công ty bảo mật của Tây Ban Nha. Virustotal tổng hợp nhiều công cụ chống virus và các công cụ quét trực tuyến. Virustotal cho phép chúng ta quét các đường dẫn,tệp,ip độc hại. Các tệp với kích thước tối đa 650MB có thể được tải lên để virustotal phân tích xem có độc hại hay không. Virustotal sử dụng Cuckoo sandbox để phân tích động các tệp độc hại. Ngoài ra virustotal còn cung cấp các API để hỗ trợ việc tự động hóa phân tích các tệp,ip,url. Tuy nhiên nếu sử dụng API key miễn phí thì sẽ bị giới hạn số yêu cầu được gửi lên là chỉ có 500 yêu cầu được gửi lên 1 ngày và giới hạn với tốc độ 4 yêu cầu/phút. [13]

1. ***Cơ chế hoạt động của hệ thống***

Khi mà các gói tin từ trong đi ra hay từ ngoài đi vào hệ thống thì đều đi qua hệ thống ngăn chặn xâm nhập Suricata. Ở trên máy chạy suricata có cài filebeat để có thể vận chuyển nhật ký của suricata lên graylog, ngoài ra với filebeat có bật module suricata để có thể chuẩn hóa nhật ký của suricata. Sau khi chuẩn hóa thì thông điệp sẽ có thêm một số trường như destination\_ip,source\_ip,destination\_port,source\_port và sau đó sẽ được gửi lên graylog. Ở trên graylog, chúng ta có thể cấu hình để nó gọi API tới máy chủ chạy FastAPI (192.168.10.170) để có thể xử lý một số dấu hiệu được gửi lên từ các thông điệp như ip,hash. Cái cách thức để graylog gọi được API thì cũng tùy vào các trường hợp. Có thể cấu hình pipeline trên graylog để gọi API hoặc cũng có thể gọi API khi một sự kiện nào đó trên graylog được kích hoạt, chi tiết cách thức gọi API sẽ ở trong từng kịch bản phía dưới. Sau khi gọi tới api trên FastAPI (192.168.10.170), các api làm nhiệm vụ sẽ kiểm tra xem các dấu hiệu tấn công có phải là độc hại hay không bằng cách sẽ gửi API lên virustotal. Ngoài ra các api này còn gọi tới các API trên máy chạy suricata để chặn các dấu hiệu hoặc cho phép không chặn nữa và nó còn có tác dụng làm giàu dữ liệu sau đó trả về graylog,graylog có thể dựa vào các dấu hiệu này để đưa cảnh báo tương ứng tới telegram. Để làm được như vậy thì sẽ cài đặt thêm FastAPI (Python framework) trên máy chạy suricata. Các API trên này có nhiệm vụ là tự động viết các bộ luật trên suricata để có thể ngăn chặn các dấu hiệu độc hại, chi tiết về bộ luật sẽ nói trong từng kịch bản. Để có thể gửi cảnh báo tới telegram thì chúng ta cần cấu hình cảnh báo và con bot trên telegram để khi gặp sự kiện bảo mật nào đó trên graylog nó sẽ gọi tới con bot trên telegram và con bot này sẽ đưa ra cảnh báo với định dạng mà chúng ta cấu hình.

1. ***Môi trường và các công cụ***
2. *Máy chủ có ip 192.168.10.157*

* Hệ điều hành Ubuntu 20.04
* Hệ thống ngăn chặn xâm nhập Suricata
* Python3
* FastAPI
* Filebeat

1. *Máy chủ có ip 192.168.10.166*

* Hệ điều hành Centos 7
* Java 8
* Graylog
* Elasticsearch
* MongoDB

1. *Máy chủ có ip 192.168.10.170*

* Hệ điều hành Centos 7
* Python 3
* FastAPI framework

1. ***Cài đặt các thành phần***
2. *Suricata*
3. Cài đặt

Thêm thông tin về kho phần mềm của Open Information Security Foundation (OSIF) vào ubuntu

sudo add-apt-repository ppa:oisf/suricata-stable

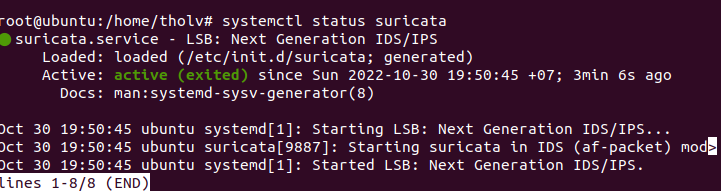
Cài đặt suricata:

sudo apt install suricata

Kiểm tra trạng thái và cho chạy lúc máy khởi động:

system enable suricata

systemctl status suricata

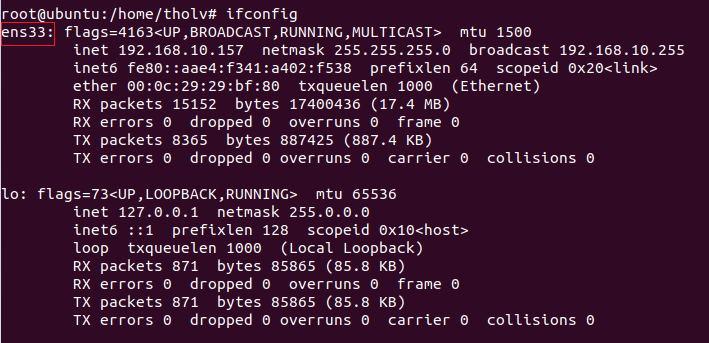


*Hình 4.3. Trạng thái của suricata*

1. Cấu hình suricata

Kiểm tra các interface đang dùng của máy:

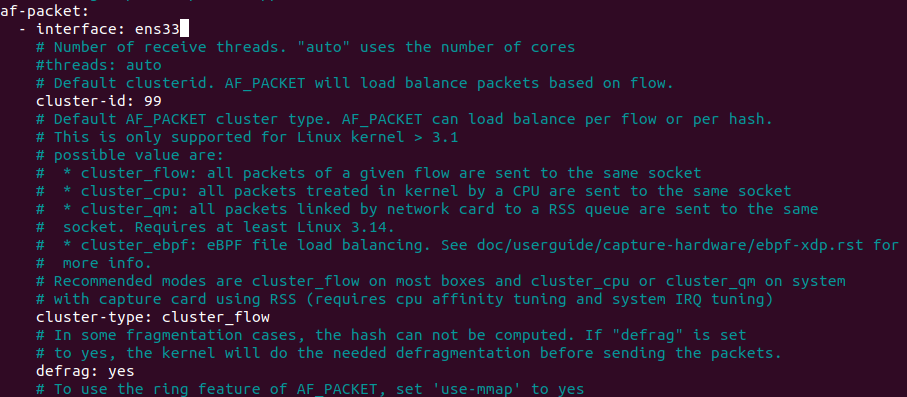
ifconfig



*Hình 4.4. Xác định interface máy đang dùng*

Mở file /etc/suricata/suricata.yaml và sửa interface tương ứng

nano /etc/suricata/suricata.yaml



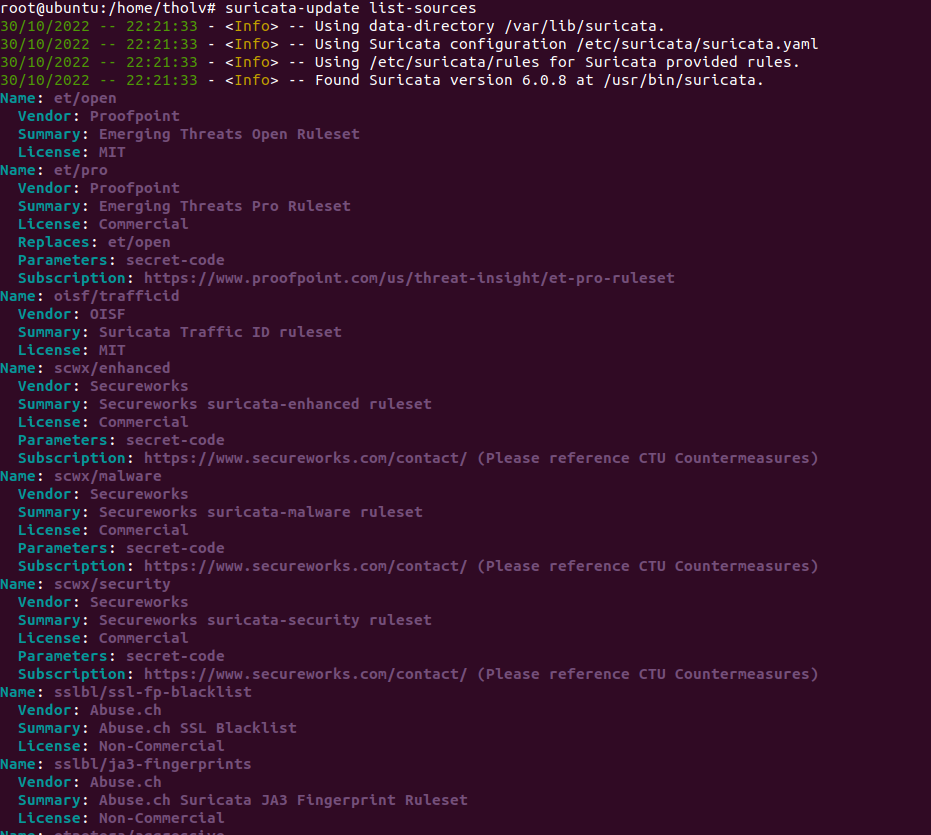
*Hình 4.5. Cấu hình interface cho suricata*

Cập nhật,tải các bộ luật được cập nhật mới nhất từ các nhà cung cấp luật về:

sudo suricata-update

Liệt kê các tập luật được cung cấp bởi các nhà cung cấp:

suricata-update list-sources

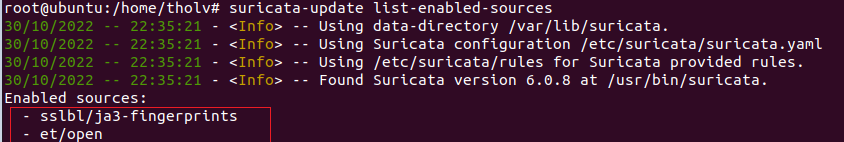


*Hình 4.6. Các bộ luật được cung cấp bởi các nhà cung cấp*

Bật các bộ luật et/open, sslbl/ssl-fp-blacklist

suricata-update enable-source et/open  
suricata-update enable-source sslbl/ja3-fingerprints

Kiểm tra các bộ luật được bật



*Hình 4.7. Các bộ luật được bật*

Cập nhật các rule từ source-list đã được enable:

suricata-update

Kiểm tra xem cảnh báo đã chạy chưa:

<http://testmynids.org/uid/index.html>

Kết quả trả về có chứa “uid” là một dấu hiệu hệ thống có thể đã bị thỏa hiệp thông qua web shell

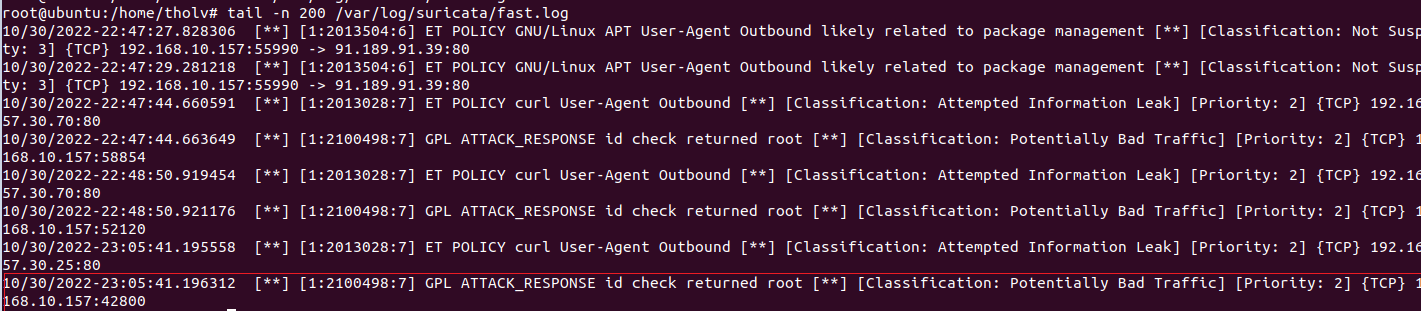


*Hình 4.8. Kết quả thử request*

Kiểm tra cảnh báo:

tail –n 200 /var/log/suricata.log

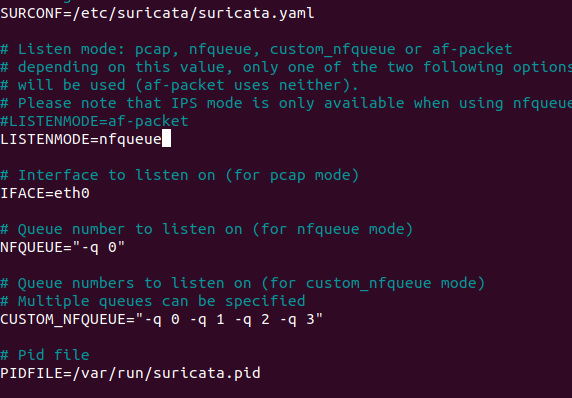
2100498 là rule để cảnh báo nếu trong response có chứa “uid”



*Hình 4.9. Cảnh báo tương ứng với yêu cầu ở trên*

Cấu hình để suricata chạy ở chế độ chặn

LISTENMODE=nfqueue

*Hình 4.10. Cấu hình suricata ở chế độ chặn*

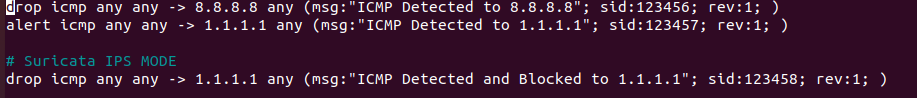
Cấu hình để iptable gửi traffic tới suricata

sudo iptables –I FORWARD –j NFQUEUE

sudo iptables –I INPUT –j NFQUEUE

sudo iptables –I OUTPUT –j NFQUEUE

Tạo bộ luật tùy chỉnh

 *Hình 4.11. Luật tùy chỉnh để thử chế độ ngăn chặn*

Kiểm tra xem đã block chưa

 *Hình 4.12. Kết quả suricata chặn ip 8.8.8.8*

Thử kiểm tra nhật ký cảnh báo trong tệp nhật ký của suricata

 *Hình 4.13. Nhật ký cảnh báo chặn ip 8.8.8.8 của suricata*

1. *FastAPI*

Cài đặt môi trường ảo cho python:

apt upgrade –y

apt install –y python3-pip

apt install build-essential libssl-dev libffi-dev python3-dev

apt install –y python3-venv

Tạo môi trường ảo:

python3 –m venv env\_test

source /{dir}/env\_test/bin/activate

Cài đặt fastapi:

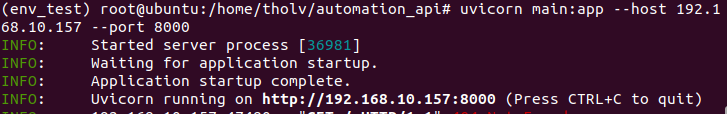
pip install fastapi

Cài đặt uvicorn:

pip install uvicorn[standard]

Chạy server

uvicorn main:app –host 192.168.10.157 --port 8000 –reload

 *Hình 4.14. Trạng thái của máy chủ python*

1. *Filebeat*
2. Cài đặt

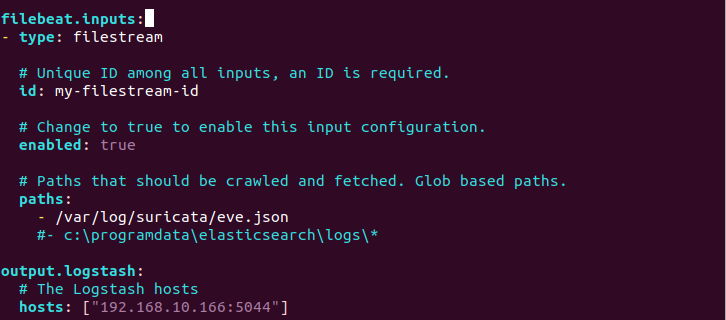
curl -L -O <https://artifacts.elastic.co/downloads/beats/filebeat/filebeat-7.17.7-amd64.deb>

sudo dpkg -i filebeat-7.17.7-amd64.deb

1. Cấu hình

Vào file /etc/filebeat/filebeat.yml và cấu hình đường dẫn tới tệp nhật ký và địa chỉ để đẩy nhật ký tới như sau:

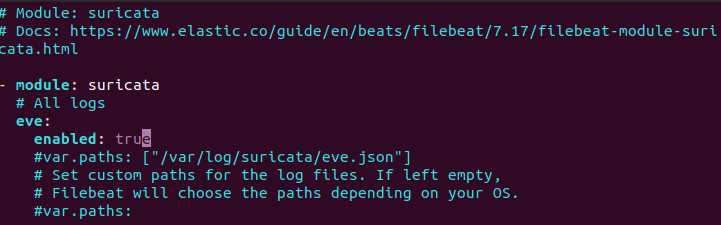
nano /etc/filebeat/filebeat.yml

 *Hình 4.15. Cấu hình của filebeat*

Cho phép bật module suricata. Đây là module để chuẩn hóa nhật ký của suricata. Nó sẽ thêm một số trường ý nghĩa như source\_ip,destination\_ip,source\_port…. Thực hiện với câu lệnh sau:

filebeat modules enable suricata

Vào file /etc/filebeat/modules.d/suricata.yml và chỉnh cấu hình như sau:

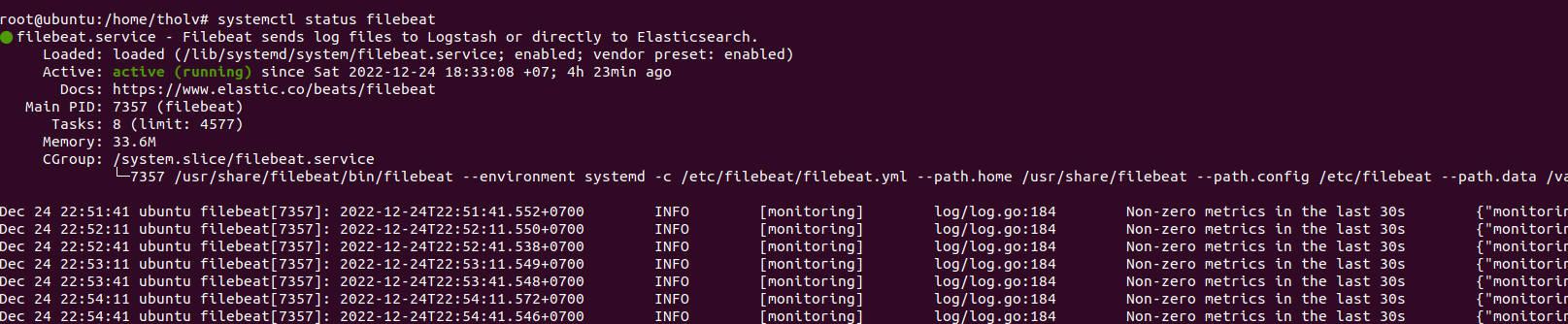
 *Hình 4.16. Cấu hình module suricata của filebeat*

Khởi động filebeat:

systemctl start filebeat

Kiểm tra trạng thái của filebeat:

systemctl status filebeat

 *Hình 4.17. Trạng thái của filebeat*

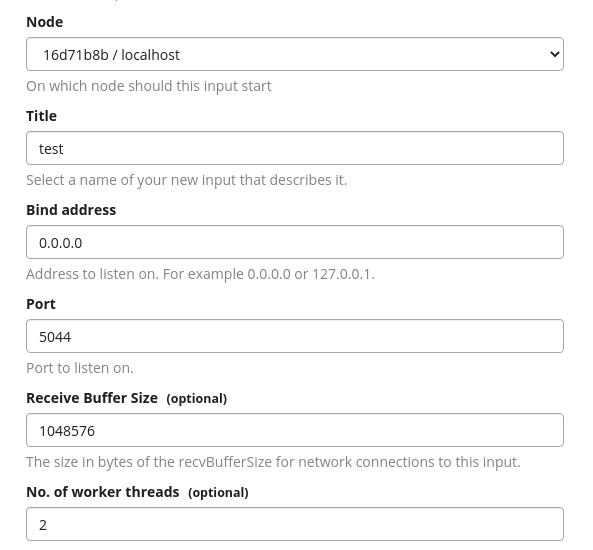
1. *Graylog*

Phần cài đặt graylog đã được trình bày ở phần 3.6

Ở phần này sẽ trình bày cấu hình đầu vào cho graylog và cấu hình để gửi thông báo tới telegram

1. Cấu hình đầu vào cho graylog

Tạo input Beats để nhận log từ filebeat thông qua cổng 5044

****

*Hình 4.18. Tạo input beats*

1. Cấu hình thông báo về telegram

Cài đặt plugin telegram cho graylog

Di chuyển vào thư mục chứa plugin của graylog

cd /usr/share/graylog-server/plugin

Tải plugin telegram từ github về

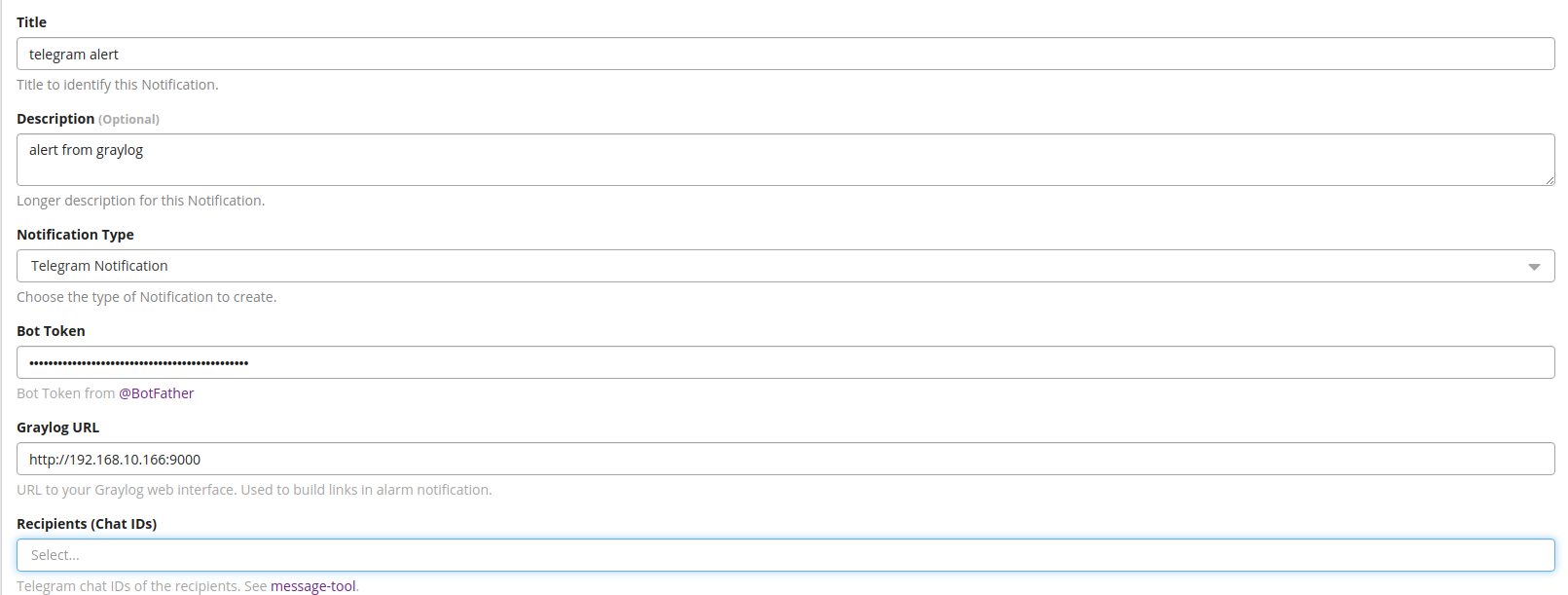
wget <https://github.com/irgendwr/TelegramAlert/releases/download/v2.3.6/graylog-plugin-telegram-notification-2.3.6.jar>

Khởi động lại graylog

systemctl restart graylog-server

Cấu hình bot trên telegram và lấy thông tin chat\_id,bot token

Cấu hình thông báo về telegram trên graylog



*Hình 4.19. Cấu hình thông báo về telegram*

* 1. **Thử nghiệm**

1. ***Kịch bản 1***
2. *Mô tả*

Trong kịch bản 1, tất cả các ip đi qua suricata được gửi lên trên graylog đều sẽ được kiểm tra qua virustotal. Nếu kết quả là nếu khoảng 4 tổ chức trở lên báo ip đấy là độc hại thì sẽ thực hiện ngăn chặn tự động ip đó trên suricata và sẽ đưa cảnh báo ngăn chặn tương ứng về telegram. Trong trường hợp ip đang trong trạng thái bị chặn mà kiểm tra trên virustotal không có kết quả nào báo là độc hại thì sẽ thực hiện việc cho phép ip tự động trên suricata và sẽ gửi thông báo tương ứng về telegram

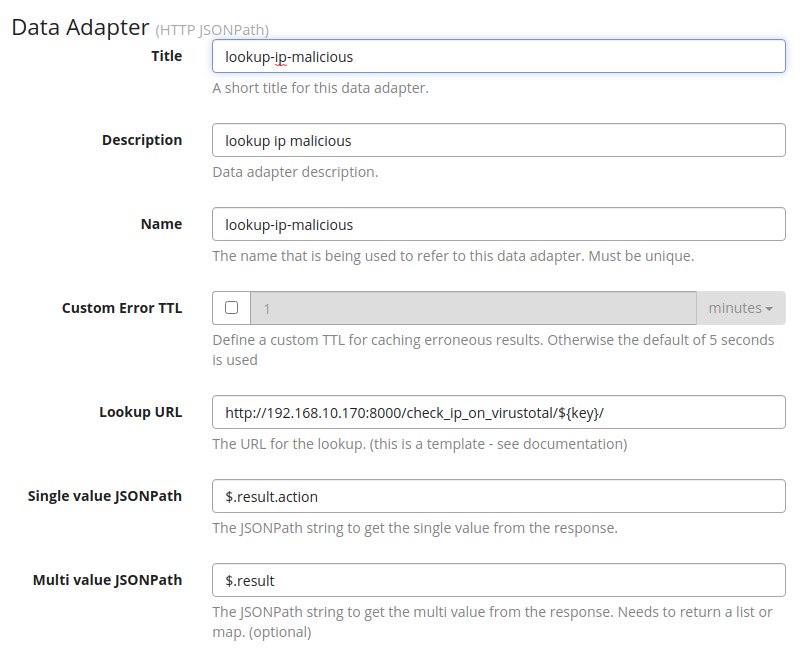
1. *Luồng hoạt động*

Đầu tiên nhật ký của suricata sẽ được đẩy lên graylog

Ở trên graylog sẽ viết các bộ luật pipeline để lấy ra giá trị của các trường source\_ip,destination\_ip và sẽ tra cứu bảng. Chúng ta sẽ tạo bảng tra cứu thực hiện việc request api tới 192.168.10.170. API trên 192.168.10.170 sẽ thực hiện request tới API của virustotal để kiểm tra mức độ độc hại của IP. Sau khi 192.168.10.170 thực hiện request API tới virustotal sẽ được nhận được phản hồi dưới dạng JSON, API trên 192.168.10.170 sẽ xử lý dữ liệu phản hồi này và trả về các trường cho graylog như: action có nghĩa là hành động tương ứng được thực hiện trên suricata, reason có nghĩa là lý do hành động tương ứng được thực hiện. Ngoài ra nếu kết quả trả về từ virustotal cho thấy có trên 4 tổ chức đánh giá đây là ip độc hại thì sẽ gọi API block tương ứng trên suricata. Ở trên suricata cũng cài FastAPI chứa các API để thực hiện việc chặn hoặc cho phép ip trên suricata. API này sẽ viết vào file rule của suricata luật để chặn ip có dạng: drop ip [list] … Trong trường hợp trạng trạng thái bị chặn của ip được gửi lên từ suricata nhưng sau đó tìm kiếm trên virustotal không thấy các tổ chức báo ip đó là độc hại nữa thì sẽ gửi API để cho phép ip tới suricata để thực hiện việc cho phép ip. Để cho phép ip thì chúng ta chỉ cần đơn giản bỏ ip đấy trong danh sách ip của luật chặn ip. Mỗi ip được gửi lên sẽ được phân tích trên virustotal trong khoảng thời gian 30 phút/lần. Ngoài ra,nếu ip đang bị block mà trước khoảng thời gian 30 phút chúng ta muốn cho phép ip đó luôn thì chúng ta có thể thêm ip đó vào danh sách trắng, ngay lập tức ip đó sẽ được cho phép trên suricata. Tất cả các hành động như ngăn chặn hay cho phép đều sẽ có thông báo tới telegram.

1. *Cấu hình các thành phần*

Cấu hình bảng tra cứu để nó gửi yêu cầu api tới FastAPI (192.168.10.170)



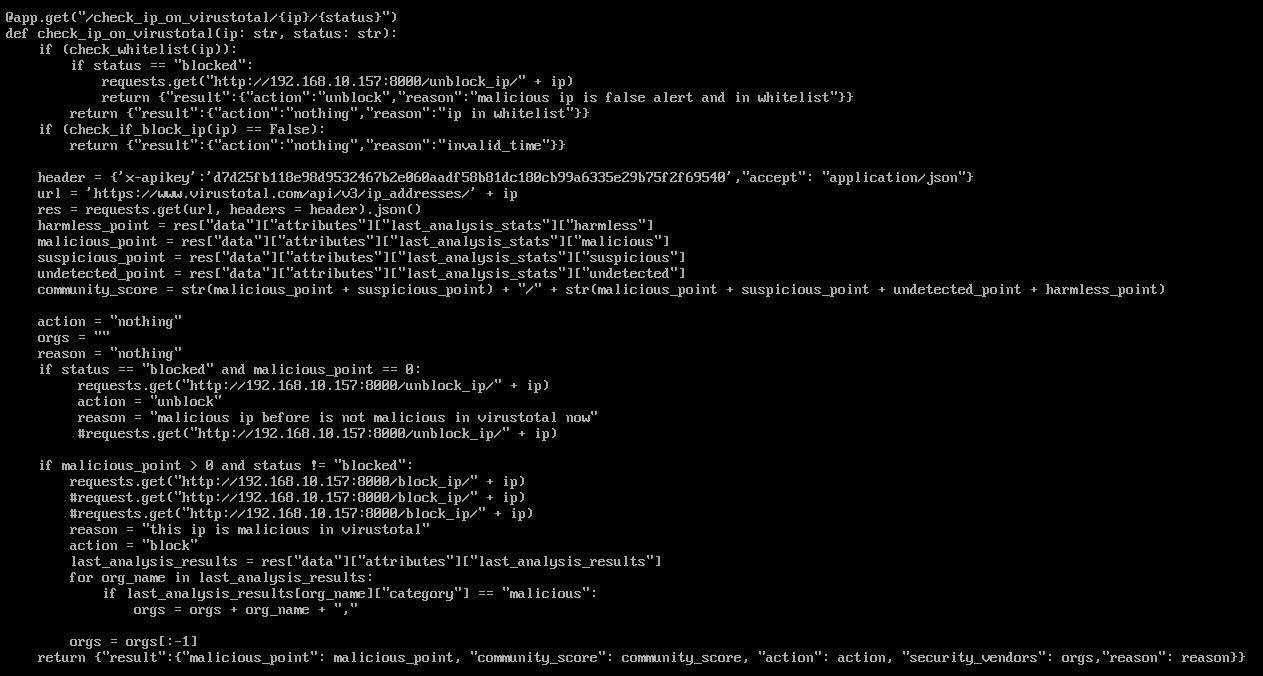
*Hình 4.20. Cấu hình bảng tra cứu*

Cấu hình các bộ luật pipeline để thực hiện tra cứu tới bảng tra cứu mỗi khi có thông điệp được gửi lên mà có trường destination\_ip và source\_ip



*Hình 4.21. Luật để tra cứu ip với bảng tra cứu*

Dưới đây là đoạn mã viết cho api check\_ip\_on\_virustotal



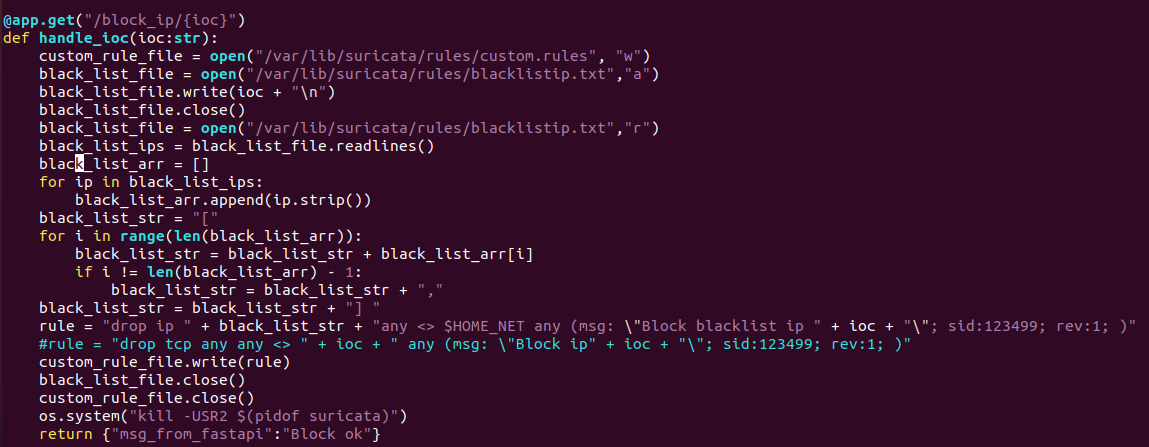
*Hình 4.22. Api check\_ip\_on\_virustotal*

Api này sẽ hoạt động như sau:



*Hình 4.23. Luồng hoạt động của api check\_ip\_on\_virustotal*

Dưới đây là api để chặn ip trên suricata được viết bằng python3 sử dụng FastAPI framework



*Hình 4.24. Api chặn ip trên suricata*

Cách hoạt động của api này đơn giản là chúng ta sẽ tạo 1 file custom.rules trong /var/lib/suricata/rules và thêm vào file cấu hình. Sau đó thì api này sẽ ghi luật vào file đấy. Luật có dạng như sau:

drop ip [list ip]

Mỗi khi có một ip mới thì api này sẽ xử lý để thêm ip này vào danh sách các ip, do đó chúng ta chỉ cần 1 luật nhưng có thể chặn được nhiều ip. Việc bỏ chặn ip thì api chỉ cần loại bỏ ip đó ra khỏi danh sách

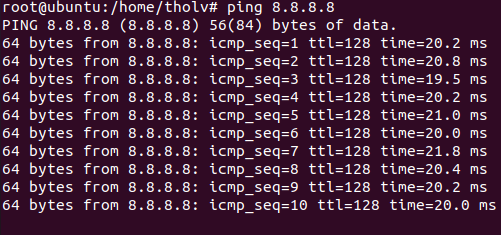
**

*Hình 4.25. Luật để chặn ip trên suricata*

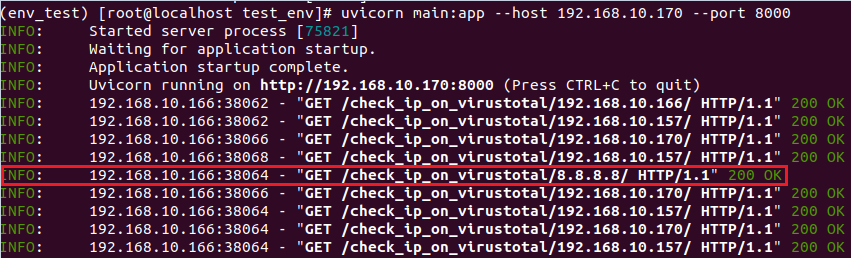
1. *Thử nghiệm*

Do phần thử nghiệm của đồ án này có hạn chế là suricata chưa đưa lên được internet nên việc nhận ip từ bên ngoài bị hạn chế. Trong quá trình thử nghiệm thì thấy ip 8.8.8.8 có 2 nhà đóng góp đánh giá là độc hại là Comodo Valkerie Verdict và Cyble. Tuy nhiên đây có thể là cảnh báo sai, trong quá trình thử nghiệm thì thấy các nhà đóng góp trên cũng cảnh báo sai khá nhiều nên trong api để ngưỡng là 4 nhà đóng góp báo là độc hại thì mới chặn. Tuy nhiên việc này cũng có thể gây ra cảnh báo sai, do đó mỗi hành động chặn hay bỏ chặn thì đều có cảnh báo tới telegram. Nếu người quản trị thấy chặn nhầm ip thì có thể vào thêm whitelist ở trên 192.168.10.170 để bỏ chặn. Riêng với thử nghiệm này sẽ hạ ngưỡng độc hại xuống là 1 nghĩa là chỉ cần 1 nhà đóng góp báo là độc hại thì cũng chặn để thấy được là hệ thống hoạt động đúng. Như ở trên trong quá trình thử nghiệm thì thấy ip 8.8.8.8 có 2 nhà đóng góp báo là độc hại nên ta sẽ thử ping tới 8.8.8.8 và xem hệ thống có tự động chặn và gửi cảnh báo không. Nếu có thì sau đó thêm whitelist để xem hệ thống có bỏ chặn và gửi cảnh báo không.

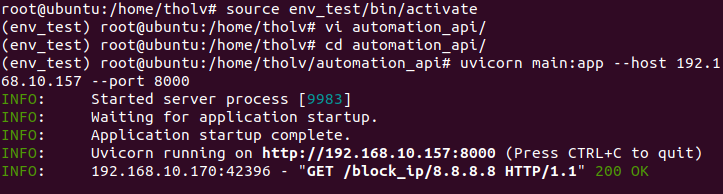
Ping thử ip 8.8.8.8 thấy vẫn ping được bình thường:

 *Hình 4.26. Trạng thái khi ping 8.8.8.8 lần đầu tiên*

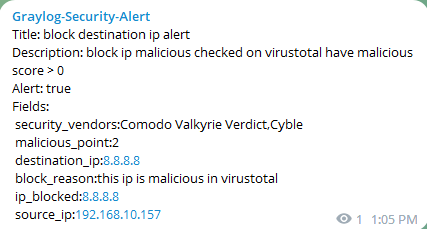
Thấy được ở FastAPI (192.168.10.170) đã nhận được request check\_ip\_on\_virustotal từ graylog

*Hình 4.27. Python server 192.168.10.170 nhận được request check ip với ip 8.8.8.8*

Thấy được ở python server đang chạy trên 192.168.10.157 cũng là máy chủ đang chạy suricata nhận được request block ip từ 192.168.10.170

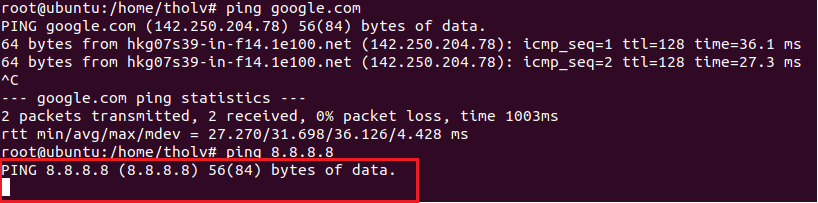
 *Hình 4.28. Python server 192.168.10.157 nhận được yêu cầu chặn ip*

Thấy cảnh báo báo về telegram

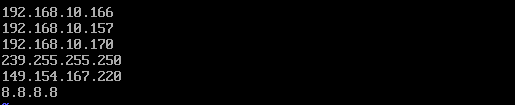


*Hình 4.29. Cảnh báo chặn ip 8.8.8.8 báo về telegram*

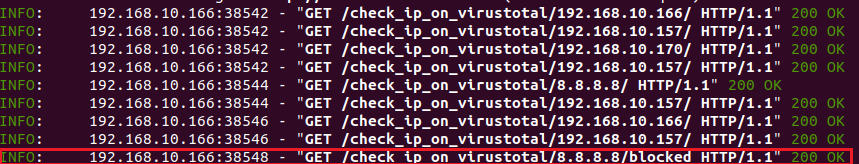
Ping lại sau khi bị chặn sẽ thấy không ping được nữa

 *Hình 4.30. Ip 8.8.8.8 đã bị chặn*

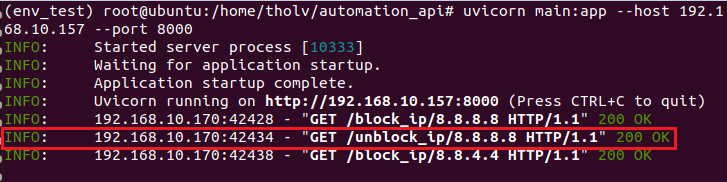
Sau khi nhận cảnh báo thấy đây là cảnh báo sai, ta có thể vào thêm whitelist ip trên 192.168.10.170

 *Hình 4.31. Thêm whitelist ip*

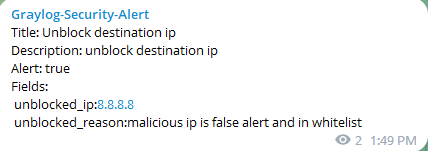
Sau khi ping lại ip 8.8.8.8 thấy python server 192.168.10.170 nhận được request với thêm tham số blocked

 *Hình 4.32. Yêu cầu nhận được sau khi chặn ip 8.8.8.8*

Thấy nhận được request unblock ip trên suricata từ python server 192.168.10.170

 *Hình 4.33. Yêu cầu bỏ chặn ip 8.8.8.8*

Nhận được thông báo bỏ chặn ip trên telegram



*Hình 4.34. Thông báo bỏ chặn ip 8.8.8.8*

* 1. **Kết chương**

Chương 4 đã trình bày về mô hình để giám sát và tự động hóa một số hành động bao gồm suricata, graylog, FastAPI, Virustotal. Chương này cũng trình bày chi tiết về các thành phần của hệ thống, luồng hoạt động của hệ thống trong các kịch bản cũng như trình bày chi tiết cách hoạt động của từng thành phần trong luồng chạy.Tại chương này cũng đã cài đặt và thử nghiệm thành công các kịch bản tự động phân tích ip và đưa ra hành động tương ứng. Cuối cùng, chương 4 đưa ra kết quả thử nghiệm của các kịch bản

# KẾT LUẬN

**Kết quả đạt được:**

Đồ án đã thực hiện được những nội dung sau:

* Trình bày tổng quan giám sát an toàn thông tin bao gồm khái niệm về giám sát an toàn thông tin,cơ chế hoạt động cũng như một số lợi ích của giám sát an toàn thông tin. Giới thiệu về các thành phần cơ bản của hệ thống giám sát an toàn thông tin như hệ thống phát hiện xâm nhập, cơ sở dữ liệu chứa các dấu hiệu tấn công, hệ thống quản lý sự kiện và thông tin bảo mật.
* Trình bày về quản lý nhật ký bao gồm giới thiệu về nhật ký và hệ thống quản lý nhật ký. Trình bày khái niệm về nhật ký,các loại nhật ký,kiến trúc và chức năng của hệ thống quản lý nhật ký.
* Trình bày về phần mềm quản lý nhật ký graylog bao gồm khái niệm, các thành phần,chi tiết các thành phần,kiến trúc,tính năng cũng như là hướng dẫn cài đặt graylog. Graylog có rất nhiều tính năng nhưng một trong những tính năng đáng chú ý là pipeline và cho phép viết luật. Pipeline sẽ bao gồm nhiều luật cho phép chúng ta chuẩn hóa nhật ký và cho phép chúng ta có thể gọi API. Việc gọi được API là tiền đề để chúng ta có thể thực hiện được việc tự động hóa các quy trình như tự động tìm kiếm sự độc hại của các dấu hiệu tấn công trên Virustotal.
* Mô tả về quá trình cài đặt hệ thống bao gồm thành phần: hệ thống quản lý log graylog, hệ thống phát hiện xâm nhập suricata, máy chủ API. Cài đặt và trình bày về các kịch bản thử nghiệm liên quan tới hệ thống quản lý nhật ký graylog kết hợp với việc tự động hóa phát hiện ra các ip độc hại và tự động ngăn chặn ip độc hại đó trên hệ thống ngăn chặn xâm nhập suricata.

**Hạn chế:**

* Hệ thống lưu trữ như elasticsearch mới có một thành phần đơn lẻ dẫn tới việc lưu trữ không được tối ưu và tốc độ xử lý dữ liệu không được nhanh.
* Các công cụ để phân tích các dấu hiệu tấn công còn ít, chính sách để ra lệnh ngăn chặn, bỏ ngăn chặn địa chỉ ip chưa được tối ưu, vẫn có thể chặn nhầm nhiều.
* Hệ thống API xử lý với tốc độ chưa nhanh.

**Hướng phát triển trong tương lai:**

* Phát triển hệ thống phân tán bao gồm nhiều nút elasticsearch để có thể lưu trữ và xử lý dữ liệu với tốc độ nhanh.
* Xây dựng, tối ưu hệ thống API để có thể xử lý được nhiều yêu cầu song song và xử lý các yêu cầu được nhanh hơn.
* Có thể kết hợp được một số các công cụ phân tích mã độc vào hệ thống và có thể xây dựng ra các chính sách để tự động ngăn chặn các dấu hiệu như ip một cách chính xác hơn.
* Phát triển hệ thống để có thể tự động phân tích và ngăn chặn nhiều kiểu dấu hiệu tấn công hơn nữa.

# TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | "Why Is Security Monitoring Important?," 2021. [Online]. Available: https://www.iomart.com/updates/why-is-security-monitoring-important-iomart/. [Accessed 29 9 2022]. |
| [2] | Y. Jaiswar, "Cybersecurity Monitoring: Importance, Tools, Process," 2022. [Online]. Available: https://www.knowledgehut.com/blog/security/cybersecurity-monitoring. [Accessed 30 09 2022]. |
| [3] | "Intrusion Detection System (IDS)," GeeksForGeeks, 2022. [Online]. Available: https://www.geeksforgeeks.org/intrusion-detection-system-ids/. [Accessed 29 09 2022]. |
| [4] | J. A. J. Beaumont, "HIDS - A Guide To Host Based Intrusion Detection Systems," Bulletproof, 2021. [Online]. Available: https://www.bulletproof.co.uk/blog/host-based-intrusion-detection-systems. [Accessed 01 10 2022]. |
| [5] | Hoàng Xuân Dậu và Nguyễn Thị Thanh Thủy. Thủy, in *Bài giảng Cơ sở an toàn thông tin*, Học viện Công nghệ Bưu chính Viễn Thông, 2016. |
| [6] | S. COOPER, "10 Best Network Intrusion Detection Systems Software & NIDS Tools," 2022. [Online]. Available: https://www.comparitech.com/net-admin/nids-tools-software/. [Accessed 11 19 2022]. |
| [7] | S. COOPER, "6 best HIDS tools," 2022. [Online]. Available: https://www.comparitech.com/net-admin/hids-tools-software/. [Accessed 19 11 2022]. |
| [8] | X. Bellekens, "What are Cyber Threat Intelligence Feeds ?," Lupovis, 2022. [Online]. Available: https://www.lupovis.io/what-are-cyber-threat-intelligence-feeds/. [Accessed 07 10 2022]. |
| [9] | K. Kent and M. Souppaya, in *Guide to Computer Security*, National Institute of Standards and Technology, 2006. |
| [10] | "Graylog Documentation," Graylog, [Online]. Available: https://go2docs.graylog.org/. [Accessed 11 2022]. |
| [11] | "FastAPI," [Online]. Available: https://fastapi.tiangolo.com/. [Accessed 15 12 2022]. |
| [12] | "Filebeat overview," [Online]. Available: https://www.elastic.co/guide/en/beats/filebeat/current/filebeat-overview.html. [Accessed 02 12 2022]. |
| [13] | "VirusTotal," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/VirusTotal. [Accessed 16 12 2022]. |
| [14] | "What is SIEM? Complete Beginner’s Guide (intellipaat.com)," Intellipaat, 2022. [Online]. Available: https://intellipaat.com/blog/what-is-siem/. [Accessed 28 09 2022]. |
| [15] | K. Gast, "What is SIEM? And How Does it Work?," 2021. [Online]. Available: https://logrhythm.com/blog/what-is-siem/. [Accessed 29 09 2022]. |
| [16] | "Security event manager," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Security\_event\_manager. [Accessed 06 10 2022]. |
| [17] | "What Is SIEM, Why Is It Important and How Does It Work?," exabeam, [Online]. Available: https://www.exabeam.com/explainers/siem/what-is-siem/. [Accessed 08 10 2022]. |
| [18] | R. S. Jr, "What Is Cybersecurity Monitoring And Why Do You Need It?," 2021. [Online]. Available: https://www.clearnetwork.com/cybersecurity-monitoring-why-you-need/. [Accessed 09 10 2022]. |
| [19] | R. Mohanan, "What Is Security Information and Event Management (SIEM)? Definition, Architecture, Operational Process, and Best Practices," 2022. [Online]. Available: https://www.spiceworks.com/it-security/vulnerability-management/articles/what-is-siem/. [Accessed 09 10 2022]. |
| [20] | "The Essential Guid To SIEM," exabeam, [Online]. Available: https://www.exclusive-networks.com/ie/wp-content/uploads/sites/19/2021/07/The-Essential-Guide-to-SIEM.pdf. [Accessed 08 10 2022]. |
| [21] | "Suricata User Guide," Open Information Security Foundation, [Online]. Available: https://suricata.readthedocs.io/en/suricata-6.0.9/. [Accessed 11 2022]. |
| [22] | "How To Configure Suricata as an Intrusion Prevention System (IPS) on Ubuntu 20.04," DigitalOcean, 2021. [Online]. Available: https://www.digitalocean.com/community/tutorials/how-to-configure-suricata-as-an-intrusion-prevention-system-ips-on-ubuntu-20-04. [Accessed 11 10 2022]. |
| [23] | P. M. Hoang, "Hướng dẫn cơ bản framework FastAPI từ A -> Z (Phần 1)," 2020. [Online]. Available: https://viblo.asia/p/huong-dan-co-ban-framework-fastapi-tu-a-z-phan-1-V3m5W0oyKO7. [Accessed 11 11 2022]. |
| [24] | Jay, "Post JSON to FastAPI," 2021. [Online]. Available: https://codehandbook.org/post-json-fastapi/. [Accessed 25 12 2022]. |