ELASTICSEARCH SIEM: TỔNG QUAN, KIẾN TRÚC VÀ CÁCH HOẠT ĐỘNG

1. Elasticsearch SIEM là gì?

Elasticsearch SIEM là một giải pháp quản lý và phân tích sự kiện an ninh (Security Information and Event Management - SIEM) được xây dựng dựa trên nền tảng của Elastic Stack (Elasticsearch, Logstash, Kibana, Beats/Elastic Agent). Mục tiêu chính của nó là:

- **Thu thập dữ liệu an ninh:** Tích hợp dữ liệu log và các sự kiện an ninh từ nhiều nguồn khác nhau như máy chủ, ứng dụng, thiết bị mạng, endpoint, …
- **Phân tích và phát hiện mối đe dọa:** Sử dụng các kỹ thuật tìm kiếm, truy vấn mạnh mẽ của Elasticsearch, kết hợp với các quy tắc (rules) và các mô hình machine learning để xác định các bất thường và dấu hiệu của tấn công mạng.
- **Điều tra và phản ứng:** Cung cấp giao diện trực quan (thông qua Kibana) cho các chuyên gia bảo mật để điều tra, quản lý các vụ việc (incident), và đưa ra các biện pháp phản ứng kịp thời.
- **Giao diện trực quan:** Kibana cung cấp bảng điều khiển mạnh mẽ để giám sát và điều tra các sự kiện bảo mật.
- **Tích hợp Machine Learning:** Nhận diện hành vi bất thường dựa trên AI/ML để phát hiện các cuộc tấn công zero-day.

Nói cách khác, Elasticsearch SIEM là một công cụ tập trung dữ liệu an ninh, cho phép người dùng theo dõi, phân tích và phát hiện sớm các mối đe dọa nhằm bảo vệ hệ thống khỏi các tấn công mạng.

2. Kiến trúc tổng quan của Elasticsearch SIEM

Kiến trúc của Elasticsearch SIEM được xây dựng dựa trên các thành phần chủ yếu của Elastic Stack, kết hợp thêm các chức năng chuyên biệt phục vụ an ninh mạng:

- **Thu thập dữ liệu - Elastic Agent/Beat:** Các agent nhẹ được cài đặt trên các hệ thống, máy chủ, thiết bị để thu thập log, metric và các sự kiện an ninh.
- **Xử lý dữ liệu:**
  - **Logstash:** Đóng vai trò xử lý, chuyển đổi và định dạng dữ liệu từ nhiều nguồn khác nhau trước khi gửi vào Elasticsearch.
  - **Ingest Node**: Cung cấp pipeline xử lý nhẹ trong Elasticsearch.
- **Lưu trữ và tìm kiếm – Elasticsearch:** Là kho lưu trữ trung tâm của toàn bộ dữ liệu an ninh. Dữ liệu được phân mảnh (sharding) và sao lưu (replication) trên nhiều node nhằm đảm bảo tính sẵn sàng và hiệu năng truy vấn cao.
- **Trực quan hóa và phân tích - Kibana:** Giao diện trực quan giúp hiển thị dữ liệu thông qua dashboard, biểu đồ, và bảng báo cáo. Trong đó, ứng dụng SIEM của Kibana cung cấp các công cụ điều tra (investigation) và quản lý vụ việc (case management).
- **Tích hợp các chức năng bổ trợ:**

- **Detection Engine và Machine Learning:** Áp dụng các quy tắc, mô hình học máy để phân tích hành vi bất thường và tự động tạo cảnh báo khi có dấu hiệu tấn công.
- **Threat Intelligence và Security Integrations:** Tích hợp thông tin từ các nguồn bên ngoài để so sánh và phát hiện các mối đe dọa dựa trên các chỉ số (indicators) đã biết. Hỗ trợ SOAR (Security Orchestration, Automation, and Response) để tự động hóa phản ứng bảo mật.

Tổng thể, kiến trúc của Elasticsearch SIEM cho phép thu thập dữ liệu từ nhiều nguồn, xử lý dữ liệu một cách hiệu quả, lưu trữ dưới dạng phân tán và cung cấp công cụ phân tích, cảnh báo nhằm giúp các chuyên gia bảo mật nắm bắt được tình hình an ninh mạng trong thời gian thực.

3. Elasticsearch SIEM hoạt động như thế nào?

Quá trình hoạt động của Elasticsearch SIEM được chia thành các bước chính:

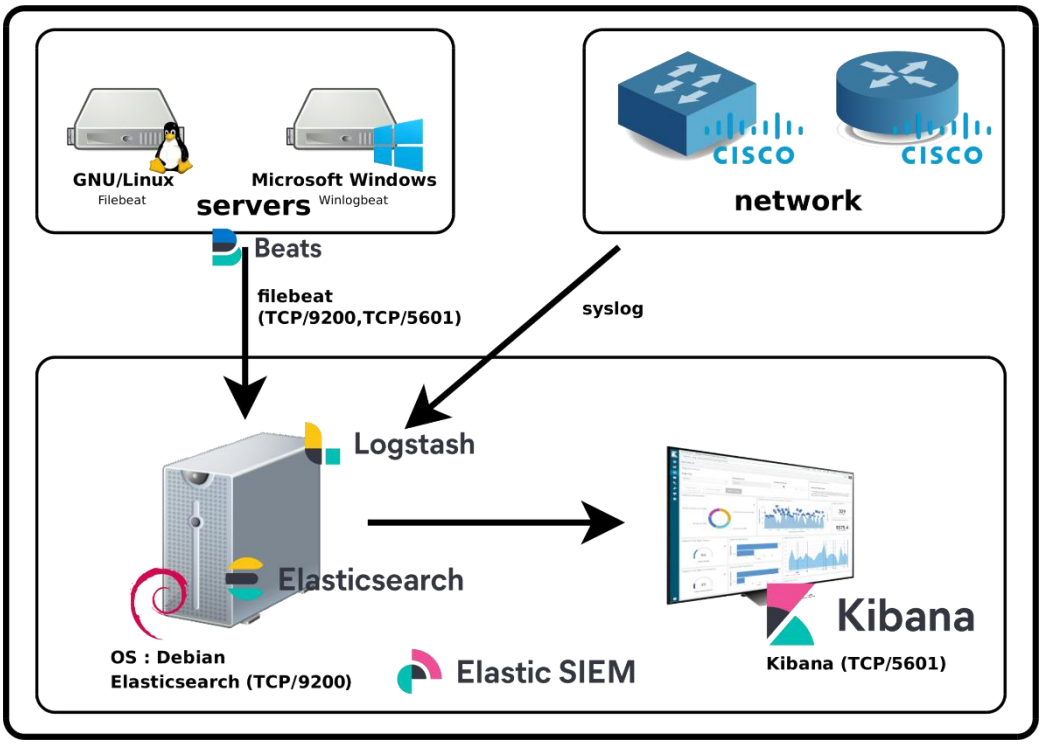
1. **Thu thập dữ liệu:**
   - Dữ liệu an ninh được thu thập từ nhiều nguồn (máy chủ, thiết bị mạng, ứng dụng, endpoint, v.v.) thông qua các agent như Elastic Agent hoặc Beats.
   - Logstash hoặc Ingest Node có thể được sử dụng để tiền xử lý, chuyển đổi định dạng và làm sạch dữ liệu nếu cần thiết.
2. **Lưu trữ và lập chỉ mục:**
   - Dữ liệu thu thập được gửi vào Elasticsearch, nơi nó được lập chỉ mục (index) dưới dạng các document JSON.
   - Elasticsearch đảm bảo khả năng tìm kiếm nhanh và phân tán nhờ vào cơ chế sharding và replication.
3. **Phân tích và phát hiện:**
   - **Công cụ truy vấn:** Các truy vấn được thực hiện thông qua Elasticsearch Query DSL, cho phép tìm kiếm và phân tích dữ liệu sâu.
   - **Detection Engine:** Các quy tắc (rules) và mô hình machine learning được áp dụng để phân tích các mẫu hành vi bất thường, giúp phát hiện các dấu hiệu tấn công hoặc sự cố an ninh.
   - **Correlation:** Các sự kiện từ các nguồn khác nhau được kết hợp lại để xác định mối liên hệ giữa các sự kiện, từ đó phát hiện các vụ việc an ninh phức tạp.
4. **Trực quan hóa và điều tra:**
   - **Kibana SIEM App:** Giao diện Kibana cung cấp dashboard chuyên biệt cho SIEM, cho phép hiển thị các cảnh báo, biểu đồ phân tích và các báo cáo về tình hình an ninh.

- **Case Management:** Các chuyên gia bảo mật có thể sử dụng các công cụ điều tra để xem chi tiết sự kiện, truy vết nguồn gốc của sự cố và quản lý các vụ việc an ninh.

5. **Phản ứng và cảnh báo:**
   - Khi một quy tắc hoặc mô hình phát hiện hành vi bất thường được kích hoạt, hệ thống tự động tạo ra cảnh báo.
   - Các cảnh báo này có thể được tích hợp với các hệ thống quản lý sự cố (Incident Response) hoặc gửi thông báo qua email, SMS, hoặc các công cụ quản lý khác để hỗ trợ việc phản ứng kịp thời.

Tóm lại:

- **Elasticsearch SIEM** là giải pháp SIEM tích hợp sẵn trong Elastic Stack, tập trung vào việc thu thập, lưu trữ và phân tích dữ liệu an ninh để phát hiện sớm và phản ứng với các mối đe dọa.
- **Kiến trúc tổng quan** gồm các thành phần thu thập dữ liệu (Elastic Agent/Beats, Logstash), lưu trữ và tìm kiếm (Elasticsearch) và trực quan hóa, điều tra (Kibana với ứng dụng SIEM), kèm theo các chức năng bổ trợ như machine learning, detection engine và threat intelligence.
- **Cách thức hoạt động** của hệ thống là từ việc thu thập và xử lý dữ liệu an ninh, lập chỉ mục dữ liệu, phân tích thông qua các quy tắc và mô hình học máy, trực quan hóa qua Kibana, cho đến việc tạo ra cảnh báo và hỗ trợ quản lý vụ việc an ninh.



Minh hoạ Kiến trúc của ElasticSearchSIEM

Ví dụ: Phát hiện tấn công brute force đăng nhập

1. **Thu thập dữ liệu:**
    - **Elastic Agent/Beats:**
        - Các agent được cài đặt trên máy chủ web và máy chủ ứng dụng để thu thập log đăng nhập.
        - Ví dụ: Filebeat được cấu hình để đọc file log chứa các sự kiện đăng nhập, bao gồm thông tin như địa chỉ IP, tên người dùng, trạng thái đăng nhập (thành công hoặc thất bại), và thời gian.
    - **Logstash (tuỳ chọn):** Nếu cần, Logstash có thể tiền xử lý log (ví dụ: trích xuất thông tin, chuẩn hóa định dạng) trước khi chuyển dữ liệu vào Elasticsearch.
2. **Lưu trữ và lập chỉ mục:**
    - **Elasticsearch:**
        - Dữ liệu log được gửi vào Elasticsearch và lưu trữ dưới dạng các document JSON trong một index (ví dụ: web-login-logs-*).
        - Cơ chế phân mảnh (sharding) và sao lưu (replication) đảm bảo dữ liệu luôn sẵn sàng và truy vấn nhanh.
3. **Phân tích và phát hiện:**
    - **Detection Engine trong SIEM:**
        - Một quy tắc (rule) được thiết lập trong Elasticsearch SIEM, chẳng hạn: "Nếu có hơn 5 lần đăng nhập thất bại từ cùng một địa chỉ IP trong vòng 1 phút, thì kích hoạt cảnh báo."
        - Hệ thống sử dụng **Elasticsearch Query DSL** để liên tục quét dữ liệu log, xác định các mẫu bất thường dựa trên các tiêu chí đã định nghĩa.
    - **Machine Learning (nếu áp dụng):** Ngoài quy tắc thủ công, một job machine learning có thể được thiết lập để phân tích hành vi đăng nhập theo thời gian, phát hiện những thay đổi đột ngột hoặc các xu hướng bất thường không rõ ràng từ quy tắc tĩnh.
4. **Trực quan hóa và cảnh báo:**
    - **Kibana SIEM App:**
        - Các cảnh báo được hiển thị trên dashboard của Kibana SIEM, với thông tin chi tiết như địa chỉ IP nguồn, số lần đăng nhập thất bại, thời gian xảy ra sự kiện và các thông tin liên quan.
        - Giao diện trực quan cho phép chuyên gia bảo mật nhấn vào cảnh báo để xem chi tiết, truy xuất các log liên quan, và thực hiện phân tích sâu hơn.
    - **Thông báo:** Hệ thống có thể tích hợp với email, SMS hoặc các công cụ quản lý sự cố để gửi thông báo ngay khi cảnh báo được kích hoạt.

5. **Điều tra và phản ứng:**

- **Phân tích sự kiện:**
  - Chuyên gia bảo mật truy cập vào Kibana để xem lại toàn bộ chuỗi sự kiện, đánh giá xem đây có phải là tấn công brute force thật sự hay không.
  - Họ có thể sử dụng các bộ lọc và query nâng cao trong **Kibana Query Language (KQL)** để truy xuất các log liên quan, so sánh với các sự kiện khác và xác nhận vụ việc.
- **Hành động:** Nếu xác định là tấn công, các biện pháp phản ứng như chặn địa chỉ IP, tăng cường kiểm soát truy cập hoặc thông báo cho bộ phận an ninh sẽ được triển khai ngay lập tức.

Tóm Lại:

- **Thu thập dữ liệu:** Elastic Agent/Beats thu thập log đăng nhập từ máy chủ web.
- **Lưu trữ:** Dữ liệu được gửi đến Elasticsearch, được lập chỉ mục và phân mảnh.
- **Phân tích:** Detection Engine trong Elasticsearch SIEM sử dụng quy tắc và mô hình học máy để phát hiện mẫu đăng nhập thất bại liên tục.
- **Trực quan hóa & cảnh báo:** Kibana SIEM hiển thị cảnh báo, cung cấp công cụ điều tra và thông báo cho chuyên gia bảo mật.
- **Phản ứng:** Chuyên gia bảo mật điều tra chi tiết và thực hiện các biện pháp phòng ngừa hoặc xử lý sự cố.

Tài liệu tham khảo:

- Elastic Security Documentation
- MITRE ATT&CK Framework