TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN – ĐHQG-HCM
Khoa Mạng máy tính & Truyền thông

SIEM và Đánh giá IDPS

NT204 – Hệ thống tìm kiếm, phát hiện và ngăn ngừa xâm nhập

GV: Đỗ Hoàng Hiển

hiendh@uit.edu.vn

Nội dung hôm nay…

Hôm nay có gì?

- **SIEM**
- **Đánh giá IDPS**

Nội dung và Mục tiêu

- **1. Security Information and Event Management**
- 2. Đánh giá IDPS

Quản lý Log

- **Quản lý log – Log Management (LM)** là hướng tiếp cận xử lý một số lượng lớn các log được tạo ra từ các thiết bị (*cũng được gọi là audit record, audit trail, log sự kiện, v.v…)*
- **LM** bao gồm các chức năng thu thập log, tích hợp tập trung, lưu trữ lâu dài, phân tích log (theo real-time và theo nhóm sau khi lưu trữ) cũng như tìm kiếm log và tạo báo cáo

Các thách thức

- Phân tích log để liên kết các sự kiện bảo mật có liên quan
- Thu tập log tập trung
- Đáp ứng các yêu cầu về tuân thủ CNTT
- Phân tích được hiệu quả nguyên nhân gốc rễ của sự kiện
- Làm các dữ liệu log có ý nghĩa hơn
- Theo dõi các hành vi người dùng đáng ngờ

Nhắc lại

SIEM

- Security Information and Event Management.
- Các chức năng: Thu thập log, Chuẩn hoá, Tương quan log, Tích hợp, Báo cáo, Tuân thủ.
- Khả năng: Thu thập, phân tích và biểu diễn các thông tin từ các nguồn:
  - Mạng và các thiết bị bảo mật
  - Các ứng dụng quản lý định danh và truy cập
  - Các công cụ quản lý lỗ hổng và tuân thủ chính sách
  - Log của các hệ điều hành, CSDL và ứng dụng
  - Các dữ liệu về các mối đe doạ bên ngoài

Nhắc lại

SIEM

- SIEM = phần mềm và dịch vụ bao gồm security information management (SIM) và security event management (SEM)
  - SEM: giám sát thời gian thực, liên kết sự kiện, cảnh báo
  - SIM: lưu trữ lâu dài, phân tích và báo cáo dữ liệu log
- Mục tiêu chính
  - Xác định mối đe doạ và các vi phạm có thể có
  - Thu thập các log để giám sát an ninh và tuân thủ chính sách
  - Thực hiện điều tra và cung cấp chứng cứ

So sánh

SIEM vs LM

Chức năng	Security Information and Event Management (SIEM)	Log Management (LM)
Thu thập log	Thu thập các log an ninh liên quan + dữ liệu có ngữ cảnh	Thu thập tất cả các log
Tiền xử lý log	Phân tích định dạng, chuẩn hoá, phân loại, thêm thông tin	Gán chỉ số index, phân tích định dạng hoặc không
Lưu trữ log	Log đã được phân tích và chuẩn hoá	Dữ liệu log thô
Báo cáo	Báo cáo liên quan đến bảo mật	Báo cáo mục đích chung, có thể sử dụng rộng rãi
Phân tích	Tương quan, đánh giá mối đe doạ, sắp xếp ưu tiên các sự kiện	Phân tích text, gán nhãn
Cảnh báo và thông báo	Báo cáo nâng cao liên quan đến bảo mật	Cảnh báo đơn giản trên toàn bộ log
Các chức năng khác	Quản lý sự cố, phân tích luồng hoạt động, phân tích ngữ cảnh, v.v…	Khả năng mở rộng cao trong thu thập và lưu trữ

Vì sao SIEM lại cần thiết?

- Số lượng các vi phạm chính sách ngày càng tăng do mối đe doạ bên trong và bên ngoài
- Kẻ tấn công ngày càng tinh vi và các công cụ an ninh truyền thống là chưa đủ
- Giảm thiểu các tấn công mạng phức tạp
- Quản lý số lượng log lớn từ nhiều nguồn
- Đáp ứng các yêu cầu nghiêm ngặt về tuân thủ chính sách

SIEM

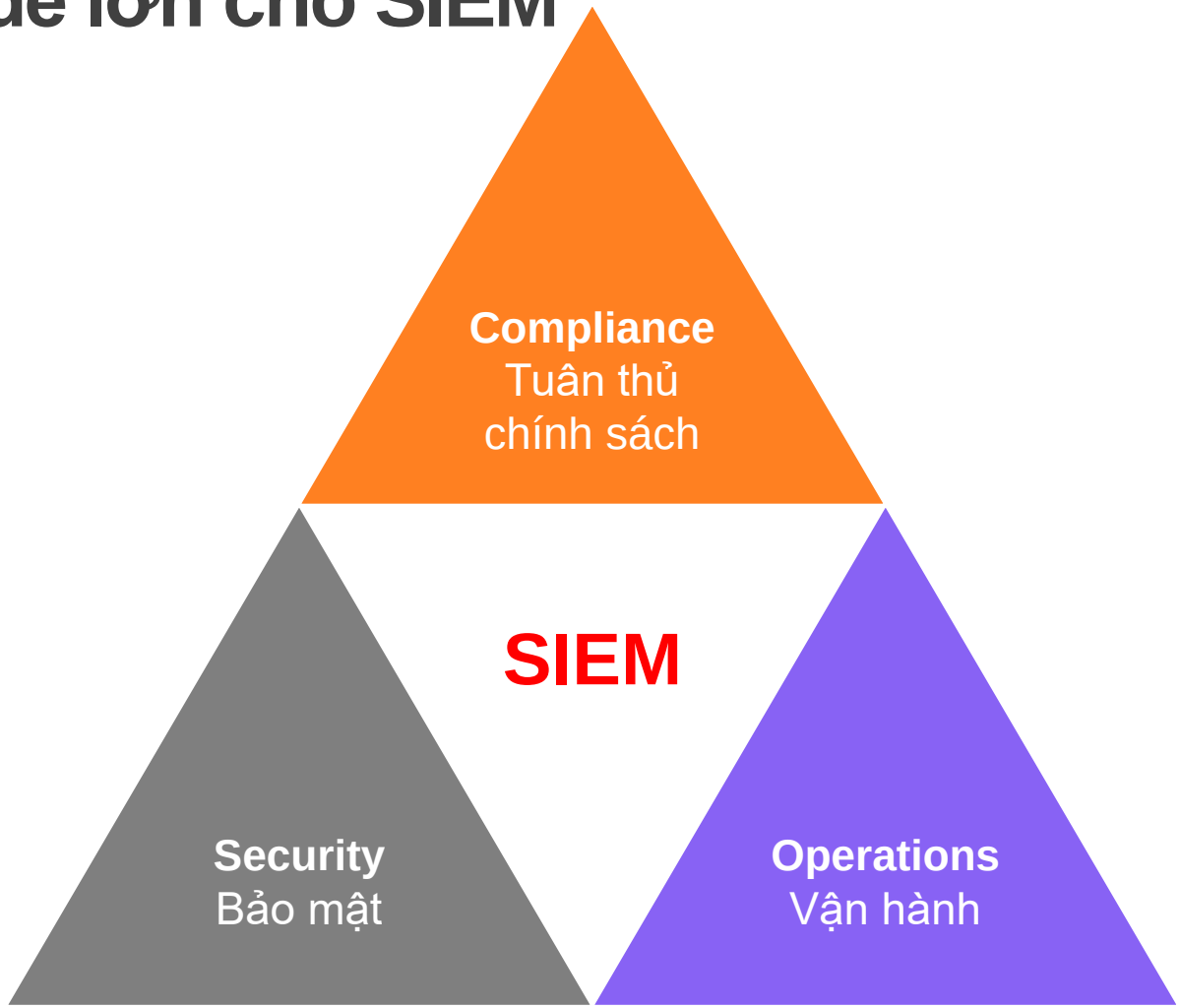
Các thành phần

- Các sự kiện giám sát được
- Bộ thu thập sự kiện
- Engine lõi
- Giao diện người dùng

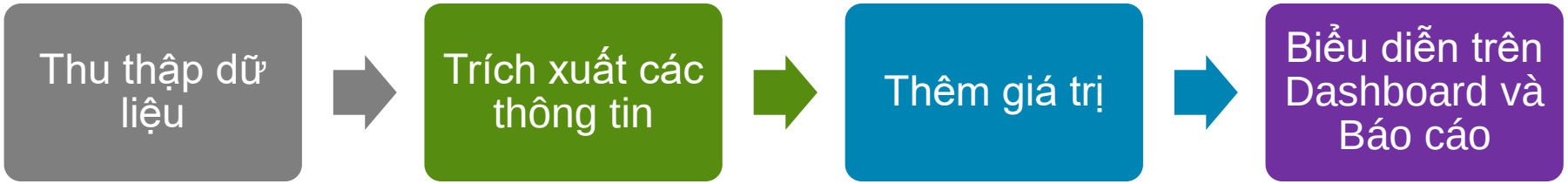
SIEM

Các chức năng tiêu biểu

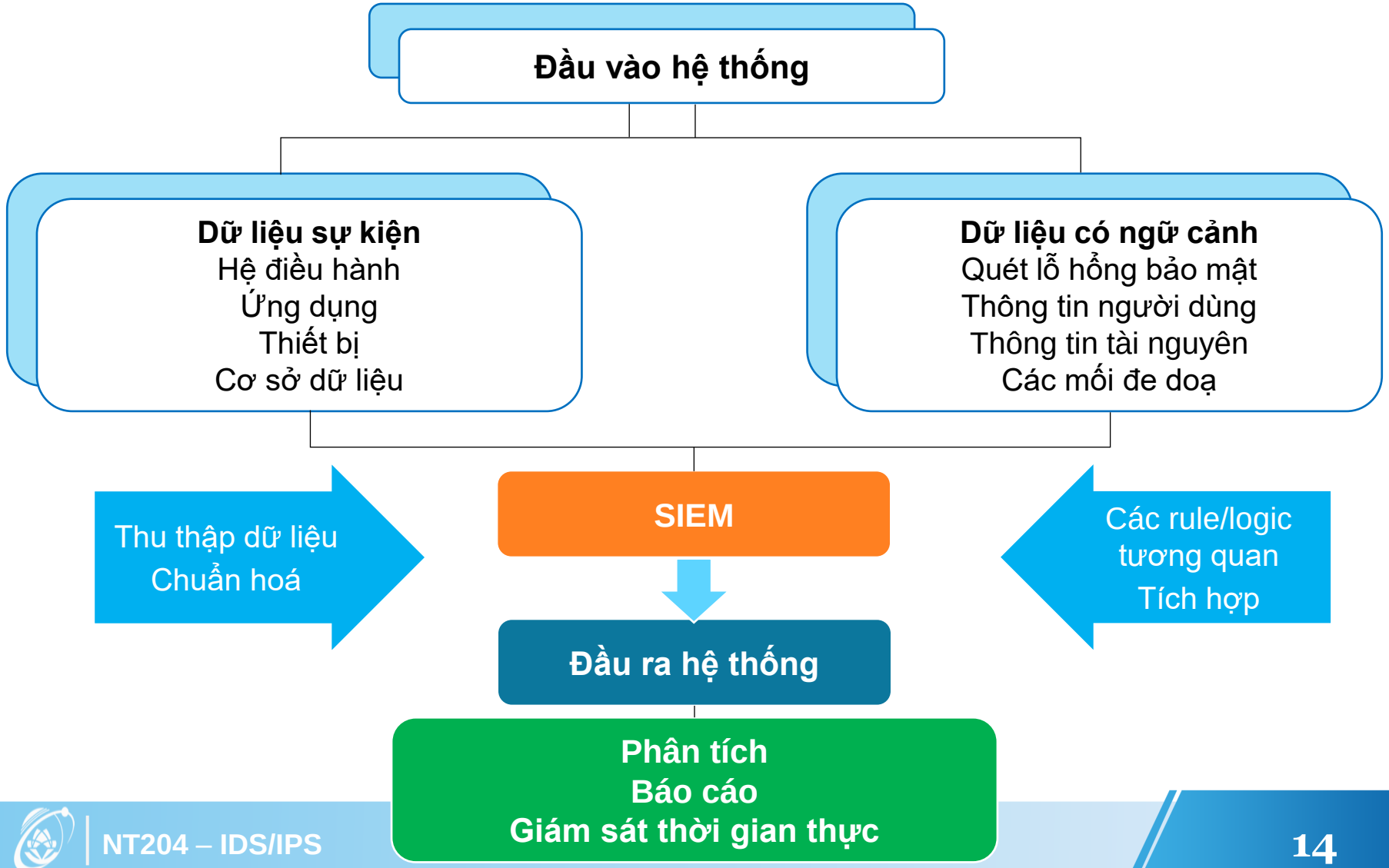
3 vấn đề lớn cho SIEM



Quy trình xử lý của SIEM

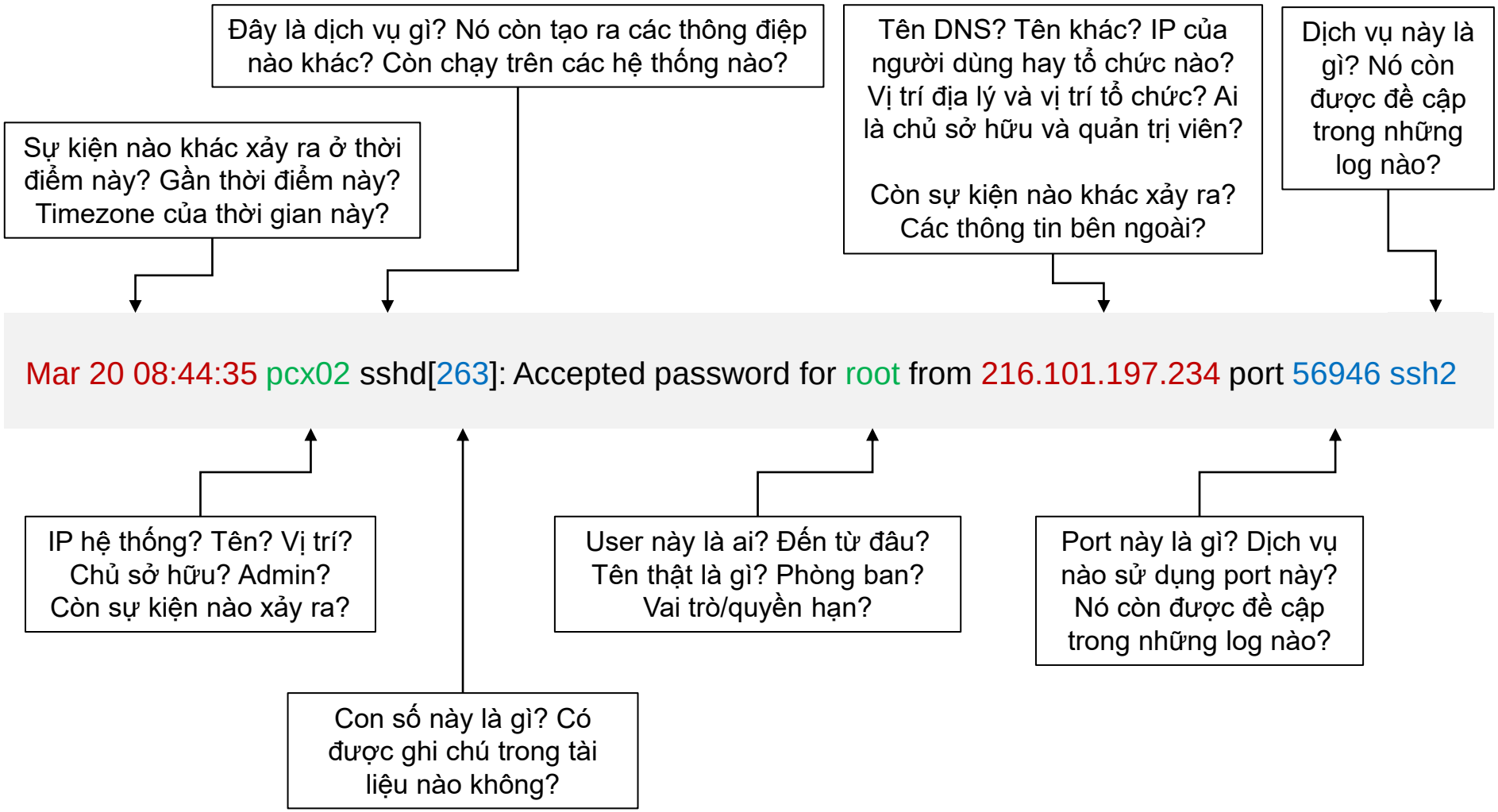


Kiến trúc của SIEM



SIEM

Ngữ cảnh



Thêm ngữ cảnh

- Ví dụ của ngữ cảnh
  - Thêm thông tin vị trí địa lý
  - Lấy thông tin từ DNS server
  - Lấy thông tin user (tên đầy đủ, công việc…)
- Thêm một số ngữ cảnh hỗ trợ việc xác định
  - Truy cập từ nước ngoài
  - Truyền dữ liệu đáng ngờ

8 chức năng quan trọng của SIEM

1. Thu thập log

- Thu thập log từ nhiều nguồn
  - Windows/Linux, ứng dụng, CSDL, router, switch và các thiết bị khác
- Phương pháp thu thập log: agent-based hoặc agent-less
- Thu thập log tập trung
- **Event Per Second (EPS)** – tốc độ gửi sự kiện của hạ tầng CNTT
  - Nếu không được tính toán phù hợp, SIEM có thể drop 1 số sự kiện trước khi lưu trên CSDL, dẫn đến các báo cáo, kết quả tìm kiếm, cảnh báo và tương quan không chính xác

8 chức năng quan trọng của SIEM (tt)

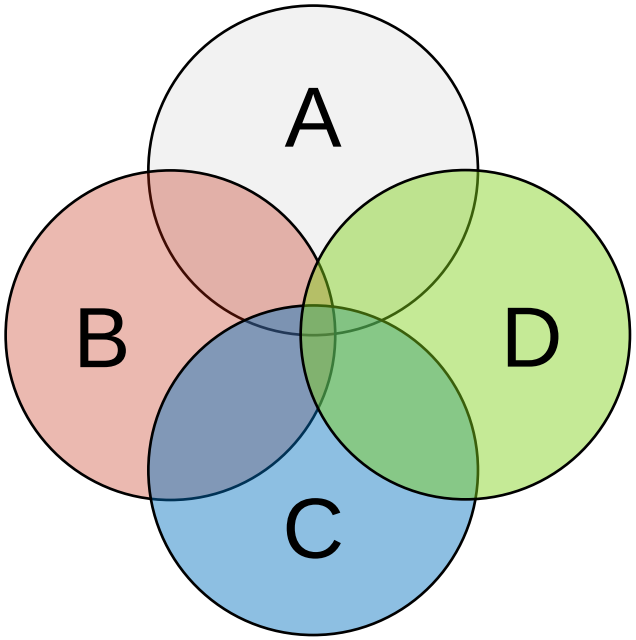
2. Giám sát hoạt động người dùng

- SIEM cần giám sát hoạt động người dùng, quyền hạn người dùng và báo cáo giám sát
- Đảm bảo SIEM cung cấp "Complete audit trail" – truy vết đầy đủ
  - Biết user nào thực hiện hành động, kết quả của hành động, diễn ra trên server nào, thiết bị người dùng nào đã kích hoạt hành động, v.v

8 chức năng quan trọng của SIEM (tt)

3. Tương quan sự kiện thời gian thực

- Hỗ trợ chủ động đối phó với các mối đe doạ
- Tăng bảo mật mạng bằng cách xử lý hàng triệu sự kiện đồng thời để phát hiện sự kiện đáng ngờ trên mạng
- Có thể dựa trên tìm kiếm log, các rule và cảnh báo
  - Các rule và cảnh báo được định nghĩa trước không hiệu quả. SIEM cần hỗ trợ các trình tạo rule và alert tuỳ chỉnh
  - Đảm bảo quá trình tương quan sự kiện thực hiện dễ dàng



8 chức năng quan trọng của SIEM (tt)

4. Lưu trữ log

- Tự động lưu trữ tất cả dữ liệu log từ các hệ thống, thiết bị và ứng dụng vào 1 kho lưu trữ **tập trung**
- Cần đảm bảo tính năng **Tamper Proof – mã hoá** và **gán nhãn thời gian** log để phục vụ tuân thủ chính sách và điều tra pháp chứng
- Hỗ trợ truy xuất và phân tích log đã lưu

8 chức năng quan trọng của SIEM (tt)

5. Báo cáo tuân thủ chính sách

- Tuân thủ chính sách là cốt lõi của SIEM
- Đảm bảo khả năng báo cáo tuân thủ chính sách như PCI DSS, FISMA, GLBA, SOX, HIPPA, v.v
- Cần có khả năng tuỳ chỉnh và dựng báo cáo tuân thủ mới để đáp ứng các hoạt động quản lý trong tương lai

8 chức năng quan trọng của SIEM (tt)

6. Giám sát tính toàn vẹn của file

- Hỗ trợ các chuyên gia bảo mật giám sát các file và thư mục quan trọng trong tổ chức
- SIEM cần giám sát và báo cáo tất cả thay đổi diễn ra như tạo, truy cập, xem, xoá, thay đổi, đổi tên file/thư mục, v.v
- SIEM cũng cần gửi các cảnh báo real-time khi có truy cập trái phép vào các file hoặc tập tin quan trọng

8 chức năng quan trọng của SIEM (tt)

7. Pháp chứng số trên log

- SIEM cần cho phép người dùng truy vết hành vi xâm nhập hoặc hoạt động của 1 sự kiện nào đó với khả năng tìm kiếm log
- Chức năng tìm kiếm log cần trực quan và thân thiện với người dùng, cho phép quản trị viên tìm kiếm nhanh chóng trong các dữ liệu log thô

8 chức năng quan trọng của SIEM (tt)

8. Dashboard

- Dashboard hỗ trợ quản trị viên thực hiện các hành động kịp thời và đưa ra quyết định cho các sự kiện đáng ngờ
- Dữ liệu an ninh cần được biểu diễn 1 cách trực quan và thân thiện với người dùng
- Dashboard cần hỗ trợ tuỳ chỉnh để quản trị viên có thể cấu hình các thông tin bảo mật cần quan sát

Vì sao triển khai SIEM lỗi

- **Không có kế hoạch**
  - Không định nghĩa trước phạm vi
- **Chiến lược triển khai lỗi**
  - Thu thập và quản lý dữ liệu log không liên tục
  - Số lượng dữ liệu không liên quan có thể làm hệ thống quá tải
- **Vận hành**
  - Thiếu giám sát, quản lý
  - Giả định plug and play

Bảo mật là một quá trình, không phải là một sản phẩm

Lợi ích thương mại

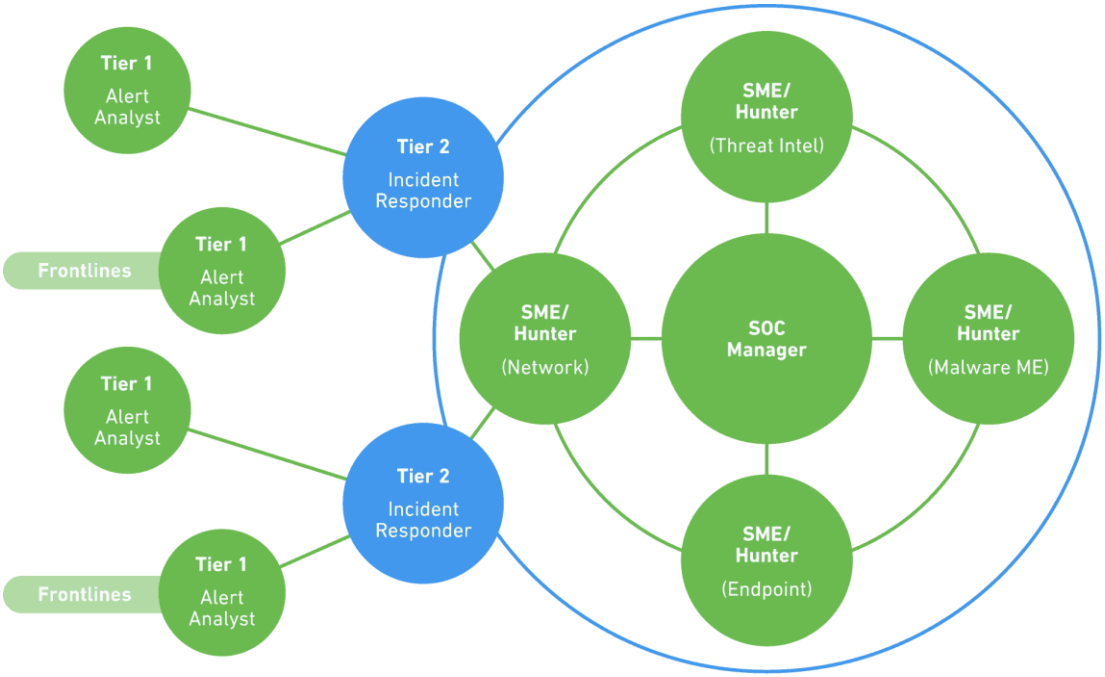
- Giám sát thời gian thực
- Tiết kiệm kinh phí
- Tuân thủ chính sách
- Khả năng báo cáo
- ROI (Return on Investment – tỷ suất hoàn vốn) nhanh chóng

Phân tích dữ liệu tấn công

Nội dung và Mục tiêu

- 1. Security Information and Event Management
- 2. Đánh giá IDPS

Sự cần thiết của việc đánh giá cảnh báo

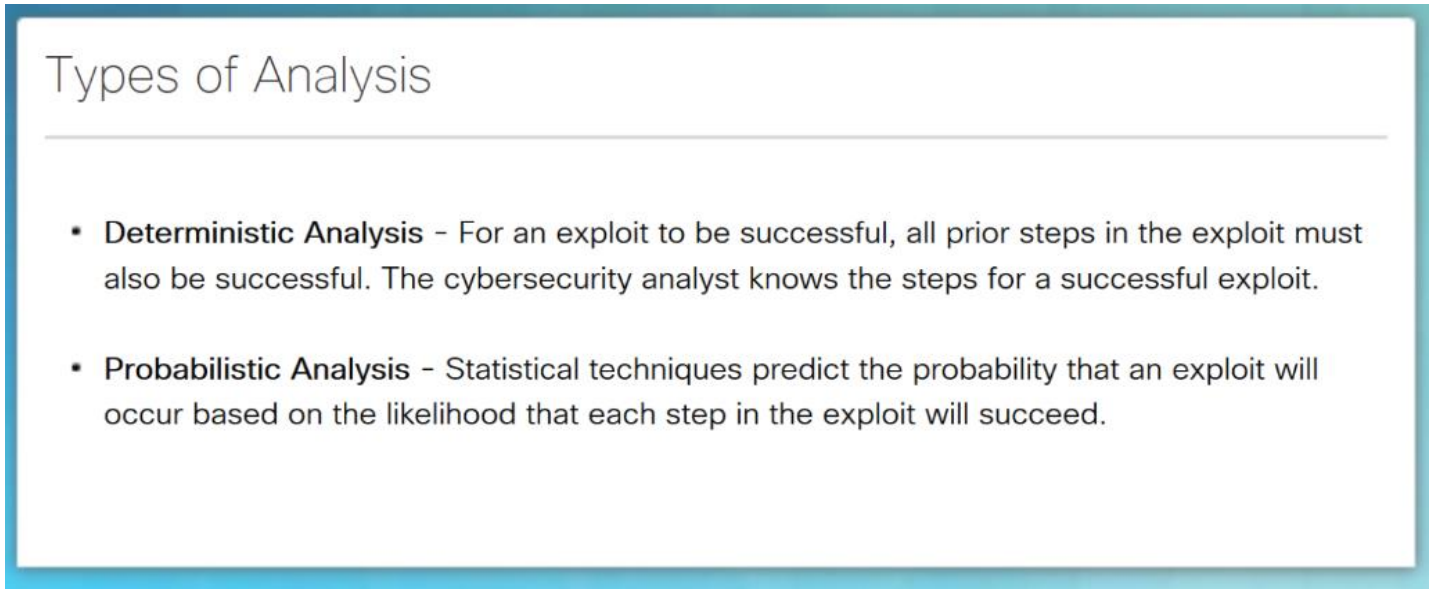


- Các hoạt động khai thác/tấn công, dù tinh vi đến mức nào, đều sẽ cố gắng qua mặt các biện pháp bảo vệ
- Các rule phát hiện cần **rất thận trọng (overly conservative)**
- Cần có các chuyên gia phân tích an ninh mạng có chuyên môn trong việc xem xét các cảnh báo để xác nhận tấn công có thực sự diễn ra hay không

- https://www.exabeam.com/siem-guide/the-soc-secops-and-siem/

Phân tích Xác định và Phân tích Xác suất

- Các kỹ thuật thống kê được dùng để đánh giá những rủi ro có thể xảy ra các tấn công trên 1 vùng mạng cho trước
  - **Phân tích xác định (Deterministic Analysis)** – đánh giá rủi ro dựa trên những hiểu biết về một lỗ hổng
  - **Phân tích xác suất (Probabilistic Analysis)** – ước tính khả năng một khai thác được thực hiện dựa trên khả năng của mỗi bước khai thác thành công



Đánh giá các cảnh báo

- Các cảnh báo có thể được phân loại thành các nhóm sau:
  - **True Positive (TP)**: Các cảnh báo đã được xác nhận thực tế đúng là 1 tấn công
  - **False Positive (FP)**: Các cảnh báo nhưng thực tế không phải là 1 tấn công
  - **True Negative (TN)**: Không có tấn công nào xảy ra
  - **False Negative (FN)**: Một tấn công nhưng không bị phát hiện

When an alert is issued, it will receive one of four possible classifications

	True	False
Positive (Alert exists)	Incident occurred	No incident occurred
Negative (No alert exists)	No incident occurred	Incident occurred

Events classified as 'true" are desired.

Đánh giá IDPS

Các chỉ số đánh giá

- **Accuracy – Độ chính xác:** có bao nhiều trường hợp được xác định đúng (là tấn công hoặc bình thường) trong tổng số các trường hợp.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

- **False positive rate (FPR) và false negative rate (FNR):**
  - FPR xác định tỉ lệ trường hợp bình thường nhưng bị cảnh báo là tấn công.
  - FNR xác định tỉ lệ trường hợp tấn công nhưng không được cảnh báo.

$$FPR = \frac{FP}{FP + TN}, \quad FNR = \frac{FN}{FN + TP}.$$

Wu C, Li W. Enhancing intrusion detection with feature selection and neural network. Int J Intell Syst. 2021;1-19.

Đánh giá IDPS

Các chỉ số đánh giá (tt)

- Để đánh giá chất lượng của một giải pháp phát hiện, chúng ta dựa trên tổ hợp 3 chỉ số đánh giá khác nhau: Recall, Precision và F1-score.
- **Precision:** tỷ lệ các phát hiện chính xác trên tổng số các cảnh báo tấn công IDPS đã tạo ra.

$$Precision = \frac{TP}{TP + FP}$$

- **Recall** *(R - detection rate):* tỷ lệ các phát hiện chính xác trên tổng số tất cả các trường hợp tấn công thực tế đã có.

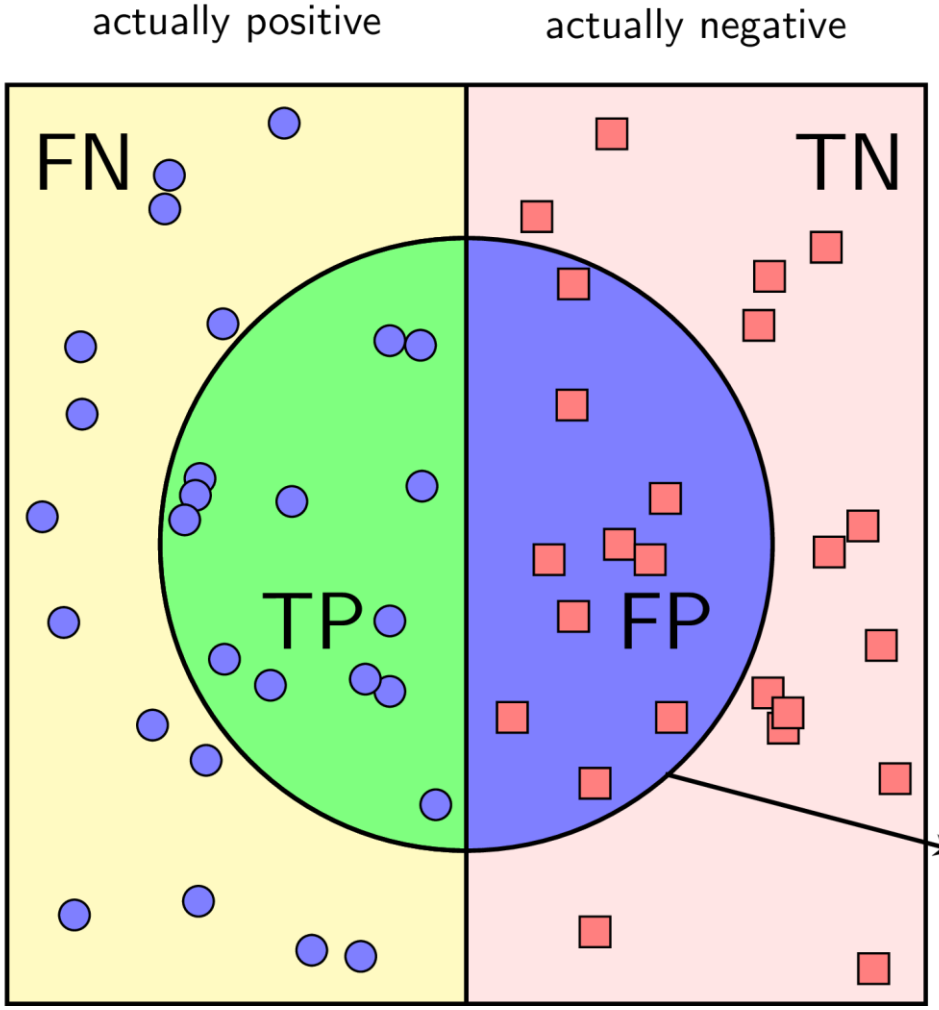
$$Recall = \frac{TP}{TP + FN}$$

- **F1-score:** tổ hợp 2 chỉ số *Precision* và *Recall*

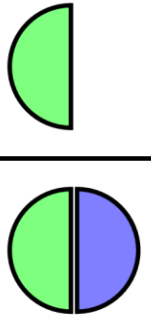
$$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Apruzzese, G., Andreolini, M., Marchetti, M., Colacino, V. G., & Russo, G. (2020). AppCon: Mitigating Evasion Attacks to ML Cyber Detectors. *Symmetry*, *12*

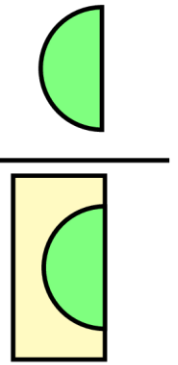
Các chỉ số đánh giá (tt)



$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} =$$



$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} =$$



https://machinelearningcoban.com/

Bài tập vận dụng

Trong một ngày, Inseclab nhận được 100 email và công cụ IDPS mà Inseclab triển khai đã phân tích và cho kết quả như sau:

- Có 30 email được phân loại là spam, trong đó có 26 email là spam và 4 email không phải spam.
- Có 70 email được phân loại là bình thường, trong đó, có 65 email là bình thường và 5 email là spam.

Xác định TP, FP, TN, FN và tính Accuracy, Precision, Recall và F1-score

Chuẩn bị cho tuần sau…

- Hôm nay: **SIEM và Đánh giá IDPS**
  - Assignment 3 – Trích xuất malware từ file pcap
    - Deadline: **23:00 03/11/2022**
- Chuẩn bị cho tuần sau:
  - **Lecture 09 – Đánh giá IDPS và biện pháp phòng tránh**
    - Tài liệu: CEHv10 – Module 12: Evading IDS, Firewalls and Honeypots

Câu hỏi/thắc mắc (nếu có)???

Today end, **See you next week!**