ỨNG DỤNG PHÁT HIỆN HÀNH VI BẤT THƯỜNG CỦA NGƯỜI DÙNG TRONG MẠNG DOANH NGHIỆP BẰNG KHAI PHÁ DỮ LIỆU & HỌC MÁY

Huỳnh Trần Trọng Nghĩa

Trường ĐH Công Nghệ Thông Tin – Đại Học Quốc Gia TP. Hồ Chí Minh

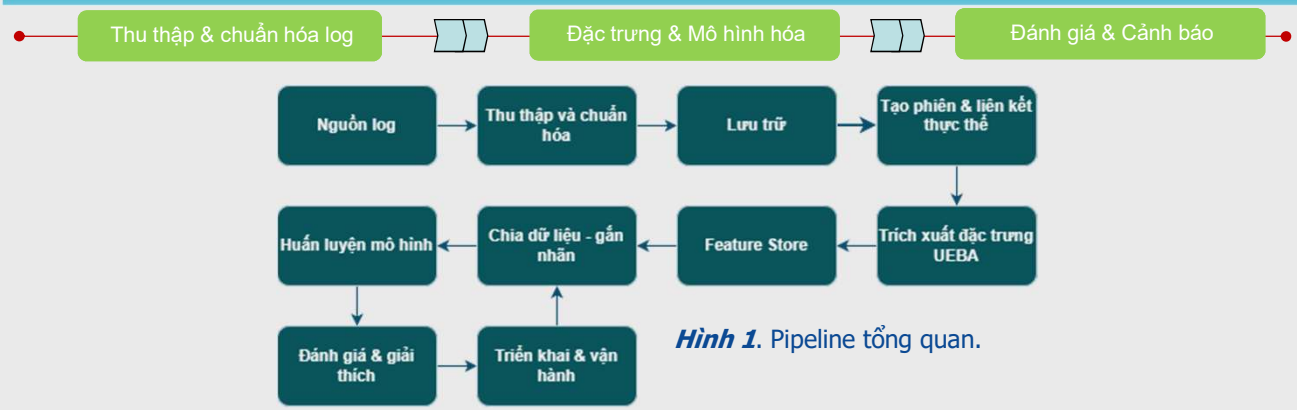
Bài toán & giải pháp?

- Xây dựng ứng dụng UEBA phát hiện hành vi lệch chuẩn từ log doanh nghiệp.
- Chuẩn hoá schema: user–host–action–resource–timestamp; gom theo time window.
- Trích xuất đặc trưng theo ngữ cảnh (tần suất, độ hiếm, chuỗi hành động, peer-group).
- Tạo điểm bất thường + cảnh báo kèm giải thích (feature đóng góp, chuỗi sự kiện).

Động lực & ứng dụng

- Rule/Signature khó bao phủ các chuỗi hành vi "hợp lệ" nhưng bất thường theo ngữ cảnh.
- Phát hiện sớm account takeover, insider threat, privilege misuse, lateral movement.
- Ưu tiên cảnh báo cho SOC (giảm FPR) và hỗ trợ điều tra bằng ngữ cảnh/giải thích.
- Khả năng tái sử dụng pipeline cho nhiều nguồn log (on-prem + cloud).

PIPELINE TỔNG QUAN



Hình 1. Pipeline tổng quan.

Diễn giải

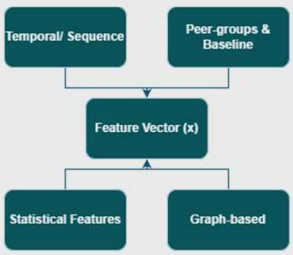
1. Thu thập & chuẩn hóa log

- Nguồn log: AD/SSO, VPN, Proxy, File, Endpoint/EDR (tuỳ chọn cloud/email).
- Parse → mapping schema chuẩn → đồng bộ thời gian → lọc nhiễu/trùng.
- Gom theo Timestamp (5m/1h/1d) tạo event table & feature table.
- Dữ liệu công khai + mô phỏng kịch bản theo MITRE ATT&CK khi thiếu nhãn.

Hình 2. ETL & chuẩn hóa log.

2. Đặc trưng & Mô hình hóa

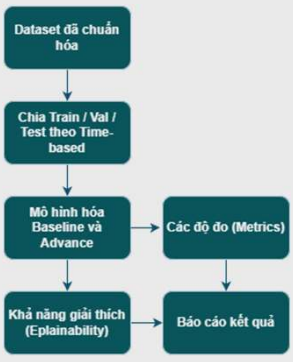
- Thống kê: tần suất đăng nhập, số file truy cập, dung lượng tải xuống, entropy/rarity.
- Chuỗi: n-gram hành động, khoảng cách thời gian, session pattern.
- Peer-group: so sánh theo vai trò/nhóm để học baseline hành vi.
- Mô hình: IF/LOF/OC-SVM; Autoencoder; XGBoost/RF.



Hình 3. Kỹ thuật tạo đặc trưng.

3. Đánh giá & Demo

- Chia train/val/test theo thời gian (tránh data leakage).
- Metrics: PR-AUC/ROC-AUC, F1, FPR, latency.
- Explainability: SHAP/importance; reconstruction error.
- Demo: alerts → drill-down user → timeline → evidence → export.



Hình 4. Mô hình hóa & đánh giá.

NII

Huỳnh Trần Trọng Nghĩa – Trường Đại học Công Nghệ Thông Tin – ĐH Quốc Gia TP. Hồ Chí Minh

TEL : 0899.2105.88 Email : nghiahtt.20@grad.uit.edu.vn