NGHIÊN CỨU VÀ PHÁT TRIỂN GIẢI PHÁP BẢO MẬT THỜI GIAN THỰC CHO HỆ THỐNG KUBERNETES DỰA TRÊN FALCO KẾT HỢP HỌC MÁY

Nguyễn Đăng Phúc Lợi - 250202012

Tóm tắt

- Lớp: CS2205.CH201
- Link Github của nhóm: https://github.com/rubyhcm/CS2205.CH201
- Link YouTube video: https://www.youtube.com/watch?v=VPHQBifxiKg

Nguyễn Đăng Phúc Lợi

Giới thiệu

- Kubernetes đã trở thành nền tảng tiêu chuẩn cho các hệ thống microservices. Tuy nhiên, sự phức tạp này đi kèm với các rủi ro bảo mật nghiêm trọng tại tầng Runtime.
- 87% Container Images chứa lỗ hổng bảo mật mức độ cao hoặc nghiêm trọng.
- Mối đe dọa Runtime: Container escape, Privilege escalation, và khai thác Zero-day.
- Hạn chế của Falco: Phụ thuộc vào luật tĩnh, dẫn đến tỷ lệ cảnh báo giả (False Positives) cao và bỏ sót các mối đe dọa chưa biết.

Mục tiêu

- Giảm False Positives, Phát hiện Zero-day
- Xây dựng cơ chế lai kết hợp bộ luật tĩnh của Falco với mô hình học máy để kiểm chứng chéo, giảm thiểu cảnh báo giả gây nhiễu.
- Ứng dụng thuật toán Anomaly Detection (như Isolation Forest) để nhận diện các hành vi lệch chuẩn mà luật tĩnh không bao phủ được.
- Hiệu năng Thời gian thực
- Thiết kế pipeline xử lý luồng tối ưu, đảm bảo độ trễ thấp và mức tiêu thụ tài nguyên chấp nhận được.

Nội dung và Phương pháp

Kiến trúc Hybrid Analysis

Hệ thống kết hợp giữa giám sát dựa trên luật (Signature-based) và giám sát dựa trên hành vi (Anomaly-based).

- Collection Agent: Sử dụng Falco/eBPF để thu thập syscalls.
- Feature Engineering: Trích xuất đặc trưng N-gram và Time-window statistics.

Nội dung và Phương pháp

Detection Engine:

- Fast Path: Luật Falco cho tấn công đã biết.
- Intelligent Path: Mô hình ML (One-Class SVM) cho bất thường

Nội dung và Phương pháp

Quy trình xử lý dữ liệu

- Thu thập: Falco lắng nghe sự kiện syscall từ Kernel Kubernetes.
- Tiền xử lý: Làm sạch, mã hóa và trích xuất vector đặc trưng.
- Phát hiện: Mô hình ML phân tích vector để tìm điểm bất thường.
- Phản ứng: Ghi log, cảnh báo và cô lập Container bị xâm nhập.

Kết quả dự kiến

- Mô hình kiến trúc: Đề xuất thành công mô hình bảo mật runtime lai cho Kubernetes.
- Bộ dữ liệu thực nghiệm: Xây dựng dataset syscall bao gồm workload bình thường và các kịch bản tấn công theo MITRE ATT&CK.
- Hiệu quả vượt trội: Giảm tỷ lệ False Positives và tăng khả năng phát hiện Zero-day so với Falco mặc định (đánh giá qua Precision, Recall).
- Prototype: Hệ thống thử nghiệm hoạt động ổn định trong môi trường Lab.

Tài liệu tham khảo

- [1] Souppaya and J. M. Scarfone, "Application Container Security Guide," NIST SP 800-190, 2017.
- [2] R. Chandramouli, "Security Strategies for Microservices-based Application Systems," NIST SP 800-204, 2019.
- [3] L. Degioanni and L. Grasso, "Practical Cloud Native Security with Falco," O'Reilly Media, 2022.
- [4] J. Kosińska and M. Tobiasz, "Detection of cluster anomalies with ML techniques," IEEE Access, 2022.

Tài liệu tham khảo

- [5] A. K. Bhardwaj et al., "AI-powered anomaly detection for Kubernetes security," Babylonian Journal of Machine Learning, 2024.
- [6] S. Wang et al., "Machine learning in network anomaly detection: A survey," IEEE Access, 2021.