A SIEM-Based Framework for Multi-Layer Data Collection and Anomaly Detection in OT-Networks

Prof. Dr. K.-O. Detken (DECOIT®), Prof. Dr. Axel Sikora (University of Applied Science Offenburg), M. Eng. Jaafer Rahmani (University of Applied Science Offenburg)

Prof. Dr. Kai-Oliver Detken
DECOIT® GmbH & Co. KG
Fahrenheitstraße 9
D-28359 Bremen
https://www.decoit.de
detken@decoit.de

Contents

Chapter 1: Introduction and Motivation (1)

- The IT-OT convergence challenge
  - Network integration creating new attack surfaces
  - Traditional IDS/SIEM inadequate for IIoT environments
  - Resource constraints on distributed edge devices
  - Protocol heterogeneity complicating security monitoring
- Real-world threat landscape
  - LogicLocker ransomware targeting industrial control systems
  - Mirai botnet compromising IoT infrastructure
  - Sophisticated multi-stage attacks spanning network layers

Chapter 1: Introduction and Motivation (2)

- Critical research problems
  - Fragmented telemetry across isolated network segments
  - Limited cross-domain detection for diverse attack chains
  - Edge processing constraints hindering real-time response
  - Dataset inadequacy for comprehensive threat modeling
- Research motivation:
  - Develop integrated, scalable, edge-optimized industrial cybersecurity framework

Chapter 2: Related Work & Research Gaps (1)

Public dataset analysis:

Dataset	Data Features	Attack Types	Format	Limitations
IoTID20 [3]	Flow data (packet header-derived)	D/DoS, MITM, scanning	CSV	Lacks sensor, host and modbus/MQTT protocol data
Kitsune [4]	Flow data (packet header-derived)	DDoS, MITM, injection, recon.	pcap, CSV	Lacks sensor, host and modbus/MQTT protocol data; no raw traffic packets
PAN2020_ICS [5]	Sensor telemetry, actuator states, HMI, PLC cmds, Modbus/TCP logs	Unauthorized access, Modbus attacks, control manipulation, replay	NA	Closed access limits utility
ICS Security [6]	SCADA time-series	Cmd injection, replay, unauthorized access	CSV	Lacks network and host data
TON_IoT [7]	Telemetry, flow, OS logs	DoS, DDoS, ransomware, web attacks	Logs, CSV	Lacks modbus/MQTT protocol data and limited host data
CIC IoT 2023 [8]	Flow data	D/DoS, recon, brute force, spoofing, Mirai	pcap, CSV	Lacks sensor, host and modbus/MQTT protocol data; no raw traffic packets
CIC APT 2024 [9]	Flow and host logs	APT (collection, exfiltration, discovery, lateral, evasion, persistence)	pcap, CSV, graphs	Lacks sensor and modbus/MQTT protocol data. Limited host data in the form of provenance logs
Edge-IIoTset [10]	Sensor data, alerts, resource logs, flow data	DoS, MITM, injection, malware	pcap, CSV	Lacks sensor and host data
X-IIoTID [11]	Flow, host logs, alerts	MITRE ATT&CK for ICS	CSV	Lacks sensor and modbus/MQTT protocol data
Our Work	Flow, sensor, host data, alerts, logs	Attacks mapped to MITRE ATT&CK for ICS	CSV, pcap	Under development; aims to integrate correlated telemetry

Chapter 2: Related Work & Research Gaps (2)

Identified research gaps:

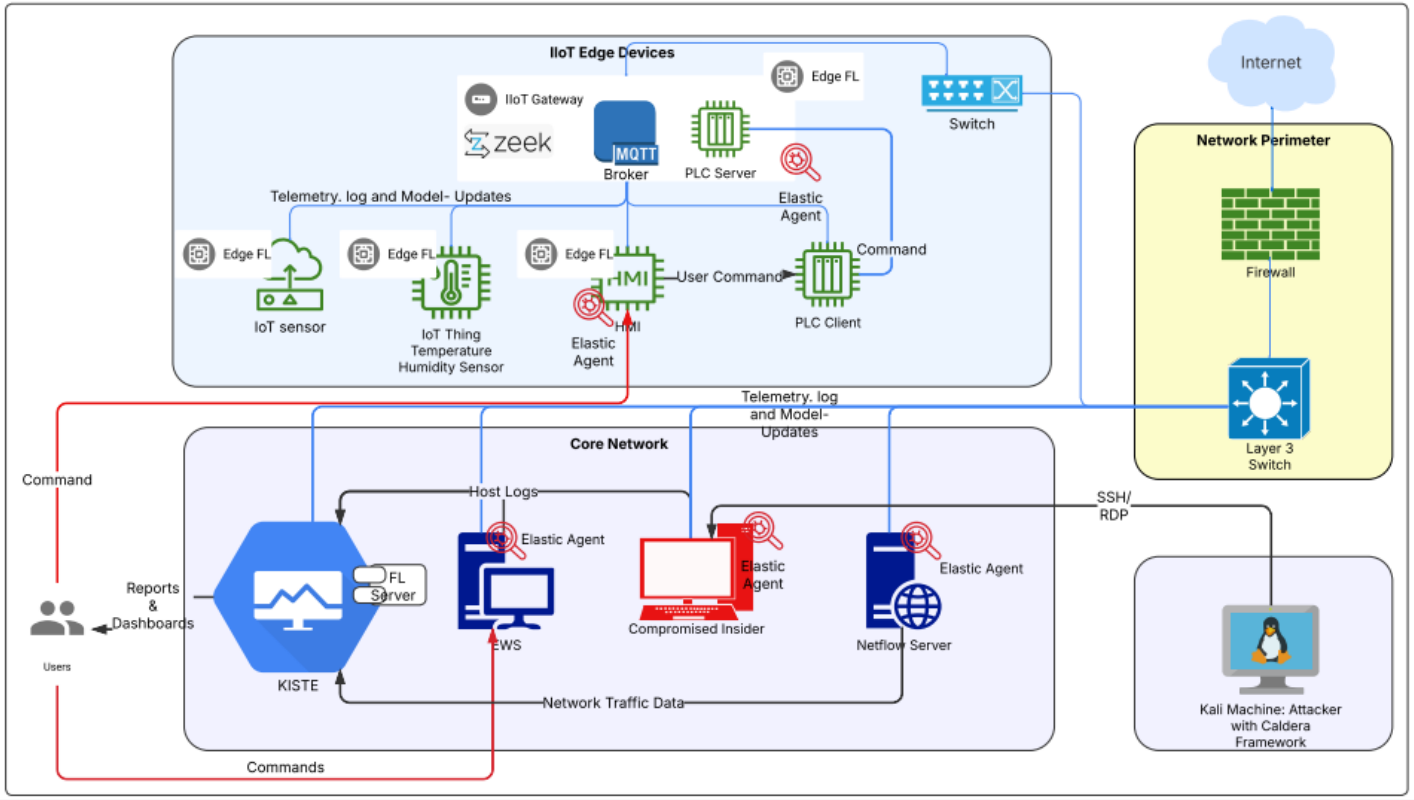
1. Dataset integration: no unified network/host/protocol telemetry
2. Attack realism: synthetic data vs. behavior-driven threats
3. Edge deployment: resource constraints largely ignored
4. Protocol awareness: missing IIoT-specific semantics

Chapter 3: Contributions and Proposed Framework (1)

Four key research contributions:

1. Multi-layer dataset generation: unified NetFlow, Zeek, auditd for comprehensive OT-specific telemetry
2. Hybrid anomaly detection: SIEM rule engine + lightweight edge ML autoencoders
3. Practical deployment: OT-IT integration with edge optimization and scalable architecture
4. Comprehensive data pipeline: raw data transformation, SIEM ingestion, ML-ready exports

Chapter 3: Contributions and Proposed Framework (2)



Framework design principles:

- Protocol-aware processing for industrial semantics
- Real-time response with low-latency detection
- Edge intelligence for distributed environments

Chapter 4: Data Collection & Attack Simulation (1)

Comprehensive multi-layer telemetry architecture:

Layer	Tool	Key Features & Capabilities
Network	NetFlow	Source/destination IPs, ports, duration, byte counts; bidirectional flow analysis for anomaly detection
Protocol	Zeek	Industrial Modbus function codes, register addresses; standard protocols (HTTP, DNS, SSL/TLS); request/response transaction correlation
Host	auditd	System events (process creation, file access, authentication); security monitoring for privilege escalation & lateral movement; behavioral pattern analysis

Chapter 4: Data Collection & Attack Simulation (2)

MITRE Caldera OT-targeted attack simulation:

Attack Tactic	MITRE ID	Implementation
Initial access	TA0101	HMI phishing with malicious ladder logic
Lateral movement	TA0108	Exploitation of PROFINET/ModbusTCP protocols
Impact	TA0109	Safety system manipulation and ransomware deployment

Chapter 4: Data Collection & Attack Simulation (3)

Simulation benefits:

- ATT&CK alignment for precise threat mapping
- Behavioral fidelity emulating realistic attacker dwell times
- Reproducible benchmarking enabling standardized ICS evaluation

Chapter 5: Data Pipeline (1)

Stage 1: collection & ingestion

- Elastic agents collect NetFlow, Zeek, and auditd data
- Initial edge filtering reduces telemetry volume
- Real-time streaming ensures continuous data availability

Stage 2: enrichment & preprocessing

- Standardizes diverse log formats
- Contextual enrichment: geolocation, device IDs, protocol mapping
- Timestamp synchronization across sources

Chapter 5: Data Pipeline (2)

Stage 3: feature extraction

Data Layer	Extracted Features
Network	Flow duration, inter-arrival times, byte-to-packet ratios
Protocol	Transaction IDs, error codes, response latencies
Host	System call sequences, process trees, user patterns

Stage 4: storage & indexing

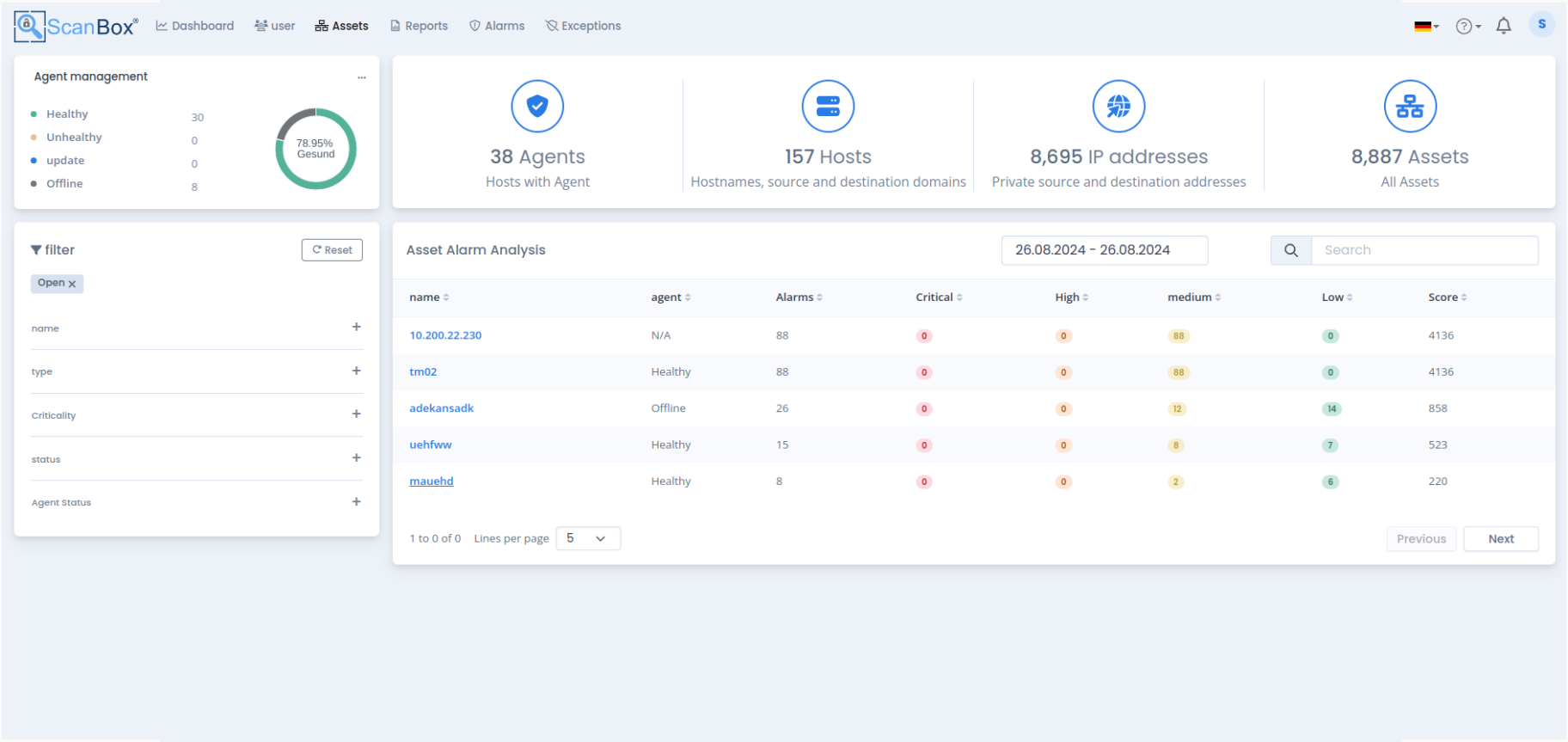
- Elasticsearch for optimized indexing and rapid correlation
- Long-term trend analysis and baseline establishment
- Sub-second querying for incident response

Stage 5: visualization & export

- Kibana dashboards for interactive correlation analysis
- Export data in CSV/JSON formats for ML training
- Automated alert notification and incident workflows

Chapter 5: Data Pipeline (3)

ScanBox® dashboards based on Elasticsearch



Chapter 6: Security & Detection (1)

Framework security analysis:

Attack vector	Target component	Mitigation strategy
Agent compromise	Elastic agents	TLS 1.3 mutual authentication, integrity attestation
Parser exploits	Zeek protocol analysis	Memory-safe languages, input sanitization
Data exfiltration	Elasticsearch/Kibana	Strong ACLs, network segmentation, encryption
Model poisoning	ML autoencoders	Provenance tracking, statistical outlier detection
Physical tampering	Edge devices	Secure boot, HSMs, tamper-evident enclosures

Chapter 6: Security & Detection (2)

Dual detection approach:

Detection method	Key capabilities	Primary function
SIEM rule-based	Signature-based IOC detection; MITRE ATT&CK mapping; real-time alerting	Known threat identification with low-latency response
Edge ML autoencoder	Behavioral baseline learning; reconstruction error analysis; resource optimization	Novel anomaly detection on constrained devices

Chapter 6: Security & Detection (3)

- Integrated analysis
  - Parallel processing: simultaneous rule-based and ML inference
  - Alert correlation: multi-layer threat event consolidation
  - Automated response: incident routing and escalation procedures
- Key architecture benefits
  - Combines precision of signature-based detection with adaptability of behavioral analytics
  - Enables comprehensive threat coverage across known and unknown attacks
  - Optimized for distributed industrial environments with varying resource constraints

Chapter 7: Future Work

Four strategic research directions:

Research direction	Key focus areas	Expected impact
Real-world industrial validation	Multi-sector deployment; performance impact evaluation; cross-domain generalizability	Proven effectiveness across diverse industrial environments
Enhanced protocol support	Emerging protocols (CoAP, MQTT-SN, OPC UA, EtherCAT); advanced threat scenarios; protocol-agnostic detection	Comprehensive coverage of modern industrial communications
Federated learning implementation	Privacy-preserving algorithms; secure model aggregation; Multi-organization collaboration; edge optimization	Collaborative security without data sharing
Advanced AI integration	Explainable AI; adaptive thresholds; predictive maintenance integration	Intelligent, operator-interpretable threat analysis

Research impact: advancing resilient, intelligent industrial cybersecurity ecosystems

Chapter 8: Conclusions (1)

Key technical contributions:

Innovation Area	Achievement	Impact
Multi-layer dataset	Network, protocol, and host telemetry unification	Comprehensive threat detection capability
Hybrid detection	SIEM precision + ML adaptability combination	Enhanced accuracy for known and unknown threats
Attack simulation	MITRE caldera with ATT&CK mapping	Realistic, reproducible threat scenarios
Edge architecture	Resource-constrained device deployment	Practical industrial implementation

Chapter 8: Conclusions (2)

- Discussion & collaboration opportunities:

Area	Focus Topics
Technical Implementation	Framework deployment strategies; SIEM integration approaches; edge device optimization
Research Collaboration	Real-world validation partnerships; protocol-specific detection rules; federated learning initiatives
Industry Applications	Sector-specific customization; regulatory compliance frameworks; cost-benefit analysis methodologies

- Acknowledgment:
  - German Federal Ministry (BMWK)
  - KISTE project: http://kiste-project.info
  - Participants: University of Applied Science Offenburg, DECOIT® GmbH & Co. KG

Thank you for your attention!

DECOIT GmbH & Co. KG
Fahrenheitstraße 9
D-28359 Bremen
https://www.decoit.de
info@decoit.de