Standardized and Usage-Controlled Alert Analysis for Improved Cyber Threat Intelligence

Hendrik Meyer zum Felde[1][0000−0002−5837−8730], Radhouene Azzabi[2][0000−0001−9123−933X], Cédric Gouy-Pailler[3][0000−0003−1298−7845], Gilles Lehmann[4][0009−0001−7483−8952], and Amaia Gil[5][0000−0002−0760−8479]

[1] Fraunhofer AISEC, Lichtenbergstraße 11, Garching near Munich, Germany
hendrik.meyerzumfelde@aisec.fraunhofer.de

[2] CEA Tech en Occitanie, 51 Rue de l'Innovation, 31670 Labège, France
radhouene.azzabi@cea.fr

[3] Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
cedric.gouy-pailler@cea.fr

[4] Télécom SudParis, 9 Rue Charles Fourier, 91000 Evry-Courcouronnes, France
gilles.lehmann@telecom-sudparis.eu

[5] Fundación Vicomtech, Basque Research and Technology Alliance (BRTA),
Mikeletegi 57, 20009 Donostia-San Sebastián, Spain
agil@vicomtech.org

Abstract. Today's IT systems are constantly at risk of being attacked. Security mechanisms and surveillance technologies were developed to detect, record, analyze, and even mitigate such attacks. However, alerts of such events are often structured using individual standards, produced by different commercial, governmental, or even open source community driven organizations. This leads to two problems. First, limited interoperability, because the attributes of the standards highly differ not just in the format but in content, also. Second, SOCs and SIEMs can not share their alert data due to regulations or the danger of leakage.

Here we show an architecture which solves both problems using the alert format IDMEFv2 and an alert sharing architecture which provides usage control of shared sensitive alerts. Our system defends against information leakage while still providing the capability to combine, aggregate, and analyze sensitive information which enables the generation of advanced cyber threat intelligence. This is a scenario that would normally be impossible in real world scenarios.

Using information sources from both cyber security contexts together with physical contexts and technically attested confidential processing of not just nonsensitive alert data, but even sensitive data, may provide necessary insights to provide countermeasures for existing threats faster. Gathered data is processed using conventional analyses and AI/ML techniques.

Since our work is still in progress, the upcoming analysis about our proof-of-concept will be used to evaluate the potential of the approach in terms of scalability, complexity, flexibility, performance, effectiveness, and most importantly security.

Keywords: Alert standardization · Usage control · CTI generation.

1 Introduction

In the field of IT security, attempts of cyberattacks are a persistent challenge. To mitigate these threats, various systems have been developed to detect, classify, and analyze security incidents. When an incident occurs, an alert is generated and transmitted to a centralized monitoring system, enabling real-time surveillance and response coordination across multiple systems.

However, two primary challenges arise in this process. First, industrial security products exhibit significant variability in the structure and content of alerts. This heterogeneity necessitates a *standardized format* that strikes a balance between providing sufficient technical detail for analysis while avoiding excessive complexity. A unified format is essential for seamless integration and centralized processing of alerts originating from diverse sources.

Second, security incident reports often contain sensitive information that must be protected against unauthorized disclosure. Security centers operate under strict constraints regarding the types of data that can be shared externally. Nevertheless, the aggregation of security data from multiple sources holds significant potential for improving threat detection and response. A solution that incorporates *usage control mechanisms* can enable secure data sharing while enforcing policies on data access and usage.

Using our system, we aim to facilitate the rapid and comprehensive generation of cyber threat intelligence by integrating data from diverse sources. Our approach enables cross-domain analysis across industrial, regional, and national boundaries, allowing for deeper insights into cyber threats and enhancing collaborative security efforts on a broader scale. Through this work, we introduce a framework that ensures both standardized incident reporting and controlled data sharing, addressing the need for interoperability while safeguarding sensitive information.

2 Background

This section details necessary background knowledge. *Usage control* is an advanced access control model that extends traditional access control mechanisms by incorporating continuous decision-making and obligations. Unlike conventional models that enforce access permissions only at the time of request, usage control dynamically evaluates and enforces policies throughout the entire session.

Key aspects of usage control include mutability, where attributes can change during usage, and continuity, which ensures ongoing compliance with policies. These characteristics make usage control particularly effective in modern security contexts, such as cloud computing, digital rights management, IoT environments, and data privacy protection. By integrating authorization, obligation, and condition-based controls, usage control provides a flexible and adaptive security framework, enabling organizations to enforce fine-grained access policies in dynamic and distributed systems.

3 Related Work

This section provides information about other initiatives with related interests and purpose. The choice of the standard depends on the use case.

3.1 Alert Format Standardization

The following standards focus on incident alerts. *The Incident Detection Message Exchange Format version 2 (IDMEFv2)*[12] is a standardized format for describing cyber and physical incidents, as well as events of interest that may contribute to such incidents. It supports the reporting and analysis of diverse security-related events, including detection of viruses, unauthorized access attempts, server downtime, reconnaissance scans, abnormal environmental conditions, and unauthorized activities near restricted facilities. Additionally, IDMEFv2 is effective in hybrid environments where cyber and physical security converge to detect complex threats.

The Common Event Format (CEF)[6] introduced by ArcSight and the *Log Event Extended Format (LEEF)*[14] introduced by IBM is a customized event format for the software QRadar. Both formats are mainly event and log related and only support information about the cyber realm. CEF and LEEF can be converted in IDMEFv2, however, the contrary is very complex. The *Vocabulary for Event Recording and Incident Sharing (VERIS)*[13] is another project which aims to structure alert information in such a way that a sufficient amount of useful information is available for further analysis without bloating the format.

The following format focuses on expressing Cyber Threat Intelligence (CTI). *Structured Threat Information Expression (STIX)*[4] developed by Oasis and IDMEFv2 are complementary. STIX is a format to model, analyze, and share CTI.

The following format focuses on emergency and cyber-physical alerts. *The Common Alerting Protocol (CAP)*[22] standardizes emergency alerts, enabling simultaneous dissemination across multiple systems while supporting threat pattern detection and integrating best practices for effective warning communication.

The *Simple Network Management Protocol (SNMP)*[1] was originally introduced in the 1990s and later improved in 2002 with SNMPv3. The standard which was developed by the IETF allows to control and surveil network devices, such as routers, servers, switches, printers, or computers. SNMP's strength is the capability to detect availability issues and system errors.

The Incident Object Description Exchange Format (IODEF)[9], is complementary to IDMEFv2. IODEF is used after detection to describe, transmit, and share information about incidents to other security teams. IDMEFv2 is used for upstreaming in probes and security management tools to detect incidents. IDMEFv2 alerts can be attached to IODEF messages for describing technical information about incidents in detail.

Standards like IDMEFv2 bridge the gap between cyber and physical security, compared to other security formats it also includes availability incidents and natural hazard incidents. This enables a holistic approach to incident detection which IDMEFv2 has in focus, which is the reason this format was chosen in this project. A more detailed explanation why the IDMEFv2 format was chosen for this project can be found in the upcoming Section 4.2.

3.2 Security Architectures and Initiatives for Alert Sharing

Several architectures, initiatives, and products facilitate IT security alerting and incident sharing across organizations, enabling real-time threat detection, response, and intelligence sharing. These can be categorized into open standards, government initiatives, commercial platforms, and community-driven projects.

Architectures and Frameworks Oasis has developed STIX and also created an application protocol to exchange STIX datasets via a predefined HTTP REST API named *Trusted Automated Exchange of Intelligence Information (TAXII)*[5]. This project also aims to use an HTTP REST API for alert transmission. Oasis has also developed *Open Command and Control (OpenC2)* which aims to standardize vendor and platform agnostic machine-to-machine communication for automated response and control commands to orchestrate security reactions after incidents occurred [23].

Malware Information Sharing Platform (MISP)[18] is an open-source platform for CTI sharing and incident response collaboration. It supports STIX, TAXII, and IDMEF formats. Enables structured threat sharing among SOC teams, Computer Security Incident Response Teams (CSIRTs), and government agencies.

Open Cyber Threat Intelligence (OpenCTI)[11] is an open-source threat intelligence management platform. It uses a STIX-native architecture for structured CTI sharing and integrates with MISP, MITRE ATT&CK, and Security Information and Event Management (SIEM) solutions, which are explained in the upcoming sections. OpenCTI provides visual analytics for attack pattern tracking.

TheHive[26] by StrangeBee is an open-source incident response and threat intelligence platform which supports STIX/TAXII and MISP. It automates SOC workflows and threat enrichment. *Open Threat Exchange*[15] by Alienvault is a community-driven threat intelligence sharing network which processes open-source threat feeds for Security Operation Centers (SOCs). It enables collaborative intelligence enrichment.

Government and Industry Initiatives *MITRE ATT&CK*[7] is a globally recognized industry knowledge base for cyber adversary behavior mapping with free accessible sources for everyone. It provides tactics, techniques, and procedures of attackers and is used for threat hunting, incident detection, SOC automation and purple teaming.

Computer Incident Response Center Luxembourg (CIRCL)[17] is a European initiative for cyber incident sharing. It supports response coordination and provides MISP-based threat intelligence sharing and GDPR-compliant cybersecurity data handling.

Cybersecurity and Infrastructure Security Agency (CISA)[8] is the US government's cybersecurity division focused on threat sharing and national defense. It operates using automated indicator sharing for real-time CTI distribution and supports STIX and TAXII-based intelligence exchange. Its main purpose is to manage national cyber awareness system alerts.

Nato Computer Incident Response Capability (NCIRC)[24] is a NATO-led initiative for cyber defense and incident response across military and defense networks. The aim is to provide intergovernmental CTI sharing.

Commercial Security and CTI Platforms *IBM X-Force Exchange*[3] is a cloud-based threat intelligence platform by IBM. It is capable to aggregate global threat intelligence feeds and provides STIX/TAXII-based integration with SIEMs. It enables collaborative threat research and intelligence sharing and focuses on enterprise-level threat intelligence, SOC threat hunting, and security research. *Anomali ThreatStream*[2] is a commercial CTI sharing platform with support for STIX/TAXII and custom threat intelligence feeds. It automates threat correlation across SOC environments and integrates with SIEM and Security Orchestration, Automation, and Response (SOAR) platforms. *Palo Alto Cortex XSOAR*[20] is a security orchestration, automation, and response platform which automates incident response workflows and provides real-time threat intelligence sharing. Formerly known as FireEye and renamed *Trellix* [16] is another commercial CTI platform which integrates real-time threat intelligence into SIEM solutions and provides automated threat scoring and adversary profiling.

Summary Numerous standards and initiatives for security alerts exist ranging from open-source platforms (MISP, OpenCTI, Open Threat Exchange) to government-led initiatives (CISA, NATO NCIRC) and commercial solutions (IBM X-Force, Anomali ThreatStream, Cortex XSOAR). However, none of the projects and initiatives mentioned deal with sensitive alert data and none apply usage control to enable CTI generation which we focus on in our work. Table 1 shows an overview of the compared formats with either none, some, more, or full support depicted using empty and filling circles for the following coverage capabilities. *Cyber Security*, *Physical Security* (e.g., intrusion, theft), *Availability* (e.g., physical failures or power outages), *Hazards* (e.g., wild fire or heat), *Logs and Events*, *Detection* (correlation capacities to improve detection monitoring), *CTI* (creation of shared cyber threat knowledge after detection), *Analysis* (deeper analysis of an incident cause and possible resolution after detection), *Command and Control* (in reaction to an incident after detection). To put it in a nutshell, IDMEFv2 covers the objectives of SNMP, CEF/LEEF and CAP and IDMEFv2 is complementary to IODEF, STIX and OpenC2.

Table 1. Comparison of related Standards and Initiatives and their coverage property support depicted in circles: ○ no, ◔ some, ◕ more, and ● for full support.

Format	Cyber Security	Physical Security	Availa-bility	Hazards	Log/ Event	Detec-tion	CTI	Analy-sis	C&C
SNMP	◔	◔	●	○	◔	●	○	◔	○
CEF	●	○	○	○	●	◕	◔	●	◔
LEEF	●	○	○	○	●	◕	◔	●	◔
CAP	○	○	○	●	○	○	○	○	○
STIX	●	○	○	○	◕	◔	●	◔	○
OpenC2	●	◕	◕	○	○	◔	○	○	●
IODEF	●	○	○	○	◕	◕	●	●	○
IDMEFv2	●	●	●	●	◕	●	◔	◕	◔

4 Concept

This section introduces the overall architecture of the alert information flow and afterwards provides details concerning the usage control architecture. Lastly, the generation of Cyber Threat Intelligence (CTI) is discussed.

4.1 Alert Centralization Architecture

The overall architecture can be seen in Figure 1. Here, the Cyber and Physical Security Information and Event Management (CPSIEM) serves as the central entity for collecting and processing security alerts. It receives alerts from multiple Security Operation Centers (SOCs), each operating in distinct domains and contexts, for instance, from military, industrial or governmental background. To ensure secure communication, a dedicated Gateway (GW) is established for each SOC, facilitating the transmission of alerts. The secure alert transmission is depicted using solid-line arrows, while CTI retrieval mechanisms are represented by dashed-line arrows. Additionally, for the retrieval and exchange of CTI, a CTI REST API is provided, enabling structured and efficient access to intelligence data. This centralized architecture allows to collect alerts from both cyber realm and physical sensors.

4.2 Alert Format Design Choice

The format for the transmission of alerts is crucial as all components depend on this decision and proper processing and compatibility must be given. The reasoning for the choice to use IDMEFv2 is described in detail in the following sections.

Standardization Across Diverse Systems Security solutions are often dependent on proprietary formats from different vendors, which requires different

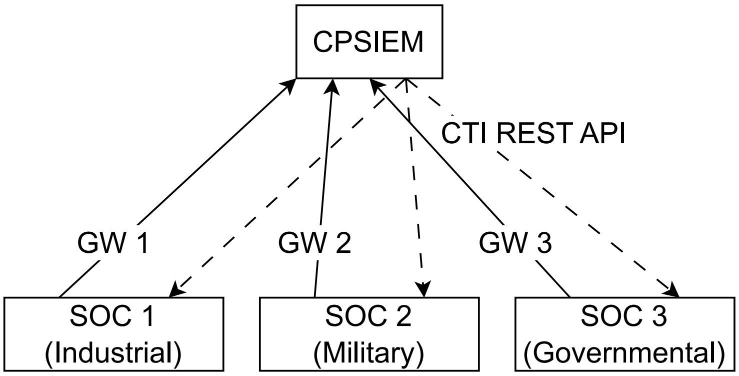

Fig. 1. General overview of the alert and CTI flow. At the top the cyber and physical security information and event management (CPSIEM) is shown which receives alerts from the security operation centers (SOCs) at the bottom, which may all work in different domains and contexts. For the alert transmission a secure gateway (GW) is established for each SOC (line with arrow). For the retrieval of cyber threat intelligence a CTI REST API is provided (dashed line with arrow).

format processing logic to enable interoperability. IDMEFv2 solves this by providing a standardized format that can be used across several security elements, such as firewalls, intrusion detection systems (IDS), Security Information and Event Managements (SIEMs), and physical security sensors.

Cyber and Physical Security Events As mentioned in Section 3.1, IDMEFv2 is designed to handle both cyber and physical security incidents. This makes it suitable for hybrid security environments, such as smart buildings, industrial control systems (ICS), and IoT/IIoT deployments, where threats may originate from multiple domains.

Flexible and Extensible Data Model IDMEFv2 messages are composed of well-defined classes (e.g., Alert, Analyzer, Sensor, Source, Target) and attributes (e.g., timestamps, IP addresses, geolocation, protocols, user identities). This structured flexibility is essential when it comes to covering requirements across numerous different security domains.

Secure and Usage-Controlled Data Exchange IDMEFv2 messages are formatted in JSON which is a structure that has a vast amount of native support in software and can be conveniently transmitted via HTTPS, ensuring confidentiality and integrity. Security centers often face limitations on data sharing capabilities due to company regulations. IDMEFv2's JSON structure can be easily integrated with usage control mechanisms as sufficient amount of information is provided for enforcing policies defining who can access, process, or store alert data while staying within legal boundaries.

Future-Proof for Emerging Threats IDMEFv2 is designed with extendability. For instance, files can be attached to the alert messages or additional fields specified to cover upcoming future cyber-physical challenges which might need coverage in areas, such as AI-driven threats, supply chain attacks, or large-scale IoT attacks.

4.3 Enabling Secure Sharing of Sensitive Alert Data

After having introduced our overall design and the reasoning for the alert format, the following sections explain the necessary technical requirements for the GWs to enable usage control for IDMEFv2 formatted alerts.

While sharing alert data from various sources and contexts is beneficial, the primary objective is to facilitate the collection and analysis of sensitive data, which would otherwise remain inaccessible due to confidentiality constraints.

For example, certain alert attributes, such as the attacker's IP address, are generally considered nonsensitive and can be shared without significant risk. In contrast, highly sensitive information, such as the impact or extent of damage caused by an attack, is typically withheld, as its disclosure could severely harm an organization's reputation.

However, if it were possible to securely share both nonsensitive and sensitive data, a more comprehensive analysis could be conducted, leading to the generation of higher-quality CTI. This would enable deeper insights into attack patterns and enhance defensive strategies. The sharing of sensitive data, however, is only feasible if the receiving system can be trusted to process the data in strict accordance with predefined and agreed-upon usage policies.

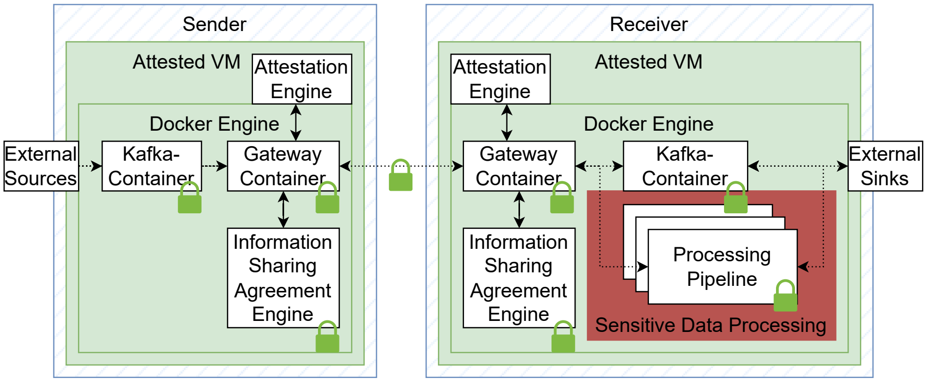

Fig. 2. Overview of the architecture for usage control on sender and receiver side shown in blue diagonal lines with attested virtual machines running an attested docker engine on both sides in light-green, attested and secured components marked as green locks, processing areas of sensitive alert data shown in dark-red, alert data flow as dotted arrows, and the communication flow shown as regular arrows.

Figure 2 illustrates the architecture for such an enforcement of usage control during the transmission and processing of security alert data. The system ensures the integrity and confidentiality of shared information by employing attested Virtual Machines (VMs) running an attested Docker engine represented in green on both the sender and receiver sides, represented in light-blue diagonal lines. Secure and attested components are marked with green locks, while processing areas for sensitive alert data are highlighted in dark-red. The alert data flow is depicted as dotted arrows, and the communication flow as solid arrows.

The remote attestation of the sender and receiver's software stack is anchored using state-of-the-art Trusted Execution Environments (TEEs). TEEs were invented to enable confidential computing in cloud environments, which are not under the control of the consuming party. Before TEEs were introduced it was necessary to trust the cloud provider to neither interfere with the execution nor extract sensitive information out of its memory. The most famous implementations of TEEs for the protection of VMs include AMD Secure Encrypted Virtualization (SEV) and Intel Trust Domain eXtensions (TDX) which could both serve as hardware trust anchor for the required usage control mechanism.

TEEs typically sandbox processes from the rest of the system using memory which is separated by hardware mechanisms and carefully separated context switches when data is loaded inside a CPU. Not even administrators with root-access can enter these protected executions. Despite numerous attacks that have been reported on TEEs[21], we expect the TEEs to work properly and build upon these security guarantees of confidentiality, integrity, and remote attestation capabilities. Especially the remote attestation feature, to verify that an expected software stack with an expected configuration is loaded and no modifications were made, is paramount for our architecture.

Sender Side Processing On the sender's side, alerts intended for transmission are first inserted into the system via a Kafka database entry. These alerts are then transferred from the Kafka container to the gateway container, where the gateway establishes a secure, encrypted connection with the receiving system's gateway endpoint.

Prior to data transmission, a mutual attestation process is conducted. During this phase, both the sender and receiver gateways communicate with their respective attestation engines to verify the integrity of their systems. The attestation mechanism ensures the authenticity and secure configuration of the virtual machines involved. Additionally, critical software components, including the Docker engine, Docker images (e.g., for the gateway and Kafka database), and the Information Sharing Agreement (ISA) engine, undergo attestation to confirm they remain untampered and in a trustworthy operational state.

If all attestation checks are successful, the gateway verifies the policy constraints defined in the Information Sharing Agreement (ISA). Only if an appropriate policy is in place and the required security level is met, the alerts are transmitted to the receiver's gateway endpoint.

Receiver Side Processing Upon arrival at the receiver's gateway, the alert data is classified as either nonsensitive or sensitive. Nonsensitive data can be processed and forwarded directly, whereas sensitive data must be handled within a predefined and strictly controlled processing pipeline. This processing pipeline is designed to:

- Aggregate and analyze complex datasets to generate enhanced cybersecurity insights.
- Enforce filtering mechanisms to ensure that no raw sensitive data leaves the attested system.

After processing, the finalized data is forwarded to the receiver's external data sink for further analysis and response actions.

Secure Gateway Communication To enable trusted message exchange between authenticated gateway senders and receivers, we introduce two core security components: a Public Key Infrastructure (PKI) for certificate management and an automated onboarding system for new gateways. This ensures that each gateway sender is properly registered and its key material verified before sending IDMEFv2 messages.

The onboarding process is as depicted in Figure 3. When a new gateway sender comes online, the system verifies its identity using pre-configured authentication methods (e.g., API key). Once verified, the gateway creates its cryptographic key pair and submits a certificate signing request to the PKI's Certificate Authority, which issues an X.509 digital certificate after validation. Before exchanging messages, both parties' gateways must perform mutual state-of-the-art TLS authentication. If none of the involved gateways' certificates were revoked a secure channel among them will be established. This prevents message flooding attacks from unregistered systems and also impersonation of other participating parties.

IDMEFv2 Message Integrity Protection To additionally protect IDMEFv2 messages and verify their integrity, the system uses JSON Web Signatures (JWS). The gateway sender signs the IDMEFv2 message by creating a JWS token, using its generated private key during the onboarding session to sign the contents. This signed message is then securely sent over HTTPS to the gateway receiver. Upon receiving the message, the gateway extracts the JWS token and verifies the signature using the sender's public key. If the signature is valid, the message is considered authentic. If the verification fails, the message is rejected.

This procedure ensures that only trusted, attested components handle sensitive security alerts, while enforcing strict policy-based data sharing and processing. This approach enhances data confidentiality, integrity, and compliance in multi-domain cybersecurity environments.

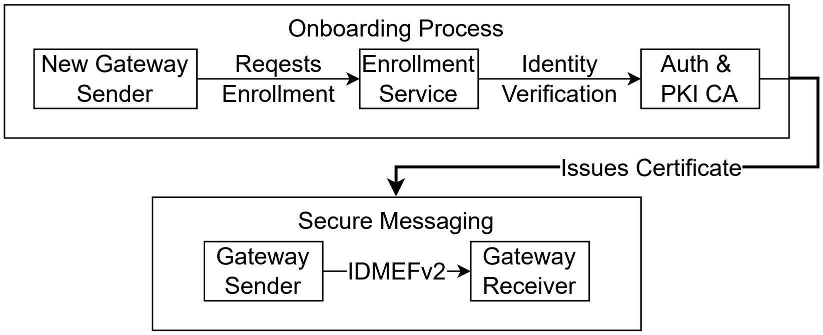

Fig. 3. Overview of the onboarding process of gateways, which must enroll using an enrollment service which must check the party's identity. Afterwards a PKI enables secure communication among accepted gateways.

4.4 AI based Threat Detection

This section details how the gathered alert data is processed further by applying AI/ML algorithms which profit from datasets containing nonsensitive data, but eventually profit even more from datasets which also contain sensitive information.

Raw Data The focus is set on detecting threats from raw data, including logs, hardware signals from embedded systems (e.g., HPC, syscalls) and summarized network data (PCAP files and netflows). The goal is to evaluate existing tools and advanced AI/ML techniques on structured data represented as an attributed heterogeneous graph. The process involves several key steps:

1. Identifying and collecting relevant raw data for project selected scenarios.
2. Exploratory analysis of the collected data to select and pre-process it for AI/ML model training. This process includes entity recognition techniques from logs, data selected from PCAP files and network topology reconstruction based on the data [25].
3. Model training and parameter fine-tuning in the training data set.
4. Application of AI/ML models in multiple controlled or simulated scenarios (or test sets) with cyber threats already annotated to validate model results.

Finally, the detected threats are converted into and IDMEFv2 messages to share AI/ML identified potential threats and generate CTI.

Threat Detection Combining Multiple Sources of Filtered Information In this case the IDMEFv2 messages are the input data, as high-level log data that will be analyzed by AI/ML models. These IDMEFv2 messages come from AI/ML detections on raw data or from translations (to homogenize messages) of different manufacturers' SIEMs logs. Correlation and clustering techniques are applied to aggregate information shared from multiple sources (with sensitive and personal information previously removed) obtaining a knowledge graph of the actual situation. The objective is to detect simultaneous incidents concerning attacks or anomalies in multiple locations or companies. Another aim is to detect which exploits can be linked to specific system vulnerabilities. The aggregation of data from multiple SOCs opens up possibilities for identifying broader attack campaigns and shared vulnerabilities across different organizations. Advanced techniques like Graph Neural Networks (GNNs) present a promising avenue for analyzing the relationships between security events represented as interconnected graphs, enabling the detection of complex, multi-stage attacks that might otherwise go unnoticed [27]. Moreover, the integration of multi-modal AI, capable of analyzing diverse data types encompassing cyber, physical, and availability events, can provide a more nuanced understanding of the threat landscape, revealing correlations that a single-domain analysis might miss. The sharing of data between SOCs, while crucial for collaborative CTI, inevitably raises significant privacy concerns. This critical aspect is advocating toward the adoption of adaptive privacy-preserving techniques when sharing events.

The Role of Contextual Information in Trustworthy AI Models The trustworthiness and effectiveness of AI models deployed in security operations are intrinsically linked to the availability and understanding of contextual information. This includes documenting the creation, operation, and lifecycle management of AI-generated events [19]. Sharing contextual information allows security analysts to better visualize and interpret the outputs of AI models, understand their potential biases or limitations, and ultimately build greater confidence in the AI-driven intelligence they receive [10].

Establishing Adaptive AI through Feedback Mechanisms To ensure that the AI systems remain effective and adapt to the ever-evolving threat landscape, the implementation of robust feedback loops is paramount. Mechanisms should be established to allow analysts to easily provide feedback on the accuracy, relevance, and usefulness of AI-generated alerts and intelligence. This feedback can then be used to adjust their parameters, or refine the underlying algorithms. For example, if analysts consistently flag certain alerts as false positives, this information can be used to adjust the model's decision thresholds or to identify and address biases in the training data. A crucial element of the feedback loop is the incorporation of CTI. Both external CTI feeds, providing information about emerging threats and vulnerabilities, and internal CTI gathered within the network, should be used to inform the AI models. This allows the models to learn about new attack techniques, indicators of compromise, and threat actors, enabling them to adapt their detection strategies accordingly. Performance monitoring and automated retraining are also essential components of an effective feedback loop. The performance of AI models should be continuously monitored using relevant metrics. When performance degrades below a certain threshold or when significant new data becomes available, automated retraining processes should be triggered to update the models and ensure their continued effectiveness. Therefore, top-down policy updates should be integrated into the AI models. High-level security policies and strategic goals defined by the initiative or by individual participating organizations should be reflected in the behavior of the AI systems. This ensures that the AI-driven intelligence is aligned with the overall security objectives.

Joint Analysis of Heterogeneous Security Data A significant challenge lies in enabling the joint analysis of heterogeneous security data. Modern security incidents often involve a complex interplay of events across different domains, including cyber networks, physical infrastructure, and the availability of critical systems. Analyzing these disparate data types in isolation provides an incomplete and potentially misleading picture of the overall security posture. Security data comes in various forms, such as network logs, system logs, physical access control records, sensor readings, and operational metrics, each with its own format, structure, time scale, and frequency. AI capabilities in this initiative are designed to integrate and analyze data from multiple modalities simultaneously, allowing for the identification of correlations and patterns that would be missed by single-modal analysis. For instance, a cyberattack might be preceded by unusual physical access attempts, or a disruption in network availability might correlate with specific physical events. Multi-modal AI can learn these complex relationships by processing data from different sources in a unified framework.

5 Future Work

As this research represents ongoing work, future efforts will focus on the implementation of the proposed framework, along with a comprehensive evaluation of applicable use cases and real-world data analysis. Additionally, an in-depth performance assessment will be conducted to evaluate the efficiency and scalability of the system.

To further validate the feasibility of the approach, a proof-of-concept implementation is planned. This will provide technical insights into the challenges associated with usage control in security-sensitive environments, enabling further refinements and optimizations of the system.

6 Conclusion

In this work, we addressed the challenges of secure and standardized alert data sharing in heterogeneous cybersecurity environments. We introduced a framework that enables the controlled exchange of security alerts and Cyber Threat Intelligence (CTI) while ensuring compliance with security policies and protecting sensitive information. By leveraging IDMEFv2 as a standardized alert format, our approach ensures interoperability across diverse security systems, including cyber and physical security domains.

To mitigate the risks associated with sharing sensitive security data, we integrated a usage control mechanism that enforces predefined policies on data access, processing, and dissemination. Our architecture incorporates attested Virtual Machines (VMs) and secure, attested communication channels, ensuring that only trusted systems can process sensitive data. The proposed design enables the aggregation of both non-sensitive and sensitive alerts, allowing for more in-depth threat analysis and the generation of advanced CTI across industrial, regional, and national boundaries.

By addressing the technical and operational challenges of secure alert data sharing, this research contributes to the advancement of collaborative cybersecurity intelligence and threat detection.

Acknowledgments. This work was partially funded by the Digital Europe Programme DIGITAL, Grant agreement no. 101145846 [6].

References

1. et al., N.W.G.J.C.: Request for comments: 3410 - introduction and applicability statements for internet standard management framework (2025), https://datatracker.ietf.org/doc/html/rfc3410/, accessed: 2025/06/02
2. Anomil: Anomali threatstream (2025), https://www.anomali.com/products/threatstream, accessed: 2025/03/27
3. Cloud, I.: Ibm x-force threat intelligence api documentation (2025), https://api.xforce.ibmcloud.com/doc/, accessed: 2025/03/27
4. Committee, O.C.T.I.T.: Introduction to stix (2025), https://oasis-open.github.io/cti-documentation/stix/intro, accessed: 2025/03/27
5. Committee, O.C.T.I.T.: Introduction to taxii (2025), https://oasis-open.github.io/cti-documentation/taxii/intro.html, accessed: 2025/03/27
6. Corporation, O.T.: Implementing arcsight common event format (cef) - version 26 (2023), https://www.microfocus.com/documentation/arcsight/arcsight-smartconnectors-8.4/pdfdoc/cef-implementation-standard/cef-implementation-standard.pdf, accessed: 2025/03/27
7. Corporation, T.M.: Att&ck matrix for enterprise (2025), https://attack.mitre.org/, accessed: 2025/03/27
8. Cybersecurity, (CISA), I.S.A.: Home page (2025), https://www.cisa.gov/, accessed: 2025/03/27
9. Danyliw, e.a.: The incident object description exchange format (2007), https://datatracker.ietf.org/doc/html/rfc5070, accessed: 2025/03/27
10. Echeberria-Barrio, X., Gil-Lerchundi, A., Mendialdua, I., Orduna-Urrutia, R.: Topological safeguard for evasion attack interpreting the neural networks' behavior. Pattern Recognition **147**, 110130 (2024)
11. Filigran: Opencti documentation space (2025), https://docs.opencti.io/latest/, accessed: 2025/03/27
12. Force, I.T.: Idmefv2 in a nutshell (2025), https://idmefv2.github.io/index.php/idmefv2-in-a-nutshell/, accessed: 2025/05/27

[6] SAFE4SoC project: https://safe4soc.eu/ and funding https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/projects-details/43152860/101145846

13. Framework, V.: Veris the vocabulary for event recording and incident sharing (2019), https://verisframework.org, accessed: 2025/03/27
14. IBM: Leef overview (2025), https://www.ibm.com/docs/en/dsm?topic=leef-overview, accessed: 2025/03/27
15. Inc., L.: Open threat exchange (2025), https://otx.alienvault.com/, accessed: 2025/03/27
16. LLC, M.U.: About trellix (2025), https://www.trellix.com/en-gb/about/, accessed: 2025/03/27
17. Luxembourg, C.C.I.R.C.: Our services (2023), https://www.circl.lu/, accessed: 2025/03/27
18. MISP-Project: Misp documentation and support (2025), https://www.misp-project.org/documentation/, accessed: 2025/03/27
19. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D., Gebru, T.: Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency. pp. 220–229 (2019)
20. Networks, P.A.: Anomali threatstream (2025), https://www.paloaltonetworks.com/resources/datasheets/cortex-xsoar, accessed: 2025/03/27
21. Nilsson, A., Bideh, P.N., Brorsson, J.: A survey of published attacks on intel sgx. arXiv preprint arXiv:2006.13598 (2020)
22. OASIS: Common alerting protocol version 1.2 (2025), https://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html, accessed: 2025/03/27
23. OASIS: Open command and control (openc2) (2025), https://openc2.org/, accessed: 2025/06/02
24. Organization, N.A.T.: Nato cyber defence (2021), https://www.nato.int/nato_static_fl2014/assets/pdf/2021/4/pdf/2104-factsheet-cyber-defence-en.pdf, accessed: 2025/03/27
25. Segurola-Gil, L., Moreno-Moreno, M., Irigoien, I., Florez-Tapia, A.M.: Unsupervised anomaly detection approach for cyberattack identification. International Journal of Machine Learning and Cybernetics **15**(11), 5291–5302 (2024)
26. StrangeBee: Thehive documentation (2025), https://docs.strangebee.com/thehive/overview/, accessed: 2025/03/27
27. Zola, F., Segurola-Gil, L., Bruse, J.L., Galar, M., Orduna-Urrutia, R.: Network traffic analysis through node behaviour classification: a graph-based approach with temporal dissection and data-level preprocessing. Computers & Security **115**, 102632 (2022)