Multi-Partner Project: CyberSecDome - Framework for Secure, Collaborative, and Privacy-Aware Incident Handling for Digital Infrastructure

Mohammad Hamad*, Michael Kühr*, Haralambos Mouratidis‡[xiii], Eleni-Maria Kalogeraki‡, Christos Gizelis¶, Dimitris Papanikas¶, Athanasios Bountioukos-Spinaris§, Charilaos Skandylas†, Evangelos Raptis **, Andreas Alexopoulos**, Grigorios Chrysos††, Mina Marmpena[xi], Sevasti Politi[xi], Konstantinos Lieros[xi], Papagiannopoulos Nikolaos‡‡, Iordanis Xanthopoulos[xii],Spyros Papastergiou[x], Sotiris Ioannidis††, Mikael Asplund†, Marc-Oliver Pahl‖, Sebastian Steinhorst*

* Technical University of Munich, Germany
† Linköping University, Sweeden
‡ Security Labs Consulting, Ireland
[xiii] University of Essex, UK
§ CyberAlytics Ltd., Cyprus
¶ Hellenic Telecommunications Organisation, Greece
‖ IMT Atlantique, France
** Aegis IT Research, Germany
†† Technical University of Crete, Greece
‡‡ Athens International Airport S.A., Greece
[x] MAGGIOLI S.P.A., Italy
[xi] Internet of Things applications and Multi-Layer development, Cyprus
[xii] Sphynx, Switzerland

***Abstract*—Digital infrastructure is vital for the economy, democracy, and everyday life, yet it is becoming increasingly vulnerable to strategic cyber-attacks. These attacks can lead to significant disruptions, resulting in widespread service outages, financial losses, and a decline in public trust. Ensuring resilience is difficult due to the infrastructure's complexity, the large volume of data involved, and the growing need for quick, coordinated responses. In the EU Horizon project CyberSecDome, we propose a multi-layered framework that provides AI-driven solutions for incident prediction and detection, automated testing, risk assessment, and rapid incident response, supporting continuity amid complex, large-scale cyber threats. Additionally, CyberSecDome introduces a virtual reality interface to enhance AI model explainability and provide real-time contextual awareness of ongoing attacks and defense mechanisms. It also enables privacy-aware model sharing across AI systems, fostering secure collaboration among different domes.**

***Index Terms*—Security, AI, Incident handling, Intrusion detection**

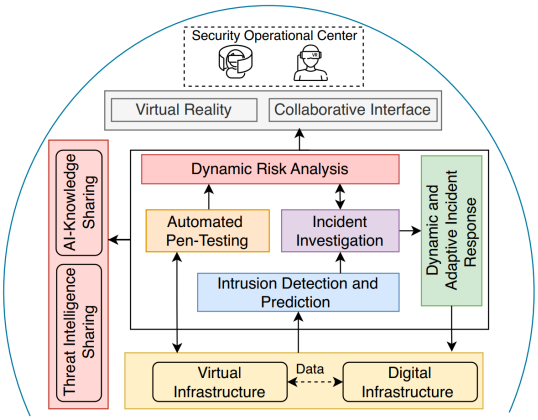

Fig. 1: CyberSecDome: High-Level Architecture

I. Introduction

The digital infrastructure has become one of the most important pillars that uphold our economy, our democracy, and our daily lives. It consists of servers, data centres, telecom exchanges, radio access networks, satellites, databases, data stores, information technology services, cloud applications, and IoT device endpoints. The ongoing war in Ukraine has shown that disrupting critical infrastructure can be a strategic objective that could be achieved even before ground invasions are conducted [1]. Cyber-attacks against digital infrastructure may cause *digital disruption*, which can in turn result in huge financial losses, reduced trust in societal services, and even loss of human life [2]. Therefore, the protection of digital infrastructure from ever-evolving cybersecurity threats is becoming increasingly important.

Yet, securing digital infrastructure is challenging for several reasons. First, while exploits and attacks are increasingly automated and able to scale massively, the protection schemes remain largely manual and resource-demanding. For example, existing proactive security protection solutions (e.g., firewalls) cannot guarantee the continuity of digital infrastructures, especially when we consider zero-day attacks. Penetration testing can help harden systems but requires expertise, which is in scarce demand. Second, the massive data that is generated from digital infrastructure makes detecting, predicting incidents, and correlating the incident information a very challenging task. Third, the need for quick and accurate responses to the detected attacks puts the security incident teams under high

stress, especially if we consider the lack of mechanisms that provide better situational attack awareness and the risk that the systems faces is taken into account. Fourth, there is a need for collaborative incident response against cross-border cyber incidents and threats that can affect multiple sectors across multiple countries. Finally, there is a lack of sharing threat information mechanisms that are compliant with data protection regulations and can address all concerns regarding mass surveillance and the protection of personal spaces.

The **CyberSecDome** framework, shown in Fig. 1, aims to provide a solution that realizes a suite of security tools for addressing all the aforementioned challenges by providing a set of *AI-Empowered security components* used to ensure that every digital infrastructure operates even in adverse circumstances and can recover quickly following cyber-attacks. The components are used to *predict and detect incidents, automate pentesting, assess ongoing risks, respond to attacks, and recover* digital infrastructure services in an efficient manner. In addition, CyberSecDome provides an interactive *advanced Virtual Reality-based interface* that enhances the Security Operations Center (SOC) analysts' understanding of ongoing attacks and provides situational awareness by visualizing detected threats and associated risks in real-time. Finally, CyberSecDome facilitates effective *collaboration and coordination* among the different stakeholders and cybersecurity agencies to prevent widespread disruption due to the domino effect of cyber-attacks and to coordinate sophisticated large-scale incident response strategies.

The CyberSecDome project, funded in 2023, aims to build and demonstrate a aferomotion framework. This project is driven by a multidisciplinary consortium of 15 partners from six EU member states (Italy, Germany, Ireland, Sweden, Greece, and Cyprus) and two affiliated countries (UK and Switzerland). The project will conduct two internal pilot tests—one with OTE (telecommunications) and the other at Athens Airport. Additionally, an open call will invite external pilots to further test and evaluate the solution.

After one year, we believe the project has reached an *intermediate stage*, allowing us to present the technical status of various framework components and the specific challenges they aim to address. In this paper, we describe how the CyberSecDome Framework handles incidents. The first step involves identifying hidden vulnerabilities through automated penetration testing and assessing the risks found (see Sec. II). During runtime, the framework continuously monitors for attacks, performs intelligence-driven investigation, and proposes mitigation strategies in a dynamic, adaptive manner (see Sec. III). To enhance analysts' understanding, a virtual reality interface is utilized for situational awareness (see Sec. IV). Finally, for effective collaborative incident handling across different domains safeguarding various infrastructures, privacy-aware incident sharing is employed (see Sec. V).

II. Incident Discovery and Assessment

A. Automated Penetration Testing

Penetration testing has been shown to be a reliable method for both vulnerability assesment and security testing of software systems. The first component of the CyberSecDome framework is responsible for automated penetration testing, operating as a black-box tool that conducts tests on digital

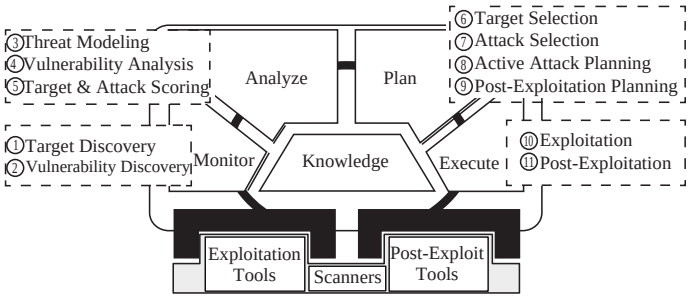

Fig. 2: Automated penetration testing component architecture

infrastructure without any prior knowledge of its architecture or functionalities. Its capabilities include attack planning and attack automation. It automates: (i) system component/asset/host and interface/service/process discovery, (ii) entry point discovery and initial access through remote exploitation, (iii) horizontal and vertical privilege escalation, (iv) persistence, pivoting and, command and control.

The automated penetration testing tool is implemeted following an autonomic manager architecture [3]. The general architecture is shown in Fig. 2. Autonomic architectures comprise two subsystems, the managing subsystem that corresponds to the top part of the figure and the managed subsystem which makes up the bottom level. The managing subsystem acts as the decision-making engine that guides the penetration testing process, while the managed subsystem is reconfigurable and provides the required tools to implement the decisions made by the managing system. The managed system corresponds to a collection of (i) network and vulnerability scanning, (ii) exploitation and attack, and (iii) post-exploitation tools that are dynamically selected and configured at runtime by the managing system to achieve its goals.

The managing system implements a MAPE-K architecture based on a Monitoring, Analysis, Planning and Execution loop with a shared Knowledge base (MAPE-K) as is common for autonomic systems [4]. The *knowledge-base* acts as the central storage and reference point that provides the rest of the MAPE-K components with access to the models, data and information they require to perform their functions. The *monitoring phase* is responsible for host and vulnerability discovery, which are achieved through a combination of domain-specific probes and sensors that observe the target system and its environment and a general enough in expressiveness system and attack model that is able to capture and analyze a variety of different system characteristics, including th the systems' behavior, their vulnerabilities, exploitation paths and network or process interactions. The *analysis phase* is responsible for the analysis and prioritization of the threats and vulnerabilities of each host. It implements two primary functions: (i) threat modeling and (ii) vulnerability analysis. Threat modeling is achieved by keeping the system and attack models updated at runtime. Vulnerability analysis is performed by analyzing the system and attack models to identify the most valuable targets and the attacks that have the highest chance of success against said targets. During the *planning phase*, the managing subsystem performs target and attack selection, first selecting a target (host) to be attacked and then a suitable attack to be performed against the selected host. Targets and attacks are selected by employing a decision theoretic evaluator that considers the

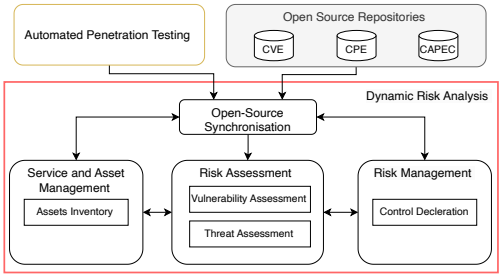

Fig. 3: Dynamic risk analysis component architecture

information stored in the knowledge base as well as the results of the analysis and prioritization performed by the analysis phase components. Once an attack or exploit is selected, the *execution phase* is responsible for configuring, coordinating, and directing the exploitation tooling to successfully penetrate the selected target and to perform any post-exploitation actions required for stealth, persistence, and further information gathering of the now successfully exploited target.

The automated penetration testing component is an information source to dynamic risk assessment. The penetration test's concrete results can be used to dynamically adjust the risk to assets or update the asset and vulnerability information when previously unknown attack paths are discovered. Finally, the penetration test results can be visualized through the VR interface by overlaying them over the topology of the infrastructure to be protected to provide increased situational awareness by highlighting the attack paths that have been demonstrated to be available to an external attacker.

B. Dynamic Risk Assessment

The main goal of this component is (i) identifying the valuable information about the system components and their vulnerabilities, (ii) revealing threats that may take advantage of those vulnerabilities to endanger the system, and (iii) estimating the possible damage and potential losses resulting from those threats [5]. The Dynamic Risk Assessment (DRA) contains the following subcomponents: (1) *Open-Source Synchronisation*: it aims to automatically update the security-related information, such as platform, weakness, vulnerability, and threat, detailed from the open-source repositories and the automated penetration testing to support the other DRA functionalities. This information is essential for identifying the accurate weaknesses, threats and vulnerabilities relevant to the assets within the digital infrastructure., (2) *Servie and Asset Management*: it aims to gather information about the identified assets. Generally, assets are identified by investigating the available services that are necessary to run the organisational business. DRA utilises the NIST CPE catalogue as shown in open intelligence, where known components are catalogued with a specific CPE URI, which can be considered as asset-specific details. Each identified asset is described in detail, specifying its type (software, hardware, operating system, or information asset), sensitivity (restricted or unrestricted), criticality (essential, required, or deferrable), as well as vendor, product, and version details. Additionally, cyber assets dependencies are also taken into consideration to visualise the asset dependencies., (3) *Vulnerability Assessment*: it identifies the confirmed vulnerabilities that are linked to the identified assets. DRA incorporates an AI-enabled model to determine the exploitation of the vulnerability, which is further used for the vulnerability assessment. Therefore, the Exploit Prediction Scoring System (EPSS) value integrates exploit attributes with temporal attributes from various data sources, such as ExploitDB and GitHub to forecast potential exploitation, which is further used by DRA and adopted by the AI model, to determine the probability of vulnerability exploitation., (4) *Threat Assessment*: it identifies and assesses the threats, which exploit the vulnerabilities related to the services and assets. Individual threats are considered potential stepping stones to security incidents (deliberate or accidental), which may affect the identified services and assets. The identified threats can be categorised through threat taxonomies and assessed in a qualitative manner using threat scales. DRA considers the widely known threat catalogue, i.e., Common Attack Pattern Enumeration and Classification (CAPEC), from open intelligence to assess threats. It provides an up-to-date catalogue of known threats, which raises overall security awareness. Each vulnerability entry-related CWE ID is enumerated through NVD, whereas in CWE, for each entry, the related CAPEC IDs are enumerated, and (5) *Control Declaration*: it aims to mitigate the potential consequence of the identified threats and related risks. It is a decision-making process to choose the right control strategy and suitable recommended controls. It implements a mechanism based on the NIST Security and Privacy Controls 800-53 Rev.5, which responds to the need by embarking on a proactive and systemic approach to identify and implement the safeguarding security and privacy measure.

III. Incident Detection and Response

A. Intrusion Detection and Prediction System

Accurate, rapid, and adaptive intrusion detection is a critical feature for any network system, whether commercial or otherwise. A key foundation of these essential characteristics in cybersecurity defense mechanisms is the use of tools designed to identify and generate alerts about malicious activities or breaches of established policies. Modern intrusion detection systems must address the following key features of advanced platforms: (i) real-time data analysis to effectively mitigate cybersecurity threats, (ii) robust predictive capabilities against emerging and unknown cybersecurity risks, (iii) high accuracy in detecting intrusions, and (iv) scalability to accommodate complex infrastructures.

The CyberSecDome Intrusion Detection and Prediction System (IDPS) is designed to monitor and analyze network traffic in real time, detecting cybersecurity intrusions and predicting emerging anomalies. The system comprises two parallel components—intrusion detection and intrusion prediction tools—that communicate and exchange information regarding detected intrusions. The Intrusion Detection System (IDS), shown in Fig. 4, is an Field Programmable Gate Array (FPGA) based tool [6], that offers real-time cybersecurity threat detection. Its primary function is to monitor network traffic, identify potential attacks, and generate corresponding security alerts. The IDS operates in three phases. First, the system undergoes pre-training using IDS signatures stored in its internal database. Next, the real-time network data processing begins on the hardware platform. The Dome network traffic is forwarded to the IDS, where a parallel hardware-based

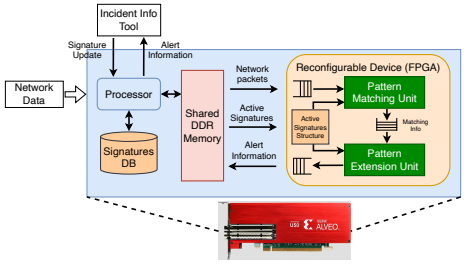

Fig. 4: Internal architecture of Intrusion detection component

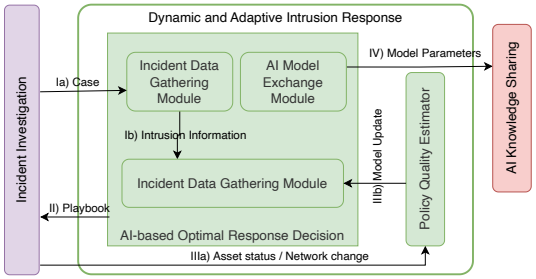

Fig. 5: Internal architecture of DAIR

pattern-matching algorithm [7] runs. If a match between the incoming data and the pre-loaded signatures is found, the network data that triggered the alert is sent to the Incident Info tool, which collects the alert details and forwards them to the Incident Investigation tool. Last, upon validation of the alert by the Incident Investigation tool, feedback is provided to the IDS module. Based on this feedback, the internal signature database is either updated with new signatures—if the alert was triggered by the Intrusion Prediction tool but missed by the IDS—or revised by removing outdated signatures in the case of a false positive, or left unchanged. The IDS processing stages are executed in a real-time streaming manner, leveraging the high-performance parallelism provided by reconfigurable hardware.

The Intrusion Prediction System (IPS) is a tool that utilizes advanced Machine Learning (ML) methodologies to detect potential attacks based on historical data. The process begins with the development of an ML model using training data. Specifically, the IPS employs AutoML [8] to generate optimized and fine-tuned models, ranging from simple linear models and decision trees to more complex boosting models (e.g., Linear Models [9], Decision Trees [10], Random Forests [11]). The AutoML component refines its machine learning pipelines through Bayesian Optimization [12], a technique that explores the hyperparameter space to identify the most effective configurations. Once the training phase is complete, the ML model is deployed to make predictions, such as identifying suspicious activity. The results of these predictions include the predicted class along with the associated probabilities of the event in question belonging to any of the available classes (e.g., malicious or benign). This information is transmitted to the Incident Info tool and, from there, relayed via a message broker to the Incident Investigation tool. Feedback from the Incident Investigation tool is then used to update the IPS ML models.

B. Incident Investigation

The Incident Investigation component (IIC) gathers diverse logs from different endpoints within the system, as well as from the integrated intrusion detection components. Beyond facilitating aggregation and archival of incident details, it leverages AI/ML capabilities to thoroughly examine each incident, uncovering hidden relationships between incidents and adding context and valuable insights. Focused alerts are then pushed to other system subcomponents for appropriate handling. Specifically, the system integrates an autoencoder, deep neural networks (DNN), and Long Short-Term Memory (LSTM) models. This approach processes Security Information and Event Management (SIEM) alert data to detect anomalies, refine risk scores, and analyze attack patterns over time, creating a comprehensive solution for threat detection and response.

After anomaly detection and clustering of similar alerts, DNN enhances risk classification by evaluating whether an alert's severity should be escalated based on recent attack patterns. The DNN model uses input features, such as latent representation , time since the last attack, and frequency of prior attacks. This model, composed of dense layers with ReLU activation functions, outputs a risk score that signals whether the alert severity should be heightened.

All similar alerts are consolidated into a *case* and shared with the Dynamic and Adaptive Incident Response (DAIR). Subsequently, the IIC receives a *playbook* containing proposed responses for the identified threats, which is then forwarded to SOC analysts. Additionally, the component updates the DAIR with any changes in network configurations or asset statuses.

C. Dynamic and Adaptive Incident Response

Playbooks are commonly used as a standard set of operational procedures to conduct incident responses [13]. By providing objective steps, they can be effective in mitigating cyber attacks [14]. One drawback of such playbooks is the human-based selection of the correct playbook for an ongoing intrusion. Playbooks can, therefore, be created by an Intrusion Response System (IRS) to provide timely playbooks and respond to incidents. Especially *adaptive* IRSs are important to dynamically adjust the response selection to changed environments [15]. At CyberSecDome we develop a *manual response system* that provides administrators or SOCs with automatically generated playbooks to enable the selection of a set of appropriate responses. Incorporating ML algorithms allows the creation of dynamic and adaptive playbooks inside the DAIR component.

As shown in Fig. 5, DAIR consists mainly of two components: The optimal response decision module will decide on an optimal response based on the alert of the preceding IIC. Based on the gathered incident information, it will incorporate an ML algorithm to select optimal responses. After providing the playbook, it will update its internal parameters with the help of the policy quality estimator. This feedback loop is typical for reinforcement learning algorithms [16]. DAIR is based on Q-learning, a model-free reinforcement learning algorithm [16] to adapt to dynamic changes in the system state via the feedback loop. The benefit of this reinforcement learning algorithm is that it can converge towards an optimal policy. Q-learning is based on so-called Q-values inside a Q-table that are adapted

through the feedback loop so that a new, optimized policy can be used in future incidents. Although an initial training phase improves the performance at the beginning of the deployment, it is not mandatory since the system learns the optimal behavior based on the feedback from the SOC.

Inside DAIR, the Q-table will have the dimensions of *amount of overall responses* $\times$ *amount of unique alerts*. For each of these entries, one Q-value represents the score. While the Q-table is empty at the beginning, a possible training process and the subsequent application, including the feedback loop, will fill the Q-table. The result of the ML model is a playbook containing optimal response suggestions as a compilation of existing steps that can be applied to the respective system. While the individual steps perform the action, DAIR selects which ones to use and in which order they should be called.

IV. eXTENDED REALITY

With the increase of digital infrastructure complexity and the vast volume of security events gathered by various cybersecurity tools, it is becoming challenging for security analysts in SOCs to have a comprehensive understanding of the system's status, which is essential for efficiently responding to incidents. eXtended Reality interfaces present a promising solution to this challenge by offering a high-bandwidth, immersive environment that allows cybersecurity professionals to visualize, interpret, and interact with complex data structures more intuitively. Their three dimensions plus time and multiple modalities such as sound, or tactile, enable the strongest non-inversive link between the human brain and algorithms. By leveraging XR's ability to present information spatially and interactively, SOCs operators gain a more comprehensive and immediate understanding of threat landscapes, effectively reducing cognitive load and enhancing situational awareness.

CyberSecDome provides advanced visualizations of cybersecurity data, displaying assets and metadata, dependencies, and topological information. Assets, which include hosts, services, and cyber-physical components (e.g., airport systems), are visualized along with their interdependencies, such as software supply chains. Topological data highlights both functional dependencies and physical locations, enabling precise evaluation of affected areas and countermeasures.

Figure 6 illustrates the central components of CyberSecDome's Virtual Reality interface, designed to optimize the interaction between security analysts and the complex datasets provided by security tools. At the foundational level lies the managed system, representing the actual IT infrastructure and assets being monitored and protected. Security tools continuously gather and generate data from this system, including threat indicators, system logs, and network activity. However, rather than relying on the distinct and often fragmented native interfaces of these individual tools, CyberSecDome's Virtual Reality interface consolidates relevant data into a unified, immersive environment, streamlining tasks and reducing the cognitive load for analysts. Recognizing that human security analysts are accustomed to traditional interfaces—and that some tasks are more efficiently handled with conventional methods—CyberSecDome offers Virtual Reality-based interfaces selectively, focusing on critical tasks where immersive visualization provides a distinct advantage.

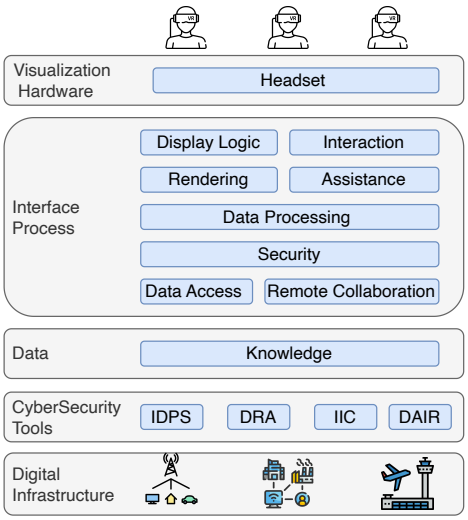

Fig. 6: Central logic components of the VR interface

The data from security tools is accessed and processed by the *Data Access* component, which acts as a bridge between the raw data sources and the Virtual Reality environment. In conjunction with the *Security* Module, this component manages the secure flow of data, handling both the retrieval of relevant information and, if necessary, transmitting inputs from Virtual Reality users back to the security tools. This bidirectional communication ensures that analysts can interact dynamically with the data, making adjustments and initiating responses directly within the Virtual Reality interface. A unique capability of the Virtual Reality environment is enabled by the *Remote Collaboration* Module, which allows geographically dispersed analysts to collaborate in real-time as if they were in the same room. The *Security* Module not only safeguards data access but also manages user authentication and the security of remote connections. Ensuring robust identity and access management within the Virtual Reality environment is essential, given that impersonation risks are elevated in virtual spaces. Meanwhile, the *Data Processing* module plays a critical role in transforming raw data into formats that facilitate analysis. The *Rendering* module provides flexible and interactive visualization options for data by offering multiple rendering templates that users can configure in real-time, allowing for tailored data views suited to specific analytical tasks. The *Display Logic* adapts these visualizations to various output devices, such as Virtual Reality headsets with differing resolutions and fields of view, ensuring a consistent experience regardless of the hardware. User interaction is facilitated by the *Interaction* Module, which interprets inputs from multiple modalities, including gestures, voice commands, and traditional controllers. This module enables analysts to manipulate and explore data intuitively within the Virtual Reality environment, further enhancing user engagement and efficiency. To support users in navigating complex data, the *Assistance* Module provides real-time guidance and recommendations, helping them focus on critical insights and navigate complex visualizations effectively. Finally, at the top of this layered system are the security analysts themselves, equipped with advanced Virtual Reality tools that empower them to work with greater speed, accuracy, and insight. CyberSecDome's Virtual Reality interface transforms the way

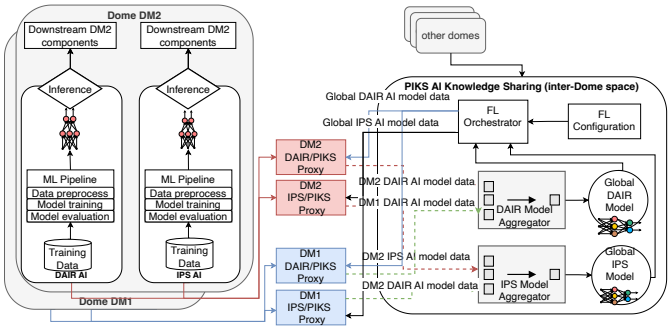

Fig. 7: Internal architecture of AIKS

analysts engage with data, providing them with a powerful, immersive platform that not only accelerates task completion but also enhances precision in identifying and mitigating cyber threats.

V. Privacy-Aware Incident Sharing

CyberSecDome enables secure collaboration between organizations through Privacy-aware Information Sharing (PIKS), an inter-dome interface. It focuses on two areas: AI Knowledge Sharing, which uses federated learning for AI model training without sharing sensitive data, and Threat Intelligence Sharing, which facilitates the exchange of cyber threat information using privacy-conscious protocols.

A. AI Knowledge Sharing

The component enables AI model training across multiple organizations without sharing sensitive data. This approach allows each participating entity to improve its local AI models while contributing to a globally trained model, all without compromising data privacy. As shown in Fig. 7, federated learning (FL) is employed to collaboratively train AI models for threat detection across multiple domes and their SOCs. Each SOC maintains control over its private, locally stored data and uses its own machine learning pipeline to train models based on internal cybersecurity events and threat patterns. A proxy service coordinates the training process, ensuring that only model parameters (e.g., weights and gradients) are shared between organizations. These parameters are aggregated on a central FL server to update the global model.

The key components of our implementation include: (1) *Decentralized Data Handling*: Each organization retains its sensitive data, ensuring that no raw data leaves its premises. Only model updates are transmitted, (2) *Training Coordination*: A FL round involves distributing the global model to all organizations to train it locally for several epochs. Subsequently, the local updates are sent back to the central FL server, which aggregates them to refine the global model. Then a new FL round can start, (3) *Privacy and Security*: We employ techniques, such as differential privacy and secure aggregation, to protect the model weights during transmission [17], (4) *Interoperability*: The system supports various machine learning frameworks (e.g., PyTorch, TensorFlow, scikit-learn), ensuring flexibility in the types of AI models that SOCs can use for local training, and (5) *Cross-Silo Collaboration and Horizontal FL*: The architecture accommodates cross-silo collaboration between multiple organizations, allowing them to participate in horizontal federated learning, where the feature space is consistent across entities, ensuring that models can be effectively generalized across diverse datasets and real-world security threats.

The collaborative and live training approach ensures that the global model benefits from a rich diversity of threat intelligence, making it highly robust and capable of addressing a wide array of security threats. Additionally, the system is scalable, allowing multiple organizations to join the federated network, continuously improving the global model's effectiveness in real-world threat scenarios. Nonetheless, challenges like handling non-IID (non-independent and identically distributed) data [18] and ensuring effective collaboration across organizations persist. Moving forward, the project will explore swarm learning to enhance decentralization and peer-to-peer model sharing [19].

B. Threat Information Sharing

Effective cyber threat intelligence (CTI) sharing across systems is essential. Various standards and protocols facilitate CTI exchange while preserving each organization's sensitive data through privacy-preserving techniques. In Thread Information Sharing component, two protocols and one standardization method are employed to facilitate the dissemination of CTI information exchange.

- *TAXII*: A protocol that enables the exchange of CTI information via RESTful API [20].
- *Traffic Light Protocol (TLP)*: A protocol for classifying the sensitivity of information being shared [21].
- *STIX*: A standardized language used for representing and sharing CTI information [22].

In the CyberSecDome system, each Dome is equipped to detect, evaluate, and respond to a wide range of cybersecurity threats. The Threat Information Sharing module is responsible for disseminating threats that possess distinctive characteristics and features. Upon the emergence of a new threat, the Threat Information Sharing module receives this threat, which is labeled with a TLP label. In case a label does not pertain to confidential information, the data is disseminated to the TIS modules of the other Domes. Consequently, the remaining Domes are able to access this information and enhance new threats with this data. Furthermore, the TIS module continuously monitors various CTI repositories and sources, employing ML-based techniques to curate relevant information for threat enrichment.

VI. Conclusion

Protecting our digital infrastructure from cyberattacks is essential for keeping our daily lives running smoothly. This work presented the CyberSecDome framework, a solution designed to help SOC analysts manage complex cyber threats. CyberSecDome provides tools for finding and assessing incidents, detecting threats in real time, investigating and responding to attacks, and enhancing situational awareness with VR tools. It also supports secure, privacy-conscious information sharing, helping different organizations work together to handle incidents effectively.

Acknowledgment

This work is supported by the European Union-funded project CyberSecDome (Agreement No.: 101120779).

REFERENCES

[1] Digital Watch Observatory, "Ukraine conflict: Digital and cyber aspects," 2023, accessed: 2024-10-27. [Online]. Available: https://dig.watch/trends/ukraine-conflict-digital-and-cyber-aspects

[2] Bitkom e.V., "German businesses under attack: losses of more than 220 billion euros per year," 2021, accessed: 2024-10-27. [Online]. Available: https://www.bitkom.org/EN/List-and-detailpages/Press/German-business-losses-more-than-220-billion-euros-per-year

[3] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," *Computer*, vol. 36, no. 1, pp. 41–50, Jan. 2003. [Online]. Available: http://dx.doi.org/10.1109/MC.2003.1160055

[4] Y. Brun, G. Di Marzo Serugendo, C. Gacek, H. Giese, H. Kienle, M. Litoiu, H. Müller, M. Pezzè, and M. Shaw, *Engineering Self-Adaptive Systems through Feedback Loops*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 48–70.

[5] M. Hamad, A. Finkenzeller, M. Kühr, A. Roberts, O. Maennel, V. Prevelakis, and S. Steinhorst, "REACT: Autonomous intrusion response system for intelligent vehicles," *Computers & Security*, vol. 145, p. 104008, 2024.

[6] D. Deyannis, E. Papadogiannaki, G. Chrysos, K. Georgopoulos, and S. Ioannidis, "The diversification and enhancement of an ids scheme for the cybersecurity needs of modern supply chains," *Electronics*, vol. 11, no. 13, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/13/1944

[7] E. Papadogiannaki, G. Chrysos, K. Georgopoulos, and S. Ioannidis, "A reconfigurable ids framework for encrypted and non-encrypted network data in supply chains," in *2023 International Conference on Engineering and Emerging Technologies (ICEET)*, 2023, pp. 1–6.

[8] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.

[9] N. Draper, *Applied regression analysis*. McGraw-Hill. Inc, 1998.

[10] W.-Y. Loh, "Classification and regression trees," *Wiley interdisciplinary reviews: data mining and knowledge discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[11] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[12] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010.

[13] Cybersecurity and Infrastructure Security Agency, "Cybersecurity Incident & Vulnerability Response Playbooks," Nov. 2021. [Online]. Available: https://www.cisa.gov/sites/default/files/2024-08/Federal_Government_Cybersecurity_Incident_and_Vulnerability_Response_Playbooks_508C.pdf

[14] A. Applebaum, S. Johnson, M. Limiero, and M. Smith, "Playbook Oriented Cyber Response," in *2018 National Cyber Summit (NCS)*. Huntsville, AL: IEEE, Jun. 2018, pp. 8–15.

[15] M. Hamad, M. Tsantekidis, and V. Prevelakis, "Intrusion response system for vehicles: Challenges and vision," in *International Conference on Smart Cities and Green ICT Systems*. Springer, 2019, pp. 321–341.

[16] S. E. Li, "Principles of RL Problems," in *Reinforcement Learning for Sequential Decision and Optimal Control*. Singapore: Springer Nature Singapore, 2023, pp. 15–40.

[17] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, p. 50–60, May 2020. [Online]. Available: http://dx.doi.org/10.1109/MSP.2020.2975749

[18] Q. Li, Y. Diao, Q. Chen, and B. He, "Federated learning on non-iid data silos: An experimental study," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022, pp. 965–978.

[19] E. T. Martínez Beltrán, M. Q. Pérez, P. M. S. Sánchez, S. L. Bernal, G. Bovet, M. G. Pérez, G. M. Pérez, and A. H. Celdrán, "Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 4, pp. 2983–3013, 2023.

[20] "TAXII version 2.1," OASIS Standard, June 10 2021. [Online]. Available: https://docs.oasis-open.org/cti/taxii/v2.1/os/taxii-v2.1-os.html

[21] F. of Incident Response and S. T. (FIRST), "Traffic light protocol (tlp) version 2.0," Standard Definition, August 2022, accessed: 2024-10-24. [Online]. Available: https://www.first.org/tlp/

[22] "STIX version 2.1," OASIS Standard, June 10 2021. [Online]. Available: https://docs.oasis-open.org/cti/stix/v2.1/os/stix-v2.1-os.html