Baltic J. Modern Computing, Vol. 10 (2022), No. 4, 689-709
https://doi.org/10.22364/bjmc.2022.10.4.06

Use Cases and Design of an Intelligent Intrusion Detection System

Evita ROPONENA[1], Jānis KAMPARS[1], Jānis GRABIS[1], Guntis MOSĀNS[1], Andris GAILĪTIS[2]

[1] Riga Technical University, Zunda Krastmala 10, Riga, LV-1048, Latvia
[2] SIA "Izglītības sistēmas", Cuibes iela 17, Riga, LV-1063, Latvia

evita.roponena@rtu.lv, janis.kampars@rtu.lv, grabis@rtu.lv, guntis.mosans@rtu.lv, andris.gailitis@e-klase.lv

Abstract. Information and communication technologies (ICT) play an important role in almost any business sector and in all aspects of modern society. Data centres are used to host ICT systems and, therefore, are also targets of cyberattacks. ICT security measures are necessary to protect information from unauthorized access. The Human-in-the-Loop approach states that cybersecurity specialists should be continuously involved in automated intrusion detection activities and should be supported by suitable tools to evaluate them. This paper proposes an overall design of the intelligent intrusion detection system with a focus on big data analysis, machine learning, knowledge management, and supporting cybersecurity specialists. The use cases and typical users of a cybersecurity system are defined to specify the requirements of such a system. The architectural design is presented that includes components and technologies supporting implementation of the system. Active learning and learning from evaluation are selected to fulfil the requirements of the Human-in-the-Loop approach.

Keywords: intrusion detection, cybersecurity, big data, human factors

1. Introduction

Cyberinfrastructure is subjected to unauthorized third parties or cyberattacks. Data centres are typically used as a host for various ICT systems and, therefore, are typical targets of cybercriminals and security threats. Threats such as viruses, malware, ransomware, spyware, spam, phishing, DDoS, and other related threats are frequent at the data centres, what shows the importance of real-time network monitoring and threat prevention (Shammugam *et al.*, 2021). According to the industry report (Trustwave, 2020), phishing threats increased from 3% in 2018 to 9% in 2019, and also the amount of compromised data has increased. In 2019, the number of C2 botnet servers has increased by 57% compared to 2018, which is usually associated with distributed denial-of-service (DDoS) attacks (ENISA, 2020). Detecting and responding timely to intrusions are essential for tackling these threats.

However, continuous monitoring of enterprise Information systems (IS) and network data, as well as incident identification and threat prevention, impose serious performance and scalability requirements on the ICT security management solution. Therefore, the

most commonly used security management solutions address this issue by processing only a portion of the available data. The available ICT security management solutions have limited capabilities to provide a full-scale complex analysis of IS systems, network data, server logs, and other unstructured and semi-structured data; therefore, the identification and mitigation of security threats is not comprehensive.

A more comprehensive monitoring and analysis of security concerns leads to elevated requirements toward cybersecurity specialists, and lack of cybersecurity specialists is an important challenge (ISACA, 2022). Currently, the management of ICT security incidents is performed by cybersecurity personnel, and the identification of the causes of the incident and the speed of response depend directly on the experience and competence of the professionals. They need in-depth knowledge of networks, their protocols and devices, device architecture, vulnerabilities, as well as cybersecurity tools, and the effectiveness of their application. Big data technologies and machine learning are promising for automating the intrusion detection. However, the participation of the human decision-maker is still important (Zhang *et al.*, 2022). The Human-in-the-Loop Security Model argues that automated machine learning-based approaches should be combined with manual intervention models to deal with low-confidence problems. Therefore, a user-oriented intrusion detection system that combines automated detection techniques with user-friendly approaches to visualize, explain, and share intrusion detection data is needed.

The objective of this paper is to formulate requirements and to elaborate design of the intelligent intrusion detection system with the focus of supporting cybersecurity specialists to fulfil the requirements of the Human-in-the-Loop approach. Typical users of the proposed Intelligent Intrusion Detection System are defined, and their use cases and interactions are identified. Architectural components supporting the implementation of the system are identified. The work is conducted as a part of the university and industry collaboration in an applied research project, and the results will serve as a basis for further implementation of the system. The proposed system is known as BICTSeMS. This paper is an extended version of the following article (Roponena *et al.*, 2022).

The rest of the paper is organized into 8 sections. Characteristics of current intrusion detection systems and the Human-in-the-loop approach are reviewed in Section 2. The requirements are identified in Section 3. The overall design of the proposed systems is presented in Section 4. The technology evaluation is presented in Section 5. The implementation of the Human-in-the-loop approach is presented in Section 6. Section 7 contains the conclusion, and Section 8 – acknowledgments.

2. Background

Intrusion detection systems (IDS) monitor network traffic by scanning a network or a system to identify suspicious activity and to generate alerts when suspicious activity is discovered. Measurement criteria such as false positives (unmalicious activity identified as malicious), false negatives (malicious activity identified as unmalicious) are used to evaluate IDS performance.

IDS can be classified into five types:

1. Network-based IDS (NIDS) - monitors packets from the network,
2. Host-based IDS (HIDS) - analyse the audit data of the operation system,

3. Protocol-based IDS (PIDS) - is located in the front end of a server and controls and interprets the protocol between user/device and server to secure a web server,
4. Application protocol-based IDS (APIDS) - is located in a group of servers and monitors and interprets the communication on application-specific protocols,
5. Hybrid IDS - is a combination of two or more approaches.

IDS uses two different detection methods: the signature-based method or the misuse detection and the anomaly-based method. Misuse detection identifies attacks based on specific patterns or signatures. It uses information on previously detected malware. One of the main problems of signature-based IDS is a large number of signatures in the database, leading to possible misses of dangerous attacks and the inability to detect unknown attacks. The anomaly-based method uses machine learning to detect previously unknown malware. The disadvantage of this method is that non-malicious behaviour can be identified as an incident that leads to a high false positive rate. Both intrusion detection methods can be combined to overcome the disadvantages of these methods (Aydın et al., 2009). In various research papers, IDS are supplemented with additional methods to improve IDS performance and overcome its limitations.

The new generation IDS that can correlate information from various resources to identify bots in the network before they inflict damage is proposed by (Catalin and Cristian, 2017). The solution consists of the Snort network intrusion detection system that performs real-time traffic analysis and packet logging and Splunk that identifies data captures, correlates real-time data, and generates visualizations.

The Signature Apriori algorithm can be used to implement the misuse detection technique and data mining in IDS. The main concept of the algorithm is to use prior knowledge of the properties of the frequent itemset. This method is used in (Zhengbing et al., 2008) research with an additional Scan-Reduction method to minimize scans of the database and by the observation of the attack signatures determine which attack signatures depend on each other and identify the new attacking signature. The system analyses data gathered from Snort and packet sniffer to find frequent items that meet minimum support, and afterwards checks all possible combinations with already known signatures. The proposed algorithm is faster than the Signature Apriori algorithm with a moderate false positive rate ~25%.

Open-Source Intelligence (OSINT) is a technique that may be used to automatically update IDS knowledge by collecting knowledge about threats from various sources. In (Vacas et al., 2018), the OSINT data are used to generate rules and blacklists that are later integrated into an IDS in the IDSoSint prototype. The proposed approach acts continuously in a loop, each iteration containing three phases: information gathering and categorizing into predefined threat categories, knowledge generation by correlating gathered information, and incident detection managed by the Pulledpork platform and updating of the Snort IDS. The proposed approach can successfully generate rules and blacklist entries using OSINT feeds and is able to eliminate traffic based on blacklists and then emit the remaining incidents through Snort's rule processor.

The enhanced signature-based IDS is proposed in the research (Alyousef and Abdelmajeed, 2019). The proposed solution resolves signature-based IDS disadvantages by distributing signature database in the several small databases based on a protocol type; by using a filtering engine, and by updating engine. The IDS captures an incoming packet, extracts its signature, and matches the extracted signatures with the signatures stored in the databases. If there is a match, the alert is sent, and the packet is blocked; if

not, the packet goes through the filtering engine, which checks the similarity of the signatures of the new packet and the stored in databases, and whether the new IP is stored on the IP blacklist. If the packet passes those criteria, it is clean; if not, it is blocked, and the signature databases and blacklist are dynamically updated.

Poltavtseva, Zeghda and Pavlenko (2019) propose an associated graph set (AGS) model-based NIDS architecture for enterprise network. The model requires the decomposition of the host into sets of MAC addresses, IP addresses, and ports associated with the host. An address of each level can be associated with one or more addresses of neighbouring levels. These sets specify a mapping function for network associations, up to the OSI protocol stack, and the analysis of these mappings allows one to define the node's internal configuration and enter security validation rules. Any attack on each OSI layer makes changes to the graph structure and can be detected. The hosts are classified according to the communication role, and the information can be used to collect data about the system. Information can be divided into structural data and communication data. The AGS NIDS architecture includes two layers of analysis modules: a module to identify attacks by known data mining methods based on behaviour, specifications, or knowledge, and a module to generate the system, system behaviour, and structure rules based on the first analysis.

The brief overview shows that different approaches can be used for IDS improvement, for example, by combining multiple open-source IDS systems.

Table 1 shows the overview of the most popular open-source, free, and real-time NIDS or HISD that performs data analysis and uses both anomaly-based and signature-based detection methods.

Table 1. The overview of selected open-source IDS

IDS	Type	Logs	Features
Zeek IDS[1] (Bro IDS)	NIDS	SNMP traffic, FTP, DNS, HTTP activity, events	Event Engine, Policy scripts
Snort[2]	NIDS	Network traffic packets	Packet logger, sniffer, and IDS modes, Rule-based configuration, Plugin framework
OSSEC[3]	HIDS	Alterations, mail, FTP, Web server data	Creation of important files checklist and validating it
Suricata[4]	NIDS	DNS, SMB, FTP, HTTP, UDP, TLS, TCP, and ICMP protocols	Integration with third-party tools like Anaval, Squil, Uses rule sets
Security Onion[5]	NIDS, HIDS	Network and host logs	Focus on log management, enterprise security monitoring. Includes tools such as Elasticsearch, Locstash, Suricata, Zeek, and others.

[1] https://zeek.org/
[2] https://www.snort.org/
[3] https://www.ossec.net/
[4] https://suricata.io/
[5] https://securityonionsolutions.com/

Table 1 shows that different IDS systems can be used in combination to analyse different network logs and obtain a complete picture of the incident. Each IDS system also has a different set of features that can be used; therefore, it is important to understand the requirements of the system to select the most suitable open-source solution if needed.

The overview of current IDS improvement solutions does not show how to integrate human interactions with the cybersecurity analysis system for the creation of the Human-in-the-loop cybersecurity system. The current IDS systems focus on technical aspects, while the importance of usability issues has increased recently. Even though artificial intelligence solutions are able to process big data quickly and provide good recognition results, these solutions are not flexible and are not able to perfectly adapt to new changes in the network or systems; therefore, human involvement can improve cyber security system performance.

The Human-in-the-loop methods can be categorized into four groups based on the nature and level of control that the human has during the learning process (Rajendran *et al.*, 2021),. These groups are active learning, learning from demonstration, learning from intervention, and learning from evaluation. In the learning from demonstration approach, the main goal is to train an artificial intelligence agent to mimic the human behaviour in a particular task. In the learning-from-intervention phase, the human takes over control during the agent-learning phase when it enters a catastrophic state. Both approaches are based on human action analysis, but the cybersecurity system learns from data, therefore, these approaches are not suitable for our system. Active learning is a semi-supervised approach where a subset of the training data is labelled, and a human interactively labels data points from the unlabelled set based on his or her knowledge when the model needs the data point. Considering that data used to detect threats in the cybersecurity system, for example, network flow, are usually unlabelled data, cybersecurity specialist knowledge could be used to label it, for example, to train the model on historical real-life data with known infected devices or threats. In the learning from evaluation approach, the human provides real-time feedback to the system, thereby shaping its policy. This approach is useful to review the system's ability to identify data anomalies and improve systems knowledge base for further use.

Zhang *et al.* (2022) defines that a Human-in-the-loop cybersecurity system should have a machine detection module (MDM), knowledge base, confidence level module (CLM), and manual intervention module (MIM). In this solution, the MDM performs data preprocessing, feature extraction, and recognition using two different machine learning methods to improve accuracy and to generate the judgment results in parallel using the knowledge base to gain the judgment basis. The MIM starts when the MDM provides unsatisfactory results, in this module the safety specialist handles the event according to his knowledge and adds a new type of event to the knowledge base to improve the MDM. The CLM connects both MDM and MIM and ensures the cooperation between these modules by determining whether MIM should be called to complete the event preprocessing based on the MDM result confidence level. The confidence level is low if two machine learning methods provide different results or the overall accuracy is low.

It is hard to explain the reasoning behind the solution provided by artificial intelligence. Machine learning models, especially artificial neural network models, are referred to as black boxes, because the algorithm's process behind decision making is

not visible and is unexplainable. Hsupeng *et al.* (2022) utilizes the open-source tool SHAP (Shapley Additive exPlanations), which provides the approach to understand the relationship between the model and data features, thereby, providing explainable AI. Szczepanski *et al.* (2020) states that IDS should deliver reliable predictions about potential threats, easy to understand explanations about its decisions, adapt towards new challenges and does not have a negative impact on performance. They add an explainer to the classification algorithm to meet these requirements.

Visualization also plays an important role in supporting the analysis and comprehension of cyber-incidents (Karami, 2018). Breve, Cirillo and Deufemia (2020) specifically focus on user interactions with IDS and provide automation of configuration activities including configuration by means of using voice activated commands.

3. Requirement Analysis

The requirements towards the Human-in-the-loop cybersecurity system are formulated to understand the required functionality of the proposed system. These requirements are elicited based on analysis of existing intrusion detection systems (Section 2), literature review of automated cybersecurity solutions (Roponena *et al.*, 2021), IT security standards, and interviews conducted with representatives from an ICT company managing data centres.

3.1. Overall Approach

The BICTSeMS platform is envisioned as an intelligent intrusion detection system that supports cybersecurity specialists. The threat prevention activity refers to the detection and blocking of malicious attacks on the ICT infrastructure (Fig. 1). It is supported by best practices providing information on the type of attacks and suitable response mechanisms. To support cybersecurity specialists, threat information is contextualized and represented in relation to the overall network topology uncovered and maintained using automated methods. The topology modelling describes relationships between the elements of the network and shows the potential attack vectors. Component topology graphs are retrieved from a cloud computing platform using a specialised adapter. The topology preprocessor extracts subgraphs (subsets of connected devices, such as one VLAN) from the entire topology graph and checks whether changes have occurred in the subgraphs. If changes are detected, a new version of the subgraph is saved in the Topology dataset.

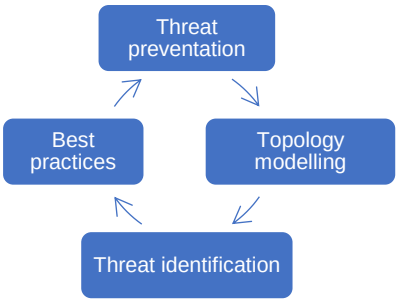

Fig. 1. BICTSeMS overall approach

Threat identification using big data analysis and machine learning methods is continuously run in the background to identify new threats and generate warnings and recommendations to human decision makers. Positive warning and recommendations are documented as best practices for providing information for treat detection. Therefore, the BICTSeMS platform provides automated threat analysis, contextualization, and accumulation of knowledge, which is also used to help cybersecurity specialists monitor and manage the ICT infrastructure.

3.2. Use Cases

The BICTSeMS system is designed as a semiautomated system that requires human interactions to perform the security management of ICT components. The identified possible system users are the following:

1. incident response team that performs intrusion detection, advisory distribution, education and awareness, information sharing, and incident response (Cichonski et al., 2012), it usually includes:
   1.1. team leader – coordinates the incident response team,
   1.2. lead investigator – performs evidence analysis,
   1.3. communication lead – provides necessary communication inside and outside of the company,
   1.4. documentation and timeline lead – documents team activities and tasks,
   1.5. legal representative - ensures that incident response activities are in line with laws and regulations,
2. system administrator – responsible for the maintenance of computer systems, for example, servers,
3. BICTSeMS administrator – responsible for user management, system performance, and system configuration of the BICTSeMS system.

In this system context, incident response team manages incident identification and prevention processes and maintains ICT security management best practices repository. The system administrator is responsible for the isolation of the infected system area and is responsible for enforcing the security measures of the ICT components. The system itself performs automatic data analysis, threat detection, and prevention if the instant actions are needed.

The BICTSeMS system provides different functionality to maintain a high ICT security level, this functionality is summarized into five use cases:

Use Case 1: Real-time full-scale integration of system data: Provides reception, preprocessing and aggregation of ICT system components data used for security threat identification.

Use Case 2: Identification of security threats: Performs system cybersecurity threat identification based on system data analysis provided by Use Case 1.

Use Case 3: Security threat prevention: Performs cyber security threat prevention based on threat identification data.

Use Case 4: Management of ICT security management best practices repository: Provides a summary of automated or recommended actions to ensure that the level of security is restored.

Use Case 5: Creation of ICT components topology model: Provides a system ICT component topology model that includes a database with IP addresses and servers to determine the area of influence of ICT device interdependence and security incident.

Three of the use cases (Real-time full-scale integration of system data, Creation of ICT components topology model and Management of ICT security management best practices repository) are described in more details in the following subsections.

3.2.1. Real-Time Full-Scale Integration of System Data

The objective of real-time full-scale integration of system data is to preprocess ICT system data from various data sources (Fig. 2). The results of the analysis are passed to the BICTSeMS threat identification module for threat detection.

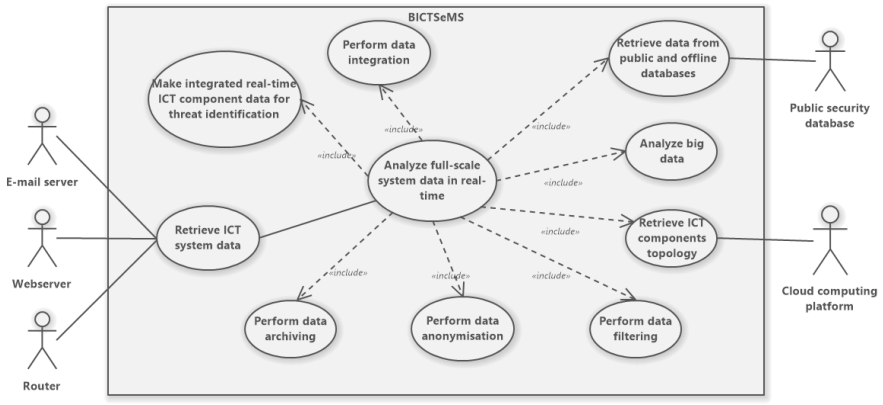

Fig. 2. Use case diagram for real-time full-scale integration of system data

This use case describes the ICT system data analysis performed by the BICTSeMS real-time data integration module using data from different sources, for example, network data, server log files, firewall log files, email log files, graph-based features, etc. that will be used to identify possible cyber security threats in the system. The data centre is built on the TCP/IP network model that can be divided into four layers: physical, link, network, and application layers, and each of them has its own security risks, and each of them produces different data (Wang and Lu, 2018). Graph-based features represent the interaction of system components; these features can be used to identify the neighbours of the infected device (Wang *et al.*, 2020). According to (Sharma *et al.*, 2011) usage of various data sources is important to understand the full picture of the events in the system, their correlation and sources. Therefore, the cyber security management system should be able to perform parallel processing of a huge amount of log data that have a different structure (Breier and Branišová, 2017). The data analysis should be done in real-time to be able to detect threats instantly and react to them (Minkevics and Kampars, 2020).

Data filtering is performed to support the context of the data that is used to interpret the data and retain interesting events from large data sets. For example, by filtering data by a timeframe, the system is able to identify the correlation of the events in this timeframe, determine the attack path and its source, and exclude unrelated data. Filtering is often used in combination with prioritization that helps determine the importance of the event and whether it should be investigated (Cinque et al., 2020).

The use case actors include relevant subsystems:

1. Public security database – contains IP and DNS addresses blacklists that are used for the analysis,
2. E-mail server, webserver, router – an example of data sources for full-scale system analysis,
3. Cloud computing platform – contains the ICT component topology of the system.

3.2.2. Creation of System Components Topology Model

The objective of this use case is to define the topology model of ICT components in a form of a graph for subsequent use of different components of the system (Fig. 3). The use-case postconditions include making the versioned ICT topology data available for other use cases.

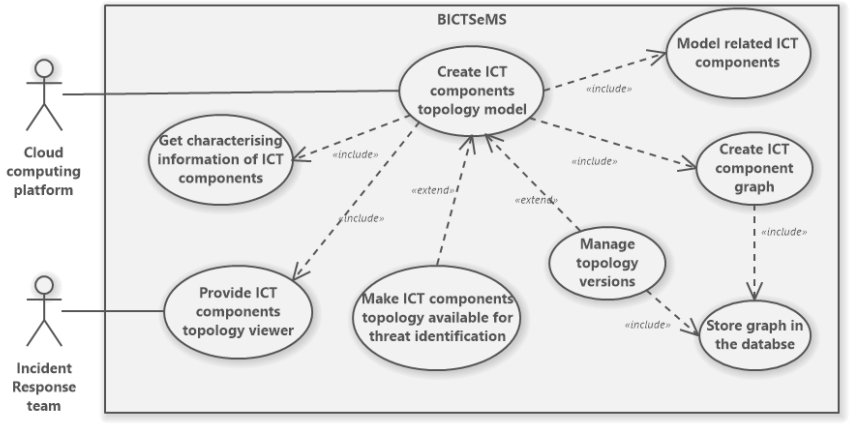

Fig. 3. Use case diagram for creation of a system components topology model

This use case describes the creation and management of the ICT components topology model. Network topology consists of nodes (users, devices, etc.), edges (relations) and degree of the node (for example, number of neighbours). The attack on certain node can cause not only disruption of this node but also could infect neighbours, therefore, topology graph is excellent way to see the relations of the nodes. When the network is large, it consists of many nodes, and it is difficult to model the interactions among all nodes. One of the solutions is to create a population game that is often used to model the strategic interaction between many players (La, 2014).

Modelling of network topology is especially helpful for botnet detection, allowing one to see how the network components change their connections and overall behaviour (Trovati *et al.*, 2017). Using graph-based features retrieved from network topology graph overcomes the limitations of using only most common flow-based features like source and destination IPs, protocol, and others that cannot capture the systems components interactions and analyse complex communication patterns (Daya *et al.*, 2020). For better analysis, graph-based and flow-based features should be used together.

The creation of ICT component graphs contains the model of data centre virtual environments (virtual machines, containers, hypervisors, etc.). The virtual environment is evolving dynamically, therefore, it is not recommended to use static topology graph

for the system data analysis, threat identification and prevention, and ICT security management best practice repository. The graph is created and versioned by the graph database management system.

The use-case actors include both users of the system and relevant subsystems:

1. Incident response team – views the topology of ICT components,
2. Cloud computing platform –used to create ICT component topology model.

3.2.3. Management of ICT Security Best Practices Repository

The main objective of the management of the ICT security best practices repository is to store the best security practices to use for threat prevention and to improve the security of the ICT system (Fig. 4). The use case post-conditions include providing threat prevention action recommendations and the review of the provided feedback on pattern suitability.

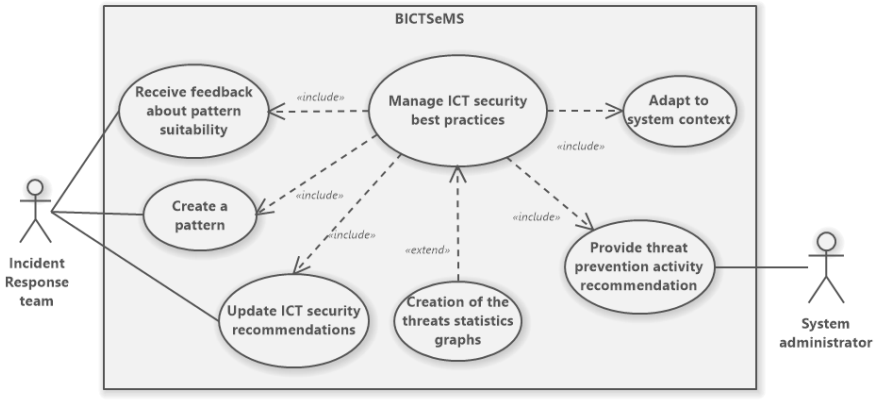

Fig. 4. Use case diagram for the management of ICT security best practices repository

This use case describes the management of the ICT security management best practice repository. Each company should have created its own disaster recovery plan designed to minimize the impact of possible disaster events on the business, including cybersecurity event recovery strategies (Srinivasan and Simna, 2017). The cybersecurity event recovery document should contain information about the people, processes, components, and dependencies between them, categorizations for all components based on their importance, identified key personnel, recommendations for managing different cyber events and recovery systems after them, and strategies for system and threat identification (Bartock *et al.*, 2016). While dealing with the incident, all information about its handling should be documented and stored in an archive (ENISA, 2010). In our case, to improve the ICT security management best practice repository, incident handling documentation could be stored there.

Identification of the behaviour pattern of ICT components in a cybersecurity context can be used to identify abnormal activity of the component by comparing the normal activity of the device with the current activity (Gu *et al.*, 2008). For example, when the device is targeted by a botnet, it usually starts to perform abnormal actions like

performing large amounts of TCP (Transmission Control Protocol) requests. Therefore, the analysis of behaviour patterns helps detect anomalous behaviour in the system.

The best practice repository will include rule-based patterns that can be easily prevented and that can lead to successful cyberattack. For example, enabling Telnet allows the user ID and password to be transmitted without any encryption. Therefore, to secure the information, it should be disabled. To evaluate the usability of the given pattern, incident response team is able to provide feedback on its suitability.

The use-case actors include:

1. Incident response team – creates initial patterns, updates ICT security recommendations, and provides feedback on pattern suitability,
2. System administrator – uses best practices from the repository to choose the best incident mitigation actions.

Fig. 5. Overall architecture

4. Overall Design

The overall design of BICTSeMS is elaborated according to the requirements following the service-oriented approach. The main components are BICTSeMS backend services, infrastructure services, adaptable ICT components, user interface workspaces and components, roles, and data sources (Fig. 5). The data sources component is responsible for the capture and pre-processing of email server, web server, DHCP, switch, router, and other cloud computing platform data. The backend services support the BICTSeMS approach and implement the specified use cases. The user workspaces and components present the core actions to the users.

4.1. BICTSeMS Roles, Workspaces, and Components

A workspace concept is used to provide a customized work environment for specific users according to their needs. The user interface consists of seven workspaces. The Team lead workspace provides cumulative statistics of incidents and prevented incidents in the form of charts. The administrator workspace allows administrators to view detected incidents and corresponding preventive actions and review recommended preventive actions. The BICTSeMS administrator workspace allows the BICTSeMS administrator to manage users, show system performance, and provide system configuration. The Communication lead workspace provides a list of threats and best practices to identify people involved in incidents and communicate with the necessary people. The Documentation lead workspace shows threat identification and threat prevention feedback to the documentation lead and allows the documentation lead to provide additional feedback if necessary, and the Legal representative workplace provides a list of all prevented threats and the parties concerned. The workspaces are composed of reusable components:

- Threat aggregation component – provides data for all incidents per day, type of incident, history of incidents, etc.;
- Threat prevention summary component – provides aggregated data on incident prevention;
- User management component – allows to configure, define and manage users and user groups across the platform;
- Platform settings component – allows editing and reviewing platform settings;
- Platform resources component – allows reviewing system resources and memory for the platform;
- Threat identification component – provides notifications and threat identification list;
- Threat prevention component – provides a list of all prevented threats;
- Topology modelling component – provides ICT device topology and its changes over time;
- De-anonymisation component – provides de-anonymisation of ICT component related data (e.g., MAC address, IP address);
- Best practices component – provides information about manual actions for restoring the level of security;

- Access log component – allows one to see the performed de-anonymisation requests by the platform users;
- Threat identification feedback component – provides feedback for identified threats (such as false positives);
- Threat prevention feedback component – provides feedback on recommended threat prevention actions.

The components mentioned above provide communication with the back-end services of BICTSeMS. The main BICTSeMS services are the real-time data integration service, the threat identification service, the threat identification and prevention service, best practise repository service, and topology modelling service. There are also IP blacklist service, anonymisation service, dashboard service, and administration service, which will provide auxiliary functions for the main services. The other services provide data collection from devices, servers, and cloud computing platforms.

4.2. BICTSeMS Backend Services and Data Sources

The real-time data integration service provides the reception, preprocessing, and aggregation of large amounts of real-time and static data, obtaining the availability of information for use in machine learning models. The service receives the data from topology modelling service, which contains topology IP mapping stream containing mapping of anonymised IPs to subgraph graph identifiers and IP blacklist preprocessor service, which provides IP blacklist dataset containing anonymised blacklisted IPs from publicly available databases. The service also receives data from data collection microservices:

- The E-mail log preprocessor service - subscribes to the Full E-mail log stream provided by the Syslog integration service.
- The Outbound GET request preprocessor service - subscribes to the Outbound GET stream provided by the Traffic analysis services.
- The Router logs preprocessor service - reads data from the Full router log stream, which is provided by the Syslog integration service.
- DHCP logs preprocessor service - reads the Full DHCP log stream provided by the Syslog integration service.
- The DNS analysis service - decrypts the DNS stream received from the Traffic analysis service, and using machine learning tries to determine whether the resolved hostname is generated by domain generation algorithm, which could potentially indicate that the device is in the communication with the head of the botnet and has been infected.
- The Server head request preprocessor service - subscribes to the Queried web server hostname stream provided by the Web server log preprocessor service. Information about the active hostnames is extracted from the stream and stored in the Active web server hostnames dataset by the Server hostname update job.
- The Port scanner service - periodically reads information about active IP addresses from Active IPs dataset and decrypts it using application-level key.
- The NetFlow preprocessor service receives NetFlow data from a switch via the NetFlow aggregator service. Information about active IP addresses is written to the Active IPs dataset used by the Port scanner service.
- The Web server log preprocessor service - subscribes to the Full E-mail log stream provided by the Syslog integration service.

The mentioned data sources are aggregated and sent to the Aggregated ICT descriptive data stream used by the Threat identification service. Also, data collection services use the Anonymisation service, which minimises the risks of exposing private information such as behavioural characteristics of a particular IP address. It provides hashing of items such as IP address, hostname, domain name, GET request path sections (e.g., last parameters or hostname). It is also used in threat prevention service, IP blacklist preprocessor service and Topology modelling service.

The topology modelling service ensures the definition of the topology model of ICT components in the form of a graph and the further use of the obtained graph for a more accurate identification of security threats and selection of the most appropriate adaptive actions. IP addresses in the subgraphs are anonymised prior to saving data using the Anonymisation service. Anonymised IP address and revision ID of the corresponding subgraph are passed to the Topology IP mapping stream which is used by the Real-time data integration service. The stored topologies can be retrieved from the API, and this functionality is used by the Topology modelling component.

The threat identification service provides specialised automated actions to identify previously known security threats (e.g., there is an unauthorised port open on the server) and previously unknown threats and suspicious activity using machine learning models. The component receives aggregated ICT component related data from the real-time data integration module and detects component-level threats based on predefined rules. The measured threat levels are then passed to the ICT component level threat stream which is processed by the Topology level threat interpretation job. It evaluates the aggregated threat level within the subgraph using a machine learning model and stores the information about detected threats in the Detected threats dataset.

The threat prevention service subscribes to both ICT component level threat stream and Topology level threat stream. The threat prevention service performs automated adaptations to the ICT components or informs the responsible person in response to the identified threats and ensures the restoration of the security level. Information about the most appropriate threat prevention action is retrieved from the Best practices repository service. It includes automatic actions such as reconfiguration of the Web server, E-mail server, Firewall, and manual actions. The information about the manual action that needs to be performed with reference to the pattern repository and threat related data is passed to the Communication lead and/or Documentation and timeline lead and/or System administrator. The mentioned roles are also informed about the execution of automatic security level restoration actions. Information about prevented threats is stored in the Threat prevention dataset and is also passed to the Threat prevention stream, which is processed by the Dashboard service.

The ICT security management best practice repository service contains automated actions or recommended actions for the responsible person to ensure that the level of security is restored in response to identified threats. The service provides API for selection of the most appropriate security level restoration action based on the detected threat, which is used by the Threat prevention service. The API also allows to provide feedback about the suitability of the selected threat prevention action and to get descriptive information of a pattern based on its ID (used by the Best practices component).

The administration service is used for monitoring and configuring infrastructure service such as Message broker (sensitive and GDPR compliant data), Distributed data storage (sensitive and GDPR compliant data), Authorization and role management service, Graph database, Job execution engine (stream processing, batch processing and

scheduled cronjobs). The Dashboard service gathers statistics about identified and prevented threats for analysis purposes. It reads thread identification information from the ICT component level threat stream and the Topology level threat stream provided by the Threat identification service.

4.3. BICTSeMS Infrastructure Services

Infrastructure services provide data transmission, messaging, data storage, and security services that the core services need. Infrastructure services are designed using open-source tools to achieve a high degree of scalability and portability. The message brokers are used for NetFlow data and system log streaming and service communication between each other. There are unlimited flows of NetFlow data and system logs and the BICTSEMS platform needs to schedule several kinds of jobs involving stream processing (windowing operations, aggregations on top of data streams) and processing of static data from distributed data storage. The Job execution engine provides continuous data processing. The data flow is intensive, and distributed data storage is required in order to ensure continuous data storage and analysis. Graph databases store the network topology. The reverse proxy and load balancer on the BICTSeMS platform is expected to redirect requests to the backend services. The BICTSeMS platform requires identity and access management, which is intended to be provided by a dedicated authentication provider. The service will provide role-based access management to backend microservices.

The architecture meets the identified requirements and conforms to the principles of the Human-in-the-loop security system (Zhang *et al.*, 2022). The threat prevention and identification services provide autonomous threat identification and prevention activities. The Best practice repository service provides the knowledge base. The topology modelling service provides explanatory capabilities. Users use customized workplaces which provide threat detection and prevention oversight.

5. Technology Evaluation

A technology evaluation was performed to select the most suitable technologies for the development of the BICTSeMS system. The evaluation process was based on the Analytical Hierarchy Process (AHP) method for organizing and analysing complex decisions (Hummel et al., 2014). The technology evaluation process consisted of several steps:

1. Identification of the evaluation criteria based on (Lněnička, 2015), (Belinda *et al.*, 2021),
2. Selection of the final criteria,
3. Evaluation of the selected criteria – using pairwise comparison (Fig. 6),
4. Technology evaluation – using the five-point scale, where 1 is the lowest compliance with the criteria, and 5 – the highest,
5. Technology selection.

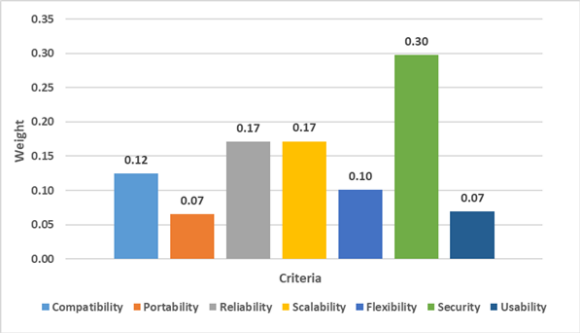

Fig. 6. Weights assigned to the defined criteria

The evaluation of the criteria and four selected technology groups (network data ingest technologies, document database, real-time streaming platforms, development and analytics tools) were carried out by 8 experts from Riga Technical University, who are experts in software development and integration or cybersecurity. As a result of technology evaluation, the most suitable technology for each group was obtained. All selected technologies are open-source technologies:

1. Network data ingest technologies - software tools like GoFlow, Fprobe, and SoftFlowBD,
2. Document database - Apache Cassandra,
3. Real-time streaming platforms - Apache Spark,
4. Development and analytics tools – Python programming language.

Software tools such as GoFlow, Fprobe, and SoftFlowBD provide a high level of flexibility and portability and are distributed under permissive licences. They support different transports, conversion to other formats, decoding of different samples, and different metrics. Apache Cassandra is an open-source, row-oriented, NoSql database that is distributed, decentralized, highly available, and has high performance. Apache Spark is an open-source, general-purpose, scalable, and in-memory computing platform. Spark is a powerful tool that reduces processing time by using memory instead of HDFS. Python is an interpreted high-level general-purpose programming language, and it has a wide variety of data analysis and machine learning libraries usable that are also compatible with Apache Spark.

6. Implementation of the Human-in-the-Loop Approach

The BICTSeMS system is designed as hybrid IDS system that uses various machine learning models to identify unknown threats and rule-based detection to identify known threats, e.g. configuration deficiencies. Learning from evaluation and active learning approaches are implemented to ensure that the system is the Human-in-the-loop cybersecurity system (Fig. 7).

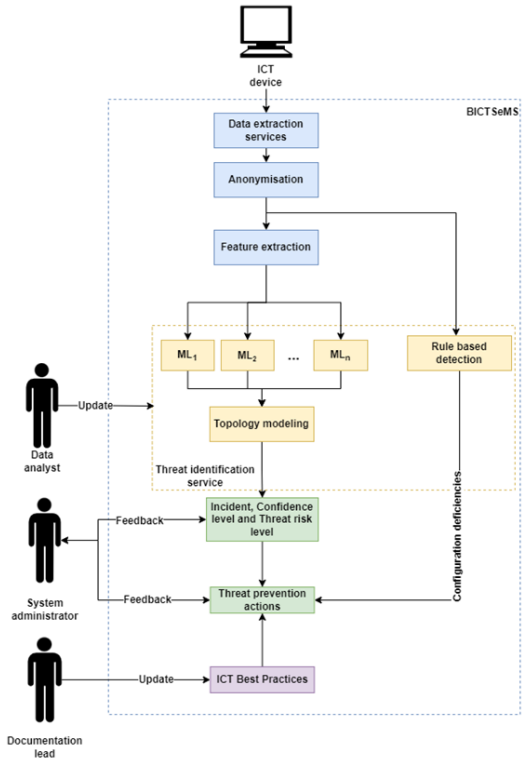

Fig. 7. BICTSeMS Human-in-the-Loop approach

In our approach (Fig. 7) data from the ICT devices are extracted and anonymised. If the extracted data are used for the rule-based detection where the Threat identification service is searching for configuration deficiencies or open ports, the system immediately after threat detection initializes threat prevention actions. In other cases, data are used to extract features for unknown threat detection using several machine learning models that provide component-level threat identification for each ICT component that are further used for the topology-level threat identification.

The topology level threat identification provides threat the risk level that measures the impact of the identified threat on the neighbouring ICT components and the confidence level that measures worthiness of results provided by machine learning models (e.g., precision, voting result, etc.). Human actions are closely integrated into the whole threat identification and prevention process (Fig. 8).

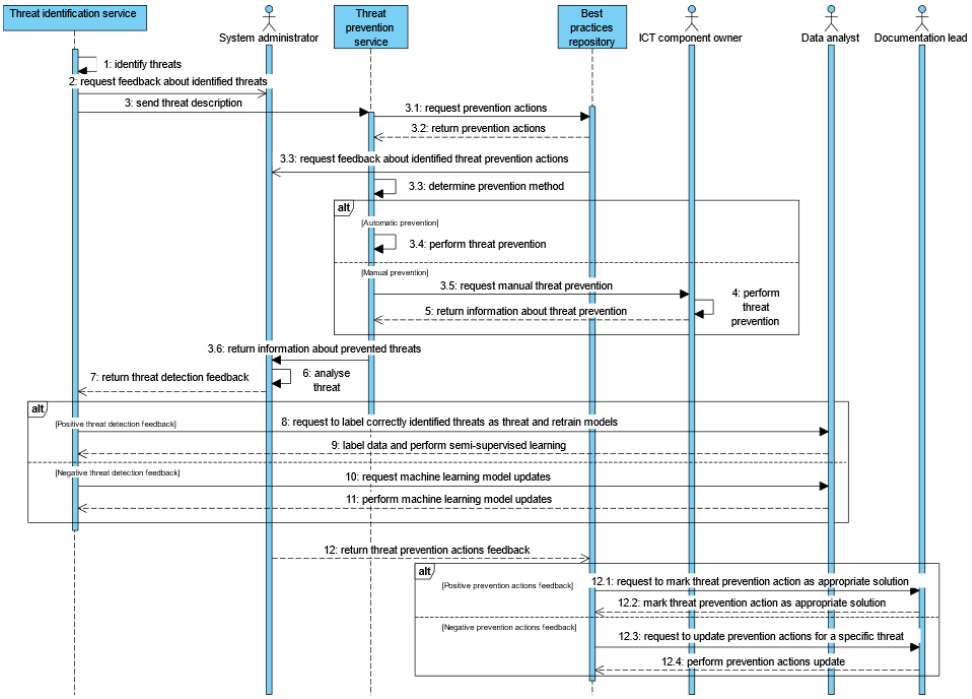

Fig. 8. BICTSeMS Human-in-the-Loop approach sequence diagram

As shown in Fig. 8, the Threat prevention service and the Best practices repository request feedback about an identified threat and preventive actions immediately after it was identified; however, the system does not wait until the feedback is send back and continues to prevent threats to ensure that ICT devices are protected. The system administrator provides feedback on the identified threats to ensure that the machine learning models did not provide false positive results, and the recommended threat prevention actions to ensure that they are up-to-date and suitable for the specific threat prevention. If the feedback on identified threat is positive, data analyst is requested to label correctly identified threat to retrain and improve machine learning models using semi-supervised learning method. If it was negative, the data analyst performs a retraining of the machine learning models labelling the falsely identified threat as normal data. If the feedback on threat prevention actions is positive, the documentation lead marks the action as suitable for this threat. If not, documentation lead performs the update of the ICT best practices repository adding new threat prevention actions or improving not suitable actions.

The threat prevention process may be done using automatic prevention methods or manual, where an ICT component owner who is a part of the incident response team prevents the threat using actions provided by the Best practices repository or his/her own knowledge if the repository does not have information about the specific threat prevention actions.

7. Conclusion

The paper proposes the BICTSeMS security management platform as a user-oriented solution for automated and intelligent intrusion detection. At the same time, it follows the Human-in-the-loop approach to involve human decision-makers and to support their activities. One of the key challenges is finding the right balance between automated intrusion detection and human involvement to maintain efficiency and continually develop the cybersecurity competences of users.

The proposed solution involves humans in the BICTSeMS threat identification and prevention cycle using active learning and learning-from-evaluation approaches. Active learning is implemented by using threat identification feedback results as data labels that are used to retrain machine learning models. Learning from evaluation is performed by requesting feedback about identified threats and invoked threat prevention actions and using the feedback to improve the BICTSeMS threat identification and prevention process.

The BICTSeMS platform provides user-friendly means to present contextualized cybersecurity information and support users with intrusion detection knowledge maintained in a best practice repository. The identified requirements and the overall architecture designed serve as the main inputs to ongoing implementation of the platform.

Acknowledgements

This research is funded by European Regional Development Fund Project Nr. 1.1.1.1/20/A/020 “Development of Big-data-driven Information and Communication Technology Security Management Solution (BICTSeMS)” Specific Objective 1.1.1 “Improve research and innovation capacity and the ability of Latvian research institutions to attract external funding, by investing in human capital and infrastructure” 1.1.1.1. measure “Support for applied research” (round No.4).

References

Alyousef, M.Y., Abdelmajeed, N.T. (2019) Dynamically detecting security threats and updating a signature-based intrusion detection system’s database, *Procedia Computer Science*, 159, pp. 1507–1516. Available at: https://doi.org/10.1016/j.procs.2019.09.321.

Aydın, M.A., Zaim, A.H., Ceylan, K.G. (2009) A hybrid intrusion detection system design for computer network security, *Computers & Electrical Engineering*, 35(3), pp. 517–526. Available at: https://doi.org/10.1016/j.compeleceng.2008.12.005.

Bartock, M., Cichonski, J., Souppaya, M., Smith, M., Witte, G., Scarfone, K. (2016) Guide for cybersecurity event recovery, *NIST Special Publication*, pp. 800–184.

Belinda, B.I., Emmanuel, A.A., Solomon, N., Kayode, A.B. (2021) Evaluating Software Quality Attributes using Analytic Hierarchy Process (AHP), *International Journal of Advanced Computer Science and Applications*, 12(3), pp. 165–173. Available at: https://doi.org/10.14569/IJACSA.2021.0120321.

Breier, J., Branišová, J. (2017) A Dynamic Rule Creation Based Anomaly Detection Method for Identifying Security Breaches in Log Records, *Wireless Personal Communications*, 94(3), pp. 497–511. Available at: https://doi.org/10.1007/s11277-015-3128-1.

Breve, B., Cirillo, S.. Deufemia, V. (2020) *An Intrusion Detection Framework for Non-expert Users (S)*. Available at: https://doi.org/10.18293/DMSVIVA20-014.

Catalin, M., Cristian, A. (2017) An efficient method in pre-processing phase of mining suspicious web crawlers, in *2017 21st International Conference on System Theory, Control and Computing, ICSTCC 2017*, pp. 272–277. Available at: https://doi.org/10.1109/ICSTCC.2017.8107046.

Cichonski, P., Millar, T., Grance, T., Scarfone, K. (2012) Computer Security Incident Handling Guide : Recommendations of the National Institute of Standards and Technology, *NIST Special Publication*, 800–61, p. 79. Available at: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf.

Cinque, M., della Corte, R., Pecchia, A. (2020) Contextual filtering and prioritization of computer application logs for security situational awareness, *Future Generation Computer Systems*, 111, pp. 668–680. Available at: https://doi.org/10.1016/j.future.2019.09.005.

Daya, A.A., Salahuddin, M.A., Limam, N., Boutaba, R. (2020) BotChase: Graph-Based Bot Detection Using Machine Learning, *IEEE Transactions on Network and Service Management*, 17(1), pp. 15–29. Available at: https://doi.org/10.1109/TNSM.2020.2972405.

ENISA (2010) Good Practice Guide for Incident Management, *Enisa*, p. 110.

ENISA (2020) *Threat Landscape 2020 - Botnet*, *European Union Agency for Cybersecurity*. Available at: https://doi.org/10.2824/552242.

Gu, G., Perdisci, R., Zhang, J., Lee, W. (2008) BotMiner: Clustering analysis of network traffic for protocol- and structure-independent botnet detection, *Proceedings of the 17th USENIX Security Symposium*, (January 2008), pp. 139–154.

Hsupeng, B., Lee, K.W., Wei, T.E., Wang, S.H. (2022) Explainable Malware Detection Using Predefined Network Flow, *International Conference on Advanced Communication Technology, ICACT*, 2022-Febru, pp. 27–33. Available at: https://doi.org/10.23919/ICACT53585.2022.9728897.

Hummel, J.M., Bridges, J.F.P., IJzerman, M.J. (2014) Group Decision Making with the Analytic Hierarchy Process in Benefit-Risk Assessment: A Tutorial, *The Patient - Patient-Centered Outcomes Research*, 7(2), pp. 129–140. Available at: https://doi.org/10.1007/s40271-014-0050-7.

ISACA (2022) *State of Cybersecurity 2022: Global Update on Workforce Efforts, Resources and Cyberoperations*.

Karami, A. (2018) An anomaly-based intrusion detection system in presence of benign outliers with visualization capabilities, *Expert Systems with Applications*, 108, pp. 36–60. Available at: https://doi.org/10.1016/j.eswa.2018.04.038.

La, R.J. (2014) Role of network topology in cybersecurity, *Proceedings of the IEEE Conference on Decision and Control*, 2015-Febru(February), pp. 5290–5295. Available at: https://doi.org/10.1109/CDC.2014.7040216.

Lněnička, M. (2015) AHP Model for the Big Data Analytics Platform Selection, *Acta Informatica Pragensia*, 4(2), pp. 108–121. Available at: https://doi.org/10.18267/j.aip.64.

Minkevics, V., Kampars, J. (2020) Methods, models and techniques to improve information system's security in large organizations, *ICEIS 2020 - Proceedings of the 22nd International Conference on Enterprise Information Systems*, 1(Iceis), pp. 632–639. Available at: https://doi.org/10.5220/0009572406320639.

Poltavtseva, M.A., Zeghda, D.P., Pavlenko, E.Y. (2019) High - performance NIDS Architecture for Enterprise Networking, *2019 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom*, pp. 2019–2021.

Rajendran, P.T., Espinoza, H., Delaborde, A., Mraidha, C. (2021) *Human-in-the-Loop Learning Methods Toward Safe DL-Based Autonomous Systems: A Review*, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer International Publishing. Available at: https://doi.org/10.1007/978-3-030-83906-2_20.

Roponena, E., Kampars, J., Gailitis, A., Strods, J. (2021) A Literature Review of Machine Learning Techniques for Cybersecurity in Data Centers, *ITMS 2021 - 2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University, Proceedings* [Preprint]. Available at: https://doi.org/10.1109/ITMS52826.2021.9615321.

Roponena, E., Kampars, J., Grabis, J., Gailitis, A. (2022) Towards a Human-in-the-Loop Intelligent Intrusion Detection System, in *CEUR Workshop Proceedings*, pp. 71 – 81. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134334893&partnerID=40&md5=aea833fad14b8f01938fdd7912c9480a.

Shammugam, I., Samy, G.N., Magalingam, P., Maarop, N., Perumal, S., Shanmugam, B. (2021) Information security threats encountered by Malaysian public sector data centers, *Indonesian Journal of Electrical Engineering and Computer Science*, 21(3), pp. 1820–1829. Available at: https://doi.org/10.11591/ijeecs.v21.i3.pp1820-1829.

Sharma, A., Kalbarczyk, Z., Barlow, J., Iyer, R. (2011) Analysis of security data from a large computing organization, *Proceedings of the International Conference on Dependable Systems and Networks*, pp. 506–517. Available at: https://doi.org/10.1109/DSN.2011.5958263.

Srinivasan, J., Simna, S. (2017) Disaster Recovery, an Element of Cyber Security-a Flick Through, *International Journal of Management*, 8(October), pp. 125–133.

Szczepanski, M., Choras, M., Pawlicki, M., Kozik, R. (2020) Achieving Explainability of Intrusion Detection System by Hybrid Oracle-Explainer Approach, *Proceedings of the International Joint Conference on Neural Networks* [Preprint]. Available at: https://doi.org/10.1109/IJCNN48605.2020.9207199.

Trovati, M., Thomas, W., Sun, Q., Kontonatsios, G. (2017) Assessment of Security Threats via Network Topology Analysis: An Initial Investigation BT - Green, Pervasive, and Cloud Computing, in M.H.A. Au, A. Castiglione, K.-K.R. Choo, F. Palmieri, and K.-C. Li (eds). Cham: Springer International Publishing, pp. 416–425.

Trustwave (2020) *2020 Trustwave Global Security Report*, *Security Report*. Available at: https://www2.trustwave.com/2013GSR.html.

Vacas, I., Medeiros, I., Neves, N. (2018) Detecting Network Threats using OSINT Knowledge-Based IDS, in *Proceedings - 2018 14th European Dependable Computing Conference, EDCC 2018*, pp. 128–135. Available at: https://doi.org/10.1109/EDCC.2018.00031.

Wang, W., Lu, N. (2018) Security risk analysis and security technology research of government public data center, in *Proceedings - 2nd IEEE International Conference on Energy Internet, ICEI 2018*. Institute of Electrical and Electronics Engineers Inc., pp. 185–189. Available at: https://doi.org/10.1109/ICEI.2018.00041.

Wang, W., Shang, Y., He, Y., Li, Y., Liu, J. (2020) BotMark: Automated botnet detection with hybrid analysis of flow-based and graph-based traffic behaviors, *Information Sciences*, 511, pp. 284–296. Available at: https://doi.org/10.1016/j.ins.2019.09.024.

Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F., Choo, K.-K.R. (2022) Artificial intelligence in cyber security: research advances, challenges, and opportunities, *Artificial Intelligence Review*, 55(2), pp. 1029–1053. Available at: https://doi.org/10.1007/s10462-021-09976-0.

Zhengbing, H., Zhitang, L., Junqi, W. (2008) A novel Network Intrusion Detection System (NIDS) based on signatures search of data mining, in *Proceedings - 1st International Workshop on Knowledge Discovery and Data Mining, WKDD*, pp. 10–16. Available at: https://doi.org/10.1109/WKDD.2008.48.

Received November 14, 2022, accepted November 29, 2022