Human-in-the-loop Anomaly Detection and Contextual Intelligence for Enhancing Cybersecurity Management

Cinzia Cappiello
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Italy
cinzia.cappiello@polimi.it

Thomas Schaberreiter
CS-AWARE Corporation OÜ
Estonia
thomas.schaberreiter@cs-aware.com

Jerry Andriessen and Mirjam Paradijs
Wise & Munro
The Netherlands
jerry@wisemunro.eu, mirjam@wisemunro.eu

Alexandros Papanikolaou
InnoSec P.C.
Greece
a.papanikolaou@innosec.gr

Fotios Gioulekas
Athanasios Tzikas
Konstantinos Gounaris
Evangelos Stamatiadis
5th Regional Health Authority of Thessaly & Sterea
Greece
fogi@dypethessaly.gr, atzi@uhl.gr, kgounaris@ghv.gr, vstam@dypethessaly.gr

Abstract

Cybersecurity management is a sociotechnical problem comprising organisational knowledge management of humans and technology. Focusing on risk and incident management, we present our approach for enhancing cybersecurity awareness in organisations and ecosystems. By augmenting our cybersecurity awareness platform with human-in-the-loop anomaly detection and machine learning, we are able to handle the dynamics of organisational human activity, as well as the continuous developments in the cybersecurity domain. We illustrate the potential impact of our approach with a realistic example in the healthcare context.

1 Introduction

Broadly speaking, there are two major cybersecurity management activities for organisations: risk management (including Business Continuity and Disaster Recovery – BC/DR) and incident management. Risk management activities are more static, with risk identification and implementing relevant risk controls and policies performed regularly. In practice, it can be observed that many organisations do not follow formal and proactive risk management processes but only implement relevant processes after a major incident has happened (Securities and Comission, 2023). Incident management involves detecting, diagnosing, and recovering from IT system anomalies caused by accidental or malicious activity. It is mostly the responsibility of the IT personnel in an organisation, either by dedicated staff or as a task of the IT administrators, and it is usually a manual and labour-intensive task.

Cybersecurity management is to a large extent a knowledge management problem. From the relevant knowledge domains, both risk and incident management require a solid understanding of current state-of-the-art practices and developments and the ability to adopt and implement adequate solutions/mitigations in the relevant context (Melaku, 2023). Information about the cybersecurity context is readily and digitally available through, e.g., best practices and procedure documents, threat intelligence, and less obvious sources like social media. In contrast, knowledge about the organisational context is a more complicated story. While there is a lot of knowledge and experience available within an organisation, e.g., how the systems and services work and interact in the day-to-day operation and what steps are taken in order to address (security) issues, this knowledge is rarely formally written down (Jasimuddin and Saci, 2022). Due to European legislation like the General Data Protection Regulation (GDPR) (European Commission, 2016)

Proceedings of the 1st International Conference on NLP & AI for Cyber Security, pages 127–136
July 29–30, 2024.

and the more recent Network and Information Security Directive 2 (NIS 2 (The European Parliament and the Council of the European Union, 2022)), many companies may actually be forced to consider and implement formal risk management for the first time.

Cybersecurity management can benefit from making tacit knowledge in an organisation explicit and digitally available (Cho et al., 2020). Artificial Intelligence and Machine Learning (AI/ML) can play an active role in the knowledge management process and ensure that organisations are made aware of anomalies in their systems that are unique to their systems and business processes. It can make sure that the right information is available to the right person to address or mitigate cybersecurity issues. This paper presents novel uses of AI/ML in cybersecurity management, enabled by the information and data available through the CS-AWARE/CS-AWARE-NEXT cybersecurity management approach.

CS-AWARE/CS-AWARE-NEXT (Andriessen et al., 2022; Luidold et al., 2023) is a research effort that provides a novel approach to cybersecurity management based on a novel socio-technical approach (Kupfersberger et al., 2018) that allows creating an understanding of an organisation that identifies and visualises its social and technical assets and dependencies, as well as the information flows generated by the day-to-day business operations of humans and technology. The approach is designed to be applicable not only to large organisations, but also to smaller organisations, e.g., municipal utility providers or SMEs covered by the NIS2 directive. The CS-AWARE approach provides a platform that enables risk- and incident-management tasks, such as monitoring for anomalies in real-time and defining policies and business continuity/disaster recovery (BC/DR) tasks, and to allow for incident handling (Schaberreiter et al., 2023). The platform includes applications to detect and report the spread of attacks, anomalies, and incidents. For considering the organisational context, we exploit a Human In The Loop (HITL) approach as described in Figure 1. The creation of the applications starts with the requirements collection and analysis (step 1). The requirements in the project are collected periodically through workshops and focus groups. The application's design (step 2) considers the organisational needs and exploits the available knowledge to prepare data and select and design the algorithm in a customised way. Note that the available knowledge includes the internal data (i.e., organisation knowledge) and the ecosystem knowledge shared among several organisations. The ecosystem knowledge includes insights gathered from data contained in public repositories. It includes, for example, cybersecurity news extracted from social media and/or threat intelligence feeds. The design phase is followed by the implementation (step 3) and the go-live of the applications (step 4). In the operation phase, during which the applications are used, the users' feedback is requested to assess the relevance of the sent information and triggers. In this way, the application continuously learns and can automatically evolve over time (step 5). Steps 4 and 5 are repeated during an application lifetime.

The remainder of the paper is organised as follows: Section 2 presents related work. Section 3 details core knowledge management support provided by CS-AWARE/CS-AWARE-NEXT and introduces the resulting organisation/ecosystem-related information sources made available through the approach. Section 4 presents how AI/ML can help to improve the efficiency and effectiveness of cybersecurity management by supporting human-in-the-loop anomaly detection as well as data contextualisation for increased awareness and decision support. Section 5 presents a realistic use-case in the healthcare sector to illustrate how the CS-AWARE/CS-AWARE-NEXT knowledge management and AI/ML can support organisations and ecosystems in cybersecurity management tasks. Finally, Section 6 concludes the paper, discusses the current implementation status, and provides an outlook for future work.

2 Related work

Especially in Europe, cybersecurity management in organisations is increasingly driven by legal requirements. Starting with the European cybersecurity strategy of 2013, updated in 2020 (European Commission, 2020), the European Union has put significant effort into developing a legal framework fostering a more secure cyberspace. For businesses and organizations, the most relevant ones are currently the GDPR and NIS/NIS2, which obliges a significant portion of European organizations to manage cybersecurity in a formal and legally compliant way. Compared to NIS, the scope of NIS2 was extended so that SMEs and smaller utility

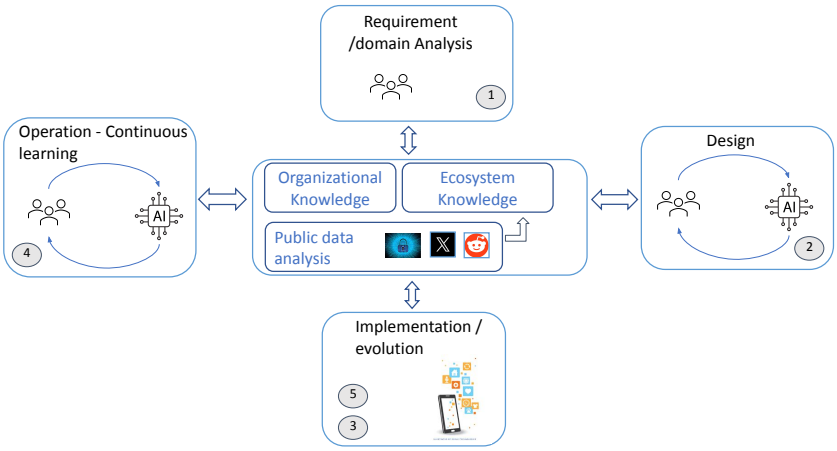

Figure 1: Application Development lifecycle.

providers are also required to follow the directive and implement formal cybersecurity procedures. The GDPR is even more significant, as all organisations that handle personal data of any kind are required to comply with the regulations and protect the data they manage.

There are various approaches to formal organisational risk management, the most prominent among them being the NIST cybersecurity framework (National Institute of Standards and Technology, 2024) and the ISO/IEC 27000 family of standards (ISO/IEC 27000:2016, 2016). They are seen as being mostly applicable to large organisations that have the resources to implement them, and it remains to be seen if smaller organisations (who need to start implementing formal cybersecurity procedures for the first time due to NIS 2) will be able to adopt such approaches at a large scale.

Security Information and Event Management systems (SIEMs) are a widely adopted sophisticated technology for supporting incident management through real-time monitoring for anomalies (Granadillo et al., 2021). However, they predominantly cater to large organisations' needs and are often not justifiable in smaller organisations due to their high cost. Furthermore, current SIEM systems typically only allow monitoring for network and infrastructure-related anomalies caused by technological aspects; more in-depth monitoring that also takes into account the social component of an organisation and the complex behaviours caused by organisation-specific business processes is not available.

The use of AI/ML in cybersecurity management is justified by the fact that conventional data analysis methods have difficulty keeping up with the complexity and speed of modern cyber threats (Shukla et al., 2022). Artificial Intelligence (AI) systems, particularly those using Machine Learning (ML) and big data architectures, have the potential to detect and mitigate these threats. Intelligent cybersecurity management applies various AI methods that eventually seek intelligent decision-making in cyber applications or services (Sarker, 2021). AI (or, more specifically, machine learning) has been widely used in cybersecurity for decades in well-known application areas, including malware detection, intrusion detection, and spam detection. Typically, such algorithms work on security-related data gathered from different relevant sources, such as network behaviour, database activity, application activity, or user activity. In (Sarker, 2021), a survey of ML methods for cybersecurity is provided. Supervised techniques can be mainly used for anomaly detection. They can classify and predict malware attacks or cyber anomalies (e.g., decision trees (Vu et al., 2019), logistic regression, random forest (Leevy et al., 2021)). In general, unsupervised learning can be used to find hidden patterns and structures from unlabeled data (Sarker, 2021). It is possible to use clustering to find groups of similar data (Landauer et al., 2020). Instead, association learning can be used to create recommendations for adopting rule-based machine learning models for incident response and risk management like (Ozawa et al., 2020).

In the CS-AWARE-NEXT project, we are exploiting such algorithms, adding a context-aware perspective. The organisational knowledge and context will be considered to operationalise threat intelligence in organisational risk and incident management.

3 Cybersecurity knowledge management at organisational and ecosystem level

The CS-AWARE approach is rooted in systems thinking, based on the core principle that a system needs to be seen as the sum of its components in order to understand all relevant implications, especially in the cybersecurity context. Looking at an organisation as a system, it is not only composed of infrastructure and services, but also of people that operate and maintain the infrastructure and services. An organisation is a human activity system and in order to understand it in its full complexity, a socio-technical approach is required to capture all the dynamics within an organisation that influence cybersecurity. We have adopted the Soft Systems Methodology (SSM) developed by Peter Checkland (Checkland, 1998; Checkland and Scholes, 1991) for this purpose, in which we work with the people of an organisation (users, technicians, managers, ...) in dedicated workshops to identify the core assets of an organisation, its interdependencies, as well as the detailed information flows that are generated through the different assets in the day-to-day operation of business processes the organisation is concerned with. Furthermore, the process includes identifying relevant monitoring sources – log files available through the different assets that can describe the state of an asset over time – and allows monitoring in real-time for potential anomalies and incidents. The output of this process is used for creating the so-called "system dependency graph". The main advantage of this proven method over other cybersecurity assessment methods is that it unlocks the tacit knowledge of the people working in the organisation as to how the systems really work, operate and are maintained in day-to-day operation, which is almost always very different from what is written down in manuals and documentation. It is essential to derive baselines for customised anomaly detection based on the unique knowledge of the people working with the systems on a day-to-day basis – something that will become relevant in the following sections of this paper.

The CS-AWARE approach brings risk and incident management closer together and makes risk management more dynamic. Traditionally, identifying organisational assets and critical information flows is a risk management task (performed at regular intervals but not dynamically), whereas monitoring for anomalies is an incident management task traditionally performed using different tools like SIEMs.

Notable features of CS-AWARE in this context include:

- The ability to manage knowledge about assets, dependencies, business processes, and information flows and to define context-specific baselines for real-time anomaly detection.
- The ability to monitor and alert not only for anomalies detected by traditional tools (integration with network monitoring tools like Suricata or Zeek or other cybersecurity tools), but to also monitor and alert based on context-specific patterns defined by the users. AI/ML supports this feature and will be further detailed in the following sections.
- The ability to create and assign policies (like security policies as well as BC/DR policies), and monitor the effectiveness and efficiency of those policies in real time.
- Generate threat intelligence in STIX format based on detected incidents and allow sharing of this information in standardised and automated form with authorities, e.g., to fulfill NIS/NIS2 or GDPR information sharing requirements.
- The ability to contextualise detected anomalies with existing threat intelligence to provide awareness as well as suggestions for mitigations to the user and even allow to invoke automatic mitigation of incidents in case a purely technical solution to an incident was detected (a feature we call self-healing). This feature is supported by AI/ML, which will be further detailed in the following sections.

In summary, through the features that the CS-AWARE platform provides, the data listed in Table 1 about organisations/ecosystems is utilised by AI/ML in order to provide contextualised cybersecurity management support, to improve efficiency and effectiveness within an organisation or an ecosystem.

4 Human-in-the-loop anomaly detection

This section illustrates the complexity of decision-making in cybersecurity management and highlights how humans can complement AI/ML-based cyberthreat detection. It presents a framework for

Organisation
(a) A system dependency graph: How assets and dependencies of an organisation interrelate, including any contextual knowledge management information the organisation wants to provide per asset or dependency.
(b) Contextual information about business processes and information flows, and how they map to the systems in day-to-day operation, including relevant behaviour patterns that depict or influence the cybersecurity state of the organisation.
(c) Log files (e.g., network, service/application logs, database logs, security appliance logs) and their role in monitoring behaviour patterns identified in (b).
(d) Organisational policies (e.g., security policies, BC/DR policies).
Ecosystem
(a) Ecosystem graph (organisations, services they provide, and how services between organisations depend on each other).
(b) Ecosystem policies (e.g., security policies, BC/DR policies that encompass multiple organisations/services).
(c) Discussions about cybersecurity problems and support.
Public data
(a) Threat intelligence (e.g., MISP, ...).
(b) Cybersecurity relevant social media.
(c) Guidelines, best practices and other cybersecurity relevant data.

Table 1: Data sources available through the CS-AWARE-NEXT cybersecurity management approach.

providing cybersecurity situational awareness to the security analyst and then provides a detailed description of how anomalies/attack detection is performed in CS-AWARE-NEXT.

4.1 Cybersecurity situational awareness based on the Cynefin framework

The wide adoption of cloud services and web apps has dramatically increased the attack surface for adversaries. Hence, any administrator responsible for an information system connected to the Internet should also expect to deal with a substantial number of incidents. At the same time, the available time for reacting is anticipated to be relatively short. The complexity of current cyber-attacks is quite high, thus having the best possible understanding of the situation in a very short time is of utmost importance in order to take the appropriate decisions and actions to respond to it. The Cynefin framework (Snowden and Boone, 2007) can be used for guiding decision-making and problem-solving during cybersecurity incidents. Cynefin has five so-called dimensions or contextual definitions that can be applied to a cybersecurity context, so as to provide better understanding about the encountered threats and attacks, as explained below (Papanikolaou et al., 2023).

- **Simple (known knowns)**. In the simple dimension, problems are well-defined, and there is a clear cause-and-effect relationship between the problem and the solution. In the context of cyber attacks, the simple domain could be applied to routine security tasks such as patch management, security configuration, and access control. Indicators of Compromise (IoCs) are unambiguous and can attribute the threat.

- **Complicated (known unknowns)**. In the complicated dimension, problems reach a state where there may be multiple potential solutions that require expert knowledge and analysis. In the context of cyber attacks, the complicated domain could be applied to tasks such as incident response, malware analysis, and vulnerability assessments. IoCs are somewhat unambiguous and can attribute the threat with a bit of effort.

- **Complex (unknown unknowns)**. In the complex dimension, problems are unpredictable and emergent, and there may be no clear cause-and-effect relationship between

the problem and the solution. In the context of cyber attacks, the complex domain could be applied to threat hunting, threat intelligence, and adaptive security measures. IoCs are ambiguous and not as trustworthy.

- **Chaotic (unknowables)**. In the chaotic dimension, problems are unpredictable and rapidly changing, and immediate action is required to stabilise the situation. In the context of cyber attacks and the kill chain, the chaotic domain could be applied to the initial response to a major cyber incident, where there is a need for rapid triage, containment, and recovery. IoCs cannot be defined; if they do so, they have almost no value as they will be too generic or untrustworthy.

- **Confusion (or Disorder)**. This domain represents situations without clarity about which of the other domains apply.

In cases of relatively low uncertainty, the whole process can significantly benefit from AI support, which can prove to be fully autonomous in identifying and mitigating the threat, or it can simply help the human operator get a better understanding of the situation. In high-uncertainty situations, a higher human intervention is anticipated, and it may not even be possible to determine which stage the attack will be at. As the attack progresses through its stages, it suggests that the security controls were not effective and, therefore, a higher degree of human intervention and participation is required. Therefore, the Human in the Loop (HITL) aspects should be considered when deploying any system with a substantial machine learning or AI component.

4.2 Anomalies/attacks detection in CS-AWARE-NEXT

Currently, the CS-AWARE-NEXT platform uses log files from different organisations to detect anomalies/attacks. In particular, the data collection flow is performed via the installation of log agents on various servers of the pilots, which capture a wide range of logs, including Windows Event Logs from channels such as security, application, and system, as well as syslog entries from Linux and other systems. For the detection of traditional attacks, such as denial of service and brute-force attacks, we trained ML-based methods (e.g., k-Nearest Neighbours and Random Forest). In this respect, the training phase was performed using public datasets (e.g., CSE-CIC-IDS2018[1]) containing both normal and malicious traffic. This was possible since the behaviour of these threats is well known. We are able to detect anomalies with good results: F1 score ranges between 0.88 and 0.97). We are also designing algorithms to detect anomalies in log entry counts by analysing trends, seasonality, and unexpected spikes.

However, it is necessary to highlight that this task is still complex due to the dynamic nature of cyber threats and the complexity of modern IT environments. For example, the most common problems are high false positive rates and data quality issues. As regards the former, anomaly detection systems may flag legitimate activities as anomalies, leading to alert management overload and decreased trust in the system. The latter refers to the fact that incomplete, inaccurate, or insufficient data can hinder the effectiveness of anomaly detection algorithms. In our experience, recent experiments show that data are affected by several issues, such as wrong formats, incompleteness, redundancies, and design problems (e.g., no primary keys and no correlated tables). For this reason, we are not just focusing on the design of analytics tools but also on the design of a robust data preparation pipeline in order to guarantee high-quality input data. Data cleaning techniques are used to handle common data quality errors. For example, missing values are addressed by applying standard data imputation, useless columns (i.e., those with constant values or redundant columns) are deleted, and duplicated rows are deleted. In this way, we aim to guarantee the reliability of the analysis output and, thus, a higher informative value.

Moreover, collecting and processing large volumes of data in real-time can be very resource-intensive, and the system must be able to handle a high volume of continuous incoming data. In order to better organise the analysis, we use a Lambda architecture (Kiran et al., 2015) in which we have a (i) *Speed layer* that performs real-time anomaly detection on the incoming data, classifying them as normal or anomalous behaviour and a (ii) *Batch layer* that processes the incoming data and stores them in a repository (e.g., data lake) for the analysis of historical data.

The main problem is that cyber threats are con-

[1] https://www.unb.ca/cic/datasets/ids-2018.html

stantly evolving to evade detection. Anomaly detection systems must adapt to these changes and continuously update their models to detect new types of anomalies. Moreover, as described in the previous sections, organisations have their own process models and policies. Therefore, additional context-aware anomalies need to be considered.

For this reason we propose the human-in-the-loop anomaly detection support. In fact, one of the core aspects provided by the CS-AWARE methodology is the ability to achieve and model a holistic understanding of how business processes of an organisation work, how they map to infrastructure, and how their behaviour can be monitored in day-to-day operation. This allows the monitoring of behaviour patterns of particular interest to the organisation and is based on realistic baselines provided by the employees/users of the organisation.

The CS-AWARE platform is able to gather the following data in machine-readable form:

- The asset(s) a monitoring pattern is related to.
- The log files and individual log file parameters that allow monitoring for specific behaviour.
- The baseline/range that defines normal behaviour.

The users defining this information have full control over the process and can change/adapt the monitoring patterns to implement the user-in-the-loop anomaly detection (steps 3 and 4 in Figure 1).

By exploiting this information, we aim to enrich our existing anomaly detection tools to enhance their accuracy and effectiveness. This approach, in fact, leverages the strengths of machines and humans, addressing the limitations of purely automated or manual approaches. We aim to incorporate rules-based methods and process mining techniques. The former are based on rules defined by the analysts on the basis of organisational and ecosystem knowledge. Rules-based systems can tailor the anomaly detection procedure to the organisation's requirements. Process mining techniques can analyse event logs to identify patterns and deviations from expected workflows, aiding anomaly detection. In this way, they can identify hidden patterns and anomalies in organisational processes that may not be apparent through manual inspection.

Note that as shown in Figure 1, we are going to customise applications also on the basis of available public data. For example, we gather from social media and threat intelligence feeds the list of the most spread malware and threats and classify them on the basis of the geographical areas and components/applications they affect. In this way, we can improve the effectiveness of the anomaly detection applications. In fact, on the one hand, organisations will receive only alerts about the relevant cybersecurity threats that might affect them. On the other hand, the applications will be adapted on the basis of such information.

5 An example use-case

The relevance and the potential impact of our approach are demonstrated with a (simulated, albeit) realistic scenario within the context of healthcare ecosystems. Several cyber-incidents have been reported lately in the health sector (McGlave et al., 2024). The healthcare ecosystem comprises a plethora of components, such as all of its departments and clinics that provide services to patients, and all operational flows require dedicated access policies. Patient medical data (personally identifiable information) is stored and updated according to local, regional, national, and European regulations. We focus our analysis on the case of the hospitals' radiology department operational flow.

Figure 2 delineates a typical deployment business process of the radiology department. It demonstrates the socio-technical nature of such business processes involving humans and machines. Not only are the Radiology Assessment and the associated DICOM Image from the outpatient's MRI-scanner (magnetic resonance imaging) stored in local RIS (Radiology Information System) and PACS (Picture Archiving and Communication System) servers but also in a remote backup infrastructure for business continuity purposes. As a result, patient care improves by allowing specialists to access the information they need when they need it. Therefore, upon the patient's approval, an external physician could grant access to these data via a temporary web service. Additionally, a patient's Personal Electronic Health Record is updated and accessed nationally or cross-border-wise within the EU health Dataspace. European Medical Devices Regulation 2017/745 (MDR) dictates general safety and performance requirements (GSPRs) to which the MRI Scanner manufacturer shall demonstrate compliance while in parallel the European Union Agency for Network and Information Security (ENISA, 2016; Eichelberg et al., 2021) recom-

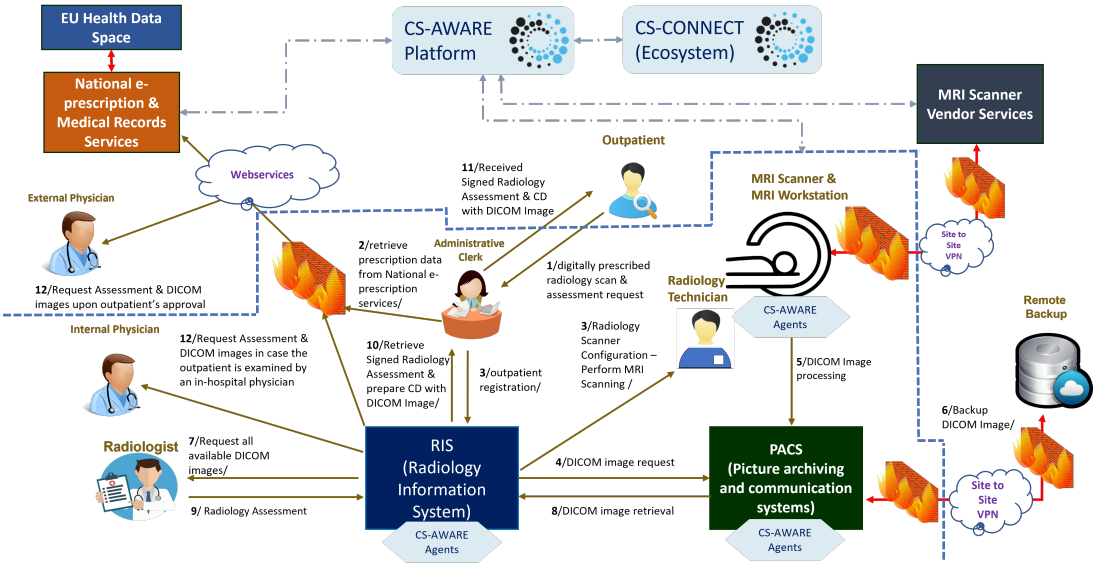

Figure 2: Monitoring of Radiology Business Process.

mends that regular updates/patches to Networked Medical Devices, including virus scanners, well-defined access rights and the usage of encryption mechanisms where is necessary (e.g., external access to DICOM images) are implemented. To this end, service engineers on the vendor's sides are regularly called for and their work may lead to leaked passwords or make the system more vulnerable to hackers. The critical assets in the business process, depicted in Figure 2, constitute the MRI Scanner along with its associated Workstation and the RIS/PACS system. During the last decades, EU hospitals have invested through the National Health Ministry's procurement policies in MRI scanners to enhance the health services provided to the citizens.

5.1 An incident

There suddenly appeared to be an issue with the MRI scanners of a specific vendor for EU hospitals. First, the healthcare professionals in one hospital started to notice a decline in the scanner's performance, and they informed the hospital's Biomedical and IT departments and contacted the MRI scanner supplier. The supplier claimed the MRI scanner was working perfectly despite being a 17-year-old version. Notwithstanding the supplier's assurances, the following day, the MRI scanner began to continuously transmit MRI images in a loop, causing the overload of the PACS server in a matter of minutes. The hospital shut down the MRI scanner to interrupt the image transmissions, but it had already lost access to the millions of CT scans, MRIs, and X-rays stored in the PACS, significantly affecting patient care delivery. This event was reproduced in several hospitals over the next few days. Two days later, the hospital's IT staff managed to bring back the PACS server online but continued to work with the MRI scanner vendor technicians to reconnect the MRI scanner. They identified that a change in the configuration code caused a malfunction: it had set a ceiling of 50 million scans for the machine. When the equipment reached this figure, it overloaded and entered a continuous loop transmitting MRI imaging. This change was caused by a sophisticated cyber-attack that exploited known vulnerabilities of the MRI scanner/workstation system's obsolete operating system. By providing the necessary information to the supplier, they were able to develop a software patch to amend the problem. Two weeks later, after the software patch was applied, the MRI scanner became operational again.

5.2 Impact

The aforementioned scenario directly impacted the hospital's care delivery, as well as the reputation of the medical equipment manufacturer, whose equipment severely affected the healthcare organisation's IT infrastructure and the Biomedical Department's procedures. Further, the hospital's operations were impacted due to the loss of availability of the MRI scan service as well as access to the PACS server because of the MRI scanner system's failure. Recovering from the attack took approximately 2 working days in order to put the PACS server back online

and restore the healthcare professionals' access to millions of stored CT scans, MRIs, and X-rays, and another 2 weeks to patch the vulnerability found in the MRI scanner and reconnecting it to the hospital's network.

5.3 Improved efficiency through AI/ML

In such an incident scenario, CS-AWARE NEXT can provide several AI/ML-supported mechanisms for improved efficiency and effectiveness of the hospitals' cybersecurity management:

- The anomaly could have been detected earlier because of the monitoring of user-defined baselines by people who know the processes very well (through human-in-the-loop AI).
- The root cause can be discussed and finally discovered/mitigated through discussions on the relevant ecosystem. The generated knowledge could then be shared and highlighted to everyone with the same asset (through AI/ML contextualisation).
- If the manufacturer or the security community has already reported a vulnerability or threat report to the threat intelligence or vulnerability database about this specific bug, CS-AWARE can highlight this to everyone having this asset (using AI/ML contextualisation).
- There could have been early warning discussions about this behaviour and eventually also pointers to mitigations or solutions on social media in relevant channels, and CS-AWARE-NEXT can alert everyone with the same asset about something that is going on (through AI/ML contextualisation).
- One affected organisation could have solved the issue and shared this information with authorities or the public through CS-AWARE-NEXT information sharing capability. Other organisations with the same asset can be alerted (through AI/ML contextualisation).

6 Conclusion and outlook

The paper describes the approach designed in CS-AWARE/CS-AWARE-NEXT. We aim to extend anomaly detection and improve cybersecurity awareness by adopting a HITL approach and considering contextual information. This aims to design applications able to identify well-known and unknown anomalies on the basis of organisational rules and knowledge. In this way, the organisation would benefit from earlier detection, earlier deployment of the patch, improved procurement policies, more effective risk management procedures, and so on. Future work will focus on the validation of the approach with the organisations involved in the project.

Limitations

HITL and context-aware anomaly detection offers significant advantages over traditional anomaly detection methods but they also come with certain limitations. First of all, we have to consider the complexity of Context Modeling. Managing and processing diverse types of contextual data (temporal, spatial, categorical, etc.) can be challenging. Another challenge is Data Quality: in real-world scenarios, data may be noisy or missing. Such issues can lead to erroneous anomaly detection results. Data preparation pipelines need to be well-defined in order to guarantee data reliability. The computation complexity is also not negligible: dealing with high-dimensional datasets (e.g., logs, social data) can result in a high computational overhead and poor scalability, making real-time anomaly detection complex.

Ethics statement

This material is the authors' own original work, which has not been previously published elsewhere.

Acknowledgements

Funded by the European Union (Grant Agreement No 101069543). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

References

Jerry Andriessen, Thomas Schaberreiter, Alexandros Papanikolaou, and Juha Röning, editors. 2022. *Cybersecurity Awareness*, volume 88 of *Advances in Information Security*. Springer, Cham.

P. B. Checkland. 1998. *Systems Thinking, Systems Practice*. John Wiley & Sons Ltd. 1981.

P. B. Checkland and J. Scholes. 1991. *Systems Thinking, Systems Practice*. John Wiley & Sons Ltd.

Selina Y. Cho, Jassim Happa, and Sadie Creese. 2020. Capturing tacit knowledge in security operation centers. *IEEE Access*, 8:42021–42041.

Marco Eichelberg, Klaus Kleber, and Marc Kämmerer. 2021. Cybersecurity protection for pacs and medical imaging: Deployment considerations and practical problems. *Acad Radiol.*, 28(12):1761–1774.

ENISA. 2016. Smart hospitals: Security and resilience for smart health service and infrastructures.

European Commission. 2016. Regulation of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Regulation (EU) 2016/679.

European Commission. 2020. The EU's Cybersecurity Strategy for the Digital Decade. Joint Communication to the European Parliament and the Council - JOIN(2020) 18 final.

Gustavo Granadillo, Susana González-Zarzosa, and Rodrigo Diaz. 2021. Security Information and Event Management (SIEM): Analysis, trends, and usage in critical infrastructures. *Sensors*, 21:4759.

ISO/IEC 27000:2016. 2016. Information technology – security techniques — information security management systems – overview and vocabulary. Technical report, ISO/IEC.

Sajjad M. Jasimuddin and Fateh Saci. 2022. Creating a culture to avoid knowledge hiding within an organization: The role of management support. *Frontiers in Psychology*, 13.

Mariam Kiran, Peter Murphy, Inder Monga, Jon Dugan, and Sartaj Singh Baveja. 2015. Lambda architecture for cost-effective batch and speed big data processing. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2785–2792.

Veronika Kupfersberger, Thomas Schaberreiter, Christopher Wills, Gerald Quirchmayr, and Juha Röning. 2018. Applying soft systems methodology to complex problem situations in critical infrastructures: The cs-aware case study. *International Journal on Advances in Security*, 11:191–200.

Max Landauer, Florian Skopik, Markus Wurzenberger, and Andreas Rauber. 2020. System log clustering approaches for cyber security applications: A survey. *Computers & Security*, 92:101739.

Joffrey L. Leevy, John T. Hancock, Richard Zuech, and Taghi M. Khoshgoftaar. 2021. Detecting cybersecurity attacks across different network features and learners. *J. Big Data*, 8(1):1–29.

Christian Luidold, Thomas Schaberreiter, Christian Wieser, Adamantios Koumpis, Cinzia Cappiello, Tiziano Citro, Jerry Andriessen, and Juha Röning. 2023. Increasing cybersecurity awareness and collaboration in organisations and local / regional networks: The CS-AWARE-NEXT project. In *Sustainable, Secure, and Smart Collaboration (S3C) Workshop 2023*.

Claire McGlave, Hannah Neprash, and Sayeh Nikpay. 2024. Hacked to pieces? the effects of ransomware attacks on hospitals and patients.

Henock Mulugeta Melaku. 2023. A dynamic and adaptive cybersecurity governance framework. *Journal of Cybersecurity and Privacy*, 3(3):327–350.

National Institute of Standards and Technology. 2024. The NIST Cybersecurity Framework (CSF) 2.0. Cybersecurity White Paper (CSWP) 29.

Seiichi Ozawa, Tao Ban, Naoki Hashimoto, Junji Nakazato, and Jumpei Shimamura. 2020. A study of iot malware activities using association rule learning for darknet sensor data. *Int. J. Inf. Sec.*, 19(1):83–92.

Alexandros Papanikolaou, Christos Ilioudis, and Vasilis Katos. 2023. Cyber-pi: Intelligent cyberthreat detection and supervised response. In *Proceedings of RCIS 2023*, Corfu, Greece.

Iqbal H. Sarker. 2021. Cyberlearning: Effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. *Internet Things*, 14:100393.

Thomas Schaberreiter, Christian Wieser, Adamantios Koumpis, Christian Luidold, Jerry Andriessen, Cinzia Cappiello, and Juha Röning. 2023. Addressing critical issues and challenges for dynamic cybersecurity management in organisations and local/regional networks: The CS-AWARE-NEXT project. In *2023 TransAI*, pages 232–236.

Australian Securities and Investments Comission. 2023. Spotlight on cyber: Findings and insights from the cyber pulse survey 2023. Report 776.

Sanjana Shukla, José Ignacio Parada, and Keri Pearlson. 2022. Trusting the needle in the haystack: Cybersecurity management of ai/ml systems. In *Advances in Information and Communication*, pages 441–455. Springer.

David J. Snowden and Mary E. Boone. 2007. A leader's framework for decision making. *Harvard business review*, 85(11):68.

The European Parliament and the Council of the European Union. 2022. Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022. Official Journal of the European Union L333/80.

Quang Hieu Vu, Dymitr Ruta, and Ling Cen. 2019. Gradient boosting decision trees for cyber security threats detection based on network events logs. In *2019 IEEE International Conference on Big Data (IEEE BigData), Los Angeles, CA, USA, December 9-12, 2019*, pages 5921–5928. IEEE.