# D3.2 SPHINX Cyber Situational Awareness Framework fitness/suitability- Real Time Risk Assessment Models v1

## WP3 – Cyber security risk assessment & Beyond – Sphinx Intelligence

**Version: 1.00**

SPHINX

A Universal Cyber Security Toolkit for Health-Care Industry

## Disclaimer



## Copyright message



## Document information

| Grant Agreement Number | 826183 | Acronym | SPHINX |
|---|---|---|---|
| **Full Title** | A Universal Cyber Security Toolkit for Health-Care Industry | | |
| **Topic** | SU-TDS-02-2018 Toolkit for assessing and reducing cyber risks in hospitals and care centres to protect privacy/data/infrastructures | | |
| **Funding scheme** | RIA - Research and Innovation action | | |
| **Start Date** | 1st January 2019 | **Duration** | 36 months |
| **Project URL** | http://sphinx-project.eu/ | | |
| **EU Project Officer** | Reza RAZAVI (CNECT/H/03) | | |
| **Project Coordinator** | Dimitris Askounis, National Technical University of Athens - NTUA | | |
| **Deliverable** | D3.2. SPHINX Cyber Situational Awareness Framework fitness/suitability- Real Time Risk Assessment Models v1 | | |
| **Work Package** | WP3 – Cyber security risk assessment & Beyond – Sphinx Intelligence | | |
| **Date of Delivery** | **Contractual** M18 | **Actual** | 18 |
| **Nature** | R - Report | **Dissemination Level** | P - Public |
| **Lead Beneficiary** | HMU | | |
| **Responsible Author** | Yannis Nikoloudakis | **Email** | nikoloudakis@pasiphae.eu |
| | | **Phone** | |
| **Reviewer(s):** | Dana Oniga (SIMAVI), Patrik Karlson (AiDEAS), Serafeim Moustakidis (AiDEAS) | | |
| **Keywords** | Situational Awareness | | |

### Document History

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

| Version | Issue Date | Stage | Changes | Contributor |
|---|---|---|---|---|
| 0.10 | 10/12/2019 | Draft | ToC | Yannis Nikoloudakis (HMU) |
| 0.20 | 28/5/2020 | Draft | Content | Yannis Nikoloudakis (HMU) |
| 0.25 | 12/6/2020 | Draft | Content | Yannis Nikoloudakis (HMU) |
| 0.30 | 12/6.2020 | Draft | Incorporate AiDEAS's contribution | Serafeim Moustakidis (HMU), Yannis Nikoloudakis (HMU) |
| 0.40 | 15/6/2020 | Draft | Incorporate NTUA's contribution | George Doukas (NTUA), Yannis Nikoloudakis (HMU) |
| 050 | 17/6/2020 | Draft | Corrections in content | Yannis Nikoloudakis (HMU) |
| 0.60 | 23/6/2020 | Draft | Integration of PDMFC contribution | Stelios Karagiannis (PDMFC), Louis Landeiro Ribero (PDMFC), Yannis Nikoloudakis (HMU) |
| 0.70 | 23/6/2020 | Draft | Integration of TECNALIA contribution | Erkuden Rios (TECNALIA), Yannis Nikoloudakis (HMU) |
| 0.80 | 25/6/2020 | Draft | Internal Review 1 (SIMAVI) | Dana Oniga (SIMAVI) |
| 0.85 | 25/6/2020 | Draft | Internal Review 2 (AiDEAS) | Patrik Karlsson (AiDEAS), Serafeim Moustakidis (AiDEAS) |
| 0.90 | 25/06/2020 | Draft | Addressing Reviewer's comments | Yannis Nikoloudakis (HMU) |
| 0.95 | 29/6/2020 | Pre-final | Rady for quality review | George Doukas (NTUA), Michael Kontoulis (NTUA) |
| 1.00 | 29/6/2020 | Final | Final | Christos Ntanos (NTUA) |

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.*

# Executive Summary

The purpose of this deliverable is to perform and present a desk research on the term “Situational Awareness” (SA) and Real-Time Risk Assessment Models. Towards that, an in-depth literature review was performed that revealed several research initiatives and numerous approaches, addressing different issues and barriers. This document aims to present the findings of this research and shed some light by presenting the state of the art on SA, discover existing standards, frameworks, tools, and methodologies that address the situational awareness in the cybersecurity domain. Moreover, it aims to act as a guiding tool for Task T3.3 and the *Distributed Cyber Situational Awareness Framework & Real Time Risk Assessment* Module.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# Contents



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# Table of Figures



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# Table of Tables



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# Table of Abbreviations

SA : Situational Awareness

ID : Intrusion Detection

CVE : Common Vulnerability Exposures

SCAP : Security Content Automation Protocol

ISCM : Information Security Continuous Monitoring

ISMS : Information Security Management Systems

STIX : Structured Threat Information Expression

CIP : Critical Infrastructure Protection

IoC : Indicators of Compromise

UCO : Unified Cybersecurity Ontology

RDF : Resource Description Framework

OWL : Web Ontology Language

CNN : Convolutional Neural Network

ISO : International Organization for Standardization

OSPF : Open Shortest Path First

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# 1 Introduction

Situational awareness (SA) has been a buzz word for several years, within the scientific community. Although the term itself is somewhat new, the history of SA goes back to the military theory[1], as its first appearance was in Sun Tzun's "Art of War". It was thereon used in the aviation domain, and in the recent years, is has been used in the cybersecurity domain. SA, is based on a three-layered model (Endsley's model) [1]. Namely, the "Perception", the "Comprehension", and the "Projection" (Figure 1).

- **Perception**

The perception layer pertains the monitoring, acquisition, and initial processing of the status, attributes, and dynamics of the elements in the surrounding environment. In a network environment, the elements that can be monitored are the network flows, the users' behaviour, etc.

- **Comprehension**

The comprehension layer involves the cross correlation of seemingly unrelated events, measurements, etc., to create patterns. In this layer, feature extraction and pattern recognition are envisioned, in order to interpret the findings and evaluate the results.

- **Projection**

The projection layer involves the projection/prediction of the consequent developments on the environment's elements' actions. In a network environment, a possible result would be the prediction of a cyber-attack such as Denial of Service (DoS), based on the network data that have been collected and processed (e.g. increased rate of incoming packet size).

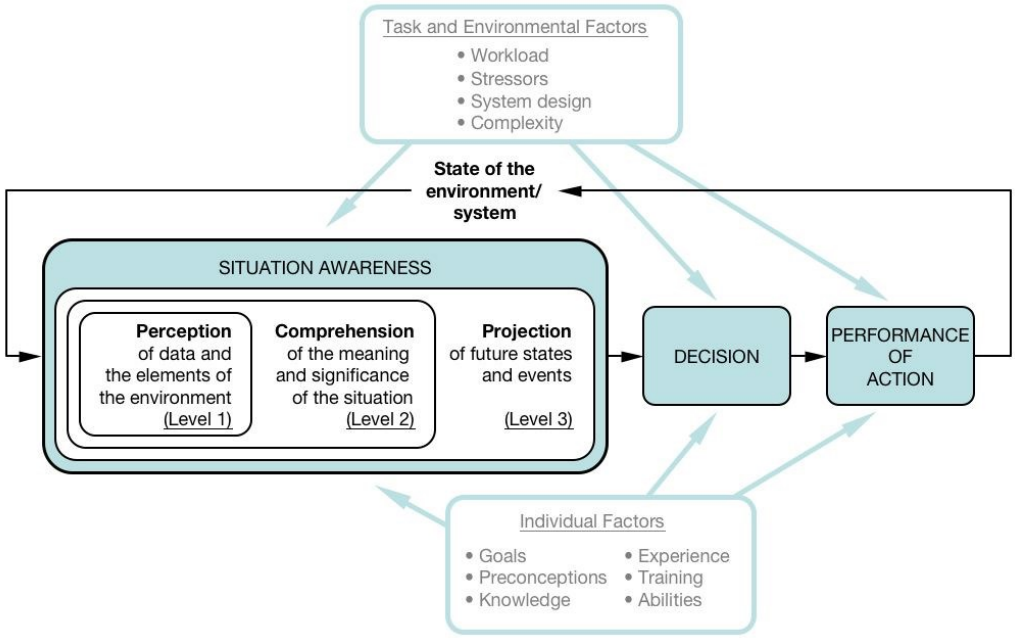


***Figure 1 Endsley's situational awareness model***

[1] https://en.wikipedia.org/wiki/Situation_awareness

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

## 1.1 Purpose & Scope

The purpose of this document is to present the current state of the art, concerning SA, and unveil its intricacies. Moreover, existing standards, platforms, tools, frameworks, and methodologies, as well as existing risk models have been researched. All the above, will help us discover potential issues and barriers towards efficient SA within an ICT environment, and provide useful input for task T3.3.

## 1.2 Structure of the deliverable

The rest of this document is structured as follows. Section 2, presents the literature review results, as well as the researched risk models. Additionally, an analysis on the research results is performed, and a research of existing risk models is presented. Finally, the potential risks and barriers are analysed and the existing standards, platforms and methodologies are detailed.

## 1.3 Relation to other WPs & Tasks

This task (T3.1 Research Analysis on Situational Awareness Approaches for Advanced Threats Management) and as a result of it, this document is closely related to task ***T3.3 SPHINX Distributed Cyber Situational Awareness Framework & Real Time Risk Assessment Module.*** This document will act as a direct input for Task T3.3 and the corresponding deliverable ***D3.1 Distributed Situational Awareness Framework v1*** of WP3**.**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# 2 Cyber Situational Awareness

## 2.1 Literature review

### 2.1.1 Situational Awareness

#### *2.1.1.1 Anomaly Detection in Cyber Security Situational Awareness*

Alsmadi et al [2], presented a framework that dynamically extracts models and uses contextual information to detect both known and zero-day attacks. To potentially detect zero-day attacks, their framework combines semi-supervised anomaly detection with attack-profile similarity. Additionally, the framework uses data transformations with linear discriminant analysis, thus leading to a decrease in time of possible intrusions at system runtime. Lastly, to detect known attacks the framework is able to describe a specific environment in order to select and use numerous types of context profiling and semantic networks of attacks.

The simultaneous use of Traffic Circle (visualization tool that complements CLIQUE) and CLIQUE (behavioural summarization tool)[2] in a near-real-time environment, provided by MeDICi was presented by D. Best et al [3], to allow visualization of network traffic as it occurs. Numerous potential issues can be investigated and therefore, prevented as soon as behaviour deviates from normal. Traffic Circle allows us to detect potential threats, contained within raw flow records with different attribute spaces and colour encoded filters, while CLIQUE (based on LiveRac [4]) provides aggregated flows to a higher-level abstraction, to help analysts cope with data scale. To combine the two afore-mentioned applications/tools while reducing the complexity and ease the development of high-performance analytic applications over numerous domains, the Middleware for Data Intensive Computing (MeDICi) was developed. The MeDICi Integration Framework (MIF), is used for the production of the analytic pipeline for the network visualization

C Zhong et al [5], performed a literature review, regarding theory and models in Situational Awareness, in the Cyber Security domain. While D'Amico et al [6] described six broad analysis roles, namely triage analysis, escalation analysis, correlation analysis, threat analysis, incident response and forensic analysis, the authors went in depth, focusing mostly on Data Analysis and Data Triage. C Zhong et al did not propose/develop a specific framework/tool for situational awareness in cybersecurity, but they identified the human part in Security Operations Centres (SOCs), and proposed virtualization tools for anomaly-based intrusion detection analysis [7], wherein they assist the analysts regarding monitoring, analysis, and response.

In order to minimize the data storage issue regarding situational awareness data, W. Yu et al [8] proposed a cloud computing based architecture. In addition, they implemented a cloud-based threat detection system that identifies attacks based on their signature with anomaly detection techniques.

#### *2.1.1.2 Data Fusion for Cyber Security Situational Awareness*

L. F. Sikos et al [9], proposed a novel framework that collects and fuses heterogeneous network data, using the Resource Description Framework (RDF)[3], wherein they augment the fused data with provenance data to provide rich semantics with highly specialized ontology terms, therefore leading to highly contextual, uniform data. Having uniform data, allows for the development of an automated network data framework, which is tasked with the analysis of the data. The developed framework enhances the RDF descriptors with annotations from controlled vocabularies and ontologies [10]. Description Logics (DL) reasoners such as HermitT[4] and

[2] http://vacommunity.org/Traffic+Circle+and+CLIQUE
[3] https://www.w3.org/RDF/
[4] http://www.hermit-reasoner.com

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

FaCT++[5] are also used, in order to have a proper trade-off between expressivity and reasoning complexity, while ensuring decidability. The DL axioms are implemented in RDF from the Web Ontology Language (OWL) [6] ontologies. In terms of Cyber-Situational Awareness, the framework implements tagged graphs with terms from the Communication Network Topology and Forwarding Ontology (CNTFO)[7], which is specifically designed for this.

Another approach concerning fusion of heterogeneous network data and the understanding of network topologies, is delivered from S. Voigt et al [11]. None of the current literature has developed ontologies for the Internet Protocol (IP)[8], the Open Shortest Path First (OSPF)[9] and the Border Gateway Protocol (BGP)[10]. Therefore, they developed three ontologies that can be used to represent complex communication concepts, namely the Internet Protocol Ontology, The OSPF Ontology and the BGP Ontology. These ontologies provide the means to combine heterogeneous data from different sources (network diagrams, router configuration files, and routing protocol messages) and to be clearly represented. Their proposal also uses the OWL.

A semantic approach that combines traditional Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS), but is also equipped with new sensors, to derive new attack signatures based on zero-day attacks, is proposed by M Matthews et al in [12]. Their framework apart from scanners, antivirus, etc., also includes sensors that scan online forums, blogs, and vulnerability databases for textual descriptions of attacks. The framework is a combination of ontology, a knowledge base, and reasoners. Their ontology is an extension of their previous work [13][14]. The network data is encoded as OWL and RDF, wherein the data and events from the data streams are represented in the ontology. Afterwards, the knowledge base verifies whether the alert from the IDS is a false positive or not, and based on the report, they identify attacks using a network traffic flow classifier.

Y. Gao et al [15], proposed a network security situational awareness model that fuses information from multi-sources.. The multi-source information is extracted using a rules library, which normalizes the raw collected data, and a knowledge base, by pre-processing multi-source information. This model is based on analysing the theoretical model of network security situation perception and research initiatives.

Moreover, a visual analytics solution that binds data together was proposed by M. Angelini et al. [16]. The proposed idea is to separate the events using security profiles (network security officer, network security manager, etc.), thus clarifying the network state and the impact an attack or a specific risk will have on the system and the business. Their proposal was focused more on risk analysis in the actual implementation, but they aim to extend it by using the relationship of attacks and vulnerabilities and creating extra layers of analysis.

L. Zareen Syed et al [17], proposed the Unified Cybersecurity Ontology (UCO). UCO fuses heterogeneous data and knowledge schemas, from various cybersecurity systems and standards in order to share and collect related information. UCO allows data sharing across different formats and standards. The vital classes of the ontology are:

- Means - contains information regarding various way to execute attacks
- Consequences - describes the possible outcomes of attacks
- Attack - characterizes a cyber-attack

[5] http://owl.man.ac.uk/factplusplus/
[6] https://www.w3.org/OWL/
[7] https://lesliesikos.com/ontology/network.ttl
[8] https://en.wikipedia.org/wiki/Internet_Protocol
[9] https://en.wikipedia.org/wiki/Open_Shortest_Path_First
[10] https://en.wikipedia.org/wiki/Border_Gateway_Protocol

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

- Attacker - identification of the attacker
- Attack Pattern - information regarding the methods used and ways to mitigate the attack
- Exploit - information about a specific exploit
- Exploit Target - contains exploit targets that are vulnerable or have weaknesses in software, systems, networks or even configurations that can be targeted
- Indicator - pattern identifying conditions

Their approach uses semantic web languages, which are preferable for security situations (RDF, OWL). They both have a decentralized philosophy, and OWL provides rich semantic constructs for schema mapping and combines it with robust reasoners. UCO offers more coverage in contradiction to other isolated cybersecurity ontologies since it has been mapped to publicly available ontologies.

#### *2.1.1.3 Frameworks/tools that assist Cyber Security Analysts for Situational Awareness*

K. Huffer et al [18], presented Situational Awareness of Network System Roles (SANSR) tool. SANSR's role in the cybersecurity domain is to feed security analysts and network administrators, with information regarding the role and operations of every network-enabled entity near the handler. Leading to an information system that will help security analysts and network administrators to prioritize intrusion alerts, and easily detect possible changes in the underlying network. The tool uses a collection of network flow data, that discovers the roles of each entity by using both clustering and categorization techniques.

R. Graf et al [19] presented an experimental setup, combined with a management method based on Artificial Intelligence (AI) that can support cyber analysts in establishing cyber situational awareness, in order to quickly deploy countermeasures in case of an attack. The aim of their proposal is the replacement of human input, for cyber incident analysis tasks (triage). With that aim in mind, the AI eliminates the need for the security analyst to classify cyber incident reports, find related reports, eliminate irrelevant information, and `produce reports regarding the life cycle management in an automated manner. This approach increases accuracy and performance, while also reduces the number of manual operations. For the adoption of this experimental setup, they used a blockchain-based technique along with neural networks. The blockchain's role in the setup is to provide an automated trusted system for incident management workflow, which allows automatic classification, acquisition, and enrichment of incident data.

In order to tackle multistage attacks in real-time, S. Mathew et al [20] analysed the content of event streams produced by network sensors (IDSs), using a comprehensive situational awareness tool ECCARS(Event Correlation for Cyber Attack Recognition System). The ECCARS tool categorizes attack patterns, which represent the semantic stages of typical zero-knowledge and multistage attack scenarios. The semantic categories that also contain a criticality value, are related to the alerts in the signature sets from the sensors (IDSs).

G. Settanni et al [21], presented and evaluated three different Vector Space Models (VSM)-based information correlation methods, the Artifact-based, the Word-based, and the Dictionary-based Linking , to compare security information. The main aspect of this paper is the correlation of natural language documents, to identify similarities in order to detect and handle cybersecurity-related incidents. Depending on the computational power required, the methods are described as follows: the Artifact-based Linking method balances between accuracy and time consumption, wherein the Word-based Linking method benefits accuracy over time requirements, and the Dictionary-based Linking method is faster but less precise.

A prototype fuzzy-logic-based application was proposed by E. Allison et al [22] that uses the joint knowledge of the Computer Network Defence (CND) and Information Assurance (IA) [23][24], to produce an Alert Priority Rating (ARP), with the use of computational intelligence. The Fuzzy Logic Utility Framework (FLUF) mentioned

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

in [23], also takes under consideration the damage a compromised asset would impose to the system in terms of confidentiality, integrity, and availability. Through the tested dataset, they noticed an increase in accuracy, regarding prioritization, compared to Snort prioritization, presenting the severe alerts in a more "suitable" order.

To aid security analysts, W. Matuszak et al [25] developed Cyber Situational Awareness for Visualization (CyberSAVe). CyberSAVe's role is to establish and maintain trust between the system and the sensors of the topology. Providing the necessary tools that can be deployed to allow the investigation of cybersecurity-related incidents, administrators can determine if sensors are working as intended and they have not been compromised.

V. Lenders et al [26], proposed a cyber-situational awareness framework based on the "observe, orient, decide, act" (OODA) decision support model that can provide cognitive mapping, combining raw data from sensors and detailed analysis of threats and vulnerabilities. In more detail, the framework collects information with sniffers, extracts them from system log files, net tools, and databases. To create a dynamic framework, the authors rely on Semantic Web technologies to support reasoning with an integrated decision support system. Their framework contains all the phases of the OODA decision support model.

The advantages that deep learning architectures can offer to classify and correlate malicious activities that are detected, led R. Vinayakumar et al. [27] to present ScaleNet, a framework that analyses and correlates events from DNS, Email, and URLs, therefore eliminating the need for an ontology to describe the large volume of raw data. Their framework is also easily extensible to handle data from other resources.

A practical way of detecting Indicators of Compromises (IoC) is with the use of regular expressions. Despite the usefulness of regular expressions, most algorithms avoid using full Perl-Compatible Regular Expression (PCRE[11]) features, since the usage of regular expressions is time-consuming for the framework/tool. While most regular expressions processing is time consuming, Rematch tool [28] can match thousands of regular expressions against a data stream at line speeds,, thus leading to earlier detection and identification with total inspection in each network. For that purpose, H. PARK et al [29], evaluated the features and performance of regular expression processing algorithms.

The traditional situational awareness methodologies rely on static network topologies, while most of them rely upon a unified communication protocol. For the afore-mentioned reasons and because in the IoT domain, power consumption should be included in the parameters, F. He et al [30] defined a Stochastic Coloured Petri Net (SCPN), focused on the IoT domain and then proposed a game model for cybersecurity situational awareness. Through SCPN, coloured tokens represent different types of threats, therefore even collaborative attacks are clearer to understand and mitigate. The game process includes players making decisions (while simultaneously each decision affects the other player (attacker/defender)), and selecting strategies, considering the current state.

H. Zhang et al [31] proposed a system composed of IDS sensors, an anomaly detection algorithm, and firewalls. While active sensors will monitor the traffic, passive sensors will exist within hosts and network-enabled entities to gather logs linked with cyber threats. Through the info provided by the sensors (active and passive) combined with detection schemes, the mitigation will occur utilizing the firewalls

### 2.1.2 Machine Learning

In Machine Learning (ML), a sample is represented by several features forming a multidimensional feature vector. ML systems operate in two phases: the learning phase (training) and testing one. The role of the pre-processing unit is to normalize data, remove noise and apply any other function or routine that will contribute

[11] https://en.wikipedia.org/wiki/Perl_Compatible_Regular_Expressions

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

to the formulation of a more compact representation of the samples. During the training phase, the feature extraction/selection unit attempts to generate and/or identify the most informative feature subset in which the learning model will be applied [32]. The feedback loop allows adjustments of the pre-processing and feature extraction/selection units that will further improve the performance of the learning model. During the testing phase, the trained model is utilized to take an appropriate decision (classification or regression) for each one of the testing samples based on the selected features. Deep learning [33], which is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain, sets an alternative architecture by shifting the burden of feature engineering to the underlying learning system. From this perspective, pre-processing and feature extraction or selection are omitted, leading to a fully trainable system that begins from raw input (e.g. image pixels or time-series) and ends with the final output of recognized objects or predicted values.

Learning can be classified as supervised, unsupervised or reinforcement learning. In supervised learning, each data sample is represented by a pair consisting of an input (typically a multi-dimensional feature vector) and the desired output value (e.g. a label). The training phase involves the task of learning a function that maps every input to its associated output. The generated inferred function is used to map unknown inputs during the testing phase. Unsupervised learning is a class of ML techniques that operate with unlabelled data to discover interesting structures or patterns in the dataset. In reinforcement learning, a model learns through trial and error interactions with its environment using reward and penalty assignments.

In the terminology of ML, classification is considered as an instance of supervised learning. In short, it is the task of identifying to which of a set of categories (sub-populations) a new example belongs, based on a training set of data (experience) containing examples whose label is known. Regression constitutes another supervised learning task, which aims to provide a prediction of an output variable according to the input variables, which are known. The most known regression algorithms are linear regression and logistic regression [34], as well as, stepwise regression [35]. Also, more complex regression algorithms have been developed, such as ordinary least squares regression [36], multivariate adaptive regression splines [37], multiple linear regression, and locally estimated scatterplot smoothing [38].

Dimensionality reduction (DR) is a task that belongs in both families of supervised and unsupervised learning types, to provide a more compact lower-dimensional representation of a dataset preserving as much information as possible from the original data. It is usually performed prior to applying a classification or regression model in order to avoid the effects of the curse of dimensionality. Some of the most common DR algorithms are the following: (i) principal component analysis [39], (ii) partial least squares regression [40] and (iii) linear discriminant analysis [41]. Finally, clustering is an application of unsupervised learning typically used to find natural groupings of data (clusters). Well established clustering techniques are the K-means technique [42], hierarchical clustering [43], and the expectation-maximization technique [44].

A relatively new area of ML research is Deep learning (DL) [33] which is allowing computational models that are composed of multiple processing layers to learn complex data representations using multiple levels of abstraction. One of the main advantages of DL models is that in some cases the step of feature extraction is performed by the model itself. Currently, DL models have dramatically improved the state-of-the-art in many different sectors and industries including healthcare. A deep neural network (DNN) is an ANN with multiple hidden layers between the input and the output layers and can be either supervised, partially supervised or even unsupervised. A common DL model is the convolutional neural network (CNN), where feature maps are extracted by performing convolutions in the image domain. A comprehensive introduction to CNNs is given in [45]. Other typical DL architectures include deep Boltzmann machine, deep belief network [46], and auto-encoders [47].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

#### 2.1.2.1 Machine Learning in Automated Intrusion Detection (AID)

In the last few decades, ML has been used to improve intrusion detection. There is a large number of related studies using various synthetic datasets (such as KDD-Cup 99 [48] or DARPA 1999 [49] datasets) to develop and validate ML-empowered Automated Intrusion Detection (AID) systems. Before proceeding with the presentation of the studies in the recent literature, a short description of the AID problems and challenges is provided below.

*Definition of the problem*: In a common AID system, machine learning, statistical-based or knowledge-based methods are used to define a normal model of the behaviour of a computer system. Any significant deviation between the observed 'normal' behaviour can be regarded as an anomaly, which can be then interpreted as an intrusion. The main assumption of the aforementioned approaches is that malicious behaviour differs from typical user behaviour. One simplistic method to decide whether a behaviour is normal or abnormal is by comparing it with the standard deviation of the normal user behaviours in the training dataset. Any example exceeding the pre-determined threshold (e.g. three times the standard deviation) could be classified in the intrusion category. ML provides a more sophisticated method for decision making overcoming the deficiencies of the heuristic approaches (such the manual selection of the threshold etc). Development of ML-based AID systems comprises of two phases: the training phase and the testing phase.

1. In the training phase, the normal traffic profile is used to learn a model of normal behaviour,
2. In the testing phase, a new data set is used to validate the system's capacity to generalize to previously unseen intrusions.

AIDS can be classified into several categories based on the method used for training, for instance, statistically based, knowledge-based and machine learning-based [50].

*Advantages of AID*: The main advantages of ML-empowered AID systems are:

- Their ability to identify zero-day attacks without relying on a signature database [51]. A danger signal can be triggered when the examined behaviour differs from the usual behaviour.
- Their capability to discover internal malicious activities. An alarm will be created in cases where an intruder starts making transactions in a stolen account that are unidentified in the typical user activity.
- The normal user behaviour is hidden to intruders and thus it becomes more difficult for them to remain undetected.

The objective of using machine learning techniques is to create IDS with improved accuracy and less requirement for prior human knowledge. However, one of the main *challenges* of current AIDS is the high false-positive rates because anomalies may just be new normal activities rather than genuine intrusions.

#### 2.1.2.2 Knowledge extraction in Automated Intrusion Detection Systems

One of the crucial phases in today's ML pipelines is the process of extracting knowledge from large quantities of data. To effectively extract knowledge from raw data, ML relies on a set of rules, methods, or complex "transfer functions" that are applied to find interesting data patterns or to recognize and predict behaviour [52]. Many ML algorithms (such as clustering, neural networks, association rules, decision trees, genetic algorithms, and nearest neighbour methods) have been recently applied in the area of AIDS for discovering knowledge from intrusion datasets [53], [54].

Some prior research in data mining has examined the use of different algorithms to extract meaningful information for intrusion data. Two feature selection algorithms were investigated by Chebrolu et al. employing Bayesian networks (BN) and Classification Regression Trees (CRC). The outputs of the aforementioned algorithms were finally combined to increase accuracy[55]. Bajaj et al. proposed a technique for feature selection using a hybrid approach that combines Information Gain (IG) and correlation attribute evaluation. To validate the discrimination capacity of the selected features, the authors applied several classification algorithms such as C4.5, naïve Bayes, NB-Tree and Multi-Layer Perceptron ([56]-[57]). Genetic-fuzzy rule mining has been also explored to evaluate the importance of IDS features in [58]. Thaseen et al. proposed a Random

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

Tree model to improve the accuracy and reduce the false alarm rate [59], whereas Subramanian et al. also studied the performance of decision tree algorithms on the NSL-KDD dataset [60].

#### 2.1.2.3 Supervised Learning in Intrusion Detection Systems

Supervised learning-based IDS techniques detect intrusions by using labelled training data. Specifically, relevant features and classes are identified in the training phase and the algorithm learns from these data samples. This means that each record is a pair, containing a feature or a feature set (e.g. features extracted from a network or host data source) and an associated output value (i.e., label), namely intrusion or normal. Next, feature selection is applied for (i) ranking features concerning their importance as well as (ii) eliminating unnecessary features. A supervised learning technique is then trained on the selected features to learn the inherent relationship that exists between the input data and the labelled output value. In the testing stage, the trained model is used to classify the unknown data into intrusion or normal class. The resultant classifier then becomes a model which, given a set of feature values, predicts the class to which the input data might belong.

Given that there are many classification methods (e.g. decision trees, rule-based systems, neural networks, support vector machines, naïve Bayes and nearest-neighbour), selecting the most appropriate classification method is not a straightforward task. Each technique uses a learning method to build a classification model. However, the 'best' learning model should not only handle the training data, but it also should identify accurately unknown records (not included in the training sets). Thus, identifying the classification model with reliable generalization ability is another important factor that should be taken into account. The Table below cites several supervised ML models that have been applied to the AID problem in the recent literature

| Method | Principles | Features | Reference |
|---|---|---|---|
| Naïve Bayes | Using conditional probability formulas, it can calculate the probability that a particular kind of attack is occurring, given the observed system activities.<br>It relies on the features that have different probabilities of occurring in attacks and in normal behaviour. | Ease of use and calculation efficiency, both of which are taken from its conditional independence assumption property.<br>It does not operate well if this independence assumption is not valid; It has reduced accuracy for large datasets | Yang & Tian, 2012 [61] |
| | | A more sophisticated Hidden Naïve Bayes (HNB) model that can be applied to IDS tasks that involve high dimensionality, extremely interrelated attributes, and high-speed networks | Koc et al., 2012 [62] |
| Fuzzy logic | Proper choice for IDS problems as they include vagueness, and the borderline between the normal and abnormal states is not well identified | The system derives a group of fuzzy rules to describe the normal and abnormal activities in a computer system, and a fuzzy inference engine to define intrusions | Elhag et al., 2015 [58] |
| Support Vector Machines (SVM) | They use a kernel function to map the training data into a higher dimensional space so that intrusion and normal | Feature selection was applied to reduce the feature dimensionality and an SVM-RBF classifier was applied to | Li et al., 2012 [63] |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

| | | | |
|---|---|---|---|
| | user data are linearly classified | classify the KDD 1999 dataset into predefined classes. | |
| Nearest Neighbours (KNN) classifiers | An example is classified by a plurality vote of its neighbours, and therefore is assigned to the most common class among its k nearest neighbours | Typically applied as a benchmark for other classifiers.<br>It provides a moderate classification performance in most IDSs | Lin et al., 2015<br>[64] |
| Artificial Neural Networks (ANN) | Computing systems vaguely inspired by the biological neural networks; ANNs are the most broadly applied ML methods and has been shown to be successful in detecting different malware | Increased detection capabilities: It can be biased to the majority class and this makes it difficult for ANNs to learn the properties of less frequent attacks correctly. As a result, detection accuracy might deteriorate for the minority class. They could also be time consuming. | Wang et al., 2010<br>[65] |
| Hidden Markov Models (HMM) | HMM is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobservable (i.e. hidden) states. | HMMs are trained against known malware features (e.g., operation code sequence). The trained model is then applied to score incoming traffic. Decisions are made on a predefined threshold. Scores above the threshold indicate malware and vice versa. | Annachhatre et al., 2015<br>[66] |
| Genetic algorithms (GA) | Heuristic approaches to optimization, based on the principles of evolution. | GA was used to evolve simple rules for network traffic. Every rule was represented by a genome and the primary population of genomes was a number of random rules. | Murray et al., 2014<br>[67] |
| | | In this paper, each genome was comprised of different genes, which correspond to characteristics such as IP source, IP destination, port source, port destination and protocol type. | Hoque & Bikas, 2012<br>[68] |
| Ensemble methods | Ensemble learning helps improve ML results by combining several weak models | Random forest improvement was applied for intrusion detection on the Kyoto dataset accomplishing increased accuracies compared to basic classifiers (e.g. Bayes). | Jabbar et al., 2017<br>[69] |

***Table 1 Supervised ML Models applied to AID problem***

#### *2.1.2.4 Deep learning in Automated Intrusion Detection*

Unlike ML approaches that require the extraction of features, Deep learning (DL)-based detection methods learn feature automatically in an end-to-end fashion (directly from raw data to decisions). DL is gradually attracting more interest in AID studies. A CNN-based AID methodology was presented by Potluri et al. [70] conducting experiments on the NSL-KDD and the UNSW-NB datasets. In the pre-processing phase, the features of the datasets were transformed into images of 8*8 pixels. Then, a three-layer CNN was trained to classify the attacks. Pre-trained deep networks (ResNet 50 and GoogLeNet) were also explored as alternative solutions to

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

the task of extracting new informative features. The proposed CNN performed best, reaching accuracies of 91.14% on the NSL-KDD and 94.9% on the UNSW-NB 15.

A sparse autoencoder was also proposed by Zhang et al. [71], to extract features from the NSL-KDD dataset. The extracted features were supplied to an XGBoost model with the objective to detect attacks. To overcome the observed data imbalance problem, data resampling was employed (using SMOTE). The SMOTE algorithm oversamples the minority classes and divides the majority classes into many subclasses so that every class is balanced.

Data augmentation with GANs has been also explored by Zhang et al. [72]. The GAN model was used to generate data similar to the flow data of KDD99. Adding this generated data to the training set increased the generalization capacity of the detection model that was able to identify not only attacks but attack variants as well.

### 2.1.3 Risk Models

Cyber risk is traditionally considered as part of operational risk in corporation risk management. The approach of seeing cyber risk only applicable on operational level was limiting the effectiveness of risk management, mostly because it was not taking into consideration several factors that play significant role in the core value generation process of business. Under today's business context, it is increasingly evident that cyber risk should be embedded into all parts of critical business risk. Moreover, the rapid pace of increasing complexity and the potential impact levels of cyber threats demand better prioritisation, availability of resources and prompt reaction than any other type of risk corporations face today.

Risk management is a central concern for every organisation. Risk can take different forms and originate from either inside or outside the organisation. IT security is amongst one of the concerns that drive strategy at every corporation, including the risk of non-compliance, data breaches, infrastructure outages, legal penalties and more.

Risk management can be described as consisting of four core processes: Context definition, Risk Assessment, Actions needed and Monitoring. National Institute of Standards and Technology (NIST) describes these processes as risk framing, risk assessment, risk response, and risk monitoring, while International Organization for Standardization (ISO) highlights a bit more some areas and describes the processes as Communication and consultation, Scope- context and criteria, Risk assessment, Risk treatment, Monitoring and review and Recording and reporting.

Information security regulations are getting stricter. They are heavily focused on risk management and putting controls in place to prevent potential threats. The General Data Protection Regulation (GDPR), for example, was approved by the EU parliament to strengthen data protection regulations. Noncompliant organisations can face massive fines. This is where threat modelling comes into play to address all the underlying sub-threats and root causes of higher-level threats.

Risk assessment is the initial step of risk management and constitutes the most critical and difficult phase. In order to assess the scenarios that compose the threats, a risk assessment model needs to be structured. Using a simplified interpretation, a risk assessment model can be seen as a set of rules by which we aim to predict the future performance of a system from a risk perspective. Threat modelling, combined with risk management, should give answers to the question of who will attack your own systems, and how or where the attack will originate from. Threat modelling will provide valuable insights on IT risks facing organisations, and then outline necessary measures and sufficient controls to stop the threat before it becomes effective.

The main goal of any risk assessment model is to provide a relative or absolute quantification of risks. Models try to encapsulate in a comprehensible structure, the aspects of a real problem using simplification in contrast

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

with simulation techniques that try to reproduce a specific set of conditions of the problem. All models, from the simplest to the most complex ones, make use of probability theory and statistics. In simple applications where expert reasoning drives the assessment this is not so clear, while in mathematically rigorous models is obvious. Risk assessment brings together all aspects of the threat model with an environmental model (i.e. a representation of the operational and technical environment in which threats could occur), so that the likelihood and consequence severity of threat scenarios or individual threat events can be estimated or evaluated.

> *"Risk models define the risk factors to be assessed and the relationships among those factors. Risk factors are characteristics used in risk models as inputs to determining levels of risk in risk assessments. Risk factors are also used extensively in risk communications to highlight what strongly affects the levels of risk in particular situations, circumstances, or contexts. Typical risk factors include threat, vulnerability, impact, likelihood, and predisposing condition. Risk factors can be decomposed into more detailed characteristics (e.g., threats decomposed into threat sources and threat events). These definitions are important for organizations to document prior to conducting risk assessments because the assessments rely upon well-defined attributes of threats, vulnerabilities, impact, and other risk factors to effectively determine risk." [NIST 2012]*

#### 2.1.3.1 General Approaches

Models can be classified into three general categories. From simplest to most complex, are matrix, probabilistic, and indexing models.

Matrix models

One of the simplest risk assessment structures is a decision-analysis matrix. It ranks risks according to the likelihood and the potential consequences of an event by a simple scale, such as high to low, or 1 to 10. Each threat is assigned to a cell of the matrix based on its perceived likelihood and consequences. This approach may simply use expert opinion or more complicated applications through quantitative information to rank risks. While this approach cannot consider all pertinent factors and their relationships, it does help to clarify thinking at least by breaking the problem into two parts for separate examination.

Probabilistic models

The most rigorous and complex risk assessment model is a modelling approach commonly referred to as probabilistic risk assessment. These models use mathematical and statistical techniques that relies heavily on historical failure data and event-tree/fault-tree analysis. Initiating events such as equipment failure and safety system malfunction are flowcharted forward to all possible concluding events, with probabilities being assigned to each branch along the way. Failures are backward flowcharted to all possible initiating events, again with probabilities assigned to all branches. All possible paths can then be quantified based on the branch probabilities along the way. Final accident probabilities are achieved by changing the estimated probabilities of individual events.

These models are technologically more demanding to develop, require trained operators, and need extensive data. A detailed probabilistic risk assessment is usually the most expensive of the risk assessment techniques.

The output of a probabilistic risk assessment is usually in a form whereby its output can be directly compared to other risks. However, in rare-event occurrences, the lack of historical data leads to an arguably blurred view. The technique therefore makes extensive use of failure statistics of components as foundations for estimates of future failure probabilities. However, as statistics can provide part of the probabilistic relationships between the nodes, many probabilities must still be assigned by experts. In order to minimize subjectivity, applications

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

of this technique became increasingly comprehensive and complex, requiring thousands of probability estimates.

Indexing models

The most popular risk assessment technique in current use is the *index model* or some similar scoring technique. In this approach, numerical values (scores) are assigned to important conditions and activities that contribute to the risks. In order to calculate these scores risk-reducing and risk-increasing variables are introduced. Weightings are also assigned to each variable, which reflects the importance of the specific item in the risk assessment and which is based on statistics, if available, or on experts' opinion where data are limited or not available.

These models are very comprehensive and less demanding compared to the probabilistic ones. They also provide output that can be directly compared to other risks. As greatest challenges, the efficiency, scalability, and performance are central factors to consider at any level when designing or building an index.

Selection of risk assessment approach

Any or all the above-described approaches can be applied in risk assessment/management. Understanding the strengths and weaknesses of the different risk assessment methodologies gives the decision-maker the basis for choosing one. For example, a simple matrix approach helps to organise thinking and is the initial step towards formal risk assessment. If the need is to evaluate specific events at any point in time, a narrowly focused probabilistic risk analysis might be the answer. Correspondingly, an index models should be used for weighing immediate risk trade-offs or perform inexpensive overall assessments.

In principal, the pros and cons of qualitative and quantitative risk assessments are summarised in the following table

| | Pros | Cons |
|---|---|---|
| **Quantitative** | • The results are based on independently objective processes and metrics, which removes the amount of subjectivity.<br>• It provides greater insight of asset value determination and risk mitigation.<br>• Cost/benefit assessment.<br>• The results can be expressed in management specific language (e.g., monetary value, percentages, probabilities etc.). | • Calculations can be complex and time-consuming.<br>• Higher costs.<br>• Requires large amounts of preliminary work in collecting and quantifying the different risk analysis components.<br>• Participants cannot be coached easily through the process. |
| **Qualitative** | • Simpler without complex calculations.<br>• It is not necessary to determine the monetary value of assets.<br>• It is not necessary to quantify threat frequency.<br>• It is easier to involve non-security and non-technical staff. | • Subjectivity.<br>• Results and quality of the risk assessment depend solely on the expertise and quality of the risk management team.<br>• Limited effort to develop "value" for targeted assets<br>• No basis for the cost/benefit analysis of risk mitigation. |

***Table 2: Qualitative vs Quantitative approaches***

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

#### 2.1.3.2 Choosing a Technique

According to ISO, the techniques apply differently to each step of the risk assessment process, and can be classified as follows:

- risk identification
- risk analysis – consequence analysis
- risk analysis – qualitative, semi-quantitative or quantitative probability estimation
- risk analysis – assessing the effectiveness of any existing controls
- risk analysis – estimation the level of risk
- risk evaluation.

Risk assessment may be undertaken in varying degrees of depth and detail and using one or many methods ranging from simple to complex. The form of assessment and its output should be consistent with the risk criteria developed as part of establishing the context. In general terms, suitable techniques should exhibit the following characteristics:

- it should be justifiable and appropriate to the situation or organisation under con-sideration
- it should provide results in a form which enhances understanding of the nature of the risk and how it can be treated
- it should be capable of use in a manner that is traceable, repeatable, and verifiable.

The reasons for the choice of techniques should be given, regarding relevance and suitability. When integrating the results from different studies, the techniques used, and outputs should be comparable. Once the decision has been made to perform a risk assessment and the objectives and scope have been defined, the techniques should be selected, based on applicable factors such as:

- the objectives of the study
- the needs of decision-makers
- the type and range of risks being analysed
- the potential magnitude of the consequences
- the degree of expertise, human and other resources needed
- the availability of information and data
- the need for modification/updating of the risk assessment
- any regulatory and contractual requirements.

<u>The nature and degree of uncertainty</u>

It is important that a risk assessment identifies the role of uncertainty in its use of assumptions and identifies how the state of limited or no information is assessed. The nature and degree of uncertainty requires an understanding of the quality, quantity, and integrity of information available concerning the risk under consideration. Available data do not always provide a reliable basis for the prediction of the future. For unique types of risks, historical data may not be available or there may be different interpretations of available data by different stakeholders. In such cases risk assessment need to understand the type and nature of the uncertainty and appreciate the implications for the reliability of the risk assessment results, which clearly affects the final selection of the appropriate technique.

<u>Complexity</u>

Consequential impacts and risk dependencies need to be understood to ensure that in managing one risk, an intolerable situation is not created elsewhere. Understanding the complexity of a single risk or of a portfolio of risks of an organisation is crucial for the selection of the appropriate method or techniques for risk assessment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

Risk assessment during life cycle phases

Risk assessment can be applied at all stages of the life cycle and is usually applied many times with different levels of detail to assist in the decisions that need to be made at each phase. Life cycle phases have different needs and require different techniques. Where several options are available, risk assessment can be used to evaluate alternative concepts to help decide which provides the best balance of risks.

Capture all aspects

Most qualitative risk models lack granularity and objectivity, while quantitative models lack efficiency, statistical robustness, and reliable asset valuation. Moreover, most methods focus on technology and are limited in covering people, process, and socio-economic risk factors. An information system is comprised of technology, people, processes, and data. Therefore, effective risk analysis must examine each of these aspects. Traditional risk models are inadequate as they are technology-driven and focus primarily on known threats to types of computing assets employed by an organisation [73]. A tech-centric approach does not involve business-users to the extent necessary to identify a comprehensive set of risks, or to promote risk awareness throughout an organisation. Finally, the lack of solid business case for risk evaluation means the lack of accountability and prioritisation in implementing mitigation actions. In practice inefficient approaches of risk assessment and measurement tend to be considered as a barrier rather than a necessity.

#### *2.1.3.3 Cybersecurity Frameworks*

Various frameworks have been developed to assist organisations in achieving robust cybersecurity programs. Cybersecurity frameworks refer to defined structures containing processes, practices, and technologies, which organisations can use to secure network and computer systems from security threats. The key points of the most known cybersecurity frameworks are:

- **Joint Task Force Transformation Initiative (NIST SPs)**

The risk management process as defined in NIST SP 800-39 consists of four activities: risk framing, risk assessment, risk response, and risk monitoring. NIST SP 800-39 defines a risk frame as “the set of assumptions, constraints, risk tolerances, and priorities/trade-offs that shape an organisation’s approach for managing risk.” The assumptions about threat sources and threat events – specifically including the types of adversarial tactics, techniques, and procedures (TTPs) to be addressed, and adversarial characteristics (e.g., capability, intent, targeting) – implicitly or explicitly define the organisation’s threat model. This threat model is further refined and populated when risk assessments are performed, and the populated values are updated as part of risk monitoring.

NIST SP 800-30R1 provides a representative threat model as part of an overall risk assessment methodology. That threat model includes

- A taxonomy of threat sources, with accompanying characteristics for adversarial threats (capability, intent, and targeting) and for non-adversarial threats (range of effects)
- A representative set of adversarial threat events, using the structure of a cyber campaign (i.e., a cyber-attack lifecycle), and a representative set of non-adversarial threat events
- A taxonomy of predisposing conditions (i.e., environmental factors which affect the likelihood of threat events occurring or resulting in adverse consequences) Because vulnerabilities are characterized in a wide variety of ways, NIST SP 800-30R1 does not include a taxonomy of vulnerabilities.

NIST SP 800-30R1 does not prescribe this threat risk model.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

- **NIST Cybersecurity Framework for Improving Critical Infrastructure Cybersecurity**

A revision of NIST's Framework for Improving Critical Infrastructure Cybersecurity was published in 2018. The Cybersecurity Framework (CSF) defines a high-level approach to risk management, to complement the cybersecurity programs and risk management processes of organisations in critical infrastructure sectors. The CSF does not define cyber threat modelling terms, but uses the following terms: cybersecurity threats, threat exposure, threat environment, evolving and sophisticated threats, and cyber threat intelligence. CSF describes five functions that manage the risks to data and information security. The functions are "identify", "protect", "detect", "respond", and "recover".

- **ISO IEC 27001/ISO 27002**

The ISO 27001 cybersecurity framework consists of international standards, which recommend the requirements for managing information security management systems (ISMS). ISO 27001 observes a risk-based process that requires businesses to put in place measures for detecting security threats that impact their information systems. To address the identified threats, ISO 27001 standards recommend various controls. An organisation should select proper controls that can mitigate security risks to ensure it remains protected from attacks. In total, ISO 27001 advocates a total of 114 controls, which are categorized into 14 different categories. Some of the categories include information security policies containing two controls; information security organisation with seven controls that detail the responsibilities for various tasks; human resource security category with six controls for enabling employees to understand their responsibility in maintaining information security; among others.

ISO 27002 framework comprises of international standards that detail the controls which an organisation should use to manage the security of information systems. The ISO 27002 is designed for use alongside ISO 27001, and most organisations use both to demonstrate their commitment to complying with various requirements required by different regulations. Some of the information security controls recommended in the ISO 27002 standard include policies for enhancing information security, controls such as asset inventory for managing IT assets, access controls for various business requirements and for managing user access, and operations security controls.

- **CBEST Framework**

The CBEST framework was created, developed, and it is run by the Bank of England. CBEST provides a structured and controlled approach for intelligence-led security testing within the financial sector.

The CBEST Threat Assessment identifies two things:

- Targetable information on organisation that can be used by adversaries
- The contextualised information and intelligence assessment on organisation's most likely adversaries, including their capabilities, motives, and intent

The CBEST approach focuses on identification of specific threat actors and their common attack patterns to generate actionable cyber reconnaissance. Using as much intelligence as is available, analysts using the CBEST approach analyse each specific threat actor's identity and motivations more deeply than in most models. It models what is known about the threat actor's phases of operation; TTPs; countermeasures against discovery; timing and coordination of activity. The CBEST approach is intended to enable analysts, given adequate cyber threat intelligence data, to derive a model of threat actors rigorous and precise enough to be predictive of likely threat events. Though this level of threat intelligence may often not be available, the CBEST approach seeks to generate the most realistic threat scenarios possible given the information at hand.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

- **COBIT**

Control Objectives for Information and Related Technologies (COBIT) is a cybersecurity framework that integrates a business's best aspects to its IT security, governance, and management. ISACA (Information Systems Audit and Control Association) developed and maintains the framework. The COBIT cybersecurity framework is useful for companies aiming at improving production quality and at the same time, adhere to enhanced security practices. The factors that led to the creation of the framework are the necessity to meet all stakeholder cybersecurity expectations, end to end procedure controls for enterprises, and the need to develop a single but integrated security framework.

COBIT is based on components of the ISO standards, including incorporation of the ISO 38500 model for the corporate governance for IT and an ISO 15504 aligned COBIT Process Capability Assessment Model. Security controls are based on the ISO 27001 series of control objectives. This includes assessment considerations aligned with operational practice, implementation guidance, measurement, and risk management. COBIT is accompanied by the Risk IT framework for managing business risks of IT. Risk IT consists of a risk model together with a process model; processes are defined for the domains of risk governance, risk evaluation, and risk response. The model underlying risk evaluation in Risk IT is not a security risk model but does identify security risk as a class of risk to be considered. A risk scenario is described in terms of threat type (which includes malicious threats), actor, type of event (i.e., type of impact), asset or resource affected, and time. In addition, the scenario planning approach in Risk IT's risk assessment framework allows for risk consideration beyond an individual organisational or system view.

- **Federal Information Systems Management Act (FISMA)**

FISMA is a framework designed for federal agencies. The compliance standard outlines a set of security requirements that government agencies can use to enhance their cybersecurity posture. The security standards aim at ascertaining that federal agencies implement adequate measures for protecting critical information systems from different types of attacks. Moreover, the framework requires vendors or third parties interacting with a government agency to conform to the stipulated security recommendations. The main aim of the security standard is to enable federal agencies to develop and maintain highly effective cybersecurity programs. To achieve this, the standard consists of a comprehensive cybersecurity framework with nine steps for securing government operations and IT assets.

- **SOC 2**

The American Institute of Certified Public Accountants (AICPA) developed the SOC 2 framework. The framework's purpose is to enable organisations that collect and store personal customer information in cloud services, to maintain proper security. Also, the framework provides SaaS companies with guidelines and requirements for mitigating data breach risks and for strengthening their cybersecurity postures. Also, the SOC 2 framework details the security requirements which vendors and third parties must conform. The requirements guide them in conducting both external and internal threat analysis to identify potential cybersecurity threats. SOC 2 contains a total of 61 compliance requirements, and this makes it among the most challenging frameworks to implement. The requirements include guidelines for destroying confidential information, monitoring systems for security anomalies, procedures for responding to security events, internal communication guidelines, among others.

- **CIS v7**

CIS v7 lists 20 actionable cybersecurity requirements meant for enhancing the security standards of all organisations. The framework categorizes the information security controls into three implementation groups. Implementation group 1 is for businesses that have limited cybersecurity expertise and resources. Implementation group 2 is for all organisations with moderate technical experience and resources in

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

implementing the sub controls. Implementation group 3 targets companies with vast cybersecurity expertise and resources. CIS v7 stands out from the rest since it enables organisations to create budget-friendly cybersecurity programs. It also allows them to prioritize cybersecurity efforts.

- **Committee of Sponsoring Organizations (COSO)**

COSO is a framework that allows organisations to identify and manage cybersecurity risks. The core points behind the development of the framework include monitoring, auditing, reporting, controlling, among others. The framework consists of 17 requirements, which are categorized into five different categories. The categories are control environment, risk assessments, control activities, information, communication, monitoring and controlling. All the framework's components collaborate to establish sound processes for identifying and managing risks. A company using the framework routinely identifies and assess security risks at all organisational levels, thus improving its cybersecurity strategies. Also, the framework recommends communication processes for communicating information risks and security objectives up or down in an organisation. The framework further allows for continuous monitoring of security events to permit prompt responses.

- **Health Information Trust Alliance (HITRUST) CSF**

HITRUST cybersecurity framework addresses the various measures for enhancing security. The framework was developed to cater to the security issues organisations within the health industry face when managing IT security. This is through providing such institutions with efficient, comprehensive, and flexible approaches to managing risks and meeting various compliance regulations. In particular, the framework integrates various compliance regulations for securing personal information. Such include Singapore's Personal Data Protection Act (PDPA) and interprets relevant requirement recites from the General Data Protection Regulation (GDPR). HITRUST cybersecurity framework is regularly revised to ensure it includes data protection requirements that are specific to the HIPPA regulation.

- **North American Electric Reliability Corporation Critical Infrastructure Protection (NERC CIP)**

NERC CIP is a cybersecurity framework that contains standards for protecting critical infrastructures and assets. The framework has nine standards comprising of 45 requirements. The critical cyber asset identification standard makes it mandatory for an entity to document all cyber assets considered to be critical. Also, personnel and training standard requires employees with access to critical cyber assets to complete security and awareness training. Other standards included in the NERC CIP framework are electronic security perimeter, incident response, managing systems security, and maintaining recovery plans.

#### *2.1.3.4 Threat Modelling to Support Identification*

In the previous list of Cyber Security Frameworks some of them incorporate a threat model within the overall risk model. At the system implementation or operations level, a threat risk model can highlight the necessity for the selection of specific security controls or/and courses of action and support decisions or security operations. At the business function level, a threat risk model can support the organisation's information security architecture, and its business function architectures. At the organisational level, a threat risk model reflects and expresses the organisation's assumptions about its threat environment; these are an integral part of the organisation's risk frame.

Threat modelling for risk assessment can follow three different approaches:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

- Start with modelling the threat, generally or specifically, and then apply it to a relevant environment
- Start with modelling the systems, data, and boundaries in the environment and then determine what threats are relevant
- Start by identifying the assets that could be affected by threats, characterizing the threats that could affect or target those assets, and situating the assets in terms of systems [74].

These three approaches are illustrated in Figure 1. Independently of which the starting point is, all approaches take into consideration all the other aspects of risk either implicitly or explicitly.

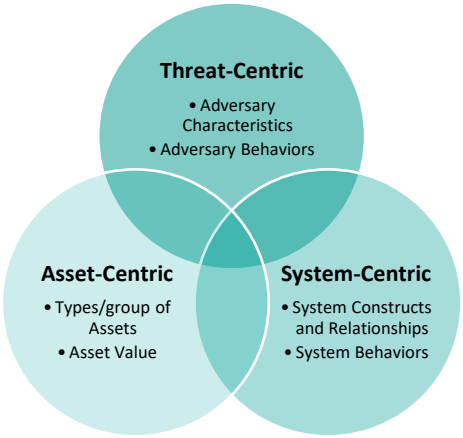


***Figure 1: Threat Modelling Approaches [MITRE]***

Several commonly used terms are utilised in threat modelling, including "threat", "threat actors", "threat vector", "threat scenario", "threat event", "attacker" and "attack vector". According to NIST Special Publication (SP) 800-30R1 the definition of some standard terms related to threat is:

Threat: Any circumstance or event with the potential to adversely impact organisational operations and, assets, individuals, other organisations, or the Nation through an information system via unauthorized access, destruction, disclosure, or modification of information, and/or denial of service.

Threat sources: adversarial, accidental, structural, and environmental.

Threat actors: Individuals, groups, organisations, or states that seek to exploit the organisation's dependence on cyber resources (i.e., information in electronic form, information and communications technologies, and the communications and information-handling capabilities provided by those technologies).

Threat event: An event or situation that has the potential for causing undesirable consequences or impact.

Threat scenario: A threat scenario is a set of discrete threat events, attributed to a specific threat source or multiple threat sources, ordered in time, that result in adverse effects.

Attack vectors: The behaviours or actions of an adversarial threat actor can be characterized in terms of the threat vector or avenues of attack that are general approaches to achieving cyber effects [NIST 800-61].

This terminology comprises only part of a larger setting of terminology about risk, which, indicatively, also includes:

Information Asset: a body of knowledge that is organised and managed as a single entity. Like any other corporate asset, an organisation's information assets have financial value.

Attack Surface: the sum of the different points (the "attack vectors") where an unauthorized user (the "attacker") can try to enter data to or extract data from an environment.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

Likelihood: the possibility of a threat event occurring where a threat actor will exploit a weakness. The likelihood of threat events resulting in adverse impacts estimates the possibility that a threat event would result in an actual outcome. The combined analysis of both threat assessment vectors impacts established an overall threat likelihood.

Impact: the potential damage (physical, logical, monetary loss, etc.) of a threat event.

Control: a safeguard or countermeasure to avoid, detect, counteract, or minimize security risks to information, computer systems, or other assets.

Mitigation: A systematic reduction of risk or likelihood's impact to an asset.

Tractability Matrix: a grid that allows documentation and easy viewing of what is required for a system's security.

Threat modelling approaches depend on assumptions about the technological and operational environment in which risk will be managed, therefore there are differences in the definition of those terms. Using any of these threat modelling approaches, risk is estimated by assessing identified threat events or scenarios, in the context of relevant vulnerabilities and environmental assumptions, as to likelihood of occurrence and severity of impact. The resulting measure is the result of any inherent risk, mitigated by the implemented controls, and constitutes a measure of residual risk. This process may iterate as additional controls are identified and implemented, and as evolving threat capabilities are identified and reported. Measuring risk levels and identifying operational processes that support ongoing mitigation of cyber threats should result in a reporting capability for significant risk-based metrics. Risk metrics are critical to providing executive managers with oversight capabilities to establish a cyber program baseline to manage acceptable residual risk to the institution.

#### *2.1.3.4.1 Threat Modelling frameworks and methods*

Attack modelling techniques (AMT) are used to model and visualise the sequence and/or combination of events that enable a successful cyber-attack on a computer or network. AMTs can be broadly divided into three categories: methods that are based on the use case framework, methods that present a cyber-attack from a temporal perspective, and graph-based methods. These methods are highlighted in Figure 2 [75].

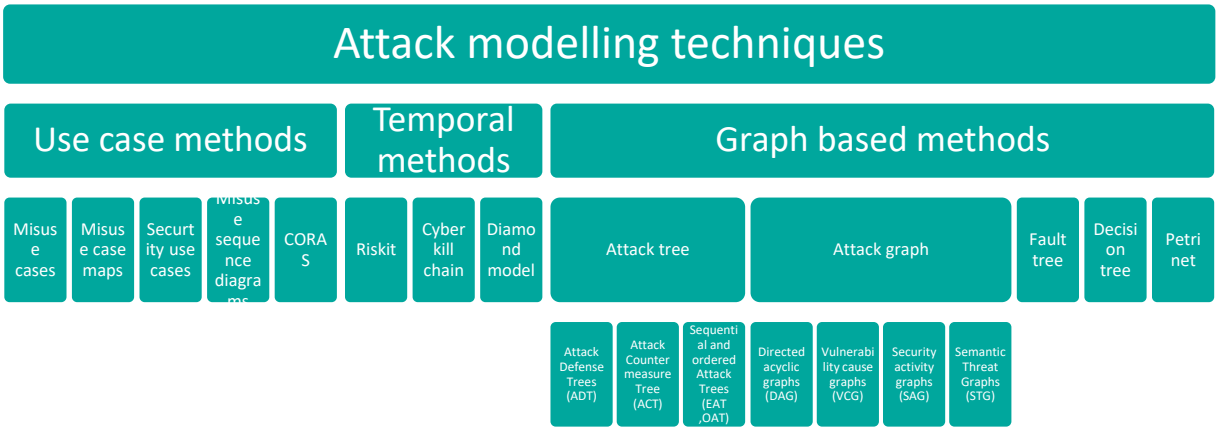


***Figure 2: Attack modelling techniques***

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

- **Graph based methods**

Attack graphs, attack trees and their variants, which include: OCTAVE, event trees and decision trees, are graph-based representations of a cyber-attack. Of these, attack graphs and attack trees are the most well-established approach to developing threat scenarios for risk assessment.

AMTs enable observers to evaluate the salient information in a diagram [76][77][78][79] and help remove the intellectual burden from security experts — who must evaluate cyber-attack scenarios and evaluate potential mitigations [80]. Consequently, security problems can be presented in a manner that enables a decision maker — whether an expert or non-expert, to grasp the problem more quickly, to better perceive risk landscapes [81], and easily perceive complex concepts [82]. In such circumstances, AMTs provide effective tools and workspaces [83], they make this process clearer and simpler and thereby facilitate easier discussion and debate and can aid the perception of cyber-attacks with little reference to logical models [28].

- **Temporal methods**

Cyber kill chain

The recognition that attacks or intrusions by advanced cyber adversaries against organisations or missions are multistage, and occur over periods of months or years, has led to the development of multistage models which can be used to "trace" or characterize attack events. Such a multistage model is frequently referred to as a "cyber kill chain. An initial cyber kill chain model was developed by Lockheed Martin [12]. Cyber-attack lifecycle models are most commonly defined for external attacks on enterprise IT and command and control systems. Palo Alto Networks Accredited System Engineer (PSE) use a seven-phase cyber-attack lifecycle model, as illustrated in

Figure 3.

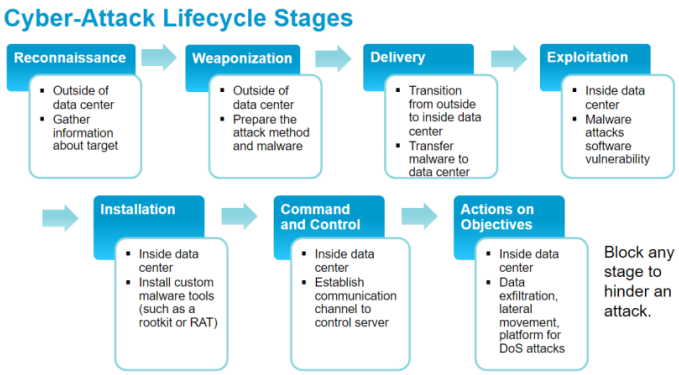


***Figure 3: Cyber-attack life cycle (Palo Alto)***

[12] Cloppert, M. "Security Intelligence: Attacking the Kill Chain," 14 October 2009.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

Variant attack lifecycles have been proposed. For example, an Advanced Research and Development Activity (ARDA) Workshop designed a version to characterize activities by insiders[84]: reconnaissance, access, entrenchment, exploitation, communication, manipulation, extraction & exfiltration, and counterintelligence. Dell SecureWorks identifies 12 stages: define target, find and organise accomplices, build or acquire tools, research target infrastructure/employees, test for detection, deployment, initial intrusion, outbound connection initiated, expand access and obtain credentials, strengthen foothold, exfiltrate data, and cover tracks and remain undetected[13]. Microsoft researchers have identified a set of ten "base types" of actions: reconnaissance, commencement, entry, foothold, lateral movement, acquire control, acquire target, implement / execute, conceal & maintain, and withdraw[85]. The CIS Community Attack Model defines nine stages: Initial Recon, Acquire / Develop Tools, Delivery, Initial Compromise, Misuse / Escalate Privilege, Internal Recon, Lateral Movement, Establish Persistence, and Execute Mission Objectives[14].

Diamond model

The Diamond model is one of the novel models for cyber intrusion analysis [86] where an adversary attacks a victim depending on two key motivations rather than using a series of steps like the kill chain or the attack graph. This model consists of four basic elements such as adversary, infrastructure, capability and victim. An adversary is an actor (or set of actors) who attacks a victim after analysing their capability against the victim. Initially the adversary starts with no knowledge of the capability of the victim. After analysing the capability of a victim, the adversary may find that he/she has more capability than the victimdoes, to attack or not. This model is important when dealing with more advanced attackers such as those who have already gained some control over the network. The adversary also analyses the infrastructure of his/her technical and logical ability to command and control any of victim's network.

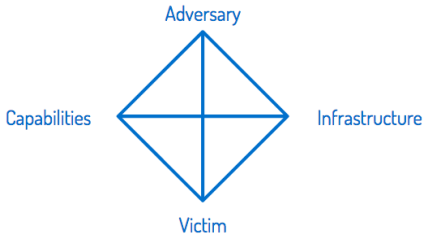


***Figure 4: Diamond model***

The diamond model is also associated with some meta- features such as timestamp, phases, result, directions, methodology and resources. In the event of an attack, the diamond model identifies phases in a timestamp. Components of the diamond model can be found in the Figure 4 which illustrates that the adversary looks for opportunity to attack a victim depending on the capability or the infrastructure.

- **Use case methods**

(Mis-) Use cases

A use case is a list of actions or event steps typically defining the interactions between a role (known in the Unified Modelling Language (UML) as an actor) and a system to achieve a goal. The actor can be a human or other external system. In systems engineering, use cases are used at a higher level, often representing

[13] SecureWorks. 2016. "Advanced Persistent Threats: Learn the ABCs of APTs - Part A," September 27, 2016
[14] CIS Critical Security Controls - The Center for Internet Security Community Attack Model

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

missions or stakeholder goals. Use case analysis is an important and valuable requirement analysis technique that has been widely used in modern software engineering since its formal introduction.

Misuse case [87] is a business process modelling method and derives from and is the inverse of use case. These methods highlight something that should not happen (i.e. a Negative Scenario) and the threats hence identified, help in defining new requirements, which are expressed as new Use Cases. While these approaches facilitate the design of functional and non-functional requirements (e.g. security requirements, platform requirements, etc.), their most important weakness is simplicity.

CORAS

CORAS [88] is a method for conducting security risk analysis. CORAS provides a customised language for threat and risk modelling and comes with detailed guidelines explaining how the language should be used to capture and model relevant information during the various stages of the security analysis. In this respect CORAS is model based. The Unified Modelling Language (UML) is typically used to model the target of the analysis. For documenting intermediate results, and for presenting the overall conclusions we use special CORAS diagrams which are inspired by UML. The CORAS method provides a computerised tool designed to support documenting, maintaining and reporting analysis results through risk modelling.

In the CORAS method a security risk analysis is conducted in eight steps (Figure 5):

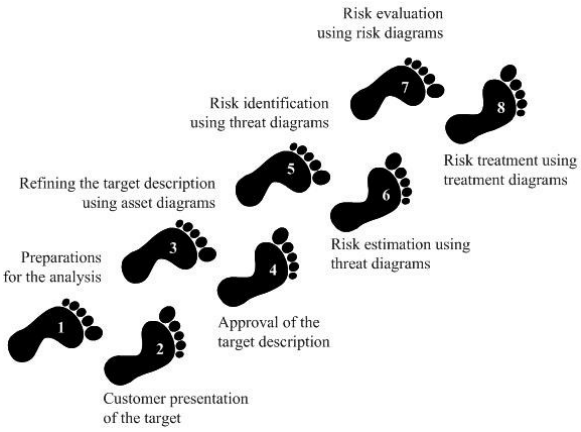


***Figure 5: CORAS method***

- **Cyber Prep Adversary Characterization Framework**

MITRE's Cyber Prep methodology [89] provides concepts, terminology, and characteristics that an organisation can use to articulate its risk frame for cyber risks – its assumptions about the cyber threat it faces and the potential consequences of greatest concern, the constraints on its cyber risk management decisions, its cyber risk tolerance, and its risk-related strategic trade-offs. Cyber Prep enables an organisation to characterize the class of threat it faces and its overall approach to cyber preparedness. This high-level characterization provides motivation for the organisation's cybersecurity strategy. The Cyber Prep framework defines fourteen aspects of organisational preparedness, in three areas: Governance, Operations, and Architecture & Engineering. Different adversary characteristics motivate different aspects of preparedness. Adversary characteristics include goals, scope or scale of operations, timeframe of operations, persistence, concern for stealth, stages of the cyber-attack lifecycle used, cyber effects sought or produced, and capabilities. In addition to the modelling constructs indicated in Figure 6, Cyber Prep identifies a representative set of high-level attack scenarios. The characteristics of an organisation make different scenarios more or less attractive to adversaries with different characteristics.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

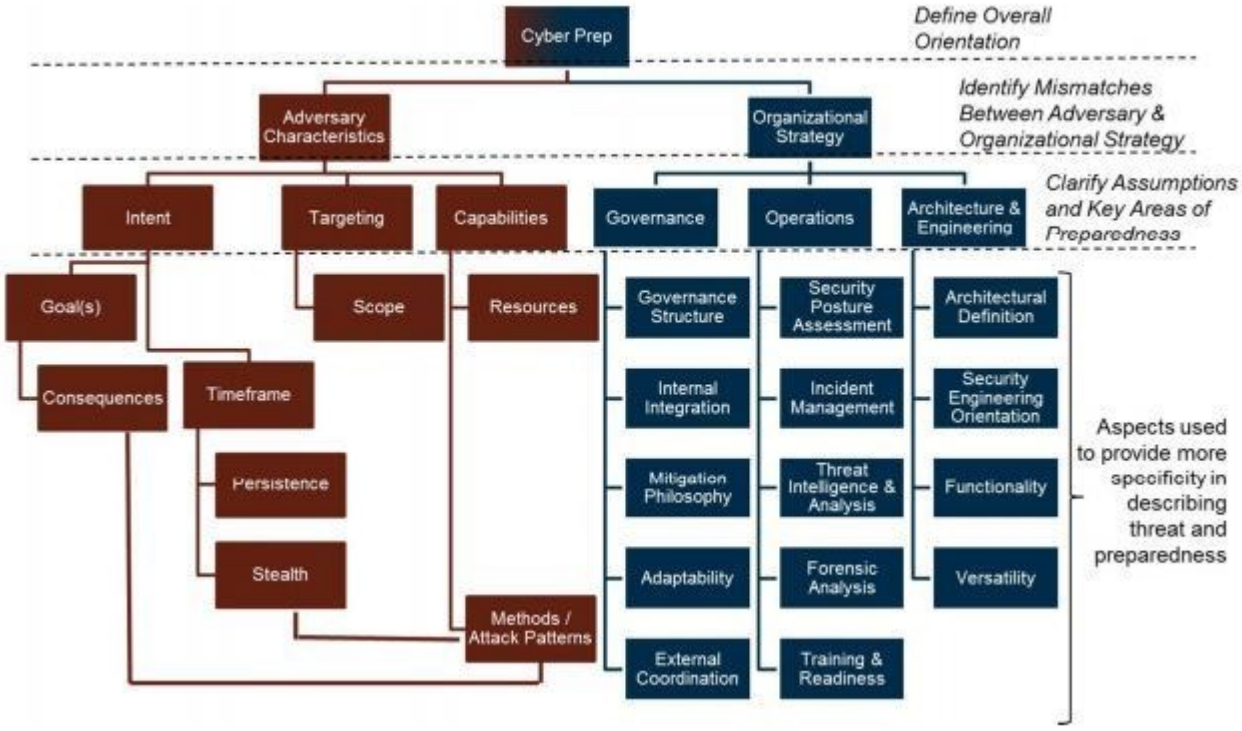


***Figure 6: Cyber Prep Framework in Detail***

- **Insider Threat Modelling**

Insider threat modelling includes models of insider behaviour intended to help identify indicators of insider activity [90]. Insider threat modelling also includes models intended to predict whether and how an insider could become malicious, and to analyse and predict the effects of organisational actions on insider behaviour. Such predictive analysis and modelling emphasize psycho-social factors [91]. Insider threat modelling overlaps with cyber threat modelling, insofar as insiders act in and on an organisation's cyber resources. However, there are areas in which the two forms of modelling are distinct:

- insider threat modelling considers external threat actors only with respect to their efforts to influence or suborn insiders and focuses on actions that an individual user can take.
- insider threat modelling can include purely non-cyber threat scenarios (e.g., theft).

#### *2.1.3.5 Threat Modelling to support Analysis*

Several highly structured threat modelling approaches have been developed to support decisions. These most established approaches are discussed below.

- **Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege (STRIDE)**

STRIDE was developed for internal use at Microsoft, as part of their push to produce more secure software. STRIDE mnemonically identifies six risk categories for assessed threats:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

- Spoofing [identity] -- identifying authentication threats
- Tampering [with data] -- identifying threats to data integrity
- Repudiation
- Information disclosure -- identifying data stewardship threats and data leaks
- Denial of service -- identifying threats to availability
- Elevation of privilege -- identifying authorization vulnerabilities

While sometimes referred to as a threat model or threat modelling framework, STRIDE serves primarily as a categorization of general types of threat vectors to be considered, helping analysts identify a complete threat model, for example using attack tree analysis [92]. STRIDE does not directly address level of detail or specific attack methods. Based on the findings, an analyst might conclude that there is an attack vector that needs to be mitigated in some other way (e.g. additional security component or policy change)

- **Damage, Reproducibility, Exploitability, Affected Users, and Discoverability (DREAD)**

DREAD was also created at Microsoft for use in their software development process to improve the security of their products. DREAD provides a scheme by which threat vectors identified using STRIDE or other methodologies are evaluated and prioritized. Scores for each element of the title are determined on a scale of 1 to 10. Each individual threat vector is scored on the five elements and an average taken, which can then be used to compare its severity and likelihood to those of other threat vectors. DREAD thus goes part of the way beyond threat modelling to risk assessment.

- **Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE)**

OCTAVE, although jointly developed by the U.S Computer Emergency Readiness Team (US-CERT) and Carnegie Mellon's Software Engineering Institute, is less of a methodology for assessing technological risk and more of a methodology for assessing organisational risk. The goal of OCTAVE/Allegro is to produce more robust risk assessment results without the need for extensive risk assessment knowledge by focusing on information assets in the context of how they are used, where they are processed /stored, and how they are exposed to threats, vulnerabilities, and disruptions. OCTAVE is more flexible. Probability analyses are "optional," the only requirement being thoroughness; analysis teams are directed to consider a variety of factors that can influence probability, as well as to explicitly determine the exact numerical thresholds for "high," "medium" and "low" probabilities.

The threat modelling portion of the OCTAVE/Allegro approach consists of identifying areas of concern (threats, threat sources, impacts on information assets) and developing threat scenarios, using threat trees. Key attributes of a threat in the OCTAVE/Allegro threat modelling approach include actor, asset, access or means, motive, and outcome (disclosure, modification, destruction, loss, or interruption).

- **Threat Agent Risk Assessment (TARA) and Threat Agent Library (TAL)**

Intel Corporation published, in December 2009, its Threat Agent Risk Assessment (TARA) methodology, which is designed to distil possible information security attacks into a digest of only those exposures most likely to occur. Its objective is to identify threat agents that are pursuing objectives which are reasonably attainable and could cause losses. The methodology identifies which threat agents pose the greatest risk, what they want to accomplish, and the likely methods they will employ. These methods are cross-referenced with existing vulnerabilities and controls to pinpoint the areas that are most exposed. The security strategy inherent in TARA then focuses on these areas to minimize efforts while maximizing effect. Intel also published a library of threat agents [Intel 2007] to serve as a starting point for enterprise development of an organisation-specific characterization of threat agents. The site at which the library white paper can be found was updated in 2015. The Threat Agent Library (TAL) defines 22 archetypes, using eight key attributes or parameters: intent, access, outcome, limits, resources, skill, objective, and visibility. Intel subsequently modified its list of key parameters

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

to include motivation. In addition, Intel identified 10 elements of the motivation parameter (ideology, coercion, notoriety, personal satisfaction, organisational gain, personal financial gain, disgruntlement, accidental, dominance, and unpredictable), and modified its model so that each agent can have multiple motivations (defining motivation, co-motivation, subordinate motivation, binding motivation, and personal motivation). The concept of multiple motivations has been carried into the definition of the Threat Actor Domain Object in Structured Threat Information eXpression (STIX™).

- **Visual, Agile & Simple Threat Modelling (VAST)**

The present VAST methodology came into light mainly to address the limitations and shortcomings of other threat methodologies. The principle of VAST methodology is the importance of scaling the threat modelling process across infrastructure and the systems development life cycle (SDLC) and achieving a seamless integration into an agile software development methodology. VAST aims to provide valuable and actionable insights to various involved parties including senior executives, developers, and security professionals.

- **Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK™)**

Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK™, [MITRE]) is a framework for describing the actions that an adversary may take while operating within an enterprise network. It provides a detailed characterization of adversary behaviour post-access, i.e., after initially gaining entry via a successful exploit. ATT&CK is intended to assist in prioritizing network defence by detailing the post-initial access that advanced persistent threat actors use to execute their objectives while operating inside a network. Ten tactics categories for ATT&CK were derived from the later stages (control, maintain, and execute) of the seven-stage Cyber Attack Lifecycle [MITRE] or the Cyber Kill Chain. Each category contains a listing of techniques that an adversary could use to perform that tactic, including technical description, indicators, useful defensive sensor data, detection analytics, and potential mitigations. Some techniques can be used for different purposes and therefore appear in more than one category. ATT&CK continues to be populated and updated as new techniques are reported.

- **Common Attack Pattern Enumeration and Classification (CAPEC™)**

CAPEC aims to provide a publicly available catalogue of attack patterns along with a comprehensive schema and classification taxonomy. Attack patterns are “descriptions of the common elements and techniques used in attacks against vulnerable cyber-enabled capabilities.” Each pattern defines a challenge that an attacker may face, provides a description of the common techniques used to meet the challenge, and presents recommended methods for mitigating an actual attack. Attack patterns help categorize attacks in a meaningful way to provide a coherent way of teaching designers and developers how their systems may be attacked and how they can effectively defend them.

- **Web Application Threat Models (WASC)**

WASC developed a classification of weaknesses in and threats against web applications. Its 34 classes of attacks include, among others, buffer overflow, cross-site scripting, and denial of service. Classes of weaknesses include improper input handling and abuse of functionality. At the time being, these classes are used in the Open Web Application Security Project (OWASP) WASC Web Hacking Incidents Database, which continues to be updated. The OWASP effort is reflected in the Process for Attack Simulation & Threat Analysis (PASTA) threat modelling methodology [93]. The OWASP Automated Threat Handbook currently describes 20 threat events. For each threat event, the following information is included: sectors targeted (e.g., financial, health), parties affected, data commonly misused, related threat events, description, other names and examples, CAPEC category, WASC threat identifiers, Common Weakness Enumeration (CWE) identifiers, OWASP attack category, possible symptoms, and suggested countermeasures.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

- **Process for Attack Simulation and Threat Analysis (PASTA)**

PASTA is a risk-centric threat-modelling framework developed in 2012. It contains seven stages, each with multiple activities, which are illustrated in Figure 7 below:


***Figure 7: Threat Modelling w/PASTA: Risk Centric Threat Modelling Case Studies***

PASTA aims to bring business objectives and technical requirements together. It uses a variety of design and elicitation tools in different stages. This method elevates the threat-modelling process to a strategic level by involving key decision makers and requiring security input from operations, governance, architecture, and development. Widely regarded as a risk-centric framework, PASTA employs an attacker-centric perspective to produce an asset-centric output in the form of threat enumeration and scoring.

- **Common Vulnerability Scoring System (CVSS)**

CVSS captures the principal characteristics of a vulnerability and produces a numerical severity score. CVSS was developed by NIST and is maintained by the Forum of Incident Response and Security Teams (FIRST) with support and contributions from the CVSS Special Interest Group. The CVSS provides users a common and standardized scoring system within different cyber and cyber-physical platforms. CVSS consists of three metric groups (Base, Temporal, and Environmental) with a set of metrics in each.

A CVSS score is derived from values assigned by an analyst for each metric. The metrics are explained extensively in the documentation. The CVSS method is often used in combination with other threat-modelling methods.

- **Trike**

Trike [94] was created as a security audit framework that uses threat modelling as a technique. It looks at threat modelling from a risk-management and defensive perspective. Trike starts with defining a system. The analyst builds a requirement model by enumerating and understanding the system's actors, assets, intended actions, and rules. This step creates an actor-asset-action matrix in which the columns represent assets and the rows represent actors. Each cell of the matrix is divided into four parts, one for each action of CRUD (creating, reading, updating, and deleting). In these cells, the analyst assigns one of three values: allowed action, disallowed action, or action with rules. A rule tree is attached to each cell.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

After defining requirements, a data flow diagram (DFD) is built. Each element is mapped to a selection of actors and assets. Iterating through the DFD, the analyst identifies threats, which fall into one of two categories: elevations of privilege or denials of service. Each discovered threat becomes a root node in an attack tree.

To assess the risk of attacks that may affect assets through CRUD, Trike uses a five-point scale for each action, based on its probability. Actors are rated on five-point scales for the risks they are assumed to present (lower number = higher risk) to the asset. Also, actors are evaluated on a three-dimensional scale (always, sometimes, never) for each action they may perform on each asset.

- **IDDIL/ATC**

  IDDIL/ATC is a mnemonic:

  - Identify the assets
  - Define the attack surface
  - Decompose the system
  - Identify attack vectors
  - List threat actors
  - Analysis & assessment
  - Triage
  - Controls

IDDIL/ATC methodology [95] provides a structured process for applying its cyber kill chain model, together with its variant of STRIDE (STRIDE-LM, which adds Lateral Movement), and attack trees. Key modelling constructs include assets, threat actors, and attack vectors. A threat profile (a tabular summary of threats, attacks, and related characteristics) identifies the asset or threat object; threat types; the attack surface; attack vectors; threat actors; the resultant condition; vulnerabilities, and controls.

- **MIL-STD-882E**

MIL-STD-882E, a Department of Defence methodology, uses two scales for rating risks: a four-point severity rating scale and a six-point likelihood scale. Unlike with other threat models, MIL-STD-882E goes ahead and specifically defines what each point on the scale means. MIL-STD-882E is designed to be applied throughout the life cycle of a system. MIL-STD-882E is considered too simplistic methodology for assessing malicious threats. Still, offers the benefits of being straightforward and well defined. Thus, MIL-STD-882E is a great "starter" methodology for those less experienced with threat-rating theory, while remaining useful for more experienced security analysts as well.

#### *2.1.3.6 Threat Modelling to Support Information Sharing*

Threat information sharing is a key aspect of many cyber risk management approaches and is needed to facilitate all stages of risk assessment. Key points of STIX, OpenDXL and PRE-ATT&CK efforts are included in this report.

- **Structured Threat Information eXpression (STIX)**

STIX is a language and serialization format used to exchange cyber threat intelligence [96]. STIX enables organizations to share CTI with one another in a consistent and machine-readable manner, allowing security communities to better understand what computer-based attacks they are most likely to see and to anticipate and/or respond to those attacks faster and more effectively. The STIX domain model defines data structures to characterize or describe an adversary and adversary activities. STIX Domain Objects include Threat Actor; Malware; Tools; Attack Pattern, Vulnerability and Intrusion Set. The Threat Actor object has several optional associated properties, including goals, sophistication, resource level, primary motivation, secondary

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

motivations, and personal motivations. Attack patterns, malware, and tools are all forms of TTPs. Information about adversary reasons for acting and how they organise themselves is described via the threat actor, intrusion set, and campaign domain objects.

- **Open Data Exchange Layer (OpenDXL)**

Open Cybersecurity Alliance (OCA) announced the availability of OpenDXL Ontology, the first open-source language for connecting cybersecurity tools through a common messaging framework. OpenDXL Ontology enables any tool to automatically gain the ability to communicate and interoperate with all other technologies using this language.

OCA was launched in October 2019 to connect the fragmented cybersecurity landscape with common, open-source code and practices. Governed under the auspices of OASIS, the OCA now includes more than 25 member organisations and has brought two major interoperability projects into the open-source realm, with OpenDXL Ontology (contributed by McAfee) and STIX Shifter (contributed by IBM Security) now available for cross-industry collaboration and development.

The Open Data Exchange Layer (OpenDXL) is an open messaging framework to develop and share integrations with other tools. The release of the OpenDXL Ontology offers a single, common language for these notifications, information, and actions across security products that any vendor can adopt in order to communicate in a standard way with all other tools under this umbrella. This provides companies with a set of tooling that can be applied once and automatically reused everywhere across all product categories, while also eliminating the need to update integrations as product versions and functionalities change.

- **PRE-ATT&CK**

MITRE PRE-ATT&CK™ is an emerging framework for categorizing and characterizing adversary activities in the early stages of the cyber-attack lifecycle. Seventeen categories of high-level tactics are currently defined, primarily covering techniques external to the enterprise. Tactics can be technical, human, or organisational; examples include People Information Gathering, Adversary OPSEC (Operations Security), Persona Development, and Test Capabilities.

PRE-ATT&CK can be used by cyber defenders to prioritize cyber threat intelligence of data acquisition and analysis. Pre-exploit adversary activities, such as gathering information from the Internet about potential targets of attack, are largely executed outside of a potential victim's purview, making it significantly more difficult for defenders to detect.

However, PRE-ATT&CK could provide a common lexicon to allow cyber defence to understand, detect, mitigate, and share information about adversary activities across the FSS. This could then be used to shift to a more proactive/predictive analytic capability to support elements of attribution and defensive responses.

## 2.2 Analysis

Each theoretical perspective has its strengths and weaknesses, particularly in their application to the cyber SA. Much of the research on cyber SA has primarily taken an algorithmic perspective, focusing mostly on the automation and the development of new defensive tools for protection, detection, and response[97]. Examples of this work include data visualizations[98], data fusion methods for tracking cyber-attacks [99][100], identification of internal and external threats using intelligent agents [101][102], and the use of probabilistic models to assess network vulnerabilities [103]. Although valuable, this body of work overlooks perhaps the most crucial component of cyber defence analysis: the human component [104]. These approaches paid little attention to how operators perform with existing technologies let alone whether or not these new technologies actually improve SA in human operators.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

## 2.3 Gaps and Barriers

- Existing approaches to gain cyber situation-awareness focus mostly on vulnerability analysis, intrusion detection and alert correlation, attack trend analysis, causality analysis and forensics, information flow analysis, damage assessment, and intrusion response. These approaches however only work at the lower (abstraction) levels. Higher level situation-awareness analyses are still done manually by a human analyst, which makes it labour-intensive, time-consuming, and error-prone. Strong, theory-based approach to measure and analyse human situation awareness in cyberspace would address a critical gap. There are several remaining difficulties.
- There is still a big gap between human analysts' mental model and the capabilities of existing cyber situation-awareness tools. SA can increase system resilience by giving advance warning of network problems or analytic slowdowns through displays.
- Existing approaches need to handle uncertainty better. Uncertainty in perceived data could lead to distorted situation awareness.
- Lack of data or incomplete knowledge may raise additional uncertainty management issues. Apart from the facts of a situation, actionable information, indications, and warnings in the form of intelligent, dynamic, multimedia components are needed in ongoing monitoring and response to a threat.
- Existing approaches lack the reasoning and learning capabilities required to achieve full situation awareness. Agents can implement capabilities for making connections by continuously pursuing knowledge discovery (looking for relationships between all types of available information). Moreover, agents can augment human pattern recognition by learning new threat patterns and presenting them to the analyst for validation.
- All aspects of cyber situation awareness have been treated in most cases as separate problems, but full cyber situation awareness requires all these aspects to be integrated into one solution. Such a solution is still missing.

The promising area of distributed situation awareness cannot be implemented easily. Different agents can enhance system scalability by adapting to highly distributed architectures.

## 2.4 Standards, Platforms and Methodologies

The internal processes of cybersecurity situational awareness have been presented in Section 2.1.1 and 2.1.2. As described, the cybersecurity situational awareness requires several inputs to enhance the ability to understand the environment and to be able to predict potential security issues.

**Methodologies:** The process of cybersecurity situational awareness involves three key areas that include network components, threat information and mission dependencies. Therefore, cybersecurity situational awareness has three different dimensions [105]:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

1. **Network Awareness** [106]**:** Includes vulnerability auditing, port scanning, anomaly detection and data deriving from the IDs, patch management and any other information of the status of elements that compromise the network.
2. **Threat Awareness:** Consisting of information regarding possible attack vectors including current and past attacks based on existing exploit repositories like common vulnerabilities and exposures (CVEs). Furthermore, this key area regards the identification and tracking of internal incidents and suspicious behaviour and the incorporation of knowledge related to external threats. Threat sharing technologies and procedures are also included on this specific aspect.
3. **Mission/Operation Awareness:** The scope of this specific key area is to develop a comprehensive picture of the critical dependencies. To identify how decreased or degraded network operations will affect the mission of the network is the main outcome of this aspect. Mission awareness is very important and includes the perception and summation of the required activities which ensure that appointed tasks are performed in accordance with the intended purpose or plan [107].

The combination of the various inputs (Figure 8) determine the risks and the cybersecurity situational awareness process is to provide all the incoming information together in a clear and meaningful way. The main goal for cybersecurity situational awareness is to enable us to stay ahead of the potential threats, to enhance security operations, to provide real-time alerting on observed risks and threats associated with assets of interest (network awareness).

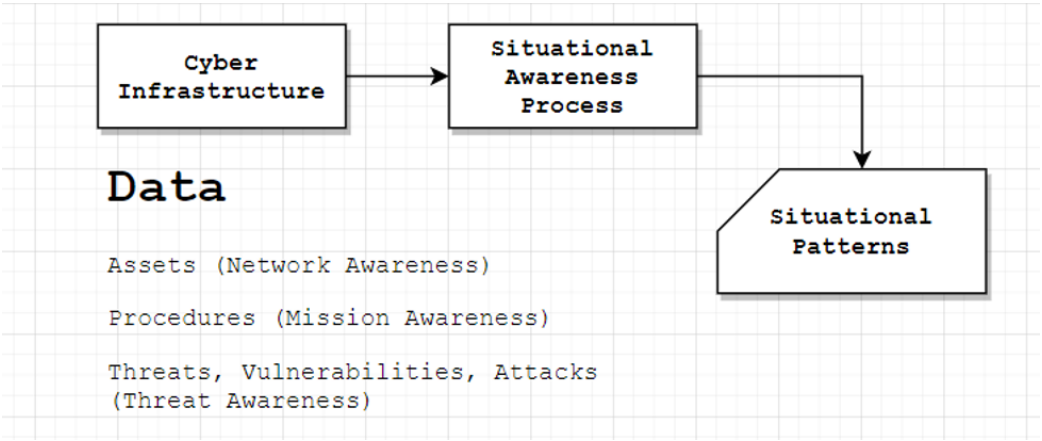


***Figure 8. Cybersecurity situational awareness process***

Other approaches highlight more details to be considered for situational awareness such as [108]: Vulnerability Management, Patch Management, Event Management, Incident Management, Malware Detection, Asset Management, Configuration Management, Network Management, License Management, Information Management, Software Assurance

The above 11 domains include different data which are analysed and forwarded to a diverse set of security tools to perform scoring and to provide the overall situational awareness [108]. However, the option to summarize the required process using the three key areas seems more clear and coherent approach. For example, Park et al., 2019 [109] mention and use these key areas as the foreground to provide a situational awareness model focused on the android environment and connectivity of IoT devices.

MITRE[15] has also released a document presenting an overview for their cyber situational awareness solutions[16] MITRE provides a range of technical solutions that could be used for any toolkit for providing efforts and data towards the cybersecurity situational awareness (Figure 9).

[15] mitre.org
[16] mitre.org/publications/technical-papers/an-overview-of-mitre-cyber-situational-awareness-solutions

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

| | |
|---|---|
| **Threat Analysis** | CRITs<br>ATT&CK™<br>STIX™, TAXII™ |
| **Dependency & Impact Analysis** | CyCS<br>CJA |
| **Analysis of Alternatives (AoA)** | TARA |
| **Emerging Solutions** | FACT<br>CyGraph<br>AMICA |

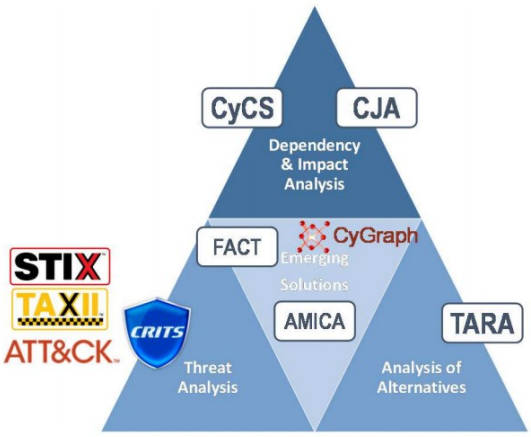


***Figure 9. MITRE efforts by cyber defence situational awareness (CDSA)***

Within Cyber Threat Intelligence area, MITRE has developed a Tactics, Techniques and Procedures (TTP) knowledge base, based on real-world observations, called Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK)[17]. It supports the end user to understand how an attack might happen, i.e., to understand the tools and patterns of required actions associated to a specific threat. Using that information, the decision-making of an organization can be improved. Related specifications to ATT&CK initiative are[18]:

(i) Common Attack Pattern Enumeration and Classification (CAPEC) to provide a publicly available catalogue of common attack patterns classified in an intuitive manner

(ii) Malware Attribute Enumeration and Characterization (MAEC™) to define and develop a standardized language for sharing structured information about malware.

MITRE has developed another innovative system for real-time situational awareness, improving network security posture and focusing on protection of mission-critical assets [110]. It offers a decision support system based on an ongoing overall status of the security.

Besides MITRE, NIST has released a document for facilitating continuous monitoring by presenting a reference model, known as Continuous Asset Evaluation, Situational Awareness, and Risk Scoring (CAESARS) Framework Extension (FE), to provide an overall situational awareness [108]. In order to provide such functions, it is important to consider and to rapidly identify and discover new and emerging cyber threats (threat awareness) and proceed to the required actions accordingly (mission awareness). The CAESARS FE promotes the use of the security automation standards such as standards-based methods for performing data collection and Security Content Automation Protocol (SCAP).

The situational awareness approach, driven by security continuous monitoring is also described by other main stakeholders in the cybersecurity arena as the Software Engineering Institute (SEI)[19] or by Gartner, which is a main global expert consultant for breaking-through tools in IT in general and cybersecurity in particular. As NIST defined CAESARS FE, Gartner defined a strategic approach for "continuous adaptive risk and trust assessment", also called CARTA[20]. This approach is based on the concept that the security is adaptive, everywhere and all the time. It requires that risk and trust need to be continuously monitored and assessed, therefore enabling actionable security decisions,

[17] https://attack.mitre.org/
[18] https://attack.mitre.org/resources/related-projects/
[19] https://insights.sei.cmu.edu/sei_blog/2020/05/situational-awareness-for-cyber-security-architecture-tools-for-monitoring-and-response.html
[20] https://www.gartner.com/teamsiteanalytics/servePDF?g=/imagesrv/media-products/pdf/Forcepoint/Forcepoint-1-4YCDU8P.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

NIST defines Information Security Continuous Monitoring (ISCM) as maintaining ongoing awareness of information security, vulnerabilities, and threats to support organizational risk management decisions [111]. Specifically, NIST produced a report on status of international cybersecurity standardization for the Internet of Things (IoT) [112] which includes a list of security approved, draft Security Automation and Continuous Monitoring (SACM) standards on the IoT field. Moreover, it also contains standards related to Information Security Management Systems (ISMS), which cover standardized processes and corresponding security controls to establish a governance, risk, and compliance structure for information security of an organization. ISMS standards include the well-known ISO/IEC 27001:2013 that sets the requirements for establishing, implementing, maintaining, and continually improving an information security management system.

With regards to Cyber Incident Management, NISTIR 8200 [112] released existing standards which include ISO/IEC 27035:2016 and ISO/IEC 27035-2:2016 that provide guidelines for information security incident management, including detection, assessment and reaction to incidents. Furthermore, OASIS Structured Threat Information Expression (STIX) and OASIS Trusted Automated Exchange of Indicator Information (TAXII) are approved specifications for cyber threat information sharing.

Tools and platforms that support ISCM together with ISMS are the key to have an operational situational awareness framework. ISO/IEC 27035 identifies multiple technologies related to detection of security vulnerabilities and events, collection of information on the events and vulnerabilities detected, and reporting on the events and vulnerabilities [113]. Mentioned technologies include: Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), honeypots, log monitoring systems, security information, event management systems and network monitoring systems.

As defined by Gartner[21], a Security Information and Event Management (SIEM) technology provides a unified platform for the following:

i) threat detection
ii) event collection, correlation, and visualization
iii) security incident management.

The aforementioned technologies are usually part of a SIEM system as security event data sources or sensors. SIEM technology is one the most widely deployed solution to provide situational awareness in an organization[19].

Because of the importance of SIEM technology, Gartner annually publishes the technical and market analysis of information and event management solutions [114]. Splunk and IBM SIEM solutions are considered as market and technical leaders. Splunk includes unsupervised Machine Learning-driven User Behaviour Analytics (UBA) capabilities, Security Orchestration Automation and Response (SOAR) capabilities, enhanced real-time monitoring, the ability to implement security automation with threat intelligence, and healthcare-specific vertical content to address prescription theft and patient privacy violations. IBM SIEM solution covers multiple capabilities such as: vulnerability management, risk management including threat simulation, UBA and SOAR capabilities, incident forensics support and advanced-analytics-based root cause identification engine. Other well positioned SIEM solutions are Exabeam, Securonix, Rapid7 and LogRhythm. Exabeam and Securonix have included solutions based on MITRE ATT&CK framework besides UEBA and SOAR capabilities.

Even though SIEM technology covers wide range aspects for a successful situational awareness in an organization, there are other complementary technologies such as cyber incident information sharing solutions. Cyber incident information sharing enables the improvement of organisations' cyber threat intelligence in general, and aids organisations in the detection and prevention of incidents and attacks. Information sharing practices and supporting tool platforms are an essential part of continuous improvement in cyber security frameworks and keeping organisations up-to-date in their cybersecurity strategy as they serve to enrich

[21] https://www.gartner.com/en/information-technology/glossary/security-information-and-event-management-siem

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

organisation's knowledge about cyberattacks, understand the techniques used by the attackers, learn on actual occurrence of incidents and potential solutions, etc.

Information sharing platforms support the core activity of Information Sharing and Analysis Centres all around the world. They are non-profit organisations that provide resources for centralising information on cyber threats to critical industries and infrastructures, helping this way in domain-specific Critical Infrastructure Protection (CIP). In the healthcare domain, the H-ISAC [22] in the USA is the main community of critical infrastructure owners and operators within the Health Care and Public Health sector (HPH). There is no such organisation at the European level.

The major exponent of open source Information sharing platforms, currently available in the market is the MISP platform, MISP - Open Source Threat Intelligence Platform & Open Standards for Threat Information Sharing[23]. This solution is a software platform for sharing, storing, and correlating multi-attribute data around cyber incidents, including not only malware information but also other Indicators of Compromise (IoC) of targeted attacks. The platform was initiated in 2011 and NATO took charge of it in 2012. Later, CERT-EU, CIRCL and many other organisations started to adopt the software and promote it among CERTs worldwide. Today, more than six thousand organisations worldwide are using MISP and there exist multiple public and private MISP communities to which organisations can adhere and benefit from cyber intelligence shared therein.

Other platforms include Cyware Threat Intelligence eXchange[24], which is defined as a smart, client-server threat intelligence platform (TIP) for ingestion, enrichment, analysis, and bi-directional sharing of threat data within your trusted network. Another example is the FS-ISAC intelligence exchange platform[25] recently launched by the Financial Services Information Sharing and Analysis Centre (FS-ISAC) for cyber information sharing in the banking sector and mostly in Asia Pacific.

[22] https://h-isac.org/
[23] https://www.misp-project.org/
[24] https://cyware.com/ctix-stix-taxii-cyber-threat-intelligence-exchange
[25] https://www.theedgesingapore.com/news/cybersecurity/fs-isac-launches-intelligence-exchange-help-financial-services-tackle-cyber

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# 3 Conclusion

This deliverable, has presented an in-depth literature review on the Cybersecurity Situational Awareness paradigm, focusing on the existing theoretical background, algorithms, methods, frameworks, and standards. The research revealed several gaps and barriers towards providing an automated, seamless, secure, and precise SA framework. Moreover, this research revealed several state-of-the-art platforms, tools and methodologies concerning SA. These results will be well-received and utilised by Task T3.3.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

# Annex I: **References**


[1] C. D. Wickens, "Situation awareness: Review of mica Endsley's 1995 articles on situation awareness theory and measurement," *Human Factors*, vol. 50, no. 3. SAGE PublicationsSage CA: Los Angeles, CA, pp. 397–403, 01-Jun-2008.

[2] A. AlEroud and G. Karabatis, "A framework for contextual information fusion to detect cyber-attacks," in *Studies in Computational Intelligence*, vol. 691, Springer Verlag, 2017, pp. 17–51.

[3] D. M. Best, S. Bohn, D. Love, A. Wynne, and W. A. Pike, "Real-time visualization of network behaviors for situational awareness," in *ACM International Conference Proceeding Series*, 2010, pp. 79–90.

[4] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, "LiveRAC: Interactive visual exploration of system management time-series data," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1483–1492, 2008.

[5] C. Zhong, J. Yen, P. Liu, R. F. Erbacher, C. Garneau, and B. Chen, "Studying analysts' data triage operations in cyber defense situational analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10030, Springer Verlag, 2017, pp. 128–169.

[6] A. D'Amico and K. Whitley, "The Real Work of Computer Network Defense Analysts," in *VizSEC 2007*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 19–37.

[7] R. Rtoty and R. Erbacher, "A Survey of Visualization Tools Assessed for Anomaly-Based Intrusion Detection Analysis," no. April, p. 50, 2014.

[8] W. Yu, G. Xu, Z. Chen, and P. Moulema, "A cloud computing based architecture for cyber security situation awareness," *2013 IEEE Conf. Commun. Netw. Secur. CNS 2013*, pp. 488–492, 2013.

[9] "Automated Reasoning over Provenance-Aware Communication Network Knowledge in Support of Cyber-Situational Awareness | SpringerLink." [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-319-99247-1_12. [Accessed: 13-Nov-2019].

[10] L. F. Sikos, *Mastering Structured Data on the Semantic Web*. Berkeley, CA: Apress, 2015.

[11] S. Voigt, C. Howard, D. Philp, and C. Penny, *Proceedings of the 5th International Workshop on Graph Structures for Knowledge Representation and Reasoning {(GKR2017)}: Revised Selected Papers, Melbourne, Australia, August 21, 2017*, vol. 10775. Springer International Publishing, 2018.

[12] M. Mathews, P. Halvorsen, A. Joshi, and T. Finin, "A Collaborative Approach to Situational Awareness for CyberSecurity," in *Proceedings of the 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2012.

[13] J. Undercoffer, A. Joshi, T. Finin, and J. Pinkston, "Using DAML + OIL to classify intrusive behaviours," *Knowl. Eng. Rev.*, vol. 18, no. 3, pp. 221–241, 2003.

[14] J. Undercoffer, A. Joshi, and J. Pinkston, "Modeling computer attacks: An ontology for intrusion detection," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2820, pp. 113–135, 2003.

[15] Y. Gao and S. Zhang, "A Network Security Situation Awareness Method Based on Multi-source Information Fusion," vol. 130, no. Ifmeita 2017, pp. 273–276, 2018.

[16] M. Angelini and G. Santucci, "Cyber situational awareness: from geographical alerts to high-level management," *J. Vis.*, vol. 20, no. 3, pp. 453–459, 2017.

[17] L. M. and A. J. Zareen Syed, Ankur Padia, Tim Finin, "UCO-AUnifiedCybersecurityOntology," *Assoc. Adv. Artif. Intell.*, no. Figure 1, pp. 195–202, 2016.

[18] K. M. T. Huffer and J. W. Reed, "Situational awareness of network system roles (SANSR)," *ACM Int. Conf. Proceeding Ser.*, pp. 3–6, 2017.


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

[19] R. Graf and R. King, "Neural network and blockchain based technique for cyber threat intelligence and situational awareness," *Int. Conf. Cyber Conflict, CYCON*, vol. 2018-May, pp. 409–425, 2018.

[20] S. Mathew, S. Upadhyaya, M. Sudit, and A. Stotz, "Situation Awareness of multistage cyber attacks by semantic event fusion," *Proc. - IEEE Mil. Commun. Conf. MILCOM*, pp. 1286–1291, 2010.

[21] G. Settanni, Y. Shovgenya, F. Skopik, R. Graf, M. Wurzenberger, and R. Fiedler, "Correlating cyber incident information to establish situational awareness in Critical Infrastructures," *2016 14th Annu. Conf. Privacy, Secur. Trust. PST 2016*, pp. 78–81, 2016.

[22] E. Allison Newcomb, R. J. Hammell, and S. Hutchinson, "Effective prioritization of network intrusion alerts to enhance situational awareness," *IEEE Int. Conf. Intell. Secur. Informatics Cybersecurity Big Data, ISI 2016*, pp. 73–78, 2016.

[23] E. A. Newcomb and R. J. Hammell, "A fuzzy Logic Utility Framework (FLUF) to support information assurance," *Stud. Comput. Intell.*, vol. 654, pp. 33–48, 2016.

[24] T. P. Hanratty, E. Allison Newcomb, R. J. Hammell, J. T. Richardson, and M. R. Mittrick, "A fuzzy-based approach to support decision making in complex military environments," *Int. J. Intell. Inf. Technol.*, vol. 12, no. 1, pp. 1–30, 2016.

[25] W. J. Matuszak, L. DiPippo, and Y. L. Sun, "CyberSAVe - Situational awareness visualization for cyber security of smart grid systems," *ACM Int. Conf. Proceeding Ser.*, pp. 25–32, 2013.

[26] V. Lenders, A. Tanner, and A. Blarer, "Gaining an edge in cyberspace with advanced situational awareness," *IEEE Secur. Priv.*, vol. 13, no. 2, pp. 65–74, 2015.

[27] R. Vinayakumar, K. P. Soman, P. Poornachandran, V. S. Mohan, and A. D. Kumar, "ScaleNet: Scalable and hybrid framework for cyber threat situational awareness based on DNS, URL, and email data analysis," *J. Cyber Secur. Mobil.*, vol. 8, no. 2, pp. 189–240, 2018.

[28] "REmatch: High-performance Regular Expression Matching for Network Security Petabi Inc,."

[29] H. K. Park, M. S. Kim, M. Park, and K. Lee, "Cyber situational awareness enhancement with regular expressions and an evaluation methodology," in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, 2017, pp. 406–411.

[30] F. He, Y. Zhang, H. Liu, and W. Zhou, "SCPN-based game model for security situational awareness in the intenet of things," *2018 IEEE Conf. Commun. Netw. Secur. CNS 2018*, pp. 1–5, 2018.

[31] H. Zhang *et al.*, "Towards an integrated defense system for cyber security situation awareness experiment," *Sensors Syst. Sp. Appl. VIII*, vol. 9469, p. 946908, 2015.

[32] A. Zheng and A. Casari, *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.," 2018.

[33] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[34] D. R. Cox, "The regression analysis of binary sequences," *J. R. Stat. Soc. Ser. B*, vol. 20, no. 2, pp. 215–232, 1958.

[35] M. A. Efroymson, "Multiple regression analysis," *Math. methods Digit. Comput.*, pp. 191–203, 1960.

[36] B. D. Craven and S. M. N. Islam, *Ordinary least-squares regression*. Sage Publications, 2011.

[37] J. H. Friedman, "Multivariate adaptive regression splines," *Ann. Stat.*, pp. 1–67, 1991.

[38] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J. Am. Stat. Assoc.*, vol. 74, no. 368, pp. 829–836, 1979.

[39] K. Pearson, " LIII. On lines and planes of closest fit to systems of points in space ," *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.

[40] H. Wold, "Partial Least Squares," *Encyclopedia of Statistical Sciences*. Aug-2006.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

[41] kue tradisional khas Aceh, “The use of multiple measures in taxonomic problems,” pp. 1–9, 1954.

[42] S. P. Lloyd, “Least Squares Quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[43] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[44] A. P. Dempster, N. M. Laird, and D. B. Rubin, “ Maximum Likelihood from Incomplete Data Via the EM Algorithm ,” *J. R. Stat. Soc. Ser. B*, vol. 39, no. 1, pp. 1–22, 1977.

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[46] R. Salakhutdinov and G. Hinton, “Deep Boltzmann machines,” *J. Mach. Learn. Res.*, vol. 5, no. 3, pp. 448–455, 2009.

[47] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, “Stacked denoising autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion,” *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.

[48] “19.”

[49] “20.”

[50] E. Darra and S. K. Katsikas, “A survey of intrusion detection systems in wireless sensor networks,” *Intrusion Detect. Prev. Mob. Ecosyst.*, pp. 393–458, 2017.

[51] A. Alazab, M. Hobbs, J. Abawajy, and M. Alazab, “Using feature selection for intrusion detection system,” *2012 Int. Symp. Commun. Inf. Technol. Isc. 2012*, pp. 296–301, 2012.

[52] S. Dua and X. Du, “Data Mining and Machine Learning in Cybersecurity,” *Data Min. Mach. Learn. Cybersecurity*, 2016.

[53] N. Kshetri and J. Voas, “Hacking Power,” no. December, pp. 91–95, 2017.

[54] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, “IoT Security Techniques Based on Machine Learning,” pp. 1–20, 2018.

[55] S. Chebrolu, A. Abraham, and J. P. Thomas, “Feature deduction and ensemble design of intrusion detection systems,” *Comput. Secur.*, vol. 24, no. 4, pp. 295–307, 2005.

[56] Khraisat, Ansam, Gondal, Iqbal, and Vamplew, Peter, *An Anomaly Intrusion Detection System Using C5 Decision Tree Classifier*, vol. 1. Springer International Publishing, 2018.

[57] K. Bajaj and A. Arora, “Improving the Intrusion Detection using Discriminative Machine Learning Approach and Improve the Time Complexity by Data Mining Feature Selection Methods,” *Int. J. Comput. Appl.*, vol. 76, no. 1, pp. 5–11, 2013.

[58] S. Elhag, A. Fernández, A. Bawakid, S. Alshomrani, and F. Herrera, “On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems,” *Expert Syst. Appl.*, vol. 42, no. 1, pp. 193–202, 2015.

[59] S. Thaseen and C. A. Kumar, “An analysis of supervised tree based classifiers for intrusion detection system,” *Proc. 2013 Int. Conf. Pattern Recognition, Informatics Mob. Eng. PRIME 2013*, pp. 294–299, 2013.

[60] S. Subramanian, V. B. Srinivasan, and C. Ramasa, “Study on classification algorithms for network intrusion systems,” *J. Commun. Comput.*, vol. 9, no. 11, pp. 1242–1246, 2012.

[61] X. Yang and Y. Tian, “EigenJoints-based Action Recognition Using Naïve-Bayes-Nearest-Neighbor,” pp. 14–19, 2012.

[62] L. Koc, T. A. Mazzuchi, and S. Sarkani, “A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier,” *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13492–13500, 2012.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

[63] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 424–430, 2012.

[64] W. C. Lin, S. W. Ke, and C. F. Tsai, "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors," *Knowledge-Based Syst.*, vol. 78, no. 1, pp. 13–21, 2015.

[65] G. Wang, J. Hao, J. Mab, and L. Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering," *Expert Syst. Appl.*, vol. 37, no. 9, pp. 6225–6232, 2010.

[66] C. Annachhatre, T. H. Austin, and M. Stamp, "Hidden Markov models for malware classification," *J. Comput. Virol. Hacking Tech.*, vol. 11, no. 2, pp. 59–73, 2015.

[67] S. N. Murray, B. P. Walsh, D. Kelliher, and D. T. J. O'Sullivan, "Multi-variable optimization of thermal energy efficiency retrofitting of buildings using static modelling and genetic algorithms - A case study," *Build. Environ.*, vol. 75, pp. 98–107, 2014.

[68] M. Sazzadul Hoque, "An Implementation of Intrusion Detection System Using Genetic Algorithm," *Int. J. Netw. Secur. Its Appl.*, vol. 4, no. 2, pp. 109–120, 2012.

[69] M. A. Jabbar, R. Aluvalu, and S. S. Reddy, "RFAODE: A Novel Ensemble Intrusion Detection System," *Procedia Comput. Sci.*, vol. 115, pp. 226–234, 2017.

[70] S. Potluri, S. Ahmed, and C. Diedrich, *Convolutional Neural Networks for Multi-class Intrusion Detection System*, no. December. Springer International Publishing, 2017.

[71] B. Zhang, Y. Yu, and J. Li, "Network intrusion detection based on stacked sparse autoencoder and binary tree ensemble method," *2018 IEEE Int. Conf. Commun. Work. ICC Work. 2018 - Proc.*, no. 61702046, pp. 1–6, 2018.

[72] H. Zhang, X. Yu, P. Ren, C. Luo, and G. Min, "Deep Adversarial Learning in Intrusion Detection: A Data Augmentation Enhanced Framework," pp. 1–10, 2019.

[73] M. Gerber and R. von Solms, "Management of risk in the information age," *Comput. Secur.*, vol. 24, no. 1, pp. 16–30, Feb. 2005.

[74] B. Potteiger, G. Martins, and X. Koutsoukos, "Software and attack centric integrated threat modeling for quantitative risk assessment," in *Proceedings of the Symposium and Bootcamp on the Science of Security - HotSos '16*, 2016, pp. 99–108.

[75] H. S. Lallie, K. Debattista, and J. Bal, "A review of attack graph and attack tree visual syntax in cyber security," *Computer Science Review*, vol. 35. Elsevier Ireland Ltd, p. 100219, 01-Feb-2020.

[76] D. Staheli *et al.*, "Visualization evaluation for cyber security: Trends and future directions," in *ACM International Conference Proceeding Series*, 2014, vol. 10-Novembe, pp. 49–56.

[77] T. Keller and S. O. Tergan, "Visualizing knowledge and information: An introduction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005, vol. 3426 LNCS, pp. 1–23.

[78] J. Homer, A. Varikuti, X. Ou, and M. A. McQueen, "Improving attack graph visualization through data reduction and attack grouping," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, vol. 5210 LNCS, pp. 68–79.

[79] G. Dondossola and L. Pietre-Cambacedes, "Modelling of cyber attacks for assessing smart grid security," in *Cigre Study Committee D2 Colloquium. Buenos Aires, Argentina. 19th - 20th October 2011*, 2011.

[80] S. Roschke, F. Cheng, and C. Meinel, "A new alert correlation algorithm based on attack graph," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6694 LNCS, pp. 58–67.

[81] I. Hogganvik and K. Stølen, "Investigating Preferences in Graphical Risk Modeling," 2016.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

[82] D. Schweitzer and W. Brown, "Using visualization to teach security," *J. Comput. Sci. Coll.*, vol. 24, no. 5, pp. 143–150, 2009.

[83] G. A. Fink, C. L. North, A. Endert, and S. Rose, "Visualizing cyber security: Usable workspaces," in *6th International Workshop on Visualization for Cyber Security 2009, VizSec 2009 - Proceedings*, 2009, pp. 45–56.

[84] M. Maybury *et al.*, "Analysis and detection of malicious insiders," *Int. Conf. Initelligence Anal. McLean, VA*, no. May 2014, pp. 3–8, 2005.

[85] J. Espenschied and A. Gunn, "Threat Genomics," in *Metricon*, 2012.

[86] S. Caltagirone, A. Pendergast, and C. Betz, "The Diamond Model of Intrusion Analysis," *Threat Connect*, vol. 298, no. 0704, pp. 1–61, 2013.

[87] G. Sindre and A. Opdahl, "Capturing security requirements through misuse cases," *NIK 2001, Nor. Inform. 2001*, p. 12, 2001.

[88] "Lund, Mass Soldal, Bjørnar Solhaug, and Ketil Stølen. Model-driven risk analysis: the CORAS approach. Springer, 2010."

[89] D. Bodeau and R. Graubart, "Motivating Organizational Cyber Strategies in Terms of Preparedness," *Mitre Corp.*, vol. Case Numbe, pp. 1–82, 2017.

[90] D. L. Costa, M. J. Albrethsen, M. L. Collins, S. J. Perl, G. J. Silowash, and D. L. Spooner, "An Insider Threat Indicator Ontology," 2016.

[91] Greitzer, Kangas, Noonan, Brown, and Ferryman, "Psychosocial Modeling of Insider Threat Risk Based on Behavioral and Word Use Analysis," *e-Service J.*, vol. 9, no. 1, p. 106, 2013.

[92] T. Xin and B. Xiaofang, "Online banking security analysis based on STRIDE threat model," *Int. J. Secur. its Appl.*, vol. 8, no. 2, pp. 271–282, 2014.

[93] T. Ucedavélez and M. M. Morana, *Risk Centric Threat Modeling: Process for Attack Simulation and Threat Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2015.

[94] P. Saitta, B. Larcom, and M. Eddington, "Trike v.1 Methodology Document [Draft]," 2005.

[95] M. Muckin and S. C. Fitch, "A Threat-Driven Approach to Cyber Security Methodologies, Practices and Tools to Enable a Functionally Integrated Cyber Security Organization," *Lockheed Martin*, pp. 1–45, 2017.

[96] S. Barnum, "Rights in Technical Data-Noncommercial Items clause at DFARS 252," 1995.

[97] E. McMillan and M. Tyworth, "An alternative framework for research on situational awareness in computer network defense," in *Situational Awareness in Computer Network Defense: Principles, Methods and Applications*, IGI Global, 2012, pp. 71–85.

[98] A. D'Amico and M. Larkin, "Methods of visualizing temporal patterns in and mission impact of computer security breaches," in *Proceedings - DARPA Information Survivability Conference and Exposition II, DISCEX 2001*, 2001, vol. 1, pp. 343–351.

[99] A. Stotz and M. Sudit, "INformation Fusion Engine for Real-time Decision-making (INFERD): A perceptual system for cyber attack tracking," in *FUSION 2007 - 2007 10th International Conference on Information Fusion*, 2007.

[100] S. J. Yang, A. Stotz, J. Holsopple, M. Sudit, and M. Kuhl, "High level information fusion for tracking and projection of multistage cyber attacks," *Inf. Fusion*, vol. 10, no. 1, pp. 107–121, Jan. 2009.

[101] D. Ha, S. Upadhyaya, H. Ngo, S. Pramanhik, R. Chinchani, and S. Mathew, "Insider threat analysis using information-centric modeling," in *IFIP International Federation for Information Processing*, 2007, vol. 242, pp. 55–73.

[102] J. Yen, M. McNeese, T. Mullen, D. Hall, X. Fan, and P. Liu, "RPD-based hypothesis reasoning for cyber

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.

situation awareness," *Adv. Inf. Secur.*, vol. 46, pp. 39–49, 2010.

[103] P. Xie, J. H. Li, X. Ou, P. Liu, and R. Levy, "Using Bayesian networks for cyber security analysis," in *Proceedings of the International Conference on Dependable Systems and Networks*, 2010, pp. 211–220.

[104] N. Lau, R. Pastel, M. R. Chapman, J. Minarik, J. Petit, and D. Hale, "Human Factors in Cybersecurity - Perspectives from Industries," in *Proceedings of the Human Factors and Ergonomics Society*, 2018, vol. 1, no. 1, pp. 139–143.

[105] P. Barford *et al.*, "Cyber SA: Situational Awareness for Cyber Defense," in *Advances in Information Security*, vol. 46, 2010, pp. 3–13.

[106] A. Jakalan, "Network Security Situational Awareness," *Int. J. Comput. Sci. Commun. Secur.*, pp. 61–67, 2013.

[107] K. Jabbour and S. Muccio, "The Science of Mission Assurance," *J. Strateg. Secur.*, vol. 4, no. 2, pp. 61–74, Jun. 2011.

[108] P. Mell *et al.*, "Reports on Computer Systems Technology NIST Interagency Report 7756 (Second Draft) CAESARS Framework Extension: An Enterprise Continuous Monitoring Technical Reference Model (Second Draft)," 2012.

[109] M. Park, J. Han, H. Oh, and K. Lee, "Threat Assessment for Android Environment with Connectivity to IoT Devices from the Perspective of Situational Awareness," *Wirel. Commun. Mob. Comput.*, vol. 2019, pp. 1–14, Apr. 2019.

[110] S. Noel, E. Harley, K. H. Tam, M. Limiero, and M. Share, "CyGraph," in *Handbook of Statistics*, vol. 35, 2016, pp. 117–167.

[111] K. L. Dempsey *et al.*, "Information Security Continuous Monitoring (ISCM) for federal information systems and organizations," Gaithersburg, MD, 2011.

[112] M. Hogan, B. Piccarreta, M. Hogan, and B. Piccarreta, "NISTIR 8200 Interagency Report on Status of International Cybersecurity Standardization for the Internet of Things (IoT)," *NIST*. Gaithersburg, MD, pp. 1–185, Nov-2018.

[113] I. A. Tøndel, M. B. Line, and M. G. Jaatun, "Information security incident management: Current practice as reported in the literature," *Comput. Secur.*, vol. 45, pp. 42–57, Sep. 2014.

[114] K. M. Kavanagh, T. Bussa, and G. Sadowski, "Magic Quadrant for Security Information and Event Management," 2020.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826183 - Digital Society, Trust & Cyber Security E-Health, Well-being and Ageing.