UNIVERSIDAD POLITECNICA DE VALENCIA

DEPARTAMENTO DE COMUNICACIONES

Decision support elements and enabling techniques to achieve a cyber defence situational awareness capability

Departamento de Comunicaciones

Universitat Politècnica de València

A thesis submitted for the degree of

Doctor por la Universitat Politècnica de València

Valencia, March 2023

Author: Salvador Llopis Sánchez

Director: Dr. Manuel Esteve Domingo

Co-Director: Dr. Wim Mees

Page intentionally left in blank

Abstract

This doctoral thesis performs a detailed analysis of the decision elements necessary to improve the cyber defence situation awareness with a special emphasis on the perception and understanding of the analyst of a cybersecurity operations center (SOC). Two different architectures based on the network flow forensics of data streams (NF3) are proposed. The first architecture uses Ensemble Machine Learning techniques while the second is a variant of Machine Learning with greater algorithmic complexity (λ-NF3) that offers a more robust defense framework against adversarial attacks. Both proposals seek to effectively automate the detection of malware and its subsequent incident management, showing satisfactory results in approximating what has been called a next generation cognitive computing SOC (NGC2SOC). The supervision and monitoring of events for the protection of an organisation's computer networks must be accompanied by visualisation techniques. In this case, the thesis addresses the representation of three-dimensional pictures based on mission-oriented metrics and procedures that use an expert system based on fuzzy logic. Precisely, the state-of-the-art evidences serious deficiencies when it comes to implementing cyber defence solutions that consider the relevance of the mission, resources and tasks of an organisation for a better-informed decision. The research work finally provides two key areas to improve decision-making in cyber defence: a solid and complete verification and validation framework to evaluate solution parameters and the development of a synthetic data set that univocally references the phases of a cyber-attack with the Cyber Kill Chain and MITRE ATT & CK standards.

Resumen

La presente tesis doctoral realiza un análisis en detalle de los elementos de decisión necesarios para mejorar la comprensión de la situación en ciberdefensa con especial énfasis en la percepción y comprensión del analista de un centro de operaciones de ciberseguridad (SOC). Se proponen dos arquitecturas diferentes basadas en el análisis forense de flujos de datos (NF3). La primera arquitectura emplea técnicas de Ensemble Machine Learning mientras que la segunda es una variante de Machine Learning de mayor complejidad algorítmica (λ-NF3) que ofrece un marco de defensa de mayor robustez frente a ataques adversarios. Ambas propuestas buscan automatizar de forma efectiva la detección de malware y su posterior gestión de incidentes mostrando unos resultados satisfactorios en aproximar lo que se ha denominado un SOC de próxima generación y de computación cognitiva (NGC2SOC). La supervisión y monitorización de eventos para la protección de las redes informáticas de una organización debe ir acompañada de técnicas de visualización. En este caso, la tesis aborda la generación de representaciones tridimensionales basadas en métricas orientadas a la misión y procedimientos que usan un sistema experto basado en lógica difusa. Precisamente, el estado del arte muestra serias deficiencias a la hora de implementar soluciones de ciberdefensa que reflejen la relevancia de la misión, los recursos y cometidos de una organización para una decisión mejor informada. El trabajo de investigación proporciona finalmente dos áreas claves para mejorar la toma de decisiones en ciberdefensa: un marco sólido y completo de verificación y validación para evaluar parámetros de soluciones y la elaboración de un conjunto de datos sintéticos que referencian unívocamente las fases de un ciberataque con los estándares Cyber Kill Chain y MITRE ATT & CK.

Resum

La present tesi doctoral realitza una anàlisi detalladament dels elements de decisió necessaris per a millorar la comprensió de la situació en ciberdefensa amb especial èmfasi en la percepció i comprensió de l'analista d'un centre d'operacions de ciberseguretat (SOC). Es proposen dues arquitectures diferents basades en l'anàlisi forense de fluxos de dades (NF3). La primera arquitectura empra tècniques de Ensemble Machine Learning mentre que la segona és una variant de Machine Learning de major complexitat algorítmica (λ-NF3) que ofereix un marc de defensa de major robustesa enfront d'atacs adversaris. Totes dues propostes busquen automatitzar de manera efectiva la detecció de malware i la seua posterior gestió d'incidents mostrant uns resultats satisfactoris a aproximar el que s'ha denominat un SOC de pròxima generació i de computació cognitiva (NGC2SOC). La supervisió i monitoratge d'esdeveniments per a la protecció de les xarxes informàtiques d'una organització ha d'anar acompanyada de tècniques de visualització. En aquest cas, la tesi aborda la generació de representacions tridimensionals basades en mètriques orientades a la missió i procediments que usen un sistema expert basat en lògica difusa. Precisament, l'estat de l'art mostra serioses deficiències a l'hora d'implementar solucions de ciberdefensa que reflectisquen la rellevància de la missió, els recursos i comeses d'una organització per a una decisió més ben informada. El treball de recerca proporciona finalment dues àrees claus per a millorar la presa de decisions en ciberdefensa: un marc sòlid i complet de verificació i validació per a avaluar paràmetres de solucions i l'elaboració d'un conjunt de dades sintètiques que referencien unívocament les fases d'un ciberatac amb els estàndards Cyber Kill Chain i MITRE ATT & CK.

Page intentionally left in blank

Preface

Cyber Situational Awareness (CySA) is a very broad research topic. The acquisition of a **´cognitive state of mind`** is a long-term endeavour where human operators strive to understand, in a perfect harmony, all the internal and external elements when confronted with a cyber incident. The ideal solutions to achieve a situation awareness (SA) sometimes recall on aspects which may be considered science fiction due to its forward-looking perspective about the moment in which this could be feasible. CySA has been a recurrent shortfall for the civilian, military, and industry in the past years. It has urged the scientific community to make advances in related disciplines such as cognitive computing, artificial intelligence or human reasoning to name only a few. Numerous publications in conferences or workshops dedicated to this subject show an extensive research effort whose expectations will accelerate even more in the future. Notwithstanding that the implications of a wide recognition of cyberspace as an operational domain for military operations equally important to the traditional physical domains - land, air, maritime and space – are yet to be experienced. This scientific field conforms well with a dual-use approach between civilian and military. The benefits of adopting a comprehensive and mature product may impact as well any security operation centre (SOC) at many organisations and ultimately could assist to a rapidly evolving digital society.

SA in cyber defence may be interpreted in simple words as the necessity to understand completely the operational environment where a mission is planned and conducted. The foundational principles can be understood thanks to other disciplines like psychology - which is very much related with the cognitive progress and human factors. Definitively, operators and decision makers are at the heart of any technological development and therefore their performance and efficiency need to be compared with the human limitations in addressing perception, comprehension, and projection. That is the ‘beauty’ and the complexity in this thesis, the human-centered design when approaching solutions for decision making. With the emergence of disruptive technologies, enabling techniques are becoming indispensable to accelerate innovation. Network administrators and operators rely on automated processes to ease their daily business workload on routine activities. Automation is seen as an inevitable consequence if a more effective and quick response need

to be articulated to counter cyber threats in real-time. SA was conceptually incepted in the 90´s due to the concerns raised by the United States Air Force to adapt modern cockpits demanded by a digital transformation of the aircrafts. This adaption was thoroughly analysed by a multi-disciplinary team of engineers to create a fit for purpose environment for experienced pilots. This piece of work set the fundamental basis of today´s knowledge applicable to other disciplines with salient ramifications to cyberspace. Dr. Endsley´s model is generally accepted as a valid approach able to provide a comprehensive overview of the main tenets to achieve SA. Lessons learned in the implementation of SA in the air domain paved the way for a plethora of various topics which fall under the category of human factors like visualisation, attention and decision. A decision support mechanism in support of a CySA needs to be seamlessly integrated into an overall flexible architecture which comprises other functional parts. Although the decision making can be interpreted as the core function that reaches an ultimate goal, a modular approach in an open architecture would permit to connect other meaningful modules such as visualisation, data processing, analysis on cyber threats, etc.

To facilitate a reader´s comprehension, this doctoral research work comprises 7 chapters. The first *Chapter* titled *Introduction and Objectives* will provide a more detailed explanation on the research plan approved by the doctoral academic programme in telecommunications engineering in 2016. *Chapter 2* will address the *state-of-the-art technology* in CySA notably on which emerging trends are subject to research. This thesis reflects years of experience working on this subject by the today´s doctoral student complemented by several publications in congresses and journals to shed some light on the research possibilities and challenges. *Chapter 3* will address a *description of technology modules and components* of a decision-driven design which uses fuzzy logic for a three-dimensional representation of an operational picture. *Chapter 4* will describe the *proposed architecture* subject to a further implementation. Chapter 5 will delve into the *architecture validation and tests*. Chapter 6 will analyse enabling techniques for decision support systems and *Chapter 7* will describe some *conclusions and future work*.

Acknowledgements

To my wife and children for their outstanding support in conducting my research studies.
Special Thanks to my Director and co-Director for their support in making this PhD a reality.
In memoriam, to Salvador Bernad for always encouraging me to take on this endeavour and never give up.
To Belen for her friendship and kindness.

Page intentionally left in blank

Contents

Chapter 1
Introduction and Objectives

Page intentionally left in blank

1.1 Introduction

The thesis at hand comprises various research activities that have evolved along the years (2015-2022) at the same time that I am improving my skills and getting a better understanding of the technology challenges to face in order to approach a **decision system for CySA**. Along the thesis´ chapters, a reader would notice the extensive research done - evidenced by the number of publications and in collaboration with other European researchers from industry and academia - to address different elements that fit into the problem statement and ultimately, shed some light on the complex information infrastructure, techniques and data analytics tools which are required to shape a new generation of cognitive computing Security Operations Centres (SOC). Novel frameworks and tools for incident handling processes in support of cybersecurity decision-making were investigated by using some machine learning algorithms for data classification and analysis as explained in chapter 4. Besides that, visualisation, human factors, metrics and datasets were meaningful topics of research being addressed to close identified gaps as described in chapter 3. The author, in pursuit of obtaining tangible research results, strived to narrow the scope of the conducted research tasks while recognising implicit dependencies with other areas of cyber defence to achieve the goals of the thesis. In this 'research journey', the rapid pace of technology evolution is bringing new possibilities to comprehend the characteristics of the cyberspace as a virtual man-made domain. Cyber situational awareness is tied with the cognitive process of the human brain to recognise patterns of a given situation in cyberspace. A certain level of automated tasks is required due to the fast speed at which actions occur in cyberspace. In this endeavour, education and professional experience along these years has been paramount to focus the research efforts.

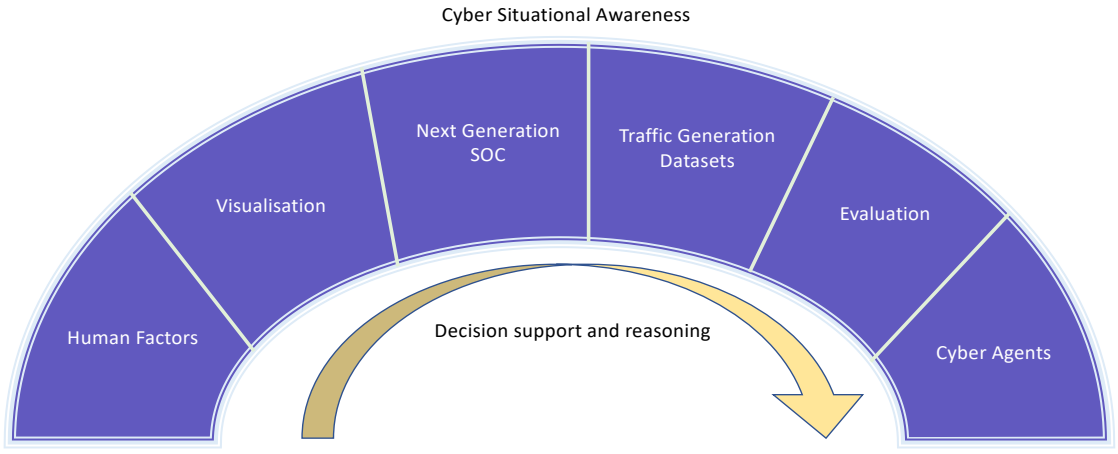


Figure 1. Perimeter of the research activity and its evolution clockwise.

1.2 Objectives of the thesis

This doctoral research comprises the following goals:

1. Understand and analyse cyber defence situational awareness and its research challenges;
2. Conduct research activities to bridge computing and data processing with operational aspects (business needs) of a cyber decision system;
3. Identify shortfalls and elucidate future lines of research;

The above-mentioned objectives led to the following research tasks:

- Get acquainted with the foundational basis of Situational Awareness including its application to cyberspace by performing a profound analysis of the literature and research advances;
- Study international initiatives addressing CySA;
- Study and analyse human factors and human system integration (HSI) concepts;
- Conduct research on cyber defence visualisation tools;
- Elaborate metrics and apply fuzzy logic to approximate reasoning mechanisms for mission planning and execution;
- Conduct research on machine learning (ML) algorithms;
- Study the applicability of the ML algorithms by automating processes in support of the identification of cyber threats in a SOC;
- Conduct research in view of creating synthetic datasets through modelling and simulation;
- Design thorough techniques for validation and verification of CySA and their corresponding definitions to guide experimentation;
- Study the intelligent cyber defence agents as a prominent area of future research;

Chapter 2
State-of-the-art Technology

Page intentionally left in blank

2.1 Cyber Defence Situation Awareness.

SA has been characterised in simple words as “understanding what is happening”. SA has an implicit link to the art of deciding. This notion was expanded to encompass a condition of human consciousness that enables decision makers to "identify and grasp the elements of their environment in terms of their physical and temporal context, and estimate what would be their evolution over time" (Endsley et al., 1998). This term has become noticeably more relevant in the context of the cyber defence, necessitating an understanding on how perceived cyber-situations might damage critical infrastructure, services, and assets integrated in physical domains. Despite its relevance, there is presently no solution entirely capable of accommodating a military´s commander tactical, operational, and strategic demands due to the dimensions of technological disruption, as well as the dynamism of the Information Technology (IT) industry. Because different functionalities can be organised into modules in an open architecture with numerous information exchange interfaces, system integrators are primarily responsible for resolving the issue.

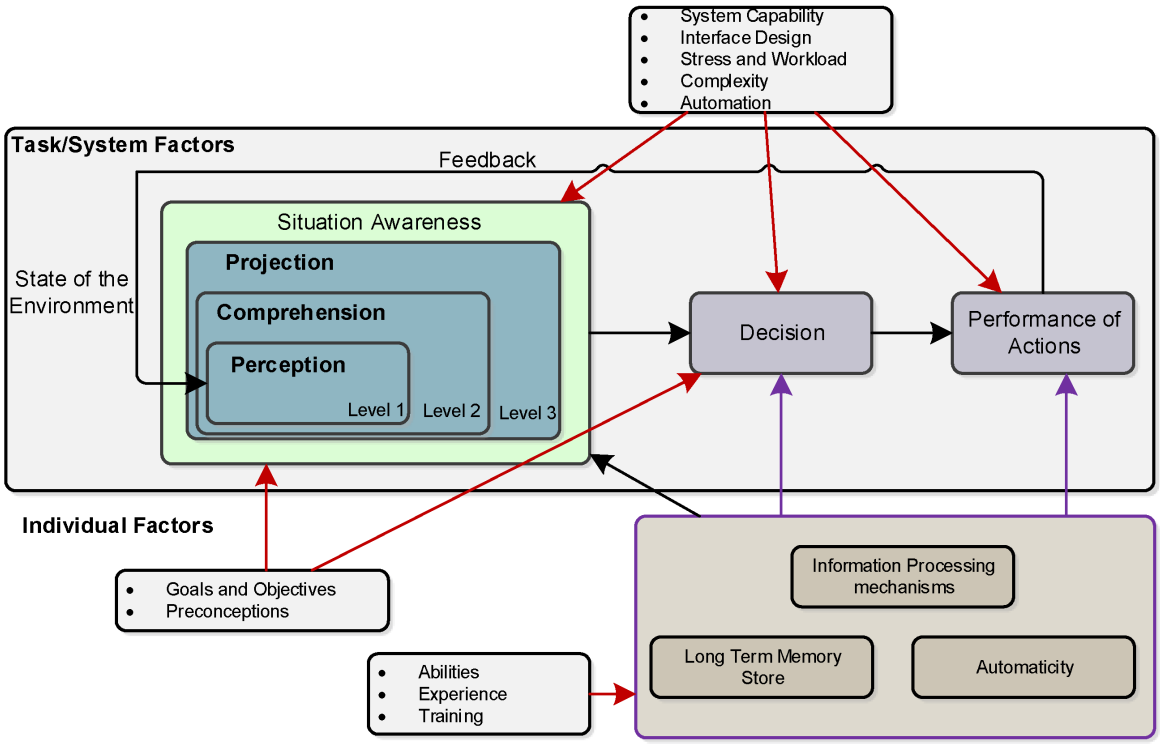


Figure 2. Situational awareness building blocks and decision-centered design (based on Endsley, 1995)

The following paragraphs will provide a description about the existing knowledge and latest technology trends on cyber defence situation awareness with particular emphasis on **decision support systems** and enabling techniques with the objective of illustrate which areas are of

high interest for researchers to fill the identified gaps. CySA is not a mere collection of threats from which to derive a cyber defence posture, neither is a dashboard that shows software vulnerabilities to patch. This is a simplistic reduction of the potential associated with a CySA capability. For tasks such a network monitoring, there are commercially available solutions while for the most of the required features of a complete CySA system, no solutions or products meet today´s increasing user needs.

2.1.1 Situation Awareness Assessment

Endsley hypothesised three types of approaching reasoning in a given situation: insight of external elements, understanding of their meaning, and prediction of future status once the previous two had been incorporated as human knowledge. The initial step (in a sequential order) includes duties such as observation of the environment and anomaly detection while the other two include: data processing, correlation, and foresight to determine the occurrence of similar events in the future. These tasks to be executed in order to protect a given system allow network administrators to take stock of where to dedicate more resources to risk mitigation in such a way that it is possible to acquire information on the state of the network through various sensors, the analysis of the risk level, action planning and evaluation of the measures to be implemented in a continuous information cycle. Given the logic and simplicity of this model, its application has been studied for several years including new ways of conceptualising SA. One of these attempts is known as the observe-orient-decide-act (OODA) loop (see Figure 3). OODA loop is a method to systematise the decision process. Although its phases are successfully used in other fields of science, its usefulness in cyber defence is not commonly accepted, perhaps due to the speed at which incidents occur and therefore the need to adopt an even more expeditious process that could shorten the OODA phases. It is true, however, that the OODA loop follows the above described Endsley´s logic to assist in the identification of the most suitable countermeasures by analysing all the judgment elements available to system administrators and operators in order to better assess the situation and propose remediation actions according with the risk level. Therefore, it can be considered as a valid mechanism for first response which is key to reduce the impact at the early stages of a cyber-attack. The OODA loop repetition over time would ensure that new observations may trigger some changes in the analysis. This approach would allow for near-real-time reaction to complicated circumstances while also boosting the effectiveness of first-response actuations.

A detailed description of these terms is provided as follows:

• **Observe**. The perception of the operating environment is achieved by the fusion of heterogeneous sources of information, which are immediately acquired by sensors.

• **Orient**. Provide a baseline of understanding through the use of modelling and simulation, previous experience or accumulated expertise to discern among small traces or hints.

• **Decide**. Make use of previous findings to determine those actions that are deemed necessary to mitigate damages in a timely manner. The decision implies ordering all the acquired knowledge so far to infer possible countermeasures and finally plan the consequences of every recommended course of action.

• **Act.** It means proceeding with the implementation of the previously agreed measures and verifying their effectiveness for the best fit depending on the effects to be achieved.

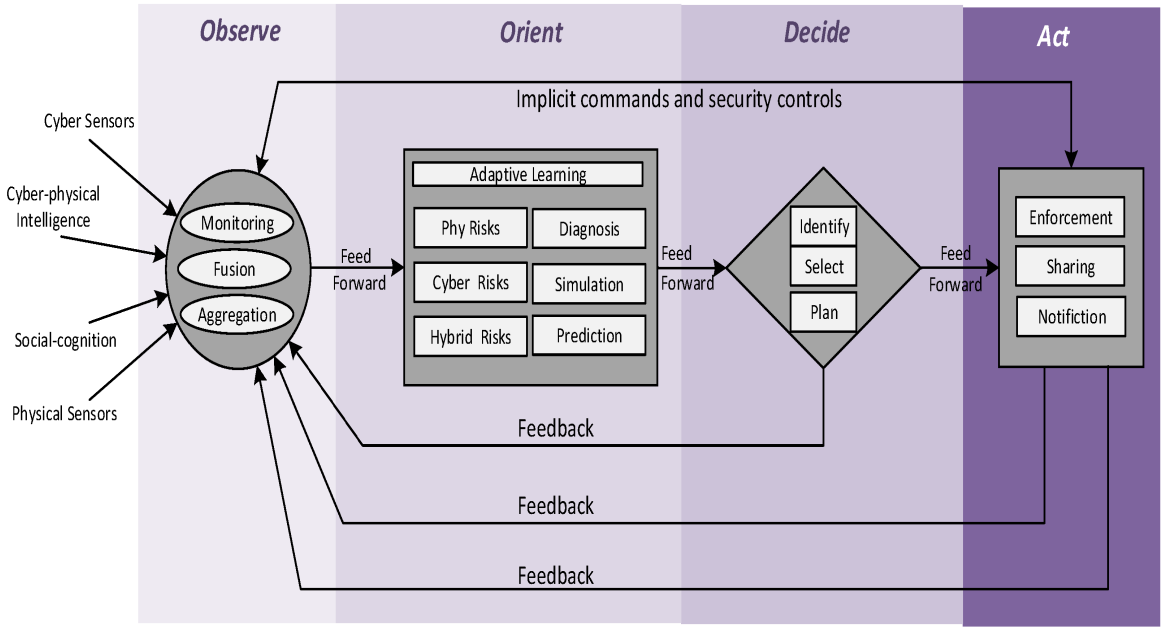


Figure 3. Description of OODA loop and its similarities with Endsley's model.

This approach enables a near-real-time reaction to complicated events while also providing a systematic feedback to tailor recommended measures at all times (Lenders et al., 2015). Several of the associated preliminary contributions were included in (Franke et al., 2014), along with an in-depth review of the CySA landscape. According to (Chatzimichailidou et al., 2015), the three most common flaws of an information security risk management (ISRM) process are: (1) ISRM regularly performs superficially; (2) risk management requires an in-depth knowledge of the digital infrastructure to be defended, which is often overlooked,

resulting in poorly valued cyber assets; and (3) risk analysis is considered a static process that only takes into account certain reporting periods which results in a miscalculation of potential dynamic forecasts. On these considerations, the OODA loop has emerged as a significant facilitator to structure the processes involved in capturing SA when reaction teams must adjust their strategies, tactics, and procedures as the operation advances. The practical applicability of the OODA loop to cyber physical systems (CPS) is not automatic since they show a series of drawbacks such as their different hardware and software configuration to monitor and control failures. These characteristics may lead to different CPS risks which shall be treated in conjunction. The dual-use of CySA linked with CPS brings the opportunity to develop a fit for purpose solution adaptable to the CPS information infrastructure but with the added value of accommodating aspects such as the operating environment or the organisational setting. (Llopis et al., 2019).

The term measure of effectiveness (MoE) refers to the measurements that characterise a cyber defence team's operational effectiveness in fulfilling its objectives during a mission. Several cutting-edge mission-mapping technologies with cyberspace dependencies were examined in (Schulz et al., 2015). An analysis of the relevant cyber assets according with mission needs, their appropriate score with respect to mission objectives are tasks which fall under the responsibility of mission planners and ultimately contributes to a CySA (de Barros et al., 2013). These assets are usually known as cyber key terrain (CKT) with an interesting approach made by (Price et al., 2017). The way cyber assets are clustered per service or per category may influence the overall comprehension of their mission impact, being the most critical component in establishing the CKT. All these discussions are trying to shed some light to address the issue of the valuation of those groups of cyber assets in a traceable manner more typical of system engineering practices. Their associated cyber risks will determine the feasibility to conduct an operation with given means or by the contrary to request additional support to reduce the accumulated risk level. Cyber-ARGUS introduced by (Damico et al., 2009) in another attempt to perform the mathematical calculations of cyber assets, their dependencies and the overall impact based on a cyber threat assessment. After a review of the state-of-the-art technology, the existing frameworks are still in its infancy to go for a rapid implementation and prototyping. The theoretical foundations strive for operationalising the identification of cyber assets and their impact over missions as a valuable tool for mission planners. There is most likely no other way than to consider the use of a powerful modeling and simulation (M&S) system shaped like a cyber range to

reduce the workload of analysts and the establishment of meaningful metrics for the evaluation of cyber risks and impact during the mission analysis. In summary, it is proposed the following areas for further research: (1) development of a standard taxonomy of cyber assets and cyber capabilities; (2) development of optimised M&S tools to perform mission analysis in a pre-production environment including mathematical calculation on risks, impact before and after the occurrence of a cyber-incident. The calculation baseline would take into account the improvements when carrying out recommended actions; (3) create a knowledge database to learn from past experiences and potentially integrated in a M&S environment to assist mission planners.

The importance of the cyber threat intelligence cannot be underestimated in order to acquire an informed SA (Endsley, 1995). SA includes a progression of getting basic information to the ability to understand and combine facts to create new knowledge, foresee and anticipate events, and produce a list of recommended mitigating measures. SA follows an incremental approach, with higher levels of comprehension being somewhat dependent on lower levels of awareness (Endsley, 2016). However, as noted by (Brynielsson et al., 2016), objective quantification to measure if a cyber defence product aids an operator to acquire a certain level of SA is a well-known shortfall. Measures become more complex when trying to discern what is important from what is superfluous in terms of assessments about the situation. Existing methodologies contemplate a comparison of small events which are assessed over time. This arduous task requires specific preparation and training for instance in the frame of a cyber exercise. In this vein, (Stevens, 1968) highlighted the benefits of a quantitative comparison of SA levels with a prior defined reality. This is a well-known claim that has received a lot of support from the academic researchers. It is important to note that to be able to measure SA levels, a reference on the true values must be obtained in order to confront disparate perceptions with the reality. (Parasuraman et al., 2018) advocates for its existence while others (Dekker et al., 2008) criticises the viability of the process to come to terms with this comparison. The research community has created and verified many methodologies that frequently define models tailored to their application area in order to address this issue. (Salmon et al., 2008) provides a solid illustration of its feasibility to other sectors related with critical infrastructures.

SA can be considered in (Endsley, 2015) as an indicator of efficiency. Decisions may be made more accurately and based on tangible facts. The following items provide a set of recommendations to support a valid and reliable methodology:

1. Creating metrics that exclusively evaluate the thing the method is intended to evaluate.
2. Using tact and diagnostic methods to offer the required comprehension.
3. Using a balanced probing method for each distinct objective.
4. The build shouldn't be drastically changed while the method is being done.

These criteria have become especially important when evaluating cyber situational awareness since the collected image exhibits both human-related and technical characteristics carried out by automatisms. The execution of cyber defence exercises (CDX) could provide a context where to measure CySA. According to (Lif et al., 2017), the focus is only on technology instead of personnel and processes to infer participant's understanding about the environment and how this understanding is impacted by the presented operational picture radically different from the point of view of how well the supporting technological enablers perform and are accepted. This approach is not exempt of criticism (Buczak et al, 2015). To counter this, (Gutzwiller et al., 2016) is in favour of a human cognition perspective for CySA. According to (Endsley, 2016), the common perception of CySA as the sole outcome of technological oriented experiments would be a mistake. Despite an increase in publications discussing the relevance of the human element in CySA-related themes (Mahoney et al., 2010), few studies have employed the actual CySA measuring technique (Malviya et al, 2011) and assessed it in real-world situations. The contributions of (Giacobe et al., 2012) should be highlighted in this study, where a cyber situation awareness global assessment method (SAGAT) questionnaire was developed to measure the participant's acquired CySA. Similar to this, situation assessment rating technique were tailored to tests and experimentation in (Evangelopoulou et al., 2014). Situation Present Assessment Measure (SPAM) was used to evaluate CySA in an attempt to lessen the level of disturbance of the training audience when completing the tasks. Other investigations, like that of (Lif et al., 2017), claims that frozen probing measuring strategies are workable in some circumstances, including large-scale CDX. Previous task assessments (Dressler et al., 2014) offer a detailed description of which data can enhance CySA collection. The research addresses what should be visualised by log analysers (Shiravi et al., 2012). A CySA system must employ various types of information to provide a generic SA. They represent CySA´s comprehensive grasps, although they still require additional clarification and analysis.

Network security visualisations, however, partially address CySA (Dressler et al, 2014). The relationship between the mission needs and the technical data layer is also not taken into account by these methodologies, which results in a severe shortcoming when used exclusively to assist cyber defence activities.

2.2 Research and Technology Challenges

2.2.1 Human Factors

Cyber threats are fast moving and continuously evolving, thus it intrinsically requires establishing a quick reaction mechanism to mitigate the risk when a cyber-attack takes place in different cyber defence systems. This quick reaction mechanism involves early detection and *mature decision cycle*. The possible consequences of a cyber-attack - if there is not an immediate action - urge communications and information systems (CIS) administrators and users to play an active role in defending and protecting the networks since they are responsible for performing the necessary actions to reduce vulnerabilities. It is a demanding challenge to have the human knowledge and deep expertise to patch any "avenue of approach" that a potential attacker would be able to exploit. The latter not only applies to cyber defence technology but also has a strong link with the areas of procedural, physical and personnel security. Moreover, automated tools partially help to understand and create situational awareness in multiple scenarios, thus the human being is the core of the process. Operators interpret the information provided by automatisation to comprehensively assign tasks. Human intelligence must be inserted in the system in addition to the threat landscape and as added value to military commanders and their respective staff in accomplishing missions.

The consideration of the human nature within a system is key since it is usually an operator who acts in view of defending the system against a potential attacker. Humans and not the machines are the ones who bear responsibility for the actions undertaken. HSI studies the appropriate insertion of the human nature in the correct functioning of any system, accepting human criteria as part of the system design. Human cyber awareness, decision making under uncertainty and supporting individual resiliency becomes obvious by analysing and compiling current research literature on human factors in cyber defence. Fact finding talks with experts demonstrated a considerable need for HSI regarding tool support at operator's

level. In particular, daily tasks in network monitoring and maintenance are rule-based activities causing high workload and distracting an operator´s attention from knowledge-based tasks such as threat analysis and incident anticipation.

The fundamental challenge of cyber defence systems lies in the understanding and handling of the highly dynamic, non-transparent and non-deterministic environment that becomes apparent in the increasing quantitative and qualitative rise of cyber threats. Following this rationale and in the author´s opinion, a human center design must be considered notably by introducing expert systems in the architecture of a decision support module with various levels of autonomy depending on the criticality of the mission.

The basis for discussion on processes shared between human and automation is the *decision ladder*, which matches cyber defence tasks with decision processes (Özyurt et al, 2013), (Rasmussen & Goldstein, 1987). Results of original works based on this method in the context of control stations in high-risk industries are applied for cyber defence. It can be interpreted as a mapping exercise of an operator's cognitive processes on a set of information processing activities and cognitive states.

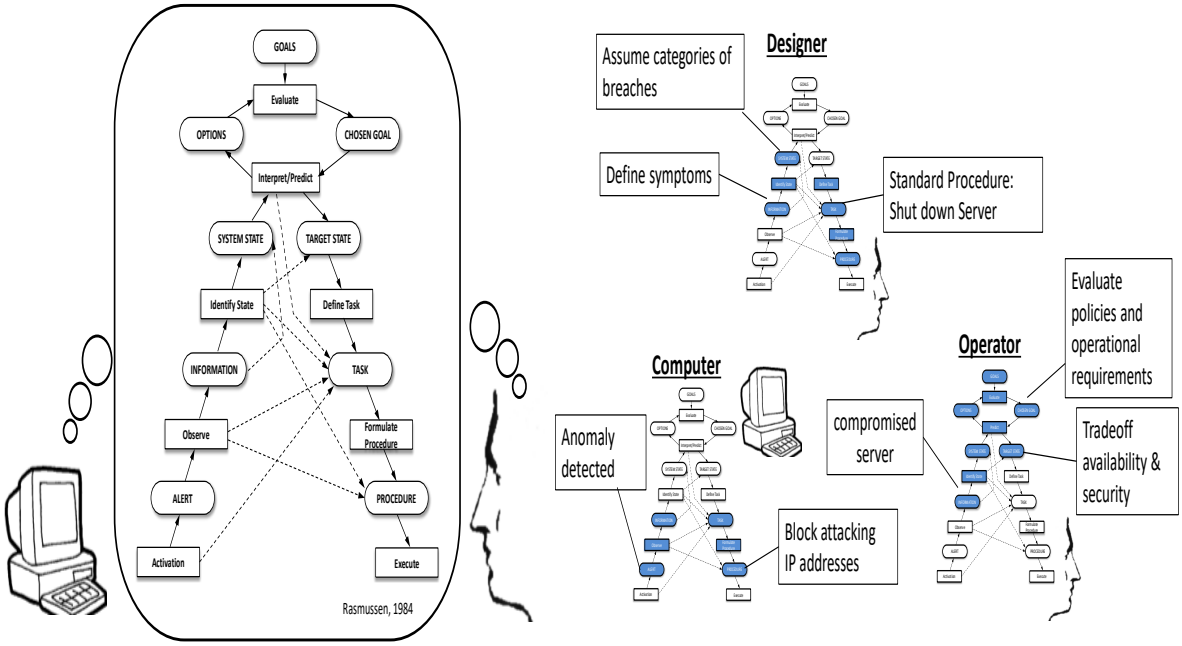


Figure 4. Decision allocation of cyber defence tasks.

The decision ladder consists of three stages: the situation assessment (up at the left-hand side), options analysis (across the top), and planning and execution of an appropriate action (down at the right-hand side). Decisions are not only allocated to operator and computer but also much more important to the designer/developer of the system. The designer tries to incorporate information to make decisions in advance for assumed possible hypothetical situations. Whenever the designer thinks that he is not able to presuppose all necessary information to decide on the next steps, he will design the system to get the human into the loop. The human may provide further information to the system the computer is not able to provide, like do classifications. This classification might possibly be sufficient to proceed with automatic procedures, but in complex situations it might not be clear if the needed information is sufficient to decide on selecting a rule triggering the right actions. In this case, the designer will also leave the decision on the next actions to the operator and might provide a set of possible actions for which he/she has defined automatic procedures from which the operator might choose. The operator sets the course into appropriate automatic procedures at critical points on the decision ladder. The designer needs to communicate with the operator through the user interface implemented into the computer. Actually, this is the primary communication. The computer is a surrogate for the designer. This perspective of thinking about the design of the communication between human and computer puts the focus on a more human-like communication and lead to more intuitive interfaces.

There are different patterns of possible allocation of decisions. The designer and the operator can collaborate by taking care of different sub-goals. The designer has defined tasks and procedures for the shutdown. The operator is informed about the shutdown, but is not involved in executing and planning the procedures for the shutdown. The operator is maintaining and supervising the system during the shutdown. The shutdown of the whole system might not be necessary, but possibly shutting down the infected components to prevent the spreading of malicious code or possible spying activities originated from the infected host or server. The designer has defined appropriate tasks like starting backup activities and the actual shutdown of the server. Moreover, the allocation of different sub-goals goes a step further, since the operator keeps track of the availability of important services during the reconfiguration of the system to achieve a smooth state transition. Different subtasks could also be allocated between operator and designer. Whereas by the collaboration on different sub-goals, designer and operator act parallel for system monitoring, now they act sequentially. While the designer has defined and implemented

procedures for the identification of the security incident, the operator monitors the performance of the system. The designer leaves the diagnostic part of event-identification to the operator. The computer displays the outcome of algorithms aggregating symptoms of the system state. The operator interprets the displayed symptoms and is able to include further context information to identify the event more reliably, since the events might vary a lot. In this example, operator and designer also cooperate at the stage of choosing the right tasks. The designer is still in charge for setting goal priorities and providing plans for actions stored in a database. In a military context, the operational requirements are dominant in setting goal priorities. The designer cannot consider all possible operational requirements, which might be very dynamic with a high variety. The operator/commander possibly needs to decide on a trade-off with respect to availability and security. Whenever the designer wants to share information of the automation processes with the operator, he/she needs to choose the right level of abstraction appropriate to the level on the decision ladder. This question is related closely to provide visualisations and user interface designs supporting SA. The decision ladder is an appropriate tool in the development process to discuss how automation is planned to be integrated and how responsibilities with respect to failures are allocated. Participatory methods in the conceptual phase of automation processes are recommended; these take into account that not only the computer and the operator are collaborating, but also the system developer and the operator, and that this communication supported by the computer should be designed accordingly. It is proposed to focus on risk management, training and tool development to strengthen resilient behaviour of computer emergency response team (CERT) units. The study on human factors recommend investigating and developing further operational and technical solutions to anticipate and understand long term cyber trends and developments e.g. monitoring security trends and future threats or implementing a cyber-advisory function, to learn from past events e.g. gain awareness and knowledge of new routines and attackers behaviour; to monitor short-term developments and revise threats and risk models continuously e.g. development of key measures on their indicators for experts and for a common operational picture (COP); and respond to regular and irregular conditions in an effective, diverse and flexible manner.

Chapter 3
Description of technology modules and components

Page intentionally left in blank

3.1 CySA Technology Demonstrators

3.1.1 ARMOUR

ARMOUR is a technology demonstrator commissioned by the Defence Research and Development Canada that constituted a first attempt towards a platform of cyber security applications. The reason for such development was the opportunity to create automated configuration options for network administrators based on an assessment of the threat landscape. These tools aimed to offer a response instrument that analyses remediation actions according with available resources and establishes priorities for action in a semi-automatic fashion. ARMOUR design has influenced subsequent initiatives to address decision support tools in response of network vulnerabilities. The goal is to merge various sources of information, create a knowledge database to dynamically secure networks. To fulfil its mission, the system was conceptually addressing different steps from data gathering, data correlation, data abstraction including its combination with logical rules learned from known cyber-attacks techniques to derive possible attack paths and to pre-empt similar cyber incidents in the future (Sawilla & Wiemer, 2011). This approach was too much dependent on the type of network infrastructure and the security perimeter devices such as firewalls or routers. Today´s computer networks are more diverse, decentralised and even the cyber security protection is evolving towards data centric instead of network centric. The originality was related to the integration of disparate functionalities into a unique platform. By using data automatic and programmable interfaces, various information repositories are connected to the platform e.g. vulnerability databases in addition to an engineering effort to come up with response options to mitigate the impact of specific software and hardware flaws is a major implementation endeavour that goes in the right direction according with the general principles of obtaining a CySA. Figure 5 shows the logical architecture of the components to achieve the required data processing able to produce courses of action or mitigation plans to be chosen by the security network administrator. ARMOUR kind of thinking has boosted many associated research lines for industry and academia along the years such as the study of mission impact, effectors, mission planning, data connectors, network monitoring, action plans, configuration plans, definition of attack paths, decision support and a large etcetera.

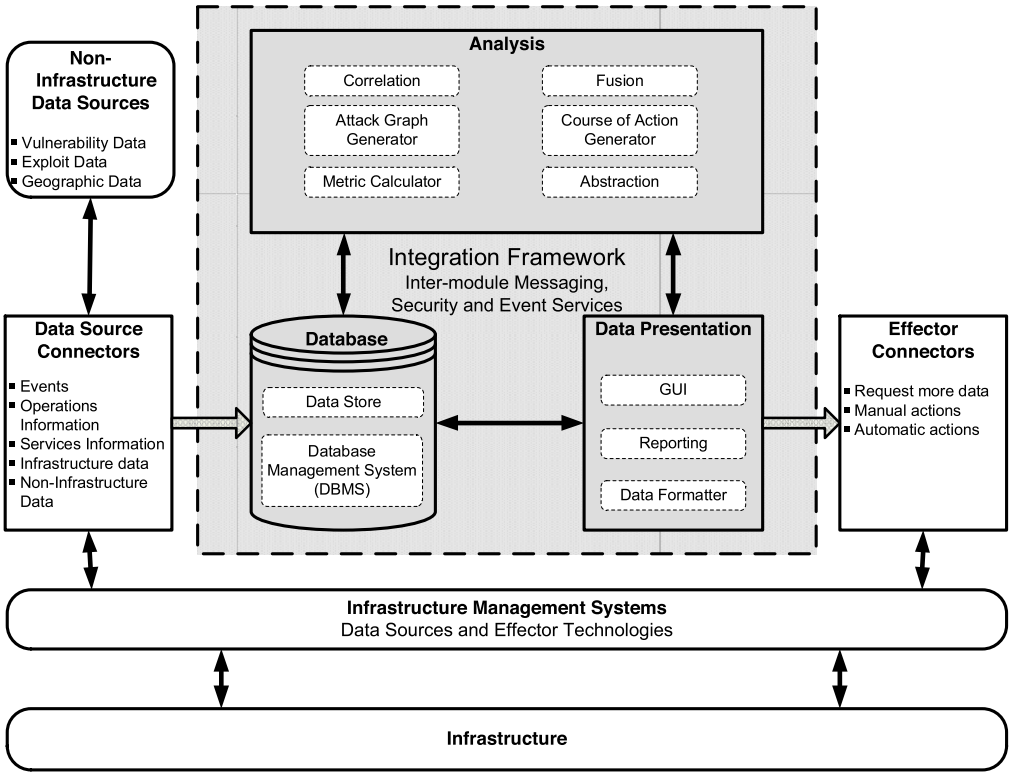
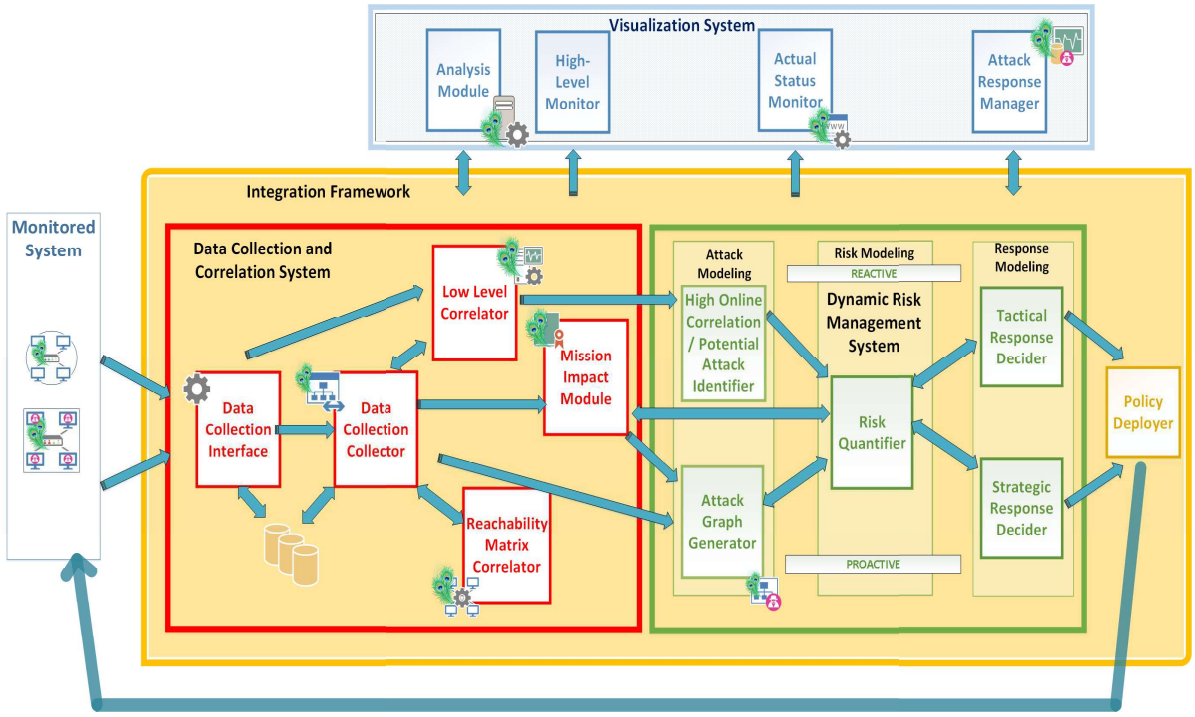


Figure 5. ARMOUR logical architecture framework (Sawilla & Wiemer, 2011)

3.1.2 PANOPTESEC

The PANOPTESEC system is as well a demonstration activity to build a set of specialised cyber decision tools very much aligned with ARMOUR in its conceptual design. This project was funded under the EU Framework Programme 7. The project provides an innovative dynamic risk assessment to base subsequent automated decision planning and support of a critical infrastructure. Although the overall construct tackles profoundly the data collection and correlation, visualisation and dynamic risk management, the response system is designed as a final phase of the data work flow to propose tactical or strategic decisions based on the inputs received and executed by a policy deployer as shown in Figure 6.



Source: FP7 PANOPTESEC

Figure 6. PANOPTESEC data collection and correlation

3.2 Visualisation

Visualisation and design of user interfaces play a crucial role but - despite various approaches of visualising networks and network activities - this is still an open issue where more effort should be brought to future research activities. The research domain of visual analytics is promising in designing appropriate cyber situational pictures for various stakeholders as the military commander, the cyber response team, civil or industrial experts. For operators in a SOC or cyber defence unit, automation is applied in the aggregation and analysis of data from monitoring hosts and network traffic by using a security information and event management (SIEM) system. The data is highly dimensional and abstract and difficult to visualise. However, since automation pre-processes the data, it is important to determine on which abstraction level the human operators need the information to build their mental model to achieve an appropriate CySA. User-training should be applied to induce appropriate mental models to visualise and make judgements on cyber situations.

3.2.1 A multi-aspect three-dimensional operational picture

Through an effective visualisation, human operators can interpret a certain situation. The issue here is that the displayed elements must represent the cyberspace and the cyber assets. In the scope of my research, I developed a three-dimensional (3D) operational picture specific to the cyber defence domain together with a group of researchers. One of the challenges to tackle was the overwhelming number of information sources that a military commander is confronted with that often hampers his/her decision-making. The aim was to facilitate an intuitive evaluation of the situation by visualisation means. With the development of the visualisation tool, the decision-making process can be more agile due to the fact that only a single glance at the images and some preparatory training about the colour codes and shapes used, the mission risk can be understood. Additionally, it provides variable decision elements based on metrics to assess the criticality of the cyber defence assets. Using a reference scenario, the telecommunications and cyber defense elements that are needed for the effective fulfillment of the mission are being designed. The research carried out demonstrates the possibility of inserting various information elements jointly integrated into a 3D visualisation solution with several views and perspectives from different angles. The displayed pictures reflect the key information at various levels of abstraction, proposing changes in the way the components of the assets are encoded as the situation evolves. The 3D visualisation is based on a “Mission – Attacker – Controls" (MAC) triangle for each cyber asset where planning elements of interest for the commander and his/her staff support element are analysed. The distribution of forces in the MAC triangle is determined by low-level security metrics which are calculated based on measurements. The technological basis of all the calculations is **fuzzy logic** which, through an aggregation process, allows the quantification of final values of the force component that is reflected in the sizes of the figures.

The interest of this type of visualisation is the ability to identify the risk and impact on the mission very rapidly (Mees & Debatty, 2015). For decision making, the current state of information security is of less interest if it has no influence on the mission. A common operational picture (OP) is the most widely adopted method to understand at a glance what is happening while at the same time it is a tool to further refine the mission components or dependencies.

The conducted research introduces a 3D visualisation which offers to eliminate the existing barriers in the computerised treatment of data and to focus on acquiring SA.

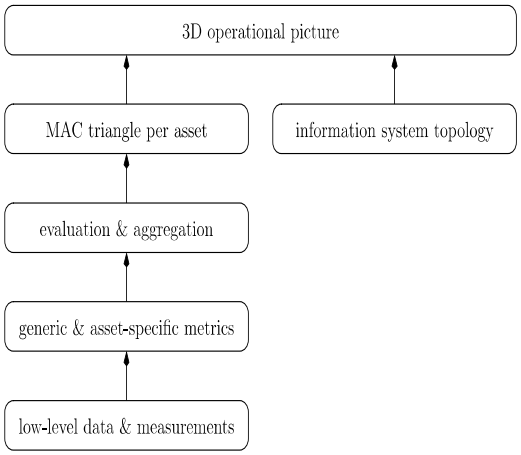


Figure 7: Conceptual design to build an operational picture based on fuzzy logic (Mees et al., 2016).

One of the misconceptions to deal with such developments is to display a bunch of quantifiers and statistics which are not often used by decision-makers. This research is instead focused on providing graphical views with a seamless interpretation of their meaning. The 3D visualisation is built to respond to the following questions:

- Do I have the necessary cyber assets to conduct the mission?
- How the cyber situation may affect the commander's intent?
- May I deduce from the pictures what are the cyber risks?
- Does the visualisation assist in the comprehension of the situation?

The operational views use a predefined codification represented by symbols such as rectangles, cylinders, etc. For instance, services are represented by rectangles and computer networks are represented by cylinders. An indication of the threat level is determined by the height of the cyber asset in the picture. A higher altitude of the cyber asset means a higher vulnerability and therefore a high risk can be deduced. The application of counter measures will decrease the risks and therefore a decrease of the height of the cyber asset. This intuitive method allows for multiple configurations to better understand the cyber situation.

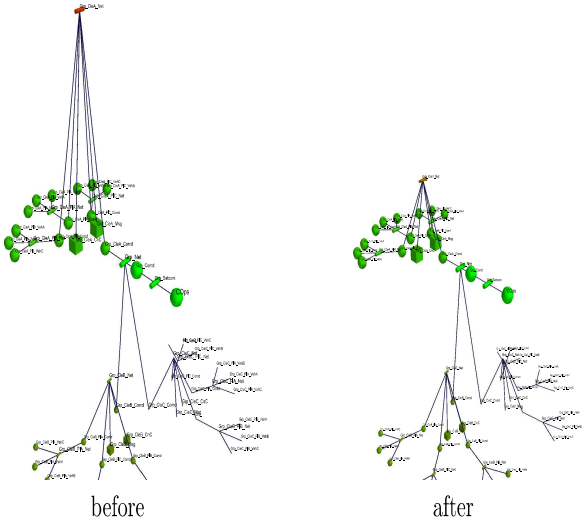


Figure 8: 3D visualisation before and after applying changes to cyber assets (Mees et al., 2016).

The triangle of forces called MAC is the core element for representing the 3D visualisation. It depends entirely on the designed metrics and measures to quantify the forces´ values. Only after quantitative variables have chosen, then technicians may undertake comprehensive evaluation to figure out whether preventive steps and defenses yield the best outcomes in reducing overall information security risks to meet mission´s objectives whenever necessary. Moreover, measurement criteria get matured due to being routinely checked to ensure also that acquired findings remain significant as generally acknowledged and pertinent by a mission planning team. The suggested test can take the form of a "wargaming" simulation used by fictional adversarial and friendly forces in order to engage together as a fit for purpose component of the military preparation procedure for reaching mutually agreeable findings prior to decision-making. Cyber security indicators highlighted in (NIST 800-55, 2003) serve to speed up selection, augment, or ensure efficiency through gathering, processing, and dissemination of important results evidence. As more than just results, measurement criteria are calculations that utilize an array of variables that try to characterise network attributes such as data security controls which are now be expanded to cover as well communications or processes with some limitations.

Low-level indicators, notably differentiated from high-level measures, are considered the immediate outcomes of assessments on components or independent network parts (Hecker, 2008). A collection of parameters is subjected to a shortlisting to decide which of them are the most appropriate for the measurement model. Perhaps linked with the chosen scenario and difficult to extrapolate to other cases unless there is a suitable adaption. What becomes paramount in terms of deciding which metrics apply in each use case is that any poor choice will have a significant impact on the outcomes, rendering the experts' efforts useless. Furthermore, the entire experience emphasises the importance of what is known as operational art which relies on dedicated expert´s knowledge. A team of cyber defence personnel, in collaboration with other computer and telecommunications technicians, assesses the cyber assets and given means to complete tasks, and consequently, supervises the achievement of CySA as a component of the wider mission. The mission requirements must have been evaluated against the protection measures. In this context, assessments are made with a specific mission in mind. Theoretically, a reproduced setting of secure scenarios throughout periods might produce important data that could be used to confirm or reject mission hypotheses. While evaluating the shared operating conditions, the following points might emerge: a) Are still the security measures in place sufficient to meet the commander's key requirements? b) How may scenario-dependent parameters be enhanced? b) Are there any circumstances which could be considered mission-critical and potentially lead to a mission cancellation? It is anticipated to address solutions to such concerns by designing, verifying, modifying, and discovering evaluation criteria deemed relevant to the chosen scenario. To establish cybersecurity measures, a thorough examination of what ought to be assessed and the reason would be required. Each result should be closely researched in order to validate truth with relevant facts and to exhibit certain practical outcomes relevant to analysts.

Table 1. Shortlist of variables.

components of F_C	control areas	associated specific variables per control area (meant to be more or less significant to one or several controls in an identification process)
pro-active human controls	• user training & awareness programs, • controls management	• people trained on cyber defence, • reporting and control echelons, • subordinate/ supporting units, • human factors, • asset usability
re-active human controls	• trained incident handling staff, • forensics experts	• people trained on incident handling or forensics, • incident handling or forensics mechanisms, • human factors, • asset usability, • response time to a security event
pro-active technical controls	• asset management, • configuration & change management, • vulnerability and patch management, • rapid localization of jamming sources	• network vulnerabilities, • vulnerability severity, • impact, • cyber hygiene mechanisms, • asset availability, • asset survivability, • protection level, • compromised devices, • time to operate, • network failures or malfunctions, • communication losses
re-active technical controls	• system monitoring, • attack detection	• network critical points, • network strengths, • cyber incidents detected
continuity / resilience	• service continuity management, • disaster recovery capability, • alternate circuits	• service availability, • operational capacity, • network readiness, • network resilience, • redundancy, • contingency links, • communications diversity

The set of variables in Table 1 isn't a complete set; in some other use cases, additional functions might prove better suitability. It is possible to handle any or more options within a single area due to the measures' versatility in relation to the influencing factors. Despite the fact that they are essential for taking into account by analysts, further managerial issues including policies, doctrines or institutional arrangements as well as extra defence concepts such as data protection are not in the scope of this initial demonstration. Some suite of protection and efficiency criteria that become suitable for the whole empirical work were suggested by (Cheng et al, 2014), with a major emphasis on system risks identification, incident monitoring and review, and damage assessment. Even if preexisting measurable parameters might well be accessible, it's rarely simple to find a measure that encompasses a

particular attribute. Following this rationale, Table 2 shows a shortlist of low-level measurements:

Table 2. low-level security measurements.

metric & type	description	score/value
cyber training of staff *(soft)*	expert assessment of the cyber readiness level of the staff as a result of training	set of terms (e.g. "substandard", "average", "good", "expert")
average response time *(hard)*	average time needed to handle a security event or incident	measured average response time (in hours)
average time to operate (cyber deployable assets) *(hard)*	average time needed to operate cyber deployable assets under the planned conditions (e.g. individual vehicles or dismounted patrols)	measured average time to operate deployable assets (in minutes)
network readiness *(soft)*	is CieA_network ready to accomplish the mission? e.g. all required services are supported by available servers	set of terms (e.g. "not ready", "some critical services down", "critical services available", "all services available")
asset survivability *(soft)*	survivability aspect of the CieA_network after being degraded, attacked or compromised	set of terms (e.g. from "no survivability" to "fully survivable")
communications diversity *(hard)*	number of direct communication links able to establish by different means simultaneously	number of direct communication links by different means simultaneously
network critical points *(hard)*	a revision of the network architecture can identify critical points subject to be exploited by an attacker	number of system critical points
resource redundancy *(soft)*	is there any redundant (backup) resources assigned or allocated for a critical task/mission?	set of terms (e.g. "no backup solutions", "some services have backup solutions", "critical services have backup solutions", "all services have backup solutions")

A combination of application metrics on a given scenario can lead to a complete analysis of the situation and give certainty to an operation commander in achieving his/her goals. The challenge resides in the correct configuration of the deployable resources rather than in the parameters of the fixed network that are commonly known in other studies. That would mean to expand the current cyber threats to a more mission-oriented aspects such as mobility, entry into service, and so on. Leaving out of the realm of information security metrics applied to computer networks to be focused on ad hoc solutions emerged from an analysis of operational constraints. This is the bridge to establish between traditional or classical cyber security controls with others which are less covered by the state-of-the-art.

It would be definitively interesting to analyse from a mission perspective (business-oriented approach) what is known as "operation assessment" (JDN, 2015) which is a series of activities that take place when the operation is in progress or in its execution phase. The mentioned methodology serves as a diagnosis to evaluate the progress of the lines of action.

This type of analysis is based on obtaining feedback on how the objectives are being met in a linear way and most important to know if corrections are necessary to achieve decisive conditions. These decisive conditions are key performance indicators (KPI) to determine changes of the situation, in particular MoE and MoP explained in chapter 2. To advance in the computational treatment of these subjects, low-level security measures, as depicted in Table 2, can be classified as hard and soft. Soft metrics are calculated by human expert´s expressions while hard metrics require an arithmetical quantification. A complicated actual issue such evaluating performance levels cannot be determined by using exclusively a domain expertise defined in a numerical manner. The generalised norms and limitations that apply to this topic are understood at a subjective and descriptive degree by the analyst, and are often represented employing ambiguous language phrases. A hazy range of issue knowledge is what this represents. Fuzzy logic control sentences serve to convey this unclear degree of comprehension. Those latter arguments may specify a sharp or an imprecise measure and a fuzzy set which indicates the anticipated numbers for such criterion in sequence to properly function for every action. After which, estimated inference is used to perform a specific evaluation (Jang et al., 1997), (Klir, 1995).

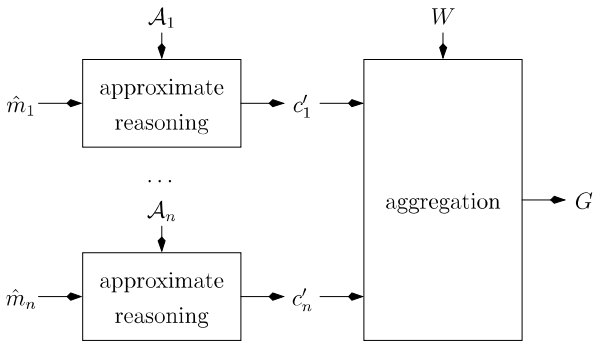


Figure 9: domain knowledge overview (Mees et al., 2016).

In summary, once the parameters extracted from the scenario have been identified, the proposed 3D visualisation provides the judgement elements instantly to intuitively know the risks associated with the mission and, ultimately, provide a quick overview about the cyber defense situation. This initial research shed some light on the possibilities about decisions based on visualisations in an attempt to bridge the gap between technical aspects and mission-centric analyses of a given cyber defence context. The 3D visualisation can be organised as follows: firstly, display images that encode the relevant high-level parameters using a simple approach that facilitates the recognition of cyber assets and resources that require a posteriori assessment. Secondly, the distribution of forces using the MAC triangle

describes for each cyber defence asset those mission key factors that are deemed critical with a value that is determined by the sum of some low-level but scenario-based metrics.

At the end, the suitability of low-level measures is experimented through numerical calculations that uses an aggregated factor which conforms a priori expert judgement. Although not specifically addressed along the explanation of the methodology, the visualisation would be interoperable with a dynamic evolution of the cyber situation, the configuration of assets or the emergence of new risks. If properly computed, metrics can change to adapt to new circumstances. A preferable management system to deal with this rapidly evolving cyber threat landscape would be a command and control (C2) system integrated in a SOC or connected with a network operations centre (NOC) - a supporting element that provides network monitoring and performs network operations. The design should follow the principle of ease of use aimed to reduce the cognitive workload of operators in an increased data rich landscape. This simplicity may assist in a quick implementation. The existing efforts are more devoted to the integration of a suite of tools which can offered distinct network and security services, being the visualisation part only a layer which is fed by data coming from various sources. The ultimate goal is to discern what is important over a massive amount of information. This characteristic would be exploited by a planning team to propose courses of action tailored to mission needs. There is no limitation in the proposed mathematical model to address simultaneous activities and their dependencies to calculate risks. An interesting line of research would consist on determine which cyber assets contribute to obtain a capability – considered at a higher abstraction level. Instead of showing cyber assets, the visualisation would provide a translation mechanism to understand the impact on capabilities by transferring the risks. Moreover, a M&S environment is becoming essential to perform an evaluation of the results by a team a mission planners. A further experimentation may leverage the results including its potential integration with the cognitive computing SOC framework described in chapter 4. This research was carried out during my research stay at the Belgium Royal Military Academy (Mees et al., 2016).

(Llopis et al., 2018) provide a comparison of incident handling visualisation methods utilising the operational picture concept. Authors predicted that the classification of data using AI algorithms will improve visualisation strategies and aid in decision-making. Visual tools must be accompanied by a full CySA solution in order to help operators with technical

issues which could be increasingly automated or even smartly managed. Pictures generated might individually or together reflect technical data or mission-relevant data. The authors' research aims to test if various existing visualisation techniques may result in various levels of knowledge and understanding of cyber-situations. Aspects of complementarity are highlighted to discover further consequences in case of progressing their maturity towards a prototype. Unquestionably, a mapping exercise of respective functionalities derived from an analysis of both visualisations is key to establish common grounding. A graphical representation must be contrasted with user demands. The reality that operators and military commanders view differ due to human factors like expertise when they are faced with stressful situations that call for quick response actions. Approaching useful visualisation approaches that might increase SA, particularly during incidents, is of great value. An information baseline able to reflect the reality of a situation is referred as "ground truth," which is the greatest awareness threshold.

In addition to the underlying human-related variables that are the focus of cognitive research, tested technology can also contribute to assist network administrators in routine tasks. While creating a visualisation approach based on a broad cyber defence framework, two important concerns arise: (1) which are the best practices to accurately define user requirements? and (2) how to evaluate the suitability of the visualisation to increase user performance? To meet the demands of the operators, a protracted iteration process is anticipated. Also, a potential architecture's technical components must have data connections with multiple repositories, for supporting decisions.

A comparative analysis was made between CyCOP (Esteve et al., 2016) and the previous mentioned three-dimensional COP. CyCOP developed various interfaces to receive meaningful information in order to provide a cyber hybrid SA. Using its connectivity to a set of tools and data repositories is able to provide a timely data processing including a geographical context for cyber assets (See Figure 10). The following conclusions were drawn from a qualitative investigation that examined the pertinent features of each visualisation technique: (i) the two visualisations work well together. A military commander may acquire real-time data from external interfaces implemented by CyCOP, such as C2 systems, Open Source SIEM (OSSIM), and Malware Information Sharing Platform (MISP), and portray mission-critical features from a 3D operational picture. An external service provider or data source imports a vulnerability/threats assessment; (ii) various perspectives

(representations) contribute to meeting various user needs. Both tools meet the visual needs of a decision-maker and an operator or technical staff member. In order to give operators freedom in their reporting, CyCOP provides various templates for communicating results or to perform analysis about the cyber situation.

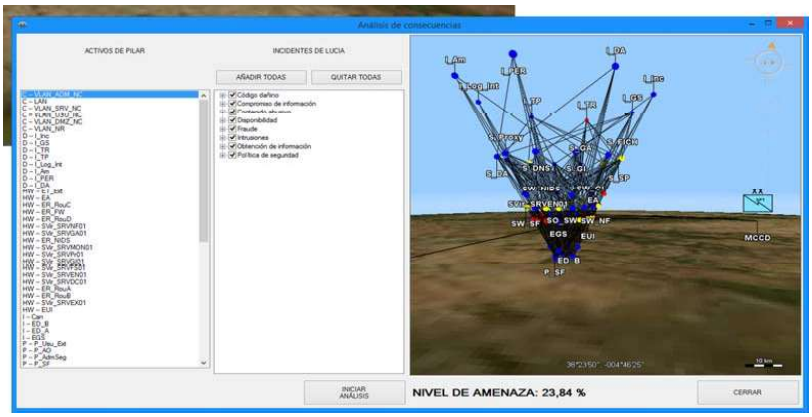


Figure 10: CyCOP views.

(iii) the continuation of a narrative is crucial for evaluating prior judgements and the success of remedial activities; (iv) one of the important advantages of CyCOP is its granularity, which allows users to pick which information to depict, which visualisation approach to use, and where to display data as key features of the system; (v) there is a significant problem with the solutions offered by comprehensive decision-support systems for achieving an improved CySA. The ultimate objective of a visualisation is to support technical workers and decision makers in their understanding of the cyberspace; (vi) mission focused. The method used determines the mission's level of criticality. In that way, a military planner provides the connections between the kind of operation and the cyber assets used. As a result, the representation needs to be simple to understand and allow for the quick identification of risks. Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 2018) is a technique which supports experimentation of the human cognitive aspects. In this particular case, SAGAT must be tailored to the training audience and change some of the features. Information categorisation and decision-making are two other areas where visualisation tools for cyber situation awareness may be improved. (MacQueen, 1967; Scholkopf et al., 1998) suggested the employment of artificial intelligence to boost performance. According to (Bertini et al., 2010), automatic refining of visualisation is a promising technique that uses artificial intelligence to analyse and classify data based on risk levels. The possibility may increase incident management effectiveness and save up time for decision-making.

Page intentionally left in blank.

Chapter 4
Design and Implementation of a proposed Architecture

Page intentionally left in blank

4.1.The Next Generation Cognitive Computing Security Operations Center

This section describes two novel approaches and methodologies to build a smart *automated SOC* which is the result of applying complex algorithms in an incremental manner. During my PhD research, I contributed with a research team to produce two publications (Demertzis et al., 2018), (Demertzis et al., 2019) about a cognitive SOC. Both publications explain in more depth the mathematical analysis and results. My contribution to these researches were in the validation, formal analysis, review and editing. The proposed architectures, their validation and experimentation are subject of discussion in chapter 4 and 5.

A SOC is a team of skilled professionals that employs cutting-edge malware detection solutions to proactively avoid, discover, and handle cyber-attacks within a specific organisation. In general, SOC's efficacy is largely contingent on its ability to scrutinise and analyse a significant volume of information and to link various cybersecurity occurrences. The oversight and classification of data is a crucial step that enables the planning, administration, and control of computer networks, as well as identifying and investigating potential cybersecurity breaches. Additionally, a SOC is accountable for continuously observing, examining, evaluating, and protecting an organisation´s security infrastructure. SOC personnel has different cyber security profiles from highly specialised malware reverse engineers to security administrators. They are responsible to ensure the timely detection, analysis, and resolution of security issues. The SOC employs advanced applications which must be tested and certified prior to enter into service. These mentioned tools aid in the detection of security threats.

However, commercial-off-the-shelf tools that assist in the network analysis or in the remote configuration of hosts and security perimeter devices have significant drawbacks. One of the consequences of these drawbacks is that they are often resource-intensive, particularly for the classification and inspection of ciphered traffic, necessitating the reconstruction of data packages inside complex classification protocols. Moreover, these software packages exhibit an increased incidence of misclassification brought on by a deficiency in accurate prediction mechanisms. Consequently, these traditional approaches are often inadequate for identifying unknown vulnerabilities or zero-day attacks. A set of available practices used in responsive cybersecurity depend upon the knowledge and interpretation of practitioners to determine the potential effects and minimise the attack vectors. The thesis proposes the employment of cybersecurity procedures to overcome identified shortfalls in an attempt to improve the

current situation of inconsistent and repetitive applications by streamlining the required functionalities in combatting threats. These proposed frameworks - to be used by modern and highly automated SOCs - are designed to detect and propose remediation actions according with identified threats proactively and timely, using advanced analytics, machine learning, and automation technologies. This particular type of *SOC* employs a subset of algorithms including other advanced analytics to retrieve meaningful insights from inspected data. This approach enables the SOC to identify new and emerging security threats, as well as to respond quickly and effectively to detected incidents. In summary, the proposed cognitive computing SOC is an advanced cybersecurity resource that combines proactive monitoring, real-time threat detection, and automated response capabilities to counter the evolving threat landscape.

4.1.1 Network Flow Forensics

The proposed Network Flow Forensics Framework (NF3) leverages advanced, fully automated intelligence methods to optimise poor power resources and computational capacity utilisation. A framework is designed for the next wave of efficient and fully-automated SOC - formulated as Next Generation Cognitive Computing SOC (NGC2SOC) - and represents a combination of ML approaches to perform an in-depth analysis of network data, malware detection and discovery of obfuscated content. In order to better understand the complex information environment that a SOC is facing in its daily operations, it is worth to highlight the following aspects:

1. Capturing, examining, and reviewing network traffic flow to manage network services, ensure security, and optimize performance is known as network traffic analysis. Primary methods for handling data analysis in a computer network encompass: the payload-based classification strategy, which involves sorting the data flows and protocols to be analysed according with payload attributes associated with ISO layers 2-4. It means to consider medium access control (MAC), IP address and source/destination ports including an empirical packet analysis to classify data based on interpacket arrival, session, timestamp, and other parameters.

2. Malicious software colloquially known as malware may gain unauthorised access to network infrastructures, disrupt computer operations and facilities, and collect personal information. Malware is able to exploit software or system vulnerabilities. This includes exploiting weaknesses in the source code that a program may use to handle events. Once the vulnerabilities are exploited, the malware can gain access to sensitive data, take control of the affected system or perform other malicious tasks. Malware often use sophisticated techniques to obfuscate and remain concealed, including the use of tactics such as changing file attributes or pretending to be legitimate to avoid being detected by antivirus software or other security measures. For instance, malware may have a name that resembles a legitimate file or application, making it more difficult to identify as malicious. Moreover, malware often tries to evade detection by hiding its processes, network connections and communications from dubious registry values or uniform resource locators (URLs). The use of ciphering tools is becoming widespread to conceal cybercriminals´ activities. This encryption hinders the analysis of malware behaviour and increases the difficulty of detecting it. Malware is specifically created to remain hidden for long periods, enabling cybercriminals to establish control over the infected system and communicate through encrypted channels with their command and control servers. This allows for various malicious activities such as data theft. Thus, implementing adequate and strong cybersecurity measures that prevent malware infection, detect and eliminate any malware that does penetrate the system, and keep operating systems up to date is essential.

NF3 is specifically designed for a SOC that solely relies on dynamic automated processes driven by software algorithms. This framework comprises an efficient and precise group of ML solutions in view of analysing network flow in real-time, using low processing calculations and means capable of producing fast identifications over encrypted and malware traffic. The mentioned framework is based on a new wave of smart systems which make use of a fusion of ML models, incorporating four different algorithms, including Support Vector Machine (SVM) (William et al., 2007), Artificial Neural Network (ANN) (Hubel et al., 2005), Random Forest (RF) (Breiman, 2001), and k-Nearest Neighbours (k-NN) (Hall, 2008), to examine data logs to detect the presence of malware in the system of study. The adoption of ensemble techniques is motivated by the complex multifactorial nature of the

problem being considered, which requires the tolerance of the intertwined models to analyse and solve. Furthermore, the ensemble model is highly effective in expressing the numerical modelling of data traces that disclose complex dependencies, like those that differentiated usual data logs from suspicious network activity. By utilising a combination of four ML models, the data analysis becomes streamlined, facilitating cyber resilience and hastening the merge of the proposed algorithms into a unique and less disruptive platform. Besides an overfitting risk mitigation strategy enabled by ML, the proposal promotes generalisation. Consequently, this proposed network forensics framework provides an innovative approach to the detection of malicious activities on the network. By utilizing an ensemble architecture that combines several machine learning algorithms, the framework enhances the accuracy of the system, provides a generalisation that avoids overfitting, and ensures the sensitivity of the overlapping models used to identify malicious traffic.

The particularities of each of the algorithms employed in the ensemble construct can be described as follows:

- SVM is a supervised machine learning algorithm that is often used for classification and regression analysis. SVM is based on the concept of finding a hyperplane that best separates the data points of different classes in a high-dimensional space. This algorithm is particularly useful when working with datasets that have a large number of features, as it can efficiently classify data by mapping it into a high-dimensional space.
- ANNs are machine learning models that are inspired by the structure and function of the human brain. ANNs consist of interconnected nodes, or "neurons," that process and transmit information. Each neuron receives input from one or more other neurons, processes that input, and then sends output to one or more other neurons. ANNs are often used for classification and regression analysis, and can be used to recognize complex patterns in data.
- RF is an ensemble machine learning algorithm that constructs a multitude of decision trees during the training phase. The algorithm randomly selects subsets of features and data points to build a set of decision trees. During the testing phase, each decision tree predicts the outcome, and the forest outputs the most frequent prediction. The random forest algorithm is known for its ability to minimise overfitting and augment accuracy.

- k-NN is a non-parametric machine learning algorithm that can be used for both classification and regression analysis. The algorithm works by identifying the k-nearest data points in the training set to a given data point and using those neighbours to make a prediction. The value of k is chosen by the user and can affect the accuracy of the algorithm. k-NN is often used in data mining and pattern recognition applications.

The ensemble architecture used in the proposed NF3 combines these four machine learning algorithms to increase the system´s precision and resilience. It becomes straightforward that different algorithms complement each other, and the ensemble model can produce more accurate predictions than any single algorithm alone. Furthermore, the ensemble model helps to mitigate the risk of overfitting and promotes generalisation, which is critical for machine learning applications.

Network flow analysis software is essential for detecting cyber threats and malware communications. However, such applications have limitations that may affect the accuracy of the analysis. One significant drawback is the lack of access to more elaborated analysis and inspection on the data attributes, since these applications are not able to check all the features and produce the level of detail necessary for an exhaustive and complete overview. Another critical issue is the accuracy of the interpretation, which relies on the sampling rate chosen. Not only the frequency in obtaining samples in a valid criterion that leads to accurate analysis, but also the sample form used influences the results. Commercial applications manage differently the specifications for sample rates (Demertzis et al., 2016). Moreover, there is a variety of network protocols that should be feasible to analyse by a network flow forensics capability. This includes the bandwidth overhead and computer resource demands to conduct examinations, which at the end may cause some system resources adjustments (Demertzis et al., 2014). Additionally, when analysing big data, technicians are supported by graphics, and the meaning of the displayed information displayed is directly linked with the user´s expertise. Therefore, there is a need to further investigate human factors in order to increase the performance of CySA visualisation products usually integrated as a key component to enable decision-making as shown in chapter 3.

One of the significant challenges of network flow analysis applications is the use of signatures to identify threats. While signature-based malware identification can recognise

well-known events, up-to-date malicious code may be visible and not masked from normal network activity by employing behavioural tests on top of other phased techniques (Demertzis et al., 2015). The following is a description of how advanced persistent threats or other sophisticated malware operates. New forms of malware while resident in networks maintain an external connectivity with a centralised control using data burst to send discreet pieces of information. These communications serve the purpose of updating the payload or amend initial instructions if so required. The security network perimeter is composed by IDS/IPS which may overlook the existence of malicious activity due to the extensive use of dynamic DNS (Yadav et al., 2012). Moreover, malware programming code may obfuscate certain rules for ciphering information which together with Blind Proxy Redirection (BPR) technique – it redirects traffic through a proxy server capable of intercepting and analysing network traffic – make hard to detect active C&C servers. Demystifying malware traffic is, therefore, one of the suitable instruments that prevents the occurrence of cyber-incidents and with the goal to conduct an exhaustive screening of suspected connections. It is a valuable procedure to obtain indicators of compromise derived from the analysis of malware tactics.

Researchers have proposed various methods to detect and trace botnets and identify encrypted traffic in networks. (Hsu et al., 2010) developed a fast reaction method that uses anomaly detection to inspect thoroughly the hypertext transfer protocol (HTTP)/secure (HTTPS), which has produced notable results. Additionally, (Haffner et al., 2005) used several ML algorithms to classify the secure shell (SSH) protocol, albeit with reduced characteristics of the workload. (Alshammari et al., 2007) suggested a fusion model which accurately classifies SSH data avoiding to extract parameters from the load. (Holz et al., 2008) researched a concise procedure concise for tracing advanced malware, while (Almubayed, 2015) introduced a process for measuring effectiveness of programming code to detect obfuscated data flows. These studies are essential for developing effective techniques to detect and mitigate cyber threats. By leveraging different detection methods and tracing techniques, it is possible to enhance network security and ensure protection against malicious activities. NF3 assists a SOC operator in automating specific tasks. A cutting-edge NGC2SOC should include a combination of diverse algorithms, immediate updates, graphical applications, and advanced techniques to reduce network vulnerabilities to key resources including the exercise of regular backups to return to normality in case of a major disruption caused by cyber-attacks. The advantages over other approaches (Mercaldo et al., 2017), (Montieri et al., 2017) is that NF3 minimises overfitting at the same time that

ensures efficiency in computational data processing. It proposes an ensemble ML model that produces a parametrisation of network data samples, concurrently checked by each model in order to compute an empirical aggregation of the findings. Organisations can leverage these capabilities to effectively monitor and secure their networks and mitigate the risks to critical assets.

The NF3 architecture involves a dynamic traffic character recognition method in Stage 1. (Figure 9) shows that each learning algorithm concurrently checks these features in Stage 1 resulting in an aggregated score of the cluster as reflected in Stage 2. An ensemble averaging model calculates the average predictions of each instance that forms part of the experimentation data collection. The proposed analysis involves determining whether the network traffic is normal or suspicious (network traffic analysis). If this test yields a confirmed output in Stage 3 and the network flow is deemed suspicious, then it will be subject of further analysis to determine the specific type of suspicious origin (demystification of malware traffic) that occurs. Normal traffic will be subjected to additional checks in Stage 4 to identify any instances of ciphered data (encrypted traffic identification) and determine the type of communication being used e.g. SSH, SSL, P2P in Stage 5. In cases where the traffic is not encrypted, the analysis will identify the specific application e.g. network, transport protocol being used in Stage 6. This approach allows for a comprehensive analysis of network traffic that can assist to detect and mitigate abnormal activity. By leveraging machine learning algorithms and advanced analytics, SOCs can develop effective cyber defence mechanisms that can automate the restoration of cybersecurity issues and minimise the risk to critical assets.

NF3 implementation involves optimal use and fusion of high-performance intelligent software models to build a leading-edge combination of ML methods to deal with cybersecurity problems. A combination of diverse algorithms involves a seamless integration of multiple ML techniques to create a single one, more effective predictive method. This can be done to decrease variance (using bagging), reduce bias (using boosting) or enhance the accuracy of predictions (using stacking) (Bonab et al., 2016). Combining multiple methods offers several advantages, including increased consistency, improved forecasts, and the ability to generalise to new and unseen data. This is crucial for ensuring that machine learning models are flexible to changing circumstances. While using a combination of predictive models may not always result in the highest level of effectiveness,

it can greatly reduce the occurrence of poor results and help ensure more reliable outcomes overall. However, a detailed examination of the ensemble model's elements, structure, and critical decision points is essential to optimize performance and ensure effective use of the approach (Zhou, 2012). This can help to identify the strengths and limitations of the approach, and guide decision-making for effective cyber defence mechanisms.

The effectiveness of a fusion method depends greatly on the number of predictors used in its construction. For the implementation of the NF3 model, four techniques have been employed based on design principles established by the "law of minimising costs in combination framework" (Kuncheva, 2004), (Dietterich, 2001). In other words, while combining several ML models can lead to improved performance, there comes a point where adding more models does not result in a significant improvement in performance. This is due to the fact that the marginal gain obtained by each additional model decreases as the number of models increases. Therefore, it is important to strike a balance between the number of models used and the expected performance gains. The final ensemble size was determined using a trial and error method and statistical tests. When selecting appropriate predictors for an ensemble method, the settings and configurations of the restrictions should be taken into consideration to account for different decision boundaries.

Choosing algorithms based solely on their performance with minimal inaccuracies in learning activities is not always ideal for creating combinations. This is because effectiveness on learning may not accurately reflect a prospect on the algorithm performance over never seen traffic (Webb et al., 2004). To ensure effective selection of individual classifiers, they should exhibit a specific degree of variety and employ several workable parameters and training data sets, enabling the creation of different decision boundaries that will work in common to minimise the complete inaccuracies.

The selection of predictors for the NF3 model was founded on a probabilistic strategy that took into account the key features of the algorithms and the way different situations are handled. For example, parametric models such as ANN parametric and non-parametric Kernel SVM may be suitable, and techniques like RF may assist to mitigate the impact of exceptions or excessive quantities. k-NN has some advantageous properties for handling noise, including the ability to recognise regular or irregular distributed data and get successful outcomes with a bunch of data variables.

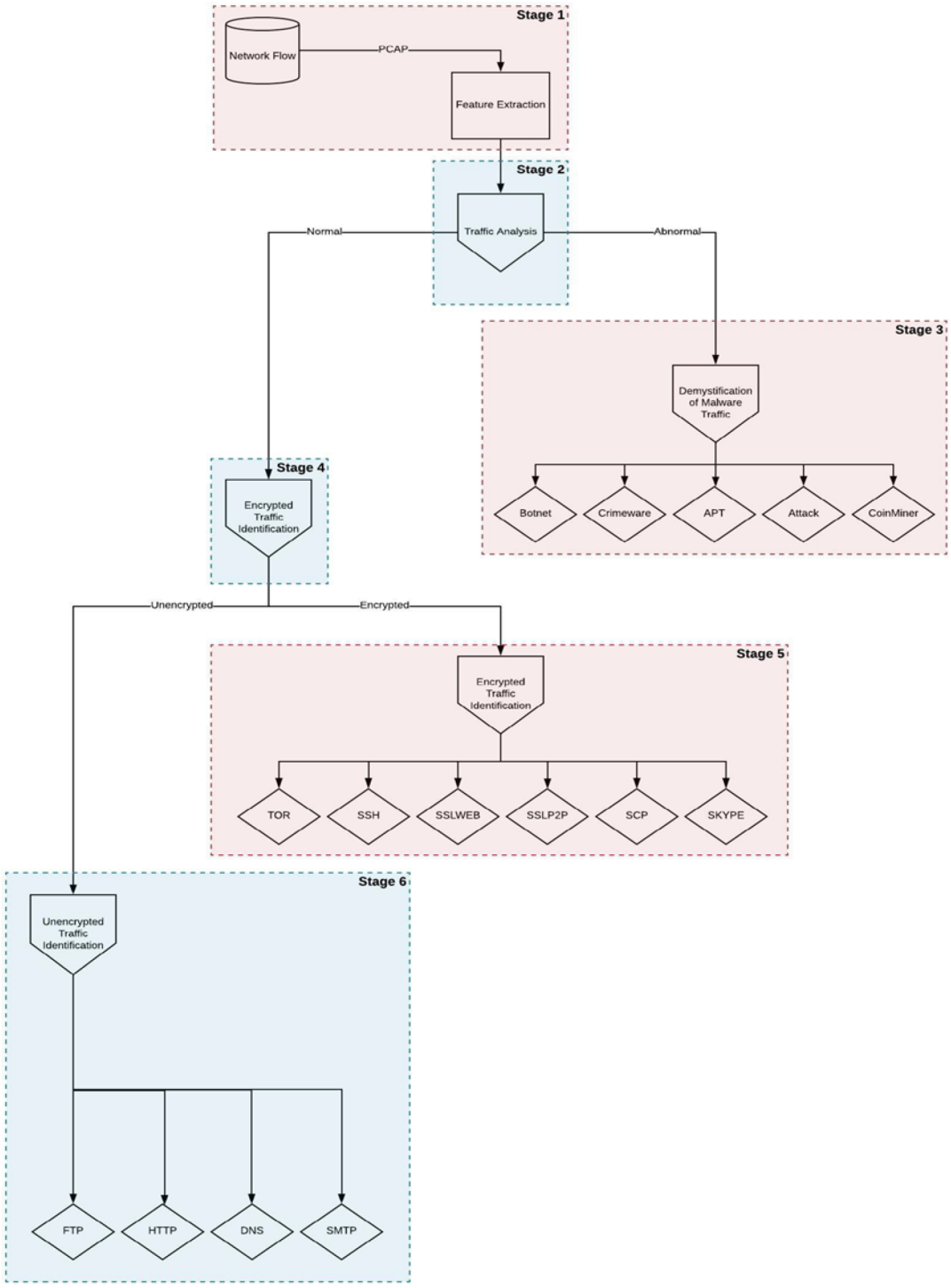


Figure 11. Description of the phases of the NF3 model (Demertzis et al., 2018).

4.1.1.1 Configuration of the NF3 Ensemble model

To optimise an ensemble model, an analysis should be conducted to determine the weights of the algorithms used. The weight vector plays a crucial role in the learning procedure of a combined method, since it evaluates predictors' reliability and classification's

trustworthiness. Higher weights raise questions about their function for specifying a procedure of ensemble learning algorithms and establishing the entire method´s trust. A common process is to use equal scores in the ML techniques and average the predictions (Tsoumakas et al., 2005), arguably an empirical process but exempted from academic rigour which is not based on scientific evidence. In creating NF3 framework, the determination of the variables assigned to the algorithms is specified determined through a quantitative test and failure procedure. This method considers various factors such as how algorithms handle different situations, including parametric models, non-parametric models, outlier and noise handling techniques. The selection process of predictors in the ensemble model considers the fundamental features of classifiers and how every use case is managed. These factors ensure that individual classifiers show a specific degree of variety and that different decision boundaries are created to minimise holistic inaccuracies.

The precision of a machine learning model's forecasts is a crucial attribute for validating the trustworthiness of a combination method. A slight difference in performance between forecasts is a significant factor in measuring the trustworthiness and consistency of the combined method. Consistent prediction solutions with low dispersion rates increase reliability, while high dispersion rates suggest a higher degree of uncertainty in the final forecast. The ideal scenario is for the expected error to be concentrated around an average error value. To increase the method classification performance, it is not recommended to use fixed categorizers for creating an ensemble predictive model. Instead, diversity among the chosen classifiers, including various settings, configurations of attributes, and learning processes, is an essential reliability factor for the ensemble model. In creating NF3 framework, several classifiers are to be chosen linked with their role model, including their settings, that utilised various frameworks, configurations of features, and learning processes.

Below is a short summary of the algorithms employed in the ensemble framework (Demertzis et al., 2018) provide more details about the specific determination and usage parameters for each algorithm.

1. SVM is a type of classifier that builds hyperplanes in multi-dimensional space to create decision boundaries between different classes. SVM assumes that data is linearly separable and uses a repetitive learning processes for creating optimal variables that minimizes the inaccuracies and maximizes the profitability, subject to

a collection of regular limitation. The process is interpreted like an optimisation issue, which is solved by a polynomial function. When dealing with irregular measurements, the SVM transforms the data to a different parameter to achieve regular distance.

2. ANNs are computational models that mimic the functionality of the intellect. ANNs follow an irregular modelling approach able to manage vague problem formulations. It uses Back-Propagation (BP) as a learning model to train the algorithm. BP calculates the necessary attributes for the various neural layers for reducing inaccuracies in the results.
3. RF is a ML algorithm used for prediction and classification. It creates multiple decision trees by randomly selecting subsets of features and data from the original dataset. RF uses a learning model based on bagging – tree-learning process, where the samples are repeatedly drawn from the dataset. Predictions of new data are made through a mathematical calculation of the forecasts. This process helps in improving the forecasts' precisions and reducing the variance of this method, while still maintaining a good level of interpretability. The RF algorithm is widely used in many domains, including finance, healthcare, and cybersecurity.
4. k-NN model can be used when the probability distribution of data is unknown or difficult to determine. It involves searching for the k nearest neighbours to a new data point within the training dataset, using the Euclidean distance function to quantify the proximity between the test sample and each of the learning samples. A common label among the k neighbours is then assigned to the novel source of information. The parameter k is determined by the user and is a crucial factor in the classification accuracy of the algorithm. If k is too small, the algorithm may become too sensitive to noise in the data, whereas if k is too large, the algorithm may lose its ability to discern between classes.

4.1.2 Adaptive Analytical *λ*-Architecture in support of cyberdefence

The λ-Architecture Network Flow Forensics Framework (λ-NF3) introduces some modifications to the previous mentioned ensemble machine learning model. This new framework incorporates the lambda machine learning architecture, which allows for the analysis of two data classes by employing using new ML models. First algorithm is Extreme Learning Machine neural network with Gaussian Radial Basis Function kernel (ELM/GRBFk), which is used for batch data analysis. The second algorithm, Self-Adjusting

Memory k-Nearest Neighbours classifier (SAM/k-NN), examines real-time data streams to identify patterns. The λ-NF3 framework is intended to enhance SOC's automation to better cope with real-time incident handling. Although in the author´s opinion, the man-in-the-loop should always exist to validate findings and make decisions. Perhaps this is one of the reasons to reflect on the convenience of a "hybrid posture" combining expert systems with artificial intelligence techniques. This model also addresses the same "desired end state" of achieving a NGC2SOC.

In summary, ML comprises methods for creating complex models and frameworks which make reliable and reproducible judgements and uncover underlying trends via training past information. Machine learning methods are reactive to minor variations in input or transformations (Davi et al., 2004), and they are sensitive to adversarial examples. Adversarial examples are inputs that have been intentionally crafted to mislead the machine learning algorithm, making it classify the input incorrectly. This is a major security concern, and it is utilised by attackers to compromise ML methods. Adversarial attacks manipulate input data or the weaknesses in the ML algorithms to impair the protection of the network. For instance, a neural network algorithm may assign which group a specific data point belongs by establishing links with a learning compedium of known data points. However, if the input data is modified, it can lead to a wrong classification. Neural networks are reactive to overfitting, overly regular, and marked by the ambiguity of their forecasts.

Understanding the protection aspects of training models in environments with an opponent requires addressing several important issues. Firstly, it is essential to identify possible threats on ML models during training and categorisation. Secondly, suitable cyber-attacks need to be designed in accordance with the existing vulnerabilities and their risks to compromise a network should be evaluated. Finally, countermeasures must be proposed to enhance the protection of ML models to overcome cyber-incidents already analysed. Two primary defence strategies are typically employed to face complex cyber-attacks. On one side, the responsive procedure that involves constructing a new algorithm using a variety of execution modes and restriction parameters that could result in a variation of the limits to decide, while keeping other limitations constant. Diverse classifiers with various functional settings and different training sets should be selected to create different decision boundaries, which can then be grouped to minimise the estimated inaccuracies. On the other side, is the preventive approach, that trust in deploying appropriate preemptive learning to establish precise

decision boundaries. Investigating the learning method is critical in discovering the optimal weights. This weight vector is a critical parameter, used to define the reliability of algorithms and the trustworthiness of the information processing procedure. Higher weights could have an essential role in regulating the type of limitations in the mock-up. Therefore, it would be paramount to research cyber resilient ML models to deal with identified weaknesses.

λ-NF3 stages are similar to the ones that NF3 proposes for efficient network traffic analysis, suspicious data flows demystification, and ciphered data recognition with the goal to enhance cyberdefence against adversarial attacks. The λ-architecture was chosen for its ability to handle multifactorial problems of great difficulty with numerous traffic samples. This implementation follows a responsive cybersecurity procedure by learning from two opposing algorithms to identify possible attacks and reject them. Furthermore, the explained framework provides fast training, minimal complexity, reduces human footprint, and employs reduced computing processing and means.

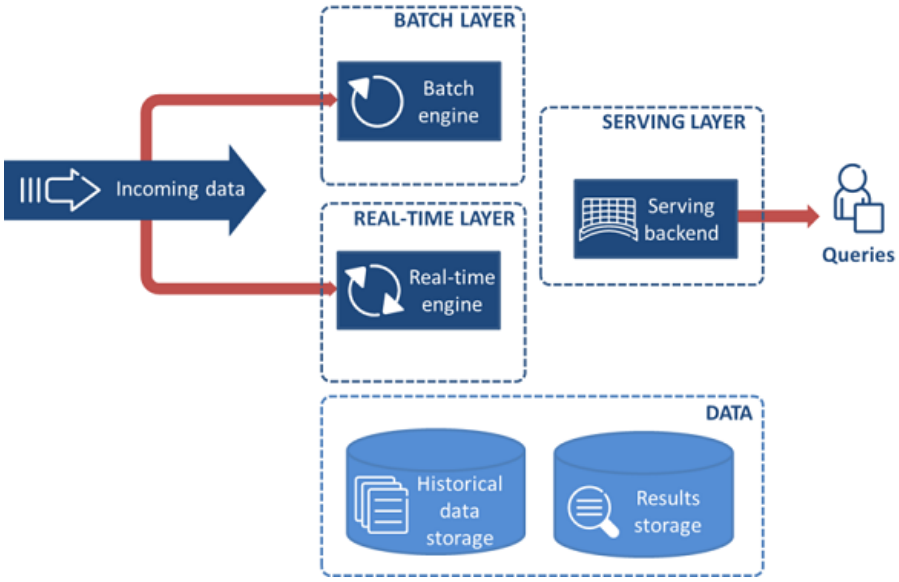
- ELM/GRBFK is a type of artificial neural network that can be used for regression or classification tasks. The ELM algorithm uses a single hidden layer feedforward network, and the weights between the input layer and the hidden layer are randomly initialized. The output layer weights are then analytically calculated to minimize the error between the predicted and actual output. GRBF kernel is used to transform the input data into a higher-dimensional space, where linearly inseparable data may be linearly separable. This is done by computing the distance between the input data and a set of centres, and using the distance as the input to the GRBF.
- Additionally, SAM/k-NN is a type of instance-based learning algorithm that classifies new data based on its proximity to labelled examples in the training set. SAM/k-NN is a variant of the traditional k-NN algorithm that uses a self-adjusting memory to adapt the size of the neighbourhood used for classification. This memory is used to estimate the true error rate of the k-NN algorithm, and adjusts the size of the neighbourhood based on the estimated error rate. The SAM/k-NN algorithm is particularly well-suited for streaming data, as it can adapt to changes in the distribution of the data over time.

4.1.2.1 Configuration of the λ-NF3 model

To effectively process large amounts of data, two distinct approaches can be used: batch processing and stream processing:

- Batch processing involves processing a large amount of data all at once, typically collected during a specific time period. This type of processing requires significant computational power and hardware infrastructure to handle the processing and analysis of these large datasets. However, it can be concluded or scheduled during off-peak hours to avoid wasting system resources and to increase overall utilization rates.
- Data streams are generated continuously and in real-time from multiple network infrastructures, such as sensors and IoT devices. Stream processing involves handling the data in sequence and incrementally, either by processing it over sliding time windows or by using specialized data processing techniques to extract hidden knowledge. Stream processing is used for quick study and information retrieval and consultation environments.

Machine learning algorithms can be used for both batch and stream processing. Overall, batch processing and stream processing are important approaches for processing large amounts of data, and ML models can be used to effectively analyse and extract knowledge from these datasets in both cases. The extraction of real-time information from large network flows presents a challenge, which includes consumption, storage, and modification procedures for large data volumes. Unfortunately, analysing huge quantities of information requires time and is unable to be performed in timely manner, necessitating a significant amount of data warehousing to store the results of queries for future use. This introduces latency, which can be mitigated by using the λ-architecture, which provides two inspection lines for analysing network logs. A batch layer (cold path) analyses incoming original information and performs batch treatment on it. This analysis's results are saved as a batch report. In real-time, a speed layer (hot path) examines unlimited data flows, trading off accuracy for low latency (Yamato et al., 2016). The λ-architecture balances delay, efficiency, and high availability by employing the cold path for full and precise representations of past data and the hot path for live data stream processing of incoming information. The two projection outputs can be merged to enhance the process and add accuracy to the entire model. During the initial step, the lambda-NF3 computational technique incorporates the characteristic retrieval process from data traffic. In the second step, both algorithms evaluate such properties to reduce the potential of being fooled by adversarial attacks. The outputs are aggregated with a preference towards the cold path (batch processing) for a better audit trail.



Source: Ericsson

Figure 12. Building blocks of a λ-framework

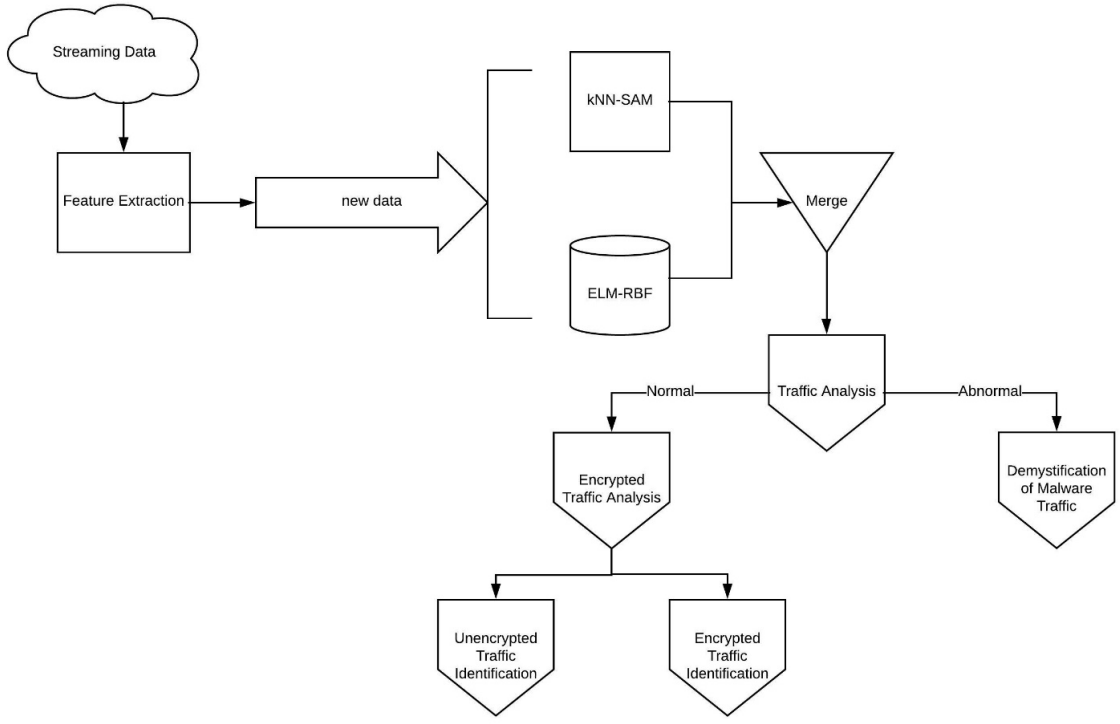


Figure 13. λ-NF3 phased approach (Demertzis et al., 2019).

The λ -NF3 proposes the integration of various advanced algorithms with diverse operating modes, configurations, architectures, and training techniques. This coupling of algorithms requires different implementations and hyper-parameter settings. The features of the algorithms are as follows:

1. An Extreme Learning Machine (ELM) is a type of Single-Hidden Layer Feed Forward Neural Network (SLFFNN) that consists of N hidden neurons (Cambria et al., 2013). ELMs are notable for their unique feature of randomly assigning input weights and bias in the hidden layer (Huang et al., 2014). This random initialization allows the ELM to achieve a much faster training speed than traditional SLFFNNs while still achieving high accuracy in many applications. Additionally, the ELM has been shown to have superior generalisation performance and good scalability.
2. The SAM/k-NN is an AI model that mimics a human brain, particularly the short and long-term remembering (Losing et al., 2016). The short-term memory (STM) is responsible for keeping information for a short period of time, while the long-term memory (LTM) stores information indefinitely. The information from STM is transferred to LTM through the memory consolidation process, which involves recurrent reactivations that encode memory information, leading to the integration of new knowledge. In this process, knowledge transforms over time, becoming a permanent memory in LTM. The SAM architecture is partly inspired by this biological memory model, and it is used for time-series analysis. For example, the general statement of new inputs (streaming data) is more related for current estimates that can be associated with temporal trends or time-based events. On the other hand, the batch processing from historical data can lead to much better prediction results while offering generalization. SAM helps the algorithm to remember the previously seen data by assigning weights to each data point based on their relevance, thus improving the accuracy of the classification.

Chapter 5
Architecture validation and experimentation

Page intentionally left in blank

5.1 Architecture Validation

The validation of the NF3 models described in chapter 4 were executed independently by using specific datasets with the following results:

5.1.1 Results of the NF3 Ensemble model and discussion

The below information reflects the results achieved for testing the NF3 method (Demertzis et al., 2018).

Table 3. Measurement of methodologies.

Network Traffic Analysis (Binary) (208.629 Instances)						
Classifier	**Classification Accuracy & Performance Metrics**					
	TAC	**RMSE**	**PRE**	**REC**	**F-Score**	**ROC_Area**
SVM	98.01%	0.1309	0.980	0.980	0.980	0.980
MLFF ANN	98.13%	0.1295	0.981	0.981	0.981	0.994
k-NN	96.86%	0.1412	0.970	0.970	0.970	0.970
RF	97.12%	0.1389	0.972	0.971	0.971	0.971
Ensemble	97.53%	0.1351	0.976	0.975	0.975	0.979

Table 4. Measurements of methodologies.

Demystification of Malware Traffic (Multiclass) (168.501 Instances)						
Classifier	**Classification Accuracy & Performance Metrics**					
	TAC	**RMSE**	**PRE**	**REC**	**F-Score**	**ROC_Area**
SVM	96.63%	0.1509	0.967	0.967	0.968	0.970
MLFF ANN	96.50%	0.1528	0.981	0.981	0.981	0.965
k-NN	94.95%	0.1602	0.970	0.970	0.970	0.950
RF	95.91%	0.1591	0.972	0.971	0.971	0.960
Ensemble	95.99%	0.1557	0.972	0.972	0.973	0.961

Table 5. Measurements of methodologies.

Encrypted Traffic Analysis (Binary) (166.874 Instances)						
Classifier	**Classification Accuracy & Performance Metrics**					
	TAC	**RMSE**	**PRE**	**REC**	**F-Score**	**ROC_Area**
SVM	98.99%	0.1109	0.989	0.990	0.990	0.990
MLFF ANN	99.12%	0.1086	0.998	0.998	0.998	0.998
k-NN	97.84%	0.1372	0.975	0.975	0.978	0.980
RF	98.96%	0.1107	0.989	0.989	0.989	0.990
Ensemble	98.72%	0.1168	0.987	0.987	0.988	0.989

Table 6. Measurements of methodologies.

Encrypted Traffic Identification (Multiclass) (214.155 Instances)						
Classifier	**Classification Accuracy & Performance Metrics**					
	TAC	**RMSE**	**PRE**	**REC**	**F-Score**	**ROC_Area**
SVM	90.31%	0.1906	0.905	0.905	0.906	0.950
MLFF ANN	92.67%	0.1811	0.930	0.930	0.928	0.960
k-NN	85.19%	0.2032	0.890	0.890	0.890	0.935
RF	91.56%	0.1800	0.920	0.916	0.916	0.930
Ensemble	89.93%	0.1887	0.911	0.910	0.910	0.943

Table 7. Measurements of methodologies.

Unencrypted Traffic Identification (Multiclass) (186.541 Instances)						
Classifier	Classification Accuracy & Performance Metrics					
	TAC	RMSE	PRE	REC	F-Score	ROC_Area
SVM	99.92%	0.1003	0.999	0.999	0.999	0.999
MLFF ANN	99.91%	0.1008	0.999	0.999	0.999	0.999
k-NN	98.98%	0.1020	0.989	0.989	0.990	0.995
RF	99.93%	0.1001	0.999	0.999	0.999	0.999
Ensemble	99.68%	0.1008	0.996	0.996	0.997	0.998

In this discussion about ML algorithms' results, it must be explained that the classification is a type of problem where the goal is to predict a categorical or discrete output variable based on input variables. In this case, the two main types of classification problems are binary classification and multi-class classification. Binary classification involves predicting one of two possible outcomes. On the other hand, multi-class classification involves forecasting one of three or more possible outcomes. The error can be measured using a likelihood frequency of all attributes (Mao et al., 2000), (Fawcett, 2006). In binary classification, the error is calculated using metrics such as accuracy, precision, recall, F-score and Receiver Operating Characteristic (ROC). Accuracy is the most simplified measurement and represents the ratio of correct predictions to the total number of predictions made. It is defined as:

Total accuracy (TAC) = (true positives + true negatives) / (true positives + false positives + true negatives + false negatives)

Precision (PRE) measures the proportion of correctly identified positive instances (true positives) out of all the instances that were classified as positive, whether they are true positives or false positives. It is defined as:

precision = true positives / (true positives + false positives)

Recall (REC) measures the proportion of correctly identified positive instances (true positives) out of all the instances that are truly positive, whether they were correctly classified as positive or not. It is defined as:

recall = true positives / (true positives + false negatives)

F-score is a harmonic mean of precision and recall and balances both metrics. It is defined as:

F-score = 2 * precision * recall / (precision + recall)

ROC measures the ability of the model to distinguish between positive and negative instances. It plots the true positive rate against the false positive rate at different classification thresholds and calculates the area under the curve. ROC ranges from 0 to 1, with higher values indicating better model performance. In multi-class classification, confusion matrix is added to the calculation of the above-mentioned measurements. Confusion matrix shows the number of true positives, true negatives, false positives, and false negatives for each class. Accuracy measures the ratio of correctly predicted labels to the total number of samples.

It is defined as:
accuracy = (sum of diagonal elements of confusion matrix) / (total number of samples)
Precision, recall, and F-score can be computed for each class separately using the confusion matrix. Macro-averaging or micro-averaging can be used to compute the overall metrics across all classes. Macro-averaging calculates the metrics for each class separately and takes the average across all classes. Micro-averaging aggregates the confusion matrix across all classes and computes the metrics from the aggregated matrix. RMSE stands for Root Mean Square Error and is a commonly used metric in ML for regression tasks, where the goal is to predict a continuous output variable based on input variables. RMSE measures the average deviation of the predicted values from the actual values. RMSE is computed as the square root of the mean of the squared differences between the predicted values and the actual values. RMSE is preferred over mean absolute error (MAE) because it penalizes larger errors more than smaller errors due to the squared term. This means that the model is penalised more severely for large deviations from the actual values, which is often desirable in many applications. RMSE is expressed in the same units as the output variable, which makes it easy to interpret. This makes it easy to compare the performance of different models and choose the one with the lowest RMSE. In summary, RMSE is a popular metric in machine learning for regression tasks because it measures the average deviation of predicted values from actual values and penalises larger errors more severely.

According with results, the combination algorithm shows an equal or minor deviation on the effectiveness using various network flows, in contrast with the algorithm that presents a better TAC. This circumstance does not diminish the optimistic results of the ensemble model in particular due to its advantages on decreasing consumption, data processing and overfitting. These characteristics permit to confirm the validity of the proposal to deal with

multifactorial issues which refer to problems that involve multiple interacting factors or variables, making them difficult to analyse, understand, and solve. These problems are characterised by their complexity, unpredictability, and non-linearity. In complex multifactorial problems, the factors or variables that influence the problem are often interdependent and interact with each other in nonlinear ways. This means that changes in one factor or variable can have unexpected effects on other factors or variables, making it challenging to predict the outcome of the problem. Solving complex multifactorial problems often requires the use of advanced analytical and computational methods, such as machine learning, network analysis, and simulation modelling, to capture the complexity of the problem and identify the most effective solutions. With regards to the rest of parameters, the ensemble models show as well a positive outcome in terms of precision (PRE) and recall (REC). They measure different aspects of the model's performance in predicting the positive class. In practical terms, precision measures the model's ability to avoid making false positive predictions while recall measures the model's ability to detect all positive instances, including those that are missed. The choice of whether to optimise for precision or recall depends on the specific problem and its associated costs and benefits.

In the simulation, the ensemble obtained high rates which permits to confirm that the technique is secure and reliable and yields significant outcomes. Acknowledging that F-score depends on the precision and recall, it can be stated that the values obtained for the ensemble method are close to the ones which obtained better scores. F-score is useful when there is an imbalance between the number of positive and negative instances in the dataset. In such cases, accuracy may not be a good measure of model performance because a model that always predicts the majority class can have a high accuracy even if it performs poorly on the minority class. F-score, on the other hand, takes into account both precision and recall and provides a more balanced evaluation of model performance. F-score is used when both precision and recall are important and need to be considered together. The ROC curve is a useful tool for evaluating the performance of a binary and multi-class classification model because it allows us to visualise the trade-off between sensitivity and specificity at different threshold settings. The scores follow the same tendency than the other metrics.

Therefore, tables 3–7 evidence that the combination algorithm has quite strength in comparison with individual models. The scores are pretty close which it means that the performance would be quite similar but more robust. Returning to the visualisation

techniques, the NF3 detection method fits well with the three-dimensional representations in support of the NGC2SOC (chapter 3) without excluding other type of graphical interfaces. At the end, they are tools to assist the effectiveness of operators or administrators in a highly human cognitive demanding environment that requires quick reactions. Indeed, the model can be enhanced by the integration with vulnerability databases or cyber threat intelligence information like MISP.

NF3 framework is composed on ensemble ML algorithms that may assist to transform human related tasks to an automated NGC2SOC. The benefit of the NF3 framework is to reduce the cognitive burden to human operators but more importantly to accelerate the transformation to a high-performance SOC able to interact at the speed of relevance in cyber defence. NF3 may aid organisations to adapt its cyber posture to counter cyber threats. In general, a SOC performs several key functions that includes:

- Threat detection: The SOC uses a variety of security tools, such as intrusion detection systems (IDS), security information and event management (SIEM) systems, and network traffic analysis (NTA) tools to detect potential security threats in real-time.
- Incident response: When a security incident is detected, the SOC team works to investigate and contain the incident, including identifying the scope and severity of the attack, containing the damage, and restoring normal operations.
- Vulnerability management: The SOC team regularly assesses the organisation's systems and networks to identify vulnerabilities and recommends steps to remediate them.
- Threat intelligence: The SOC team continuously monitors threat intelligence sources to stay up-to-date on emerging threats and adjusts their security posture accordingly.

Automation can help to improve the efficiency and effectiveness of SOC operations. Here are some examples of how tasks within the SOC can be efficiently automated:

- Threat detection: ML algorithms can be used to analyse vast amounts of security data and identify patterns that may indicate a security threat. Automated tools can also be used to triage and prioritise security alerts, reducing the workload on SOC analysts. In this area of work is what this research can mostly contribute as described in chapters 4 & 5.
- Incident response: Automated incident response tools can be used to respond to low-level security incidents, such as disabling user accounts, blocking IP addresses, or

quarantining infected machines. This frees up SOC analysts to focus on more complex security incidents that require human intervention.

- Vulnerability management: Automated vulnerability scanners can be used to identify vulnerabilities in systems and networks and recommend steps to remediate them. This helps to ensure that vulnerabilities are identified and addressed quickly and efficiently.
- Threat intelligence: Automated threat intelligence tools can be used to collect, analyse, and disseminate threat intelligence data.

Furthermore, NF3 method can assist in making the inspection of data flows long before a cyber-attack can take place enabling a dynamic cyber defence which incorporates the extracted knowledge to better respond to incidents even in cases where a previous knowledge does not exist. That circumstance provides new advantages to perform prevention and introduce new rules in security perimeter devices such as intrusion detection systems/intrusion prevention systems (IDS/IPS) notably for the early contention of zero-day vulnerabilities. NF3 provides a higher level of cyber resilience to the monitored infrastructure with rapid, automated, intelligent-driven detection actuations including reduced levels of human involvement and less computational requirements for network data inspection, demystification of malware data flows and ciphered data recognition. Its benefits can be expanded to the early precise detection of Denial of Service (DoS)/Distributed Denial of service (DDoS) attacks. It is usually required the confluence of various techniques:

- Network Monitoring: This involves analysing network traffic to detect any unusual patterns or spikes in traffic. The sudden increase in traffic could be a sign of a DoS attack.
- Intrusion Detection Systems (IDS): IDS can detect malicious traffic or patterns of behaviour that are consistent with a DoS attack. IDS can be set up to alert network administrators or automatically block the offending traffic.
- Log Analysis: Monitoring server logs can help detect a DoS attack by tracking unusual or excessive requests to a server. This can help identify the source of the attack and the type of attack being launched.
- Traffic Filtering: Traffic filtering involves blocking traffic from known sources of DoS attacks or blocking traffic that meets certain criteria, such as excessive requests from a single IP address.

- Load Balancing: Load balancing distributes traffic evenly across multiple servers, reducing the impact of a DoS attack by preventing a single server from being overwhelmed with traffic.
- Testing and Simulation: Regular testing and simulation of DoS attacks can help detect vulnerabilities and identify potential weak points in a network or system. It is important to have a combination of these detection methods in place to effectively detect and respond to DoS attacks.

A DoS/DDoS attack (Sagduyu & Ephremides, 2007), (Sagduyu et al., 2010), (Tsiropoulou et al., 2016) is characterised by the irruption of huge amounts of incoming data flows that can greatly disrupt network services or resources by demanding an increased number of connections at the same time. Applying statistics to the packet attributes assist the NF3 framework to detect malicious attempts within the network flow to rapidly recognise DoS/DDoS attacks.

5.1.2 Results of the λ-NF3 model and discussion

The below tables show the obtained measurements of the various classifiers (Demertzis et al., 2019).

a) Batch Data processing effectiveness

Table 8. Measurements of methodologies.

Network Traffic Analysis (Binary) (208.629 Instances)							
Classification Accuracy and Performance Metrics							
Classifier	**TA**	**RMSE**	**Precision**	**Recall**	**F-Score**	**ROC Area**	**Time**
SVM	98.01%	0.1309	0.980	0.980	0.980	0.980	273.6 s
MLFF ANN	98.13%	0.1295	0.981	0.981	0.981	0.994	300.2 s
k-NN	96.86%	0.1412	0.970	0.970	0.970	0.970	100.7 s
RF	97.12%	0.1389	0.972	0.971	0.971	0.971	72.2 s
ELM/GRBFK	97.78%	0.1322	0.977	0.977	0.977	0.977	1.9 s

Table 9. Measurements of methodologies.

Demystification of Malware Traffic (Multiclass) (168.501 Instances)							
Classification Accuracy and Performance Metrics							
Classifier	**TA**	**RMSE**	**Precision**	**Recall**	**F-Score**	**ROC Area**	**Time**
SVM	96.63%	0.1509	0.967	0.967	0.968	0.970	101.1 s
MLFF ANN	96.50%	0.1528	0.981	0.981	0.981	0.965	148.3 s
k-NN	94.95%	0.1602	0.970	0.970	0.970	0.950	61.8 s
RF	95.91%	0.1591	0.972	0.971	0.971	0.960	38.7 s
ELM/GRBFK	96.59%	0.1523	0.970	0.980	0.975	0.975	0.91 s

Table 10. Measurements of methodologies.

Encrypted Traffic Analysis (Binary) (166.874 Instances)							
Classification Accuracy and Performance Metrics							
Classifier	**TA**	**RMSE**	**Precision**	**Recall**	**F-Score**	**ROC Area**	**Time**
SVM	98.99%	0.1109	0.989	0.990	0.990	0.990	91.5 s
MLFF ANN	99.12%	0.1086	0.998	0.998	0.998	0.998	116.6 s
k-NN	97.84%	0.1372	0.975	0.975	0.978	0.980	59.2 s
RF	98.96%	0.1107	0.989	0.989	0.989	0.990	40.1 s
ELM/GRBFK	99.20%	0.1056	0.990	0.990	0.990	0.990	0.88 s

Table 11. Measurements of methodologies.

Encrypted Traffic Identification (Multiclass) (214.155 Instances)							
Classification Accuracy and Performance Metrics							
Classifier	**TA**	**RMSE**	**Precision**	**Recall**	**F-Score**	**ROC Area**	**Time**
SVM	90.31%	0.1906	0.905	0.905	0.906	0.950	288.9 s
MLFF ANN	92.67%	0.1811	0.930	0.930	0.928	0.960	312.5 s
k-NN	85.19%	0.2032	0.890	0.890	0.890	0.935	100.9 s
RF	91.56%	0.1800	0.920	0.916	0.916	0.930	78.6 s
ELM/GRBFK	92.65%	0.1813	0.930	0.930	0.930	0.955	2.28 s

Table 12. Measurements of methodologies.

Unencrypted Traffic Identification (Multiclass) (186.541 Instances)							
Classification Accuracy and Performance Metrics							
Classifier	**TA**	**RMSE**	**Precision**	**Recall**	**F-Score**	**ROC Area**	**Time**
SVM	99.92%	0.1003	0.999	0.999	0.999	0.999	119.5 s
MLFF ANN	99.91%	0.1008	0.999	0.999	0.999	0.999	162.9 s
k-NN	98.98%	0.1020	0.989	0.989	0.990	0.995	82.7 s
RF	99.93%	0.1001	0.999	0.999	0.999	0.999	51.5 s
ELM/GRBFK	99.94%	0.1000	0.999	0.999	0.998	0.999	1.84 s

The λ-NF3 model approximately uses the same parameters used in the NF3 ensemble model. Tables 8–12 reflect the outputs of the λ-NF3 model and those from ML algorithms (Support vector Machine (SVM), Multi-Layer Artificial Neural Network (MLFF) ANN, k-Nearest Neighbor (k-NN) and Random Forest (RF)). In this case, ELM/GRBFK proposal shows a very good effectiveness that may be interpreted as promising results considering the benefits of the batch processing model in reaching higher speed rates (hundreds of times) to get these outputs. It is assessed as a reliable tool for performing big data analysis.

b) Data streams processing effectiveness

A tailored metric used in ML algorithms would be required to calculate accuracy of data streams inspections. (Žliobaité et al., 2015) studied the Kappa statistic which is particularly useful when dealing with imbalanced data or when there are multiple possible labels for each instance. It can be used to assess the performance of classification models, evaluate inter-

annotator agreement in natural language processing tasks, or measure the similarity between different clustering algorithms. Overall, Kappa statistic is a valuable tool for assessing the reliability and consistency of data labelling and model predictions in machine learning applications. Kappa statistic measures the agreement between two annotators or evaluators when labelling data and is useful to determine the accuracy of data streams analysis. Tables 13–17 show the outputs about data streams processing together with the outputs resulted from other similar approaches (Hoeffding Adaptive Tree (HAT) (Corrêa et al., 2017) and the primal estimated sub-gradient solver for support vector machine (SPegasos) (Shalev-Shwartz et al., 2011). HAT is a decision tree algorithm for online learning and data stream mining. It is designed to efficiently and accurately process high-volume, high-velocity data streams, where data arrives sequentially and needs to be processed in real-time. In the primal formulation of the SVM, the optimization problem is to find a hyperplane that maximizes the margin between two classes while minimizing the classification error. The primal estimated sub-gradient solver is a widely used optimization algorithm for SVM because it is relatively simple to implement, computationally efficient, and can handle large-scale datasets. However, it may not always converge to the global optimum and can be sensitive to the choice of the step size and working set selection. 10,000 instances constituted the training dataset. The utilisation of a prequential evaluation method involves evaluating the model's performance on each new incoming instance of data, before incorporating it into the training process. Specifically, the method involves predicting the class label of a new instance using the current model, and then immediately evaluating the model's prediction accuracy on that instance. The prediction is then used to update the model, and the process is repeated for each new instance of data. By evaluating the model's performance on each new instance of data, the prequential evaluation method provides a more accurate and timely measure of the model's performance than traditional batch evaluation methods. It also provides a measure of how well the model is able to adapt to changes in the data distribution over time (Vinagre et al., 2014). The training windows used were 5000 and 1000 instances.

Table 13. Measurements of methodologies.

Network Traffic Analysis				
Performance Metrics				
Classifier	Window Size 5000		Window Size 1000	
	Kappa Stat	Kappa Temp Stat	Kappa Stat	Kappa Temp Stat
SAM/k-NN	76.90%	77.96%	88.12%	89.64%
HAT	76.87%	77.95%	84.55%	86.19%
SPegasos	76.89%	77.29%	85.02%	87.38%

Table 14. Measurements of methodologies.

Demystification of Malware Traffic				
Performance Metrics				
Classifier	Window Size 5000		Window Size 1000	
	Kappa Stat	Kappa Temp Stat	Kappa Stat	Kappa Temp Stat
SAM/k-NN	77.02%	78.10%	83.24%	84.36%
HAT	77.06%	78.12%	83.20%	84.01%
SPegasos	77.00%	78.01%	83.02%	84.18%

Table 15. Measurements of methodologies.

Encrypted Traffic Analysis				
Performance Metrics				
Classifier	Window Size 5000		Window Size 1000	
	Kappa Stat	Kappa Temp Stat	Kappa Stat	Kappa Temp Stat
SAM/k-NN	79.00%	79.94%	86.39%	87.76%
HAT	78.96%	79.91%	82.11%	83.81%
SPegasos	78.98%	79.89%	82.68%	83.52%

Table 16. Measurements of methodologies.

Encrypted Traffic Identification				
Performance Metrics				
Classifier	Window Size 5000		Window Size 1000	
	Kappa Stat	Kappa Temp Stat	Kappa Stat	Kappa Temp Stat
SAM/k-NN	77.11%	77.54%	84.05%	85.18%
HAT	77.02%	77.35%	80.89%	81.23%
SPegasos	77.03%	77.36%	83.14%	84.16%

Table 17. Measurements of methodologies.

Unencrypted Traffic Identification				
Performance Metrics				
Classifier	Window Size 5000		Window Size 1000	
	Kappa Stat	Kappa Temp Stat	Kappa Stat	Kappa Temp Stat
SAM/k-NN	76.70%	77.87%	83.91%	85.22%
HAT	76.67%	77.86%	81.08%	81.92%
SPegasos	77.15%	77.95%	82.50%	83.04%

In general, SAM/k-NN algorithm performs better than other methodologies by getting lower inaccuracies. Furthermore, one of the main benefits derived from the outcomes is kappa reliability, which can be interpreted as the output of the data filtering, permitting the conservation of an increased amount of noteworthy information for future prediction. The kappa reliability is shown in Table 18.

Table 18. Kappa reliability.

Kappa	Reliability
0.00	no reliability
0.1–0.2	minimum
0.21–0.40	little
0.41–0.60	moderate
0.61–0.80	important
≥0.81	maximum

The kappa reliability related with SAM/k-NN model is featured as "important" in the simulations with windows size 5000 instances and "maximum" with windows size 1000 instances. The outputs of SAM/k-NN evidences a better performance on the window size 1000 instances due to the fact that data samples of past instances vanish faster with the shorter window. This demonstrates that the model is well-suited for data streaming analysis due to its resistance to imprecise information such as shifting data flows. λ-NF3 is a substantial contribution towards a NGC2SOC (Demertzis et al., 2018) which utilises ML technology to deal with fast evolving cyber threats and cyber events. It provides a higher degree of automation which, at the end, facilitates to decrease the human involvement specially for routine activities. The model makes use of a flexible and effective smart methodology for batch processing and a new developed classifier model for data streams processing in view of providing a novel vision of a multi-faceted cyber security problem affecting the protection of networking devices and systems.

The framework proposes a singular and hybrid configuration of ML models notably the ELM/GRBFK and SAM/k-NN methods. This specific ensemble provides faster training rates, simpler conduct, minor human intervention and it minimises computational data processing and data monitoring, demystifies malware data flows and identifies ciphered information. The results of this experimentation influence the way to consider new datasets able to reflect as well the considerations about the context (mission) as a valuable input for decisions. Consequently, λ-NF3 is the proposed solution to overcome increased computational demands for big data analysis of network traffic. Lambda architecture

proposes two different but sophisticated algorithms able to manage a large bunch of information divided into batch and data streams. In practical terms, it utilises batch data inspection procedures to enhance the observation of previous knowledge and immediate data stream procedures to detect new features from incoming data. The analyst would receive the findings derived from a combination of the results.

It must be noted that the λ-NF3 model should contribute with an adaptive cyber security governance model instead of relying exclusively to a reactive mode of operation against an adversary. It fuses learning outputs from two contrary algorithms to identify malware. Learning on the environment is accomplished through the use of advanced collections of data aligned with practical situations. The functioning context presented with the fusion of batch and data streams comprehensively enable a mathematical model setup as well as an increased categorisation or comparison. One identified drawback is related to the degree of difficulty of the mathematical computation which led to multifactorial questions for further exhaustive examination and reproduction. The testing environment was setup with Laptop Intel-i7 2.4 G CPU, 16 G DDR3 RAM, Ubuntu 18.04 LTS, Anaconda Python Data Science Platform and TensorFlow Python environment.

5.2 Verification and Validation (V&V) Framework to evaluate CySA

One of the tangible contributions of my PhD research is to approach a meaningful framework to perform V&V in a general context for CySA tools and solutions including its applicability to decision support systems. This research was conducted together with a research team which results led to a publication (Llopis et al., 2022) where all the various tests that comprise the framework are explained in more depth.

While working with CySA, there is a dearth of widely accepted procedures for demonstrating outcomes. As a result, it was judged necessary to establish techniques that can aid researchers and practitioners in verifying and validating CySA collection, allowing for an accurate interpretation of how human operators sense, acknowledge and predict circumstances and cyber incidents. There are rising research opportunities to connect existing technology foundations - increasingly dominated by automated or intelligent algorithms - with a business-oriented approach to deal with day-to-day business operations, missions, and courses of action (CoA).

The proposed CySA V&V framework is intended to objectively assess whether existing capabilities match user needs while also assisting in the development life-cycle of new cyber defence systems. To fill this highlighted need, this section will offer a novel V&V framework that combines organisational factors with other existing V&V methodologies and may be used to guide the assessment of CySA-related products. The framework introduces three key principles in CySA evaluation: software, operational, and application testing. These evaluations cover many associated parameters like assessing a deployment of emergent dual-use solutions up to a capacity that permits a collection of comprehensive factors about the situation that a cyber analyst can grasp. The proposed V&V framework aims to establish a consistent process for assessing and certifying the set of tools designed to aid in the procurement of CySA solutions. There are various reasons why commercially available solutions do not meet the expectations of the user: (i) the significance of cutting-edge data processing methods in tailoring situational images according with the demands of various roles (user profiles), such as technical, decision-makers, staff, and so on; and (ii) the critical necessity to avoid overabundance of valuable data and boost performance levels.

The above is caused by a variety of factors, including the significant disruption of cyber operations, the challenge of bringing the technical, organisational and procedural strategies into alignment, including the rapid pace of technological evolution and innovation of the digital market around key sectors of economy. Although many cyber security managers are interested in obtaining CySA tools and products that fit into their needs, performing V&V in this field of cyber defence is not straightforward (Lif et al., 2017). This, in turn, implies that there is a significant shortfall on collecting key requirements and manufacturing CySA products with the sufficient dynamism, interoperability and scalability for their adaption to different services in an era of constant technological advances.

In response to these difficulties, this section will outline a standard evaluation methodology that makes use of well-documented processes in order to carefully analyse all the associated parameters of cyber security mechanisms to guide the achievement of CySA. They are related with the life cycle functions from design, operation and disposal, with the objective of evaluating efficacy before being taken into consideration for deployment on actual scenarios. The conducted research contributes to:

- Introduce a basic set of assessment principles and check processes for performing a validation of CySA products particularly at low technology readiness levels (TRL) where software development is driven by agile methodologies;
- Propose novel V&V guidelines for mission-centric in order to assess cyber risks when bridging the technical orientation with a mission focused guidance;
- Outline innovative CySA principles for developing solutions and provide a sample survey for gathering expert opinions.

The sub-sections that follow will discuss the design assumptions, the V&V concepts and their associated challenges in full depth to get a complete insight into the complexity of the subject.

5.2.1 Hypothesis and Research Questions

Quality control over system requirements and V&V procedures, is often considered separately involving diverse approaches (Katasonov et al., 2006). It is an area where much research would be needed in the coming years to come to terms accuracy, plausibility, rationalism and objectivity in developing unbiased CySA tools. Using (Zimek et al., 2006) on analysing data mining viewpoints, the conclusion could be that experts were capable to determine the right taxonomy of the problem by combining their expertise while independently no clear output could be drawn. Each of the proposals that were reviewed for CySA V&V offered a separate and meaningful portion of the parameters of the solution under analysis. Once these parts are merged in a systematic manner, queries established by (Jackson, 1995) can be responded for a proper evaluation about the fulfilment of software needs and conditions, such as:

• How the requested environmental impact can be achieved?

• How would the machine behave at the border (interface) in order to have the intended environmental impact?

• How will the machine's inner activity and architecture affect the required behaviour at the boundary?

The specification of the system's outer needs followed the response to the initial query. Also, the internal specifications of the system are established by responding to the second question. The third question, which must be addressed by the assessment procedure, concerns the engineering guidelines affecting the product that must be created in order to fulfil both the outside and inner criteria (see Figure 14).

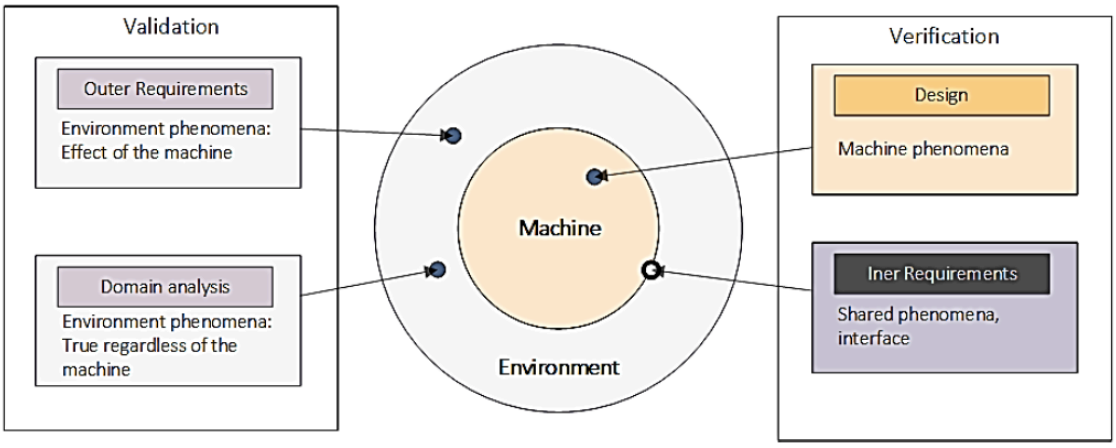


Figure 14. Validation and verification in relation to machine and environment

This means that although software designers and developers pay special attention to internal requirements, application customers and consumers prefer to concentrate on outward demands. An updated classification links inner needs to usable functionality and external needs to non-functional characteristics; these two categories must be established independently during the design stage of CySA-related instruments. To sum up:

• Functional requirements, which relate to the capacity or features necessary to directly assist users in achieving their objectives.

• Non-functional requirements, which are frequently implicit and technical in nature, emerge as system requirements to meet the users' functional demands. Examples include the quality of the service, its accessibility, promptness, and correctness.

Both types of needs, after being specified, it is assigned a unique identification number that enables a precise connection between the demand and the application´s conception, programming and checking procedures. Confirmed criteria must be quantifiable or qualitatively measurable. Quantitative measures, on the other hand, frequently use numbers. Counting events, measurements, or numerical key performance indicators (KPIs) that are directly acquired from trials, presentations, or the operational performance of the solution can be done using a quantitative method. Validation and verification notions may be distinguished clearly in the context of the aforementioned claims, and they both serve to address the following inquiries:

• Validation: Is the product I am developing appropriate in light of responding to user requirements?

• Verification: Is the solution being built properly?

Verification operations often evaluate inner requirements, whereas validation processes are focused on verifying the accomplishment of outside needs. As a result, the definition of validation is "the assurance that a product, service, or system meets the needs of the customer and other identified stakeholders," while the definition of verification is about checking if a solution satisfies a rule, a necessity or a pre-condition.

5.2.2 V&V Model

V&V techniques must be applicable to the entire life cycle of CySA solutions which could be composed either by isolated functionalities as individual components or in combination in major developments. That would mean to address the inception, the documentation of requirements, research, development, simulation, testing, engineering, fielding, upgrading, transformation or disposal. The mentioned tasks are not an exhaustive list of all the actions that should be part when developing a capability that involves people, processes and technology. Following this rationale, the newly presented methodological approach separates—testing, operation, and application—that will be thoroughly investigated by going from the bottom up. Each of these concepts focuses on a different set of assessment criteria. Testing procedures will be carried out in order to validate firstly the developments functionality for CySA products (Figure 15), with the aim of determining if the software programming tasks fulfil the needs or not with the possibility of detecting software glitches at early stages for quality and security purposes. It encompasses the operationalisation of individual or group testing with various objectives prior to determine if the developed software has the appropriate level of assurance for its use in operational environments.

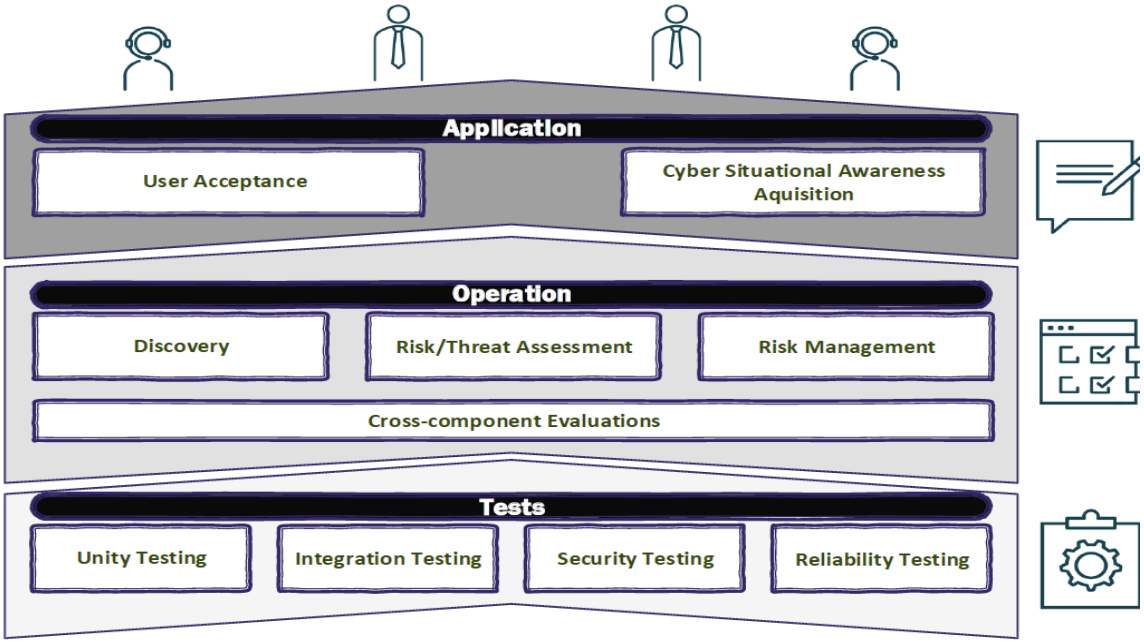


Figure 15. CySA verification and validation framework (Llopis et al., 2022)

The second step is to evaluate the intended operational capabilities. These processes include a prior review of cross-functionalities that could emerge during the integration in a platform such as protection against malfunctioning or incidents, precision, effectiveness, speed of response, upgradeability, hardware compatibility, etc. They are evaluated independently or as part of a whole CySA solution. Then, a specific evaluation will be performed, focusing on the capabilities for mapping the cyberspace, analysing a transmission of risks. Lastly, the final phase involves assessing the examined methods for correctly enabling the acquisition of CySA, which is supposed to be their primary goal. This involves assessing customer responses and objectively measuring human characteristics. Moreover, user acceptability is subject to a confirmation through a survey.

5.2.3 Proposal´s and Research´s constraints

Initial discoveries of the limits encountered during the research are explained as follows:

• Concepts suggested do not particularly address the challenges of responsible and safe data handling. As a result, prior to implementation, additional measures are to be considered for enforcing privacy.

• This proposed technique provides a high-level overview but must be tailored to the specifics of each application case. The data offered above should aid analysts in evaluating

and fine-tuning cyber defence capabilities as well as the information system subject of deployment.

• Before adoption of the model in resource-constrained contexts, an appropriate frugal adaption may be necessary.

• Previous experimentation to tackle CySA from a mission perspective is severely deficient in diverse and significant undertakings.

• The research omits potential antagonistic techniques that might undermine the planned V&V techniques. It concerns adversarial attacks on the networking devices used to measure the effectiveness of the methodology.

5.2.4 Initial Risk Analysis of the research findings

The adoption of the evaluation model could be impacted for some of the following assumed events:

• The sensors utilised, as well as how they interact inside, may influence the measurements taken.

• Appropriate expertise would be required to perform evaluation tasks as described without excluding the need for specific training to cyber defence analysts.

• Insufficient observations may result in erroneous findings.

• There is a serious drawback in the availability of datasets in order to define and tailor the methods generated in this evaluation.

• Suggested quizzes may fail to account for differences in cognition and competence among users. As a result, the established methodologies count on the ability of the evaluators to deal with such circumstance.

• The way operators work in real situations could be may disturbed by the various tests, prospective intermediate checks, and so on, resulting in a clearly visible artificial working setting. This might have an effect on the evaluation findings.

• If users are aware that they are being monitored, they may act differently.

5.2.5 Support to Decision Making

Decision support is the most difficult thing to verify and evaluate since it covers the basic operation of a CySA ecosystem including human specialists and automation. Creating an adequate framework (Shameli-Sendi et al, 2016) to enable response options according with the situation understanding has proven not being an easy problem to solve, as demonstrated

by the NF3 models in Chapter 4. Research efforts are underway to mature new avenues of approach which often depend on sophisticated data analytics being operational research one of the salient areas for further applicability to CySA. Much of the difficulties in their design arises from the need for a highly exact comprehension of multiple parameters. For instance, the adequate calculation of risk related to the protection of network or communications assets. Existing research separates distinct patterns based on the way the issue is posed. (Miehling et al., 2018) suggest to address decision support using mathematical graphics. Consequently, prospective risks are classified into phases that aims to reduce impact on the information infrastructure. Other methods approach decision from the viewpoint of a complex optimisation issue. The choice of the possible methods depends on the problem´s characteristics and available resources. Mathematical optimisation, evolutionary algorithms, artificial neural networks, and swarm intelligence, among others are methods that can handle complex objective functions involving multiple variables, constraints, and conflicting objectives. As a result, the CySA assessment framework instantiation would advise investigating hybrid solutions similar to (Llansó et al., 2019), but tailored to the specific organisational needs, where a weight is specified for each decision dimension. The scoring process is normalised, and its formulation would occur along the mission design phase, especially while it takes place a definition of probable CoAs. A variation of this proposal was subject of the visualisation explained in chapter 3 with fuzzy logic and the aggregation Ordered Weighted Averaging (OWA) operator.

5.2.6 User acceptance

To assess the level of knowledge gained throughout the evolution of the assigned tasks to a SOC operator, it becomes vital to understand the role of human factors which might impact information acknowledgement. It is feasible to reduce the influence of external disturbance on decision making through practice and training. Nevertheless, a perceived lack of objectivity inherent in this technique, requires an evaluation process divided into three parts is required: baseline, test performance, and final assessment. The following practices can be used to achieve this goal: (i) show assignments for accomplishment; (ii) create a silent environment to permit concentration from the user on the tasks; and (iii) to provide surveys that measure the individual's first perceptions in a subjective manner (Evangelopoulou et al, 2014). It is advised, among other things, to conduct a final survey on the technique used to evaluate and analyse the situation assessment. The proposed concept has not been used in

circumstances when users pledge to utilise a technology but fail to provide evidence of usage. User acceptability is linked with motivation in undertaking this analysis on the feasibility of CySA solutions for routine tasks and how their performance is enhanced by the use of tailored products. The suggested technique includes an updated survey to fit cyber situational awareness (Chin et al, 1988). The user acceptability level is measured using the following performance indicators:

• Usability. How effectively, efficiently, and satisfactorily end users can do their activities using CySA products in a certain usage environment

• Simplicity. CySA programmes' learnability and intuitiveness for operators.

• Presentation.

• Efficiency.

• Satisfaction. End users are asked in the surveys to rate the importance of each performance indicator. These and other queries include "Do the assessed functionalities present a precise image of the cyber domain?" and "Will using the CySA capabilities increase my work performance?" An activity map of the assessment workflow is summarised in Figure 16. As a result, testing, operations, and applications are the three distinct lines of evaluative activity. Quality managers, however, will be in charge of making any modifications they think necessary in light of the breadth of testing or the resources at their disposal. The needed tests are often integrated separately with respect to the operational environment since it is advised to implement this line before evaluating the operational efficacy of the proposal. They might be paralleled with any other CoA because of this characteristic. Nevertheless, any problems discovered during testing may lead to declare non-valid the acceptability. An evaluation loop is activated once per stage. The application tests would be reviewed at the conclusion of each step, including their capacity to achieve CySA and operator´s approval. Findings could be recorded and saved appropriately to facilitate future updates, integrations, or deployments in various operational situations. Because operational evaluations investigate the unique capabilities provided by CySA solutions, effort of specialists in determining, fusing and creating conformance standards is critical to their performance. Subsequently, user acceptance will be evaluated, which is anticipated to depend on how well the solutions under consideration can facilitate the acquisition of CySA.

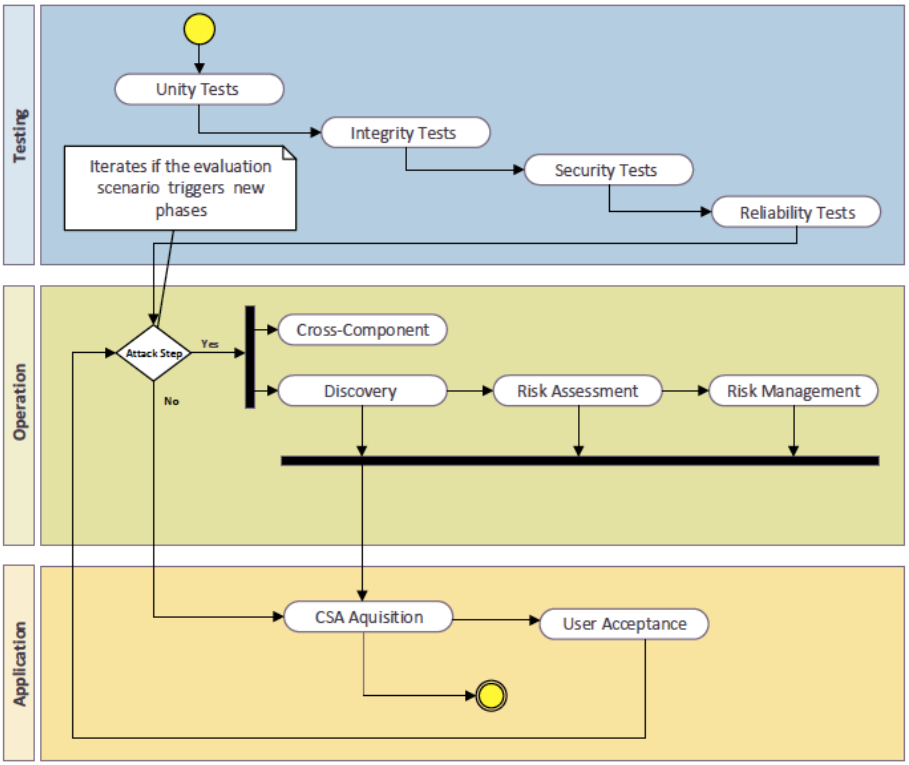


Figure 16. Evaluation workflow (Llopis et al, 2022)

Notwithstanding this reliance, it is not preferable to advocate paraphrasing the questions in a way that would lessen the influence of operational outcomes on the acceptance indicator findings, despite the tool's shown efficacy. Direct communication with operators CySA solutions will be used to gauge acceptance. The research explained here delves into the problems, obstacles, and shortfalls associated with the validation of CySA techniques in terms of software tests, operations and applications.

5.3 Research on Datasets for operationalising a mission-centric CySA

Another contribution of my PhD research is to approach the creation of meaningful datasets for CySA tools and solutions. This research was conducted together with a research team which results led to a publication (Medenou et al., 2022) where all the features of the synthetic datasets are explained in more depth.

One of the most important gaps for developing CySA products is the lack of trustworthy and complete sets of data where to leverage learning to train specific tools. According to the communication layers of the open system interconnection (OSI) architecture (physical to application), the state-of-the-art in cyber defence displays a list of network logs with labelling mistakes, a lack of variety, and incompleteness (Milenskoski et al, 2015),

(Zimmermann, 2015). They are created artificially under circumstances that are far apart from the realities experienced to counter cyber threats in dedicated cyber defence teams. A gap exists in the linkages of the available network data with the characteristics of obtaining a CySA (Liu et al., 2017). As a result, they frequently overlooked the dependencies of the cyber incidents with the impact that they may cause to the mission and objectives of the organisation. This subsection presents CYSAS-S3, a set of data developed to explore the primary requirements in support of the training aspects in a SOC. This investigation led to the creation of a group of network data flows that reflects the events of cyber-attacks and their progression using the cyber kill chain standard procedure. The cyber kill chain is a framework used to describe the various stages of a cyber-attack, from the initial reconnaissance to the final exfiltration of data. The framework was originally developed by Lockheed Martin and has since been widely adopted in the cybersecurity industry. The Cyber Kill Chain comprises the following seven stages:

1) Reconnaissance: In this stage, the attacker gathers information about the target, such as identifying potential vulnerabilities, researching employees, and mapping out the target's network architecture.
2) Weaponization: In this stage, the attacker creates or acquires a weapon, such as malware, that can be used to exploit a vulnerability in the target's system.
3) Delivery: In this stage, the attacker delivers the weapon to the target, often through email, social engineering, or other means.
4) Exploitation: In this stage, the weapon is used to exploit the vulnerability in the target's system, allowing the attacker to gain access.
5) Installation: In this stage, the attacker installs a backdoor or other means of maintaining access to the target's system.
6) Command and Control: In this stage, the attacker establishes a command and control channel that allows them to communicate with the compromised system and issue commands.
7) Exfiltration: In this final stage, the attacker exfiltrates the data or information they are seeking to steal.

By understanding and mapping out each stage of the cyber kill chain, cybersecurity analysts can identify potential vulnerabilities and take steps to prevent or mitigate an attack. The framework can also help organisations to develop effective security strategies and improve incident response planning.

Due to these reasons, CYSAS-S3 dataset was developed to close identified gaps in assessing remediation actions with a precise understanding on the cyber-attack progression coupled with mission, objectives and resources. This should relate to, (1) assessing which resources are necessary and propose mitigation actions; (2) exposing deficiencies in the cyber resilience of own network and services; and (3) evaluating the capabilities to incident handling including identify and prevent cyber-attacks. In a virtualised modelling and simulation infrastructure (Cyber Range), a CIS-level CYSAS-S3 dataset was created, which included three key use cases: DoS, exfiltration of information and theft of network credentials. In this experimentation CYSAS-S3 benefitted from an environment that allows to connect technical features with business-related operations. Throughout the past years, various traffic generators were created based on various techniques. Among other things, the specific characteristics of artificial network data is addressed by (Milenkoski et al., 2015), which distinguishes three types of workloads: benign, malware and a combination of both traffic classes (Muralidharan et al., 2022). Three phases of development have been identified (see Figure 17). The initial phase was dedicated to data collecting. The second block of activities was produced based on the previous stage's analytical study, with the goal of defining a method for verifying CySA. It means determining the appropriate use cases in support of a dataset production strategy that connects the various CySA impact assessment layers may be developed. Finally, the produced collection of data is subject to a validation process using experimental findings from well-known event detection methods.

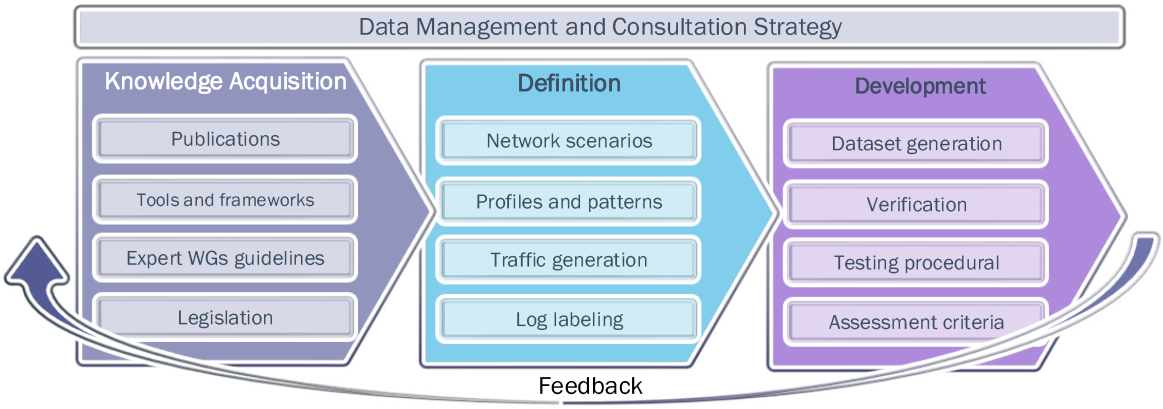


Figure 17. Dataset research method (Medenou et al, 2022).

In general, information needed for analysing and developing a CySA system (CYSAS) should be taken from the network security devices and protection mechanisms, standard network flows and other sources of cyber security information.

• An assumption was made in the sense that the commercial-off-the-shelf (COTS) solutions used in the CYSAS-S3 dataset creation process functioned normally. This includes the accuracy of the logs, events, and warnings produced by such solutions.

The following constraints have been observed as a result of the study activities:

• Because of the high amount of network activity created by running each scenario, generating big datasets using packet capture (PCAP) files was almost impracticable in terms of manageability, hence aggregated information was delivered via comma separated values (CSV) files.

• The wide range and variability of simulated impartial attitudes used as a reference for cyber-incidents use cases could provoke a misinterpretation of findings. A thorough investigation regarding the influence of created material is a difficult endeavour that is beyond the scope of the research. The dataset must be capable of covering everything from detection of malicious samples to the provision of recommendations. To provide an engineering support to these tasks, and to go beyond the boundaries of current state-of-the-art datasets and evaluation procedures, mission level reports were synthetically replicated in tandem with cyber-attack scenarios.

Each cyber-attack phase is appropriately marked and distinct. The procedure depicted in Figure 17 will first validate the capability for detecting and analysing CIS-level risks/threats in cyberspace. The functionality capable of inferring the extension of occurrences linked with organisation, resources or objectives is next examined (left hand side from a bottom-up cycle). The capacities for recognising, choosing, organising, and translating the derived mitigation measures are assessed in the next step (right hand side from a top-down cycle).

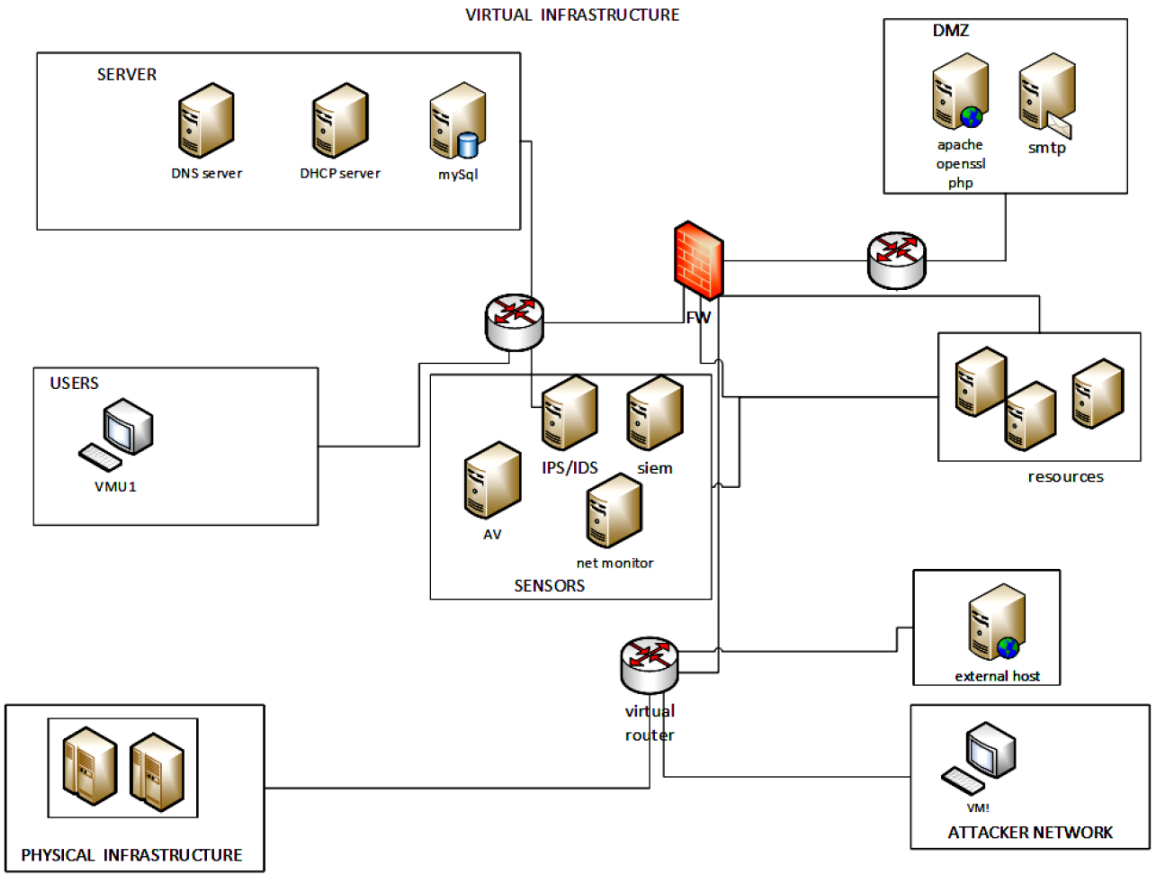


Figure 18. Network view of the deployed infrastructure (Medenou et al, 2022).

The research discussed CYSAS-S3, a dataset created and developed to help the testing, assessment, and calibration of CySA applications. The goal of the undertaken research was to link in a hyperrealistic simulated environment, the influence of known and well-documented cyber-incidents on the cyber defence assets that provide the capabilities required for mission accomplishment. This challenge was unprecedented in the state-of-the-art. The databases relate cyber events to how they affect the tasks, objectives, etc. of each mission. Each cyber-attack developed a detailed cyber kill chain.

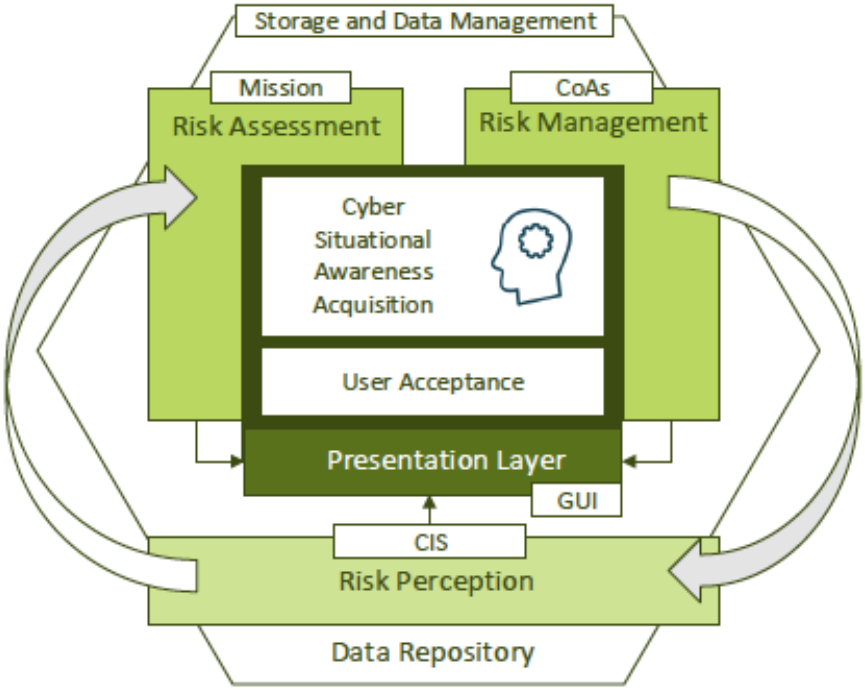


Figure 19. Evaluation loops (Medenou et al, 2022).

More analytic steps may be taken to build on the findings offered. These may comprise, for example, an analysis of the influence of developed solutions under review, as well as a more detailed explanation of the fictitious situations used (command executed, scripts, behavioural models, etc.). Additional interesting research directions may emerge from the following challenges:

• Incorporate a broader range of techniques supported by standard cyber security incident frameworks. Investigate innovative approaches to comprehend adversarial attacks according with internationally adopted taxonomies e.g. MITRE.

• Investigate new scenarios and add operational planning terminology like dependency of actions, objectives, etc.

• Create samples with various profiles, both on the attacking and defending sides. Certain factors may control elements such as initiative, predictability, stress level, and so on.

Chapter 6
Enabling techniques for Decision Support Systems

Page intentionally left in blank

6.1 Autonomous Intelligence Cyber Defence Agents (AICA)

This section is about autonomous intelligent software or hardware agents. These agents are code-based objects built on artificial intelligence algorithms. Its creation and subsequent deployment are subject to a vertiginous pace involving study and experimentation in fields other than telecommunications systems, computer networks and cyber defence. The description of its architecture, capabilities, potential applications, and future evolution are topics of interest that contribute to an understanding of the state-of-the-art and where future research and innovation efforts in this field of science may be focused. Arguably, its most well-known aspect is the defence and security of computer networks where vulnerabilities exist, including the detection and response to cyber-attacks. Detection of anomalous activity patterns, anticipation and rapid response within information systems only illustrate a part of the existing projects to develop cognitive agents that can perform complex tasks autonomously with a minimum human intervention. The set of rules that an agent must follow can be static or dynamic, allowing a greater or lesser autonomy of its actions. The lessons learned will lead to substantial improvements of their performance and efficiency in the coming years.

They are virtual elements of greater or lesser complexity in their software programming and designed to execute a series of tasks in a specific environment. In the field of cyber defence, these agents are closely related to cyber resilience or dynamic resilience of computer networks to unintentional errors and cyber-attacks. The advent of such agents must be understood in the context of the digital transformation and the emergence of a pervasive interconnection of platforms and systems.

Among these platforms to be networked are those that are autonomous such as unmanned and sensors. All of them have an increasing exposure and dependence on software and hardware. Specifically, air, naval, ground and unmanned underwater systems, sensors and others face several operational challenges such as ensuring effective command and control, the availability of long-distance communications, the ability to operate in degraded environments - understood as those in which not all the technical features are available and therefore it is required an alternative use of means for contingency actions such as environments where there is a high congestion of the radio spectrum or in presence of electromagnetic jamming, harsh weather conditions, insufficient energy supply, and so on - or the ability to process information on board to relieve workload to operators.

Cyber threat protection needs are growing, including the need to adopt a system engineering practice in the design, evaluation of functionalities and manufacturing of all civilian and military platforms. Among its many applications, artificial intelligence (AI) would make possible to check, through modelling and simulation techniques, the presence of possible design flaws or discover any deviation of user requirements. These procedures established from the beginning of the development of any system or platform would avoid having to resort to risk mitigation actions to solve any essential cybersecurity requirement that has been overlooked in the specifications. Both the hardware and the integrated software of any platform form a single entity when operating, but it is desirable that they could be decoupled with the aim of having a flexible, modular and open architecture. This decoupling between hardware and software would be due to the need to make modifications of the latter to adapt to changing environments, where employment requirements could evolve, leading to updates or improvements of future functionalities once deployed in missions. To this end, the decoupling among the hardware platform, the autonomy and on-board control software together with the software of the operator control system would allow the autonomy software update to be addressed with greater agility without the need to create new ad hoc platforms (Defence Science Board, 2012). The proliferation of autonomous platforms and their use has not always had an adequate operator training for their correct use. In their technological evolution, the first autonomous systems, rapidly incorporated in support of missions, may have been underutilised due to lack of confidence and uncertainty about their autonomous mode of operation.

In this regard, it would be advisable to always carry out a post-mission study of the characteristics of the behaviour and the actions carried out by the autonomous system through 'indicators of understanding' on the 'black box' that on many occasions makes up the operation of its AI algorithm. An unequivocal distinction must be made between autonomy – understood as the ability for a system action to be governed independently by the system itself in accordance with specified limits in its programming – and automation, which does not have the said ability to compose and select between several courses of action, but is governed by pre-established rules. Likewise, autonomy must be understood within the man-machine collaboration. For greater effectiveness in such collaboration, natural language processing techniques as a mean of communication must be improved. The language contains a greater richness of information to facilitate understanding when commands are

transmitted to autonomous systems compared to teleoperation through interfaces (Defence Science Board, 2016).

6.1.1 Artificial intelligence and cyber defence

Cyber defence is paramount to secure AI, just as AI enhances cyber defence capabilities by adding a component of autonomous understanding necessary to the speed at which operations occur in cyberspace. Deep neural networks (DNN), reinforced learning (RL), machine learning (ML) and generative adversarial networks (GAN) are some of the advanced techniques of AI that may represent great advances for agents and the possibility of expanding its use to other fields such as energy, land or air traffic management, medicine or industrial systems.

One of the fields of application in the use of agents is the detection of malicious software (malware) and cyber threats (Mees & Debatty, 2014). The speed of their actions exceeds that of an operator of a security operations centre (SOC). Furthermore, the efficiency and performance of their activity largely depends on the data set that has been used for the training of their decision algorithms. The availability of a reliable data set is one of the preconditions for optimal agent behaviour at an early stage of operation. As the agent learns from the environment, it may adjust its parameters to better suit the circumstances of the system in question. The complexity is greater when it comes to multi-agent systems (MAS) where there is a distribution of tasks for the sake of a common goal and an information fusion from several sources is required to know in detail the situation. In this form of multi-agent work, it is important to obtain consensus when determining if a system is compromised or if it is a false alarm.

The use of agents for autonomous tasks and cyber defence operations has been called Adaptive Cyber Defence (ACD) and includes the development of advanced reasoning methodologies. The cyber defence of autonomous systems involving a significant use of means and resources could not easily be carried out in a centralised manner. This is due to the complexity of establishing a single network connectivity that also allows high mobility and dynamism of the users of the systems themselves. In this case, the option to adopt would be a distributed cyber defence. This circumstance has an impact on the functions to be developed by deployable SOCs since it will be very likely that it will exceed their monitoring and incident response capabilities even if they are established in a distributed manner.

These situations can motivate a technological evolution towards the use of local cyber defence agents or groups of such agents that adapt to frequent changes in the situation and are able to autonomously offer effective protection even in the absence of a control centre responsible for sending orders in real-time to organise their actions. The implications of such complexity when it comes to keeping autonomous systems in a network would be translated into transferring monitoring, maintenance and cyber resilience to the agents installed in a decision and cyber defence module of these systems.

The transfer of knowledge in AI is undoubtedly an area of research interest since it can save time in the training of agents even those which foresee to carry out isolated actions with the challenge of knowing if the accumulated experience could be replicated. The certification of its operation would also open the interesting discussion about the possibilities of explainability of the behaviour of agents. It is of particular interest the detailed study of MAS compared with the competencies of a single agent in the virtual domain. This comparison presents obvious analogies with drones and swarms respectively in the physical domain. While an agent can be considered as a software or hardware element capable of deciding for itself; the MAS, as a set of agents, multiply the possibilities of action. For example, MAS have the ability to establish a coordination of tasks between agents with the same or different profiles, they are normally in charge of a computationally complex mission that might require a reorganisation of own resources depending on the circumstances – even without interaction with a supervisory control centre – and could face unforeseen situations, etc. Based on these responsibilities, it becomes more important to provide "algorithmic robustness" to the MAS to have a high resistance against cyber-attacks, with the capacity to generate contingency and safeguard plans that allow to protect – the system or a part of it – in case of corruption of its programming code, ability to activate a fail-safe mode with minimal disaster recovery capabilities and many other future capabilities that demonstrate the magnitude of the technological challenge - in line with the discussion of autonomy and human control of intelligent machines, it is well known the situation that happened in the science fiction film "2001: A Space Odyssey", where the computer on board the spacecraft (HAL 9000) challenged human control due to an ethical dilemma.

The implementation of autonomous capabilities also significantly impacts command and control systems understood as the set of applications and procedures that facilitate the supervision and transmission/reception of orders. Operations planning is a linear methodology, established in phases and requiring a certain time of analysis. From

understanding the operational environment to choosing courses of action (COA), risk analysis and determination of plans, many reflection sessions mediate aimed at providing advice for decision making. Autonomous capabilities in decision support would increase the pace of information analysis and the generation of intelligent proposals as a change of plans. Its acceptance by the human team of analysts would always start from the fact that each proposal must be justified with objective reasoning.

Similar to command and control systems, in disconnected, intermittent and limited (DIL) environments, telecommunications and the electromagnetic environment would be bolstered and optimised with autonomous cyber defence capabilities that would counter an adversary's capabilities. An example would be cognitive electronic warfare that uses advanced learning systems based on AI techniques that analyse and interpret information from sensors to reason and adapt to a changing electromagnetic environment. The spectrum of radio signals is finite and is in constant competition for its use. This circumstance raises the need for dynamic spectrum management that reduces the saturation and congestion to which own signals are subjected. The future evolution of the cybersecurity architecture of information systems also has a lot to do with the inherent autonomous capabilities of agents.

Zero trust (ZT) security systems consider that no network user can have implicit guarantees to access without exhaustive control. Both local and remote users will need to go through a rigorous authentication process in order to access services. The premise is to assume that the network could be compromised, so it is necessary to monitor its status at all times. This advanced security system applies to all network resources, both computers and software applications that must be authenticated in each transaction. The analysis and evaluation of the risks lead to consider that minimum access privileges are to be established for any user, proceeding to make a complete diagnosis of that trust with each transaction or session. For this purpose, the device that ensures compliance with security policies relies on advanced decision elements such as trust algorithms that allow effective authentication and authorisation of users. The architecture control panel (Ros, 2020) will have to respond to two main challenges. On one hand, the viability of its architecture in networks that do not have enough information processing capacity to rightly carry out all access control and verification operations. This would result in specifications that would be applicable to lighter and less resourceful networks. On the other hand, the control plane is unique and seems to be the weakest point of the architecture where all decisions on the implementation of agent-supported security policies converge.

Special attention also merits the specific use of cyber deception as a source of cyber threat intelligence and as a technique through which an adversary can be deceived to induce errors in his assumptions. Cyber deception helps improve cyber defence by providing insight into the tactics, techniques, and procedures used by an adversary to conduct a cyber-attack. The implementation of cyber deception services takes place in honeypot systems. The fight against deception carried out by an adversary and aimed at confusing own AI-based systems is another area of incipient research that seeks to strengthen the defences of ML applications to make them more resistant to deception attacks. As a sign of the importance of protecting AI algorithms against deception-based attacks, the Defence Advanced Research Projects Agency (DARPA) launched in 2019 the "Guaranteeing AI Robustness against Deception" (GARD[1]) programme whose objectives are to: identify ML vulnerabilities through use cases, characterise the properties that improve its robustness and create the defences that counteract the aforementioned adversary´s deception. Many expectations are placed in the advances of adversarial reasoning and in the adversarial wargaming to deduce the intentions of the adversary. The progress of AI must always attend to two different perspectives, the point of view of the one who defends and the one who attacks.

In that respect, as part of the mentioned programme, GARD researchers from Two Six Technologies, IBM, MITRE, University of Chicago, and Google Research generated a virtual testbed (Armory), a toolbox (Adversarial Robustness Toolbox or ART), a benchmarking dataset, and training materials that are available to broader research community via GitHub. Armory is a testbed for running scalable evaluations of adversarial defences. ART is a library on python programming language for securing ML algorithms and is hosted by the Linux Foundation AI & Data Foundation. ART has been designed to offer assistance to computer programmers and experts to assess the cyber resilience of ML and in particular its protection against adversarial advances. ART uses most common ML platforms for its analysis. This valuable information is complemented with a dataset of named APRICOT including a Google Research self-study repository.

The race to gain a technological advantage will always be counter-balanced by an adversary's potential capabilities to prevent it. DARPA has also launched in 2022 a program called "Cyber Agents for Security Testing and Learning Environments" (CASTLE[2]) that aims to: create environments as real as possible through the use of "cyber ranges" and training cyber

[1] https://www.darpa.mil/program/guaranteeing-ai-robustness-against-deception
[2] https://www.darpa.mil/news-events/2022-10-24

defence agents to counter Advanced Persistent Threats (APT) - threats based on the exploitation of vulnerabilities which are unknown to computer network administrators and in this way, they are able to infiltrate these networks and remain silent for long periods of time to obtain information. APTs act stealthily and are very difficult to detect as they mask their exchanges of information with external command centres as normal network activities. CASTLE will address the robustness of cyber defence mechanisms by studying problems based on ML algorithms and provide training of cyber defence agents in evolving adversary environments.

6.1.2 Intelligent and autonomous agents in cyberspace

Situational awareness must be understood as a modular set of many functionalities and not only as a capacity dedicated to obtaining and managing cyber intelligence. There is no doubt about the added value that agents may bring for the improvement of situational awareness in cyberspace. Their performance in the detection of malware and learning to prevent cyberattacks, the immediate repair of those systems or networks affected or even the modification of network parameters or their security mechanisms to obtain more efficient and better protected system architectures would produce a remarkable advance in cyber defence. The way of presenting these situations and their impact in the course of operations for a better comprehension of the mission context is precisely one key objective to achieve SA in cyber defence. In particular, understanding the situation and comparing it with the expected situation according to the agreed plans, general advice of the operation, prediction of possible future scenarios, recommendation of actions that achieve the desired effects, proposal of additional actions that have left unnoticed by an analyst and many others, would effectively support the achievement of business objectives.

The IST-152 research group of the Information Systems Technologies Panel of the NATO Science and Technology Agency has developed a reference architecture and a technical roadmap about agents (Theron et al., 2020) that provides greater definition of the expected functionalities for the benefit of providing a common research and development framework focused on agents' interoperability. Although the theoretical conception of the agents is not new, it is new the increased research in recent years, due to advances in information processing with big data techniques and (above all) due to the disruptive nature of AI. The agents and in general the autonomy of the systems and platforms are governed by the principle of trust that the system in question will work, and act as planned allowing

the correction of its instructions, in any case. The introduction of autonomous systems requires enhancing human-machine communications and collaborative teamwork. Autonomous systems should be supervised by humans at different levels of autonomy depending on the mission. The added value is to have a tool that can offer continuity and persistence in the performance of tasks without diminishing their capabilities for long periods of time while assisting the operator.

The Reference Architecture proposed by (Kott et al., 2019) for Autonomous Intelligent Cyber Defence Agents (AICA) establishes five high-level functions:

- Perception of the environment and identification of its state
- Planning and selection of actions
- Collaboration and negotiation
- Execution of actions
- Learning and improvement of knowledge

From this high-level distribution, the architecture descends in detail to describe all the general and specialised components of the agent including a complete list of the research challenges associated with each function. It should be noted that one of the components that should be addressed in greater depth is the one relating to mission analysis. The chosen context for the definition of the content of AICAs and the sequential order of high-level functions seems to be better suited to the characteristics of classical computer networks and their cybersecurity, without necessarily responding to other larger-scale and highly dynamic information systems. Mission analysis could be performed within the planning function, which is critical to determining the selection of possible actions from a predefined catalogue. Another solution that could bring benefits to a high-level function study would be the comparison with the operational planning process (OPP) of military missions to undertake an expeditious process adapted to the needs of the agent/MAS. In addition, the developed architecture should be a starting point towards a standardisation process that allows different final implementations, but without altering the essence of a common interoperability.

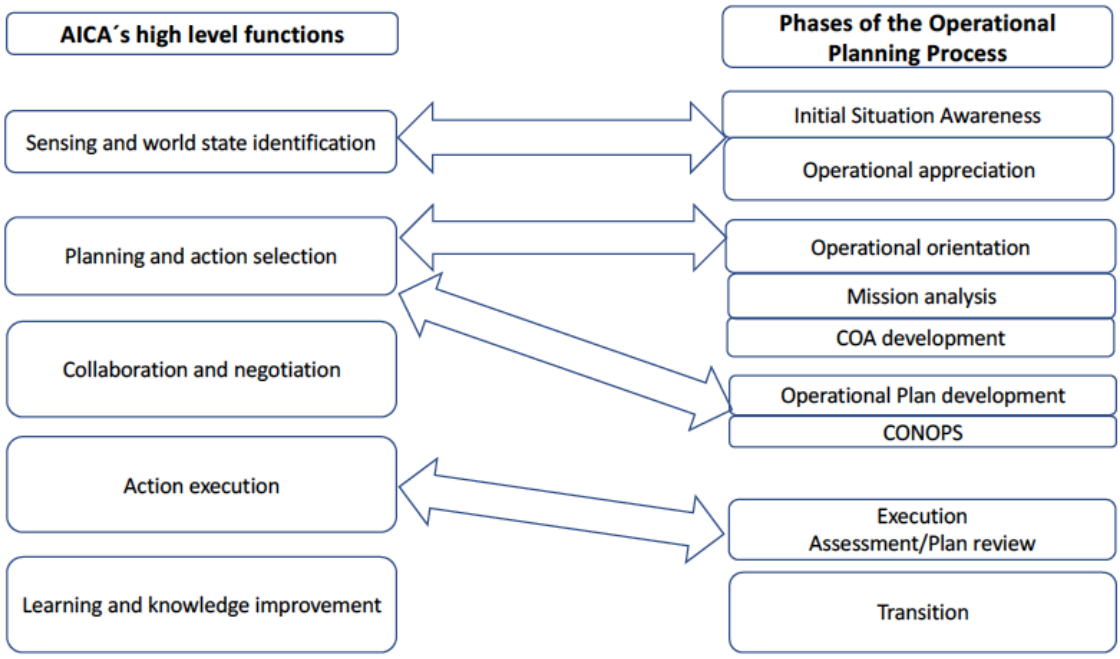
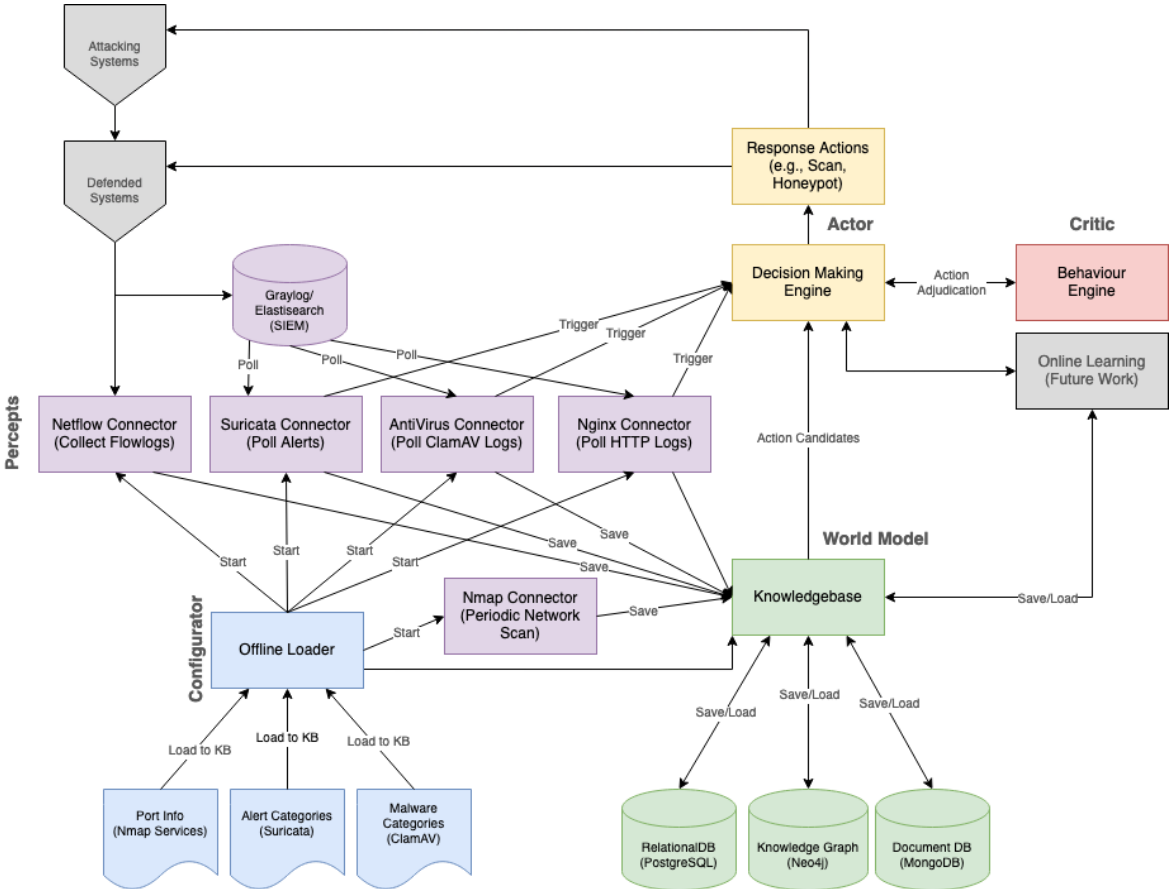


Figure 20. Comparison between AICA´s high-level functions and OPP´s phases.

The agents are installed in the network as a software application or integrated into the hardware of platforms to monitor control devices, navigation, decision, communications or information exchange interfaces. In the course of commissioning the agent, its functions should tend to be computationally efficient, resulting into the least possible effort when processing information to reserve resources. Hardware components will have to be energy-efficient or look for another alternative energy sources to enable agents´ tasks in situations where they must adapt to drastic changes in their environment. The AICA architecture is in testing phase through the GitHub platform with an initial distribution considered as a functional version of a minimum viable product (MVP). Precisely, the verification and validation of agents in experimental environments is a complex task. Operational validation is a prerequisite for commissioning and serves to maintain operators' confidence in the specific use of agents. The latter can be addressed by using the CySA V&V framework as described in sub-section 5.2. AICA is a cutting-edge research endeavour which aims to build a smart agent model compatible with different environments. Figure 21 shows a typical agent´s building blocks:



Source: https://github.com/aica-iwg/aica-agent

Figure 21. High-level structure of the AICA agent

The introduction of the agents described in this chapter will be the technological focus in the coming years due to a drastic digitalisation of networks, platforms and services. Countering cyber threats in the form of sophisticated cyber-attacks that could prevent the normal functioning of telecommunications and information systems will be a major challenge. The increased presence of autonomous systems and the pervasive use of machine learning algorithms, may bring closer a reality of a "robot battle" competition that would result in an unprecedented acceleration of the action-reaction between defender and attacker. Given this reality, the agents / MAS would be the asset capable of countering malicious software in any of its forms in time and opportunity. This would mean an own capacity for early warning, prompt reaction to incidents, planning, decision-making, execution of actions and almost immediate damage assessment. The process of planning and conducting mitigation measures in the advent of compromised networks could benefit from intelligent advice on decision-making. Agents would analyse the information and provide an alternative point of view. Systems engineering would also benefit from agents conducting user requirements check and a safety assessment of the design of military platforms. As a result of the above and with

the intent of serving as a guide for researchers and experts in cyber defence, the following possible lines of research are proposed, complementing those described by (Kott, 2019) and (Defense Science Board, 2016):

- The creation of reliable datasets for training, learning and obtaining knowledge of different critical information environments. If the dataset is referenced to the MITRE ATT & CK phases, the analyst could better follow the evolution of the alerts. MITRE ATT & CK is a standardized knowledge base that categorizes the actions of the adversary according to the phases of the life cycle of a cyber-attack progressing from pre-attack recognition to impact determination. In the CYSAS-S3 dataset MITRE ATT & CK and cyber kill chain were referenced;
- Determination of agent/MAS objectives in quantifiable and programmable orders;
- Development of error correction techniques for agent instructions, operation in fail-safe modes and contingency action;
- The development of techniques for comparing short-term objectives with long-term ones, analyse conflict resolution issues when dealing with mutually exclusive objectives and address the redefinition of agent/MAS objectives;
- The identification of priorities in achievement objectives;
- Collaboration, coordination and division of tasks in the MAS;
- The development of efficient architectures for the exchange of information in dynamic MAS;
- Optimisation of response plans identified by the MAS;
- The development of learning federation techniques to overcome the difficulties of scarce and reliable training data (frugal learning);
- The feasibility of knowledge transfer between agents;
- The development of advanced techniques for the processing and understanding of man-machine natural language serving as a source of information for transmitting commands and obtaining machine responses about language comprehension;
- The anticipation of failures and the realisation of self-diagnosis of the agent / MAS

Some authors establish an approximate period of 10 years to obtain tangible results in the proposed lines of research in order to mature the results obtained to technological demonstrators. Many of the challenges above-described might require a longer technological leap in time which could be estimated by 2035 without ruling out that some significant technological advance in AI in the coming years might accelerate the expected pace of getting these results in the field of AICA/MAS.

Page intentionally left in blank

Chapter 7
Conclusions and future work

Page intentionally left in blank

7.1 Conclusions

Throughout the doctoral thesis, a reader will find several research tasks (human factors, visualisation, data analytics, fuzzy logic to compute security metrics, novel machine learning models for inspecting network data flows, development of synthetic data sets, a validation and verification framework to evaluate CySA tools and solutions, etc.). All of these research activities have a common nexus which is the author´s attempt to bridge the gap between the CIS/IT sphere of knowledge in cyber defence (information infrastructure) and the related mission impact/decision mechanisms, using the foundational basis of SA to certainly understand the challenges and progress towards a common understanding about a specific situation. It is indeed a dual-use endeavour aimed to assist operators, administrators and decision-makers in shaping a cognitive computing SOC/CERT capacity in response of an evolving cyber threat that affects systems, computer networks of critical infrastructure and the society in general. In the author´s opinion, more work shall be devoted to building CySA capabilities tailored to user needs where research and experimentation will be paramount to understand the human cognitive processes to achieve what is the "ground truth". The challenge is to include in the SA analysis, more elements of the decision and the performance of action based on the situation, thus better evaluate mission dependencies, cyber assets and resources. The research performed during my doctoral programme has improved the current state-of-the-art by addressing concrete aspects of the complex cyber defence ecosystem whose results are presented over the various chapters of this thesis. A further reflection on the issues encountered and a forward-looking perspective will be discussed. Figure 22 describes the relationships of the conducted research for decision support. The picture shows some arrows in the boundaries which are meant to incorporate new functionalities, modules of information sources.

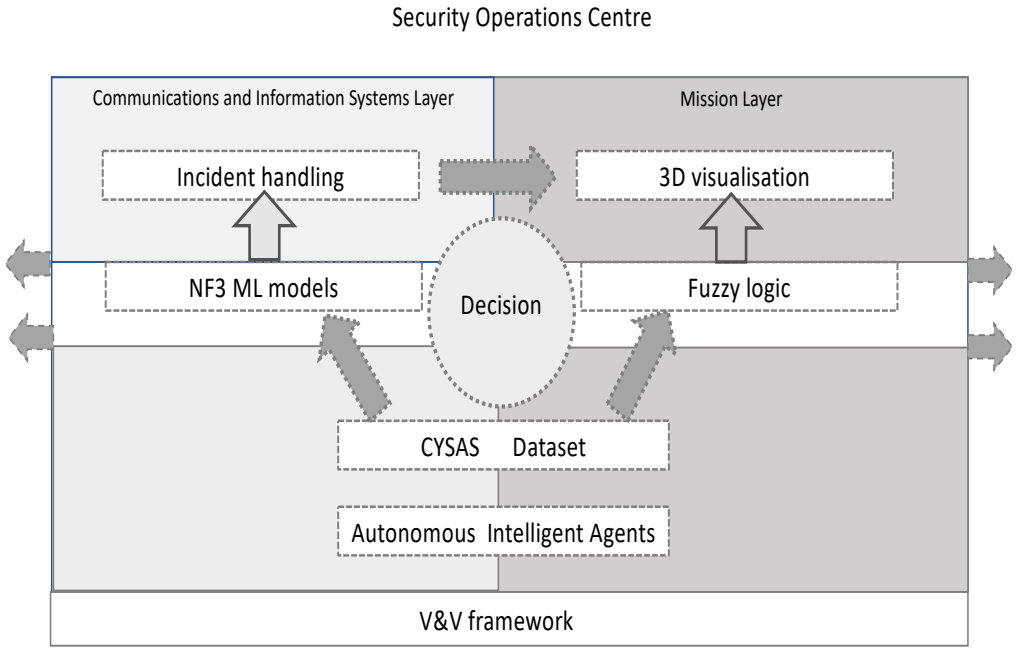


Figure 22. Decision elements elaborated during the thesis and their relationships.

7.1.1 Conclusions of the NF3 machine learning models

The proposed NF3 model is adaptive, trustworthy, and extremely functional digital forensics solution built around the concepts of ML and AI. When typical malware products are increasingly demanding higher computing power with the risk of becoming obsolete in their detection capability, NF3 model arises as an alternative. Its ensemble ML architecture leverages the most effective fusion of four extremely effective and quick classifiers. Its contrasted added value is suggested as a future-proof modern SOC. This complex technique made a convincingly new suggestion again for specification and development of enhanced computer security networks, especially when paired with the encouraging outcomes that have been observed. Moreover, this scheme employs a data science strategy that makes use of unified and precise observations of new entry data packets in an effort to compensate delay, capacity, and high availability. It is important to note that NF3 presents a revolutionary algorithmic fully distributed forensics mechanism that leverages minimal computer usage capabilities to analyse network activity, contextualise infected behaviour, and identify cipher text. With respect to the information richness, it appears to be adequate among available deep learning methods; nevertheless, deep learning techniques are highly computationally inefficient and time demanding notably throughout its learning curve. Advanced models require a significant amount of time to train requiring powerful computer

graphics devices. It is a question of minutes to training the NF3 model from its initial status. Nevertheless, the selection of variables, architecture, practicing techniques, and so on still remains an uncertain issue. The classifiers employed in the NF3 framework make it easier to deal with data and orientate towards the right architecture.

7.2 Future Work

A promising area for future research is related with embedded intelligent algorithms shaped as autonomous intelligent cyber defence agents due to its scalability and major disruptiveness as mentioned in chapter 6. Cyber defence mechanisms are evolving from perimeter-based security devices to more decentralised network architectures where cyber security aspects might be delegated to software agents or by enhancing the robustness of data protocols.

7.2.1 Discussion on future research initiatives with regards to NF3 models

The ensemble framework inside a composite system, which would be anticipated to manage numerous data employing batch and stream workflow techniques (lambda model), might be further analysed in a future study (lambda architecture). Moreover, to manage and extract hidden information from the heterogeneous data that emerge from network flow analysis, semi-supervised algorithms and learning algorithms techniques may be applied. Moreover, NF3 might be improved by further refining the ensemble framework's settings in order to get an even more accurate, rapid, and effective categorization. SOC technicians would benefit from a personalised display included in the suggested NF3 in comprehending the cyber defence scenario. It would also be crucial to research how this construct may be expanded by utilising the same design in a scalable large data analysis system using big data analytics. In addition, the functioning of NF3 with auto-optimisation and recursive transfer learning to automatically execute the protection versus complex cyber-attacks might be taken into consideration in the direction of potential development.

It is yet to be seen the possible use of bio-inspired intelligent systems to complement the proposed architecture. An area of improvement would consist of finding a straightforward strategy to lessen the arithmetical difficulties of the ensemble format and the fusion of algorithms. Further experimentation initiatives are anticipated in order to expand the λ architectural model's capabilities under a unique framework. Moreover, the λ -NF3 model might be enhanced to further strengthen the limitations of the algorithm utilised by the λ

architecture, resulting in projections that is even more effective, exact, and quick. A C2 system with sophisticated reports and representations that improve the overall decision mechanism may be supported by multi-format visualisation choices. The process of the λ - NF3 model with consciousness and contextual techniques to completely optimise the defence against oppositional cyberattacks is the last extra piece that might be taken into consideration as a continued expansion. The latter being a promising area to discover weaknesses and mitigation measures in view of achieving a necessary cyber resilience of artificial intelligence algorithms and vice versa. The study of the applicability of operations research, case-based reasoners, Bayesian analytics, Hidden Markov Chains and fuzzy logic in combination with new constructs of intelligent algorithms is also an area of interest to make progress on the decisions aspects that should be integrated in building a CySA system.

Acronyms

ACD	*Adaptive cyber defence*
AI	*Artificial intelligence*
AICA	*Autonomous intelligent cyber defence agents*
APT	*Advanced Persistent Threats*
CERT	*Computer Emergency Response Team*
CDX	*Cyber Defence Exercise*
CKT	*Cyber Key Terrain*
CPS	*Cyber Physical Systems*
CySA	*Cyber defence situation awareness*
COA	*Courses of action*
CONOPS	*Concept of operations*
COP	*Common operational picture*
CP	*Command post*
DIL	*Disconnected, intermittent and limited*
DNN	*Deep neural networks*
GAN	*Generative adversarial networks*
HSI	*Human System Integration*
ISR	*Intelligence, surveillance and reconnaissance*
MAS	*Multi-agent systems*
ML	*Machine learning*
M&S	*Modelling and Simulation*
OPLAN	*Operational plan*
OPP	*Operational planning process*
RL	*Reinforcement learning*
SA	*Situation Awareness*
SIEM	*Security Information and Event Management*
SOC	*Security operations centre*
ZT	*Zero trust*

Table of Figures

Table List

References

Almubayed, A.; Hadi, A.; Atoum, J. A Model for Detecting Tor Encrypted Traffic using Supervised Machine Learning. Int. J. Comput. Netw. Inf. Secur. 2015, 7, 10–23.

Alshammari, R.; Zincir-Heywood, N.A. A flow-based approach for SSH traffic detection, Cybernetics, ISIC. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Montreal, QC, Canada, 7–10 October 2007; pp. 296–301.

Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. Big Data Cogn. Comput. 2018, 2, 35 15 of 17.

Bonab, R.H.; Can, F. A Theoretical Framework on the Ideal Number of Classifiers for Online Ensembles in Data Streams. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; p. 2053.

Buczak, A.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Commun. Surv. Tutor. 2015, 18, 1153–1176.

Cambria, E.; Huang, G.B.; Kasun, L.L.C.; Zhou, H.; Vong, C.M.; Lin, J.; Yin, J.; Cai, J.; Liu, Q.; Li, K.; et al. Extreme learning machines [trends & controversies]. IEEE Intell. Syst. 2013, 28, 30–59.

Chatzimichailidou, M.; Stanton, N.; Dokas, I. The Concept of Risk Situation Awareness Provision: Towards a New Approach for Assessing the DSA about the Threats and Vulnerabilities of Complex Socio-Technical Systems. Saf. Sci. 2015, 79, 126–138.

Chin, J.; Diehl, V.; Norman, K. Development of an instrument measuring user satisfaction of the human-computer interface. In Proceedings of the SIGCHI conference on Human factors in computing systems, Washington, DC, USA, 15–19 May 1988; pp. 213–218.

Corrêa, D.G.; Enembreck, F.; Silla, C.N. An investigation of the hoeffding adaptive tree for the problem of network intrusion detection. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 4065–4072.

Dalvi, N.; Domingos, P.; Sanghai, S.; Verma, D. Adversarial classification. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Seattle, WA, USA, 22–25 August 2004; pp. 99–108.

D'Amico, A.; Buchanan, L.; Goodall, J.;Walczak, P. Mission Impact of Cyber Events: Scenarios and Ontology to Express the Relationships between Cyber Assets, Missions, and Users; Tech. Rep. OMB No. 0704-0188; AFRL/RIEF, US Defence Technical Information Center; Fort Belvoir, VA, USA, 2009.

De Barros Barreto, A.; Costa, P.; Yano, E. Using a semantic approach to cyber impact assessment. In Proceedings of the 8thConference on Semantic Technologies for Intelligence, Defense, and Security (STIDS 2013), Fairfax, VA, USA, 13–14 November 2013; pp. 101–108.

Defense Science Board, 2012. Task Force Report: The Role of Autonomy in DoD Systems, Washington, D.C.: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics.

Defense Science Board, 2016. Summer Study on Autonomy, Washington, D.C.: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics.

Dekker, S.; Hummerdal, D.; Smith, K. Situation awareness: Some remaining questions. Theor. Issues Ergon. Sci. 2008, 11, 131–135.

Demertzis, K.; Iliadis, L. Evolving Computational Intelligence System for Malware Detection. In Advanced Information Systems Engineering Workshops; Lecture Notes in Business Information Processing; Springer: Cham, Switzerland, 2014; Volume 178, pp. 322–334.

Demertzis, K.; Iliadis, L. Evolving Smart URL Filter in a Zone-based Policy Firewall for Detecting Algorithmically Generated Malicious Domains. In Statistical Learning and Data Sciences;

Lecture Notes in Computer Science; Gammerman, A., Vovk, V., Papadopoulos, H., Eds.; Springer: Cham, Switzerland, 2015; Volume 9047.

Demertzis, K.; Iliadis, L. Ladon: A Cyber-Threat Bio-Inspired Intelligence Management System. J. Appl. Math. Bioinform. 2016, 3, 45–64.

Demertzis, K.; Tziritas, N.; Kikiras, P.; Llopis Sanchez, S.; Iliadis, L. The Next Generation Cognitive Security Operations Center: Network Flow Forensics Using Cybersecurity Intelligence. Big Data Cogn. Comput. 2018, 2, 35.

Demertzis, K.; Tziritas, N.; Kikiras, P.; Llopis Sanchez, S.; Iliadis, L. The Next Generation Cognitive Security Operations Center: Adaptive Analytic Lambda Architecture for Efficient Defense against Adversarial Attacks. Big Data Cogn. Comput. 2019, 3, 6.

Dietterich, T.G. Ensemble methods in machine learning. In Multiple Classifier Systems; Kittler, J., Roli, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2001; Volume 1857, pp. 1–15.

Dressler, J.; Bowen, C.; Moody,W.; Koepke, J. Operational data classes for establishing situational awareness in cyberspace. In Proceedings of the 6th International Conference On Cyber Conflict (CyCon 2014), Tallinn, Estonia, 3–6 June 2014; pp. 175–186.

Endsley, M. Measurement of situation awareness in dynamic systems. Hum. Factors J. Hum. Factors Ergon. Soc. 1995, 37, 65–84.

Endsley, Mica. (1995). Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32-64. Human Factors: The Journal of the Human Factors and Ergonomics Society. 37. 32-64. 10.1518/001872095779049543.

Endsley, M. Situational awareness misconceptions and misunderstanding. J. Cogn. Eng. Decis. Mak. 2016, 9, 4–32. Stevens, S. Measurement, Statistics, and the Schemapiric View. Science 1968, 161, 849–856.

Endsley, M.; Selcon, S.; Hardiman, T.; Croft, D. A Comparative Analysis of Sagat and Sart for Evaluations of Situation Awareness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Chicago, IL, USA, 5–9 October 1998; pp. 82–86.

Enrico Bertini and Denis Lalanne, Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery, SIGKDD Explor. Newsl. 11, 2 (May 2010), 9-18.

Esteve M., Pérez I., Palau C., Carvajal F., Hingant J., Fresneda M.A., Sierra J.P., Cyber common Operational Picture: A tool for Cyber Hybrid Situational awareness improvement, NATO IST-148 Symposium on Cyber Defence Situational Awareness, October 2016.

Evangelopoulou, M.; Johnson, C. Attack Visualisation for Cyber-Security Situation Awareness. In Proceedings of the 9th IET International Conference on System Safety and Cyber Security, Manchester, UK, 15–16 October 2014; pp. 937–942.

Fawcett, T. An introduction to ROC analysis. In Pattern Recognition Letters; Elsevier Science Inc.: Amsterdam, The Netherlands, 2006; Volume 27, pp. 861–874.

Flemisch, F. O., Schindler, J., Kelsch, J., Schieben, A., & Damböck, D. (2008). Some Bridging Methods towards a Balanced Design of Human-Machine Systems, Applied to Highly Automated Vehicles. Applied Ergonmics International Conference. Las Vegas, USA.

Franke, U.; Brynnielsson, J. Cyber Situational Awareness—A Systematic Review of the Literature. Comput. Secur. 2014, 46, 18–31.

Giacobe, N. Measuring the Effectiveness of Visual Analytics and Data Fusion Techniques on Situation Awareness in Cyber-Security. Ph.D. Thesis, The Pennsylvania State University, State College, PA, USA, 2012.

Gutzwiller, R.; Hunt, S.; Lange, D. A task analysis toward characterizing cyber-cognitive situation awareness (CCSA) in cyber defense analysts. In Proceedings of the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), San Diego, CA, USA, 21–25 March 2016.

Haffner, P.; Sen, S.; Spatscheck, O.; Wang, D. ACAS: Auto-mated Construction of Application Signatures. In Proceedings of the ACM SIGCOMM, Philadelphia, PA, USA, 22–26 August 2005; pp. 197–202.

Hall, P.; Park, B.U.; Samworth, R.J. Choice of neighbor order in nearest-neighbor classification. Ann. Stat. 2008, 36, 2135–2152.

Hecker, A. On system security metrics and the definition approaches. In 2008 Second International Conference on Emerging Security Information, Systems and Technologies, 2008.

Holz, T.; Gorecki, C.; Rieck, K.; Freiling, F. Measuring and detecting fast-flux service networks. In Proceedings of the Network & Distributed System Security Symposium, San Diego, CA, USA, 10–13 February 2008.

Hsu, C.-H.; Huang, C.-Y.; Chen, K.-T. Fast-flux bot detection in real time. In International Workshop on Recent Advances in Intrusion Detection; Springer: Berlin/Heidelberg, Germany, 2010.

Huang, G.-B. An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. Cogn. Comput. 2014, 6, 376–390.

Hubel, D.H.; Wiesel, T.N. Brain and Visual Perception: The Story of a 25-Year Collaboration; Oxford University Press: Oxford, UK, 2005; p. 106. ISBN 978-0-19-517618-6.

Jackson, M. Software Requirements & Specifications: A Lexicon of Practice, Principles and Prejudices; CM Press/Addison-Wesley Publishing Co.: New York, NY, USA, 1995.

Joint Doctrine Analysis Division. Operation assessment. Technical Report 1-15, Joint Doctrine Note, 2015.

Jyh-Shing Roger Jang, Chuen-Tsai Sun, and Eiji Mizutani. Neuro-Fuzzy and Soft Computing. Matlab Curriculum Series. Prentice Hall, Upper Saddle River, NJ (USA), 1997.

Katasonov, A.; Sakkinen, M. Requirements quality control: A unifying framework. Requir. Eng. 2006, 11, 42–57.

Klir, George J. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, Upper Saddle River, NJ (USA), 1995.

Kott, A. et al., 2019. Autonomous Intelligent Cyber Defense Agent (AICA) Reference Architecture, Release 2.0, Adelphi, MD: US Army Research Laboratory, ARL SR-0421, September 2019, available from https://arxiv.org/abs/1803.10664

Kuncheva, L. Combining Pattern Classifiers: Methods and Algorithms; Wiley: Hoboken, NJ, USA, 2004.

Lenders, V.; Tnner, A.; Blarer, A. Gaining an Edge in Cyberspace with Advanced Situational Awareness. IEEE Secur. Priv. 2015, 13, 65–74.

Llansó, T.; McNeil, M.; Noteboom, C. Multi-Criteria Selection of Capability-Based Cybersecurity Solutions. In Proceedings of the 52nd Hawaii International Conference on System Sciences, Maui, HI, USA, 8–11 January 2019.

Llopis S. et al., A comparative analysis of visualisation techniques to achieve cyber situational awareness in the military, 2018 International Conference on Military Communications and Information Systems (ICMCIS), 2018, pp. 1-7.

Llopis Sanchez, S.; Mazzolin, R.; Kechaoglou, I.; Wiemer, D.; Mees, W.; Muylaert, J. Cybersecurity Space Operation Center: Countering Cyber Threats in the Space Domain. In Handbook of Space Security; Springer: Cham, Switzerland, 2019.

Llopis Sanchez, S.; Sandoval Rodriguez-Bermejo, D.; Daton Medenou, R.; Pasqual de Riquelme, R.; Torelli, F.; Maestre Vidal, J. Tackling Verification and Validation Techniques to Evaluate Cyber Situational Awareness Capabilities. Mathematics 2022, 10, 2617.

Lif, P.; Granasen, M.; Sommestad, T. Development and validation of technique to measure cyber situation awareness. In Proceedings of the International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA), London, UK, 19–20 June 2017; pp. 1–8.

Liu, P.; Jajodia, S.; Albanese, M.; Subrahmanian, V.; Yen, J.; McNeese, M.; Hall, D.; Gonzalez, C.; Cooke, N.; Reeves, D.; et al. Computer-Aided Human Centric Cyber Situation Awareness. Theory and Models for Cyber Situation Awareness; Springer: Cham, Switzerland, 2017.

Losing, V.; Hammer, B.; Wersing, H. KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016; pp. 291–300.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1. University of California Press. pp. 281–297.

Mahoney, S.; Roth, E.; Steinke, K.; Pfautz, J.;Wu, C.; Farry, M. A Cognitive Task Analysis for Cyber Situational Awareness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Francisco, CA, USA, 27 September–1 October 2010; Volume 54, pp. 279–283.

Malviya, A.; Fink, G.; Sego, L.; Endicott-Popovsky, B. Situational Awareness as a Measure of Performance in Cyber Security CollaborativeWork. In Proceedings of the 8th International Conference on Information Technology: New Generations, Las Vegas, NV, USA, 11–13 April 2011; pp. 937–942.

Mao, J.; Jain, A.K.; Duin, P.W. Statistical pattern recognition: A review. IEEE Trans. Pattern Anal. Mach. Intell. 2000, 22, 4–37.

Medenou Choumanof, R.D.; Llopis Sanchez, S.; Calzado Mayo, V.M.; Garcia Balufo, M.; Páramo Castrillo, M.; González Garrido, F.J.; Luis Martinez, A.; Nevado Catalán, D.; Hu, A.; Rodríguez-Bermejo, D.S.; Pasqual de Riquelme, G.R.; Sotelo Monge, M.A.; Berardi, A.; De Santis, P.; Torelli, F.; Maestre Vidal, J. Introducing the CYSAS-S3 Dataset for Operationalizing a Mission-Oriented Cyber Situational Awareness. Sensors 2022, 22, 5104.

Mees W. and Debatty T., "An attempt at defining cyberdefense situation awareness in the context of command & control," 2015 International Conference on Military Communications and Information Systems (ICMCIS), 2015, pp. 1-9.

Mees W. and Debatty T., “Multi-agent system for APT detection,” in 2014 IEEE International Symposium on Software Reliability Engineering Workshops, Nov 2014, pp. 401–406.

Mees W., Llopis S., Debatty T., Achieving cyber situation awareness through a multi-aspect 3D operational picture, NATO IST-148 Symposium on Cyber Defence Situational Awareness, October 2016.

Mercaldo, F.; Martinelli, F. Tor traffic analysis and identification. In Proceedings of the 2017 AEIT International Annual Conference, Cagliari, Italy, 20–22 September 2017; pp. 1–6.

Miehling, E.; Rasouli, M.; Teneketzis, D. A POMDP Approach to the Dynamic Defense of Large-Scale Cyber Networks. IEEE Trans. Inf. Forensics Secur. 2018, 13, 2490–2505.

Milenkoski, A.; Vieira, M.; Kounev, S.; Avritzer, A.; Payne, B. Evaluating Computer Intrusion Detection Systems: A Survey of Common Practices. ACM Comput. Surv. 2015, 48, 1–41.

Montieri, A.; Ciuonzo, D.; Aceto, G.; Pescapé, A. Anonymity Services Tor, I2P, JonDonym: Classifying in the Dark. In Proceedings of the 2017 29th International Teletraffic Congress (ITC 29), Genoa, Italy, 4–8 September 2017; pp. 81–89.

Muralidharan, T.; Cohen, A.; Gerson, N.; Nissim, N. File Packing from the Malware Perspective: Techniques, Analysis Approaches, and Directions for Enhancements. ACM Comput. Surv. 2022.

NIST. 800-55, security metrics guide for information technology systems. 2003.

Özyurt, E., Döring , B., & Flemisch, F. (2013). Simulation Based Development of a Cognitive Assistance System for Navy Ships. IEEE CogSIMA. San Diego.

Parasuraman, R.; Sheridan, T.; Wickens, C. Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. J. Cogn. Eng. Decis. Mak. 2018, 2, 140–160.

Price, P.; Leyba, N.; Gondreey, M.; Staples, Z.; Parker, T. Asset criticality in mission reconfigurable cyber systems and its contribution to key cyber terrain. In Proceedings of the 50th International Conference on Systems Sciences (HICSS 2017), Waikoloa Village, HI, USA, 4–7 January 2017; pp. 446–456.

Rasmussen, J., & Goodstein, L. P. (1987). Decision Support in Supervisory Control of High-risk Industrial Systems. Automatica, 23(5), 663-671.

Rose, S., Borchert, O., Mitchell, S. and Connelly, S. (2020), Zero Trust Architecture, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.SP.800-207

Sagduyu, E.; Ephremides, A. A Game-Theoretic Analysis of Denial of Service Attacks in Wireless Random Access. In Proceedings of the 2007 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, Limassol, Cyprus, 16–20 April 2007; pp. 1–10.

Sagduyu, Y.E.; Berryt, R.A.; Ephremides, A. Wireless jamming attacks under dynamic traffic uncertainty. In Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Avignon, France, 31 May–4 June 2010; pp. 303–312.

Salmon, M.; Stanton, N.;Walker, G.; Baber, C.; Jenkins, D.; Mcmaster, R.; M.S., Y. What really is going on? Review of situation awareness models for individuals and teams. Theor. Issues Ergon. Sci. 2008, 9, 297–323.

Sawilla R.E and Wiemer D.J., Automated computer network defence technology demonstration project (ARMOUR TDP): Concept of operations, architecture, and integration framework, 2011 IEEE International Conference on Technologies for Homeland Security (HST), Waltham, MA, USA, 2011, pp. 167-172.

Schölkopf B. et al., Nonlinear component analysis as a kernel eigenvalue problema, Neural Computation, Volume 10 Issue 5, July 1, 1998, Pages 1299 – 1319, MIT Press.

Schulz, A.; Kotson, M.; Zipkin, J. Cyber Network Mission Dependencies; Technical Report 1189; Massachusetts Institute of Technology, Lincoln Laboratory: Lexington, MA, USA, 2015.

Shalev-Shwartz, S.; Singer, Y.; Srebro, N.; Cotter, A. Pegasos: Primal estimated sub-gradient solver for SVM. Math. Program. 2011, 127, 3–30.

Shameli-Sendi, A.; Louafi, H.; Wenbo, H.; Cheriet, M. Dynamic Optimal Countermeasure Selection for Intrusion Response System. IEEE Trans. Dependable Secur. Comput. 2016, 15, 755–770.

Shiravi, H.; Shiravi, A.; Ghorbani, A. A Survey of Visualization Systems for Network Security. IEEE Trans. Vis. Comput. Graph. 2012, 18, 1313–1329.

Tambe, M., Casey,W., Eds.; GameSec 2016; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9996.

Theron, P. et al. (2020). Reference Architecture of an Autonomous Agent for Cyber Defense of Complex Military Systems. In: Jajodia, S., Cybenko, G., Subrahmanian, V., Swarup, V., Wang, C., Wellman, M. (eds) Adaptive Autonomous Secure Cyber Systems. Springer, Cham.

Tsiropoulou, E.E.; Baras, J.S.; Papavassiliou, S.; Qu, G. On the Mitigation of Interference Imposed by Intruders in Passive RFID Networks. In Decision and Game Theory for Security; Zhu, Q., Alpcan, T., Panaousis, E.,

Tsoumakas, G.; Angelis, L.; Vlahavas, I.P. Selective fusion of heterogeneous classifiers. Intell. Data Anal. 2005, 9, 511–525.

Vinagre, J.; Jorge, A.M.; Gama, J. Evaluation of recommender systems in streaming environments. In Proceedings of the Workshop on 'Recommender Systems Evaluation: Dimensions and Design' (REDD 2014), Silicon Valley, CA, USA, 10 October 2014.

Webb, G.I.; Zheng, Z. Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques. IEEE Trans. Knowl. Data Eng. 2004, 16, 980–991.

William, H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. Section 16.5. Support Vector Machines. In Numerical Recipes: The Art of Scientific Computing, 3rd ed.; Cambridge University Press: New York, NY, USA, 2007; ISBN 978-0-521-88068-8.

Yadav, S.; Reddy, A.K.K.; Reddy, A.L.N.; Ranjan, S. Detecting Algorithmically Generated Domain-Flux Attacks with DNS Traffic Analysis. IEEE/ACM Trans. Netw. 2012, 20, 1663–1677.

Yamato, Y.; Kumazaki, H.; Fukumoto, Y. Proposal of Lambda Architecture Adoption for Real Time Predictive Maintenance. In Proceedings of the 2016 Fourth International Symposium on Computing and Networking (CANDAR), Hiroshima, Japan, 22–25 November 2016; pp. 713–715.

Yi Cheng, Julia Deng, Jason Li, Scott A DeLoach, Anoop Singhal, and Xin-ming Ou. Metrics of security. In Cyber Defense and Situational Awareness, pages 263-295. Springer, 2014.

Zimek, A.; Vreeken, J. The blind men and the elephant: On meeting the problem of multiple truths in data from clustering and pattern mining perspectives. Mach. Learn. 2015, 98, 121–155.

Zimmermann, A. The Data Problem in Data Mining. SIGKDD Explor. Newsl. 2015, 16, 38–45.

Zhou, Z.H. Ensemble Methods: Foundations and Algorithms; CRC Press: Boca Raton, FL, USA, 2012.

Žliobaité, I.; Bifet, A.; Read, J.; Pfahringer, B.; Holmes, G. Evaluation methods and decision theory for classification of streaming data with temporal dependence. Mach. Learn. 2015, 98, 455–482.