# A Continuous Risk Management Approach for Cyber-Security in Industrial Control Systems

*Submitted in partial fulfilment of the requirements for the degree of*

**Doctor of Philosophy**

*By*

**CAROLINA ADAROS BOYE**


Faculty of Computing, Engineering and the Built Environment

**October 2021**

**DECLARATION**

I hereby declare that the thesis entitled "A Continuous Risk Management Approach for Cyber-Security in Industrial Control Systems" submitted by me, for the award of the degree of Doctor of Philosophy to Birmingham City University is a record of bona fide work carried out by me under the supervision of Prof. Paul Kearney, Prof. Mark Josephs, and Mr. Hans Ulmer.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Stuttgart
Date: 13th of Oct, 2021

**Signature of the Candidate**

**Abstract**

In industrial networks, a cyber-incident can have, as a consequence, the interference with physical processes, which can potentially cause damages to property, to humans' health and safety, and to the environment. Currently most safeguards built into Industrial Control Systems provide mitigations against accidents and faults but are not necessarily effective against malicious acts. Moreover, even if cyber-threats can be contained, significant costs will be incurred whenever operations have to shut down in response to a cyber-attack. As there are important gaps in Industrial Control Systems, they have increasingly been targeted over the past decade, creating concern among the cyber-security and the process control engineering communities. Operators may be reluctant or unable to implement standard cyber-security controls in this type of systems because they might interfere with time-sensitive control loops, interrupt continuous operation or potentially compromise safety. This situation calls for a more proactive approach to monitor cyber-risks since many of them cannot be totally eliminated or properly controlled by preventative measures. Traditional risk management approaches do not address this, since they are not conceived to work at the same speed that changes can occur in cyber-security operations. This thesis aims to facilitate the adoption of Continuous Risk Management in industrial networks by proposing a risk assessment methodology focused mainly on the aspect of risk likelihood updates.

The approach proposed is based on a Continuous Risk Assessment Methodology, which is derived from a typical Risk Management process and modified to work in a continuous basis. The methodology consists of workflows and a description of each process involved, including its inputs and outputs. Additionally, a number of resources to support the implementation of the methodology on industrial environments were developed. These resources consist of the introduction and categorisation of the concept of "Indicator of Risk" (IoR), a knowledge base, containing a set of different categories of IoRs, named as the "IoR Library" and the implementation of this knowledge base on a Bayesian Network template. Finally, behavioural anomaly detection using sensors data is demonstrated to illustrate the use of IoRs based on data from physical processes as a resource to detect possible cyber-risks. These resources provided concrete means to address issues in industrial cyber-security risk management such as the availability and quality of information, the complexity of defining rules and identifying normal and abnormal states, the limited scope of academic work, and the lack of integration between risk management and cyber-security operations.

**Acknowledgements**

I want to thank my Director of Studies, Professor Paul Kearney for his patience, dedication, and good advice given throughout the research process, as well as Professor Mark Josephs and Mr. Hans Ulmer for their contributions and positive interactions as thesis supervisors. I would also like to thank everyone at Birmingham City University that made it possible for me to undertake my PhD studies, especially Professor Ali E. Abdallah and members of BCU staff who have been always caring and helpful, particularly Sue Witton, Ian McDonalds, and Noel Alonso. This thesis would also have not been the same without having the opportunity to exchange ideas with and learn from many researchers, industrial control engineers, and security experts. Among BCU researchers, I would like to name Dr. Lida Ghahremanlou, Dr. Esther Palomar, Professor Sharon Cox, Dr. Adel Aneiba, and Dr. Sheeren Fouad. It is also important to thank my former colleagues in Chile Juan Montecinos and Andres Lazo, who helped me get a good overview of the ICS world, and also people from the Bosch group, British Telecom (BT), and the IoT Security Foundation for their collaboration and input.

Taking the decision to do PhD studies takes you into a journey full of enriching experiences and discovery, but it also comes with many personal costs and difficulties. Hence, this journey would not have reached a good destination without the support of my family and friends to help me to go through the hardest parts. I would like, firstly, to give credit to my parents Patricia Boye and Pedro Adaros for never letting me doubt that I am able to accomplish any goal I put my mind into. This recognition extends to many other members of my family who have been there for me in different episodes of my life helping me to build strength, confidence, resilience, and develop so many other resources which have come handy in this stage. Special thanks to my friends who did their PhDs before me, especially Dr. Rodrigo Maldonado, Dr. Maria Vanessa Montañez, and Dr. Oana Jaycob, who inspired me and shared with me their experiences, and also to those friends who have been wise enough to never do a PhD but still encouraged me to keep going. They are too many to name them all. Finally, a big thank you to all my fellow doctoral students who shared with me part of this journey, particularly Dr. Antonio Nehme and Dr. Thomas Wagner.

**Contents**

**List of Figures**

## List of Tables

List of Publications

C. Adaros Boye, P. Kearney and M. Josephs, “Collective responsibility and mutual coercion in IoT botnets: a tragedy of the commons problem”. In International Workshop on Behavioral Analysis for System Security-BASS, 2018.

C. Adaros Boye, P. Kearney and M. Josephs, “Cyber-Risks in the Industrial Internet of Things (IIoT): Towards a Method for Continuous Assessment”. In International Conference on Information Security, 2018.

C. Adaros-Boye, P. Kearney and M. Josephs, “Continuous Risk Management for Industrial IoT: A Methodological View”. In International Conference on Risks and Security of Internet and Systems (CRiSIS), 2019.

C. Adaros-Boye, P. Kearney, M. Josephs and H. Ulmer, “An Indicators of Risk Library for Industrial Network Security,” In The 16th International Conference on Availability, Reliability and Security, 2021

# 1. Introduction

For decades, process control and electromechanical systems evolved independently from information systems. However, Operational Technologies (OT) nowadays can communicate using standard Information and Communications Technologies (ICT), in addition to more specific industrial communication protocols. This brings endless possibilities of reinventing almost every process, product and service and of creating new ones, but it also exposes industries to several cyber-security threats.

Increased connectivity has provided new opportunities for malicious actors, who have brought to the scene emergent cyber-threats, including some that specifically target OT. A cyber-incident in an industrial environment can stop production, interrupt critical services, damage infrastructure, cause environmental damage and put lives at risk [1] [2]. Consequently, a proactive approach to risk management is desirable to enable timely action to be taken, not only when an incident occurs but also when the conditions for a potential incident are met. In this thesis, a methodology and modelling framework are proposed for adapting standard cyber-risk management practices to work on a continuous basis. One of the key aspects of this thesis is proposing the identification of observations, which we will call "Indicators of Risk" (IoRs). The use of the terms "continuous", "near real-time", or even "real-time" risk assessment in the context of this thesis mean that under this approach, risks can be re-assessed at any given time based on these observations. The novelty of this is that processes to support this are not explicitly described in traditional risk management models.

The World Economic Forum in its 2019 report rated cyber-security risks as the second highest risk type after environmental and natural disasters [3]. In 2020, only water crises exceeded cyber-attacks in terms of impact [4]. McKinsey also recognizes cyber-security as an essential topic on the agenda of most top managers [5]. However, traditional risk management methodologies lack mechanisms to capture changes in risk factors at the velocity that they occur in cyber-security. This is critical when considering industrial networks and OT that are connected to ICT systems. Automation of industrial processes is expected not only to grow, but to become more integrated with ICT, as part of what is known as the Fourth Industrial Revolution. The heterogeneity, the complexity, and the lack of standardisation of these systems, most of which are not designed with security in mind, is exposing them to an increasing variety of cyber-threats. In order to be prepared, organisations not only need to minimise the vulnerabilities of their systems but they should also develop means to create and maintain continuous awareness of risk in what is known to be a fast-changing landscape.

The US National Institute for Standards and Technology (NIST) defines Information Security Continuous Monitoring (ISCM) as "maintaining ongoing awareness of information security, vulnerabilities, and threats to support organizational risk management decisions." [6]. Typically, organisations need to accept a certain level of risk to undertake their normal operations. This also applies to cyber-risks. In the case of industrial environments, the objective of reducing cyber-security risks can conflict with other business objectives, such as safety, performance, or availability [7]. Therefore, acknowledging and monitoring existing risks is considered to be a compensating control [8], which is crucial to prevent an incident happening or escalating. Nevertheless, usually a risk assessment is an off-line activity, which is performed prior to deployment of a system to determine appropriate security controls, and repeated at scheduled intervals or when significant changes are made. This gives space for the risk level to increase under certain circumstances without being detected.

The motivation for this research project is to contribute towards filling the current gaps in industry for effective assessment, monitoring and management of cyber-security risks in industrial systems. The scope of the project includes Industrial Control Systems (ICS) and the Internet of Things (IoT) as a broad concept, and more specifically the Industrial Internet of Things (IIoT). IoT is defined by the European Union Agency for Network and Information Security (ENISA) as an "emerging concept comprising a wide ecosystem of interconnected services and devices, such as sensors, consumer products and everyday smart home objects, cars, and industrial and health components. These technologies collect, exchange and process data in order to dynamically adapt to a specific context" [9]. Concepts that are tightly related to this definition in which this research has specific focus are OT, ICS, Industrial Internet of Things (IIoT), Cyber Physical Systems (CPS), and Industry 4.0. The differences between these overlapping terms are mainly due to their historical origins and are detailed in the Glossary of terms at the end of this thesis. It is assumed in this research that many industry sectors will face a transition period in which different generations of technologies will coexist during several years or even decades. Thus, it was considered relevant to include them all under the scope.

This thesis proposes a methodology for continuous cyber-risk management on systems that have cyber-physical components, with particular focus on industrial environments. This approach allows making use of existing

monitoring, detection, data analytics, and event management tools to provide timely and relevant information about how different events can relate to the possible development of a threat scenario. The main postulate is that continuously re-assessing risks can help to improve preparedness for preventing and mitigating emerging security incidents. By providing an up-to-date picture of the cyber-risk situation of a system security gaps that remain as "accepted" or as residual risks can be compensated through the monitoring of variables that can give indication of an increased risk probability.

While it is expected that the number and diversity of connected devices will grow in both the consumer market and industrial applications [10] [11], there are still important cyber-security challenges to be solved. Concerns about security are being ignored in the face of market drives for innovation, efficiency and productivity, since in most industry sectors technological changes and interruptions in operations are very costly. This means that, for example, stopping operations to do security updates or implement security controls is considered too costly and risky. Even if there was a clear understanding of cyber-security risks involved on not doing these updates or applying these controls (which is often not the case) it is likely that they will not be applied because their interference or potential interference with industrial operations. Hence, there is an urgent need for security solutions that can adapt to the requirements of industrial networks.

Security professionals shall also catch up and overtake attackers in their understanding of the potential for misuse of these interconnected technologies. One accepted fact is that the development of cyber-security frameworks, codes of practice, technical controls and mitigations for IoT and IIoT systems is far behind enterprise IT security [8] [12]. Even in the light of efforts made by both public and private organisations to develop security standards throughout the whole development lifecycle of industrial systems products and operations, there are several difficulties in applying these standards in practice. The complexity of these systems, the lack of standardisation, and in many cases having a large number of heterogeneous devices spread within a wide physical perimeter makes it difficult to ensure that all devices and communications are handled securely at all times. OT, which are increasingly adopting the IIoT paradigm, are employed in a wide variety of industries including manufacturing, energy and utilities, building management, transportation and aeronautics, health-care, mining and other heavy industries. Sabotage of one of these systems can not only cause considerable money losses but also cause physical damage and endanger health and safety, which means that consequences go beyond the information systems realm.

Despite the expectations about the future introduction of new generations of technologies in the industrial domain, which are meant to have built-in security features, many legacy systems cannot be easily replaced. These systems are still widely used and deployed and will need to coexist with the concept of Industry 4.0. One important concern is that their original design did not consider appropriate security controls for the current levels of connectivity. Many communication dynamics in industrial protocols are based on the assumption that if a device is connected to another device in the network, then it can be trusted, giving the chance for compromised devices to be connected. Remote connections used for maintenance can also be an important attack vector, in the cases in which external connections are allowed, however, even air-gapped systems can be breached with the help of an insider or by a compromise in the supply chain. Security mechanisms typically used in ICT could be insufficient or even infeasible when it comes to industrial and IoT applications [1]. Specialised security mechanisms are steadily being developed and adopted, however, internal and external conditions are always changing, and what is considered secure at a given moment can be proven to be insecure by a single breach. Therefore, it is important to find methods to improve preparedness for preventing and mitigating emerging security incidents and minimise their consequences.

One important aspect to understand about IIoT systems is that they embody several layers of IT and data processing and communication sub-systems, resulting in a complex architecture, which provides more opportunities for attack [13]. The computers and servers that run in manufacturing workstations represent an important attack vector. They often have outdated Windows operation systems with known vulnerabilities that can be used by an attacker to gain a foothold in the system. Even if they are not connected to public networks, it just takes an infected USB drive or laptop to download malicious files and software. Industrial computers and controllers, such as PLCs, are another attack vector since they work with insecure communication protocols that are vulnerable to unauthenticated connections, command injection, and denial of service attacks. Also, nowadays there is malware available that specifically targets this type of device. Network gateways such as routers, hubs, and couplers can also be compromised and used maliciously, as can smart sensors and actuators whose firmware can be tampered with in the supply chain. Maintaining an up-to-date inventory of all the devices on an industrial network, is quite challenging. The inventory might include equipment from different manufacturers, often supported by third parties. Knowledge is often distributed among different specialists,

each dedicated to a specific part of the system. Often there is no role within an organisation that has both an end-to-end view of all the processes and an appropriate cyber-security training, which makes it complex to have a holistic view of the security status of an industrial system. In addition, these systems usually cannot be monitored based only on traditional security monitoring tools such as intrusion detection and prevention systems.

Another important challenge for IIoT cyber-security is the conflict that can exist between cyber-security measures and operational objectives. For example, security updates can compromise continuity of operations [14], defences against brute force attack can lock out valid users in the middle of a crisis, and control flow integrity checks can restrict throughput [15]. Specific cyber-security threat detection and monitoring tools for industrial networks are emerging on the market. However, these solutions are still at an early stage of adoption. As several reports demonstrate, most companies do not yet have a mature strategy for their cyber-security operations or even a strategy at all, and they continue to make basic cyber-security hygiene mistakes [8] [15] [16]. Even when an organisation is aware and well informed about cyber-security risks, it might need to make trade-offs between security and other operational objectives. If an informed decision to accept a degree of cyber-security risks is made, it is prudent to monitor these risks in case assumptions are erroneous or conditions change.

Generally speaking, conventional risk management is an off-line activity in which risk assessments are performed on a periodic basis and/or triggered by a perceived change in risk factors. For example, there might be a security risk management board that meets on a monthly basis to review risk factors and decides whether any mitigating actions need to be taken. In many organisations, the information presented to such a board will be prepared manually, using spreadsheets or other general-purpose tools that have no integrated data feeds. It is not unusual for risks to be controlled through a dashboard displaying key indicators that are stored on a document-sharing platform and updated by the relevant responsible areas. Even in cases in which more sophisticated risk management tools are used, such as RSA Archer suite [17], the conclusions and consequent decisions made based on risk analysis can only be as good as the information available at the time assessments are performed. When speaking of technology-related risks, the fast pace of changes calls for risks to be re-assessed continuously in the light of new information. Therefore, risk monitoring should be performed in such a way that it can keep up with the speed of changes, which implies introducing a high degree of automation on data feeds collection and processing. Integration of security operations with security risk management becomes essential in the implementation of continuous risk management since real time information from security operations can provide an updated view of the risk landscape. At the same time, an updated evaluation of risks should support security operations by allowing a better prioritization of security controls, as well as timely reactions when there is a high risk for a threat to develop and transform into an imminent issue.

Over the last decade, several guidelines, methods, and codes of practice focused on cyber-security for Operational Technologies have been published. Some of these have a specific focus on critical infrastructure such as nuclear plants, the electrical grid, and oil and gas industries [18] [19], while others are more general in scope, such as ISA-99/IEC-62443 which covers Industrial Automation and Control Systems Security [20]. Approaches such as the Framework for Improving Critical Infrastructure Cyber-security from the US National Industry of Standards and Technology (NIST) [21] are meant to help industries to implement a cyber-security management system that addresses the needs, challenges, and requirements of an industrial network environment. However, even a well-prepared organisation is vulnerable to determined and sophisticated opponents, particularly when they use novel attacks employing zero-day exploits. Examples of such attacks are targeted ICS malware such as Stuxnet, TRITON, Conficker, Industroyer, LockerGoga, and Ekans. Another example is PLC Blaster, which is a proof of concept worm that operates on PLCs and is undetectable by anti-malware software and conventional network monitoring systems.

Frameworks for cyber-security management systems include risk management as an essential area to be implemented in an organisation, since it allows identifying, understanding and estimating the criticality level of potential security incidents. However, there is not much availability of detailed guidelines to integrate risk management with security operations and to introduce methods for continuous risk monitoring. Academic work in this subject is also relatively new and generally presents methods of narrow scope, which lack guidance for practical application.

## 1.1. Aim, objectives, and research questions

The aim of this thesis is to define a methodology to monitor cyber-risks during the operation of an Industrial Internet of Things or Industrial Control System in order to have a better visualisation of the risk landscape in real time. As a consequence, this should allow risk-based decisions to be made with updated information and overall improve the organisation's capability to prevent and respond to rapidly evolving threats. The expected outcome is describing a holistic approach that can be adapted to a broad variety of contexts and applications. In order to achieve this, the research project focused on the following specific objectives:

i. To increase cyber situational awareness related to ICS and IIoT systems in terms of giving visibility of the risk landscape at all times.

ii. To define a risk management approach that allows applying a continuous security risk assessment for ICS and IIoT through adapting and extending existing methods.

iii. To develop a risk assessment methodology and calculation algorithm to perform security risks calculations in ICS and IIoT during operation.

iv. To validate the methodology by developing worked examples and validating assumptions with different sources of reference in the industrial sector.

In order to achieve these objectives, a series of research questions were formulated, which are the following:

i. What information is needed in order to monitor security risks in IIoT/ICS?

ii. How can that information be derived from what can actually be measured?

iii. How can existing cyber-risk management frameworks be adapted for a more dynamic risk monitoring?

iv. How can the modifications on the traditional risk management paradigm be introduced?

The main contributions of this work are the following:

- A blueprint to introduce continuous risk management by proposing a modified version of the traditional risk management process.
- Use of real time information to extend the concept of Indicator of Compromise (IoC) to "Indicator of Risk" (IoR) as a means to monitor cyber-risks in quasi-real-time.
- A method of automated recalculation of risk probabilities based on Bayesian Networks, which makes use of IoRs to provide dynamic updates of the risk landscape.
- Inclusion of IoRs from different levels of an industrial system, including field devices and physical variables, such as input and output data from sensors and actuators which allows monitoring IIoT and ICS security considering all variables of the system beyond the ones that traditionally only refer to ICT.
- Guidelines on how to map indicators to different attack techniques, and consequently to risks.
- To promote a shift for a new generation of risk assessments which should be fully integrated with operations management to allow a better informed strategic and tactical decision making.

## 1.2. Why continuous risk management in ICS is important

According to risk management experts, organisations that conduct risk assessments are better prepared to deal with cyber security issues than those that do not [22]. A risk management process focuses on providing decision makers the "best possible information" for dealing with risks [23]. Risks are shaped by different internal and external factors that cannot always be controlled and which continuously change. Furthermore, information about threats and vulnerabilities which would define the likelihood of a system being targeted is often based on incomplete information and assumptions. New vulnerabilities in industrial devices and their associated software are found on a regular basis by security researchers, and product vendors and malicious actors are increasingly targeting industrial environments [15] [8] and developing new exploits, including Advance Persistent Threats

(APT). Additionally to this, having signs of an attack attempt can modify the initial belief of a risk being highly unlikely and lead to consider a review and re-evaluation of the risk and the decisions made based on this analysis. This could imply for an organisation to change their cyber-security posture.

Continuous risk management is important because there are constantly changing conditions that might require the modification of existing risk analysis, evaluations, and the security plans based on them may quickly become obsolete. Hence, it is necessary to challenge and to question the estimated probability of occurrence of an event, based on the evidence available. Cyber-risks situational awareness is crucial for visualising the current state of security in a system and to add context to different events that can be detected by security monitoring tools. In the case of IoT, IIoT, and ICS, there are more attack vectors and less visibility of the system from end-to-end in comparison with IT systems. Because cyber-security has been neglected during many years in industrial control applications and OT, which are systems that are often connected to IT networks, industrial environments have proven to be highly vulnerable to cyber-attacks.

As industrial equipment is a long-term investment and it is designed to last many decades, a large amount of legacy systems are still operative in which it is not possible to do security updates. This means that many vulnerabilities in ICS devices usually can only be mitigated by using network security controls, but they cannot really be fixed. Furthermore, even firmware and software security updates on newer or state-of-the-art devices are not possible since they might require stopping critical operations that run 24x7. In these cases it is often that security patches only can be applied during programmed maintenance windows which are only done a couple of times a year. While long term fixes and mitigations are planned, some security risks can be retained or accepted for long periods of time, which means that risk monitoring acts as a compensatory security control.

Continuous risk management allows having an up-to-date quantitative estimation on the likelihood of an undesirable event based on internal and external changes that can influence the risk level. The importance of this is that it can be continuously re-assessed if the risk level is still tolerable or if further actions are required to adjust the security posture. The continuous risk monitoring approach proposed is based on the correlation of events of different nature, so it can raise attention to situations that might remain unnoticed in a typical security monitoring use case. Since the conditions that are monitored are not necessarily an evidence of malicious actions by themselves but can give an indication of a risk when they are observed at the same time, it is made possible the investigation of issues based on a set of factors, which, in isolation, might be dismissed. Thus, it also could allow detecting an attack in its early stages or even novel or advanced attack procedures that might be harder to identify. Continuous risk monitoring in ICS is also meant to be sensitive to insider threats, which are performed by malicious actors with access to the infrastructure and privileged knowledge of the system and are in a better position to disguise their actions.

### 1.3. Overview of the proposition

Given the importance of cyber-risks situational awareness and the fact that industrial systems have proven to be highly vulnerable to cyber-attacks, this research project looks into how to do a continuous risk assessment for ICS and IIoT. The proposition developed in this thesis is based on gathering information on near real-time, which can allow detecting changes in the risk landscape of the system and check for signs of emerging risks. By integrating cyber-security risk management and cyber-security operations, it is intended to make a more efficient use of the available information than when these two disciplines operate separately and provide a holistic view of the cyber-risks to which the system is exposed and how they evolve in time. The main focus is on using this information to update risk likelihood estimations.

Companies nowadays have large amounts of data available, however they do not always have the capabilities to process it in a way that can provide meaningful results or all the necessary human resources to analyse it. By using a risk-based approach it is aimed to transform this data into information that can be prioritised according to its meaningfulness and relevance on a given context. For example, a single IoR can be manifested in a negligible increase in one or more risks, however a combination of IoRs can result in a more significant increase on risk levels. Hence, the approach proposed should give decision makers information that is both relevant and timely. In order to provide a systematic method to process security-related data IoRs are defined as the observation of events and conditions of different nature that, depending on their state, can be related to hazardous or insecure conditions, and ultimately to the likelihood of an attack technique being used against the system.

When an IoR can provide evidence of the perpetration of an attack using a defined technique with a higher degree of certainty, it can be considered to be an “Indicator of Compromise” (IoC) which needs to be evaluated not only as part of the risk analysis but also as part of the security operations process. When an indicator or combination of indicators can reveal an attack in progress, they are considered to be “Security Alerts”, which means that an immediate response is required, then it becomes part of the incident management processes. Figure 1.1 shows the different sorts of indicators and their relationship, and use. Hence, IoCs are considered as a sub-set of IoRs. If an IoR can reveal in a specific context the presence of a threat with an established degree of certainty, then it can be considered an IoC.

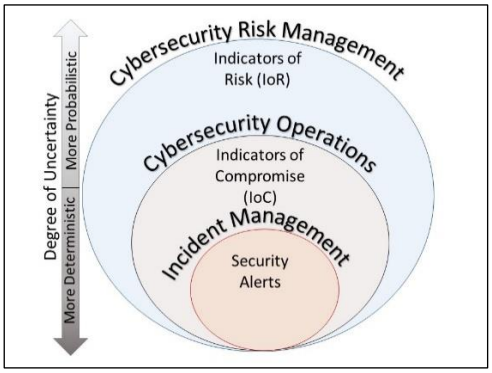



*Figure 1.1: Hierarchy of Security Risk Indicators*

To generate IoRs in real-time, information from monitoring and detection tools is required, which is processed through rules that which define when the IoR is observed. Figure 1.2 shows the main building blocks of the architecture proposed for continuous risk assessment. The resulting outputs are updated risk scores, whose behaviour should provide the risk analyst evidence based on operational data that can be used to adjust the risk treatment plan accordingly and to advise decision makers on security related issues. The left hand side of the figure shows a SIEM (Security Information and Event Management) based infrastructure such as is used in many enterprise Security Operations Centres (SOCs), and will increasingly be used for IIoT systems. State-of-the-art SIEM tools can also have analytics plug-ins embodying more sophisticated data analytics algorithms that can be used e.g. to implement anomaly and misbehaviour detection techniques. This would allow monitoring not only of security-related variables from computers, servers, and network devices, but also of physical variables from sensors and actuators. This last part constitutes one of the novel aspects of this proposal, which would allow use of information from field devices to provide complementary information that is independent from the one provided by the security monitoring tools. On the right-hand side of the diagram, the Risk Assessment module makes use of security alerts and event detection obtained from the SIEM. A risk calculation engine updates risk scores based on IoRs. The effect of the IoRs on the risk scores is determined by a Bayesian Network (BN), which is defined in a setup process establishing relationships between IoRs and known attack techniques.

The inputs for the risk scores adjustments will come from a variety of sources, which includes system variables, such as configuration parameters from endpoints, field devices, and networks at different levels of the system, security detection tools both IT and OT specific, input and output data (I/O) and logs. Additionally, information from external security sources such as threat intelligence or new vulnerability information can also be used to adjust risk scores. However, this research will focus more on the monitoring of internal sources. Data feeds can be online or batch, but the expectation will be at least for information coming from security detection tools to be online.

One of the main aspects of the present work is to describe a Continuous Risk Assessment methodology to implement the proposed approach by adapting traditional risk management standards to work in a continuous fashion. A process-oriented view of the methodology is described in terms of workflows and activities with their expected inputs and outputs. Through this perspective, activities that are covered by a traditional risk management process are identified, as well as those which need to be specifically defined for a continuous monitoring and reassessment of risks. Using the ISO/IEC 27005 [24] standard as a reference, a blueprint of how to integrate standard cyber-risk management practices with a continuous risk assessment paradigm is presented. The methodology considers three phases: the Baseline (or initial) Risk Assessment, the Transition phase and the Continuous Risk Assessment phase. The baseline risk assessment follows a standard risk assessment process, but includes additional activities and specific work products in order to enable the

development of the other two consecutive phases. In the transition phase the inputs for the continuous risk assessment are defined, Bayesian Networks are built, and the supporting tools are configured. In the Continuous Risk Assessment phase the risk scores calculated in the baseline risk analysis get updated in the light of new information during operational time.

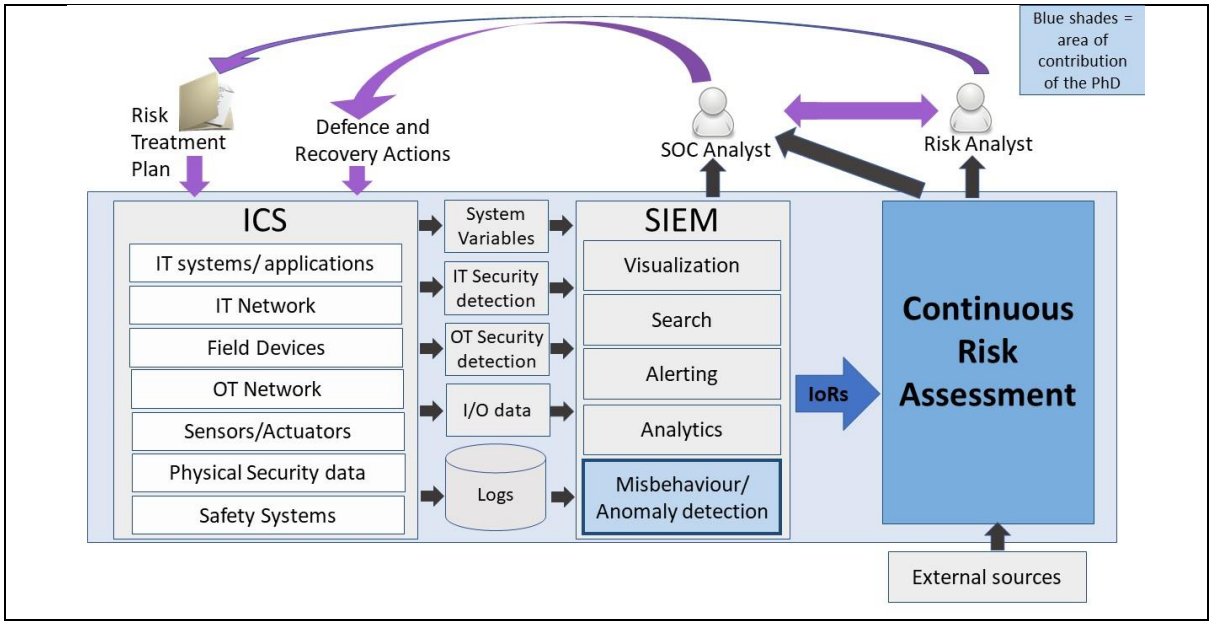



*Figure 1.2: Proposed Architecture for Continuous Risk Assessment*

The purpose of the methodology proposed is to allow improved visibility of cyber-security risks and their updated status in IIoT and ICS systems. Since lack of visibility and cyber-situational awareness is an important problem in an industrial environment, the opinion that was formed during this research was that this project is addressing a relevant problem in this field. The incorporation of intrusion detection based on misbehaviour of expected data patterns obtained from physical components of the system is also a key aspect since these devices are usually a blind spot for traditional detection tools. In addition, the integration of detection techniques in risk analysis, in general, is an idea not fully implemented currently in industry, particularly when speaking of risks that are contextualised within the business impacts rather than the consequences at the system level.

The scope of this thesis covers only modifications on risk scores based on changes on the estimations of risk likelihood while considering the impact component as constant. This premise is not necessarily always true since impacts could also be modified dynamically based on changes on the system's conditions. However, for simplification matters, these scenarios are not explicitly considered in the processes and examples described. Proper consideration of changes on impacts, including methods and instruments to define under which conditions impacts change and how this is measured and used to modify risk scores would be interesting as a complement of the present work.

## 1.4. Structure of this thesis

This thesis is divided into three parts, each one of which comprehends one or more chapters. The first part, which consists of Chapters 1, 2, and 3 introduces the problem and provides a state-of-the-art review including relevant literature that the research was inspired on. The main goals of this state-of-the-art review are to illustrate the value that the ideas presented can create in industrial control networks security. The second part of the thesis consists on Chapters 4, 5, and 6, which are the central part, and expose different aspects of the continuous risk management approach for ICS and IIoT proposed in this investigation work. Finally, Chapter 7 provides the conclusions derived from the research and development undertaken. The following is the specific content of each one of the chapters:

**Chapter 1: Introduction**
Chapter 1 is dedicated to introduce the problem that is addressed in this thesis, and describes the aim and objective of the research project, its focus, and brief description of the solution proposed.

**Chapter 2: Cyber-security risk management**
In Chapter 2, a state-of-the-art review of cyber-security risk management, and continuous risk monitoring is provided, in order to identify some of the challenges and gaps that need to be addressed to provide a continuous cyber-risk monitoring methodology for industrial environments.

**Chapter 3: Cyber-security in industrial networks**

Chapter 3 describes the main aspects of the observations made through a state-of-the-art review of cyber-security in industrial environment. The main approaches for security monitoring in cyber-physical systems and the use of indicators to detect risk and compromise, as well as typical vulnerabilities and threats in industrial systems are highlighted. Following this, some security monitoring approaches that have been proposed in the literature are discussed, considering their potential to be used as part of the building blocks for continuous risk assessment. A summary of the main gaps and challenges in industrial cyber-security is also provided.

**Chapter 4: The Proposed Continuous Risk Assessment Methodology**

In Chapter 4, a detailed description of the proposed methodology is presented, including workflows and the key activities and their inputs and outputs. The methodology is illustrated by means of a use case based on a worked example.

**Chapter 5: Use of Indicators of Risk (IoRs) for Continuous Risk Assessment**

In Chapter 5, the concept of "Indicator of Risk" (IoR), which is already introduced in the previous chapters, is explored further including the proposal of an "IoR Library" which was developed as an extension of the ICS ATT&CK framework. A Bayesian Network (BN) template based in the IoR Library is proposed to further assist the implementation of the continuous risk assessment.

**Chapter 6: Physical-based anomaly detection applied to continuous risk monitoring**

Chapter 6 complements Chapters 4 and 5 by elaborating on a specific aspect of the proposal, which was considered to have a great potential on industrial networks security. This aspect is behavioural anomaly detection based on data from sensors that measure physical variables of the system. While this is considered in Chapters 4 and 5 as one of the types of possible IoRs, in Chapter 6 practical examples are developed using data from sensors obtained from both experimental and real life settings. This illustrates what is meant by sensor-based anomaly detection and how it would look in practice, highlighting the relevance of this type of monitoring for an integral security risk monitoring in cyber-physical systems since it addresses the physical part, which is often overlooked.

**Chapter 7: Conclusion**

Chapter 7 provides reflects on how the work developed in Chapters 4, 5, and 6 fulfil the aim and objectives and address the research questions presented in Chapter 1 and some of the gaps identified in Chapters 2 and 3. To conclude this thesis, the main contributions and the limitation of this research are discussed, followed by directions for further research and development.

# 2. Cyber-security risk management

One of many accepted definitions for the term "risk" is that it is "the effect of uncertainty on objectives" [25]. As the objectives of cyber-security are to preserve the confidentiality, integrity, and availability of a system, any event that might compromise these properties is considered a cyber-risk. The business impact of these negative effects should be also considered as part of the risk. Examples of business impact are the costs of a service interruption, of data loss or data exposure, and also the costs of putting in place defences and recovery actions, the cost of legal actions and of losing customers' trust. In an industrial context, this can extend to production losses, as well as to health, environment, and safety hazards. Considering risks and countermeasures to reduce and counteract them is an essential part of any business. Whether it is undertaken through a defined process based on a risk management framework, or informally on a case-to-case basis, a risk analysis will be usually done to make important decisions. In cyber-security, this is not different and security measures and controls will be put in place according to the perceived risk level of the decision makers.

Formal methods for conducting a risk assessment have the goal of making the rating of a risk level as objective as possible, avoiding biases produced by underestimating threats, by overconfidence on the existing security defences, or simply by lack of knowledge and understanding of the nature of the risks. Thus, despite its inability to give accurate results, doing risk assessments is not only widely accepted, but also mandatory in many industries and also recommended by most cyber-security management frameworks. One of the problems to be addressed in this thesis is that traditional formal methods for cyber-security risk management generally lack effective mechanisms to monitor different factors that influence the risk analysis in order to continually re-assess the risk level at any given time.

The rapidly evolving nature of cyber-security events plus the current availability of tools to monitor a great number of events and conditions in a system mean that cyber-security risk monitoring should evolve to be performed in real time. The first section of this chapter consists of a general overview of factors that define a cyber-risk. The second section is a state-of-the-art review that was conducted in order to understand better the most used risk management practices and the methods and tools for continuous cyber security risk assessment that have been proposed so far. In the third section, a discussion of the chapter is presented, including a gap analysis together with comments on the main challenges of conducting a continuous cyber security risk assessment in the context of industrial networks.

## 2.1. Factors for quantifying cyber-risks

In this section, the factors that shape a cyber-risk are discussed. These factors can be used to estimate the magnitude of a risk either qualitatively or quantitatively. As the main characteristics of a risk are uncertainty and the potential of causing damage, risks are normally defined by two factors: the likelihood and the impact. The first refers to how much chance there is for an event to happen, and the second to the effects that this event can have on an organisation's objectives. The term "likelihood" is used as opposed to the term "probability", because in the English language "probability" often implies a mathematical quantification [24] whereas "likelihood" has a more general interpretation. However, on some occasions these terms can be used interchangeably. It must be noted that in other languages, there is a single word for likelihood and probability. Another important clarification is that, in this context the impact will only refer to adverse effects, which means that we are speaking about "pure risk" in contrast to "speculative risk" which can also lead to positive outcomes [26], examples of this are gambling or investing in the stock market.

The main purpose of doing a risk analysis is to provide information to make rational decisions with incomplete data. By definition, risks are events with potentially-significant consequences that might or might not occur. It is necessary for an organisation to establish a risk management framework to identify and assess the risks that are considered relevant, and to define methods to rate the risks either in a qualitative or quantitative manner. A risk analysis should reflect how likely is for a particular event to happen and how significant would be the impact. Different approaches propose their own methods to estimate the likelihood of cyber-risks, however, most of them, one way or another define them as a function of the threat and the vulnerability [27] [28] [29]. The threat is essentially the possibility of a system to get targeted and depends on the motivation of the attacker and the availability of tools, techniques and knowledge to exploit the vulnerabilities of a system. Similarly,

vulnerability are the weaknesses that allow the possibility that an attack, if it occurs, will be successful; the weaker the defences of the system, the higher its vulnerability. According to this reasoning, a commonly used mathematical definition for cyber-risks is as the product of "threat", "vulnerability", and "impact", for which the simplest representation would be the following:

$$Risk = Threat \times Vulnerability \times Impact$$

Despite the exact formulation of cyber-risk can present variations from method to method, describing the likelihood as the product of the vulnerability and threat levels appears to be widely accepted in the security community. Regardless of the method used, the quantification of risks is not always straight forward or absolute, since they need to be considered in an appropriate context. Hence, the same risk could have a different likelihood quantification for different systems, depending on their implementation, the environment, the motivation and capability of threat agents, and particular characteristics of the industry sector of the business. Finding well-defined and systematic methods to quantify risk factors can allow different risks to be positioned on the same scale for comparison, making it easier to be evaluated and prioritised which enables the development of a cost-effective cyber-defence and recovery strategy.

The exact definition of risk factors depends on the method or standard used. Figure 2.1 shows a hierarchical diagram, which is reasonably representative of the main factors that most methods take into account one way or the other, to quantify risks. For the threat level estimation, some things to be taken into account are the amount of resources that an attacker requires in order to exploit a vulnerability, such as time, equipment, logistics, staff, and knowledge. Considering the amount of resources an adversary needs can give indication of which types of threat agents have such a level of funding and hence, which proportion of the threat population, given the right motivation, would actually be capable of a successful attack. Typical threat agent types are hackers, hacktivists, script kiddies, nation states, cyber terrorists, organized crime, and malicious insiders. For vulnerability estimation, it is usual to consider the exploitability or susceptibility to threats, as a main rating factor, which can be defined by several metrics. Security controls and mitigations, on the other hand, can act in counteracting the vulnerability level. The quantification of impact is usually done in monetary values, which allows comparison and addition. Impacts in monetary values can allow also to compare a cyber-security risks and the costs of risk mitigation plans, and also with other figures of a company. This allows non-technical managers to make decisions, which might define a cyber-security budget.

In many cases, the lower level factors shown in Figure 2.1 will not be explicitly defined in a calculation model, but they will be implicit in other factors or represented by other sub-factors. For example, in the Common Vulnerability Scoring System (CVSS) method [30], metrics such as attack vector, attack complexity, privileges required, and user interaction are used to measure the exploitability of a vulnerability. As much as the vulnerability is about the system itself and its defences, and the threat is about the adversaries' actions, these two variables are not independent. The more vulnerable a system is, the more likely is that it will become a target for opportunistic attackers. Therefore, it will not be rare to find that some methods for vulnerability estimation consider threat related metrics or factors and that threat estimation could also take in consideration the vulnerability of the system.

When speaking about impacts, these should be considered as the sum of the different costs for the business. The FAIR approach [30] differentiates two types of impact, which are primary losses, such as productivity, response expenses, and replacement costs, and in secondary losses, such as fines and judgments, competitive advantage and reputation. Technical impacts, which are usually defined as the potential effects of a risk on confidentiality, integrity, availability, and safety should be ultimately translated into business impacts by analysing which business processes can be affected. Rather than expressing these impacts in terms of absolute monetary values, it is important to relate them to the organisation's capacity to deal with them. For example, the OWASP risk metrics [31] considers a financial impact to be minimum if it is lower than the cost of fixing the vulnerability, and maximum if the consequence is bankruptcy. Intermediate levels are a function of their effect on the expected annual profit.

The quantification of risk factors has several challenges and limitations since many calculation methods are imprecisely defined, and give room to interpretation, and the choice of input values is often subjective. Added to this, in many cases the available information is incomplete and many estimations will carry a high degree of uncertainty. It must be noted that in analytics and management science, making assumptions is a valid approach to understanding uncertain scenarios. However, for transparent results, these assumptions need always to be

declared and validated, and the analysis shall be updated if there is evidence at any point that an assumption is wrong. This means that a risk analysis that was considered valid in a given point of time can become obsolete at the moment that risk factor estimations are proven to follow a flawed analysis because of a misleading assumption.

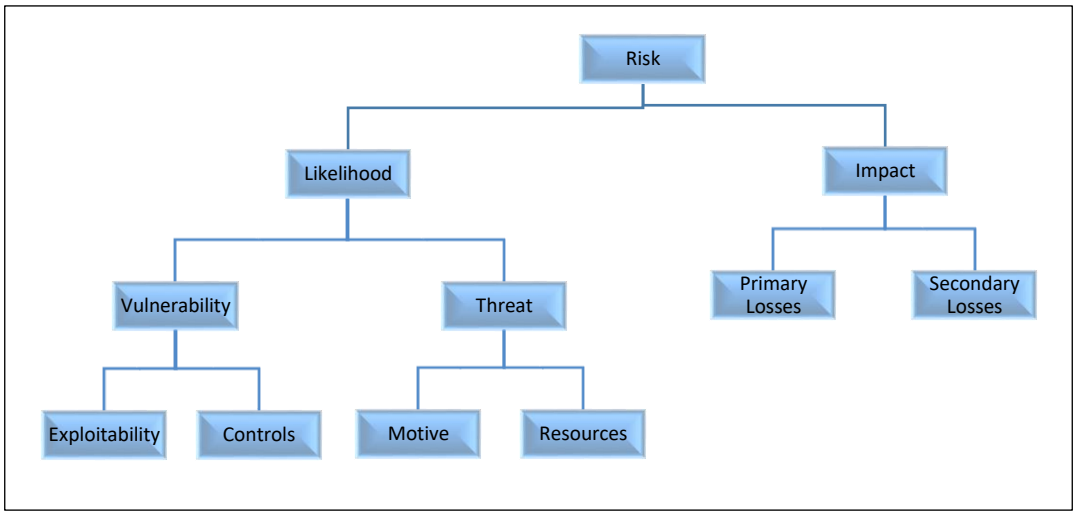



*Figure 2.1: Risk Factors*

## 2.2. State-of-the-art review of cyber-security risk management

The purpose of this state-of-the-art review is to understand how cyber-security risk management is currently being approached in the majority of organisations, and to identify existing initiatives related to continuous risk assessment.

Risks are estimated based on the conditions of the system and the information available at a given moment. However, conditions can change, and a traditional risk management system is usually not able to capture and process those changes and react accordingly in a short period of time. Establishing mechanisms to detect conditions and new information that could modify the level of security and use it to re-evaluate risks can allow making decisions based on updated information. Monitoring this information could provide "continuous awareness" about the security status of the systems, particularly focusing on those events that can inflict a significant damage to the business. Nevertheless, continuous cyber-security risk monitoring is not yet an industry practice, especially in industrial networks where cyber-security management, in general, is still immature.

In the first sub-section, an overview of cyber-security risk management frameworks is given to understand its objectives and main processes involved. In the second sub-section some of the most well-known cyber-risk analysis methods are described, followed with a mention on common industry practices in the adoption of these frameworks and methods in sub-section number three. The fourth sub-section focuses on work related to continuous risk assessment methods that was reviewed during the development of this research is given. Some of the general ideas or contributions from this work presented on continuous risk monitoring, analysis, and assessment were used as a reference for the work described in Chapters 4 to 6 of this thesis.

### 2.2.1. Cyber-security risk management frameworks

The European Union Agency for Cyber-security, ENISA, defines risk management as "the process of identifying, quantifying, and managing the risks that an organisation faces; a process aimed to obtain efficient balance between realizing opportunities for gains and minimizing vulnerabilities and losses" [32]. Thus, a good cyber-security strategy should weigh the cost of cyber-security controls against the cost of a potential cyber-security incident and decide rationally according to this comparison. In that same line, BSI specification PAS 555:2013 [33] describes the contribution of risk management as allowing an organisation to focus investment, minimize potential loss, improve operational effectiveness and efficiency, develop resilience, improve prevention and

incident management, and identify and mitigate cyber security risk throughout the organization. Thus, risk management is crucial to proper allocation of cyber-security resources and is regarded as such by most widely recognised cyber-security frameworks. To mention some examples, the ISO/IEC 27000 family of standards, includes the ISO/IEC 27005 standard for Information Security Risk Management [24], and NIST has a risk management framework for information systems and organizations [34] and also includes guidelines to self-assess cyber-security risks in its Framework for Improving Critical Infrastructure Cyber-security [21]. Other specific frameworks for Industrial Networks security such as IEC 62442 [35] also consider risk management as an essential part of a Cyber-Security Management System (CSMS).

Overall, it is not possible to put in place adequate security controls without a previous analysis of the most relevant threats for an organisation and of how well prepared it is to deal with them. Risk management provides methods to analyse and evaluate risks that are consistent over time and that deliver results that are useful in making decisions. A cyber-security risk assessment can be useful in determining how much security is good enough, and how well prepared an organisation is to deal with a cyber-security breach. The ISO/IEC 27005 standard [24] describes an iterative risk management model, which is represented in Figure 2.2 and will be used as a reference throughout this thesis. This model is widely used as a reference, and does not fundamentally conflict with other well-known frameworks, including those specific to industrial systems. Context Establishment defines the scope and the methods used in the risk assessment and the organisation's capability to deal with, and its attitude towards, risks. Risk Assessment comprises Risk Identification, in which threats, vulnerabilities and impacts are identified for each risk; Risk analysis, in which risks are quantified or rated; and Risk Evaluation, which defines which risks are acceptable and which should be treated.

A risk treatment plan should be executed, based on the risk assessment. This last should provide the decision-makers with the "best possible information" to prepare themselves for dealing with cyber-risks [23] and prepare such plan. Risk management and treatment decisions that are made as a result of a risk assessment should be based on an in depth knowledge about the system and its environment [27]. At the same time, knowledge about the business, its goals, priorities, and risk appetite [36], are also important elements to take in account when making decisions related to security risks.

Communication and Consultation, and Monitoring and Review are processes that take place throughout the whole risk management lifecycle. The communication and consultation process allows all stakeholders to agree on a common view regarding how risks are managed and the monitoring and review process identifies changes in risks and their factors as early as possible in order to maintain a complete overview of the risk landscape. The focus of this thesis is mostly on the Monitoring and Review process, which is regarded as a "Continuous Risk Assessment".

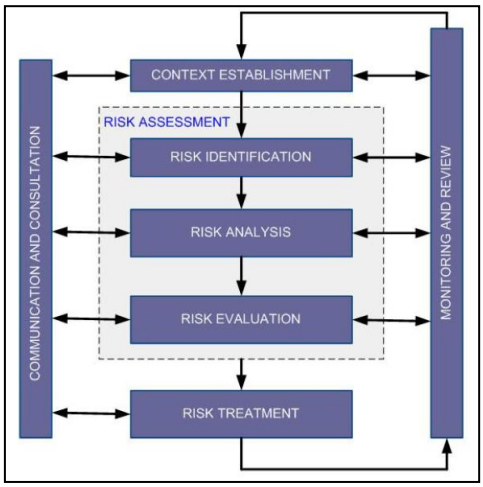



*Figure 2.2: ISO/IEC 27005 Risk Management Overview*

ISO/IEC 27005 and other frameworks mentioned in this sub-section are mostly high-level approaches that give a reference of "what" to do, leaving to each organisation to make their own choices regarding the "how" to perform the risk assessment. It is not unusual for an organisation to develop its own risk assessment method,

but most of the time it will be an adaptation of, or at least inspired by, well-known cyber-security risk assessment methods. Examples of this will be given in the next sub-section.

### 2.2.2. Cyber-security risk assessment methods

In general terms, a risk assessment method should provide information about what can go wrong, how likely it is that this will happen, and how severe would be the impact [27], which is a definition also considered in the method proposed. In cyber-security it is also very important to know how something can go wrong, which is the reason why vulnerabilities and threat assessments are an essential part of a risk assessment. A good quality of a risk assessment method is that it should define inputs and procedures that are realistic and feasible to be applied in an organisation and that it should be able to make the best use possible of the available information in order to assist decision-making. As the nature of cyber-security risk management, and risk management as a discipline in general, is to address the inability to control all factors on which risks depend, estimations cannot be expected to be accurate, but just only to appear reasonable. Hence, the focus should be on the process and the methods. On other words, the risk assessment process should be repeatable even if the results are not [37].

The core process of a risk assessment is the risk analysis, since it is where risks are quantified or rated, based on their likelihood and impact. The following well-known methods used to calculate risks, and risk likelihood factors (threat and vulnerability) were reviewed, having also in consideration that the focus of this research is on risk likelihood, rather than impact:

i. The FAIR method
ii. FAIR using Monte Carlo techniques
iii. Octave-Allegro
iv. MITRE Systems Engineering for Mission Assurance
v. CORAS
vi. CRAMM
vii. NIST Guide for Conducting Risk Assessments
viii. IRAM2
ix. OWASP Risk rating methodology
x. Compliance based risk analysis
xi. OTA (SANDIA)
xii. Common Vulnerability Scoring System (CVSS)
xiii. Cyber-threat Intelligence
xiv. Pentesting
xv. Vulnerability scan

#### 2.2.2.1. Overview of methods to quantify risks and risk factors

The following overview presents a brief description of different methods that can be used to quantify risks. In addition, methods to analyse threats and vulnerabilities were also reviewed since these methods can be combined and used to generate inputs for a cyber-risk analysis. Some methods were reviewed in more depth than others according to the interest that they presented for this research.

**i. The FAIR method**

The Factor Analysis for Information Risk method (FAIR) was developed by the Open Group and consists of specific guidelines to calculate risk scores through a set of pre-defined risk factors [23] [38] [39]. Figure 2.3 shows the different factors that are taken into account in the risk calculation.

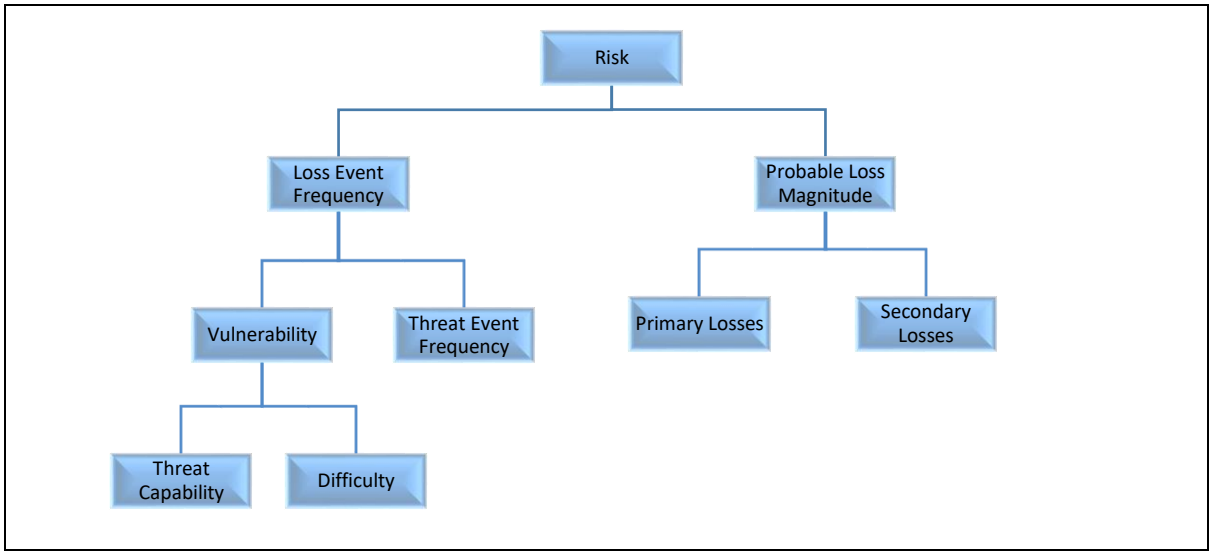



*Figure 2.3: Risk Factors defined by the FAIR method*

The Loss Event Frequency (LEF) is the rate at which successful attacks are expected to occur. It is estimated from two other factors, which are the Vulnerability and the Threat Event Frequency (TEF). TEF is the rate at which attacks are attempted, and Vulnerability is the probability that an attempted attack will succeed. The vulnerability is estimated from two other factors, the Threat Capability (TCap) and the Difficulty [23], which replaces what was formerly defined as "Control Strength" [38] [39]. This means that the level of vulnerability depends on the skill and resources of the attacker and how difficult it is to carry out a successful attack. The impact factor is termed "Probable Loss Magnitude" (PLM) and should be calculated considering both primary and secondary losses. The method also requires analysis of the Worst Case Scenario (WCS) to provide additional information, allowing one to get an idea of the possible range of losses, rather than just the amount considered as the "most likely".

For each factor from the lower level, FAIR provides a table that helps determining its appropriate level. For example, Threat Capability is categorised in five levels that go between very low to very high depending on the proportion of the threat population that is presumed to have the resources needed to perform the specific attack. Once all these lower level factors are estimated, the higher level factors are derived through evaluation matrices. Figure 2.4 shows, as an example, the matrix used to define the final risk score. FAIR is meant to be compatible with either ISO 27005 or NIST 800-37, providing guidelines or "cookbooks" to be used under each one of these frameworks.

| **Probable Loss Magnitude (PLM)** | | | | | |
|---|---|---|---|---|---|
| **Severe** | High | High | Critical | Critical | Critical |
| **High** | Moderate | High | High | Critical | Critical |
| **Significant** | Moderate | Moderate | High | High | Critical |
| **Moderate** | Low | Moderate | Moderate | High | High |
| **Low** | Low | Low | Moderate | Moderate | Moderate |
| **Very Low** | Low | Low | Moderate | Moderate | Moderate |
| **Loss Event Frequency (LEF)** | **Very Low** | **Low** | **Medium** | **High** | **Very High** |

*Figure 2.4: FAIR Risk Evaluation Matrix*

**ii. FAIR using Monte Carlo techniques**

The FAIR approach offers the alternative to do a quantitative analysis based on Monte Carlo simulation, as well as other stochastic methods [39]. This is done by representing each risk factor in terms of probability distributions rather than as single point indicators. For each factor, it is necessary to define a most likely value

and describe a distribution shape. The derive factors, and the final risk estimation will also be probability distributions obtained by doing a large number of risk calculations based on repeated random samples of values of the lower level risk factors. For example, if a probability distribution is assigned to "Threat Capability" and another probability distribution to "Difficulty" the probability distribution of "Vulnerability" will come as a distribution that is a result of multiplying many times random values coming from the Threat Capability distribution and random values coming from the "Difficulty" distribution.

A typical case would be to describe a symmetric shape distribution in which the mode, the median, and, consequently, the mean value are the same. This can be the case, if it is considered that the most likely value is the average. The distribution can also be skewed (asymmetric) to the left, if it is considered that the most likely value (the mode) is lower than the median, and skewed to the right if it is considered that the most likely value is higher than the median. Figure 2.5 shows an example describing how an attack mechanism that does not require too many resources or skills from a threat agent will have a TCap distribution skewed to the right and an attack that is more complex and hence, requires more resources and capabilities will have a TCap distribution skewed to the left.

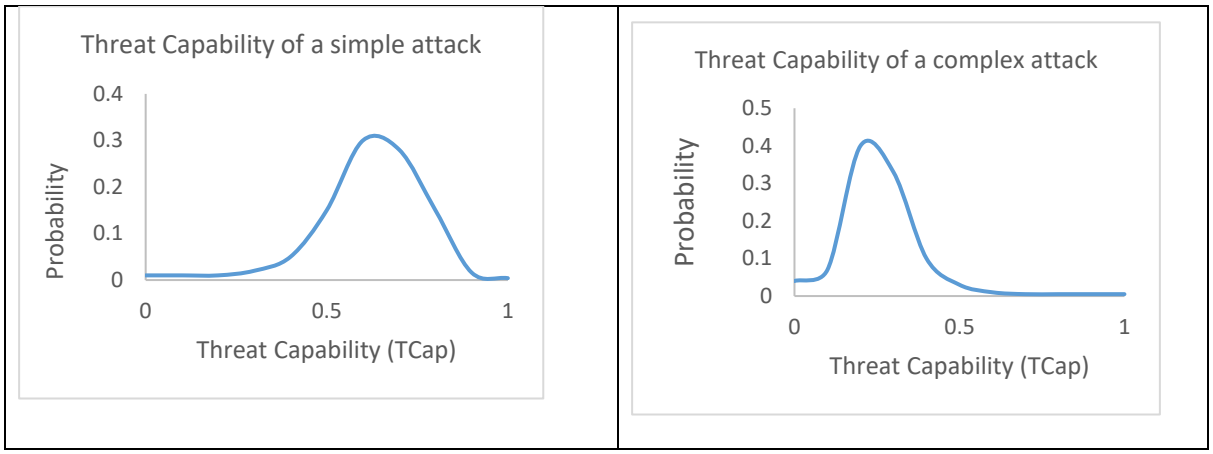



*Figure 2.5: Examples of Threat Capability Distributions*

The reason for these two distribution shapes is that if an attack does not require too many resources there will be a bigger proportion of the threat population that will be able to perform it (for example, more than 50%, which is the median). In an analogous way, for an attack that requires a high level of skills and resources is more likely that a fewer proportion of threat agents from the total population would be able to perform the attack.

Freund and Jones [23], who are recognised promoters of this approach, also suggest starting to use Monte Carlo Simulation in the level of the FAIR ontology that is considered more adequate for each case. For example, it would be also possible to directly define a probability distribution for vulnerability instead of deriving it from TCap and CS [23]. In any of these cases, the best way to build a probability distribution in absence of reliable data is to estimate a minimum, maximum, and most likely value and whether this most likely value is above or below the median.

**iii. OCTAVE-Allegro**

OCTAVE Allegro [40] was developed by the Software Engineering institute of Carnegie Mellon University (SEI) based on the Operationally Critical Threat, Asset, and Vulnerability Evaluation method (OCTAVE). This method can be performed in a workshop-style, and is supported with guidance through worksheets, and questionnaires. According to the SEI, it can also be done "without extensive organizational involvement, expertise, or input". The method is based in eight steps, which belong to four main areas where the organization starts establishing measurement criteria consistent with their drivers, identifies the assets and threats, analysed the risks and develop mitigation strategies. Risk is defined by combining a threat condition with an impact level, most of the variables involved are qualitative, and some impacts are expressed as ratios of the operational costs. Although

the vulnerability is not explicitly defined, by stating a qualitative value of likelihood of a threat scenario of occur, the stakeholders will be implicitly evaluating how vulnerable they are.

**iv. MITRE Systems Engineering for Mission Assurance**

The MITRE Corporation includes a cyber-risk management process in their Systems Engineering for Mission Assurance guides, which is part of the MITRE Systems Engineering guide [41]. Risk Assessment, as proposed by MITRE is based on the Threat Assessment and Remediation Analysis (TARA) method [42], which has two parts: the Cyber Threat Susceptibility Assessment (CTSA) and Cyber Risk Remediation Analysis (RRA). The risk analysis method recommended in the CTSA starts by identifying cyber-threats and subsequently evaluating their associated risks. Scores from 1 to 5 are assigned to different factors related to the threat and the impact. The criteria for scoring each one of the attributes is provided by a given risk scoring spreadsheet. Different "Tactics, Technics, and Procedures" (TTPs) are identified and assessed using this method assigning a risk score. To assign the risk score, the score of the different factors are consolidated through a weighted average. The weights are assigned differently to each risk according the significance of each factor depending on the specific type of risk. The goal of the Threat Susceptibility Assessment is to quantitatively assess a system's ability to resist cyber-attack. This is analysed against a range of different TTPs using a "Threat Susceptibility Matrix". Other well-known resources developed by MITRE which can be used either in the context of the TSA assessment or on their own are the Common Attack Pattern Enumeration and Classification (CAPEC) [43] and the Adversarial Tactics, Techniques & Common Knowledge (ATT&CK) [44]. This last one is a database of different attack tactics, techniques, and procedures based on real cases, which has a version for Industrial Control Systems since 2020.

**v. CORAS**

CORAS [22] is a risk assessment approach that provides a language, a tool, and a method. The language is based on graphical symbols to construct diagrams, the tool is a graphic editor to construct the diagrams, and the method is oriented to an "asset-driven defensive risk analysis". The method is performed in collaboration between the owner of the system and an expert of the method where information about the different assets is gathered to elaborate diagrams representing different threat scenarios. Likelihood of attacks can be calculated by assigning different probability values to the associated chain of events.

**vi. CRAMM**

The CCTA Risk Analysis and Management method, CRAMM [45] developed by the Central Computer and Telecommunications Agency of the UK (CCTA) was released for the first time in 1985, and its last update was in 2003 corresponding to the fifth version. The method provides an automated tool that allows one to perform a qualitative analysis based on evaluating threats, vulnerabilities and assets. The result is a score from 1 to 7 obtained from a pre-defined risk evaluation matrix [46]. No evidence was found about the use of this method in recent years, except in a modified version developed by the NATO for their own use.

**vii. NIST Guide for Conducting Risk Assessments**

The US National Institute of Standards and Technology (NIST) in their special publication 800-30 [47] provides guidelines on how to conduct a risk assessment which is also referred in their more recently published Risk Management Framework for Information Systems and Organizations, special publication 800-37 [48]. This document focuses on the risk assessment process providing a workflow and a general description of the activities. The process is very similar to the one described in other approaches and consist of identifying threats and events, identify vulnerabilities and predisposing conditions, determine likelihood of occurrence, determine likelihood of impact, and determine risk. The approach provides several tables to support the process of estimating likelihood and impact and determine the risk through the review of different factors which are all described in qualitative terms or "semi-quantitative" which means employing rules to assess factors using pre-defined scales of measurement in order to have more consistent and repeatable results than in qualitative assessments.

**viii. IRAM2**

IRAM2 guides the user through the overall risk management lifecycle, including the provision of guidance on risk treatment. This methodology is only available for members of the Information Security Forum (ISF), but according to different sources [49] [50], it provides a tool that follows a workflow of six phases that include a qualitative analysis of threats, vulnerabilities and impacts. The threats are identified and assessed according to

their likelihood of initiation after which the vulnerability is quantified based on the control strengths. These two variables allow to calculate the likelihood after which the business impact is calculated to determine the risk.

**ix. OWASP Risk rating methodology**

The Open Web Application Security Project (OWASP) also has a Risk Assessment Framework and it suggests a risk rating methodology that follows six steps [31]. The process starts by identifying the risks, followed by quantifying the likelihood and impact factors. This is done by assigning pre-defined scores to the threat agents, vulnerabilities, and impacts according to their characteristics. Scores for overall likelihood and impact are calculated and assigned a qualitative label. The risk severity is reviewed through a risk evaluation matrix and OWASP offers a risk-rating calculator [51]. The factors for estimating likelihood are based on threat factors, which are skill level, motive, opportunity, and size, and vulnerability factors, which are ease of discovery, ease of exploit, awareness, and intrusion detection. factors for estimating impact are loss of confidentiality, loss of integrity, loss of availability, loss of accountability, financial damage, reputation damage, non-compliance, and privacy violation. additionally, there are customizing options suggested for each organisation to adapt the method according to their own context by adding factors, customizing terminology, and introducing weighting factors.

**x. Compliance based risk analysis**

It is also possible to perform a risk analysis based on compliance against a cyber-security framework or model to check the level of adoption of cyber-security good practices. A type of compliance test is using automated tools that check for compliance from a secure programming and configuration perspective, such as Nessus compliance checks [52]. Specific compliance tools for ICS devices can also be found in the market such as PLC checker, Nextnine's ICS Shield, and Verve Security Centre [53]. In this case, the risk level is assessed based on the gaps a system has respect to a defined security policy or a best practices framework. The more gaps and security vulnerabilities found, the higher the risk. This approach might overlook the context and impacts, so the results can be misleading. However, it has the advantage that the owner of the system can execute them without the help of an expert [37]. The author of this thesis believes that this sort of approach rather than replacing a risk assessment based on likelihood and impact analysis, it can be a complement of it, and can generate inputs for a risk analysis.

**xi. OTA (SANDIA)**

Sandia Laboratories proposes a matrix to rate threat levels according to different criteria, which is based on the capabilities of different threat agent types [54]. Factors used in the threat estimation are intensity, stealth, time, technical personnel, and knowledge.

**xii. Common Vulnerability Scoring System (CVSS)**

A widely used approach to estimate vulnerabilities is the Common Vulnerability Scoring System (CVSS) [30], which defines a score from 0 to 10 based on metrics that define characteristics of a vulnerability that might allow exploitation under different conditions, and the degree on which its exploitation can affect confidentiality, integrity, and availability. CVSS base scores for known software, firmware, and hardware vulnerabilities can be usually found in vulnerability databases, such as the NVD, and the MITRE CVE lists, and many times the same product vendors or project developers specify them, when they discover and disclose information about a vulnerability. CVSS scores are widely used within the cyber-security community, which means that most security professionals will be familiar with them and use them to communicate and establish a common understanding respect to the severity of a vulnerability. The metrics used to define CVSS scores are attack vector, attack complexity, privileges required, user interaction, scope, confidentiality, integrity, availability. CVSS scores define the severity of a vulnerability in a particular component; the risk needs to be assessed depending on how this component is integrated in a particular environment, so CVSS alone cannot be used to determine the level of risk. In order to assess better the vulnerability in a particular context it is also possible to adjust the CVSS base scores using additional metrics, which are the temporal and environmental metrics.

**xiii. Cyber-threat Intelligence**

Cyber-threat Intelligence consists of monitoring information from diverse sources regarding threat actors and their behaviour and development of new attack procedures and exploits. A well-known threat intelligence

platform for ICS is Dragos [53] [55]. This can be done through one or multiple platforms, which provide information that might be relevant to certain organisations.

**xiv. Pentesting**

Penetration testing consists of looking into possible vulnerabilities from the outside of the system or as a lower privilege user looking into the system from an attacker's perspective. Usually pentesting analysts, also known as, "ethical hackers" will use a set of different techniques and tools, which might or not include social engineering and searching in the web for malicious websites that abuse a brand for the purpose of committing fraud or phishing. In the case of industrial systems, it is not possible to use any potentially disruptive pentesting techniques. However, it is possible to do passive reconnaissance, such as port scanning. The results of pentesting can be a useful resource to be considered as an input for vulnerability and risk analysis.

**xv. Vulnerability scans**

Vulnerability scanning through specialised tools can also provide an important input to risk analysis. A vulnerability scanning tool can be oriented to scan vulnerabilities in networks, web applications, servers, and data bases, and will look mostly at things related to unsecure configuration and known software vulnerabilities. Some well-known vulnerability scanning software are Nessus, Qualys, Accunetix, OWASP Zap, and OpenVas. However, some of these tools might not be fit to be used in most industrial networks. Examples of specific ICS vulnerability scanning tools are ABB Cyber Security Benchmark, Tripwire CCM, Verve Security Center [53].

In the case of pentesting and vulnerability scans, it is needed to use an additional method to translate results into risk scores. In general, these results should be just one of several inputs used to calculate risks with a particular method.

#### 2.2.2.2. Comparison of methods to quantify risks and risk factors

In order to have a comparative view on methods for analysing and quantifying risks and risk factors, the methods reviewed were classified through five attributes. Some of the attributes were defined based on the security metrics classification described in [56] and a classification scheme proposed by SANDIA [37]. Table 2.1 shows the classification of the methods described on section 2.2.2.1 based on the following five attributes:

**Attribute 1: results**

It described the variable that is measured, which can be risk, threat or vulnerability.

**Attribute 2: scale of measurement**

The simplest way of classifying a metric is defining if it is qualitative or quantitative. Nevertheless, within these two classifications there are different variations. According to theory of measurement, for qualitative variables, a scale can be of the nominal or ordinal type and quantitative variables the interval or ratio type [57]. If a qualitative variable is nominal, it does not have a pre-established hierarchy, as if the case of ordinal, in which there is an intrinsic order. An example of a nominal attribute is associating a vulnerability to a CWE (Common Weaknesses and Exposures) and an example of ordinal is the "Privileges required" metric (used for calculating CVSS scores), which can be "none, "low", or "high". While ordinal qualitative variables provide an order of magnitude, they do not allow mathematical calculations, such as the mean or a standard deviation, as quantitative metrics do [57] [58]. An interval (quantitative) measurement has a well-defined unit of regular size, which allows performing linear mathematical functions, statistical operations, and quantifying increments, however, it does not allow non-linear functions and coefficients of variations, as in the case if the ratio scales. In interval scales (e.g. temperature) it is not possible to say that one value is the double or 20% higher than another value, because they do not have an objective origin (non-arbitrary zero) as a reference, as in the case of the ratio scale. In [56] two quantitative scales are defined in addition to the scales mentioned, which are "absolute" and "distribution". The absolute scale is similar to the ratio but it is applicable only in a single method. The distribution scale aims to reflect the uncertainty on a measurement by assigning a set of possible values rather than a single estimate in the form of a probability distribution. Models based on Monte Carlo techniques, are an example of using variables in a distribution scale in replacement of single point variables.

Most methods to analyse risks and risks factors will combine metrics with different scales of measurement to produce a single score. It will be part of the tasks included in the method to normalise and process this

information. It also must be noted that there is a differentiation between the scales used in the analysis, and the one used to present the results [23]. While it is possible to transform a scale of measurement to one with less level of accuracy, it is not possible the other way round. Therefore, while the analysis can be done using quantitative variables, the results can be presented in quantitative terms, as well.

**Attribute 3: provision of a tool**

Some methods have associated tools that can be manual such as checklists, and evaluation matrices or automated through a software application. It might be the case that tools are developed independently by a third party based on a publicly available method. However, in this classification only the tools provided by the developers of the method were considered.

**Attribute 4: knowledge level required**

Some methods require an expert in the method itself to be able to perform the analysis and other methods provide step by step guidelines to be followed. Also some methods require detailed knowledge about the system under review while others do not. [56] makes a distinction between three types of knowledge level required. Expert oriented methods require an external assessor who is expert in the method, owner oriented methods require knowledge about the system under analysis, and collaborative methods require both to some extent. While expert-based methods have a higher level of abstraction and are able to cover a broader scope, owner-based methods offer a higher level of granularity in the results but have a narrower scope.

**Attribute 5: type of approach**

According to [56] an approach can be temporal, functional, or comparative. A temporal approach is based on the results of a test that is done at a particular moment of time. A functional approach is based on different factors that are combined to provide a result and a comparative approach will consist in performing checks against a specific standard.

*Table 2.1: Classification of risk analysis methods*

| Method | Attribute 1 | Attribute 2 | | | | | | Attribute 3 | Attribute 4 | | | Attribute 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Qual. | | Quantitative | | | | | | | | | | |
| | Result | Nominal | Ordinal | Interval | Ratio | Absolute | Distribution | Provision of a Tool | Expert | Collaborative | Owner | Temporal | Functional | Comparative |
| **FAIR** | Risk | | x | | | | | Manual | | x | | | x | |
| **FAIR + Monte Carlo** | Risk | | | | | | x | Manual / Software | x | | | | x | |
| **Octave-Allegro** | Risk | | x | | | | | Manual | | x | | | x | |
| **MITRE (TARA-CTSA)** | Risk | | x | | | x | | Manual | | | x | | x | |
| **CORAS** | Risk | x | | | x | | | Software | | x | | | x | |
| **CRAMM** | Risk | | x | | | | | Software | | | x | | x | |
| **NIST 800-30** | Risk | x | x | | | | | Manual | | x | | | x | |
| **IRAM2** | Risk | | x | | | | | Software | | x | | | x | |
| **OWASP risk rating** | Risk | | x | | | x | | Manual | | | x | | x | |
| **Compliance based** | Risk | x | | | | | | Manual / Software | | | x | | | x |
| **OTA (SANDIA)** | Threat | x | x | | | | | Manual | | | x | | x | |
| **CVSS** | Vulnerability | | x | | | x | | Software | | | x | | x | |
| **Threat Intelligence** | Threat | x | | | | | | Software | x | | | x | | |
| **Pentesting** | Vulnerability | x | | | | | | Software | x | | | x | | |
| **Vulnerability scan** | Vulnerability | x | | | | | | Software | | | x | x | | |

Table 2.1 shows that most methods used to estimate risks levels, either directly or through providing data about risk factors such as threat or vulnerability, tend to use qualitative data and to be manual. It is undeniable that for risk analysis methods to be scalable it is necessary for them to be implemented in tools that do the processing of the input data and generate the risk scores. In this sense, quantitative methods, allow better to do mathematical operations and provide results that can have an easier interpretation. One claim against

quantitative methods is that they can give a false sense of accuracy regarding the risk estimations. However, quantitative methods do not provide any advantage in this aspect either.

It has to be accepted that estimation of cyber-risks will always be inexact because it is based in several assumptions. However, using quantitative methods systematically can allow a better communication of the risk magnitude and, also continuous improvement of the methods based on empirical data. According to Hubbard and Seiersen, the authors of "How to measure anything in cybersecurity risk" [58], for all practical decision-making it is necessary to have observations that reduce uncertainty. As much as these observations can be expressed quantitatively, it has to be accepted that in risk management, as well as in management science in general, accurate measurements will not always be available. In that sense, methods such as FAIR and TARA can be easily adapted to provide quantitative scores, but not necessarily to work in operational time since they are not based on identifying observations that can give indication of changes on the risk landscape over time. Hence, it is necessary to identify additional methods that can allow establishing a relationship between system's observations and risks scores. Examples of such methods are Markov Models and Bayesian Networks.

Looking into methods specifically focused on industrial systems, the work done by Cherdantseva et al. [27] reviews 24 different methods for cyber security risk assessment specifically applied to SCADA (Supervisory control and data acquisition systems), including academic work. A series of shortcomings and opportunities of improvement are highlighted in this paper, particularly, the conclusions suggest that the main areas of improvement for the methods reviewed were related to the following issues:

- Addressing the context establishment stage of the risk management process
- Overcoming attack- or failure orientation
- Accounting for the human factor
- The capturing and formalisation of expert opinion
- The improvement of the reliability of probabilistic data
- Evaluation and validation
- Tool support

As much as the Context Establishment is regarded by most risk management framework as one of the crucial steps for a risk assessment to be relevant and properly focused, only nine of the 24 methods reviewed addressed this processes. At the same time, most methods focus on analysing possible attacks or failure forms of the system rather than on the system's goals that need to be protected. The human factor including cyber-security knowledge and awareness of key personnel is another aspect that tends to be neglected, as well as formal means for using expert's knowledge as an input of the methods. Risk management relies at a great extent in expert judgement as a mean to reduce uncertainty. To make the most of a risk assessment method these judgements and their underlying assumptions should be captured in a systematic way in order to allow improving risk estimations and the risk analysis methods themselves. Probabilistic data used in risk assessments whether if it is obtained by expert judgement, or based on experimental test beds or honeypots results or in industry sharing information, it always have a room for improvement on the light of new data and new knowledge. Means for capturing probability data and refining probability calculations should also be included in risk assessment methods. Evaluation and validation of risk assessment methods is recognised in [27] as a challenge, since researchers will rarely have the chance to test their method in reality. To address this, it is suggested that it is still possible for evaluation and validation to be more rigorous by combining different methods, such as the ones proposed in [59]. Finally, tool support is without doubt crucial to assist any risk assessment method in practice. However, the use of tools was discussed only in seven of the 24 papers.

### 2.2.3. Common industry practices for cyber-risk management

The common practice is for each organisation to adapt known frameworks and standards to fit their own conditions and capabilities. An important limitation of most well-known cyber-risk management approaches is that most methods are designed to be done offline and independently from daily cyber-security operations. Usually they are defined as a discrete activity performed according to a pre-defined frequency. In practice, full risk assessments are often done once or twice yearly, and risk monitoring can be done more often, but still it is usually not based on collection of real-time data. Additionally, some events or circumstances can trigger a risk

assessment, such as new projects, changes in an organisation or acknowledging a significant threat or vulnerability, which was previously unknown or not taken in consideration. Moreover, in many organisations risk management is done manually, using spreadsheets and other generic tools or even no tools at all [27]. Examples of experiences collected from interaction with stakeholders in industries is that they do risk assessments manually and upload the corresponding data in shared platforms such as SharePoint. This is consistent with how the process is described in most frameworks such as the ones reviewed in sub-section 2.2.1., as well as with the research work performed in the recent years by some group of academics who support the idea of introducing more dynamic processes for cyber-risk assessment [60] [61] [62].

### 2.2.4. Continuous risk monitoring and analysis

The goal of doing a continuous risk assessment is to detect changes, which can make even a recently done risk assessment less accurate, or even off base or irrelevant. The continuous risk monitoring can allow making use of additional data gathered during operation to improve the risk estimates, as well as to track changing risk values. This is relevant because, risks assessments are used in organisation to make decisions. Hence, if the results of a risk assessment can be based in updated data it would reduce the possibility of making misinformed and biased decisions. Risk monitoring should look into indicators that can reveal that malicious actions are more likely to take place or how likely is that the system has already been compromised, e.g. by installation of a Remote Access Trojan (RAT). Overall, the idea of this part of the state-of-the-art review is to find existing methods that allow repeated re-evaluation of the risk analysis results, which is the main topic of this thesis.

The first academic works that was found proposing a real-time cyber security risk assessment based on network sensors and IDS was published by Årnes and Haslum between 2005 and 2007 [63] [64] [65]. In most recent years, other theoretical approaches have been built based on this work in which the probability of a state is based on observations [61] [66]. Refsdal and Stølen published other pioneering work proposing a method to provide a dynamic cyber-risk picture in 2009 [67], which introduces the idea of using key indicators. Examples of dynamic of methods developed specifically to be applied on industrial networks are the multi-model approach proposed by Zhang et al [68] [69], the work of Gonzalez-Granadillo et al [62], the work from Chen et al. applied to the power grid [70], and the work of Kotenko et al. [71]. Approaches reviewed which can provide potentially useful techniques and methods to be used in dynamic risk calculations, even if they were not specifically developed for IoT or industrial systems were the use of Bayesian Networks [72] [73], Hidden Markov Models [66], and fuzzy logic [71]. In addition, some extent of research proposing dynamic cyber-risk assessment models inspired by immunology was also found [74] [75] [76] [77].

The first publication of Årnes and Haslum proposes a multi-agent architecture [63] that uses Hidden Markov Models (HMM) to compute probabilities of attack. Each assessed element has a security state vector “S" in which states are probabilistic and have three qualitative values: Good, Attacked, or Compromised. They base the risk assessment on “sensors" which provide information about the security of the network assuming that these sensors would provide a standardised output. A follow-up paper was published, refining the method by adding a weighted sum of sensor's input in which these weights depend on reliability of the sensors [64]. This work was extended proposing a framework that also introduces intrusion prediction based on a combination of HMM with fuzzy logic [65]. These publications have two key points in common with the present research, which are the development of the idea of computing cyber-risk probabilities in real time and the integration of security detection systems with risk management. However, it does not provide specific guidance on how to translate the information from a sensor to the probability of certain state. Furthermore, they keep a generic definition of sensor stating that it typically refers to an IDS, but it could be any other information-gathering artefact that can provide security relevant data. It was also found that this work is widely cited in several of other references reviewed during the course of the present research.

The ideas proposed by Refsdal and Stølen in [67] are very close to the work of this thesis in the sense that defines the idea of “key indicators” to provide a dynamic overview of risks. It also introduces the idea of doing an initial risk analysis of the system in order to identify which are the relevant risks that should be monitored and establish the threat models to be used in the later steps. An example is given based on the threat modelling language developed in the CORAS method [22]. This work was performed under the assumption of the availability of an adequate infrastructure for monitoring the system and producing the required indictors based on measurable observations. As the work developed by Årnes and Haslum, this publication does not provide specific examples or guidance on the generation of these indicators and it was not developed having in mind industrial systems. However, in both cases, many similarities with the ideas developed in this thesis can be found, and also both

works can be recognised as precedent of most of the work that can be found related to real-time or dynamic cyber risk assessments.

Regarding work that has a more specific focus on ICS and industrial networks the multi-model approach for incident prediction and risk assessment [68], and the dynamic cyber-security risk assessment based on fuzzy probability and Bayesian Network [69] from Zhang et al share several points in common with this research. Examples of this are the use of both attack and anomaly evidences to compute cyber-security risks and the inclusion of data related to the industrial operations control (e.g. pressure, heating). These works provide an architecture for the methods and use cases but do not discuss implementation methodologies or the use of risk indicators. Another interesting and most recent publication the work of Liu [78], which was published in 2021, and presents ideas, which are very close to part of the present work in the sense that it also proposes the use of Bayesian Networks for quantitative security analysis on ICS. However, the authors do not make a link between their work and the field of risk management.

The work of Gonzalez-Granadillo et al. [62], also shares some high-level principles and goals with the present research and provides demonstrations based prototype over an emulated SCADA environment. The proposed framework is implemented on a tool and tested in this emulated SCADA system. The risk analysis is based on attack graphs and the orientation of this paper is more focused on technical aspects of the risk analysis and the generation of automated responses and does not discuss an implementation methodology to integrate this method in a cyber-security risk management process. It also does not consider industrial operations data or the idea of risk indicators. The publication from Chen et al [70] presenting a risk warning system based on big data applied to the power grid, among other similar references collected, offers some interesting aspects to explore regarding real-time risk assessments in industrial systems, however, it is not specifically oriented to cyber-security but more to operational and safety risks. This particular work also provides a tool which has five modules and collects data from different information systems. The work of Kotenko et al. [71] describes a step by step fuzzy method for automatic decision making according to vectors whose values could be either "normal operation", "known threat" or "unknown threat". In general, the work of this author and his colleagues has been an important reference on the inclusion of behavioural-based anomaly detection, which will be discussed, as well in Chapter 3.

Another important reference regarding continuous or real-time cyber-security risk assessment is the Wide-Impact Cyber Security Risk Framework (Wiser) which was a project developed under the Horizon 2020 European Initiative [79]. This project includes the implementation, deployment and operation of a real-time security assessment infrastructure which also includes risk assessment. The focus of this project is ICT environments as most cyber security assessment solutions, however, it includes similar concepts such as the use of metrics to be processed by a “risk assessment engine.” This project is mainly based on the monitoring infrastructure and tooling rather than in the algorithm and it appears to be promoted as a learning and exercise platform in a first instance to be promoted commercially by 2022 [80].

Continuous risk monitoring and re-assessment does not imply a change in the high-level processes of traditional risk management approaches. Moreover, it comes naturally as a logical evolution of them in response to the rapid speed of changes of the risk landscape and can be based on well-known cyber-risk analysis methods. The main purpose of a continuous risk assessment is to be able to make decisions based on updated information that can be related to the status of risks. It was observed that the works reviewed related to continuous risk management do not present fundamental contradictions with well-known standards such as ISO/IEC 27005, and most of them are based on existing risk analysis methods and approaches. However, with exception of [67], most of them do not consider a methodology for continuous risk management that includes steps for implementation of the continuous risk monitoring, analysis methods. This means that there is a gap on providing a methodological view of the implementation of continuous cyber-risk management. This is important since there is currently enough availability of security monitoring tools that can be used to monitor risks. However, tools by themselves are not enough for continuous risk management since if the risks are not put in context in relation with the system and its operational goals it is not possible to prioritise different events. For this reason, the ISO/IEC 27005 regards the context establishment as a critical process for the success of a risk assessment.

Hence, a successful continuous risk management system shall consider the integration of processes and tools, and the correct identification and contextualisation of risk-related data.

## 2.3. Discussion

In Chapter 2 several aspects of cyber-security risk management have been discussed including the identification and analysis of cyber-risk factors, and qualitative and quantitative methods for risk analysis. An overview of typical industry practices in security risk management was also given. This was followed by a review of continuous risk monitoring methods that were proposed in a number of academic publications. The main takeaway regarding continuous cyber-security risk management gathered from the state-of-the-art review is that there is a considerable amount of academic work pointing in that direction, but still the general practice in industry is to perform risk assessments as a static and manual task. While the most well-known and widely used risk management frameworks do highly recommend monitoring risks continuously, they do not give suggestions or guidelines to perform continuous risk assessments. This result on a gap between an increasing offer of security tools that claim to enable monitoring risks continuously and the capacity of an organisation to implement continuous risk monitoring in their ICS environment.

A cyber-risk management system that makes it possible to assess risks dynamically would not only create a benefit to this discipline, but it would actually make a drastic change on how the overall idea of a risk assessment is conceived. To understand why continuous risk monitoring is not a common practice in most industries the current gaps and the main challenges for its implementation were reviewed and identified as the following:

**Availability and quality of information**.
Some of the data needed to feed a continuous risk assessment system are already available in real time, provided by different sorts of monitoring and detection tools. Examples of this are logs, network traffic, malware detection, intrusion detection systems, and threat intelligence feeds for which there is an increasing offer specifically dedicated to ICS. Integrating and pre-processing this data can present technical challenges, however, new generations of Security Information and Event Management (SIEM) tools provide interfaces with a wide variety of systems and tools and have processing capabilities for big volumes of data. Nowadays, large organisations can have security tools that process millions of events per second. Hence, the main challenge actually relates to transforming this data into useful information that can give a good picture about the risk landscape. This thesis addresses this by proposing the concept of Indicator of Risk (IoR) including a method to identify and use them for continuous risk monitoring. The methodology proposed also considers the full risk management lifecycle in which variables that are to be monitored shall be put in context in relation to the risks for the specific system.

**Limited scope of academic work**
In section 2.2.4. a summary of the academic papers that describe work related to continuous risk assessments was presented. Most of these present general examples without much detail on how data for the observations of the security states are obtained. For some of the approaches the proofs of concept focus mostly on the risk analysis methods and have a narrow scope, and in most cases, no evidence was found of further development built on the initial proposal. Many of them cannot be considered as a holistic cyber-risk assessment due to their lack of consideration of the context and business impact. At the same time, there is a gap in the absence of work providing a methodology, which could guide practitioners in implementing a continuous risk assessment method. In this thesis a Continuous Risk Assessment Methodology is proposed, which has been shared with the research community through the publication of two conference papers. A third paper was written introducing the concept of IoRs and their use in Continuous Risk Monitoring, which has not yet being published at the time of submitting this thesis.

**Lack of integration between risk management and cyber-security operations**.
In order to make the best use as possible of data that can be useful for a risk assessment, information sharing between cyber-security operations, business operations, and risk management operations is crucial. This means that the same data can serve a number of different stakeholders in an organisation, each one of which needs to be provided with consistent and up-to-date role-specific views. However, as commonly there is no integration

between these areas, it becomes challenging to have an effective, efficient, and timely flow of risk-relevant information. The need for this integration becomes more evident when speaking about a risk-monitoring dynamic that works in operational time, and therefore requires near real-time data. The approach presented in this thesis proposes a continuous exchange of information between risk management, and security operations to allow linking security events with business risks.

**Cyber-security maturity level**

Implementing a continuous risk management system successfully requires a certain maturity level in an organisation's cyber-security management system. It would be expected than an organisation that adopts this approach has a well-defined cyber-security strategy and processes, standardised methods, and collects and analyses security metrics to quantitatively manage and continuously improve their cyber-security operations. Organisations that have an immature or not well-developed approach to cyber-security will not be prepared to implement a continuous risk assessment. As will be explained in more depth in the next chapter, this is usually the case for industrial networks. The work developed in this thesis assumed that an organisation has reached a certain level of maturity in ICS cyber-security operations and does not address the problems that are inherent of not having established processes that are required for a risk assessment. An example of this is having an appropriate asset management system, which would allow identifying what is needed to protect in order to identify threats, vulnerabilities, and impacts.

In the next chapter, specific aspects of industrial cyber-security will be discussed, including a state-of-the-art review, as well as approaches for security monitoring and real-time generation of risk indicators.

# 3. Cyber-security in industrial networks

Having agreed that cyber-risk can be described in terms of vulnerabilities, threats, and impacts, it is important to understand that these three factors, and therefore risks, in industrial systems have fundamental differences respect to Enterprise IT. In order to develop a continuous cyber-risk assessment method that is applicable in the context of industrial networks, it is necessary to have a good level of understanding of the characteristics and current state-of-the-art of industrial cyber-security. This would allow building a continuous risk management approach that is consistent with the constraints and requirements of industrial environments. These constraints and requirements many times imply that even with a well implemented cyber-security management system, in industrial environments certain cyber-risks need to be accepted or retained for a long period of time. Hence, the monitoring of these risks as a compensatory control becomes very relevant. Having, as well, an overview of what security monitoring tools and methods exist in the industry and have been proposed in the academic literature allows to build upon validated techniques, tools, and methods, which can be used as part of the implementation of a continuous risk assessment for industrial networks.

The first section of this chapter provides a description of the main purpose of ICS and a general view of typical ICS and IIoT architecture models. This is followed by an explanation of the differences between IT traditional and industrial cyber-security. In the second section, a state-of-the-art review is presented. This review includes the most common vulnerabilities in ICS components and devices, security weaknesses in operations management, and the main threats. The third section provides a review of monitoring and detection tools and of the generation of real time security metrics and indicators. Such indicators are an important aspect of the Continuous Risk Assessment method proposed in this thesis. The fourth section discusses the main gaps and challenges in industrial cyber-security and concludes the chapter.

## 3.1. About ICS and IIoT

Whereas the goal of IT systems is to manage information, the goal of ICS and IIoT systems is to supervise and control operational processes which most of the times require interactions with the physical world. In order to be more effective and efficient, industrial systems (ICS, IIoT) require, processing and managing information. Hence, a natural evolution of this system has been to come closer to and more integrated with IT systems. This means that, nowadays, most industrial networks are connected to IT systems. Nevertheless, OT components in an industrial network cannot be treated as standard IT systems and neither can the IT systems that are connected to them since they run software that does not correspond with standard desktop applications and can allow connecting to physical processes. For example, a computer in an industrial workstation could run software applications that allow starting or stopping industrial machines or changing their operational parameters. Hence, this type of IT system even if they are based on commodity servers and computer is not considered strictly the same as enterprise IT.

In an industrial network a diversity of OT components can be found which serve functions that are specific to a particular industrial process. Hence, it is more difficult to establish a standard architecture model that can be applied across different industrial networks. However, one of the most well-known attempts to develop a general understanding across ICS is the Purdue model, which divides the system in different zones. Figure 3.1 shows an example of an ICS architecture based on the version of the Purdue model developed by the SANS institute along with their guidance for Secure Architecture in Industrial Control Systems [81]. In this model three zones are defined which include different levels of the model. The Enterprise Zone represents the IT systems used to run the business, the Manufacturing Zone represents the IT systems used to manage the industrial operations, and the Cell/Area Zone represents a combination of IT and OT systems used to control the industrial processes.

The Enterprise Zone, which comprises levels 4 and 5 is related to business operations, is considered for most purposes out of the scope of this thesis. The Manufacturing Zone or Level 3 includes production reporting and scheduling systems, engineering workstations, network file servers, remote access and other IT services which are likely to exchange information with the lower levels. Level 3 is partially considered into scope of this work, however the main focus of this research is on Levels 0, 1, and 2, which corresponds to the Cell/Area Zone. Level 0 is where the sensors and actuators that interact directly with the processes are connected, level 1 corresponds to the control devices such as PLCs and other types of industrial controllers, and Level 2 to on-premises

operator's interfaces, such as HMI (Human-Machine Interfaces) and local workstations. Other examples of OT components that can be found in an ICS are field couplers or gateways that serve as interfaces with devices that use specific OT protocols with IT systems and devices that use other standard IP communication protocols.

The Safety Zone is where safety assurance systems are implemented such as redundant process controls or fail-safe mechanisms. These systems are usually managed independently from the main ICS and are built according to the requirements set by the safety standards and regulations of each specific system. As this is a totally separate area of knowledge, the Safety Zone will be not studied in depth during this research, however it is acknowledged as relevant part of the system. Data extracted from the Safety Zone has also the potential to contribute to understand and monitor risks.

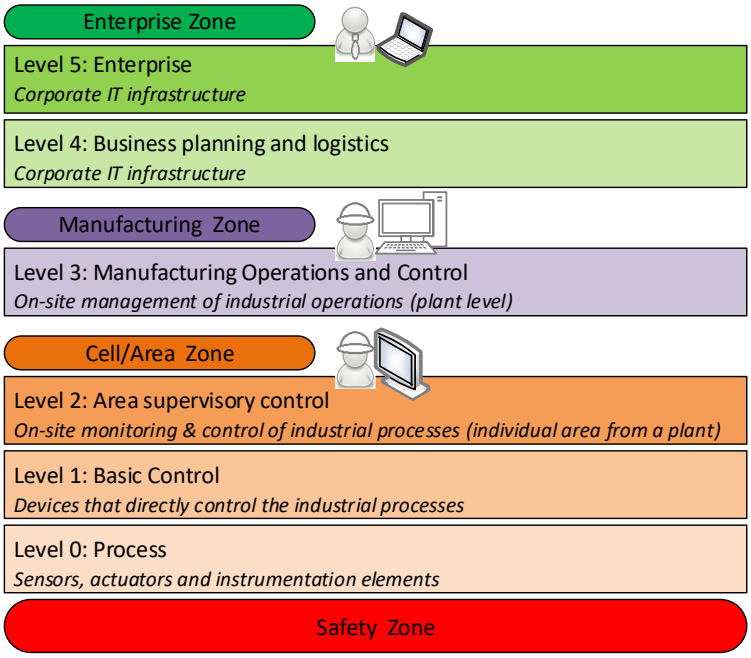



*Figure 3.1: Purdue Model*

The Industrial Internet Consortium, also proposes an architecture pattern for the implementation of smart industrial systems (IIoT) which comprehends three tiers [82]. Figure 3.2 shows a simple version of this model, extracted from [82], which provides a high level perspective of the interaction between devices that are based on one or multiple sensors and actuators or have control capabilities, which are in the most recent times also regarded as "things" and standard IT enterprise systems. The "edge tier" collects data from the edge nodes, which correspond to the field devices. This also includes the "proximity network" which field devices use to communicate between them, and the network topology and communication protocols used will depend on the specific use cases. Therefore, in many cases a gateway will be needed to communicate the edge tier with the platform tier, which is done through the "access network". The "platform tier" receives, processes and, forwards commands from enterprise to edge tier, consolidates, processes and analyses data from other tiers and it provides management and other functions like queries and analytics. The platform tier communicates with the "enterprise tier" through the "service network". In this last tier domain-specific software runs, including decision support systems and interfaces to end-users where control commands can be sent to the platform tier and subsequently to the edge tier.

It is an undeniable fact that the risk of industrial operations suffering cyber-attacks has become higher in the past few years [8] [83] [84], and security specialists believe that this trend will continue [85] [86]. However, cyber-attacks on cyber-physical systems are not new. The first known cyber-incident suffered by an industrial organisation was in 1982, when a Trojan Horse caused an explosion in a pipeline in Siberia. More details of this case can be found in the Repository of Industrial Cyber Security Incidents (RISI) [87], which provides information about 242 attacks on Industrial Systems from 1982 to 2015. This includes well-known cases such as the Maroochy sewage water spill in Australia in the year 2000 that was caused by a contractor with privileged access

to the SCADA system, and the Stuxnet case in the year 2010, in which a worm caused major damage to a nuclear plant in Iran.

Stuxnet marked an inflexion point in Industrial Systems cyber-security since it was the first malware known to specifically target ICS and to be programed to interact with industrial devices. In this case the target devices were Siemens S7 PLCs. This led Siemens to be one of the first ICS vendors to have a cybersecurity incident response team specifically dedicated for products or "Product CERT". A relevant source of information about attacks on ICS is the MITRE ICS ATT&CK knowledge base [88], which is currently maintained and kept up-to-date by members of the cyber-security community. Through it, it is possible to observe that the past few years several adversary TTPs specifically targeting industrial environments have been developed. This also allows forecasting an increase on demand for cyber-security solutions specifically developed for ICS [89].

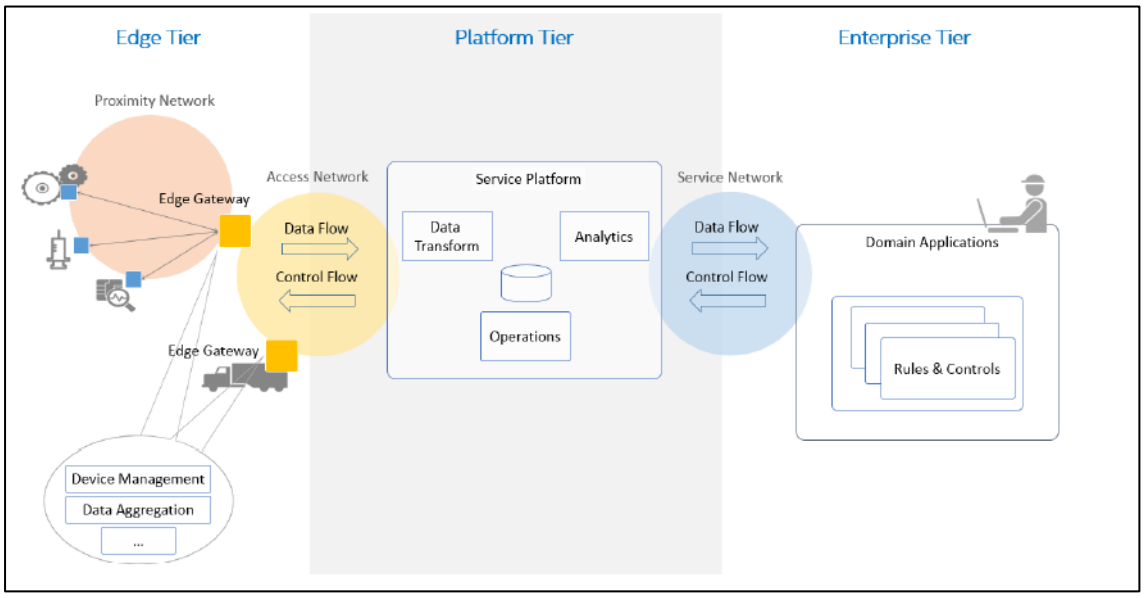



*Figure 3.2: Industrial Internet Reference Architecture [89].*

While the world of OT has made use of devices with computational and communication capabilities for decades, this field has developed in a separate track from office and home computers. Over the years, OT has become more highly connected, making industrial networks more efficient, but also more vulnerable to cyber-attacks. This is true for both traditional ICS and also for new generation technologies that are often labelled as IoT or IIoT. While there might be an expectation for cyber-security to become an integral part of the Industry 4.0 paradigm [90], it still does not seem to be a priority in many industries. Newer technologies are introducing security by design, however, most of the systems that can be currently found in industrial environments possess several security vulnerabilities, which are expensive and not straightforward to fix or mitigate.

### 3.1.1. Main differences between IT and OT cyber-security

During this research it was found that the approach to cyber-security that needs to be taken in an industrial network differs from traditional cyber-security practices for reasons that are best explained from historical, environmental, functional, and technical perspectives.

**Historical perspective**

OT has historically been developed by specialized vendors from the world of industrial control and automation, independently from the development of IT systems. The field of cyber-security initially was mostly focused on information and network security for ICT systems and did not considered other programmable devices, which have their particular languages and logic. In OT the main functionalities are collecting and monitoring data from sensors and sending commands to actuators based on control rules that are derived from operational

requirements. For example, if a sensor indicates that fluid in a container has a pressure level above certain threshold, a command can be send to open a valve to release the pressure. In many cases, industrial programmable controllers such as PLCs are designed to execute a limited amount of functions giving priority to the capability of processing data and executing commands in real time and ensuring reliability rather than allowing diverse programs and functionalities such as in the case of computers.

The same as in older computers, in many cases, industrial systems were designed assuming inherent trust between connected devices, i.e. the fact that two devices were physically connected meant that the data exchanged between them was trusted and no authentication or authorisation was required. Industrial communication protocols, were developed with focus on robustness and availability, rather than security, and many of them, which are still being used in the present, do not support authentication and encryption. Connection to IT systems used for monitoring and control purposes, was expected to be through private networks and within industry premises with strictly limited access. Therefore, it was assumed that whomever could access these systems was an authorised person.

Thus, security relied heavily on isolation and physical accesses controls. Most industrial networks were physically isolated from untrusted networks and in particular, not connected to the internet. This led to overconfidence due to what is called "the air-gap myth", which is the belief that network isolation is sufficient as a single line of defence. Cases such as Stuxnet clearly show that even so called air gaped systems can be breached if an adversary is determined to do so. It must also be noted that currently, industrial environments are increasingly been connected to networks that allow remote access, which becomes practical for remote monitoring of processes. However, these industrial networks are usually set-up and administered by control engineers who have not been trained in cyber-security. This means that there is a general lack of cyber-security skills and experience in the industrial field and at the same time cyber-security professionals will generally not know how to implement cyber-security standards in industrial environments [90].

The prevalence of proprietary communication protocols in ICS also gave a sense of "security by obscurity" since knowledge of how industrial devices work and information about what different command messages mean was not as accessible as it is nowadays. The legacy systems issue presents also important constraints regarding the implementation of security controls. In industrial systems this is not only restricted to software but mostly to hardware which is not easy to replace, because the complexity of their implementation and the high cost. Many legacy industrial machines and devices constitute expensive assets that were planned as an investment for several decades as opposed to personal computers, which are replaced to be upgraded every couple of years. Legacy components found in industrial environments include web servers, operating systems, content management systems, and other applications, which often have critical vulnerabilities, many of which have exploits that are easily made available in the internet [16].

**Functional perspective**

Industrial networks have "cyber-physical" functionalities, which allows them to interact with the "real world". This means that the system will receive information from, and make changes to the physical environment, which may be in the ICS' immediate surroundings, such as in a manufacturing cell in a factory, or as in the case of utility services spread over a wide area such as a whole city. The main functions of an ICS are not focused solely on the information processing as in IT systems, but on monitoring physical conditions, making automated decisions, and enabling interactions with physical processes. Therefore, processing information is subject to time constraints, and its value depends on its timeliness as well as its correctness. In the extreme case of real-time operation, processes are required to complete within a set time. In ICS, the main assets to protect, in most cases will not be data, but processes, business continuity, and the integrity of physical assets, and products. This means that availability and integrity will be more often the main cyber-security objectives and that confidentiality would be mostly related to the protection of intellectual property and industry secrets, rather than to financial data or customers' and collaborators'' personal information. This means that from the attacker's point of view, one of the main goals can usually be industrial sabotage rather than information theft. However, depending on the context and the type of data gathered by the system, this goal cannot be excluded since data that is valuable for a business can potentially be as well valuable for an attacker.

**Environmental perspective**

Industrial networks and IIoT systems can include a large number of connected devices, which can be distributed over a wide area, which makes it hard to have visibility of the whole ecosystem. Such is the case for applications in agriculture, electricity and water distribution, and smart cities. In such domains, equipment is often located in unmanned outdoor sites with little to prevent access by a determined attacker. Even in areas enclosed within a fence or other security perimeter it is not unusual to have physical attacks or disturbances caused by trespassers or even by drones. In manufacturing and other industrial plants, where access is usually restricted, there are still a large number of employees and contractors who have access. As industrial operations are complex and involve a large number of sub-systems, it is usual for suppliers and contractors to be responsible for some parts of the system and to have remote or unsupervised physical access. Equipment looked after by different suppliers may be co-located so that providing necessary access to a supplier's own equipment, will also grant access to others'. Lack of, or poorly implemented, privilege separation could allow access to higher privilege or out of bounds functions, since many processes are inter-related. This can allow malicious or unintended interference with normal operational conditions.

**Technical perspective**

Industrial systems are often designed to perform in real-time, i.e. there are strict constraints on the time that they take to respond to events. In many cases, the maximum delay between the reception of the input signal and generation of the output by a controller is in the order of the milliseconds. For example, nuclear power electric generation controls specifications go from 1 to 2 milliseconds and control of electrical power generation and transmission from 4 to 16 milliseconds [91]. Even when not strictly real time, industrial systems are generally time bound, such that results that are delivered late are of no use, or at least have reduced value. This means that authentication and encryption algorithms need to have low latency. Furthermore, their implementations are constrained by limited memory and processing capabilities. Other lines of cyber-security defence, such as malware scanning, firewalls, network scan, and IPS systems should also be designed and implemented in such way that they do not violate ICS operational and technical requirements and specifications.

Hand in hand with low latency requirements, many systems should also guarantee high availability. Many industrial systems are expected to work 24x7 with minimal interruptions. For example, a Tier 1 data centre is expected to have a maximum annual downtime of 28.8 hours (99.671% availability), which reduces to 26.3 minutes (99.995% availability) for Tier 4 data centres [92]. This means that the whole physical and ICT infrastructures that support the data centre, including the electrical supply and environmental control, which are managed by industrial and building management systems, should be technically capable of fulfilling this requirement. Critical infrastructure such as electrical and water supplies, communications, and healthcare are other examples that require 24x7 availability. This requirement covers the physical infrastructure, as well as OT, and ICT operations. This limits opportunities for OS and software updates and patching, which would interrupt service availability. This is one of the reasons why it is common to encounter outdated and vulnerable windows versions running on industrial workstations. Another is the risk of the monitoring and control applications having unpredictable reactions or even stopping working after an OS update. There is often no way to test the updates, since having a dedicated installation for this purpose is not realistic. For this reason, it is often that the moto "if it is working do not touch it" is used. This also refers to the fact that many older controllers run programs that have not have been properly documented and nobody still working in the company knows how they work. Due to the specific characteristics of each industrial application, controllers are often programmed in-house and software development processes are not always standardized or follow secure programing best practices.

Due to the contextual differences and special requirements compared to ICT cyber-security, industrial cyber security, can be considered, to a certain extent, a separate variant or sub-discipline of ICT cyber-security. The fact of having different requirements, functionalities, environment, technologies, actors, and tooling, makes it necessary also to have different approaches. At the same time that many security vendors, such as Symantec and McAfee, have expanded to create products directed at the ICS market, new actors have emerged specialising in ICS cyber-security, including Cyber-X, Indegy, and Dragos, just to name some examples [53]. Despite the increase in available technologies, adoption is slow due to the numerous challenges described above, plus the lack of sufficient regulations and standards. However, due to the evident need to improve cyber-security management in ICS, it can only be expected that the ICS cyber-security market will continue to grow in the future years [93].

## 3.2. State-of-the-art review of cyber-security in industrial networks

In section 2.2. a state-of-the-art review of cyber-security risk management was provided in order to explore the most well-known risk assessment methods and practices and work related to continuous cyber-risk assessments. In this section, the state-of-the-art of cyber-security in industrial networks including ICS and IIoT will be reviewed and, at some extent IIoT, since, as explained in Chapter 1, these systems have significant common ground. The state-of-the-art review in this chapter includes an overview of industrial cyber-security covering design, implementation, and operations. As security operations is the part of the lifecycle where continuous risk assessment applies, the main focus will be in this area, which includes cyber-security operations management, intrusion and anomaly detection methods, and use of real time indicators for risk monitoring.

### 3.2.1. Vulnerabilities in industrial components

As explained earlier, neither field devices nor industrial network protocols and standards were conceived to be compliant with what nowadays would be considered as cyber-security best practices. Evolution to more secure technologies has been and will continue to be slow because of a variety of challenges involved in replacing existing devices and modernising industrial operations that are already running. In addition, lack of experience of vendors and professionals in the field, plus the difficulties to make changes to systems that are already working in a continuous mode operation, and whose interruption can cause major costs and disruptions. For example, temperature could affect the quality of materials stored for use in a downstream process, meaning that interfering with the temperature control in a storage facility could spoil a whole production batch. Other industrial applications can be even more sensitive to environmental changes since they manage hazardous processes involving heavy machinery or dangerous goods, such as in the chemical or nuclear industries. Transport and navigation control systems are also vulnerable to incidents with high cost impacts including injury or loss of life.

In order to understand the current state of security in industrial devices, it is instructive to look into different ICS vendors and the security advisories that they publish. A security advisory comes as a result of the discovery of a vulnerability in a product that is already in the market. In the case of ICS some of these products can even be legacy systems that have been operating in industries for decades. Once a vulnerability is acknowledged by the manufacturer, they should analyse it and provide relevant information to customers including, ideally, a solution such as a security patch, or at least mitigation measures and guidance for secure operation. A good example of vulnerability management in industrial control devices is the German company SIEMENS, who created their ProductCERT (Product Computer Emergency Response Team) as a results of the Stuxnet incident in 2010, and who have published security advisories regularly since 2011 [94]. In general, older generation devices tend to have many security flaws, because they were designed to operate in closed and trusted networks. For newer models and versions of a device, firmware updates are usually made available when a vulnerability is found. However, it is still a challenge for industrial organisations to adapt their processes in order to allow these updates to be done in such a way that disruption to operations is minimised.

The following are some of the most common vulnerabilities that can be found in Industrial Control Systems devices [16] [95] [96] [97] [84]:

**Improper input validation**
Many ICS devices do not have define responses for unexpected inputs. Because of this, attackers can perpetrate DoS attacks by sending them random signals to devices, causing them to crash or restart.

**Out of bounds read and write**
This vulnerability is caused by improper restriction of read and write operations to within the bounds of a memory buffer. This can allow an unauthorised party to read or write commands on memory locations in which a user will not normally have access, which can lead to different types of attacks. This can allow an attacker to inject input data that can cause a random behaviour which can make the device crash or even use specially crafted code to change the system's behaviour at will.

**Plain text communication**
Some of the most widely-used protocols in industrial networks lack encryption. This can include, authentication data, allowing to harvest credentials. An example of a widely-used industrial protocol that transmit unencrypted commands are Modbus, Profibus, Profinet, and Bacnet. In the case of Bacnet, encryption is supported, but only by newer generation devices.

**Improper authentication**
In the most extreme cases, devices lack authentication mechanisms. Others have flaws such as allowing weak passwords, storing credentials in plain text, using hard-coded credentials, or lacking mechanisms to avoid a re-play attack to login.

**Improper privilege separation and access control**
Many industrial devices do not allow definition of separate profiles with different privileges.

**Firmware not digitally signed**
It is common for industrial control devices to allow the injection of firmware code from an unknown source, since firmware is often not digitally signed.

**Supply chain vulnerability management**
Many industrial systems manufacturers rely on third party hardware and software components whose security might not be properly assessed, examples of this are processors and operation systems. For example, in 2019 Wind River released a Security Advisory known as "Urgent 11", in which eleven vulnerabilities from their VXWorks RTOS (Real-time Operation System) were disclosed [98]. This product is used in several industrial and IoT devices, including medical equipment, and several of the vulnerabilities were rated with a high or critical CVSS score, which meant that manufacturers which use VxWorks in their products had to also inform their customers and release the corresponding solutions or recommendations for security risks mitigations. In the case of industrial controllers, most of the main brands such as SIEMENS, Rockwell-Automation, Schneider Electric, and ABB were affected. A similar situation happens often with microcontroller manufacturers that provide hardware used in embedded systems which are designed to be integrated in other products.

**Secure operation modes are not set by default**
Often, control devices and applications have security features that can be enabled or configured. However, they are not enabled by default as most cyber-security best practices would recommend. This leads to many industrial organisations having an insecure implementation because lack of knowledge about the importance of changing the default configurations for more secure ones.

### 3.2.2. Vulnerabilities in operations and management practices

As in any system, in ICS, adversaries do not only take advantage of vulnerabilities in individual hardware, firmware, and software items. Weaknesses are also introduced into a system by connecting devices to an insecure network, having insecure configurations, giving unnecessary privileges to processes and users, lack of adequate defences, and loose security policies. Overall, ensuring appropriate security at all levels of an industrial system depends on establishing an effective cyber-security management system. This should cover all aspects such as purchase of equipment, implementation, operation, and maintenance, as well as decommission of devices, and, of course, risk management. Often industrial networks fall partially or completely out of the scope of the cyber-security operations management and at the same time administrators of these systems do not have adequate knowledge and experience to deal with cyber-security issues. This makes industrial organisations highly vulnerable to the most basic cyber-attacks and totally unprepared to react if they have to face advanced threats.

The following are some of the typical security problems that can be found in ICS infrastructures due to the lack of an effective cyber-security management system:

**Vulnerabilities in web servers and workstations**
Most industrial systems have Windows OS running on their web servers and workstations and usually use outdated versions, exposing the system to well-known vulnerabilities [8]. Many of the applications running on

these machines also have vulnerabilities, all of which makes these machines a popular attack vector for ICS. Examples of types of attack that web servers may be vulnerable to are Cross-Site Scripting (XSS), Cross-Site Request Forgery, Path Transversal, SQL injection, and OS Command injection, which can allow remote command execution and arbitrary file uploads. According to research published by Positive Technologies in 2018 [16], in a sample of 11 companies, 43% of web applications that were used within their industrial environments presented vulnerabilities that could potentially open a door for an attacker to penetrate the industrial control network.

Windows-based ICS workstations will often use the OPC protocol (Object Linking and Embedding for Process Control), which is a technology developed by Microsoft to gather information from, and communicate with ICS. OPC is commonly used because it is vendor neutral, and it is considered as a "universal translator" between IT and ICS [99]. However, this protocol can only work securely within a proper network segmentation [100], since it is not only considered insecure, but also "firewall unfriendly" [99]. OPC can serve as a bridge for an adversary to jump from an IT system running on a workstation (Level 2 of the Purdue model) to the actual controllers and field devices (Level 1). Actually, several TTPs specially crafted for ICS make use of the OPC protocol to compromise ICS [88].

**Use of remote access points**

Remote access might be necessary for monitoring and to allow administrators to enable remote maintenance of the system, but it can also enable attackers to access the system. An example of this is an attacker who in February 2021 managed to access the ICS of a water treatment plant in Florida through a remote desktop application and increased the amount of sodium hydroxide in the water [101]. In a report published by Cyber-X in 2019, analysis of network traffic data from 1,821 production IoT/ICS networks showed that 54% of the networks had devices that could be remotely accessed using standard protocols [8].

**Poor network security**

Industrial networks do not always have appropriate network segmentation, which means that they can be accessed through enterprise information systems, increasing their exposure. This could be for legitimate reasons such as to allow monitoring and data exchange for specific purposes and by privileged roles and processes. However, bad network segmentation can allow also unauthorised entities to access ICS. Research by Positive Technologies, revealed that in 73% of the cases tested, it would have been possible for an external attacker to penetrate the network perimeter through accessing the corporate information system, and in 82% of cases to access the industrial network from the corporate network [16]. Direct internet connections are also an important security risk in industrial networks, which according to best practice should be avoided. Cyber-X found that in 2019 27% of organisations had direct internet connections as opposed to 40% in 2018 [8]. However, the industries for each one of the samples in which the reports were based, were different in each year, so the evidence of an improvement is not conclusive. In many cases, flaws such as improper network segmentation or traffic filtering can be introduced by system administrators inadvertently while providing means for remote administration [16].

**Insecure credentials management**

Basic flaws such as the use of dictionary passwords and storage or transmission of unencrypted credentials can be also commonly found in industrial systems [8] [16]. This allows attackers to obtain credentials through password guessing, or in the case of lack of encryption through Man in the Middle attacks or other credential harvesting techniques. According to Positive Technologies, almost every company that was part of their research used dictionary passwords for webserver administration systems or for remote access [16]. Password sharing is another common malpractice in ICS environments. According to a control engineer interviewed during the course of this research, it is not uncommon a workstation has a single log in account and credentials are shared by all users [102]. This not only reduces the chances of attribution in case of an insider attack but also does not allow implementing separation of privileges.

**Lack of effective anti-malware scanning**

Often, servers, workstations, and other computers and devices are not being scanned consistently for malware. Frequently, anti-malware software is disabled or outdated, and no automated system updates are performed [8]. It can be also found that procedures for scanning devices such as USB drives and laptops for malware before connecting them to the network do not exist or are not been followed.

**Insecure system configuration**
According to Positive Technologies, most security flaws on the network perimeter are caused by misconfiguration [16]. This included publicly available interfaces, weak authentication and excessive privileges inappropriate permission levels for users, which combined allow access to restricted functions such as executing admin commands and launching attacks against critical assets.

**Industrial Systems' components are out of the scope of the cyber-security operations**
Even in cases in which corporate networks in an organisation have a "good enough" information security management system in place, exceptions are often made for industrial networks, or they might even be considered out of scope. As security engineers tend not to have knowledge about OT, and control engineers tend to overlook cyber-security, industrial networks security sometimes ends up in "no man's land". It is also often that certain components of the system are neglected, for example web applications or products or services controlled or owned by third parties [16].

### 3.2.3. Main threats in industrial networks

Despite the poor security configuration and immature approach to cyber-security management frequently exhibited by industrial networks, reports of cyber-attacks on ICS and IIoT are not as frequent as might be expected. One possible explanation is that many cases might not be disclosed to the public. Another is that that many industrial operations have indeed suffered breaches, but, due to the lack of security monitoring and situational awareness do not realise it. Cyber-X reported that 22% of the networks they analysed exhibited indications of attack, such as scan traffic, malicious DNS queries, abnormal HTTP headers, excessive numbers of connections between devices, and known malware [8]. According to Positive Technologies, some of the vulnerabilities they found in their research allow attacks that are easy to implement [16], which means that in their current situation industrial networks are have a high risk of suffering a cyber-incident caused by opportunistic attackers.

Targeted attacks appear to be less common, since compared to attacks on information systems, attacks to ICS are harder to monetise [103] and require more effort and resources from the attacker's side. For this reason, the most sophisticated attack procedures known so far have been attributed to threat agents with enough motivation and resources such as nation-state sponsored actors. However, since the discovery of Stuxnet, malicious code used for industrial sabotage has become available, and based on current trends, Kaspersky believes that ICS will become increasingly a target in the near future [97].

Most ICS threat intelligence is based on research on malware found on the wild and ex-post analyses from real-life past attacks, such as the information found in the MITRE ICS ATT&CK knowledge base [88]. Possible threats can also be inferred by looking into the possible means of exploitation of known vulnerabilities, for example, vulnerabilities related to improper authentication can be related to unauthorised access threats. Another approach to understanding ICS threats is to build an ICS honeypot. In 2020 the security company Trend Micro published a report with the results of running their so called "Smart Factory Honeypot", which was a setup that included actual ICS devices and workstations [104]. This report provides an overview of real threats that can affect industrial systems if they do "everything wrong", as they did with their honeypot in order to attract the attention of opportunistic attackers. In other words, they set up a system with as many vulnerabilities, as possible. Based on the information gathered during this research and in the judgement of the authors of this report, that sort of insecure setting might not be very unrealistic. According to their timeline, after three months of operation, attacks started to become more frequent. These included reconnaissance activities, use of the system for malicious cryptocurrency mining and fraud, robotic station shutdown, logging off legitimate users, attempts to download files from the system, ransomware attacks, and sending of control commands, among others.

### 3.2.4. Intrusion and anomaly detection methods in industrial networks

During a system's operation, several conditions can indicate the likelihood of a cyber-security incident, either malicious or unintended. Part of the challenge of the methodology that will be proposed in this thesis is to identify those conditions and put them in the context of continuous monitoring of cyber-security risk. Identifying

which are the observable variables that can provide significant evidence of risky conditions and, can be effectively monitored is a key aspect of the present approach. For this reason, different methods for intrusion and anomaly detection were explored as part of this research, in particular those with a focus on cyber-physical systems, including IoT, IIoT, and ICS. It must be noted, that conditions that can be perfectly normal in an enterprise IT system, could be considered as an anomaly in industrial environments. For example, in an enterprise context it generally is considered normal that new devices are connected to the IT network, whereas in OT environments this could be considered an unusual event. Another characteristic property of ICS is that devices only need to communicate with each other for a limited number of specific purposes. Having a clear map of which devices are allowed to communicate with each other and how the traffic between them should look like in terms of frequency, volume, and protocol used, should allow abnormal traffic to be detected [12].

As part of the different anomaly detection approaches reviewed, use of data from sensors to detect abnormal behaviours that could be attributed to an ICS targeted attack was explored. This also included using correlation checks to infer anomalies in one variable from the detection of an abnormal behaviour in another variable. This type of technique can allow anomalies to be found in the process even if other variables from the process are appear to behave normally. The goal of this is to allow one to detect that something is wrong even if an attacker spoofs inputs or outputs of the control system, for example, by replaying “normal” data. In this case, other variables could expose indications of an attack and trigger an alert. For example, environmental variables such as temperature, humidity and pressure should be expected to be correlated in a closed system. In a controlled environment, usually the higher the difference between the internal and the external environmental conditions, the more energy is needed to maintain the system’s optimal conditions. Therefore, energy consumption can also be a variable that indirectly reveals an anomaly.

The National Institute of Standards and Technologies Interagency Report, NIST IR 8219 on Behavioural based Anomaly Detection (BAD) for Industrial Control Systems [12] demonstrates how detecting anomalous conditions can improve the security and reliability of ICS. The report describes how tools from four different vendors can identify different sorts of anomalies in two test environments: a collaborative robotics-based manufacturing system and a process control system (PCS) based on commonly used ICS in chemical industries. The scope of the report is limited to BAD capabilities, which are also mapped to the NIST Cyber-security Framework. The report highlights that these sorts of capability, in contrast to traditional or signature based detection tools, look for evidence of compromise rather than for traces of the attack itself, allowing unknown threats (zero days) to be detected as well as know ones.

NIST IR 8219 defines three different types of BAD: network based, agent based, and historian and sensor based, which are all non-intrusive ways of monitoring an ICS. Network based BAD requires aggregation of all traffic in a single collection point where passive monitoring is performed whereby the traffic is compared with a "normal behaviour" baseline. The agent based BAD is also passive (non-intrusive), but the detection is distributed across "agents", which are installed in, or close to, end-point devices. The agents can, in addition, collect data at a device level such as the use of removable media. Finally, historian and sensor based BAD analyse the behaviour of the ICS by monitoring of sensor data. The different use cases for anomaly detection presented in NIST IR 8219 address particularities of ICS which typical IT network IDS, and IPS will normally do not detect.

Because communications patterns for monitoring and control can have distinctive profiles, such that traffic volume, protocols, and device interactions can be predicted for a given state of the system, detection use cases based on expected data traffic could be implemented successfully in [12]. Despite the fact that the report also highlights the value of using data from sensors, only one of the tools was used for this type of demonstration. Most of the anomaly use cases described are network based and do not consider data from sensors and actuators. As one of the main pain-points of industrial security is the potential compromise of the reliability of operations, it is perhaps surprising that data from sensors is not taken into consideration more for security monitoring, especially since this data can be directly linked to the main business objectives of an ICS, which is to have under control an industrial process. The OSIsoft tool was the only one that appeared to be capable of processing data from sensors, or at least the only one that demonstrated these capabilities in the proof of concept described in the report. However, all of the anomalies were based on binary variables, such as absence or presence of data or if a device was on or off or on observation of values above thresholds or outside operational ranges. More advanced approaches such as detection techniques based on correlation between

variables were not considered. Some SIEM tools, such as Splunk can directly ingest sensor data allowing actions to be triggered according to rules based on the behaviour of the data. Another interesting publication related to behavioural anomaly detection in ICS is the work done by Vargas [105], which describes a proof of concept of an Industrial Intrusion Prevention System (IIPS) based on the definition of security policies.

Some interesting examples of academic research related to anomaly detection for industrial or cyber-physical systems in general based on sensor data were found in the literature. Particularly, the work of Desnitsky and Kotenko proposes that a misbehaviour in one variable can be inferred by correlating events and anomalies from other variables [7] [71] [106]. Despite this research, which also proposes a risk-based approach, has over four years since it was published (2015-2016), it was observed that these ideas are not well-known. While physical based anomaly detection appears to be an incipient practice, which is starting to be considered and recommended [12] [107], correlating data from different sensors has not been found to be promoted by industry related sources but appears more often in academic literature. Another author that proposed correlation of sensors data for anomaly detection for cyber-security, in this case applied to electrical substations was Chee-Wooi Ten in 2011 [108]. In this work, possible events are ranked based on their potential impact, which also suggests an alignment with the concept of risk management. Also in 2011, Linda et al. published their work related to anomaly detection in industrial networks based on fuzzy logic. Part of this work relates to applying fuzzy logic rules to network parameters such as the average interval between packets, data length, number of protocols, and number of flag codes [109] [110]. Other work from the same authors is based on detecting anomalies in data from sensors by analysing in every PLC different signals and comparing them to thresholds, normal parameters of behaviour and with data forecasts [111]. The work published by Sicard in 2018, bases anomaly detection for ICS on defining and detecting invalid or “prohibited" states of the system [112]. The work published by Kim et al, in 2020 proposes a real-time detection algorithm, for attacks to sensors in Cyber-Physical Systems (CPS), which is also based on anomaly detection [113].

Other work proposing different approaches for IDS and anomaly detection for industrial systems provides insight regarding information that can be gathered from operational systems to allow a risk to be identified, and how it can be collected and processed. Examples of these are IDS which are distributed across several devices or “agents” [114], and IDS based on hybrid methods that compare a model of the current state of the system derived from real-time data with specifications for normal operation [115] [116] [117] [118] or detect discrepancies between estimated and measured data [119]. These methods use time-series data from physical processes to analyse whether a state is normal. In general, they require a good level of knowledge of the system, as opposed to some of the other methods previously mentioned, which also include AI techniques, such as neural networks [111]. The majority of the papers reviewed offered a proof of concept based on simulation, and a smaller proportion of them exhibited results from experimentation on a physical test-bed or a real system. Other papers offered only a theoretical overview of their approach and no proof of concept was provided.

After conducting a state-of-the-art research, it was concluded that the potential of physical-based detection has been identified by several authors, although is not yet widely implemented in the industry. An important proportion of the studies under review present multiple algorithms including fuzzy logic, data mining, knowledge based approaches, and statistical, and machine learning techniques. Most of these methods have not been properly tested in a real industrial environment; however, they provide interesting ideas for future development.

### 3.2.5. Use of real time indicators for risk monitoring

An indicator is a measurement, which provides an estimate or evaluation [120]. Some of the work reviewed in previous sections of this thesis proposes methods to generate dynamic security metrics, including risk metrics. While the concept of “Indicator of Risk” (IoR), which will be used in the present work is not yet widely used explicitly in these works, the term “Indicator of Compromise” (IoCs) is widely used in the cyber-security field. IoCs are observations of artefacts, which can be attributed to a particular attacker’s TTPs, tooling or infrastructure and that are evidence of a security breach. Typical examples of IoCs in an enterprise IT network are IP addresses, domain names, TLS Server Name Indicator values and certificate information associated with known malicious entities, and malware signatures [121]. As in ICS not all threats can be identified at an IT network or host level, additional IoCs need to be considered, which will tend to be more complex [122]. IoRs, as

opposed to IoCs are not necessarily deterministic (i.e. crisp) variables, but will often have a measure of confidence or probability associated with them. Depending on their degree of confidence they might need to be complemented by additional evidence in order to be conclusive or more accurate. Hence, an IoR does not necessarily reveals an actual compromise but provides information regarding the level of exposure to a given operational risk [123], or in our case to a cyber-risk, which can evolve with time.

IoRs can be related to insecure configuration such as open ports or a remote connection, to unusual conditions such as higher network traffic or the use of certain commands over the network, the presence of unknown files or programs, unusual log entry, or misbehaviour of sensor or actuator data. There is an overlap between IoRs and IoCs since when an IoR can allow a threat to be with inferred with a high level of confidence, then it can be considered an IoC. In addition, the combination of several IoRs can also increase the level of certainty about malicious activity taking place. Concisely, while an IoC should be able to answer the question "are we compromised?" [124] and IoR should answer the question "what is the likelihood that we can be compromised?". If the latter question is applied to the present moment or the immediate future, and the answer is that the likelihood is very high, then it is reasonable also to infer a compromise will occur. It must be noted that this means that the IoRs are related to two separate issues here. The first is about whether evidence for or against whether a breach has occurred is sufficient to draw a conclusion and the second about hoe this evidence reveals exposure to operational risk which concerns the possibility of this happening in the future.

One of the first publications found that suggests the idea of IoRs was the work of Refsdal and Stølen in 2009 [67], already mentioned in Chapter 2 in the context of work related to risk management. In this work it is proposed employing key indicators to provide a dynamic risk picture in computer systems. Although the exact term "IoR" is not used, as mentioned earlier, there are many similarities between this work and the ideas developed in this thesis. This work was declared to be done under the assumption of the availability of a monitoring infrastructure, which might have not been always the case at the time. However, the current availability of tools for monitoring and processing indicators makes it more feasible to apply these ideas in practice than at the time this paper was published. Particularly in the case of ICT, there are more tools for monitoring and detection available and currently there is an increasing offer for ICS, as well. Furthermore, SIEM tools provide capabilities to process and correlate security information and generate different sorts of indicators.

With the exception of the work of Refsdal and Stølen [67], also highlighted in Chapter 2 (which is not specific for industrial networks) most of the work on monitoring risk indicators lack a holistic overview of the full risk management process. There is also very limited work regarding indicators covering all the levels of the Purdue model since most work on OT security monitoring mostly focus only in one or two aspects among IT network, OT network and endpoint security, and very few of them focus on behavioural anomaly detection or on using sensors data for security risk monitoring. However, a complete overview of the security risk landscape can be only be done by considering all levels of the system.

### 3.3. Discussion

In Chapter 3 several aspects of the state-of-the-art of cyber-security in industrial networks were reviewed and its main challenges were discussed from historical, environmental, functional, and technical perspectives, and the most relevant security issues were described. This included technical vulnerabilities in the design and implementation of the systems, vulnerabilities introduced by insecure operational management practices, main threats for industrial systems, solutions for security monitoring, and the use of these data for generating IoRs. This overview was considered necessary to develop a deep understanding on OT and industrial systems cyber-security and developing a continuous risk assessment approach that can be applicable and useful in the context of ICS. This complements the discussion on cyber-risk management and different risk analysis approaches in Chapter2.

As ICS cyber-security, needs a different approach from enterprise IT security, and there is less knowledge and experience in this field, basic security hygiene deficiencies exist in many industrial environments. These deficiencies cover all the aspects of the system including security vulnerabilities in devices and equipment, in their implementation, and the operation of the systems. A review of industry reports and academic literature leaves little or no doubt, that ICS and IoT cyber-security has, generally speaking, a low maturity level. Raising maturity level involves the systematic acquisition of the capability to execute, manage, define, measure, and optimize processes. All the sources consulted, such as [8] [16] [96] [97] and [84], report that it can be found that,

overall, industrial cyber-security is neither well-managed nor does it have properly defined processes. Given that situation, intrusion and anomaly detection methods for industrial networks are not yet well-known or widely used. The concept of IoRs is also not as as that of IoCs, which have been used in several industrial cyber-security monitoring tools oriented to security operations, but not to risk management.

The references described in section 3.2.4 and 3.2.5 confirms that monitoring risk factors in operational or near real time is not a misbegotten idea. Particularly, in the case of cyber-physical systems (IoT, IIoT, ICS), physical-based detection methods was found to be proposed by several researcher, which could be a sign of this approach becoming part of the next generation of ICS security monitoring. However, there are several practical challenges to adopt these methods [125] [126]. The most relevant challenges to implementing security monitoring tools and techniques in ICS and using them as an input to generate IoRs are the following:

**Multi-layered architecture**
As described in Figures 3.1 and 3.2, ICS architecture presents several layers which leads to more opportunities for attack. The variety of technologies involved goes beyond the information technology realm and in some cases typical security mechanisms could be insufficient or unfeasible to apply [1] [127]. This makes security and risk monitoring more relevant since it acts as a compensating control for an in-depth security strategy but at the same time makes it more complex. By proposing an IoR Library, which includes IoRs for each layer (or level) of an ICS system this thesis addresses the need of end to end monitoring of observable of risky conditions.

**Less knowledge and experience**
IoT manufacturers tend to have less knowledge and experience in cyber security than professionals from the software and telecommunications world [128], and at the same time, most cyber-security experts do not have knowledge in operational processes and technologies. This last is very important not only to apply the appropriate cyber-security strategy but also to understand all the broader impacts of a potential cyber-incident. The IoR Library presented in this thesis can help improving understanding of cyber-risks in ICS by linking conditions that can be observed in a system to known attack techniques, which are well documented in the MITRE ICS ATT&CK knowledge base.

**Lack of tools and techniques for physical-based anomaly detection**
While typical detection mechanisms could contribute to stopping cyber-attacks at the manufacturing zone, when an adversary manages to bypass these mechanisms and targets the Cell/Area zone they could remain undetected unless additional detection measures are in place. As noted in 3.2.4, physical-based anomaly detection methods have been identified by several authors as an approach that has potential in ICS and IIoT security. However, no properly-tested anomaly detection methods for ICS are currently available. The increasing availability of commercial security monitoring tools specific to industrial networks makes it possible for companies to implement anomaly detection capabilities in ICS by customising off the shelf solutions. Vendor-independent entities such as NIST have reviewed some of these tools giving a valuable insight into the results that they can provide [12]. However, demonstrations of physical-based anomaly detection are so far very limited. This is not surprising, considering that industries need first to fully implement a cyber-security management system, before moving to more advanced approaches. In private conversation between 2017 and 2021, several security professionals specialised in ICS all agreed that their main challenges are related to asset inventory, network mapping and implementing basic cyber-security hygiene measures. This means that they might not yet be ready for physical based anomaly detection. In this thesis, some simple demonstrations of modelling sensor data to create an anomaly detection system are developed for better illustration of these ideas.

**Complex behaviour**
Some industrial systems can present a variety of states, which makes defining parameters for normal operation not a trivial task. As much as advanced machine learning techniques could also capture this behaviour it will require a training period and still there might be some states that are exceptional but still valid which will be harder to learn. Some authors suggest hybrid methods which consider both machine learning and the input from experts and to combine the strengths of different techniques. It is complex defining rules to identify normal and

abnormal states. This could be resource intensive requiring important amounts of time and expert's involvement.

**Lack of standardisation**

Equipment from different manufacturers can use different communication protocols and have different characteristics to function normally. For example, how devices manage authentication, handle memory, create and use inputs and output images, programming features, operation modes, and different files that are used in normal operation are not exactly the same in devices from different manufacturers or even in different models or generations of products within the same brand. Hence, the lower level implementation of anomaly detection rules and IoRs can present variations depending on the different technologies used.

**Legacy systems**

It is expected that in most organisations, migration to new technologies will be progressive and so they will have to deal with the coexistence of equipment of different generations. The reason for this is that ICS, in contrast to IT systems, are designed to work for decades, and replacement of equipment tends to be highly disruptive and expensive. Therefore ICS cyber-security operations needs to establish different approaches to protect newer systems that might have some built-in security features, and legacy systems, which are likely to be insecure and need to be isolated and connected only in a trusted network.

**Need to establish a trust model**

IoRs based on correlation between operational variables would work only under the assumption that at least some of the variables can be trusted. An advantage of using correlated indicators for physical-based detection is that it would require a lot more knowledge about a specific ICS on the part of the attacker to be able to compromise all potential IoRs at the same time. However, if all the correlated variables are managed by the same entity the attacker could just replay historical input and output images, as in the case of Stuxnet, then they do not need to know their expected behaviour.

Cyber-security in industrial networks is still at an immature stage of development, and despite the increasing interest in the cyber-security industry to address ICS security, there are still many opportunities for introducing new solutions and improving the existing ones. An important challenge is also to increase the level of adoption of existing security solutions in industries. However, even when these solutions are adopted, and risks are properly assessed, it is expected that there will be an important amount of risks that will be accepted or retained for certain period of time. Many examples of this situation can be found; some of them are the following:

- Security patches would require to be properly certified so they do not interfere with the current functionality of the system and they would be applied only in periods of scheduled maintenance or not applied at all. Hence, an industrial system can have known vulnerabilities for a long or indefinite period.
- Some industrial devices might also have compatibility with older Operative Systems, which will have known vulnerabilities. Also new vulnerabilities are often discovered in outdated OS, if this happens, the increased level of risk has to be acknowledged.
- Remote connections might need to be established for administration or maintenance purpose for an established time interval, period during which the risk surface increases. This increased level needs to be acknowledged, also to prevent the remote connections to be enabled for more time than necessary.
- Suppliers and contractors might have privileged access to the system for maintenance purposes or setting up new devices since they often have a more specific know-how about certain aspects or devices of the system. It will not be rare that these third parties will connect their own devices, such as laptops, to access the system, for which also related risks need to be acknowledged.
- Security controls such as anti-malware and firewalls might be temporally or indefinitely disabled to address operational issues. Until they are again activated or compensatory measures are put in place the risk level will increase.

Existing frameworks and risk models can be adapted to operate continuously and get data feeds directly from security monitoring tools and operational data coming from sensors and actuators. Additionally, the results of data analysis done for other purposes such as safety and predictive maintenance can also be used to build behavioural anomaly detection models. In this thesis, the main challenges that will be addressed are the heterogeneity and complexity of the ICS architecture by proposing a continuous risk assessment approach that gathers information at different levels if the system. By including inputs based on behavioural anomaly detection within this information it is also aimed to improve visibility and cyber-security awareness in ICS.

Through the exploration of both industrial and academic sources, an overview of the cyber-security landscape in ICS and IIoT has been provided allowing to align this research with current trends and build upon resources already made available in these references. Among the solutions for security monitoring, particular interest was given to methods of physical-based anomaly detection. This is considered as a relevant aspect of the approach proposed in this thesis since it addresses the monitoring of levels of the system that are exclusive of ICS and do not exist in IT systems, and will be illustrated further in Chapter 6. In the next chapter, the continuous risk assessment methodology will be presented including a description of all the main processes involved and a worked use case.

# 4. The Proposed Continuous Risk Assessment Methodology

Monitoring cyber-security risks in industrial systems addresses the fact that it is not possible to prevent or even identify all possible threats. In addition, the mitigations for some risks might have technical constraints or a high cost. This can lead an organisation to retain or accept certain risks, particularly when they have a very low probability of occurrence. However, if their potential impact is considerable, then risk monitoring becomes necessary as a compensatory control. Continuous risk monitoring aims to detect the presence of vulnerable conditions or threatening events that can increase the likelihood of an incident. In some cases, this could also lead to the exposure of an attack attempt on its early stages. Hence, the methodology proposed is based on information exchange between the risk management and security operations processes. While the continuous risk management process makes use of data from the security monitoring process to estimate risks, it could also provide information to trigger incident response actions.

In order to implement a continuous risk monitoring method, it is necessary to create a baseline reference for the risks that will be monitored. In this chapter, a methodology is proposed to implement continuous risk management, which is based on extending the traditional risk management process. This aims to answer two of the research questions of this thesis: “how can existing cyber-risk management frameworks be adapted for a more dynamic risk monitoring?” and “how can these modifications be introduced?” To demonstrate the methodology, a worked use case based on a temperature control is developed, including as well examples of potential risks and how they would be monitored by the approach proposed in this thesis.

The methodology was built taking in account the activities of a traditional risk management process. The result was a blueprint for how to integrate a continuous risk assessment paradigm with standard cyber-risk management frameworks as exemplified by the ISO/IEC 27005 standard [24]. The reason for focusing on ISO/IEC 27005 is that it provides a definition of activities that are performed in a risk management process with their expected inputs and outputs. This standard describes the “what” and not the “how” of the different activities allowing use of different methods to perform each activity. Other frameworks for cyber-risk management, such as NIST SP 800-37 [34] and NIST SP 800-39 [129], including those which are specific for industrial systems such as IEC 62422 [35] mostly tend also to be aligned with the ISO/IEC 27005. Particularly, it was found that these standards do not present fundamental differences that would affect on the definition of this methodology.

Two papers were published based on the content of this chapter, discussing this proposal for a Continuous Risk Assessment method for ICS and IIoT and demonstrating the ideas through a use case based on a temperature control in a data centre. In [130] an overview of the methodology is given together with a proposed architecture for its implementation, and [131] proposes a process oriented view of the approach based on workflows and descriptions of activities with their expected inputs and outputs. It must be noted that the focus of these publications, as well as of this thesis is on the updates of risk likelihood, leaving considerations on impact updates out of the scope. Hence, for the effect of this research work, impacts are considered unchanged.

In order to walk through the different steps of the methodology, a worked example was developed based on a building temperature control system. Temperature control is a use case that can be found in domestic, commercial, and industrial environments. However, in different domains the system will typically present different characteristics, types of technology, and infrastructure. Temperature control can have different purposes in different industries, for example, avoiding products such as food and chemicals decomposing or degrading, or allowing various processes to perform in optimal conditions. A temperature control malfunction will have different consequences depending on the business processes involved. These consequences usually involve degradation of the quality of products and services, but there can also be safety implications. Understanding these consequences is crucial when assessing risks, and is also a key part of finding IoRs based on sensors outputs, which can be correlated with a temperature anomaly and used in continuous risk monitoring (more about this type of IoR is explained in Chapter 5 and Chapter 6). The use case developed for this example is an invented case based on an environmental control system used in a Data Centre, which was validated with professionals with experience on this type of systems.

This chapter consists of four sections. The first section explains the methodology and its phases. The second section describes the worked example and its risk landscape, which will be used to demonstrate the methodology. The demonstration of the methodology based on this worked example is done in the third section

giving also examples of expected results. The fourth section provides a discussion of the chapter, highlighting the research contributions.

## 4.1. Description of the Continuous Risk Assessment Methodology

The main objective of this research is to enable continuous monitoring of security risks and to increase cyber-situational awareness in industrial environments. The methodology described in this chapter illustrates an adaptation of the traditional cyber-risk management approach oriented to continuous risk assessment, which includes integrating risk monitoring activities with a security operations system based on a SIEM infrastructure. The proposed framework is designed for risk analysts, whose role consist on analysing and evaluating risks and triggering risk treatment plans to reduce those risks that the organisation cannot accept. However, the results can be shared with the roles responsible for security operations and incident response who usually are represented by a Security Operations Centre (SOC) analyst or its equivalent role. If there is a risk whose likelihood of occurrence increases beyond an acceptable level, the Continuous Risk Assessment process generates alerts that are forwarded both to the risk analysts and to the security operations team. Additionally, suspicious events, behaviours that deviate from normal or any abrupt changes should also be escalated to the appropriate stakeholders, such as process or asset owners.

The intention of proposing a continuous risk assessment methodology is not only to improve risk management by having continuous updates on information, but also to integrate it with security operations, which often is managed separately. Security risk managers are responsible for reviewing and reporting on the levels of risk associated with various categories of threat and, if risk appetite is exceeded, altering the make-up or setting of mitigating controls to bring the risk back within bounds. SOC staff, in contrast, watch out for signs of attacks in progress and intervene if necessary to prevent them being successful. Effectively, the SOC forms part of the controls, and its processes are influenced by policy updates from risk managers. Conversely, operational security metrics published by the SOC are important inputs to periodic risk assessment. Until recently, risk management had a characteristic timescale of months or quarters; in contrast, a SOC has to respond to an attack in minutes, hours, or days, and in the future will need to respond even more rapidly. With the transition to continuous risk monitoring, driven by the ever-more dynamic threat environment, risk management timescales should converge with those of the SOC, and it is appropriate to look upon them as concurrent communicating processes sharing information and using a common technical platform, as described in Figure 4.1.

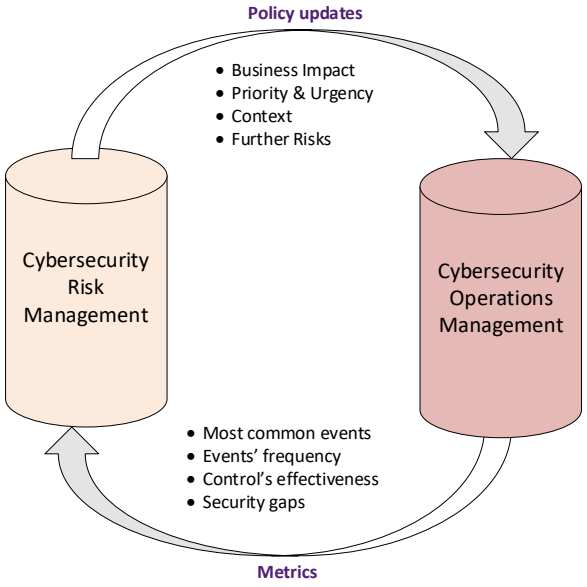



*Figure 4.1: Cyber-security Risk and Operations Management feedback*

Industry standards do not define processes for continuous risk assessment and most of them do not explicitly make any comments on the possibility of monitoring risk factors continuously. However, it was found that the approach that will be described in this chapter does not have fundamental contradictions with references such

as ISA/IEC, ISO 27005, NIST 800-37, or OG86. Most Risk Management models describe a risk assessment as a PDCA cycle (Plan-Do-Check-Act) in which risk monitoring (check) has the purpose of identifying changes in risk factors in an early stage [24]. For this, an organisation should identify a frequency for reassessing risks and triggering criteria for risk alerts, which addresses the fast-changing nature of cyber security [35]. It must be noted that shifting to continuous risk monitoring does not only means increasing the frequency of risk assessments but it is an event-driven process by which risk is re-assessed whenever a risk factor changes. Figure 4.2 presents the proposed architecture for continuous risk assessment. It shows a general overview of the continuous adjustment of risk scores based on the influence of IoRs, which serve as inputs to a Bayesian Network (BN), which is used for probability recalculation, which is used in turn to update risk scores. The outcomes of the responses to Risk Alerts are used to update the parameters of the BN. This, together with new findings that could require modifying the conditional probability estimations are part of the continuous improvement processes that shall be part of this approach.

The proposed methodology for implementation of continuous risk assessment has three phases: the Baseline (initial) Risk Assessment, the Transition Phase, and the Continuous Risk Assessment Phase. The Baseline Risk Assessment follows a standard risk management process, but includes additional activities and work products in order to establish the basis for continuous Risk Assessment. In the Transition Phase, the organisation implements all the tools and processes that are necessary to enable Continuous Risk Assessment. In the Continuous Risk Assessment phase, risk scores get updated in the light of new information. Figure 4.3 shows a workflow of the “macro-process" of the methodology using BPMN (Business Process Model and Notation), which is a widely-used notation for workflow modelling. While the first two phases are focused on the implementation of the continuous risk assessment, the third and last phase corresponds to the execution of the continuous risk assessment during normal operation. Rather than redefining the traditional risk management process this thesis proposes to extend it to support a continuous mode of working.

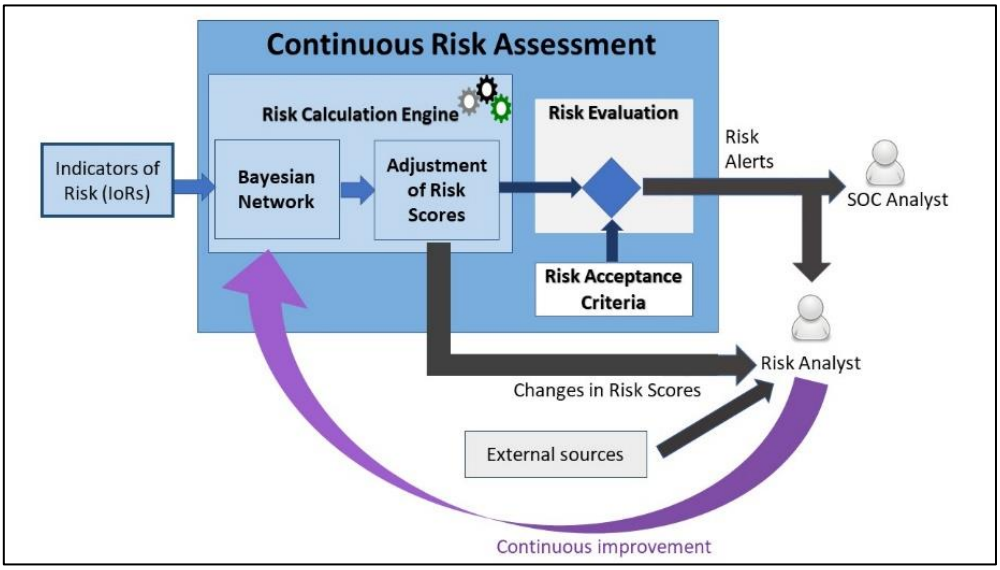



*Figure 4.2: Architecture for continuous risk assessment*

In the first phase additional activities and work products are added to the standard risk management process as a preparation for continuous risk assessment. In the following sub-sections each phase will be described with more detail including lower-level workflows. In a minor reinterpretation of BPMN visual modelling notation, lanes will be used to distinguish which activities are covered by the ISO/IEC 27005 standard and which are introduced as part of this work. This presents a difference with standard BPMN, in which lanes are used to separate roles.

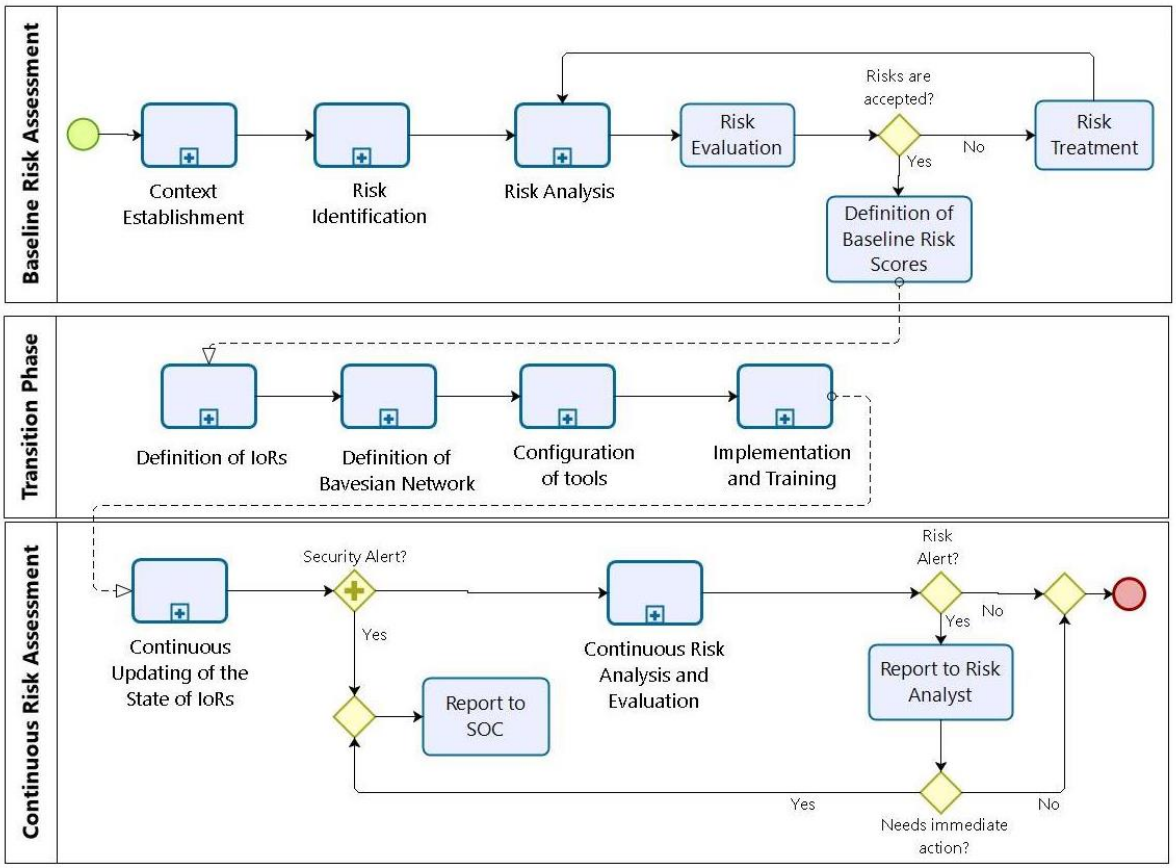



*Figure 4.3: Macro-process of the methodology*

### 4.1.1. Baseline Risk Management phase

In this phase, baseline risk scores are generated which will be used as a reference in continuous risk assessment. As can be observed in Figure 4.3 the main processes involved in this phase are context establishment, risk assessment (which encompasses risk identification, risk analysis, and risk evaluation), and risk treatment. This has no major variations with respect to a traditional risk management process, although, additional activities are added to each process in order to support the continuous risk assessment methodology. For example, information needed to model the system's functional behaviour under normal operation is gathered during Context Establishment.

#### 4.1.1.1. Context Establishment

In Context Establishment, all the information relevant to risk management is gathered and the basic criteria for the risk assessment, such as the scope of the analysis, risk acceptance criteria, and roles and responsibilities for cyber-security risk management, are defined [24]. This process is critical for the success of continuous risk assessment since it sets the priorities and objectives, and defines the risk tolerance level of the organisation. As it is not feasible to analyse all the possible risks, the scope needs to focus on the most critical assets, also known as “crown jewels" [42] [83]. In this stage of the methodology, it is crucial to capture the knowledge of experts about the industrial operations, including requirements, critical assets, business rules and expected behaviour of different system variables. When the scope is clear, it is important to establish which events and conditions of the system need to be continuously monitored. This will allow IoRs to be defined later based on information gathered from experts and historical data, which should be used to model the “normal” behaviour of the system to be used as a benchmark to identify unusual events during operation.

Characteristics of the infrastructure, network architecture, business rules, and the controllers’ programs and configurations need to be known at this stage in order to define the risks and respective IoRs to be monitored during Continuous Risk Assessment. Figure 4.4 shows the workflow for Context Establishment including the additional activities that establish the data sources that will be used during continuous risk assessment.

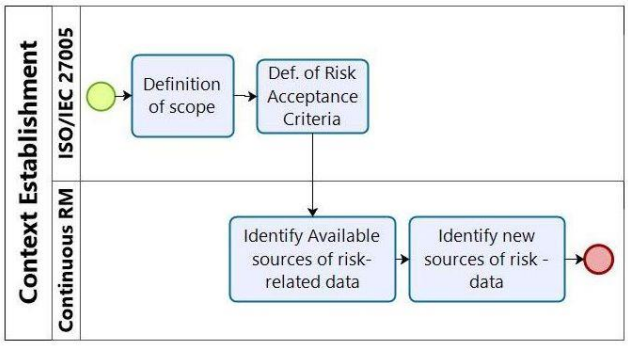



*Figure 4.4: Context Establishment for the Continuous Risk Assessment*

**Inputs:** Inputs to Context Establishment consist of all information about the organization that is relevant to information security risk management [24]. In the continuous risk management approach, this will also include information about the data that can be monitored on a continuous basis and can be associated with cyber-risks.

**Outputs:** Outputs from Context Establishment consist of documents specifying basic criteria, which includes risk evaluation criteria, scope and boundaries, and the organization of the security risk management processes [24]. In the continuous risk management approach, an additional output should be documents identifying the type of data sources currently available, which can be used at a later stage to monitor risks and also other possible mechanisms to extract data that is relevant for the risk analysis that have not been implemented yet. Examples of these sources of risk relevant data are system logs, I/O data, alerts, threat intelligence sources, etc.

#### 4.1.1.2. Baseline Risk Assessment

The baseline risk assessment consists of risk identification, risk analysis and risk evaluation as shown in Figure 4.3. Risks that can affect the assets within the scope of the assessment should be identified and rated, considering the possible threats to these assets, the security vulnerabilities that could facilitate the success of an attack attempt, and the impacts. Once identified the most important assets and processes that require protection (aka “crown jewels”), risks can be analysed respect to their significance in the given environment. Once rated, the risks are evaluated to check if there are beyond the tolerance levels of the organisation, in which case, a risk treatment plan needs to be introduced.

#### 4.1.1.3. Baseline Risk Assessment: Risk Identification

Risk identification must take into consideration all potential security issues that can affect the system or systems that are within the scope of the risk assessment. In this approach, he main risk identification activities, are the same as in a traditional risk assessment, i.e. identification of assets, identification of threats, identification of security controls and vulnerabilities, and identification of impacts.

**Inputs:** The inputs to risk identification are the scope and boundaries for the risk assessment to be conducted [24] as defined during context establishment.

**Outputs:** The outputs of risk identification are a list of assets and the business processes related to them, together with their relevance, a list of threats, a list of existing and planned security controls, a list of vulnerabilities, and a list of incident scenarios with their consequences to assets and business processes [24].

#### 4.1.1.4. Baseline Risk Assessment: Risk Analysis

Risk analysis consists of quantifying or rating risks for sub-sequent evaluation. For this, a threat model should be developed to describe the threats related to the risks that were identified. As it is important to have a common language to describe an attack and its different steps or specific objectives to be achieved, it was decided to base the vocabulary for the threat model on the ATT&CK framework from MITRE. In the ATT&CK taxonomy, attacker’s tactical goals in different stages of a “kill chain” are known as Tactics, the actions that can be

performed to achieve these goals are called Techniques and the Procedures are specific implementations adversaries have used for techniques or sub-techniques. Example of procedures are known malware or behaviour typical of specific threat groups [132]. Several security detection tools have playbooks already aligned with ATT&CK [133] [134] [135], so adopting it can facilitate the overall implementation of continuous risk monitoring.

The threat model shall consist on the identification and representation of different possible attack strategies used by an adversary in which a set of complementary and alternative TTPs (Tactics, Techniques, and Procedures) that could be used for each strategy are associated it. An example of a potentially useful and well-known way to start building the threat model is through attack trees, which provide a high-level approach, which can serve for illustration and communication purposes. Each step or intermediate goal of the tree can be associated to TTPs from the ATT&CK framework. This threat model should as well provide the basis for building the BN for continuous risk assessment, which will be fully defined in the Transition Phase.

The continuous risk assessment methodology was designed to be agnostic of the specific method used to calculate the baseline risk scores. However, it is important that the method used allows establishing the probability for each Technique and Tactic to use this as a reference later in the Bayesian Network. In the use case that will be presented next section, an example of risk analysis will be provided based on a customisation of some of the methods described in Chapter 2. Figure 4.5 shows the workflow of the risk analysis including the additional activities, which are related to build a threat model that allows the identification of IoRs.

Threat models allow analysing the risk of a threat developing at its different stages and IoRs are means of observation of the possible development of a threat, which are related to it through conditional probabilities. IoRs are derived based on the TTPs, which were identified in the threat model. For each TTP, it is required to find observable means of inference of the risk of this TTP to be executed. For example, different types of malformed network traffic could imply attempts to send malicious commands or to crush an ICS system by sending random messages. As the task of identifying several IoRs for each TTP, can be cumbersome and overwhelming, part of this research focused on facilitating this task. For this purpose, a knowledge base is proposed, which provides guidance on identifying the type of IoRs that can be associated with different adversarial Techniques from ICS ATT&CK. This instrument, which is described in more detail in Chapter 5, has the objective of assisting the activity of IoR identification and it is named as “IoR Library”. Users of the Continuous Risk Assessment Methodology can consult the IoR Library and find a list of possible types of IoRs associated with the different Techniques that they have identified in their threat model. For Tactics alone, they can consider the IoRs that are associated to the different Techniques that belong to the specific Tactic and for Procedures, lower level or more specific indicators can be derived from the corresponding Technique IoRs, as applicable.

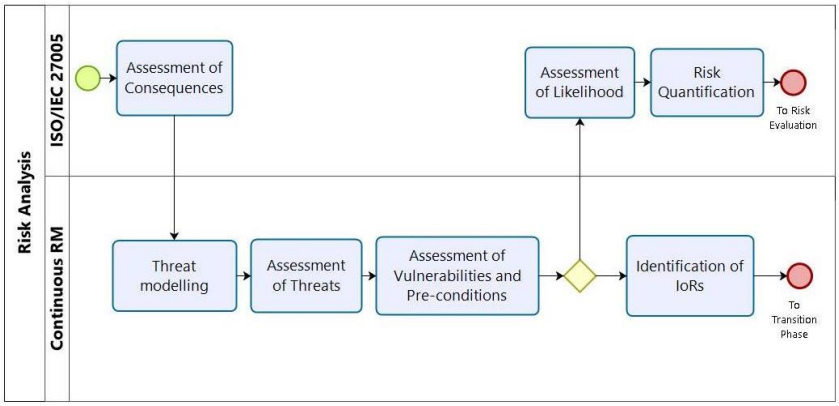



*Figure 4.5: Baseline Risk Analysis*

The estimation of risks in the baseline risk assessment should be done through a threat, vulnerability, and impact analysis based on the characteristics of the system and its environment and the information available at the moment when the analysis is undertaken. The baseline risk scores are based on the probability of occurrence and likely impact of each risk. The calculation of probability is based on the quantification of threats and

vulnerabilities. The threat models are used as a starting point for this. The overall probability of a successful attack is obtained by consolidating the individual probabilities of each step in a kill chain.

**Inputs:** The main input to risk analysis is a list of identified relevant incident scenarios, including identification of threats, vulnerabilities, affected assets, and consequences to assets and business processes [24].

**Outputs:** The outputs of risk analysis are a list of incident scenarios with an estimation (qualitative or quantitative) of their likelihood of occurrence and consequences to assets and processes [24]. This estimation has ultimately to be translated into a quantitative scale in order to define the baseline risk scores for the continuous risk assessment. Additional outputs specific to the continuous risk assessment methodology are threat models and IoRs.

#### 4.1.1.5. Baseline Risk Assessment: Risk Evaluation

In Risk Evaluation, the risks above the acceptable level are identified for further review and treatment. The definition of the “acceptable level" is given by senior management in accordance with the organisation's tolerance to risk, otherwise known as risk appetite, and should be specified in the Context Establishment process. Risk acceptance, which is the decision to “live with the risk” should mostly happen only for risks within this acceptable level, otherwise the risk level should be reduced through the treatment process.

**Inputs:** The inputs to Risk Evaluation are a list of risks with value levels assigned obtained from the risk analysis and the risk evaluation criteria obtained from the context establishment [24].

**Outputs:** The outputs from Risk Evaluation are a list of risks prioritized according to risk evaluation criteria [24] indicating which risks are considered to have an acceptable level and which are not.

#### 4.1.1.6. Risk Treatment

The objective of risk treatment is to reduce risks to a tolerated or accepted level. This can be done through three possible actions: reduction, avoidance, sharing. There is also a fourth option, namely to retain the risk. Risk retention could happen because the organisation can consider than a particular risk can be tolerated, despite exceeding the specified risk acceptance criteria, particularly if the only available mitigations can compromise other businesses’ objectives, such as operational performance. If a risk above the allowed level is retained, for example due to the high cost or infeasibility of the available treatment, the decision has to be approved by senior management. Nevertheless, most of the time, a risk treatment plan will define actions that result in a reduction of the risk level by either reducing the impact or the exposure to the risk, in other words, the likelihood. Overall, besides the risk itself, a business will also assess the cost-benefit ratio of spending on security controls and mitigations.

It must be noted that acknowledging and accepting a high level of risk is very different, at least in principle, from underestimating a risk. Optimism bias is common when risks are not properly analysed, and can lead to an organisation being unprepared should this risk arise. On the other hand, the retention of a risk implies that there is awareness about it, which allows for it to be mitigated in the future and in the meanwhile to be monitored. This enables timely reaction when there is reasonable evidence that the risk can materialise in the near future. In an industrial environment it is not uncommon to make trade-offs between cyber-security risks and other business objectives or to have serious technical constraints that limit options to reduce certain risks. In these cases, implementation of a continuous risk assessment methodology becomes particularly important as a compensating control that allows actions to be taken promptly if there is evidence that a risk is becoming an issue.

**Inputs:** The inputs of the Risk Treatment process correspond to the results of the risk evaluation in which a list of risks with value levels in generated indicating their condition in respect to the risk evaluation criteria [24].

**Outputs:** The result of the Risk Treatment process is a risk treatment plan plus a list of risks subject to the acceptance decision of the organization’s senior management [24].

#### 4.1.1.7. Definition of the Baseline Risk Scores

After implementation of the risk treatment plan, risks are quantified and evaluated again to check their value has been reduced and they can be accepted. The risks remaining after risk treatment is executed are known as "residual risks" and are expected to lie within the organisation's risk appetite. After these stages are finished, the risk scores are used as "baseline risk scores". These scores should provide a benchmark based on the information and observations of the ICS that are available at a certain point of time. For this reason, this methodology considers the results of the baseline risk assessment as referential scores, which could be modified during the continuous risk assessment.

> **Inputs:** The inputs for defining the baseline risk scores are the work products involved in the Baseline Risk Assessment, plus the results or current state of implementation of the Risk Treatment Plan.
>
> **Outputs:** The outputs of the definition of baseline risk scores are updated versions of the list of incident scenarios with an estimation (qualitative or quantitative) of their likelihood of occurrence and consequences to assets and processes generated in the Risk Analysis.

Until this point, differences between the methodology and the ISO/IEC 27005 are minimal and consist only of some additional activities, which would allow the organisation to establish the basis needed for continuous risk assessment. The next phase (transition) corresponds to the implementation of the continuous risk management system, which is designed to monitor and re-assess risks at any given time. The last phase corresponds to the continuous risk management operations.

#### 4.1.2. Transition phase

As the name suggests, this phase has the objective to transition from the initial or Baseline Risk Assessment to the Continuous Risk Assessment phase. In the Transition Phase, the events and conditions that will be monitored in Continuous Risk Assessment are linked to the risks analysed in the previous stage. Means of detecting or inferring evidence of the various stages of an attack, and the pre-conditions required for particular attack methods to be executed, are identified and linked to the threat model created during the baseline risk assessment. These observations of conditions and events are ultimately defined in terms of IoRs. An IoR is a condition, the observation of which affects the estimated probability of one or more possible threat events, which are in most cases defined in terms of TTPs (Tactics, Techniques and Procedures), as defined in MITRE ATT&CK and ICS ATT&CK data bases. Multiple IoRs can be related to multiple possible threat events through the construction of a BN.

In order to prepare for the next phase, tools need to be setup and configured in order to allow real time generation of the IoRs. Monitoring and detection tools such as firewalls, Intrusion Detection Systems (IDS), malware detection, network monitoring and, log monitoring tools need to be configured to generate inputs for the SIEM, and the SIEM itself, which will process this information and feed the BN with the corresponding IoRs. For IoRs related to physical variables, misbehaviour or anomaly detection methods specific to the ICS need to be defined. This is done by modelling the normal or expected behaviour of the system, so that the monitoring system can confirm that the real behaviour does not deviate from this model. More details about the use of anomaly detection methods to check on misbehaviour of physical variables as an input for continuous risk monitoring will be developed in Chapter 6.

The Transition Phase proposed consists of three main processes, which are the identification of IoRs, definition of the BN, and implementation of the system and training the organisation, which includes adapting tools, processes and people.

##### 4.1.2.1. Definition of IoRs

In the previous phase, TTPs were identified based on the threat model developed during the Baseline Risk Assessment. The IoRs, which are the observable variables that modify the risk likelihood estimations, can be identified based on these TTPs. However, for IoRs to be observed in a particular system, they should be defined as specific use cases, which shall be implemented in a SIEM tool. For this, it should also be taken in account the available capabilities regarding detection mechanisms and information processing. Hence, definition of IoRs will include both, specifying the IoRs identified at a lower level and deciding on which of them can be continuously monitored. The "IoR Library", which will be described in more detail in Chapter 5, provides a rationale for each IoR and examples of possible means of observations, which can assist on the IoR definition. However, as each

industrial environment is particular, the IoR definition activity would require the expertise and knowledge of the administrators and operators, as well as of those in charge of defining and implementing the SIEM use cases and associated security detection tools for the ICS. The result of this activity should be a low-level specification of each one of the IoRs to be implemented, which should include details such as specific inputs for the IoR observation, detection rules, and tools required.

**Inputs:** The inputs for the Definition of IoRs are the results of the IoR identification performed during the baseline risk assessment.

**Outputs:** The output of the Definition of IoRs is the final list of IoRs that will be implemented for the Continuous Risk Assessment including their low-level specification.

#### 4.1.2.2. Definition of the Bayesian Network.

BNs are directed acyclic graphs in which the nodes are variables and the arcs represent conditional dependencies between them [136]. They are, based on probabilistic or "Bayesian" inference theory and are used in different fields because of their potential in helping dealing with uncertainty [137]. Each node has a table of conditional probabilities associated that relates the probabilities of each one of its possible values to the possible values of the nodes it depends on.

In the Continuous Risk Assessment methodology, a BN node can correspond to an IoR, a tactic, a technique, or a procedure or any event associated to a threat that was defined in the threat model built during Baseline Risk Assessment. The user can add further nodes representing different things if this is useful, such as events that they want to monitor or specific operational or business risks derived from the successful execution of a TTP. The Bayesian network will be built in three steps: the first step is to define the nodes of the network; the second step is to establish relationships between the different nodes, and the third step is to define the conditional probabilities between nodes.

**Step 1: define the nodes of the Bayesian Network**

In the threat model developed during the Baseline Risk Assessment, TTPs and IoRs were identified and the IoRs that would be actually monitored were defined. Each IoR and TTP should be defined as a node in the BN. The user can also define additional or auxiliary nodes, which can help visualisation of events that are considered important, which are not specified in the ATT&CK framework as TTPs. Examples of this are different forms in which a TTP can be executed in a particular context. These additional nodes would allow the user to have an explicit probability estimation for threat events that they consider important to monitor. As it is not possible to capture all the possible attack mechanisms in the threat modelling, one or more nodes will be defined to represent unknown risks or zero-day attacks. This idea will be developed further in section 5.7 of Chapter 5.

**Step 2: Establish relationships between the different nodes.**

A node is said to be dependent if its state is influenced by the state of other nodes. The state of independent nodes is defined by data provided by observations made in the course of monitoring the system. The relationships between nodes are denoted by arrows in the graphical representation of the Bayesian Network. The direction of the arrow goes from the node that exerts an influence to the one that is influenced. Alternatively, this relationship can be represented through an incidence matrix, which is used to compute the probability calculations. An incidence matrix has elements which can be either "1" or "0" indicating if there is or not a relationship between variables.

For the Continuous Risk Monitoring, IoRs will be independent nodes and Techniques and Tactics will be dependent nodes. Dependencies between IoRs and techniques are established in the threat model previously defined, as well as dependencies between different TTPs. As this is proposed to be based in the ICS ATT&CK knowledge base, these dependencies can be inferred also from it. For example, a Tactic node will depend on the nodes that correspond to the Techniques that are associated to this Tactic in the ATT&CK model. To assist the definition of the relationship between nodes, in Chapter 5 a BN Template is proposed, which is based in the "IoR Library" and already contains IoR, Techniques and Tactic nodes and their links. The user can make modifications to this template such as deleting nodes that are not considered in their threat model, adding new nodes for customised event monitoring, and modifying relationships, if found necessary.

**Step 3: define the conditional probabilities**

Conditional probabilities are used to calculate the likelihood of a node having a certain state given the observation of the states of the nodes that have an influence on it. In general terms, conditional probabilities are calculated according to the following formula:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Where, if A represents the proposition that Node 1 is in state A, and B represents the proposition that Node 2 is in state B, then:

- P(B|A) is the probability of observing Node 2 in state "B" if the related Node 1 presents the state "A".
- P(A) is the probability of Node 1 having the state "A"
- P(A∩B) is the probability that Node 1 is in state "A" and Node 2 is in state "B" at the same time.

Once the conditional probabilities are defined, for example, based on historical data, expert judgement or a combination of both, the probability of the event "A" can be calculated by backwards inference, as follows:

$$P(A) = \frac{P(A \cap B)}{P(B|A)} = \frac{P(B) \times P(A|B)}{P(B|A)}$$

If P(A|B) and P(B|A) are constant, then it is valid the following simplified equation in which k is a constant that represents the quotient between P(A|B) and P(B|A).

$$P(A) = k \times P(B)$$

Then, the higher the probability of state "B" being observed in the node 2, the higher the probability that the node 1 is indeed in state "A". In the context of continuous risk monitoring, "B" might represent the observation of an IoR and "A" the successful application of an attack Technique. Figure 4.6 shows an example of conditional probabilities between three different nodes in a Bayesian Network.


| IoR | Probability of Technique's execution |
|---|---|
| False | 0.1% |
| True | 20% |

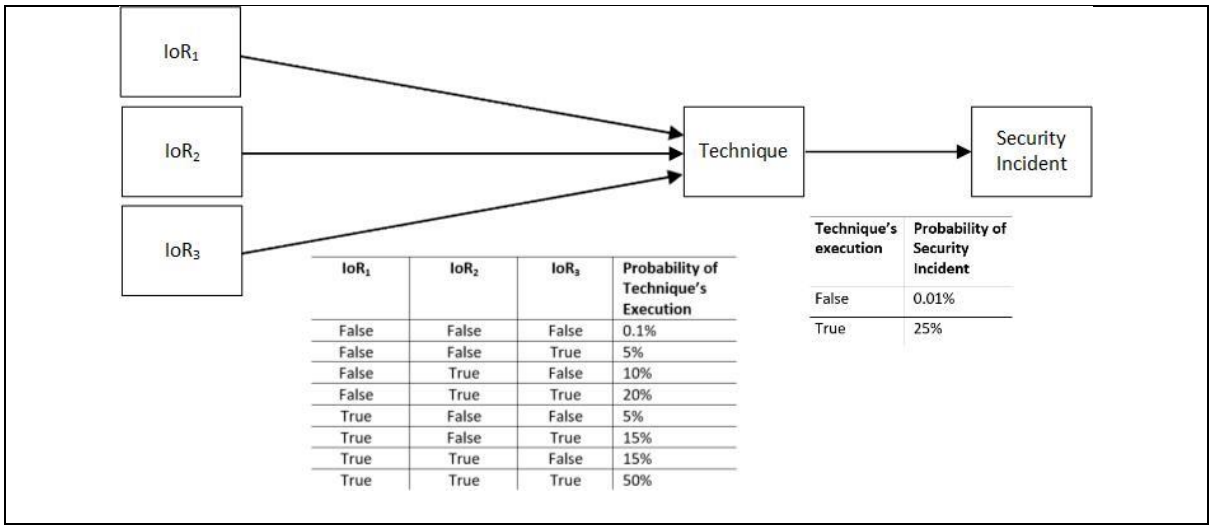
| Technique's execution | Probability of Security Incident |
|---|---|
| False | 0.01% |
| True | 25% |

*Figure 4.6: Example of conditional probabilities*

As the probability calculation of a Technique will depend on several IoRs being observed, the calculation of a Technique node in the Bayesian Network will be done as in the following example:

$$P(Technique) = k \times P(IoR_1 \cap IoR_2 \cap IoR_3)$$

Figure 4.7 shows an example of conditional probabilities between different nodes when multiple IoRs influence a Technique and subsequently this Technique is linked to a node related to a security incident.




| $IoR_1$ | $IoR_2$ | $IoR_3$ | Probability of Technique's Execution |
|---|---|---|---|
| False | False | False | 0.1% |
| False | False | True | 5% |
| False | True | False | 10% |
| False | True | True | 20% |
| True | False | False | 5% |
| True | False | True | 15% |
| True | True | False | 15% |
| True | True | True | 50% |

| Technique's execution | Probability of Security Incident |
|---|---|
| False | 0.01% |
| True | 25% |

*Figure 4.7: Conditional probabilities example with multiple IoRs*

Conditional probabilities can be estimated in different ways. If sufficient historic data is available, this can be used, although it is not always safe to assume that the past is a good predictor of the future. Hence, the judgement of one or more experts can be used either to confirm the validity of estimates based on data or to perform a probability estimation based on their knowledge and experience. A well-known method for groups of experts to agree on estimates of variables is the Delphi method [138]. Here all experts give their opinion independently, and then they agree on a consensus value. As such a method is not scalable for assigning conditional probabilities to high amount of IoR and Technique pairs, in Chapter 5, it is proposed to have a knowledge base linking IoRs to known attack TTPs which includes pre-assigned ranges of conditional probability values. This suggested conditional probability range is named as "degree of influence". Regardless of the method used to estimate conditional probability values, these values can be adjusted during Continuous Risk Assessment as part of the continuous improvement of the method. In the example of Figure 4.7, it can be observed in the table that $IoR_1$ alone increases the probability of a technique from 0.1% to 5%. This means that for states in which $IoR_2$ or $IoR_3$ are detected in addition to $IoR_1$, the probability needs to by equal or higher to 5%.

For each risk, it is important that the estimate of the probabilities of occurrence of each IoR in combination with the conditional probabilities result in a probability that is within the range of the probabilities obtained in the Baseline Risk Assessment. Any significant deviation should imply a review of assumptions and methods. As explained in Chapter 2, it is usual in English language to refer to the term "likelihood", to describe chances of occurrence through a qualitative assessment and the term "probability" when the assessment is quantitative [23]. Table 4.1 is an example of how to define an equivalence between probabilities and the likelihood estimated in the Baseline Risk Assessment for a node and the initial probabilities obtained through the BN.

*Table 4.1: probability-likelihood equivalence*

| Likelihood rating | Probability range |
|---|---|
| Very Low | Higher or equal than 0% and less than 0.01% |
| Low | Higher or equal than 0.01% and less than 0.1% |
| Moderate | Higher or equal than 0.1% and less than 1% |
| High | Higher or equal than 1% and less than 10% |
| Very High | Higher or equal than 10% and less than 100% |

It can be observed in Table 4.1, that a logarithmic scale was chosen to illustrate the likelihood variation with a better resolution since the perception of risk exposure (likelihood) will not necessarily vary in regular percentage intervals. For example, a probability of occurrence of 10% is considered already as "very high" since it means in

practice that if the conditions that lead to this probability calculation persist in time, the event occurrence will be almost imminent unless something is done about the risk. It can be found that in several scientific investigations related to cyber-risk quantification, such as [139] and [140], non-linear approaches for probability estimation are used. Another example is the FAIR method, which regardless that they define likelihood in terms of frequency instead of probability, follows the same principle of defining likelihood ratings on a non-linear scale [38].

Initial or baseline probabilities of the BN are those defined by the default state of IoRs, which is expected to be "false" in most cases, but not in all of them. For example, the IoR "outdated OS" or "unpatched firmware" could be true for a long period of time, since it is not always possible to do timely security updates in ICS environments, although the best cyber-security practices might suggest the opposite.

It is essential to do a consistency check between the likelihood estimation done in the Baseline Risk Assessment and the one obtained in the BN. Using a completely different risk analysis method from the Bayesian Network in the Baseline Risk Analysis can allow to take into account other methods to quantify risk factors. For example, vulnerability scoring and estimated capabilities of different threat actors. Additionally, comparing the probabilities obtained through both methods serves as a mean of validation.

> **Inputs:** The inputs to the Definition of the BN are the final list of IoRs, the threat models associated with different risks, the link between IoRs and the TTPs of the threat models (e.g. the "IoR Library"), and the results of the Baseline Risk Assessment.
>
> **Outputs:** The outputs from Definition of the BN is fully implemented BN that will be used for Continuous Risk Assessment . This means having all the nodes, defined, connected to their corresponding parent and child nodes, and all the conditional probabilities assigned.

#### 4.1.2.3. Implementation and Training

Once the IoRs that shall be monitored to continuously assess risks are defined, the various tools that will support the methodology should be setup and configured. This includes programming different rules to generate IoRs and use them for updating risk scores and triggering alerts. Examples of tools that need to be configured are the firewalls, IDS, malware detection, network monitoring, log monitoring, misbehaviour detection, and the SIEM. The risk calculation engine based on BNs should also be configured and initial values set according to the results of the Baseline Risk Assessment.

The Transition Phase can be complex considering the number of tools, variables, and methods involved. To overcome this, it is possible to approach the adoption of the Continuous Risk Assessment as an agile project, in which incremental changes are applied at different stages. For example, start monitoring risks within a limited scope (e.g. only key critical processes and assets) and then increase the scope when more experience in the process is gained. An adjustment period will be necessary to refine detection tool configuration and algorithms. Tool settings and parameters will need to be adjusted, and calibrated in order to maximise accuracy, and minimise false positives. This would require exploring different detection methods and working under a continuous improvement philosophy to allow lessons learned from experience to be incorporated, the low-level definition and implementation of IoRs to be improved, and conditional probabilities to be adjusted.

A training period will be necessary, as well, for users (risk analysts and security operators) and other stakeholders (such as ICS operators and managers) to get familiar with the approach and the related processes and tools. A training plan should be developed for this purpose, which will depend on the organisation's training needs and requirements. In general terms, it is expected for the training plan to consider the details of the implementation of the risk assessment methodology in the organization including tools, processes, and roles and responsibilities of relevant stakeholders.

A detailed description of the Implementation and Training related activities is beyond the scope of this methodology, since they highly depend on the context.

> **Inputs:** The inputs to the Implementation process are the results of the baseline risk assessment, the final list of IoRs and the BN and the low-level definition of the relevant processes for the Continuous Risk Assessment . The input to the Training process is a training plan.
>
> **Outputs:** The outputs from Implementation and Training consist of having tools, processes and people ready for the Continuous Risk Assessment phase.

### 4.1.3. Continuous Risk Assessment phase

The main processes that occur in this phase are Continuous Updating of the State of IoRs, Continuous Risk Analysis and Continuous Risk Evaluation. The updated state of IoRs is monitored and used as input for the Continuous Risk Analysis process, in which risk scores are adjusted. The adjusted scores are compared to the risk tolerance of the organisation and evaluated, having as a result a risk alert, if any adjusted risk score exceeds its acceptable level. This whole process is named as "Continuous Risk Assessment", based on the terminology of ISO/IEC 27005, which defines "risk assessment" as the sequence of "Risk Identification", "Risk Analysis" and "Risk Evaluation", as shown in Figure 1.3. The updates on the state of IoRs will be provided by a SIEM that continuously ingests and analyses data from various sources including security sensors and logs, and vulnerability intelligence sources (see Figure 1.2). A relevant aspect of the Continuous Risk Assessment Methodology is that it allows making use of data that is already available and in many cases also used for security monitoring. However, external sources of data can also be considered. For example, information about a zero-day vulnerability coming from another source (e.g. threat intelligence) can make it necessary to modify some parameters or to add information manually.

#### 4.1.3.1. Continuous Updating of the State of IoRs

Figure 4.8 shows the Continuous Updating of the State of IoRs process, which consists of identifying variations on the state of an IoR.

At this stage, there might be conditions, such as a combination of IoCs and IoRs that give an indication of an imminent attack attempt, which means that it would be possible to immediately trigger a security alert that is reported to the SOC or equivalent area, allowing them to make use of this information independently from the risk analysis process. While a security alert can be generated at this stage and requires immediate action, the risk alert is only generated after the risk analysis and evaluation processes and it can have different levels of urgency.

**Inputs:** The inputs to the continuous Updating of the State of IoRs are the observations that define the state of an IoR.

**Outputs:** The outputs from the Continuous Monitoring of IoRs are the current states of IoRs. The simplest example of this is that if the IoR is observed, then its state can be set as "true" and if it is not observed, the state is "false". Another type of output of this process are security alerts.

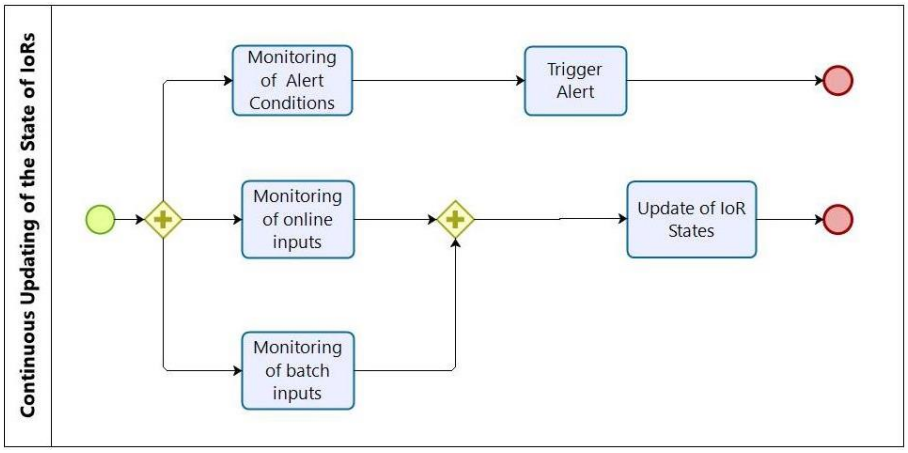



*Figure 4.8: Continuous Updating of the State of IoRs*

#### 4.1.3.2. Continuous Risk Analysis and Evaluation

A Risk Assessment as defined in the ISO/IEC 27005, consists of Risk identification, Risk Analysis, and Risk Evaluation. It must be noted that as the Risk Identification is done in the Baseline Risk Assessment, the Continuous Risk Assessment consists only of Risk Analysis and Risk Evaluation. Figure 4.9 shows the Continuous Risk Analysis and Evaluation processes, in which real time observations lead to a recalculation of conditional

probabilities in the Bayesian Network, and consequently, the corresponding risk scores are updated. As IoRs are updated and monitored continuously, risk scores can be updated at any moment.

Changes to the risk scores during the continuous risk analysis can have as a root cause that assumptions made in the baseline risk assessment were inaccurate or biased and risks were underestimated or overestimated, but also it can be due to internal and external changes that affect the system's security posture. In parallel to the process of continuous update of risk scores through the BN, new information about threats and vulnerabilities should lead to also updating the structure and conditional probabilities in the BN. For example, if information about zero day vulnerabilities is published that affects some of components of the system, the chances of success of an attacker can improve and the conditional probabilities need to be adjusted accordingly. For example, by increasing the conditional probability of the related TTPs when certain IoRs are observed. This is part of the continuous improvement of the methods, which is illustrated in Figure 4.2.

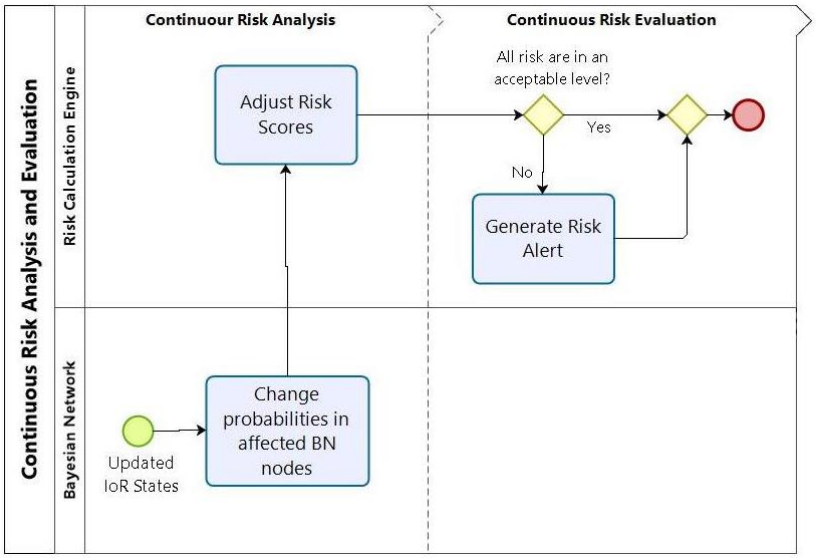



*Figure 4.9: Continuous Risk Analysis and Evaluation*

**Inputs:** The inputs of the Continuous Risk Analysis are the updated states of IoRs and security alerts.

**Outputs**: The outputs of the Continuous Risk analysis are adjusted risk scores.

If any updated risk score exceeds the acceptable level, a "risk alert" is triggered. The decision on whether an adjusted risk score is acceptable or not is made through the Continuous Risk Evaluation process. All alerts get reported to the risk analyst who should propose the implementation of additional security and risk mitigation measures. In the case of risk alerts that require some immediate reaction, the SOC also receives the alert.

**Inputs:** The inputs of the Continuous Risk Evaluation are adjusted risk scores and security alerts.

**Outputs**: The outputs of the Continuous Risk Evaluation are risk alerts.

## 4.2. Description of the Worked Example

This section briefly describes the system that will be used as example on applying the methodology, which consists on a temperature control system used in a Data Centre. The detailed description of the system and its risk landscape can be found in Appendix A. Although these descriptions do not correspond to a specific data centre they are based on information collected from the industry, including an interview with a control engineer that works implementing and maintaining HVAC (Heath Ventilation, and Air Conditioning) systems and BMS (Building Management Systems) in several data centres [102]. This information was also complemented with the previous industrial experience of the author. The plausibility of this setting, as well as of the attack scenarios

was validated with the same engineer who has wide experience in configuration, installation, and maintenance of similar systems, as well as by the administrator of a data centre in South America.

To avoid the environmental conditions being outside operating limits, data centres use an environmental control system that activates heating and cooling systems automatically based on information extracted by environmental sensors. Figure 4.10 shows a diagram of the temperature control system. The heating and cooling system controllers (labelled DDC in the figure) communicate with a Building Management System (BMS) that runs in an application server, a computer located in an on-site workstation is in charge of running the control and monitoring software interface. The BMS sends alerts via email and SMS messages when an event requires attention.

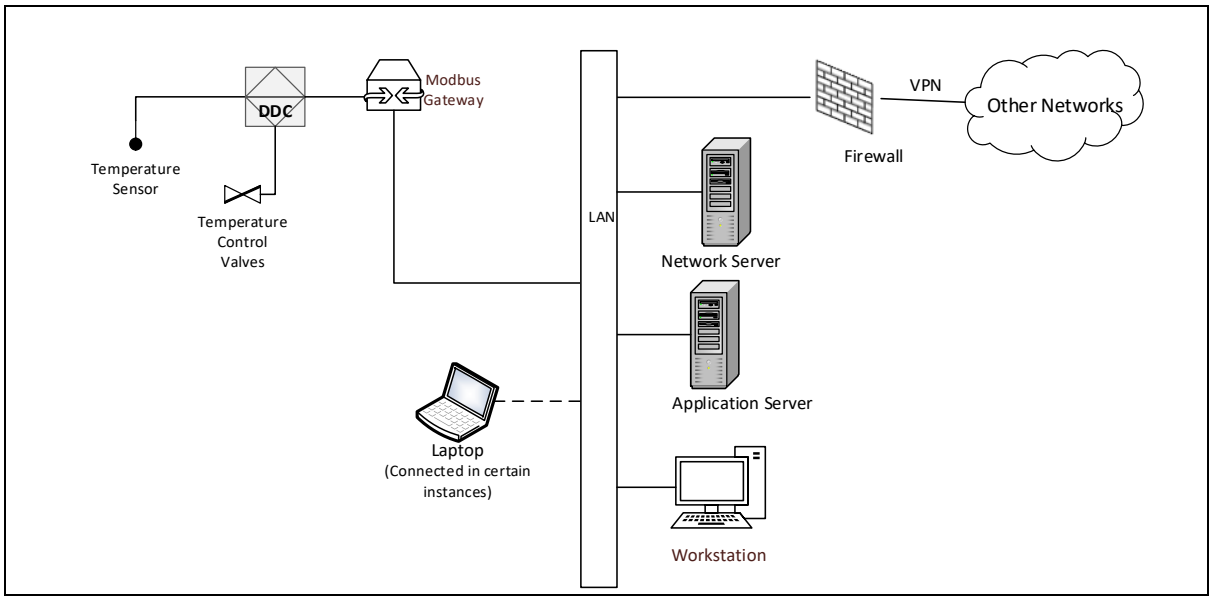



*Figure 4.10: Temperature control system*

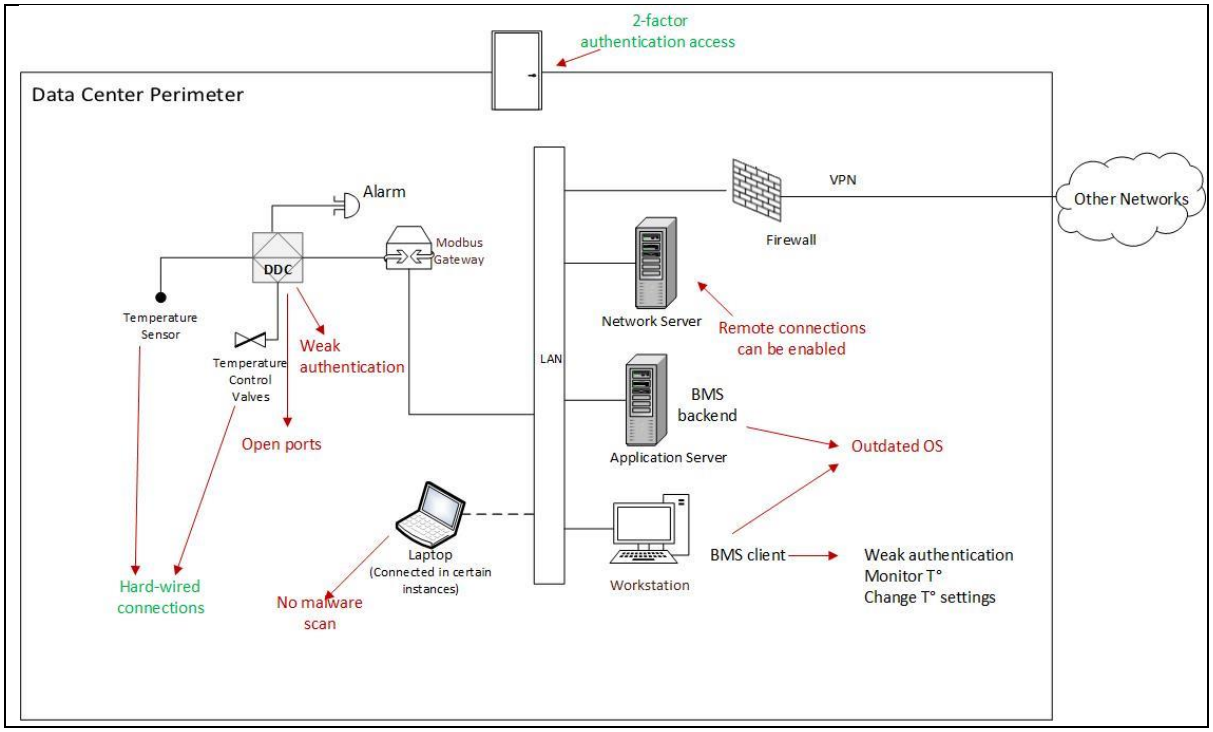



*Figure 4.11: risk landscape*

Figure 4.11 shows some of the main factors affecting security risk that were assumed to be present at the time of the baseline assessment. Some aspects of the system are highlighted in green, which represent conditions or measures that are favourable to prevent cyber-risks. For example, having hard-wired rather than wireless communications. Aspects of the system highlighted in red refer to conditions that can be associated with security vulnerabilities. For example, having an outdated OS, which can have known vulnerabilities. The purpose of this is to give a general idea of the security landscape of such a system and to use some of these aspects to develop a mock risk analysis that would be used in the demonstration of the methodology.

## 4.3. Applying the Methodology to the Worked Example

This section aims to illustrate the Continuous Risk Assessment Methodology using the worked example from Section 4.2. Each sub-section is based on the sequence of processes and activities described in each of the corresponding sub-sections of Section 4.1, starting with the Baseline Risk Management phase in Section 4.3.1, the Transition phase in Section 4.3.2, and the Continuous Risk Assessment phase in Section 4.3.3.

### 4.3.1. Baseline Risk Management phase

The following (sub)sub-sections illustrate the different processes involved in the Baseline Risk Management phase, as described in section 4.1.1.

#### 4.3.1.1. Context Establishment

One part of the context establishment process, specified in sub-section 4.1.1.1 is to define the scope of the risk assessment. This includes identifying the assets which need to be secured and the types of risks that will be considered. The other part of this process is to define the risk analysis techniques and methods that will be used and the risk acceptance criteria. Context Establishment is crucial to the analysis, evaluation and prioritisation of risks. For the Continuous Risk Assessment methodology possible sources of data for continuous risk monitoring also need to be identified, including those which are already available and additional ones that could be implemented. For the present use case, the main focus of the risk analysis is the Data Centre Infrastructure Management which is implemented through the BMS. The main assets are the servers that are housed in the data centre, since the main role of infrastructure management is to keep them safe and secure.

**Scope of the risk analysis**

The scope of the risk analysis is limited to cyber-attacks that target the temperature control. As shown in Figure 4.10, this is linked to data centre operations and to other business processes of the data centre. The assets that will be considered during threat and vulnerability analysis are those that belong to the temperature control system. For the purposes of impact analysis, the servers housed in the Data Centre and the processes that they run are also in scope, since impacts should also be evaluated at a business level. Figure 4.12 represents the scope of the risk analysis.

Other control systems in the data centre such as CCTV, fire alarm system, electrical supply and UPS, as well as servers and network equipment are considered as sources of data for risk monitoring purposes. The dependencies and correlations between the temperature control and other systems need to be established in order to identify possible IoRs and specify them so they can be included in the risk monitoring.

**Risk analysis method**

The risk analysis is based on threat models using attack trees in which probabilities of occurrence were estimated based on the combination of TTPs that would allow an adversary to gain different levels of access. The impact analysis can be done for each TTP considering that each successful execution of a malicious action can have a cost for the organization, which can be related to the execution of defence and recovery actions. However, for simplicity, in this example, only business risks will be assessed considering impacts and for technical risks (e.g. successful TTP execution), only the probabilities will be calculated.

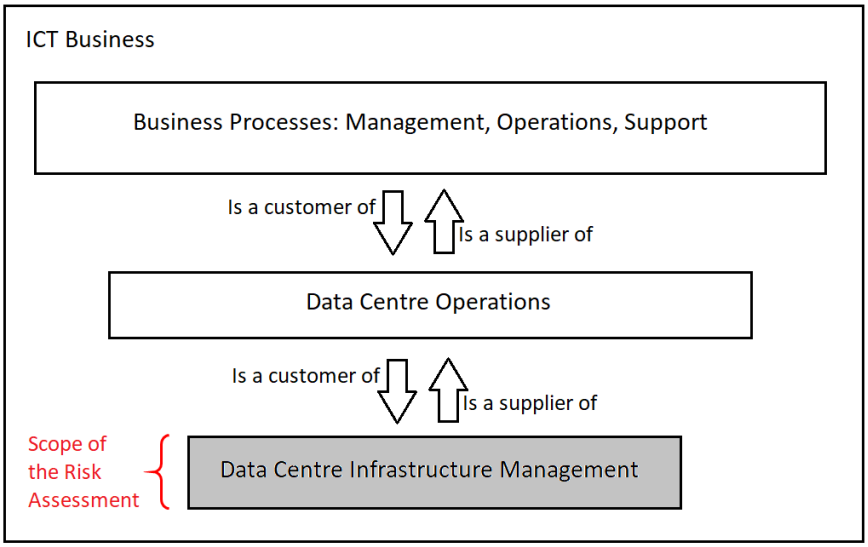



*Figure 4.12: Organisational Context Establishment*

For the threat analysis, the MITRE ATT&CK and ICS ATT&CK knowledge bases are consulted for identification of possible TTPs to be executed. For each goal of the attack trees, the combination of TTPs with the greatest chances of success (in other words, the highest threat) will be useful for quantifying the threat level.

The vulnerability analysis will be based on the security assessment done to the data centre in which main vulnerabilities of the system were identified. Each TTP identified in the threat analysis will be matched with the vulnerabilities that can facilitate the successful execution of an attack TTP. To calculate the vulnerability level the combination of those vulnerabilities that would be the easiest to exploit by an attacker (in other words, the weakest points) will be considered for the vulnerability level quantification.

For impact quantification, the variables considered are the costs of defence and recovery actions plus the costs of the business impacts. For example, cost per hour of incident response and forensics investigation for the costs related to cyber-security operations and cost per hour of downtime in DC servers for business costs. This can include variables such as revenue losses, penalties, and customers' compensations, cost of equipment damage or reduction of lifetime, cost of brand reputation damages and customer's losses, and the cost of a major disruption in the Data Centre. Costs can be expressed as ranges as a mean to acknowledge uncertainty about the real impact. As the work developed in this thesis will mostly focus on probability calculations, specifically the ones done during the Continuous Risk Assessment, the impact estimations of this example will be kept simple and will serve just illustrative purposes.

The baseline risk analysis will be based on scoring of threats, vulnerabilities to obtain an estimated probability for the achievement of different stages of the attack. Once the whole Baseline Risk Management process is performed including the risk treatment, the calculation of residual risks should follow this same method and the probabilities associated to the baseline risk scores will be used as reference to validate the BN conditional probabilities when it is initially built and no IoRs are observed yet. The probability calculations will be done by combining the vulnerability and threat levels. More details of the method used for this worked example can be found in Appendix A. It must be noted that other alternative methods can be developed to use with the Continuous Risk Assessment methodology for which the method presented in Appendix B is just an example and not a central part of this thesis work.

**Sources of data for the continuous risk analysis**

Sources of information that allow the observation of IoRs during continuous risk analysis are identified, such as updates on vulnerability databases which affect assets from the system, changes in equipment configuration, changes in equipment security policies and controls, and changes in network configurations and settings.

Available sources of data to update threat information are BMS access and event logs, workstation and servers event logs, firewall logs and network traffic information, temperature data, humidity data, electrical consumption data, performance of the servers in the data centre, safety alerts, perimeter security alerts, and threat intelligence sources.

**Risk acceptance criteria**

The risk acceptance criteria are established considering the risk appetite of data centre management regarding cyber-risks, as well as on the feasibility of applying mitigations that do not compromise other business objectives. Figure 4.13 shows a decision rules diagram with the risk acceptance criteria, which includes the following rules:

- ✓ Only low risks can be accepted without been escalated.
- ✓ Risks whose worst-case scenario impact is significant or higher shall be escalated, even if they are low.
- ✓ All identified risks should be monitored independent of their value.

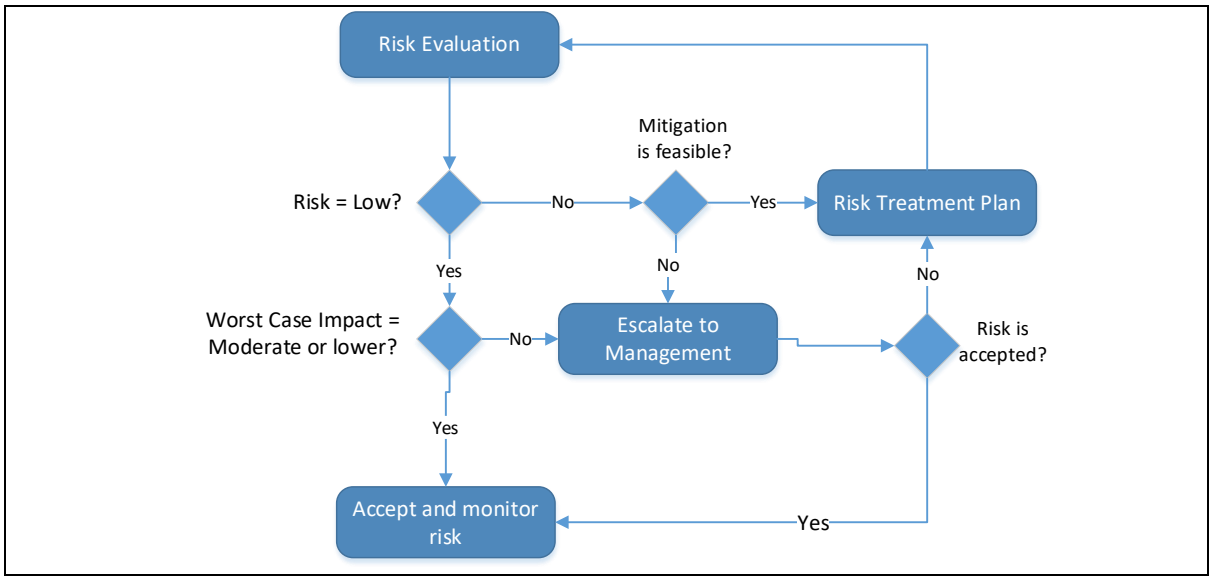



*Figure 4.13: Risk Acceptance decision diagram*

During the risk evaluation, risks that do not fit the mentioned criteria are reviewed and included in the risk treatment plan. Decisions of retaining risks outside limits need to be formally approved by senior management.

**Relevant stakeholders**

To describe how different actors are involved in the Continuous Risk Assessment, a responsibility matrix has been created, also known as a RASI matrix, which is shown in Table 4.2.

*Table 4.2: RASI matrix*

| Process / Role | Risk Analyst | SOC Analysis | CISO | Process Owner | Management |
|---|---|---|---|---|---|
| **Security monitoring** | S | R | A | I | I |
| **Risk monitoring** | R | S | I | A | I |
| **Response to security alert** | S | R | A | I | I |
| **Response to risk alert** | R | S | I | A | I |
| **Risk treatment** | S | S | I | R | A |

For each activity shown in the first column the responsibilities of each role are defined according to the following:

- Responsible (R): refers to the role that executed the task or activity
- Accountable (A): refers to the role that is accountable for the results of the task or activity
- Support (S): refers to the role or roles that provide information, resources or generate actions to help with the task or activity.
- Informed (I): refers to the role or roles that should be

#### 4.3.1.2. Baseline Risk Assessment

Following context establishment, the Baseline Risk Assessment is performed, which consists of Risk Identification, Risk Analysis, and Risk Evaluation.

#### 4.3.1.3. Baseline Risk Assessment: Risk Identification

Risk identification consists of identification of the assets that need to be protected, and associated threats, vulnerabilities and security controls, and potential impacts.

**Asset Identification**

According to the scope defined for the risk assessment, the main assets identified for the risk analysis are the data centre servers, these being the most critical asset for business operations. Additionally, the temperature control system, as a whole, was consider to be an asset since a compromise of its integrity, availability and, to a minor extent, confidentiality, can have an effect on the data centre's operations. Table 4.3 summarises the asset identification.

*Table 4.3: Asset identification*

| Asset Group ID | Asset Group Name | Asset Owner | Asset description |
|---|---|---|---|
| A001 | Data Centre Servers | DC Operations Manager | Main business assets.<br>Availability, and integrity are highly critical. |
| A002 | Processes running in Data Centre Servers | Business Process Owner | Main business assets.<br>Availability, and integrity are highly critical. |
| A003 | Data stored in Data Centre Servers | Business Process Owner | Main business assets.<br>Confidentiality, availability, and integrity are highly critical. |
| A004 | Temperature Control | BMS Manager | Integrity and availability are critical to protect the main business assets. |

**Threat Identification**

The threat identification was performed considering two possible adversarial goals. The first one was the disablement of the temperature control, and the second, tampering with the temperature control so as to raise the temperature to the highest possible value. The motivations that could drive threat agents to attempt one these two types of attacks were identified. After reviewing the possible threat population, it was considered that the threat agents with the most possibilities of perpetrating a successful attack were internal personnel, including third party contractors who have means of physical access and valid credentials to access the BMS and the VPN. Former employees were also considered to have enough knowledge to gain remote access through the VPN. Finally, there is also a chance for external parties to introduce malware by using techniques such as spear phishing or by brute forcing the VPN. An attack tree was built for each attack goal to identify how assets might be attacked linking this to TTPs from the ICS ATT&CK knowledge base see Figure 4.15 and 4.16.

For both goals, disabling the temperature control, and increasing the temperature, the following three potential motivations were identified:

**Distraction**

Disturbing the normal operation of the temperature control system can be a distraction to keep Data Centre operators focused on fixing this issue while another type of threat is developing. Increasing the temperature to a degree that is too high for humans to bear and can potentially affect a server's performance, as well making other hazards such as a fire more likely. It can be a greater distraction for operators than just disabling the temperature control.

**Sabotage**

Disturbing the normal operation of the temperature control system can be used to increase operational costs. Consider a scenario in which a data centre service provider's contract with an important customer is up for renewal. An employee could be bribed by a competitor that wants to cause customer to lose trust in the current contractor in order to win the contract. An attacker might also attempt to increase the temperature to a level that servers get overheated and performance of the services provided by the data centre get affected. This, can as well as damage or reduce the lifespan of the servers and/or other electronic equipment.

**Intermediate goal to attack other system**

Targeting the temperature control can be just an intermediate step in the development of a more complex attack, rather than a final goal. For example, targeting the fire alarms or other safety systems, security cameras or electrical power control. Alternatively, it can also be used as a proof of concept to use an analogue technique to disable these systems in the future. This is different to the case where the goal is to create a distraction, in which case the attacker wants the effect of the attack to be noticed, whereas in this case it is likely that the attacker might prefer to hide their actions as much as possible.

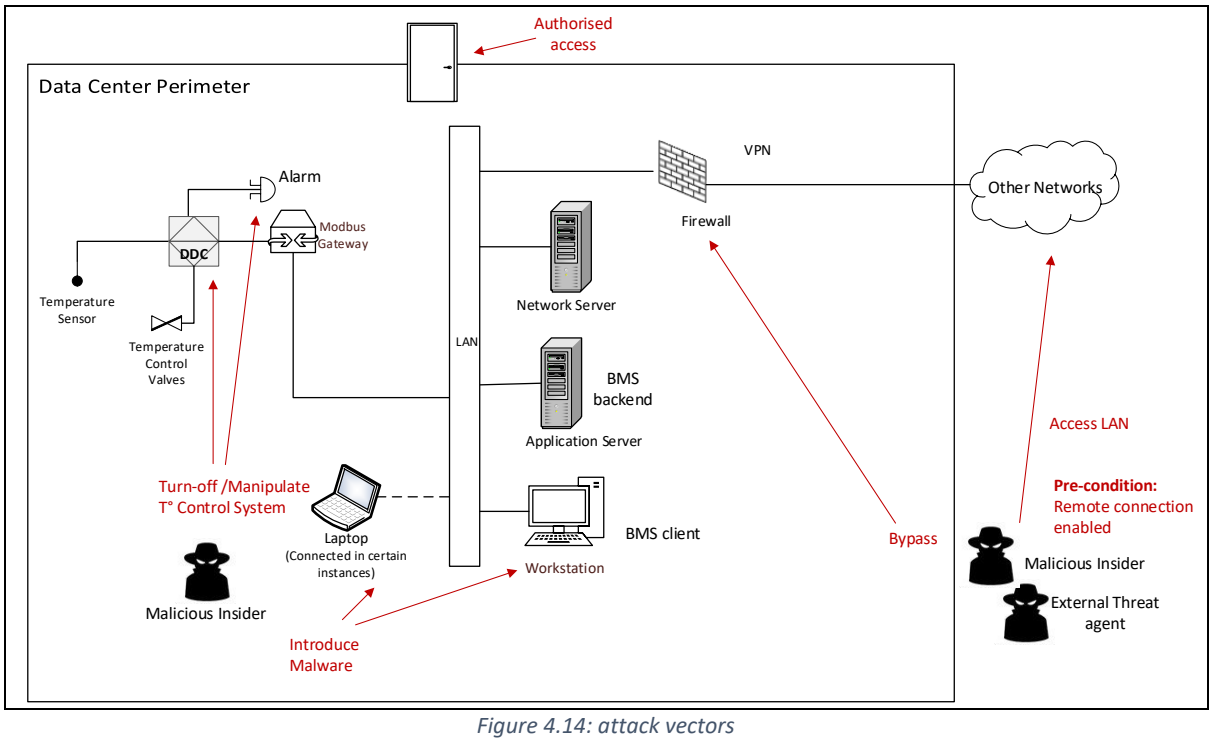



*Figure 4.14: attack vectors*

It was considered than an attack could be either conducted from inside the data centre perimeter, or using a remote connection during the periods in which it is enabled. In either case, the attacker would be able to either turn-off or manipulate the temperature control system. However, these actions would be eventually detected by the data centre operators either through an alarm been triggered or due to noticing the anomaly during

normal monitoring activities. Therefore, for an attack to be successful, the attacker should, as well, hide or disguise their action. Figure 4.14 shows the attack vectors recognised in the threat identification process.

**Vulnerability Identification**

At this stage information about vulnerabilities is gathered from the system as-is. Vulnerability identification was performed by creating a list of all the vulnerabilities found in the system and subsequently finding which of them could allow each one of the techniques to be executed. Vulnerabilities can be found by looking into specific system components (e.g. hardware, OS, software) in vulnerability databases, such as the NVD (National Vulnerability Database).

The following is a list of vulnerabilities in the example scenario:

**Vulnerabilities related to cyber-security management:**

- Lack of cyber-security policies specifically applied to the BMS
- No clear segregation of duties and permissions for different users
- It is frequent for users to share credentials
- Lack of policies for connecting mobile devices
- Lack of policies for remote connection
- No asset management practices
- Lack of mechanisms to control compliance of security policies (if they existed)
- Insufficient policies and control of equipment allowed in secure areas
- Logs for changes and updates are not registered
- Security policies regarding relationships with suppliers limited to signing a NDA
- Insufficient policies and control of work procedures in secure areas

**Vulnerabilities related to people**

- Lack of a defined role and assignation of responsibility for cyber-security in the BMS
- Lack of cyber-security awareness and training to personnel
- Excessive trust on insiders, including contractors

**Technical and implementation related vulnerabilities**

- Weak authentication to access applications
- Weak authentication to access devices
- Weak access control to program source code
- Passwords stored in plain text
- Lack of network monitoring
- Insufficient malware protection
- Open ports
- The same network server will manage other networks giving the possibility to jump from one network to another
- Outdated OS with known vulnerabilities
- Field devices with known vulnerabilities
- No control regarding enablement of remote connection
- No application blacklisting/whitelisting

The following security controls that could contribute to the mitigation of the overall vulnerability level were identified:

- Physical access control
- Logical and physical separation of LAN
- Access logs are kept

- System backups are done periodically

**Impact Identification**

Impact identification was performed by selecting the applicable Techniques listed in the ICS ATT&CK matrix under the Impact Tactic, and interpreting relevant ones in the context of this particular case to obtain the following specific impacts:

- **Impacts of disabling the temperature control**

1. T0831-Manipulation of Control

Specific Impacts:

- Operational costs of investigating and fixing the problem
- Risk of this issue deriving in other incidents

2. T0828-Loss of Productivity and Revenue

Specific Impacts:

- Operational costs of investigating and fixing the problem

3. T0826-Loss of Availability

Specific Impacts:

- The temperature control is no longer available

- **Impacts of increasing the temperature to the highest possible value**

1. T0831-Manipulation of Control

Specific Impacts:

- Operational costs of investigating and fixing the problem
- Risk of this issue deriving in other incidents

2. T0828-Loss of productivity and revenue

Specific Impacts:

- Reduction in server performance in data centre (service availability is compromised affecting business processes)
- Reduction of the lifespan of servers and other IT equipment

3. T0826-Loss of Availability

Specific Impacts:

- The temperature control is no longer available

4. T0879-Damage to Property

Specific Impacts:

- Possible damage to equipment

#### 4.3.1.4. Baseline Risk Assessment: Risk Analysis

To quantify risks in this example attack trees are used as a start point to then associate one or more TTPs (in this case mostly Techniques) to each action identified in the attack tree. This is followed by the likelihood estimation for each Technique. The method used for risk quantification, which is detailed in Appendix B, was based on threat and vulnerability scoring systems, which are combined to define likelihood scores. To define Likelihood scores for particular risks, the highest likelihood from alternative Techniques within a Tactic is chosen. When Techniques are complementary or depend on each other, then the lowest Likelihood within that set of Techniques is chosen. Risk scores are obtained by combining the likelihood and impact scores for each risk. Figure 4.15 shows the attack tree for disabling the temperature control and Figure 4.16 shows the attack tree for changing temperature settings, based on the threat identification process.

Table 4.4 shows a list of the Tactics from the ATT&CK and the ICS ATT&CK knowledge bases which can be related to each step on the attack tree and the possible Techniques that can be used by the attackers to perform an

attack strategy for turning off the temperature control or increasing the temperature levels. It can be observed that not every node in the attack trees corresponds exactly to the steps shown in Table 4.4, since this table translates the threat identification to Tactics and Techniques by which some nodes of the tree are grouped in the different steps.

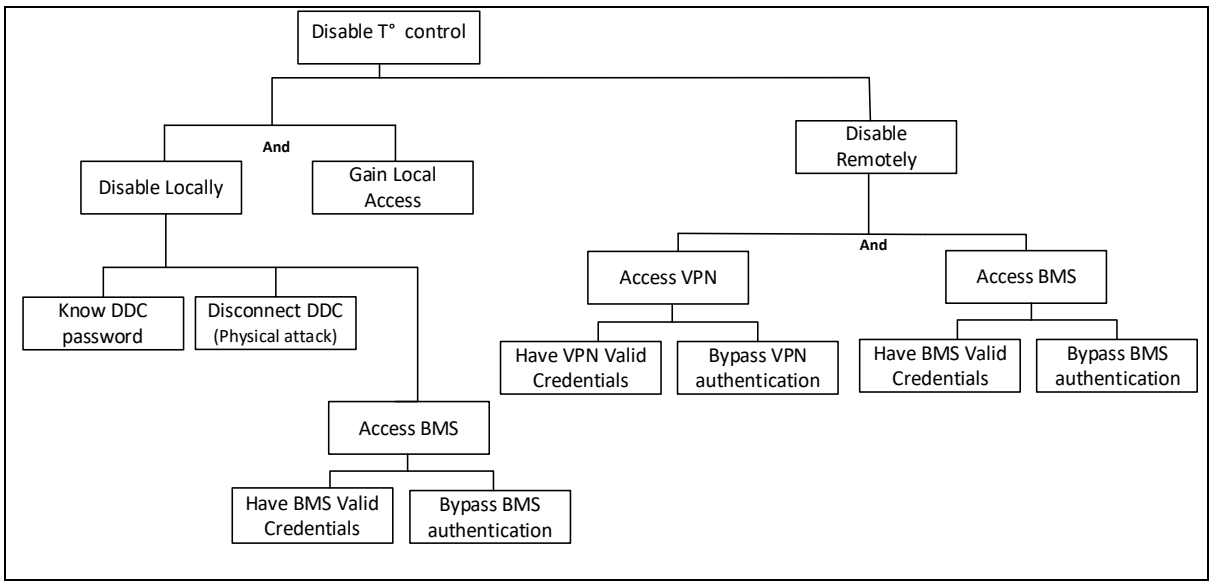



*Figure 4.15: Attack tree for disabling temperature control*

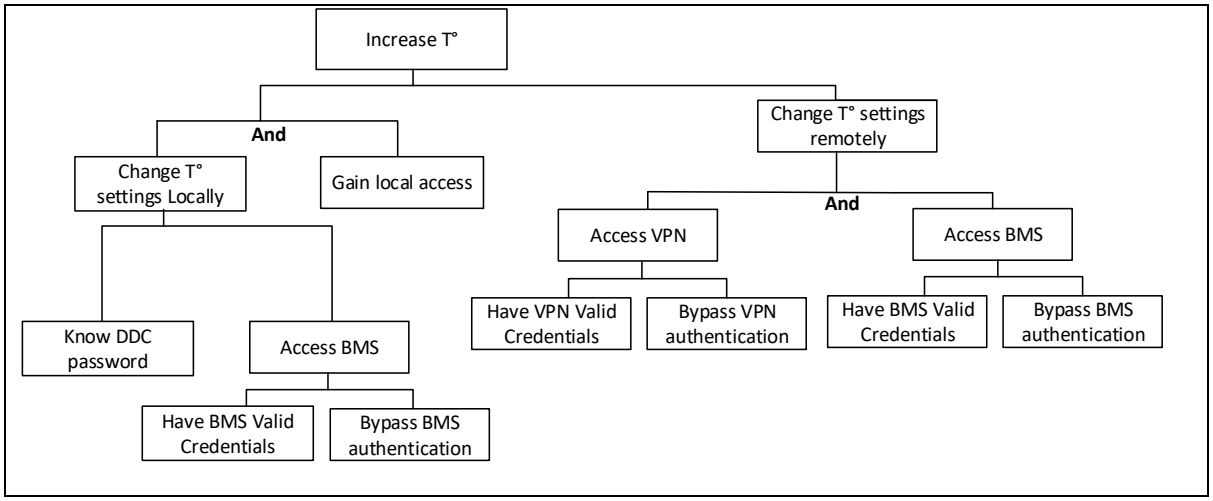



*Figure 4.16: Attack tree for increasing temperature*

Figure 4.17 provides a graphical representation of the threat model developed during the Initial Risk Assessment, which derives from the attack trees and includes Techniques, Tactics, and threat events defined specifically for modelling purpose. It can be observed that it is possible to have Tactic to Technique relationships defined, such as in the case of the "Initial Access" Tactic, which precedes or is "parent of" several Techniques and also Technique to Technique relationships. This will depend on whether a Technique depends on another specific Technique to be executed as pre-requisite or if it depends just on achieving the more general goal, which is the Tactic. A typical example of the former will be the "Initial Access" Tactic, since in many cases, the adversary needs to first access the system before attempting to perform any other Technique and it might not matter how they gained access. On the other hand, for other Techniques the pre-conditions that shall be met might depend on the previous success of other Techniques which are more specific. For example, not any Technique can be related to the "Loss of Availability" Technique but only specific ones, such as "Denial of Service".

*Table 4.4: Tactics and techniques related to the attack tree*

| Steps based on the attack tree | Tactics | Techniques |
|---|---|---|
| Access VPN with valid credentials | Credential Access<br>Initial Access | T1078 - Valid Account (VPN)<br>T0865 - Spearphishing Attachment<br>T1110 - Brute force attack<br>T0822-External Remote Services |
| Access BMS with valid credentials | Credential Access<br>Initial Access | T1078 - Valid Account (BMS)<br>T0865 - Spearphishing Attachment<br>T1110 - Brute force attack |
| Bypass BMS authentication | Initial Access | T0847-Replication through removable media (local) |
| Gain local Access | Initial Access | T1078 - Valid Account (credentials for physical access) |
| Turn off temperature control in DDC | Execution<br>Inhibit Response Function<br>Impair Process Control | T0875-Change Program State<br>T0814-Denial of Service<br>T0816-Device Restart/Shutdown<br>T0833-Modify Control Logic |
| Turn off temperature control through BMS | Execution<br>Inhibit Response Function<br>Impair Process Control | T0823-Graphical User Interface<br>T0875-Change Program State<br>T0814-Denial of Service<br>T0816-Device Restart/Shutdown<br>T0833-Modify Control Logic |
| Turn off temperature control | Impact | T0831-Manipulation of Control<br>T0828-Loss of Productivity and Revenue<br>T0826-Loss of Availability |
| Increase temperature settings in DDC (local) | Execution<br>Inhibit Response Function<br>Impair Process Control | T0875-Change Program State<br>T0833-Modify Control Logic<br>T0838-Modify Alarm settings |
| Increase temperature settings in BMS | Execution<br>Inhibit Response Function<br>Impair Process Control | T0823-Graphical User Interface<br>T0875-Change Program State<br>T0833-Modify Control Logic<br>T0835-Manipulate I/O image<br>T0838-Modify Alarm settings |
| Increase Temperature | Impact | T0831-Manipulation of Control<br>T0828-Loss of Productivity and Revenue<br>T0879-Damage to Property<br>T0826-Loss of Availability |

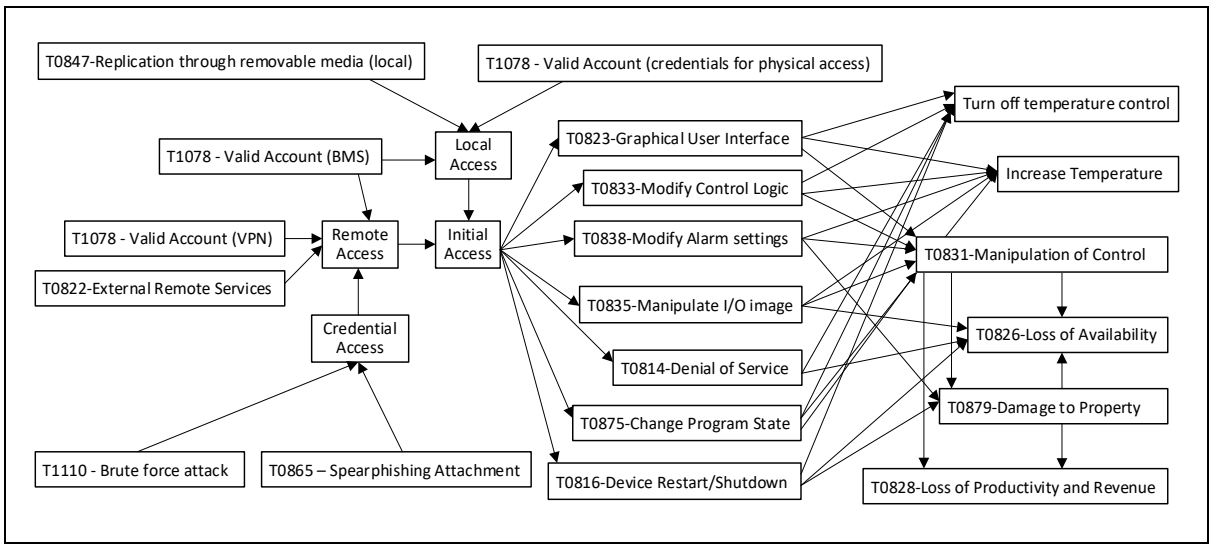



*Figure 4.17: Links between events and TTPs*

Table 4.5 shows the threat and vulnerability scores and resulting likelihood associated with each pre-cursor technique. An “attack path" corresponds to the possible combinations of techniques that will allow an attacker to achieve their final goal. The risk quantification method used to obtain a consolidated score that considers all attack paths is summarised briefly in this chapter, and described in more detail in Appendix B. The total risk

score can be expressed quantitatively as the probability of the impact costs lying within a certain numerical range or qualitative labelled as “low”, “moderate”, “high”, or “critical”. It is assumed that only the Techniques belonging to the Impact Tactic have a direct impact in the industrial process, and that their likelihoods are determined by the likelihoods of the techniques occurring earlier in the ‘kill chain’ that enable them.

*Table 4.5: Likelihood estimation for Techniques*

| Threat | Threat score | Vulnerability score | Likelihood |
|---|---|---|---|
| T1078 - Valid Account (VPN) | Moderate | Medium | Moderate |
| T0865 – Spearphishing Attachment | Moderate | High | Moderate |
| T1110 - Brute force attack | Low | Medium | Low |
| T0822-External Remote Services | Moderate | Medium | Moderate |
| T1078 - Valid Account (BMS) | Moderate | Medium | Moderate |
| T0847-Replication through removable media (local) | Moderate | Medium | Moderate |
| T1078 - Valid Account (credentials for physical access) | Moderate | Medium | Moderate |
| T0875-Change Program State | Low | Critical | Moderate |
| T0814-Denial of Service | Moderate | Medium | Moderate |
| T0816-Device Restart/Shutdown | Moderate | Medium | Moderate |
| T0833-Modify Control Logic | Moderate | Critical | High |
| T0823-Graphical User Interface | Moderate | Medium | Moderate |
| T0838-Modify Alarm settings | Low | Medium | Low |
| T0835-Manipulate I/O image | Low | Medium | Low |

Table 4.6 shows the total likelihood for each of the Impact Techniques, which is obtained by aggregating the likelihoods of Techniques by Tactic and then overall. The likelihood score for a Tactic is obtained as the maximum likelihood value of the Techniques that belong to that Tactic. This is because the Techniques belonging to the same Tactic are seen as alternatives. The total likelihood for the Impact Technique is then the minimum likelihood value of the contributing Tactics. The rationale for this is that all the Tactics contributing to an Impact Technique need to succeed for the Impact Technique to succeed. For simplicity, the Execution, Inhibit Response Function, and Impair Process Control tactics were grouped as one.

*Table 4.6: Likelihood estimation for Impact Techniques*

| Impact Techniques | Highest Likelihood TTPs | Likelihood | Total Likelihood |
|---|---|---|---|
| T0831-Manipulation of Control | Initial Access: T1078, T822, T847 | Moderate | Moderate |
| | Execution, Inhibit Response Function, Impair Process Control: T0833 | Moderate | |
| T0879-Damage to Property | Initial Access: T1078, T822, T847 | Moderate | Moderate |
| | Execution, Inhibit Response Function, Impair Process Control: T0816 | Moderate | |
| | Impact: T0831 | Moderate | |
| T0826-Loss of Availability | Initial Access: T1078, T822, T847 | Moderate | Moderate |
| | Execution, Inhibit Response Function, Impair Process Control: T0814, T0816 | Moderate | |
| | Impact: T0831, T0879 | Moderate | |
| T0828-Loss of Productivity and Revenue | Initial Access: T1078, T822, T847 | Moderate | Moderate |
| | Execution, Inhibit Response Function, Impair Process Control: T0816 | Moderate | |
| | Impact: T0831, T0879, T0826 | Moderate | |

Finally, the impact Techniques were consolidated to estimate the likelihood of each one of the two high-level risks, as shown in Table 4.7, to be assessed and further evaluated considering both likelihood and impact. This was done by selecting the highest likelihood among all the associated Impact Techniques. As in this case, all the likelihoods were “moderate”, the consolidated result for likelihood was the same.

*Table 4.7: Risk analysis results*

| Risks | Impact Techniques | Likelihood | Impact | Total Risk |
|---|---|---|---|---|
| Turn off temperature control | T0831-Manipulation of Control<br>T0828-Loss of Productivity and Revenue<br>T0826-Loss of Availability | Moderate | Very Low | Moderate |
| Increase Temperature | T0831-Manipulation of Control<br>T0828-Loss of Productivity and Revenue<br>T0879-Damage to Property<br>T0826-Loss of Availability | Moderate | Low | Moderate |

#### 4.3.1.5. Baseline risk assessment: Risk Evaluation

In this example, the risk analysis results show both risks above the acceptance level, since only low risks can be accepted without being escalated. The vulnerability analysis shows that while most of the vulnerabilities present a medium level and some of the vulnerabilities identified in Table 4.5 are high and critical.

#### 4.3.1.6. Risk Treatment

Mitigation measures need to be applied in order to reduce the risks through a risk treatment plan, which is elaborated based on the risk assessment. The following security controls are part of this risk treatment plan:

**Controls related to cyber-security management**

- Cyber-security policies specifically applied to the BMS were defined
- Each individual user is given unique access credentials
- Mobile devices, laptops and storage devices should be authorised and checked for malware before connecting to any system
- Remote connections are disabled by default and if they need to be temporarily enabled for specific tasks this should be registered
- Logs for changes and updates are maintained and monitored
- Policies of work procedures in secure areas were established
- Responsibility for cyber-security in the BMS is allocated in a specialised SOC
- The Continuous Risk Assessment method is implemented

**Controls related to people**

- Personnel are trained to increase cyber-security awareness

**Technical and implementation related controls**

- Workstation and BMS passwords are stored securely
- Passive network monitoring is implemented
- Log monitoring is implemented
- Malware scans for devices that are connected to the network, periodic scans to workstations
- Unused ports are disabled by default

#### 4.3.1.7. Definition of the Baseline Risk Scores

Following the implementation of the risk treatment plan, risks need to be reassessed by repeating the analysis and evaluation using the same method as previously. The baseline risk scores will correspond to all the residual risks, which should be monitored in the Continuous Risk Assessment phase. A residual risk is understood as the

risk remaining after risk treatment. There is an expectation for the residual risks to be under the acceptance level, however this cannot be confirmed until the reassessment is completed.

Table 4.8 shows an update of Table 4.5 with a summary of threat and vulnerability and likelihood scores, after the risk treatment plan is completed, more details can be found in Appendix B. This example illustrates that security controls can help reducing both the threat and the vulnerability scores. This is not surprising since these two variables are usually correlated, because a weak system is also a more attractive target. For example, network and log monitoring help to increase the detectability of TTPs, and a strict policy of using unique accounts instead of shared ones, increases chances of attribution of insider threats. At the same time, more skills and resources will be required for a successful attack, reducing threat levels by means of the same controls.

*Table 4.8: Likelihood estimation for Techniques after Risk Treatment*

| Threat | Threat score | Vulnerability score | Likelihood |
|---|---|---|---|
| T1078 - Valid Account (VPN) | Low | Low | Low |
| T0865 – Spearphishing Attachment | Low | Medium | Low |
| T1110 - Brute force attack | Low | Medium | Low |
| T0822-External Remote Services | Low | Low | Low |
| T1078 - Valid Account (BMS) | Low | Medium | Low |
| T0847-Replication through removable media (local) | Low | Low | Low |
| T1078 - Valid Account (credentials for physical access) | Moderate | Medium | Moderate |
| T0875-Change Program State | Very Low | Low | Very Low |
| T0814-Denial of Service | Low | Medium | Low |
| T0816-Device Restart/Shutdown | Low | Medium | Low |
| T0833-Modify Control Logic | Very Low | Low | Very Low |
| T0823-Graphical User Interface | Low | Low | Low |
| T0838-Modify Alarm settings | Low | Medium | Low |
| T0835-Manipulate I/O image | Low | Medium | Low |

Table 4.9 updates Table 4.6, showing the effect of the additional controls on the likelihood calculation for the Impact Techniques.

*Table 4.9: Likelihood estimation for Impact Techniques after Risk Treatment*

| Impact Techniques | Highest Likelihood TTPs | Likelihood | Total Likelihood |
|---|---|---|---|
| T0831-Manipulation of Control | Initial Access: T1078, T822, T847 | Low | Very Low |
| | Execution, Inhibit Response Function, Impair Process Control: T0833 | Very Low | |
| T0879-Damage to Property | Initial Access: T1078, T822, T847 | Low | Very Low |
| | Execution, Inhibit Response Function, Impair Process Control: T0816 | Low | |
| | Impact: T0831 | Very Low | |
| T0826-Loss of Availability | Initial Access: T1078, T822, T847 | Low | Very Low |
| | Execution, Inhibit Response Function, Impair Process Control: T0814, T0816 | Low | |
| | Impact: T0831, T0879 | Very Low | |
| T0828-Loss of Productivity and Revenue | Initial Access: T1078, T822, T847 | Low | Very Low |
| | Execution, Inhibit Response Function, Impair Process Control: T0816 | Low | |
| | Impact: T0831, T0879, T0826 | Very Low | |

At this point, as shown in Table 4.10, all the risk scores of this example are below the acceptance level, and we can move on to the Transition phase.

*Table 4.10: Baseline Risk Scores*

| Risks | Impact Techniques | Likelihood | Impact | Total Risk |
|---|---|---|---|---|
| Turn off temperature control | T0831-Manipulation of Control<br>T0828-Loss of Productivity and Revenue<br>T0826-Loss of Availability | Very Low | Very Low | Low |
| Increase Temperature | T0831-Manipulation of Control<br>T0828-Loss of Productivity and Revenue<br>T0879-Damage to Property<br>T0826-Loss of Availability | Very Low | Low | Low |

### 4.3.2. Transition phase

As specified in section 4.1, after the baseline risk scores are defined, the Transition phase is conducted. This starts with the definition of IoRs, followed by the definition of the Bayesian Network (BN), the implementation of the tools and processes for Continuous Risk Assessment and delivering the corresponding training. For this worked example, the focus will be only on IoR identification and on defining and building the BN, leaving the Implementation and Training part out of scope.

#### 4.3.2.1. Definition of IoRs

In this process, the IoRs that can be associated with each Technique identified in the threat model developed in the Baseline Risk Analysis (section 4.3.1.4) are defined. As explained in section 4.1.1.4, the identification and definition of IoRs is assisted by the use of a knowledge base named as the "IoR Library", which helps identifying IoRs that can be related to each Technique from the ICS ATT&CK. The structure of this knowledge base and the method for using it at this stage of the Continuous Risk Assessment Methodology are explained in more detail in Chapter 5. Table 4.11 shows the IoRs identified for the Credential Access and Initial Access Tactics. It can be observed that IoR names are prefixed by an identification code or ID, which is part of a naming structure used in the "IoR Library".

*Table 4.11: Identification of IoRs for Credential Access and Initial Access*

| Technique | IoRs |
|---|---|
| T1110 - Brute force attack | IoR001-Remote access enabled<br>IoR102-Failed login attempts<br>IoR103-Failed login attempts followed by a successful one |
| T0865 - Spear phishing Attachment | IoR109-Unknown files<br>IoR116-E-mail server suspicious activities |
| T1078 - Valid Account (VPN) | IoR001-Remote accesses enabled<br>IoR201-VPN suspicious Access log |
| T1078 - Valid Account (BMS) | IoR101-Suspicious BMS user behaviour<br>IoR105-BMS application suspicious access log |
| T0847-Replication through removable media | IoR002-Unnecesary open ports<br>IoR107-Unknown programs<br>IoR109-Unknown files<br>IoR123-Unknown USB device plugged<br>IoR601-Suspicious Physical Access log |
| T0822-External Remote Services | IoR001-Remote accesses enabled<br>IoR201-VPN suspicious Access log<br>IoR202-High volume of network traffic<br>IoR203-Unusually large inbound/outbound packets |

Table 4.12 shows the IoRs that can be associated with each Technique identified in the threat model (section 4.3.1.4) for the Execution, Inhibit Response Function, and Impair Process Control Tactics. IoRs that have an ID starting in 5 in this example are based on physical variables whose values are provided as input to the system by sensors. For example, IoR501 and IoR502 relate to temperature data and IoR502 to electrical consumption data.

These type of data is not frequently used in cyber-security monitoring despite it can have a great potential on revealing anomalies in a cyber-physical system. Hence, using sensors data has been considered as an integral part of the approach proposed. Further illustration of these ideas can be found throughout Chapter 6.

*Table 4.12: Identification of IoRs for Execution, Inhibit Response Function, and Impair Process Control*

| Technique | IoRs |
|---|---|
| T0814-Denial of Service | IoR007-Inapropriate network segmentation<br>IoR107-Unknown programs<br>IoR112-User unable to access monitor and control application<br>IoR402-High volume of network traffic (OT Network)<br>IoR509- Sensor data unavailable |
| T0816-Device Restart/Shutdown | IoR006-User with unnecessary privileges<br>IoR105-BMS application suspicious access log<br>IoR106-BMS application change logs<br>IoR301-Controller suspicious access log<br>IoR405-Unresponded commands<br>IoR406-Unexpected command sequence over OT network |
| T0875-Change Program State | IoR006-User with unnecessary privileges<br>IoR106-BMS application change logs<br>IoR107-Unknown programs<br>IoR302-Controller program change logs<br>IoR501- Temperature out of limits<br>IoR502- Misbehaviour in temperature data<br>IoR504- Misbehaviour in electrical consumption |
| T0833-Modify control logic | IoR006-User with unnecessary privileges<br>IoR105-BMS application suspicious access log<br>IoR106-BMS application change logs<br>IoR107-Unknown programs<br>IoR501- Temperature out of limits<br>IoR502- Misbehaviour in temperature data<br>IoR504- Misbehaviour in electrical consumption |
| T0823-Graphical User Interface | IoR101-Suspicious BMS user behaviour<br>IoR105-BMS application suspicious access log<br>IoR106-BMS application change logs |
| T0838-Modify Alarm settings | IoR301-Controller suspicious access log<br>IoR302-Controller program change logs<br>IoR303-Controller settings change logs<br>IoR501- Temperature out of limits<br>IoR502- Misbehaviour in temperature data<br>IoR504- Misbehaviour in electrical consumption<br>IoR701-Alarm Historian anomaly |
| T0835-Manipulate I/O image | IoR301-Controller suspicious access log<br>IoR302-Controller program change logs<br>IoR303-Controller settings change logs<br>IoR504- Misbehaviour in electrical consumption |

Table 4.13 shows the IoRs that can be associated with each technique identified in the threat model (section 4.3.1.4) for the Execution, Inhibit Response Function, and Impair Process Control Tactics.

*Table 4.13: Identification of IoRs for the Impact Tactic*

| Technique | IoRs |
|---|---|
| T0831-Manipulation of Control | IoR301-Controller suspicious access log<br>IoR302-Controller program change logs<br>IoR303-Controller settings change logs<br>IoR406-Unexpected command sequence over network |
| T0879-Damage to Property | IoR501- Temperature out of limits<br>IoR502- Misbehaviour in temperature data<br>IoR504- Misbehaviour in electrical consumption |
| T0826-Loss of Availability | IoR509- Sensor data unavailable |
| T0828-Loss of Productivity and Revenue | NA |

#### 4.3.2.2. Definition of the Bayesian Network

As described in section 4.1.3 the Bayesian Network is built in three steps, which are: defining the nodes, establishing the relationships between nodes, and defining the conditional probabilities. For simplicity, in this exercise the nodes were defined as binary variables (“true" or “false"). However, it is possible to define several possible states for each node.

**Step 1: define the nodes of the Bayesian Network**
Each Technique and IoR shown in Tables 4.10, 4.11, and 4.12 and each corresponding Tactic, is represented by a node in the BN, as well as the specific risks. Other relevant variables, may be added as auxiliary nodes to provide view in the BN of events that might be of interest. In the example scenario the additional nodes are “Initial Access Remote” and “Initial Access Local”, which are linked to the Initial Access Tactic, and “Turn off temperature control”, and “Increase temperature settings”, which are linked to Techniques from the Impact Tactic.

**Step 2: Establish relationships between the different nodes.**
Relationships between nodes can be represented by incidence matrix or through the BN graphical representation. Tables 4.10, 4.11 and 4.12 show already the relationship between IoR nodes and Technique nodes. In an analogue way, Technique nodes should be linked to their corresponding Tactic nodes, and to other Techniques to which they might be related to throughout the attack kill-chain, and to any additional nodes added in the model. For example, Techniques belonging to the Initial Access Tactic will usually be a pre-requisite for other Techniques and therefore, the higher their likelihood of occurrence, the higher the likelihood of subsequent Techniques to succeed. In the incidence matrices, the rows represent the influenced nodes (generally a Technique or Tactic) and the columns the nodes that exert the influence which can be either IoRs, Techniques, or an auxiliary node. A dependency between nodes is denoted by a “1” in the corresponding cell, and cells that are empty mean that there is no dependency. In the graphical representation, arrows are used to show the dependencies. The direction of the arrow goes from the node that exerts an influence, for example, an IoR, to the one that is influenced, for example, a Technique, for which the estimated probability of successful execution is influenced by the observation of the IoR.

The full incidence matrix for this example is very large, but it can be built up from smaller incidence matrices, each describing a region of the BN. Table 4.14 shows the portion of the incidence matrix describing dependencies of the Credential Access Tactic node, which represents the proposition that a Credential Access attack step will be successful. This depends on the success of one of two Techniques (T1110 and T0865). There are three IoRs that provide evidence for the successful use of T1110, and two that do so for T0865. Note that here, the direction of the connections between nodes reflects the direction of evidential inferences, and not causation. The sub-matrix can be derived from Table 4.11 and the structure of the ATT&CK Matrix (T1110 and T0865 belong to the Credential Access Tactic).

*Table 4.14: Incidence Matrix for Credential Access*

| | IoR001 | IoR102 | IoR103 | IoR109 | IoR116 | T1110 | T0865 |
|---|---|---|---|---|---|---|---|
| T1110 | x | x | x | | | | |
| T0865 | | | | x | x | | |
| Credential Access | | | | | | x | x |

Similarly, Table 4.15 shows the incidence matrix for the Initial Access tactic. This sub-matrix can also mainly be derived from Table 4.11 and the structure of the ATT&CK Matrix, this time extended with the two auxiliary nodes, “Initial Access Remote” and “Initial Access Local”, which are effectively sub-Tactics of Initial Access.

As a general principle, Tactics depend on Sub-tactics (if they exist), which depend on Techniques belonging to the Sub-tactic, which depend on IoRs. Three further points to note here are that:

- There are two versions of T1078 because the technique can be applied to two types of account within the scenario (VPN and BMS),
- Initial Access Remote is shown as being dependent on Credential Access, which connects the two incidence matrices described so far. This makes sense, because some credentials obtained by an attacker using Credential Access Techniques can be used to gain Initial Access.
- T1078 (BMS) belongs to both the Remote and Local Sub-tactics of Initial Access.

*Table 4.15: Incidence Matrix for Initial Access*

| | IoR001 | IoR002 | IoR601 | IoR101 | IoR123 | IoR105 | IoR107 | IoR109 | IoR201 | IoR202 | IoR203 | Credential Access | T1078 (VPN) | T1078 (BMS) | T0847 | T0822 | Initial Access Remote | Initial Access Local |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T1078 (VPN) | x | | | | | | | | x | | | | | | | | | |
| T1078(BMS) | | | | x | | x | | | | | | | | | | | | |
| T0847 | | x | x | | x | | x | x | | | | | | | | | | |
| T0822 | x | | | | | | | | x | x | x | | | | | | | |
| Initial Access Remote | x | | | | | | | | | | | x | x | x | | x | | |
| Initial Access Local | | | x | | | | | | | | | | | x | x | | | |
| Initial Access | | | | | | | | | | | | | | | | | x | x |

Thus, the ATT&CK matrix provides a starting point for the backbone structure of the BN, but the practitioner applying the methodology should adapt the structure to the specific application scenario. Furthermore, there are some implicit dependencies among Tactics that need to be reflected in the BN. We have already seen an aspect of this in the distinction made between Impact and non-impact Techniques. The dependencies reflect the fact that one Tactic may facilitate or enable subsequent use of another Tactic, and result in a partial ordering of Tactics within a kill chain.

Table 4.16 shows the incidence matrix for the Execution, Inhibit Response Function, and Impair Process Control Tactics and Table 4.17 shows the incidence matrix for the Bayesian Network nodes for the Impact Tactic. It can be observed that in this matrix Techniques are depending on other Techniques for the first time. This because in the previous matrices, there were auxiliary nodes in between Techniques. However, this relationship is not forbidden, since it is normal for the probability of a Technique or Tactic to influence on the probability of other Technique or Tactic.

*Table 4.16: Incidence Matrix for Execution, Inhibit Response Function, and Impair Process Control*

| | IoR006 | IoR007 | IoR101 | IoR105 | IoR106 | IoR107 | IoR112 | IoR301 | IoR302 | IoR303 | IoR402 | IoR405 | IoR406 | IoR501 | IoR502 | IoR504 | IoR509 | IoR701 | Initial Access | T0814 | T0816 | T0875 | T0833 | T0823 | T0838 | T0835 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T0814 | | x | | | | x | x | | | | x | | | | | | x | | x | | | | | | | |
| T0816 | x | | | x | x | | | x | | | | x | x | | | | | | x | | | | | | | |
| T0875 | x | | | | x | x | | | x | | | | | x | x | x | | | x | | | | | | | |
| T0833 | x | | | x | x | x | | | | | | | | x | x | x | | | x | | | | | | | |
| T0823 | | | x | x | x | | | | | | | | | | | | | | x | | | | | | | |
| T0838 | | | | | | | | x | x | x | | | | x | x | x | | x | x | | | | | | | |
| T0835 | | | | | | | | x | x | x | | | | | | x | | | x | | | | | | | |
| Turn off Tº Control | | | | | | | | | | | | | | | | | | | | x | x | x | x | x | | |
| Increase Tº settings | | | | | | | | | | | | | | | | | | | | | | x | x | x | X | x |

*Table 4.17: Incidence Matrix for Impact*

| | IoR301 | IoR302 | IoR303 | IoR406 | IoR501 | IoR502 | IoR504 | IoR509 | T0814 | T0816 | T0875 | T0833 | T0823 | T0838 | T0831 | T0835 | T0879 | T0826 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| T0831 | x | x | x | x | | | | | | x | x | x | x | x | | | | |
| T0879 | | | | | x | x | x | | | x | | | | x | x | | | |
| T0826 | | | | | | | | x | x | x | | | | | x | x | x | |
| T0828 | | | | | | | | | | x | | | | | x | | x | x |

For the graphical representation, sub-models were developed for each incidence matrix in order to make the methodology scalable and to provide an easier visualisation of the Bayesian Network. Figure 4.18 shows the macro-view of the BN including links between tactics. Single arrows mean that one node of the BN sub-model is link to a node to the other BN sub-model and double arrows mean that two or more nodes of both networks are linked.

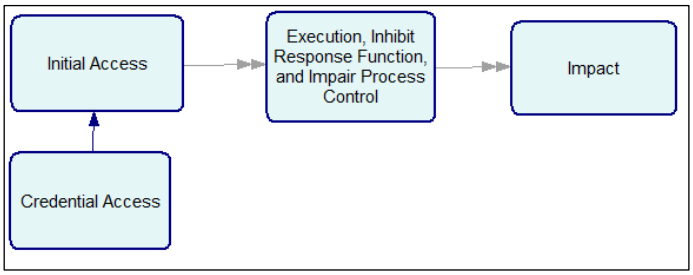



*Figure 4.18: BN macro-view*

Figure 4.19 shows the BN for the Credential Access Tactic based on Table 4.14 including the two Techniques identified that belong to this Tactic and the IoRs that will be used to monitor the likelihood of occurrence. It should be noted that the sub-models have been created only for the purpose of a better visualisation of each attack stage, but they are all interconnected and they share some IoR nodes. This means that it is not possible to capture all the IoRs that belong to each Tactic in the sub-model image capture, for which it might seem that some of the IoRs from the incidence matrices are missing. Figure 4.19 shows an example of this, in which the parent nodes of Technique T865 can be visualised by clicking on the node image. This shows that the parent nodes are IoR109 and IoR116, which corresponds to Table 4.14. However, only the node that corresponds to IoR116 can be appreciated in the diagram of the image. This same applies for other Techniques and Tactics in this and the following BN diagrams.

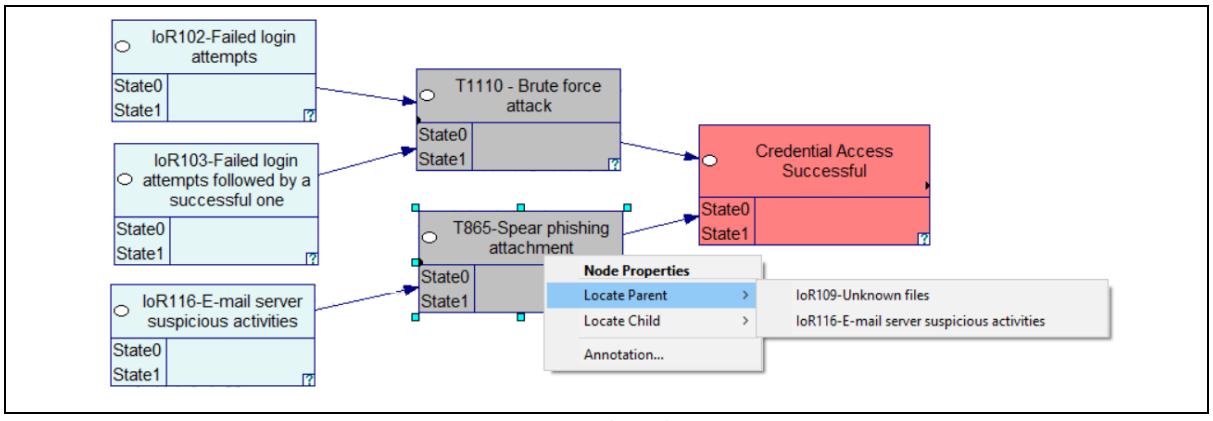



*Figure 4.19: Credential Access BN*

Figure 4.20 shows the BN for the Initial Access tactic based on Table 4.15.

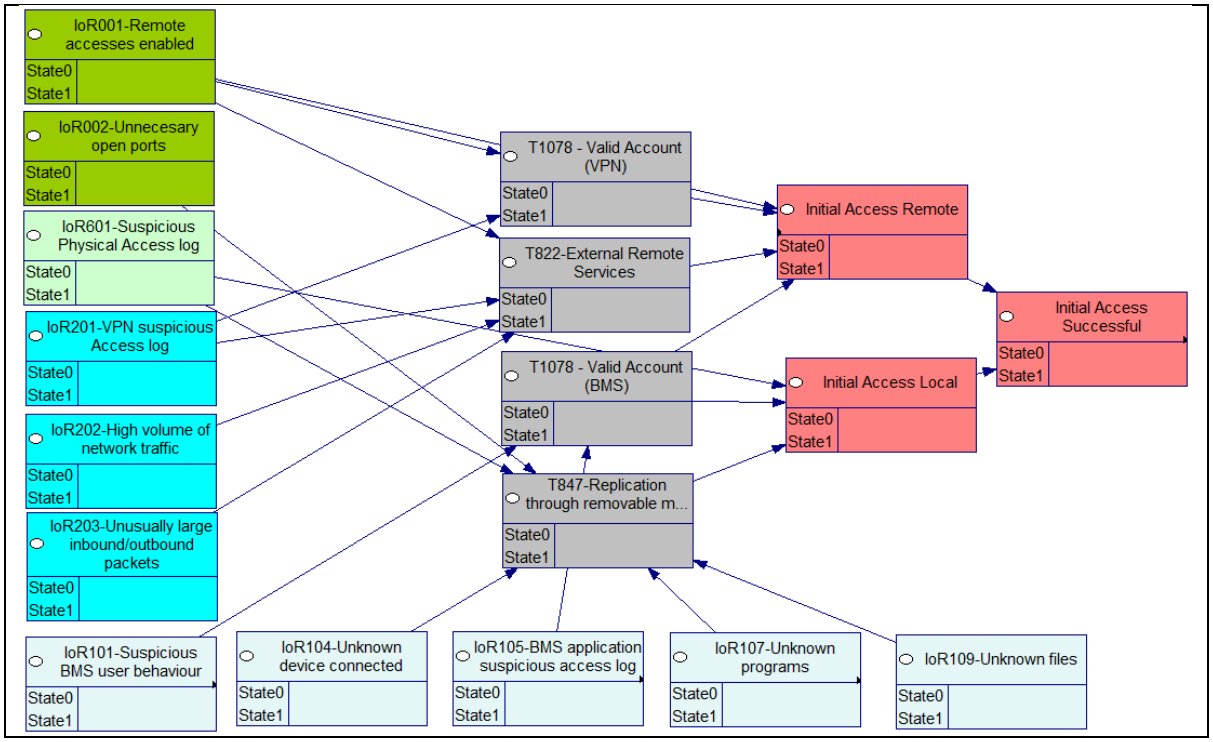



*Figure 4.20: Initial Access BN*

Figure 4.21 shows the BN for the Techniques that belong to the Execution, Inhibit Response Function, and Impair Process Control Tactics based on Table 4.17 all of them related to the achievement of the attack goals, which are disabling the temperature control and increasing the temperature levels. As attacks on ICS can affect field devices and consequently I/O that can be related to physical variables of the system, some IoRs are based on misbehaviour or anomaly detection, which will be discussed further in Chapter 6. For example, a temperature anomaly can result in other consequences such as abnormal power consumption, due to the fans of the servers reacting to the heat and spinning faster. Figure 4.22 shows the BN for the Impact tactic based on Table 4.18.

As it can be observed from the incidence matrices, the Techniques from Figure 4.22 are linked to IoRs and Techniques from the previous sub-models, which do not appear explicitly in the diagram but can be identified by clicking on each Technique node in the BN software. For example, the likelihood calculation of loss of availability is influenced by seven other nodes, from which only “Manipulation of Control" and “Damage to Property" that can be seen in the diagram because they belong to the same sub-model.

**Step 3: define the conditional probabilities**

For each dependent node, a table of conditional probabilities is defined to enable calculation of the likelihood of it being in a certain state given the observation of the states of the nodes that exert an influence on it. For simplicity, in this exercise most states were defined as binary variables such that “State 0" means “false" and “State 1" means “true". The assignment of conditional probabilities can be based on statistics, on expert judgement, or in a combination of the two. In cases such as advanced and targeted threats for ICS, it is unlikely that there will be enough data available to estimate a probability. Attacks of this type are often highly tailored to a particular victim, so they are rarely repeated exactly. Furthermore, high impact zero-day exploits may often be used only once because their effectiveness is reduced once their exploit mechanisms are revealed. However, an expert can estimate probabilities of success based on knowledge of how malware spreads (e.g. by a trusted insider or by spear-phishing) and of the effectiveness of security controls in place to defend against such techniques.

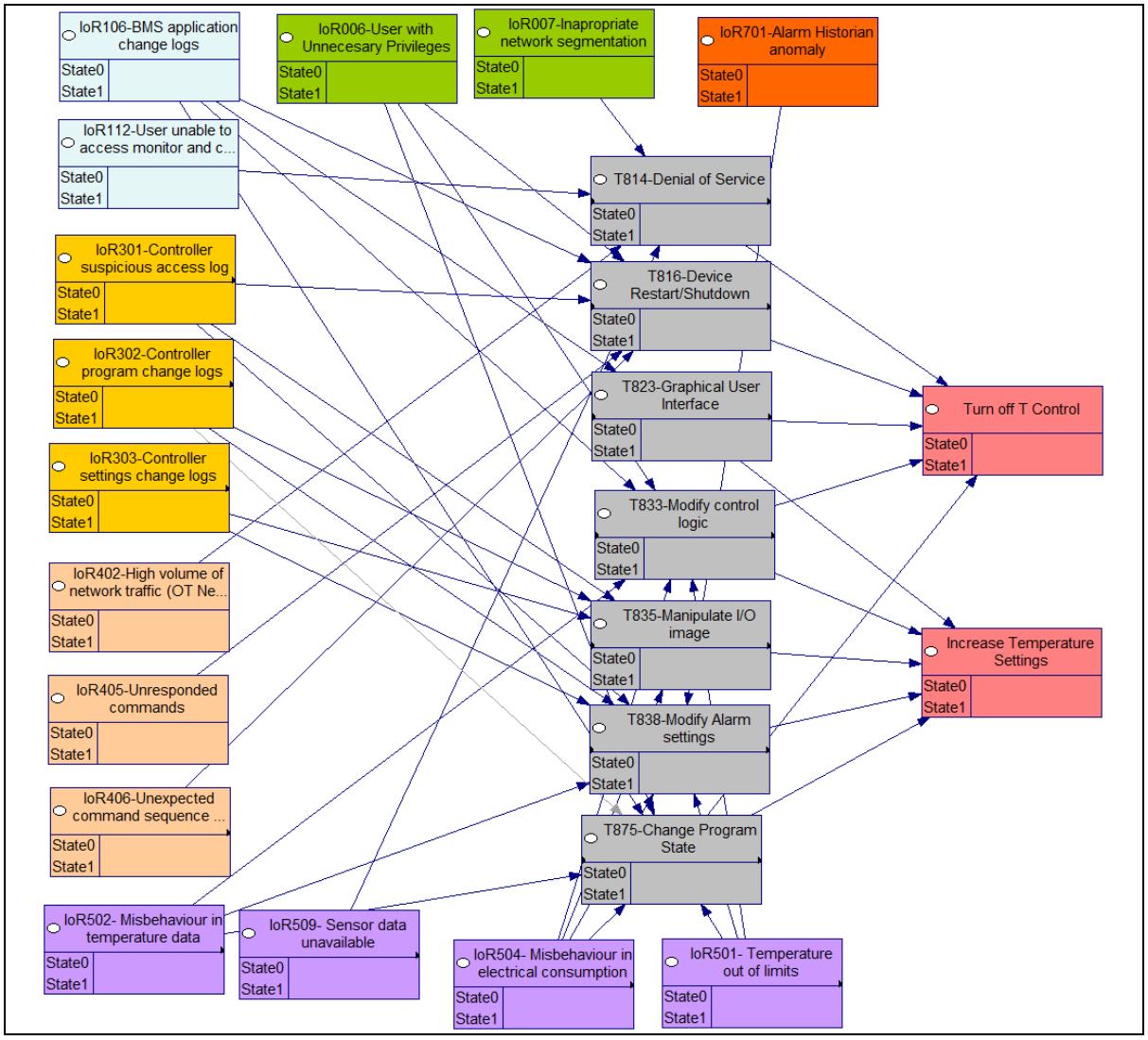



*Figure 4.21: BN for Attack Goals*

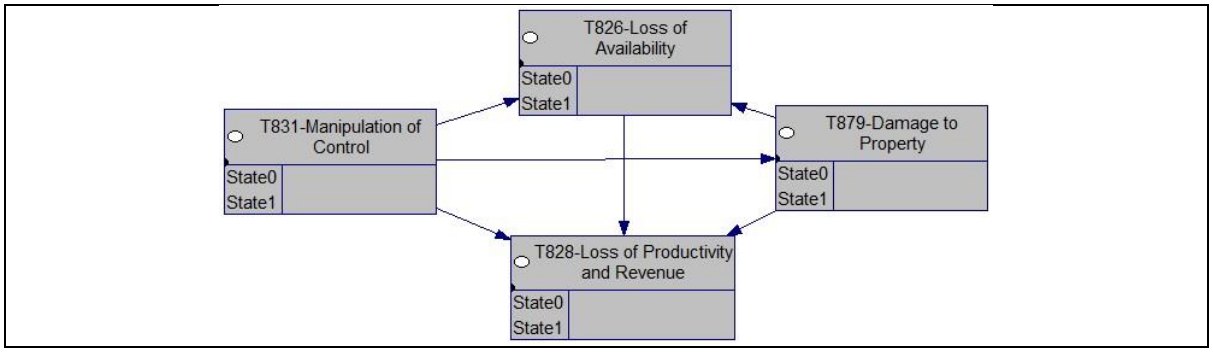


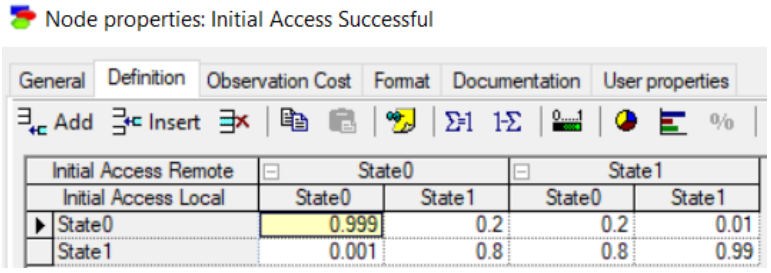

*Figure 4.22: Impact BN*

As an example, Figure 4.23 shows, the conditional probability table for the “Initial Access” Tactic, which depends on the events “Remote Initial Access” and “Local Initial Access”. It can be observed that the probability of initial access when both of these events are in “State 0” (false) is very low. However, when one of these events is in “State 1” (true) the probability increases. Note that the states of the nodes “Remote initial access” and “Local Initial Access” are also probabilistic and depend in turn on IoRs. For example, if at a given time there is a 5% probability of a malicious remote access attempt and a 0% probability of a malicious local access attempt, then the overall probability for a successful initial access will be calculated as follows:

$$P(IA) = P(IA/RA = F, LA = F) + P(IA/RA = F, LA = T) + P(IA/RA = T, LA = F) + P(IA/RA = T, LA = T)$$

Where,

- **P(IA)** is the Probability of Initial Access
- **P(IA/RA = $S_{RA}$, LA = $S_{LA}$)** is the probability of Initial Access when Remote Access (RA) has a state "$S_{RA}$" and Local Access (LA) a state "$S_{LA}$"
- $S_{RA}$ and $S_{LA}$ are independent from each other and can be on "State 1", which means True (T) or "State 0", which means False (F).




| Initial Access Remote | State0 | | State1 | |
|---|---|---|---|---|
| Initial Access Local | State0 | State1 | State0 | State1 |
| State0 | 0.999 | 0.2 | 0.2 | 0.01 |
| State1 | 0.001 | 0.8 | 0.8 | 0.99 |

*Figure 4.23: Probability table for Initial Access Tactic*

From the example statement we know that RA has a 5% probability to be in State 1 and as a consequence a 95% probability of being in State 0 and LA has a 0% probability to be in State 1 and as a consequence a 100% probability State 0. Then, if we compute all this values with the corresponding conditional probabilities in Figure 4.21 for each one of the combinations for Initial Access to be in State 1 (true), results are the following:

- $P(IA/RA = F, LA = F) = 0.001 \times 100\% \times 95\% = 0.095\%$
- $P(IA/RA = F, LA = T) = 0.8 \times 100\% \times 5\% = 4\%$
- $P(IA/RA = T, LA = F) = 0.8 \times 0\% \times 95\% = 0\%$
- $P(IA/RA = T, LA = T) = 0.99 \times 0\% \times 5\% = 0\%$

Then $P(IA) = 4.095\%$

The BN tool computes this automatically by previously configuring the probability table as in Figure 4.23 and setting the probabilities of the RA and LA nodes, as in the example. In our approach, the probabilities of different nodes are not expected to be set manually but as a consequence of system's observations (aka IoRs). The only settings that need to be configured in advance are the probability tables. Probabilities are re-calculated dynamically by the BN, which means every time that there is a change in IoR observations. This is done in an automated fashion, which means without human intervention.

In this use case, all the values used, including the ones on Figure 4.23 are just examples in which plausible values for assignment of conditional probabilities were picked to illustrate how it would be done in a case based on expert judgement. In other words, the estimation of each probability of occurrence of a technique given the observation of one or several IoRs is done through what is known in Risk Management as an "educated guess". The conditional probabilities will be higher for IoRs that can be more strongly linked to the technique and can predict it with more accuracy, and lower for those that are not strongly related but still can give some indication of risk, especially in combination with other IoRs. A suggested way to check that the probabilities assigned are reasonable is to verify consistency of the probabilities obtained in the BN when no IoR is detected with the likelihood obtained in the Baseline Risk Assessment.

### 4.3.3. Continuous risk assessment phase

The Continuous Risk Assessment consists of Continuous Updating of the State of IoRs, Continuous Risk Analysis, and Continuous Risk Evaluation. Figure 4.24 shows an example of how a Dashboard used for continuous risk monitoring applied to this case should look when no changes in the state of IoRs are observed. The key

information that is displayed is the current likelihood estimation of the risks that are being monitored, the corresponding impact and risk scores, plus the probabilities that various TTPs are currently being performed.

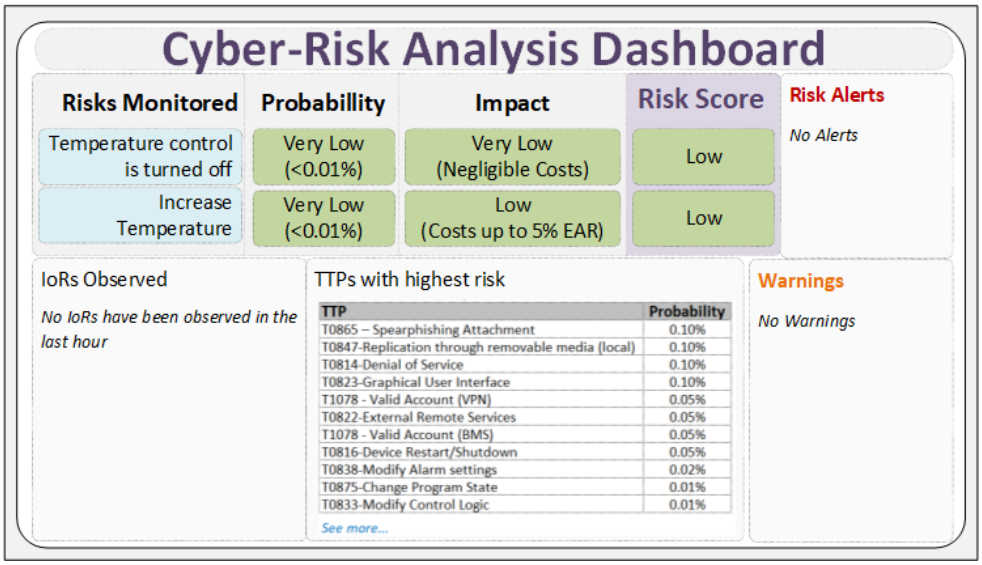



*Figure 4.24: Example 1 of a dashboard for continuous risk monitoring (no IoRs observed)*

Continuous risk analysis consists of the re-calculation of risk likelihoods based on IoRs and the consequent adjusting of the risk scores. Whenever the state of an IoR changes, the likelihood of occurrence of all the nodes representing Techniques, Tactics and risks to which it is linked directly or indirectly are automatically re-calculated through the BN. Figure 4.25 shows another example of the dashboard when two IoRs are observed: "IoR302-Controller program change logs" and "IoR303-Controller settings change logs". The information displayed in this mock dashboard is based on real calculations done using the BN that was built for this worked example. It can be observed that two Techniques have their probabilities increased and one of the risk scores now exceeds the acceptable level. In this case, the current situation, including the causes that triggered the observation of the corresponding IoRs should be investigated to find out ways to reduce the risk, as well as to prevent the excessive risk happening again.

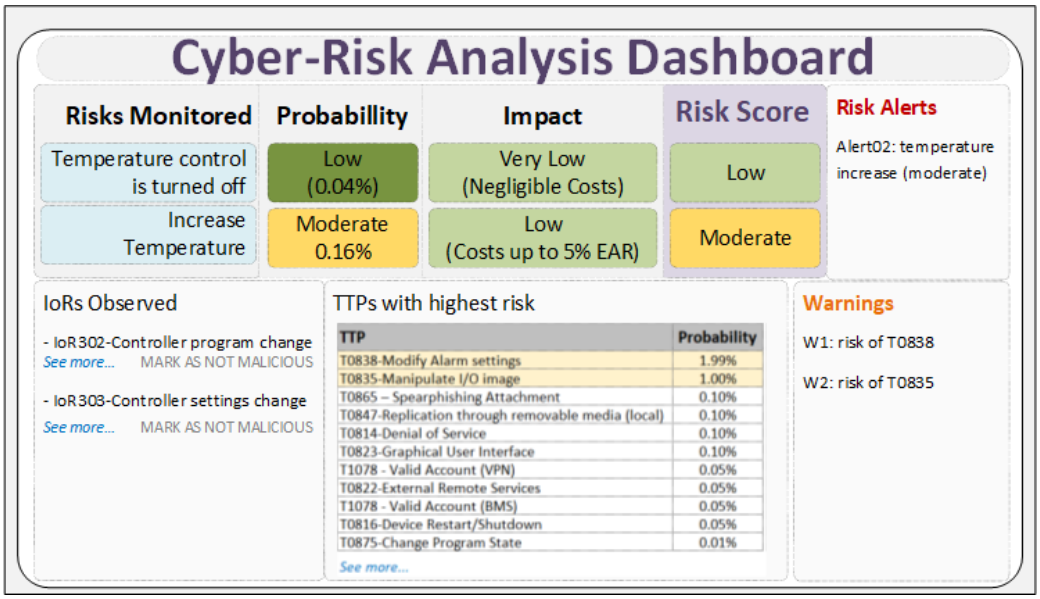



*Figure 4.25: Example 2 of a dashboard for continuous risk monitoring*

#### 4.3.4. Demonstration of probability changes based on the State of IoRs

To get a better idea of how continuous risk assessment works in practice, two examples were developed to illustrate how the BN gives updated probability calculations when IoRs are observed. The probability changes can be translated directly into a change in the risk scores. Moreover, if the impact is assumed to be the same in all cases, the focus can be restricted to the probabilities, establishing whether the probability for each risk exceeds a maximum acceptable value. The effects of a combination of IoRs on the likelihood of different risk events were observed to evaluate the behaviour the BN model. This was done by using different combinations of IoRs as an input to the BN, and checking that the output probabilities associated with various reference nodes appeared to be reasonable. For each example, a set of use cases were simulated to test the approach. The likelihood of occurrence on seven nodes of the BN were checked to verify the impact of different IoR combinations. The nodes used as reference in the simulations were the following:

- Initial Access
- Denial of Service
- Modify Control Logic
- Turn of Temperature Control
- Increase Temperature
- Loss of Availability
- Loss of productivity and revenue

##### 4.3.4.1. Scenario 1: Attack taking advantage of the remote access been enabled

Recall that the risk treatment plan described in section 4.2.6 required the remote access option to be disabled by default. This means that, under normal circumstances, the "Remote Access Enabled" IoR node is set in State 0 (false), which means that this IoR is not observed. However, remote access could be enabled temporarily for configuration and maintenance tasks. This will increase the chances that an adversary will gain unauthorised access, which will result in the risk scores exceeding the acceptable levels. This condition will remain until remote access is disabled again. This should also increase risk and warn security analysts of the fact that an issue is more likely to arise under this condition.

As can be observed in Figure 4.18, the "Remote Access Enabled" IoR node influences two technique nodes, “T1078 - Valid Account (VPN)" and “T0822-External Remote Services", and a threat event node “Remote Initial Access". To test the behaviour of the BN, the seven cases described below, which describe an evolving threat scenario, were considered. Each case is based on the previous case and adds additional IoRs. The results are displayed in Table 4.18. and Figure 4.26.

The description of each case is the following:

**Case 1: No IoRs.**
This case corresponds to normal operations, and is used as a reference against which other cases are compared.

The following are the IoR states:

- All IoRs are in State 0.

**Case 2: Remote access is enabled.**
In this case, the "Remote Access Enabled" IoR node is in state 1. When remote access is enabled through the VPN, there is an additional attack vector, which increases the probability that a malicious actor will gain initial access to the system from very low to 0.7%, which can described as medium.

The following are the IoR states:

- IoR001-Remote accesses enabled is in State 1.
- The rest of the IoRs are in State 0.

**Case 3: Case 2 plus a suspected brute force attack**
Several failed login attempts followed by a successful login in a predefined period cause IoR103 node to change to state 1, potentially indicating a successful brute force attack. Because of the evidence of a possible brute force attack on the VPN and BMS the risk of a malicious actor having valid credentials to the system and therefore

gaining initial access becomes high (4%) and the probability of other nodes starts increasing. This scenario should already require some investigation from the SOC.

The following are the IoR states:

- IoR001-Remote accesses enabled is in State 1.
- IoR103-Failed login attempts followed by a successful one is in State 1.
- The rest of the IoRs are in State 0.

**Case 4: Case 3 plus VPN and BMS access logs**

The addition of records to the VPN and BMS access logs is a normal activity and should not raise any alerts by itself. However, if access events occur while having the IoRs from Case 3 is State 1, this increases the probability of initial access to very high, which results in the other nodes having medium to high probabilities.

The following are the IoR states:

- IoR001-Remote accesses enabled is in State 1.
- IoR103-Failed login attempts followed by a successful one is in State 1.
- IoR201-VPN suspicious Access log is in State 1.
- IoR105-BMS application suspicious access log is in State 1.
- The rest of the IoRs are in State 0.

**Case 5: Case 4 plus unknown files detected**

Files can be uploaded to the system by an adversary to install or execute malicious code or to collect information from the system. If the unknown file is malicious it is most likely that the adversary has already gained access to the network, hence the probability of initial access increases significantly. Even though further evidence would be required for the file to be confirmed as malicious, in combination with other IoRs the risk levels increase.

The following are the IoR states:

- IoR001-Remote accesses enabled is in State 1.
- IoR103-Failed login attempts followed by a successful one is in State 1.
- IoR201-VPN suspicious Access log is in State 1.
- IoR105-BMS application suspicious access log is in State 1.
- IoR109-Unknown files is in State 1.
- The rest of the IoRs are in State 0.

**Case 6: Case 5 plus unknown programs detected**

An adversary can install malware or other software to perform malicious actions such as connection requests, queries, data collection, sending malicious commands, modifying the system's configuration or the control logic, among many others. As the program is unknown, but has not been identified as malicious, there could be other explanations for its presence. However, the addition of this IoR to the ones already detected increases the probability for most of the reference nodes.

The following are the IoR states:

- IoR001-Remote accesses enabled is in State 1.
- IoR103-Failed login attempts followed by a successful one is in State 1.
- IoR201-VPN suspicious Access log is in State 1.
- IoR105-BMS application suspicious access log is in State 1.
- IoR109-Unknown files is in State 1.
- IoR107-Unknown programs is in State 1.
- The rest of the IoRs are in State 0.

**Case 7: Case 6 abnormal electrical consumption**

In cases 2 to 6 the IoRs are related to events in the ICT system. In case 7 an additional IoR is triggered which is related to data from physical sensors. This data directly concerns the main process of the system and a deviation from the normal behaviour can reveal a malfunction. In this case, if the temperature control is turned off or set to increase the temperature, both the electrical consumption from the IT equipment hosted in the data centre and of the ICS itself can change. If this happens, the probability that an attack is in progress becomes higher.

The following are the IoR states:

- IoR001-Remote accesses enabled is in State 1.
- IoR103-Failed login attempts followed by a successful one is in State 1.
- IoR201-VPN suspicious Access log is in State 1.
- IoR105-BMS application suspicious access log is in State 1.
- IoR109-Unknown files is in State 1.
- IoR107-Unknown programs is in State 1.
- IoR504- Misbehaviour in electrical consumption is in State 1.
- The rest of the IoRs are in State 0.

*Table 4.18: Summary of the results for Example 1.*

| Case/ BN node | Initial Access | Denial of Service | Modify Control Logic | Turn off T° Control | Increase T° | Loss of Availability | Loss of productivity & revenue |
|---|---|---|---|---|---|---|---|
| Case 1 | 0.03% | 0.10% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% |
| Case 2 | 0.76% | 0.14% | 0.03% | 0.01% | 0.01% | 0.07% | 0.01% |
| Case 3 | 4.00% | 0.30% | 0.13% | 0.02% | 0.04% | 0.16% | 0.02% |
| Case 4 | 10.71% | 0.62% | 1.96% | 0.16% | 0.23% | 0.37% | 0.10% |
| Case 5 | 10.75% | 0.63% | 1.97% | 0.16% | 0.23% | 0.37% | 0.16% |
| Case 6 | 10.90% | 2.87% | 6.10% | 0.23% | 0.44% | 1.50% | 0.22% |
| Case 7 | 10.90% | 2.87% | 11.10% | 0.32% | 1.06% | 3.06% | 2.98% |

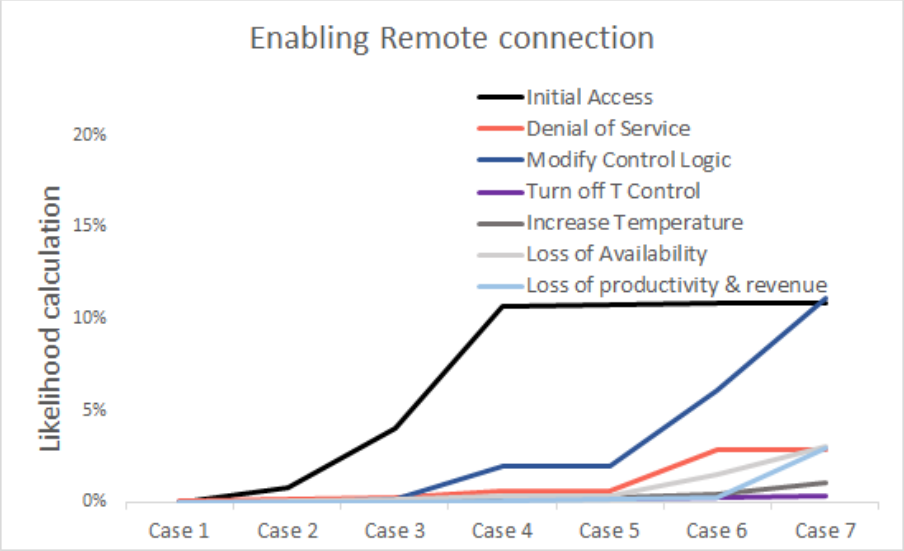



*Figure 4.26: Summary of the results for Example 1*

In cases 2 to 7 there is not enough information to confirm that an adversary has attempted a cyber-attack to the system, but the more IoRs that are observed, the higher is the likelihood that a malicious action is happening. If this likelihood reaches a level at which risks are above the acceptance level, the causes of the IoRs should be investigated to establish whether the different events that caused the simultaneous IoR observations are related between each other and whether there is a risky condition or even a security incident going on. In either case, it needs to be defined what actions need to be taken. Actions can be related to risk reduction and also to incident response, if there is reasonable evidence to presume a possible attack.

#### 4.3.4.2. Scenario 2: Attack commencing with connection of an unrecognised device

This example starts with the connection of an unrecognised removable device, such as a USB drive, to the workstation. The consequent effects on the probability calculations when successive IoRs are are observed are

described. Table 4.19. and Figure 4.26 show a summary of the results of this simulated example including some representative nodes.

The description of each case is the following:

**Case 1: No IoRs**
This case is used as a reference of normal operation, to be compared with the other cases.

The following are the IoR states:

- All IoRs are in State 0.

**Case 2: Unknown removable device is connected to the workstation computer.**
A removable device such as a USB drive can increase the risk of both accidental and intentional malware infection. Removable media can also be used to steal data or to copy malicious files from a computer terminal. Observing the connection of such device increases the chances that an adversary already has access (a typical case would be a malicious insider) and consequently the probabilities of other nodes that depend on this IoR.

The following are the IoR states:

- All IoRs are in State 0.
- IoR123-Unknown USB device plugged is in State 1.

**Case 3: Case 2 plus unknown files detected**
If to the detection of an unknown device, the detection of new files, which cannot be identified as part of the normal processes, is added, the chances of a successful initial access continue to increase.

The following are the IoR states:

- All IoRs are in State 0.
- IoR123-Unknown USB device plugged is in State 1.
- IoR109-Unknown files is in State 1.

**Case 4: Case 3 plus unknown programs detected**
As was explained in the previous example, the detection of an unknown program can increase the risk of malicious activity.

The following are the IoR states:

- All IoRs are in State 0.
- IoR123-Unknown USB device plugged is in State 1.
- IoR109-Unknown files is in State 1.
- IoR107-Unknown programs is in State 1.

**Case 5: Case 4 plus anomaly in temperature data**
An anomaly in temperature data can have several explanations such as a sensor malfunction. However, when this information coincides with other IoRs, as it does in this case, the probabilities of an adversary turning off functions of the temperature control or increasing the temperature increase.

The following are the IoR states:

- All IoRs are in State 0.
- IoR123-Unknown USB device plugged is in State 1.
- IoR109-Unknown files is in State 1.
- IoR107-Unknown programs is in State 1.
- IoR502- Misbehaviour in temperature data.

**Case 6: Case 5 plus high volume of traffic in OT network**
If an unusually high volume of traffic is detected in the OT network this could mean that there is a Denial of Service (DoS) attack in progress. This can result directly in loss of availability of the ICS and could also be a diversion to disguise a more sophisticated targeted attack.

The following are the IoR states:

- All IoRs are in State 0.
- IoR123-Unknown USB device plugged is in State 1.
- IoR109-Unknown files is in State 1.
- IoR107-Unknown programs is in State 1.
- IoR502- Misbehaviour in temperature data.
- IoR402-High volume of network traffic (OT Network)

**Case 7: Case 6 plus users unable to access BMS**

If the BMS is unresponsive or users are unable to access it, the probability of a DoS attack becomes even higher.

The following are the IoR states:

- All IoRs are in State 0.
- IoR123-Unknown USB device plugged is in State 1.
- IoR109-Unknown files is in State 1.
- IoR107-Unknown programs is in State 1.
- IoR502- Misbehaviour in temperature data.
- IoR402-High volume of network traffic (OT Network).
- IoR112-User unable to access monitor and control application (BMS).

In the same way as example 1, example 2 shows how different combinations of IoRs can give information about the probability that an attack will take place in the immediate future or even be in progress, and about the likely nature of the attack. For example, in Case 4, according to the BN model calculations, there is already a high probability that something malicious is going on, but it the attacker's goal is unclear. In Case 5 and Case 6 it becomes more clear that it is highly likely that a DoS attack is taking place.

*Table 4.19: Summary of the results for Example 2.*

| Case/ BN node | Initial Access | Denial of Service | Modify Control Logic | Turn off T° Control | Increase T° | Loss of Availability | Loss of productivity & revenue |
|---|---|---|---|---|---|---|---|
| Case 1 | 0.03% | 0.10% | 0.01% | 0.00% | 0.00% | 0.00% | 0.00% |
| Case 2 | 0.18% | 0.11% | 0.02% | 0.00% | 0.00% | 0.05% | 0.01% |
| Case 3 | 2.40% | 0.22% | 0.08% | 0.01% | 0.03% | 0.11% | 0.01% |
| Case 4 | 4.80% | 2.38% | 1.43% | 0.05% | 0.11% | 1.20% | 0.13% |
| Case 5 | 4.80% | 2.38% | 5.48% | 5.39% | 0.62% | 2.74% | 2.86% |
| Case 6 | 4.80% | 30.00% | 5.48% | 7.00% | 0.62% | 16.57% | 4.26% |
| Case 7 | 4.80% | 40.00% | 5.48% | 7.55% | 0.62% | 21.32% | 4.74% |

Although this is a simplified example, it illustrates how continuous risk monitoring based on BNs would work, and how it can provide awareness of a risky situation and of the attacker's potential goals and the consequent impacts. Additionally, as many IoRs are common to several known TTPs, and several IoRs are based on the system's behaviour rather than on specific signatures, they can reveal symptoms of an attack even if they cannot detect the specific TTP. This makes this approach potentially useful in detecting unknown or Zero Day attacks. This idea will be developed further in Section 5.7. of Chapter 5.

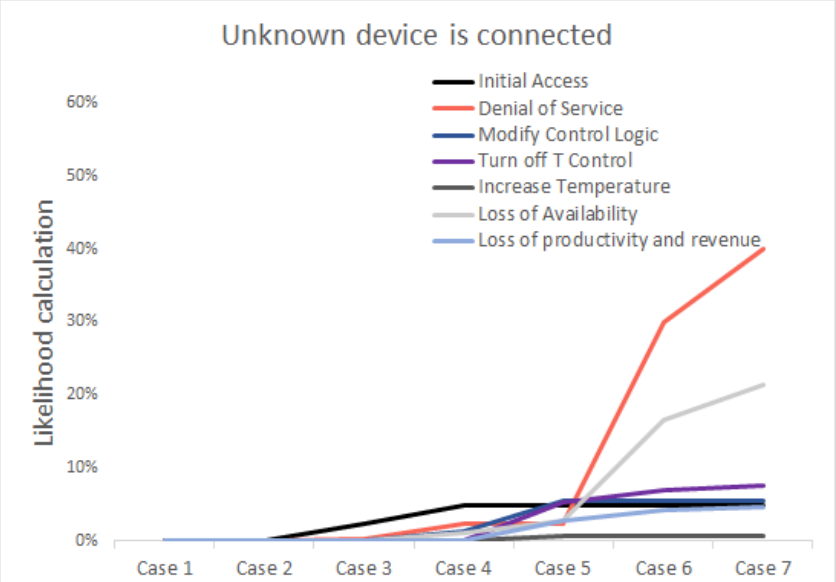



*Figure 4.27: Summary of the results for Example 2.*

## 4.4. Discussion

In Chapter 4, the Risk Assessment methodology proposed was described and illustrated through a worked use case. This example was used to describe each phase of the methodology, which starts with the Baseline Risk Management, followed by the Transition phase, to finally enter on the Continuous Risk Assessment phase. The Baseline Risk Management is used to define the context and the benchmarks for the Continuous Risk Assessment, and the Transition phase to prepare the organisation to conduct the Continuous Risk Assessment. This preparation includes identifying IoRs, defining and putting in place means to observe IoRs, building the Bayesian Network based on threat models, implementing the tools and processes for the Continuous Risk Assessment, and training people. Examples were given of different scenarios that could happen during the Continuous Risk Assessment and how the risk scores would be modified on those scenarios.

The following are the contributions from the continuous risk assessment methodology that is proposed and described in this chapter:

- A set of incremental modifications to the ISO 27005 risk management methodology that incorporate continuous risk assessment
- The concept of Indicator of Risk, which will be discussed in more depth in the next chapter
- The use of a Bayesian network to update estimates of 'instantaneous risk' (the risk associated with attacks currently in progress)

The methodology confers the following benefits:

**Increased visibility of the cyber-risks to which an industrial system is exposed**
The continuous risk assessment methodology can help provide an overview of the cyber-security risks of an ICS. Monitoring different IoR combinations can provide information about likelihood of an undesirable event happening. The overall implementation process with its three phases allows contextualising observations that can be monitored during operational time in relation with assessed risks in order to have continuous awareness of changes in the risk landscape. This should develop a better understanding of the impacts of possible cyber-incidents and the organisation's level of exposure to them. Overall, a better visibility of risks should enable an organisation to be better prepared to deal with continuously changing conditions and incomplete information.

**Validation of risk analysis assumptions**
The use of data from operations to assess risks can help to validate or refute assumptions made in the Baseline Risk Assessment. For example, in the first instance, the probability of a targeted attack could be considered as very low based on expert's knowledge, if up to the moment they have considered that the main threat to

organisation were opportunistic attacks and not targeted ones. However, the information collected during the Continuous Risk Assessment could reveal attempts to more targeted attacks based on evidence of data collection Techniques or crafted exploits. This shall result on adjusting conditional probabilities. This means that the methodology contemplates mechanisms for the results of the risk assessment to remain relevant and up-to-date at all times based on the available information.

**Adaptable and flexible**

An organisation can adopt the approach proposed in Appendix B or choose another method that suits them for the Baseline Risk Analysis, since this process is agnostic to the specific method used for risk estimations. It can also be chosen to use the BN technique in both baseline and continuous risk analysis. If it were chosen to calculate initial probabilities directly through the BN, a method to validate probability values would need to be considered, since using an independent method has the advantage to allow cross-checking probability values. Similarly, if the organisation already has suitable risk calculation method, they may adapt this to the continuous risk assessment methodology. The choice and implementation of IoRs is also non prescriptive, providing only high-level guidance to identify observations that can be useful for continuous risk monitoring from a large number of possible indicators. This allows an organisation to choose a set of IoRs that is appropriate to their system and its environment and implement it according to their own use cases. Hence, the methodology fulfils the goal of been applicable to a broad scope of industrial systems independently of their particularities.

**Enables an organisation to develop effective security strategies**

The examples developed in section 4.3 show how a risk that is initially considered low can change its probability value dynamically, through the development of events. The estimated probability increases when suspicious events are detected resulting in a higher risk score. This may lead to two complementary and concurrent courses of action. First, the SOC operator can immediately commence investigation and if necessary, carry out remediation and recovery actions. Second, the risk analyst can work on determining if this probability increase is depending on events that are likely to occur more often than expected, for which the associated Baseline Risk Score would not be representative anymore. In this case, the continuous risk assessment should give as a result a new risk treatment plan or a modification of the current one, adding security controls to reduce the risk in the future. Overall, this methodology allows the risk analyst to have a better understanding of threats and vulnerabilities and to develop better defences including automated responses to certain events. In other words, continuous risk management allows adapting security tactics and strategies over time. Overall, having visibility on the risk landscape at any given time based on up-to-date information about vulnerabilities and threats can allow identifying the most relevant security issues improving their security strategy.

**Integration between risk management and cyber-security operations**

A risk assessment methodology that is integrated with security operations will result in more effective, efficient and timely sharing of relevant security information. In this sense the methodology proposed considers in its workflows activities of information exchange between these two areas. Together with this, it introduces the concept of and use of IoRs, which will be described in more detail in Chapter 5. IoRs correspond to concrete forms of information sharing between security operations and risk management.

**Help improving cyber-assurance**

Structured approaches for cyber security assurance should answer two questions: whether security measures appropriate to meet the system's requirements have been implemented and whether they have been implemented properly [141]. The continuous risk assessment prepares an organisation to answer these two questions by giving an updated view of the security risk landscape. If increments on certain risk scores have an important magnitude or are detected frequently this can have as a root cause insufficient security measures or poor/ inappropriate implementation of controls. This means that the information that the continuous risk assessment provides should result also on security controls and measures to be reviewed and improved.

**Support decision-makers to make informed and rational choices**

Frequent updates of risks can also help decision-makers to make informed and rational choices. Think of a manufacturing company that has two plants: plant A and plant B. Plant A uses legacy systems, and plant B uses

new generation IIoT platform with built-in security. For various reasons, such as improving efficiency, the company is considering upgrading plant A to the same standard as plant B. As the investment required is high, it is important to make a good business case for this project to be approved. If there is quantitative information available showing that plant B suffers a lower level of cyber-security risk than plant A, which eventually can be realised as money savings (by preventing cyber-security incidents) this can contribute to the cost-benefit calculation [131]. Other example is the prioritisation of security controls based on gathering the information provided by the continuous risk assessment over time instead of in a specific point in time, such as it happens in the case of traditional risk assessments.

**Keep an eye on low-probability and high impact risks**

It is part of a normal risk assessment process for low risks and residual risks, which meet the risk acceptance criteria after the risk treatment plan to be accepted, since it would not be cost-effective and definitely not feasible to reduce every possible risk to zero. In ICS, cyber-risk mitigation can often introduce other issues such as affecting operations reliability or performance. For example, in cases where the operations run 24x7 in real time, delays introduced by latency of security controls, such as encryption can be problematic. Therefore, a certain level of cyber-related risk exposure can be accepted as long as it is an informed decision and there are monitoring activities in place to compensate for the vulnerability. This also addresses the fact that there is a degree of uncertainty associated with estimates of the probability of rare events. Hence, this type of risks could be also often underestimated. Particularly, risks that present a very high potential impact should be monitored continuously even if their likelihood is considered extremely low. Monitoring risks continuously through their related IoRs can allow the continuous assessment of several of these low-probability and high impact risks at the same time in an automated manner.

The next chapter will focus on the use of IoRs for Continuous Risk Assessment. A definition of IoRs based in the ICS ATTA&CK framework for which a knowledge base named the “IoR Library” will be proposed. This should allow facilitating the adoption of the Continuous Risk Assessment methodology and improve scalability of the process.

# 5. Use of Indicators of Risk (IoRs) for Continuous Risk Assessment

As discussed in previous chapters, continuous monitoring of cyber-security risks in industrial environments is highly recommended as a compensating control for some risks whose mitigations might have a technical constraint or a high cost [8]. Continuous risk monitoring aims to detect the presence of vulnerable conditions or threatening events that can increase the likelihood of an incident and it could also allow detecting an attack attempt in its early stages. In order to implement a continuous risk monitoring method, it is necessary to identify the risks that will be monitored and define how they will be monitored. For this, we have adopted the term "Indicator of Risk" (IoR) to refer to the means of observing these conditions or events. In this chapter, the IoR concept is explored further and a systematic way to identify IoRs is developed in association with adversary Techniques from the ICS ATT&CK knowledge base.

In Chapter 4, a worked example was presented to illustrate the Continuous Risk Assessment methodology. During the development of this use case, the identification of IoRs and associating them with the different stages of an attack was a laborious task. This led to think that there should be a way to systematise and standardise this process more, particularly as regards naming things such as IoRs and elements of the attack tree. At the beginning of 2020, MITRE released the ICS ATT&CK knowledge base, which gave an opportunity to use this framework to develop a more structured way to define IoRs, related to documented adversary Techniques. This led to revision of the worked example, adjusting the risk analysis method and the BN to be in alignment with ATT&CK resulting in the version that appears in Chapter 4. As a result, a method for IoR identification was developed which answers the two remaining research questions of this thesis: "what information is needed in order to monitor security risks in IIoT/ICS?" and "how can that information be derived from what you can actually measure?" As the answer can vary from system to system, a high-level approach was taken in order to provide a catalogue, named as the "IoR Library", which includes a number of IoRs that can be associated with different attack (ATT&CK) Techniques.

The IoR Library currently contains IoRs associated with Techniques from the ICS ATT&CK matrix, grouped according to the level of the Purdue architecture model to which the observations of events or conditions that they are based on apply. Its purpose is to be used as a reference when implementing Continuous Risk Assessment in a variety of contexts and to allow relevant IoRs to be associated with the risks that need to be monitored. In contrast to enterprise IT, industrial operations and their corresponding control and automation systems are diverse and each particular environment is fairly unique. This means that the specific information that needs to be processed to monitor an IoR can vary from system to system. Taking this into account, the IoR Library was structured as a catalogue to provide high-level definitions of IoRs as measurable observations of risky conditions on a system. For each IoR, the library provides guidance on its interpretation, how it relates to various techniques and examples of how it may be observed. This aims to help users to adapt IoRs to specific contexts and systems by defining more specific means of observation.

IoRs can be used to implement a Continuous Risk Assessment in a variety of industrial contexts including building management, manufacturing, and smart operations. The IoR Library allows identification of IoRs that can be linked to adversary techniques and can be associated to the risks that need to be monitored. To implement an IoR in a particular system, the generic description needs to be interpreted in terms of observations and detection rules that are meaningful in the specific context. For example, an IoR that refers to "unusually high levels of traffic in the OT network" has to be defined according to the usual range of network traffic and what level would be considered as unusually high. This interpretation is used to configure the SIEM tool to generate IoR observations when the network traffic presents abnormal characteristics.

The first section of this chapter is about defining the IoR concept and describing how IoRs are used in the context of Continuous Risk Assessment. The second section describes how the IoR library was built based on the MITRE ATT&CK TTPs and the third section describes how it is used. The fourth section presents a BN template based on the IoR Library which can be tailored to specific continuous risk monitoring use cases making the approach easier to adopt and more scalable. The fifth section has examples of BNs that include different techniques and IoRs to better illustrate how IoRs are used to monitor risks, which was already introduced in chapter 4. Section six, describes how the IoR Library was validated, section seven explains how management of IoR updates can be done, section eight describes a method to deal with unknown risks, and section nine proposes how to deal with

non-security related conditions that might trigger IoR observations. Finally, section ten provides a discussion of this chapter and highlights the contributions made and potential uses of IoRs and the IoR Library.

## 5.1. What is an Indicator of Risk (IoR)?

An important part of cyber-security operations is to develop and implement detection mechanisms that can allow an organisation to identify and counteract the actions of an adversary. Detection serves two purposes, which are investigation of possible events, and enabling the selection of appropriate responses such the blocking of adversarial actions. There are different types of threat detection that can be based either on discovering changes in the system's environment or on discovering signs of adversaries' activities [99]. Indicators of Risk, as defined in this thesis, can be environment or threat based, but can also be oriented to detect conditions that increase the vulnerability level of a system. This marks an important difference compared with some definitions of indicators of compromise, which can be restricted to specific characteristics of an attack procedure, such as IP addresses, hashes, signatures, URLs, domain names, or attack tools [53] [121] [99]. However, a wider definition of IoC [142] can include any condition that meets criteria to be considered as such regardless the detection mechanisms used. Hence, our IoRs concept covers indicators based also in techniques such as configuration analysis and behavioural analysis, which allow detecting oddities or anomalies in the system and its processes that can be related to cyber-security risks. This also follows recommended practices for ICS security monitoring [12] [99].

Overall, the term IoR can refer to any detectable condition, the observation of which alters the estimated probability that one or more possible threat events occur in the immediate future or even have occurred already. When IoRs are observed, the probability of an attack attempt being performed in the present or immediate future can increase. The magnitude of the increase of a probability estimation is determined by the degree of mutual influence between an IoR and a specific attack technique. In the IoR Library five "degrees of influence" are defined in order to give an idea of how strong the relationship between an IoR and an adversary technique is. This is ultimately expressed as conditional probabilities in a Bayesian Network

For an IoR to be implemented successfully and monitored continuously, it is important to have the appropriate tools to capture the information that would allow the conditions that the IoR is meant to expose to be inferred. In "A Survey of Security Tools for the Industrial Control System Environment" [53], different attributes are defined that characterise currently available security tools for industrial systems according to the zone in the ICS architecture in which they work, their purpose or function, their transport means or type of interfaces, and whether they are proprietary or open source. By looking into a tool's purpose or function it is possible to validate the feasibility of observing a given type of IoR. Types of tool function defined in [53] are IoC detection, network traffic anomaly detection, outlier analysis, log review, system artefact review, and reverse engineering analysis.

The idea of "Indicator of Risk" comes from a merger of the concepts of "Key Risk Indicator", commonly used in risk management, and "Indicator of Compromise" (IoC). An IoC can be defined as "one or more artefacts that relate to a particular security incident or attack" [124] or "artefacts observed on a network or in an operating system that indicate a computer intrusion with a high degree of confidence" [135]. Usually IoCs with a lower degree of confidence are dismissed because they generate a high number of false positives that can overwhelm security operators. However, risk management has as a premise the acknowledgement of uncertainty. This means that an IoR will not necessarily trigger a security alert but will result in creating awareness of the increased probability of a security event in order for key stakeholders to make better informed security decisions. Instead of ignoring or dismissing those indicators that are likely to have too many false positives, it is proposed that they are used as a means to estimate probabilities and to reduce uncertainty regarding possibility of occurrence of security-relevant events. Monitoring IoRs can provide valuable information about the state of cyber-risks, and even support cyber-security operations by improving their detection capabilities. By relating combinations of IoRs to attack techniques will give a view of possible adversary actions and their probability to be performed associated with the observation of one or more IoRs.

Figure 1.1 shows the Hierarchy of Security Risk Indicators, which is a conceptual relationship between IoRs, IoCs, and security alerts, as conceived in the continuous risk management approach developed for this thesis. In this model it can be observed that the concept of IoR covers and extends the concept of IoCs, hence, IoCs are a subset of IoRs that give a more deterministic indication of a threat. The goal of monitoring IoRs is to provide

valuable information about the state of cyber-risks, and even support cyber-security operations by improving detection capabilities.

Some IoRs will be straightforward to specify at a lower level, but some require the development of complex use cases. For example, establishing the state of an IoRs based on behavioural anomaly detection can be as simple as distinguishing normal from abnormal by comparing an input with a defined threshold, or as complex as defining normal parameters as a function of the state of the system and other system variables. Consider, for instance, network traffic. Depending on the processes run by the system there might normally be a constant flow of traffic making it easy to spot outliers by establishing a fixed threshold, or there might be batch processes that increase the volume of network traffic at defined moments. Hence, it would be necessary to have separate profiles of the normal level of traffic under different conditions. Patterns of communication in ICS can also be defined by the frequency with which two devices communicate or the use of certain commands in an OT network protocol. Some devices are expected to communicate only once during a specific interval and to send very specific messages. If these devices are detected as having frequent interactions, or if suspicious messages are detected, this could reveal a misuse of the system. Sensor data with stable behaviour will be easier to model and to use to detect anomalies. Conversely, cases in which variables have a more complex behaviour would require more complex mathematical models or use of machine learning techniques.

### 5.2. Building an IoR library based on the ATT&CK framework

The IoRs that should be implemented in order to conduct a continuous monitoring and assessment of cyber-risks in an ICS or IIoT deployment will depend on the system's baseline risk assessment and on the availability of detection and monitoring tools that can be configured for this purpose. Hence, each continuous risk assessment implementation will be unique, and depend on the ICS environment, business objectives, security posture, threat identification, and detection and monitoring capabilities. Thus, the methodology does not demand that particular IoRs be used, but provides high-level guidance to help an organization choose IoRs that are appropriate to its own context. To provide guidance in a structured and easy-to-communicate way, an artefact was developed named the "IoR library". The IoR library is a knowledge base created in alignment with the MITRE ATT&CK framework, specifically in the ICS ATT&CK matrix, which can be used to help identify IoRs appropriate to the organisation's own conditions and risk posture. Its main purpose is to facilitate the implementation of the Continuous Risk Assessment methodology in a variety of contexts and industrial environments and it could be used as part of other methodologies.

The MITRE ATT&CK framework [44] is presented as a set of matrices that give an overview of adversarial TTPs for different contexts. Currently there are three types of matrix: Enterprise (which includes information on the following platforms: Windows, MacOS, Linux, and various cloud and network platforms), Mobile (which includes information on Android and iOS platforms), and ICS. ATT&CK is commonly used as a reference to develop threat models and is also integrated in the products and services of security vendors. Several well-known SIEM (Security Information and Event Management) platforms are aligned with MITRE ATT&CK [133] [134] [135] and so are security monitoring tools specific to OT [143]. ATT&CK, thanks to the collaboration of the security community, is a rich source of information about adversary TTPs. In combination with languages such as STIX and TAXII, which also originated at MITRE, it has the potential to form the basis of a common framework used by security professionals and enabling interoperability of tools and platforms.

ATT&CK for ICS, the first version of which was released in January 2020, has its focus is on adversary TTPs whose primary goal is to disrupt an industrial control process. This disruption can cause physical damage, destruction of property, harm to the environment, injury and even death. As ICS are also integrated with or connected to enterprise IT systems, ATT&CK for Enterprise serves as a complement to ATT&CK for ICS. However, since enterprise cyber-security is a more developed area for which several playbooks and detection use cases are already known, the IoR Library is focused on the ICS matrix.

An ATT&CK Matrix is essentially a hierarchically-structured tactical library for understanding the behaviour of threat agents. A Tactic is basically a high-level part of threat agent's repertoire. It can be thought of as containing a number of Techniques, each being an alternative way of executing the Tactic that might be appropriate in a different context. The concrete way to execute a technique is known as a Procedure. An adversary formulating a general strategy will select their Tactics as the major steps in the plan and then a Technique or a set of Techniques for each Tactic. The Technique is the central concept in the ATT&CK framework. Tactics are means of grouping Techniques that are used at a similar stage in an attack. Procedures are too numerous, varied and

technology-specific to be catalogued exhaustively, and are used primarily as illustrative examples when describing Techniques. Each entry in a Matrix gives a high-level account of the technique in question, and typically also has a number of sub-sections. These vary from Matrix to Matrix, and even within a Matrix, not all are compulsory. However, entries frequently include the following sub-sections:

- Procedure Examples: outlining procedures that can be used to implement the Technique in question in a particular technical context
- Mitigations: Measures that can be taken to reduce susceptibility to the Technique
- Detection: Discussing means by which use of the Technique may be detected. Although not part of the Detection section, entries also list Data Sources that are useful in the context of detection. The Technique entries in the ICS Matrix do not have Detection sections, but nevertheless often do list Data Sources.

### 5.2.1. Main overview of the IoR library

The IoR Library is a knowledge base aligned with MITRE ATT&CK for ICS, that lists and describes IoRs and their relationships to adversary Techniques and it has been made openly available through a publicly accessible link [144]. Its main purpose is to facilitate the implementation of continuous risk assessment in a variety of contexts and industrial environments. IoRs should be selected from the library according to the risks that have been already identified and analysed, which is done in the Baseline Risk Assessment according to the methodology proposed on Chapter 4, and to implementation feasibility, which is checked in the Transition phase. As it is not possible to define the use of IoRs in a prescriptive way, the IoR Library was developed to provide a structured high-level guidance to identify and select the IoRs that are more appropriate. The latest version of the IoR library has 95 IoRs and covers 50 Techniques from the ICS ATT&CK framework, across all the 11 Tactics and can be developed further to include all the ICS ATT&CK Techniques, which at the time of writing were 81[1]. This further development would benefit also from contributions from the security community and could also be extended to techniques from other matrices of the ATT&CK framework.

The IoR library is structured as an extension to the ICS ATT&CK framework, and provides a list of IoR types cross-referenced with adversarial Techniques. Each IoR has its unique ID and name, and its entry in the library provides a rationale, examples of possible observations, examples of scenarios in which the IoR might be observed, and a list of Techniques it is applicable to. The rationale explains why and how the IoR can be related to a cyber-security risk, observations outline ways to measure or detect the IoR including possible data sources and tools, and the examples provide use cases or illustrations of processes by means of which the observations can be carried out in specific contexts. Appendix C contains an overview of the IoR definitions, an extract of which, can be found in Figure 5.1.

As mentioned earlier, the ICS ATT&CK Technique entries do not yet provide information about relevant detection means. However, it was possible to infer IoRs from other fields of the entries, including descriptions, data sources, procedure examples, and references. Additionally, other sources including ICS detection Proof of Concepts implemented by other researchers were used to help define the IoRs.

An ID scheme, detailed in Table 5.1, groups IoRs according to the level of the ICS to which they apply, based on the Purdue model. The IDs have IoR as a prefix followed by three digits, the first one represents the group to which the indicators belong and the others are an incremental counter. The third and fourth columns describe the levels in which each IoR group can originate and be observed, respectively. For example, the data from a sensor originates in Level 0 (physical process) but can be observed through a control and monitoring application in Level 2. IoRs related to physical perimeter security and safety systems are included through groups IoR6XX and IoR7XX, respectively. The main purpose of including these two last groups is to acknowledge that these systems, which are usually independent of the main ICS can provide useful data for a comprehensive risk monitoring. However, the focus of the work is on groups IoR0XX to IoR5XX.

The way the IoR groups relate to the Purdue model is described in a more graphical way in Figure 5.2, which shows the data origin related to the Purdue model for each IoR group, when applicable. The IoR Library is mostly focused on levels 0 to 2, which is also the scope of the ATT&CK for ICS model [145]. In the model adopted as

[1] At the time of submission the Techniques were reduced to 79 since the content of T0825 and T0808 was merged into other Techniques.

reference, communication between Level 3 and the Cell/Area zone is via a standard IT network. In this case, only Level 2 is capable of processing both OT and standard IT protocols.

| **IoR** | **Rationale**<br>*Why this is consider an IoR?* | **Observations**<br>*Examples of how this IoR can be observed* | **Examples**<br>*Scenarios where the IoR might be observed* | **Related Techniques** |
|---|---|---|---|---|
| IoR008-Outdated OS | Outdate OS can be missing security patches against known vulnerabilities. Also an OS that has passed the EOL will have no support and no further security updates. | Not having the latest version of the OS. | Adversaries might use known means of exploit of OS vulnerabilities, which can allow actions such as arbitrary code execution or a DoS attack. | T0810, T0818, T0866 |
| IoR117-Suspicious OPC commands | OPC is a vendor-agnostic protocol used for IT systems to communicate with ICS and access data. For this IoR it is necessary to define a baseline of normally used commands in order to identify suspicious behaviour. | Use of OPC commands blacklisted or not whitelisted.<br>Use of certain OPC commands with an unusually high frequency or under unexpected conditions. | An adversary uses an OPC command to reset all active alarms.<br>An adversary uses OPC commands to make queries about the state of the system. | T0801, T0802, T0808, T0825, T0868, T0870, T0877 |
| IoR302-Controller program change logs | A controller's program should be only changed during scheduled maintenance periods or under other pre-defined conditions, for which any changes under other circumstances can have risks. | Logs indicate a change in a controller's program outside the maintenance period. | An adversary changes a controller's program to change operational rules and disrupt industrial operations. For example, changing the doses of a component on a chemical process. | T0831 |
| IoR511-Inconsistency between different sources of data | If I/O data is inconsistent between different sources or between redundant measurements, or at different levels of the system it can mean that I/O data have been tampered with. | The I/O value given by a redundant sensor differs significantly from the main sensor.<br>I/O value displayed in an HMI to the one displayed in SCADA application in the engineering workstation. | An adversary spoofs the I/O image in the SCADA application running in the workstation. | T0832, T0835, T0878 |

*Figure 5.1: Extract of the IoR Library Definitions*

*Table 5.1: IoR naming scheme*

| **IoR ID** | **IoR group** | **Data origin** | **Data observation** |
|---|---|---|---|
| IoR0XX | Vulnerabilities | All levels | All levels |
| IoR1XX | IT system threat (servers, workstations, applications) | Level 2 | Level 2 |
| IoR2XX | IT network threat (ICT communication systems) | Levels 1 and 2 | Levels 1 and 2 |
| IoR3XX | Field devices threat (controllers, drives, smart relays) | Level 1 | Levels 1 and 2 |
| IoR4XX | OT Network (Industrial protocols) | Levels 0 to 2 | Levels 0 to 2 |
| IoR5XX | I/O data threat (sensors, actuators) | Level 0 | Levels 0 to 2 |
| IoR6XX | Physical security threat | Usually an independent system | |
| IoR7XX | Safety threat | Safety zone (independent system) | |

The scheme shown in Table 5.1 and Figure 5.2 was developed under the assumption that an ICS will have a general network architecture that corresponds to the one in the SANS version of the Purdue model [81]. However, it is expected that the system will present variations on its architecture and may have its own naming scheme for levels or zones. Hence, an organisation may choose to adapt the IoR Library IDs accordingly according to its own specific conventions and labels.

Levels 4 to 5 correspond to IT enterprise systems, which are not within the area of interest of this thesis as, in general, it is expected that OT and enterprise IT networks to either totally separated or only connected through a DMZ (demilitarized zone) [81]. However, as every organisation might have an implementation of these zones, which might differ from the Purdue model, the separation between levels might not be that rigid. Hence, Figure 5.1 also includes a variant in which the Manufacturing zone (Level 3) could be added to the scope, as a possible extension. In the same way, the model could be extended to include IoRs based on the techniques of the ATT&CK

Enterprise matrices. This extension would be useful for cases when among the risks it is considered that there is a possibility of an attacker using the enterprise network to jump to the OT environment or vice-versa.

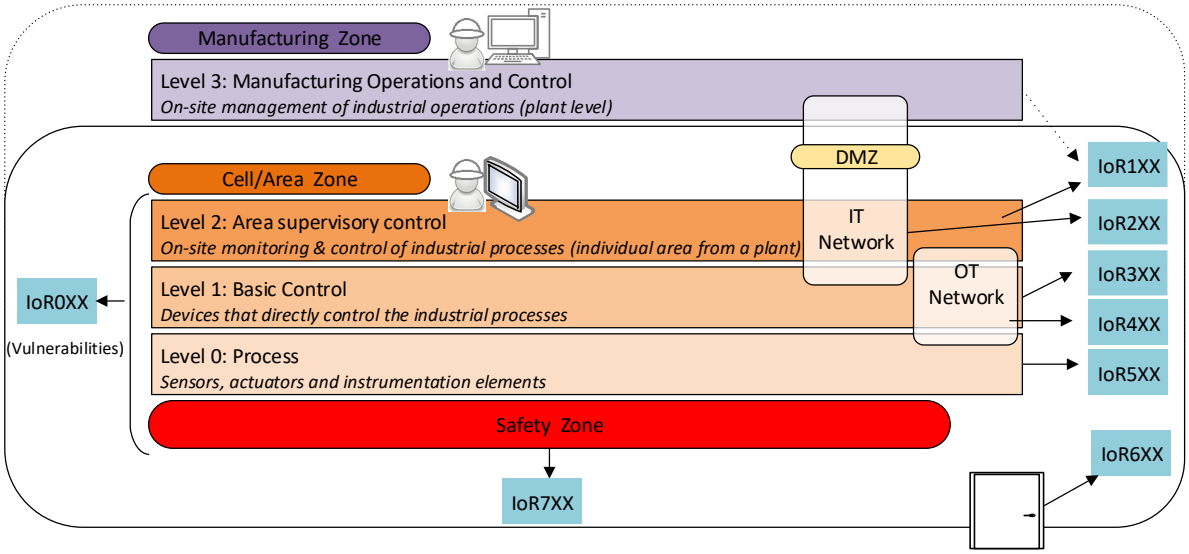



*Figure 5.2: Relating the IoR scheme to the Purdue model*

The relationship between IoRs and Techniques is summarised by an IoR-Technique matrix and described further for each Technique in a separate Technique page. It is proposed that a table be added to each ATT&CK Technique entry, which lists the applicable IoRs, explaining how each one of them is linked to the Technique, and gives an indication of the evidential strength of the IoR for this Technique. This evidential strength, or "degree of influence", can ultimately be translated into conditional probabilities. The degree of influence is an integer from 1 to 5 that suggests how much influence an IoR can have on estimating the risk of a Technique. This can also be used to represent the influence that a technique or tactic has on the likelihood of another technique being performed, as will be shown later in the Bayesian Network example. Each integer in the scale represents a 20% range of conditional probabilities. The default value is 1 (0 to 20%), used when there is not enough evidence for a higher probability.

### 5.2.2. IoR group descriptions

For each IoR Group, different types of IoRs were defined based on possible means of detecting risky conditions and indication of an increased likelihood of adversary activities. This includes pre-conditions, such as vulnerabilities that can allow a technique's execution, and post-conditions that indicate abnormal behaviour or malicious activities already in progress. Examples of post-conditions include the existence of types of log entry, and abnormal behaviour of various variables of the ICS system such as I/O data from sensors and actuators.

#### 5.2.2.1. Vulnerabilities

This group of IoRs represents pre-conditions that could increase the vulnerability level, and hence the risk of a cyber-incident. These IoRs can be found across all levels and include technical vulnerabilities such as OS, software, firmware, and hardware vulnerabilities, insecure configuration and non-compliance with a security policy. These IoRs mostly represent transitional states or states that are subject to change, otherwise they would present a permanent source of risk rather than one that appears under certain conditions or during a pre-defined period of time. For example, an IoR corresponding to a known software vulnerability can be triggered when this vulnerability is discovered and will be reset to a "false" state when the vulnerability is patched. Because software patching might be done only during maintenance windows the increment to the estimate of the risk level caused by this IoR would be sustained over the window of time in which the vulnerability can be exploited. It is also possible that during maintenance periods the system has an insecure configuration or devices are set in operating modes such as 'stop' or 'firmware update' that allow the firmware or the control logic to be modified,

or that a remote connection can be enabled temporarily. These conditions increase the vulnerability level of the system by giving adversaries more opportunities of attack.

The IoRs from the vulnerability group are the following:

- IoR001-Remote accesses enabled
- IoR002-Unnecesary open ports
- IoR003-Unnecesary devices connected
- IoR004-Internet access
- IoR005-Wireless Access Points
- IoR006-Unnecesary privileges
- IoR007-Inapropriate network segmentation
- IoR008-Outdated OS
- IoR009-Default credentials
- IoR010-Known hardware vulnerabilities
- IoR011-Known software vulnerabilities
- IoR012-Known firmware vulnerabilities
- IoR013-Known OS vulnerabilities
- IoR014-Lack of authentication
- IoR015-Insecure authentication
- IoR016-Shared accounts
- IoR017-Anti-malware is disabled
- IoR018- Firewall is disabled

#### 5.2.2.2. IT system threat

The IT system threat IoR group consists of all those IoRs that relate to the ICS servers, workstations, HMIs, and the processes and applications that they run. These IoRs will typically be similar to IoCs, conditions of alert, or SIEM cases commonly found in enterprise systems. However, there would be variations for each ICS environment. For example, blacklisting and whitelisting software applications should, in general, be easier in ICS, since it is expected that only software that is strictly necessary to perform monitoring and control of industrial operations will be installed.

The IoRs from the IT system threat group are the following:

- IoR101-Suspicious account behaviour
- IoR102-Failed login attempts
- IoR103-Failed login attempts followed by a successful one
- IoR104-Files deletion
- IoR105-Control and Monitoring application suspicious access log
- IoR106-Control and Monitoring application suspicious change logs
- IoR107-Unknown programs
- IoR108-Unknown APIs
- IoR109-Unknown files
- IoR110-File suspicious Change logs
- IoR111-Server or workstation suspicious access log
- IoR112-User unable to access monitor and control application
- IoR113-Malware detected
- IoR114-Suspicious command line parameters
- IoR115-Data Historian suspicious access log
- IoR116-E-mail server suspicious activities
- IoR117-Suspicious OPC commands
- IoR118-Unknown processes running on server or workstation

- IoR119-Poor performance of CPU
- IoR120-Suspicious change in memory usage (server or workstation)
- IoR121-OS suspicious event logs
- IoR122-unauthorised changes in project files
- IoR123-Unknown USB device plugged

#### 5.2.2.3. IT Network threat

This IoR group represents standard ICT communications, which would be mostly be based on TCP, UDP and the corresponding application layer protocols such as http, https, TLS, Telnet, etc. In the same way as in the IT system threat IoR group, these IoRs are derived from security monitoring use cases commonly used in enterprise systems. However, the model of normal behaviour should be different in an ICS environment compared to enterprise IT. For example, for automated operations most of the communication traffic should be from the field devices to the workstations since their main purpose is monitoring. Control signals would mostly be sent from the field controllers rather than from the workstations.

The IoRs from the IT Network threat group are the following:

- IoR201-VPN suspicious Access log
- IoR202-High volume of network traffic
- IoR203-Unusually large inbound/outbound packets
- IoR204-Network commands and responses do not match
- IoR205-Unresponded connection requests (port probes)
- IoR206-Unusual or unexpected commands in network packets
- IoR207-Traffic with malicious signature detected
- IoR208-Suspicious communication between devices
- IoR209-Unknown device connected to the IT network

#### 5.2.2.4. Field device threat

This IoR group refers to conditions that can be detected in field devices such as controllers, PLCs, smart relays, gateways, interfaces, and motor drives, which can give indications of a possible threat. Suspicious events related to these devices can be particular types of access or change log entry recording atypical conditions, unexpected modes or states of devices, or unexpected behaviour or responses. The IoRs from this group will require specialised ICS security monitoring tools that offer the advantage of exposing levels of the system that conventional IDS cannot monitor. However, some legacy devices do not have the appropriate ports to connect to ICS security detection tools, and so implementation of these IoRs may not be always possible for these devices.

The IoRs from the field devices threat group are the following:

- IoR301-Controller suspicious access log
- IoR302-Controller program change logs
- IoR303-Controller settings change logs
- IoR304-Controller/device in firmware update mode
- IoR305-Controller in stop mode
- IoR306-Suspicious change in memory usage (field device)
- IoR307-Process state information unavailable
- IoR308-Firmware update
- IoR309-Frequency Increase of Trouble Calls from a Machine
- IoR310-Machine Shuts Down During Normal Operations
- IoR311- Abnormal Process Variable Data Is Transmitted to the PLC

#### 5.2.2.5. OT Networks threat

This IoR group refers to communications that take place using OT protocols such as Modbus, Bacnet, or Profibus, among many others. Modelling a baseline profile for normal patterns of OT communications should be possible, since communications between field devices should follow stable and predictable patterns of behaviour, hence, anomalies could be detected and identified as IoRs. The IoRs are based on inputs from specialised ICS security monitoring tools. As in the case of the Field device threats they offer the advantage of exposing levels of the system that conventional IDS cannot monitor.

The IoRs from the OT Network threat group are the following:

- IoR401-Unknown device connected to the OT network
- IoR402-High volume of network traffic
- IoR403-Commands and responses do not match
- IoR404-Unresponded connection requests (port probes)
- IoR405-Unresponded commands
- IoR406-Unexpected command sequence over network
- IoR407-Commnication port blocked
- IoR408-Unusual connection between devices
- IoR409-Delay or timeout between connections
- IoR410-Maximum of connections exceeded
- IoR411-Use of unusual communication protocol
- IoR412-Communication through unused ports
- IoR413-File transfers between devices
- IoR414-Abnormal OT communication

#### 5.2.2.6. I/O data threat

Sensor readings, actuator status messages and control signals sent to actuators can provide important information about the system's health. This data is derived from the core functions of an ICS, and hence certain anomalies should raise an alert even if they are not caused by security related issues. It should be also possible to gather I/O data at different levels of the system allowing any inconsistency generated by manipulation of an I/O image or data spoofing to be checked. Sensors and actuators, which are in level zero, communicate with controllers, in level 1, usually through analogue electrical signals and sometimes through digital communication protocols, and controllers forward this information to HMIs and workstations in level 2. In case of inconsistencies, the most reliable data should be that directly obtained from the sensors and actuators in the field, and the data most likely to have been manipulated, that obtained from HMIs and workstations. In ICS it is also possible to check consistency of redundant measurements, which is a practice also carried out for safety reasons, and also of variables that are correlated either as a consequence of laws of physics , such as rotation speed in an electrical motor and electrical current consumption, or temperature and pressure.

The I/O IoR group covers an aspect of the ICS system that is mostly not considered by security monitoring tools, even by those that are specialised for ICS network security. Hence, monitoring these variables offers the advantage of covering a "blind spot" in the system and enabling a holistic view of its different variables at all levels. As this is not a typical indicator used in cyber-security, some practical examples of sensor-based anomaly detection models are developed in Chapter 6 for a better understanding of this IoR group.

The IoRs from the I/O data threat group are the following:

- IoR501- Sensor data out of limits
- IoR502- Misbehaviour in sensor data
- IoR503- Indicator that can be correlated to a critical input out of limits
- IoR504- Misbehaviour in data that can be correlated to a critical input misbehaviour
- IoR505- Actuator data out of limits
- IoR506- Misbehaviour in actuator data

- IoR507- Data that can be correlated to a critical output out of limits
- IoR508- Misbehaviour in data that can be correlated to a critical output
- IoR509-Sensor data unavailable
- IoR510- Actuator data unavailable
- IoR511-Inconsistency between different sources of data

#### 5.2.2.7. Physical security threat

In many ICS, security relies to great extent on securing the perimeter against intruders, since if an adversary can access a secure area they can perform several attack techniques, including side channel and physical attacks, which would not otherwise be possible. In addition, many field devices have physical interfaces with weak authentication, and vulnerabilities such as hardcoded keys. Physical access to them would allow these weaknesses to be exploited so as to manipulate of settings, parameters and control rules, or even disabling devices. Usually physical security and access control would be handled by a system that is independent of the ICS. The IoR library contains, as examples, a limited number of IoRs of this type that can provide information indicative of an increased risk of violation of the perimeter. However, more specialised knowledge on perimeter security and security systems would be necessary to have a more complete overview of the possible IoRs in this category. Additional IoRs can be added depending on the specific access security system used in an ICS installation and the specific physical security requirements.

The IoRs from the physical security threat group are the following:

- IoR601- Suspicious Physical Access log entry
- IoR602-Physical Intrusion alert
- IoR603-Misbehaviour or malfunction in Access Control System
- IoR604- Signal from security cameras is interrupted or disabled.

#### 5.2.2.8. Safety threat

Safety Instrumentation Systems are usually independent of the main ICS and may duplicate some of the functionality of the main control system in order to ensure any system failure occurs in a safe way. An example of this is the provision of additional pressure relief valves that would be activated if the pressure of a pipeline exceeds the maximum permitted value. As in the case of physical security, the IoR library contains a limited number of general examples of safety-related IoRs which can be extended for each specific cases and for the safety requirements of each particular system. The IoR Library could benefit from contributions based on specialised knowledge from the safety field.

The IoRs from the safety threat group are the following:

- IoR701-Alarm Historian anomaly
- IoR702-Anomaly in Protection Relay
- IoR703-Change logs in Safety Instrumented System
- IoR704-Anomaly in Safety Instrumented System
- IoR705-Change in alarm thresholds

### 5.3. Method used to build the IoR library

In the initial stage of this thesis work IoRs were identified based on general knowledge about typical IoCs, detection use cases, and system variables that might provide indication of a risky condition. In general, any detectable condition that could imply an increased likelihood of one or more cyber-security risks was considered to be a potential IoR. This included both pre-conditions, including vulnerabilities and insecure configurations, and other non-deterministic conditions that could be observed that could indicate the possibility of malicious attempt. Having built an initial list of IoRs, a systematic method was developed to build and populate a matrix to link techniques from the ICS ATT&CK knowledge base to the IoRs. This method was mostly based on reviewing the description of the techniques from ICS ATT&CK, their associated procedures, data sources, and references,

as well as additional references. This also allowed identification of more IoRs. This iterative exercise also included a peer review from security professionals who checked and gave feedback on the IoR Library and contributed suggestions and examples.

The rows of the Technique-IoR matrix correspond to techniques and the columns to IoRs. Each Technique-IoR coordinate has either an empty space, if the IoR and the technique are not related, or an integer between 1 and 5 if they are. This number represents the degree of influence, which suggests how reliably an IoR predicts a Technique. The degree of influence can also apply for representing how much a Technique or Tactic can affect the risk of other Technique to be performed. The IoR Library provides also for each Technique a list of Techniques and Tactics that could precede it. Each integer value represents a range of conditional probabilities, in which 1 corresponds to 0-20%, 2 to 21-40%, 3 to 41-60%, 4 to 61-80% and 5 to 81-100%. The default value of 1 (0 to 20%) is used when there is not enough information to argue a higher conditional probability. Users of the IoR library can modify the value of the degree of influence according to the information that they have available at a given time and their own degree of belief regarding the probability levels.

### 5.3.1. Description

Figure 5.3 describes the methodology used to build the IoR library, which was an iterative and incremental process. This method also allows room for the IoR library to continue to be enhanced over time. The process consisted of reviewing a Technique from the ICS ATT&CK framework, identifying applicable IoRs continuously incrementing the IoR list, and improving the IoR definitions. It is important that IoRs identified for a particular Technique, should apply directly to the Technique itself, and not indirectly via pre-conditions that depend on other Techniques. If the Technique depends on other Tactics or Techniques (e.g. from earlier stages in the kill chain) these Tactics or Techniques should be used as parent nodes.

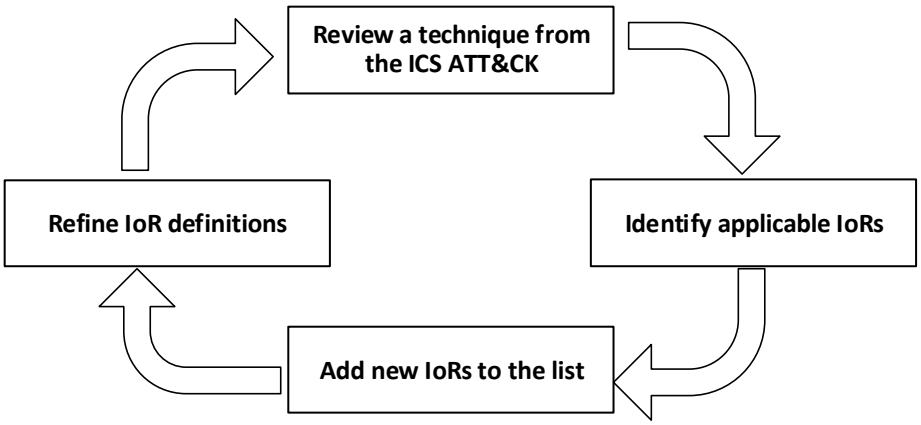



*Figure 5.3: Method used to build the IoR Library*

The following is a description of the method in four steps:

**Step 1: review a Technique from the ICS ATT&CK**
The ICS ATT&CK framework offers as a minimum, a description and references for each technique. In most cases, it will additionally provide other fields such as data sources, assets, procedure examples, and mitigations. In the enterprise knowledge bases of ATT&CK, detection strategies for techniques are also provided, which at the time of building the IoR library were not available for the ICS techniques. Hence, this information, which would have been useful for identifying IoRs, was extracted from references or inferred from the Technique description, and the IoR rationale documented in the IoR library.

**Step 2: identify applicable IoRs**
Based on the review of each Technique, the IoRs already listed that could be related to the Technique were identified, and the rationale documented. If there were specific procedures or malware that could be associated

with the Technique, then IoRs related to signature based detection were also considered (for example: IoR113-Malware detected). The degree of influence that each IoR has on the Technique under review was established, using in most cases the default value, which is the lowest. The reason for this is that it is expected that, in most cases, a significant increment on the estimated likelihood of a risk would be a result of several IoRs showing up at the same time, rather than caused by a single IoR. Exceptionally, in some cases, higher degrees of influence were assigned, when it was considered that a single IoR could allow prediction of a Technique with higher accuracy. In any case, it shall be considered that the degree of influence provided in the IoR library is only a suggestion and it is up to the user to adjust its value according to the particular implementation of the IoR library.

**Step 3: add new IoRs to the list**

If during the review of a Technique, any additional IoRs not yet listed were identified, they were added to the list and Techniques previously reviewed were checked in case the recently added IoRs could apply to them, too.

**Step 4: refine IoR definitions**

The definition of IoRs is an iterative process, since during the review of each technique, in addition to new IoRs, new specific uses of existing IoRs can be found. Each IoR of the IoR library is explained broadly in terms of a rationale, observations, and examples. Additionally, for each Technique, a more specific explanation is given about why, how, or in which context, this IoR is applicable.

### 5.3.2. Example

To illustrate better the method used to build the IoR library, the Technique "T0868-Detect Operating Mode", which belongs to the "Collection" Tactic, was chosen.

**Step 1: review T0868-Detect Operating Mode**

According to the ICS ATT&CK knowledge base, Detect Operating Mode [146] includes gathering information about a PLC's or controller's current operating mode. Knowing this state can allow an adversary to know if they could reprogram the controller at a given moment. Figure 5.4 shows a screenshot of the entry on the ICS ATT&CK knowledge base for this Technique, highlighting the sources of information used to infer IoRs.


*Figure 5.4: Example of ICS ATTA&CK entry for T0868*

**Step 2: identify applicable IoRs**

The identification of IoRs is based on reviewing the Technique including the information marked in Figure 5.4 and some of the references associated. In the Procedure Examples section, there is reference to a known procedure that uses this Technique, which is Triton [147]. Triton contains a file with definitions for program state of Triconex Safety Instrumented System (SIS) manufactured by Schneider Electric. This shows that there is specific known malware associated to this Technique and also that this malware uses a file to identify the states which means that potentially other Procedures could use a similar approach for this Technique. Table 5.2 shows

the list of IoRs that were identified after reviewing the T0868 technique, from the IoR list that was available at that time.

*Table 5.2: IoRs initially identified for T0868*

| IoR | Explanation | Degree of Influence |
|---|---|---|
| IoR017-Anti-malware is disabled | Operating mode might be detected by known malware (e.g. Triton) | 1 |
| IoR018- Firewall is disabled | A firewall can allow filtering traffic stoping malicious traffic | 1 |
| IoR107-Unknown programs | Operating mode might be detected by an unauthorised software or unknown malware | 1 |
| IoR109-Unknown files | Files might contain tables with definitions for states and operating modes (e.g. Triton) or be used to collect the data | 1 |
| IoR113-Malware detected | Operating mode might be detected by known malware (e.g. Triton) | 1 |
| IoR206-Unusual or unexpected commands in network packets | Commands to query on operating mode | 1 |
| IoR311- Abnormal Process Variable Data Is Transmitted to the PLC | Commands to query on operating mode | 1 |
| IoR406-Unexpected command sequence over network | Commands to query on operating mode | 1 |
| IoR411-Use of unusual communication protocol | Commands to query on operating mode can be part of a malware procedure and use a communication protocol that is unusual in a particular environment | 1 |
| IoR412-Communication through unused ports | Commands to query on operating mode can be part of a malware procedure and use ports that are unused in a particular environment | 1 |
| IoR414-Abnormal OT communication | Commands to query on operating mode can result on observable abnormal OT communication patterns, such as increase on frequency of connections or malformed traffic, among others | 1 |

**Step 3: add new IoRs to the list**

While researching possible methods to execute Technique T0868, it was found that certain OPC commands could also relate to this Technique since this protocol defines a specific data type called for Operating Mode Enumeration. Hence, observation of an OPC command attempting to read this data type could reveal a possible attempt to perform this Technique. A new IoR named “IoR117-Suspicious OPC commands” was created and added to the IoR. By keeping this IoR generic enough to describe the detection of any OPC commands that are not normally expected, it was possible to link this IoR to other techniques such as “T0801- Monitor Process State”, “T0802- Automated Collection”, “T0808- Control Device Identification”, and “T870- Detect Program State”.

**Step 4: refine IoR definitions**

Each IoR that was linked to T0868 was reviewed to check that its general definition was consistent with the explanation given for the particular Technique. For example, “IoR109-Unknown files”, is used for Techniques in which a malicious payload might be downloaded through a file or make use of a file as part of a process, which includes the generic use case of using a file for data collection purposes. Using a file for listing operation mode definitions or for collecting operating mode data will correspond, then, to more specific use cases for this Technique.

## 5.4. Method for using the IoR Library

The monitoring of IoRs is intended to enable the development of a near real-time cyber-risk monitoring in ICS. The IoR Library was built with the purpose of facilitating this by having a catalogue of variables common to several industrial systems that can be used as IoRs, with descriptions of how they relate to different adversary TTPs. To implement continuous risk monitoring using the IoR Library it is necessary that each organisation translates the IoRs to a lower level definition that reflects the specific characteristics of their own environment

and context. The method described here is not specific to the continuous management methodology developed in this thesis or to the use of BNs; it explains the use of the IoR Library, regardless the use given to IoRs.

### 5.4.1. Description

The following are the steps defined for the method of using the IoR Library for continuous risk monitoring:

**Step 1: Identify Techniques that are within the scope of the risks being monitored**
The techniques from the ATT&CK framework should be considered in the light of the identified risks and threats, and those that are most relevant to the system of interest for continuous risk monitoring selected. Based on this analysis, it can be decided which TTPs, and particularly, which Techniques from the ATT&CK framework are to be considered within scope.

**Step 2: Look up in the IoR Library for the IoRs that are related to those Techniques**
The IoR Library contains a Technique-IoR Matrix overview that can be used to check which IoRs relate to the techniques within the scope of the analysis. It is possible also to view for each individual Technique, an explanation of the rationale by which the correspondence between the technique and the IoR was deduced. For a more detailed description of each one of the IoRs, it is also possible to check the "Definitions" view. In this step, it is also important to review whether other potentially useful IoRs can be defined that are not mentioned in the IoR library.

**Step 3: Translate the generic IoRs into more detailed instances applicable to the specific ICS context**
As industrial environments can have very diverse types of processes and business objectives, the IoR Library provides high-level IoR definitions, which need to be translated into more concrete instances with specific rules to be implemented in monitoring tools. In the IoR Library examples of one or more ways in which the IoR can be observed or detected are given. Based on these examples, the user shall develop concrete use cases and the corresponding detection rules to observe these IoRs.

To illustrate this step, we can use "IoR201-VPN suspicious Access log", which can be decomposed into a number of lower level detection use cases such as "remote access from foreign country", "login successful after scan attempt", "possible shared accounts", and "VPN sneak attack" [148]. These detection use cases need to be defined at a lower level in order to configure the detection rules in the tools. For example, to detect access from a foreign country it is necessary to whitelist IP addresses from the local country and/or blacklist ones from foreign countries. Another example is "IoR502- Misbehaviour in sensor data", which refers in general terms to an input of the ICS that presents a behaviour that is unusual. Examples include values out of bounds, unusual trends, and oscillations between values that are not commonly observed. A specific instantiation of the means of detecting this IoR, requires identification of the specific sensors to be monitored, such as temperature, pressure, humidity, presence, and definition of a profile of normal behaviour in order to clearly define what "misbehaviour" means. Key questions for this step are "Can this IoR be decomposed into a set of more specific indicators?" and "How can different sub-types of this IoR be applied to different techniques?"

**Step 4: Define which IoRs it is feasible monitor and how can this be done**
Once the IoRs are identified and defined at a lower level, the feasibility of monitoring them continuously should be checked. Depending on the type of IoR it might be possible to forward raw data directly to a SIEM tool and use its rule engine to define and apply detection rules, or to use additional existing tools to monitor the IoRs and forward relevant observations to the SIEM. In other cases, further tools would need to be implemented or developed, which might not be immediately possible. The result of this step should be a final list of low-level IoRs to be monitored and means of performing the monitoring. Key questions for this step are "Can this IoR be monitored and how?" and "For which systems, devices or communication protocols, and in what parts of the network can this IoR be used?"

**Step 5: Review and adjust the degree of influence that each IoR has on each corresponding technique**
As explained earlier, in contrast to security alerts, IoRs are not meant to necessarily warn on an ongoing attack, but instead they provide an indication that the chance of suffering a cyber-incident is becoming more likely. IoRs

should also make it possible to estimate "how much higher is the risk", which is the reason for the degree of influence. When implementing the IoRs it is important to choose a degree of influence that reflects the strength of the indicator as a means of giving hints on risks related to a specific Technique. In the BN implementation, this is expressed as conditional probabilities.

### 5.4.2. Example

To illustrate better the method for applying the IoR library, a set of five Techniques and their corresponding IoRs were used to go through each step of the method. A real scenario would require a higher number of Techniques, however, some of them might share IoRs, which means than adding Techniques will not necessarily mean adding a high number of new IoRs to be monitored. Although different techniques might sometimes have different use cases and rules for a shared IoR, it is likely that they at least can still make use of the same detection tool and method.

**Step 1: Identify techniques that are within the scope of the risks being monitored**

The following techniques were chosen for this example:

- T0803 Block Command Message
- T0804 Block Reporting Message
- T0808 Control Device Identification[2]
- T0813 Denial of Control
- T0814 Denial of Service

Technique T0808 belongs to the "Discovery" Tactic, the Techniques T0803, T0804, and T0814 belong to the "Inhibit Response Function" Tactic, and Technique T0813 to the "Impact" Tactic. All of the Techniques chosen are related to each other and could be used in an orchestrated way to perturb and interrupt the normal performance of industrial operations.

**Step 2: Look in the IoR Library for the IoRs that are related to those techniques**

A total of 40 IoRs were related to these Techniques, as shown in Table 5.3. Most of the IoRs apply to more than one Technique.

*Table 5.3: Example of Technique-IoR mapping*

| IoR / Technique | T0803 | T0804 | T0808 | T0813 | T0814 |
|---|---|---|---|---|---|
| IoR002-Unnecesary open ports | 1 | 1 | 1 | | |
| IoR010-Known hardware vulnerabilities | | | | | 1 |
| IoR012-Known firmware vulnerabilities | | | | | 1 |
| IoR017-Anti-malware is disabled | 1 | 1 | 1 | 1 | 1 |
| IoR107-Unknown programs | 1 | 1 | 1 | | 1 |
| IoR109-Unknown files | 1 | 1 | 1 | | |
| IoR112-User unable to access monitor and control application | | | | 1 | 1 |
| IoR113-Malware detected | 1 | 1 | 1 | 1 | 1 |
| IoR117-Suspicious OPC commands | | | 1 | | |
| IoR118-Unknown processes running on server or workstation | | | 1 | | 1 |
| IoR119-Poor performance of processes or high CPU resource consumption | | | | | 1 |
| IoR120-Suspicious change in memory usage (server or workstation) | | | | | 1 |
| IoR203-Unusually large inbound/outbound packets | | | | 1 | 1 |
| IoR204-Network commands and responses do not match | 1 | | | 1 | |
| IoR206-Unusual or unexpected commands in network packets | 1 | 1 | 1 | | 1 |
| IoR207-Traffic with malicious signature detected | 1 | 1 | 1 | 1 | 1 |
| IoR304-Controller/device in firmware update mode | | | | 1 | |
| IoR305-Controller in stop mode | | | | 1 | |
| IoR307-Process state information unavailable | | | | 1 | 1 |

[2] In April 2021 this Technique was deprecated as its content has been merged into Remote System Information Discovery and Remote System Discovery Techniques.

| | | | | | |
|---|---|---|---|---|---|
| IoR308-Firmware update | | | | | 1 |
| IoR311- Abnormal Process Variable Data Is Transmitted to the PLC | | | | 1 | 1 |
| IoR402-High volume of network traffic | | | | 1 | 1 |
| IoR403-Commands and responses do not match | 1 | | | 1 | |
| IoR404-Unresponded connection requests (port probes) | | | 1 | | |
| IoR405-Unresponded commands | 1 | | | 1 | 1 |
| IoR406-Unexpected command sequence over network | 1 | 1 | 1 | | 1 |
| IoR407-Communication port blocked | 1 | 1 | | | |
| IoR410-Maximum of connections exceeded | | | | | 1 |
| IoR411-Use of unusual communication protocol | | | | | 1 |
| IoR414-Abnormal OT communication | | | | 1 | 1 |
| IoR501-Sensor data out of limits | | 1 | | | |
| IoR502-Misbehaviour in sensor data | | 1 | | | |
| IoR503-Indicator that can be correlated to a critical input out of limits | | 1 | | | |
| IoR504-Misbehaviour in data that can be correlated to a critical input misbehaviour | | 1 | | | |
| IoR505-Actuator data out of limits | | 1 | | | |
| IoR506-Misbehaviour in actuator data | | 1 | | | |
| IoR507-Data that can be correlated to a critical output out of limits | | 1 | | | |
| IoR508-Misbehaviour in data that can be correlated to a critical output | | 1 | | | |
| IoR509-Sensor data unavailable | | 1 | | | 1 |
| IoR510-Actuator data unavailable | | 1 | | | 1 |

**Step 3: Translate the generic IoRs into more detailed instances applicable to the specific ICS context**
In this example, the 40 IoRs were reviewed to identify lower level indicators and use cases for each Technique. The rationale and examples for each indicator and explanations of how they relate to each techniques can be found in [144], the following are examples of this step for eight of the IoRs:

- **IoR002-Unnecesary open ports:** Ports can be opened to prevent other processes accessing them, which is related to techniques T0803 and T0804. An example of this is Industroyer, which uses one COM port and opens the rest to prevent other processes accessing them. This results in the blocking of command and reporting messages [149]. Ports can also be opened by a malicious procedure, which is linked to T0808. A lower level, observations for this IoR can be that network security monitoring or port scan tools indicate that a typically unused port is open or logs indicating that a typically unused port has been enabled and not subsequently disabled. If the action of opening the port cannot be related to a legitimate process, there is a risk that it was done as part of a malicious procedure related to T0803, T0804, or T0808.
- **IoR010-Known hardware vulnerabilities and IoR012-Known firmware vulnerabilities:** Hardware vulnerabilities can allow the bypassing of authentication and security functions such as secure boot, which can facilitate an adversary accessing devices and performing a DoS attack. Additionally, several OT devices such as controllers, gateways and switches have known firmware vulnerabilities that can allow a DoS attack to take place. As OT devices are expected to have a lifespan of many years, and may not be updated during that time, they will remain exposed to exploitation of vulnerabilities that are discovered and published while they are in operation. Organisations will have to apply workarounds and mitigations or else simply accept the risk. While hardware vulnerabilities, in general, cannot be patched, patches are made available by ICS vendors for some firmware vulnerabilities. This will depend on the age of the device since older generation devices are less likely to be patchable. However, even when a firmware security patch is available, in many industrial operations, security updates and other changes will only be done during maintenance windows, which for 24x7 processes might be once every several months. Observations for these IoRs might come from an integrated feed from a vulnerability database, or a threat monitoring platform or manually introduced into the continuous monitoring system when an ICS vendor publishes an advisory warning of a hardware or firmware vulnerability in a device used in the system that can lead to a DoS condition.

- **IoR107-Unknown programs**: continuous risk monitoring should contemplate both the possibility of threats based on malware that is already known and hence, likely to be detected by signature based malware detection systems and on zero-day malware. Another case is to be infected with relatively new malware before the antimalware software being updated. As programs used in an ICS environment are limited, any unknown program detected could be considered suspicious until it is properly validated. This IoR can be related to several ATT&CK techniques. Examples of a concrete observation that can trigger this IoR is an attempt to install software that is not on a whitelist or an unknown process being executed on a workstation machine.
- **IoR109-Unknown files:** as in the case of unknown programs, unknown files can be related to many Techniques, since they can contain a malicious payload or used to assist a malicious procedure. For example a file can contain tables referring to ports and commands, in which case it is related to T0803 and T0804, or be used to collect data for which is also related to T0808. Concrete observations that can trigger this IoR are an increase of the number of files in a given directory or logs indicating that a new file was created.
- **IoR112-User unable to access monitor and control application:** a legitimate user not being able to access a control and monitoring application could be due to technical problems that are not security related, or as a consequence of a DoS attack or a Ransomware attack. In either case, this can have an adverse impact on normal operations and should be reported and fixed. In the case that this condition is caused by a malicious attack, it may be that preventing access is not the primary goal, but is used as a distraction or to avoid operators detecting an undesirable or risky condition in the physical processes themselves and to prevent them to take prompt actions. For this reason, this IoR is not only related to a DoS (T0814) but also to a Denial of Control (T0813). This IoR can be triggered by a user reporting this issue to IT support or directly in the continuous risk monitoring system.
- **IoR113-Malware detected:** several of the techniques described in the ICS ATT&CK framework are related to known malware that is generally listed in the technique's procedure examples. For example, Industroyer is related to all 5 techniques used in this example. Technique T0808 can also be performed by Backdoor.Oldrea (Havex), Triton, and VPNFilter. In this case, each detectable malware species should be represented by an individual lower level IoR that should be linked to all the corresponding techniques that the malware can perform. The trigger for each of these IoRs is generated by signature based detection tools (e.g. anti-malware software).
- **IoR117-Suspicious OPC commands:** this IoR is related to techniques that can be performed through OPC commands. Examples of how to trigger this IoR are the use of OPC commands that are blacklisted or not whitelisted or the use of certain OPC commands with an unusually high frequency or under unexpected conditions. For this IoR a number of lower level IoRs can be derived relating different OPC commands to different techniques. An example of a procedure in which this IoR can be related to T0808 is the one used by Industroyer, which has an OPC Payload used for identifying the system's components.

**Step 4: Define which IoRs it is feasible to monitor and how**

In Step 3, examples of concrete observations that can allow detection of conditions that would trigger an IoR were given. Followed on from this, in Step 4 reviews whether the appropriate tools and methods for gathering and processing the information of these observations are in place or are feasible to implement. The result of this step is the final list of IoRs to be monitored, for example, from the list of 40 IoRs of this example, it may be decided to implement a sub-set. The selected IoRs could be those that are considered most relevant among the ones that are feasible to implement. The selection can also be based on a mix of several criteria such as how many Techniques the IoR relates to. Table 5.4 shows an example of a selection of 19 IoRs. Those that are marked with a star (*) are the ones that require different implementations depending on the Technique.

*Table 5.4: Example of IoR selection*

| IoR / Technique | T0803 | T0804 | T0808 | T0813 | T0814 |
|---|---|---|---|---|---|
| IoR002-Unnecesary open ports | 1 | 1 | 1 | | |
| IoR017-Anti-malware is disabled | 1 | 1 | 1 | 1 | 1 |
| IoR107-Unknown programs | 1 | 1 | 1 | | 1 |
| IoR112-User unable to access monitor and control application | | | | 1 | 1 |
| IoR113-Malware detected* | 1 | 1 | 1 | 1 | 1 |
| IoR117-Suspicious OPC commands | | | 1 | | |
| IoR118-Unknown processes running on server or workstation | | | 1 | | 1 |
| IoR119-Poor performance of processes or high CPU resource consumption | | | | | 1 |
| IoR304-Controller/device in firmware update mode | | | | 1 | |
| IoR305-Controller in stop mode | | | | 1 | |
| IoR308-Firmware update | | | | | 1 |
| IoR402-High volume of network traffic | | | | 1 | 1 |
| IoR403-Commands and responses do not match | 1 | | | 1 | |
| IoR405-Unresponded commands | 1 | | | 1 | 1 |
| IoR407-Communication port blocked | 1 | 1 | | | |
| IoR411-Use of unusual communication protocol | | | | | 1 |
| IoR501-Sensor data out of limits | | 1 | | | |
| IoR503-Indicator that can be correlated to a critical input out of limits | | 1 | | | |
| IoR509-Sensor data unavailable | | 1 | | | 1 |

**Step 5: Review and adjust the degree of influence that each IoR has in each corresponding technique**
In Table 5.4 it can be observed that in most cases the default degree of influence, which is 1, is suggested. This means that if the IoR is triggered, the probability of the technique being used by an adversary is in the range 0 to 20%. Depending on the continuous risk monitoring method used, this can later be translated to a conditional probability within this range (e.g. 8%). For each implementation, the suggested degrees of influence can be modified. For example, "IoR509- Sensor data unavailable" has a default degree of influence of 1 (between 10% and 20%), but it could be adjusted to 2 (between 20% and 40%) for "T0804 Block Reporting Message" while keeping it as 1 for "T0814 Denial of Service". Cyber-security risk analysts can decide, for example, to assign a 10% probability to the unavailability of sensor data being caused by a blocked reporting message, if they consider that this IoR is triggered by other conditions on 90% of occasions.

## 5.5. A Bayesian Network template based on the IoR library

In the worked example developed in chapter 4, for 2 risks, several TTPs were identified including a relatively small number of Tactics, Techniques, and IoRs. In a real life implementation, it is expected that a larger number of risks will need to be monitored, which are based on a combination of a higher number of possible adversary Techniques. This introduces potential challenges regarding the scalability of the approach. The IoR Library partially addresses this issue by providing guidance on identifying and selecting IoRs and linking them to known TTPs. However, implementation of a risk calculation algorithm for continuous risk monitoring based on IoRs requires additional work, which could be assisted by pre-built templates. In the Continuous Risk Assessment methodology proposed in this thesis, BNs are used as means of calculating probabilities and, subsequently risks. A BN template has been built based on the IoR Library to make the implementation of continuous risk monitoring more efficient. This reusable template corresponds to a big network that includes all the Techniques, Tactics, and IoRs from the IoR Library and their relationships. After following the 5 step method for using the IoR Library, additional steps can be added to build the BN that is used for continuous risk monitoring.

### 5.5.1. Conceptual Model

The BN template based on the IoR Library includes all the Tactics, Techniques and IoRs contained in it, which requires over 400 interconnected nodes. To make the template easy to use and understand, this network was

built as interconnected sub-models, which allows a compartmentalized view and makes finding individual Techniques and IoRs easier. Figure 5.5 describes a conceptual model of the BN template in which the highest level of sub-model corresponds to Tactics. Each Tactic sub-model contains its corresponding Tactic node and one or more Technique sub-models. In the same way, each Technique sub-model has its corresponding IoRs as child nodes and the Technique itself as parent node. Additionally, it is possible that Techniques are also connected, either within the same Tactic or across different Tactics.

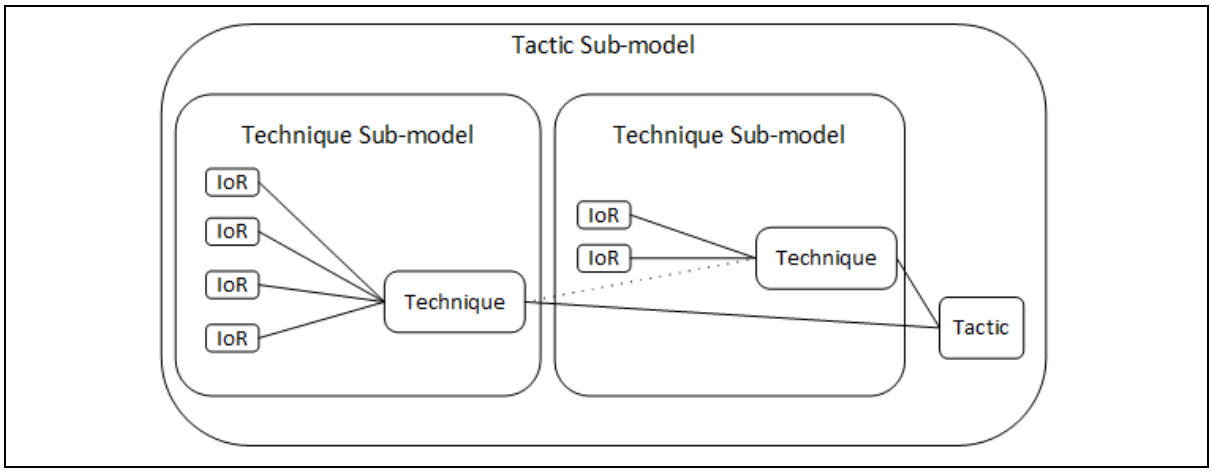



*Figure 5.5: Conceptual model of the BN template*

Figure 5.6 shows the overview of the main model of the BN Template, which contains all the adversarial Tactics from the ATT&CK framework. Each one of these Tactics is a sub-model that contains adversarial Techniques. A Tactic sub-model depends on another Tactic sub-model if any item contained on one sub-model depends on an item contained in the other.

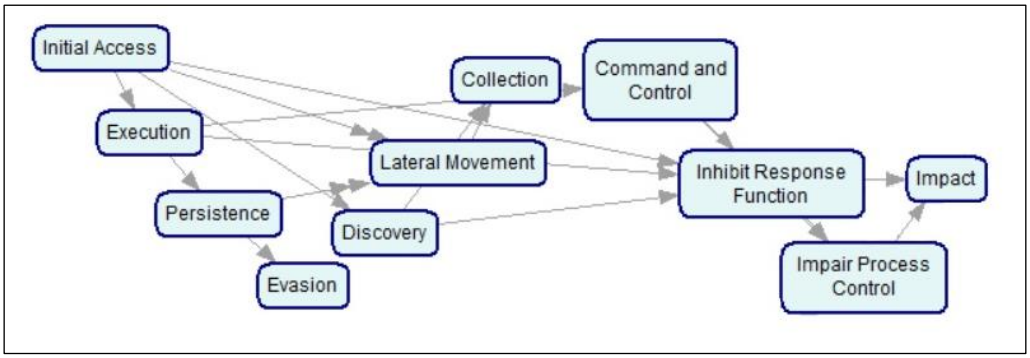



*Figure 5.6: Main Model of BN Template*

Figure 5.7 shows the "Impair Process Control" Tactic sub-model, which contains four Technique sub-models, each one of which is connected to the Tactic node through the Technique node contained in the corresponding sub-model. Finally, in Figure 5.8 an example of a Technique sub-model can be found, in which the IoRs that are linked to this Technique in the IoR Library can be observed.

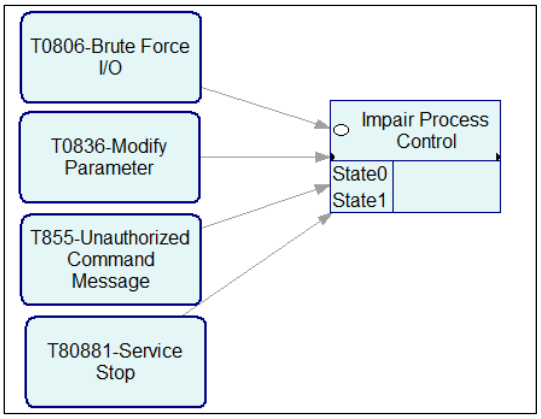



*Figure 5.7: Example of Tactic sub-model*

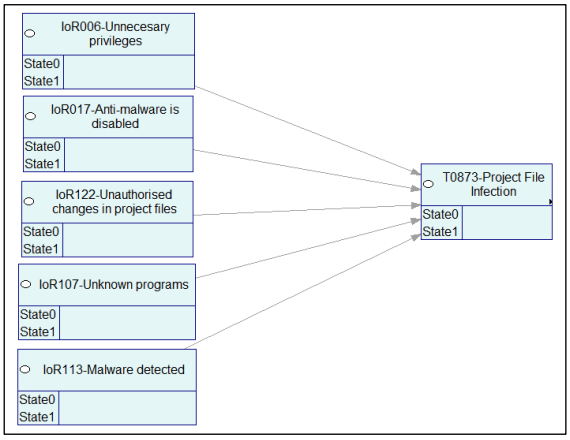



*Figure 5.8: Example of Technique sub-model*

### 5.5.2. Description of the method to use the Template

The BN template is an instantiation of the conceptual model with all ATT&CK-ICS Tactics and Techniques and all IoRs in the Library, which is then edited to obtain a BN structure to use. It is intended as a resource to help an organization to build its own BN for risk probability calculations based on IoRs. Prior to building the BN, it is necessary to identify and select the relevant Techniques and IoRs, for which the 5 step method suggested in section 5.4 can be used. Following this thread, the method for using the BN template starts with Step 6.

**Step 6: select the IoRs, Tactics, Techniques in the scope from the BN template**
Take a copy of the BN template to use as a starting point. Erase all the nodes that correspond to Techniques and Tactics out of scope. As explained in Chapter 4, the scope is defined in the context establishment process. Check the remaining Techniques and Tactics and their corresponding IoRs and erase all IoRs which are not in the list built in Step 4.

**Step 7: add new IoRs, derive specific IoRs from the generic ones, and erase redundant IoRs**
In Step 2 it was recommended that the possibility of defining new IoRs that were not included in the IoR Library should be considered. If new IoRs were defined they should be added to the BN by creating new nodes. Following this, it should to be checked whether in Step 3 it was found necessary to derive more specific IoRs from the generic ones from the IoR Library (which are the ones included in the template). For example, “IoR113-Malware detected”, can be replaced by a number of IoRs related to specific malware for each related Technique. Additionally, further IoRs for types of malware can be defined such as an IoR for Trojan, another IoR for RAT (Remote Access Trojan), and a different IoR for Ransomware. Having defined all the IoRs at a lower level, each of these specific IoRs that apply to several Techniques should be represented only once in the BN model. Hence,

redundant IoR nodes should be eliminated and each one of these shared IoR nodes should appear only once and be linked to all the Technique nodes that apply.

**Step 8: add extra nodes, if necessary**
Sometimes, the TTPs, as defined in the ICS ATT&CK might not allow having visibility of all the events that need to be monitored. Thus, if it is considered necessary, additional nodes can be added, for example, in the worked use case presented in Chapter 4, the BN had two intermediate nodes that were parents of the "Initial Access" Tactic, which were "Remote Access" and "Local Access". Nodes can also be added at the end to reflect the impacts that are specific to the industrial process including operational and business impacts.

**Step 9: adjust conditional probabilities**
Assigning conditional probabilities is probably the most laborious steps of all, since conditional probability tables for a node with several parent nodes (e.g. a technique with several IoRs) can be quite extensive. A probability table in which nodes have only 2 states (true or false), such as the ones used as an example in Chapter 4, will imply $2^n$ probabilities to be assigned as, where "n" is the number of parent nodes. This can get more complicated if some nodes are defined as multi-valued states. Hence, it is beneficial to have templates in which default conditional probabilities are previously assigned. This would make this task scalable to a BN with a large number of nodes and also for nodes with multiple states, since manual modifications would be done only to the specific probabilities that require them. These default probabilities can be based on the degree of influence proposed in the IoR Library.

### 5.5.3. Example 1 of applying the method

To illustrate better the method to use the BN template of the IoR Library, the same Techniques used in section 5.4.2 were chosen in order to give a sense of continuity to the example. They are shown in Table 5.5 along with the Tactic to which they belong.

*Table 5.5: Techniques within the scope of Example 1 and their corresponding Tactics*

| Technique | Tactic |
|---|---|
| T0803 Block Command Message | Inhibit Response Function |
| T0804 Block Reporting Message | Inhibit Response Function |
| T0808 Control Device Identification | Discovery |
| T0813 Denial of Control | Impact |
| T0814 Denial of Service | Inhibit Response Function |

**Step 6: select the in-scope IoRs, tactics and techniques from the BN template**
A copy of the BN template was made and all the out of the scope Tactics were erased, leaving only the sub-models corresponding to Discovery, Inhibit Response Function, and Impact. Each of these three Tactic sub-models was opened to erase all the out of scope Technique sub-models, and subsequently in each Technique sub-model the out of scope IoRs were erased. The remaining sub-models for Tactics and Techniques are shown in Figure 5.9.

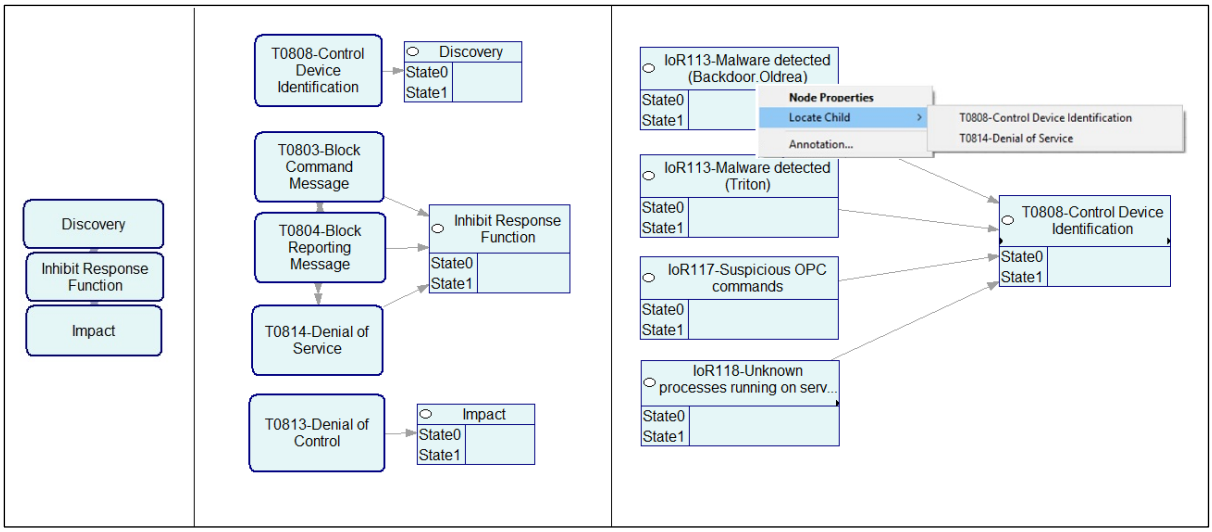



*Figure 5.9: Example of use of BN Template*

**Step 7: add new IoRs, derive specific IoRs from the generic ones, and erase redundant IoRs**

IoRs in common between Techniques were reviewed to identify redundant IoR nodes that could be erased. At the same time the "IoR113-Malware detected" was identified as an IoR that requires to be divided into multiple specific implementations. For this example, this node was broken down as shown in Table 5.6

*Table 5.6: Specific IoR implementations*

| Specific IoR Implementation | Techniques linked |
|---|---|
| IoR113-Malware detected_1 (Industroyer) | T0803, T0804, T0808, T0813 |
| IoR113-Malware detected_2 (Backdoor.Oldrea) | T0808, T0814 |
| IoR113-Malware detected_3 (Triton) | T0808 |

Figure 5.9 shows some of the details of this exercise divided into three sections. The section on the left hand side shows the Tactic sub-models remaining after eliminating the out of the scope Tactics. The section in the middle shows the three Tactics sub-models which contain the corresponding Tactic node and Techniques sub-models. Finally, the section on the right hand side shows the sub-model for T0808, as an example. Two IoR nodes are shared with other Technique sub-models. To identify the other Techniques linked to these IoRs, the options menu of the node has to be selected and the option to locate the child nodes chosen. This is also illustrated in Figure 5.9 using "IoR113-Malware detected_2 (Backdoor.Oldrea)" as an example which has both T0808 and T0814 as dependent nodes.

**Step 8: add extra nodes, if necessary**

Since this example is only an exercise based on the selection of five random Techniques, only one extra node was added to illustrate this step, which was "Delays in products delivery". This is a made-up business impact, which has as parent nodes T0813 and T0814, as shown in Figure 5.10 In a more elaborate case, more of these extra nodes would be expected in order to reflect the corresponding business impacts, as well as implications at an operational level that are specific to the system.

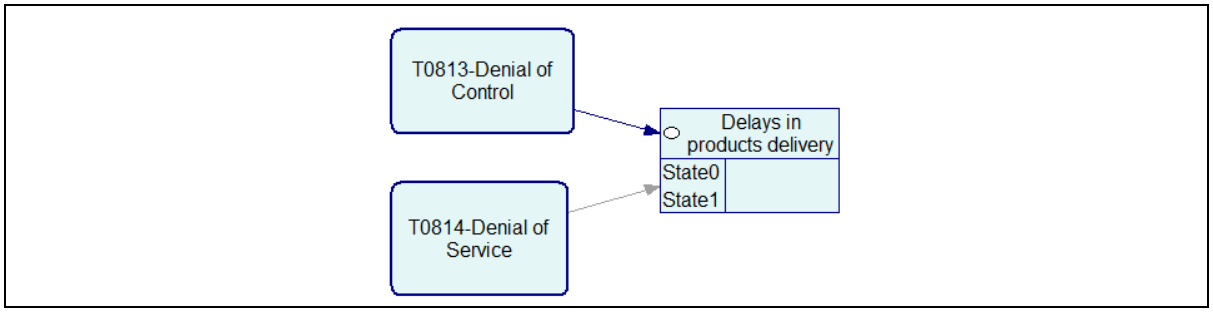



*Figure 5.10: Example of extra node*

**Step 9: adjust conditional probabilities**

Conditional probabilities are adjusted so vulnerability IoRs such as "IoR002-Unnecesary open ports", "IoR017-Anti-malware is disabled", and "IoR018- Firewall is disabled" do not by themselves increase the probability of a Technique significantly, but they do when indicators related to threats are observed. For example, if these three indicators are observed at the same time as "IoR405-Unresponded commands", then the probabilities of T0803, T0813, and T0814 have a considerable increase.

### 5.5.4. Example 2 of the method

A good starting point for an organisation to identify ICS risks is to look into the TTPs that are most often used by adversaries and include them in the scope of their Continuous Risk Assessment. This second example of the method for using the BN template is based on the Techniques identified by Dragos [150] as being among the TTPs most used by threat activity groups targeting ICS in 2020. Table 5.7 shows the Tactics and Techniques within the scope of this exercise based on the activities of two groups discovered in 2020 named as STIBNITE and KAMACITE. As described in Step 6 only the Techniques and Tactics used by STIBNITE and KAMACITE were left in the BN, with all the remaining ones being erased from the BN template. All the IoRs linked to these Techniques in the library were considered to be within scope.

*Table 5.7: Techniques within the scope of Example 2 and their corresponding Tactics*

| Technique | Tactic |
|---|---|
| T0865 Spearphishing attachment | Initial Access |
| T0866 Exploitation of Remote Services | Initial Access |
| T0816 Command Line Interface | Execution |
| T0859 Valid Accounts | Lateral Movement/ Persistence |
| T0885 Commonly Used Port | Command and Control |
| T0869 Standard Application Layer Protocol | Command and Control |
| T0884 Connection Proxy | Command and Control |

Figure 5.11 shows the Tactic view from the resulting BN. Steps 7 to 9 were not performed for this example.

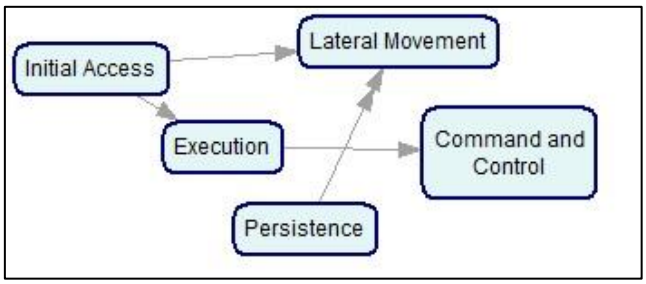



*Figure 5.11: Example of use of BN Template*

## 5.6. Validation of the IoR Library

The first presentation of the idea of using IoRs for ICS continuous cyber-risk monitoring in reference to the ICS ATT&CK framework in the context of this PhD research was done at a MITRE ATT&CK workshop in May 2020 [151]. This idea derived from the need to find a systematic way to identify IoRs, and the release of the ATT&CK ICS matrix of came as a good opportunity to relate the present work on continuous risk management to a well-known and widely used framework. These concepts received attention from the security community, which made it possible to exchange ideas about the potential utility of the IoR Library in follow-up meetings before it was actually built. After the development of the first prototype version of the IoR Library, it was shared with five cyber-security professionals from three major German companies, who checked the indicators and provided feedback, and suggestions, which were incorporated in latest version [144].

IoRs were also cross-checked against experimental ICS security detection use cases from three sources. The first source is the NIST IR8219 [12], which identifies a total of 53 ICS behavioural anomaly detection use cases using four different tools. The second is [105] that proposes an Industrial Intrusion Prevention System concept in which 15 example events corresponding to violation of security policies in ICS are listed. The third is a report on a proof of concept exercise performed in a manufacturing company in 2019 from which the main outcomes were shared under confidentially agreement.

Table 5.7 summarises these checks, indicating the tools used, the number of detection or anomaly scenarios, and the IoRs from the IoR Library [144] that matches them. It should be noted that the relationship between the PoC use cases and the IoRs is not one to one since sometimes one use case was described by more than one IoR and other times an IoR covered more than one use case. As a result, 39% of the IoRs currently in the library were matched to at least one of the practically demonstrated use cases reviewed.

*Table 5.8: Check of IoRs against existing PoCs*

| Ref | Tool | N° of anomaly use cases or scenarios | Related IoRs |
|---|---|---|---|
| BAD | Silent Defense | 15 | IoR004, IoR015, IoR102, IoR202, IoR203, IoR205, IoR302, IoR304, IoR308, IoR413, IoR414 |
| BAD | Secure-NOK SNOK | 15 | IoR004, IoR017, IoR018, IoR107, IoR113, IoR202, IoR203, IoR205, IoR207, IoR302, IoR304, IoR308, IoR408 |
| BAD | CyberX | 15 | IoR004, IoR015, IoR102, IoR113, IoR202, IoR203, IoR205, IoR207, IoR302, IoR308, IoR401, IoR402, IoR405, IoR413 |
| BAD | OSIsoft | 8 | IoR305, IoR309, IoR310, IoR311, IoR501 |
| IIPS | Industrial IPS | 15 | IoR002, IoR011, IoR012, IoR102, IoR113, IoR120, IoR123, IoR302, IoR304, IoR401, IoR408, IoR410 |
| Industry PoC | Silent Defense | 10 | IoR101, IoR102, IoR114, IoR117, IoR205, IoR402, IoR406, IoR408, IoR411, IoR412 |

As part of the validation, challenges that were identified during this research as needing to be addressed for the IoR Library to be applied in practice include the following:

- Library maintenance: Review and improve the current IoR descriptions and examples, add further IoRs, and cover more adversary Techniques.
- Develop structured guidelines: Refine the processes for using the IoR Library, including the process for generation of appropriate IoRs for a particular context.
- Exposure to a wider audience: Get more feedback and contributions.
- Engagement with MITRE: Propose discussions about possible adoption of the IoR Library as an extension to ATT&CK.
- Case studies: Develop use cases in particular contexts to test and drive development of the IoR Library and associated methods.

- Propose a standard format for exchange of IoRs: this would facilitate adoption as well as community contributions.
- Availability of appropriate tools: legacy systems cannot always be monitored by OT security monitoring tools.

### 5.7. Management of IoR updates

An important aspect that needs to be taken into account is the length of time that IoR observations remain valid, and when IoRs should change back to their default state. Some IoR observations can result in a state change of the corresponding IoR node that lasts until another observation affecting the same node is made. For example, if a means of remote access were enabled, this would remain as a source of risk until this access is disabled, or if there were an unpatched component in the system, this would also constitute a source of risk until the component is patched. Hence, in both cases, the corresponding IoR will be in a "true" state, as long as the risky conditions persist. However, in other cases, sources of risk are temporary and their corresponding IoRs should at some point be reset to their default state once they are no longer valid.

IoR states can be reset in two ways depending on the type of IoR. The first way is to manually reset the IoR after its causes are investigated and the risk is dismissed. For example, if unknown files are detected, this would trigger an IoR observation. However, if after investigating the issue it is found that these files were not malicious, then the IoR corresponding to unknown files should be turned back into a "false" or "not observed" state. The second way is to have a pre-defined lifespan for an IoR (e.g. a 'use by' date) after which the corresponding observation is deleted from a cache. If the IoR node states are based on the current contents of this cache they would be updated whenever the cache contents are, causing a "refresh" of the IoR states. This second method could be used, for example, when an unusual event is added to an access log, or when unusually high network traffic is detected. If there are no further consequences of these events, and there are no other IoRs detected after certain period of time, it would be reasonable that these observations no longer constitute an indication of risk, and neither should they be correlated with future IoR observations. It should be noted that the two ways of managing IoR updates are not mutually exclusive. An IoR observation could have a pre-defined expiration period and yet, it could be prematurely reset to its initial state if it is reasonable to assume that there is not currently a risk associated to this IoR. Figure 5.12 shows a schematic of the IoR updates management system just described.

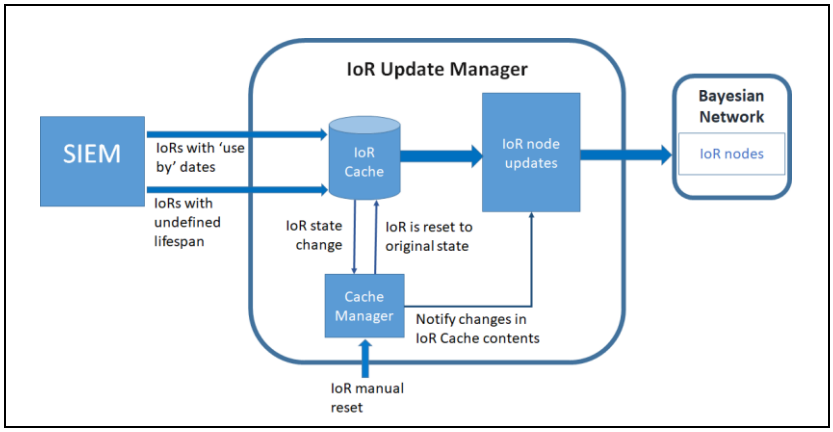



*Figure 5.12: Management of IoR updates*

### 5.8. Unknown security risks

The method suggested in section 5.4 indicates that for using IoRs in continuous risk monitoring, the Techniques within the scope of the risk analysis should be identified. However, it is important that this method also has a way to consider unknown TTPs, such as exploits of zero day vulnerabilities. The IoR Library is based on the ATT&CK framework, which includes, of necessity, only known adversary TTPs. Even if all the possible TTPs from ATT&CK were monitored, there would be always new Techniques and new variations of the existing ones

emerging. Hence, no matter how many TTPs are monitored explicitly, the risk associated with unknown or out of scope TTPs should be also be monitored somehow. This section explains how by connecting IoR nodes to Tactic nodes directly, it can be taken into account a background of implicit and unknown Techniques. This enables IoR-based continuous risk assessment to deal with the presence of undefined risks.

As explained earlier, the MITRE ATT&CK framework defines Tactics as tactical goals that an adversary needs to accomplish to be able to achieve their major objective. The ICS ATT&CK framework already provides a comprehensive overview of possible adversarial Tactics based on evidence collected from a large number of attack cases and reviews by security experts. Hence, it can be assumed that any new threats would correspond to novel Techniques, Sub-Techniques or Procedures associated with one or more existing Tactics, rather than requiring a new Tactic to be defined. Under this assumption, the most common IoRs for Techniques belonging to the same Tactic were reviewed in order to propose a BN model that would calculate the probability of execution of a Tactic, regardless of the specific Technique used. This would allow any Techniques left out of scope and also unprecedented Techniques and Procedures to be taken into account.

Table 5.8 shows the most representative IoRs for five of the ICS ATT&CK Tactics; this table can be also found in the IoR Library with the complete name of each IoR for ease of reference *[144]*. A BN Template with direct links from IoRs to Tactic nodes was also implemented by means of which IoRs to be monitored can be selected from the preloaded alternatives and new IoRs can be added.

*Table 5.9: IoRs grouped by Tactic*

| **Tactic** | **IoRs** |
|---|---|
| Initial Access | IoR001, IoR002, IoR003, IoR004, IoR006, IoR007, IoR008, IoR017, IoR018, IoR101, IoR107, IoR109, IoR113, IoR118, IoR119, IoR120, IoR121, IoR202, IoR209, IoR412, IoR601, IoR602, IoR603, IoR604. |
| Execution | IoR006, IoR017, IoR101, IoR107, IoR113, IoR206 |
| Collection | IoR017, IoR018, IoR105, IoR107, IoR109, IoR113, IoR117, IoR118, IoR119, IoR120, IoR206, IoR207, IoR209, IoR301, IoR311, IoR406, IoR411, IoR412, IoR414 |
| Lateral Movement | IoR001, IoR004, IoR008, IoR009, IoR011, IoR013, IoR015, IoR016, IoR017, IoR018, IoR101, IoR102, IoR103, IoR105, IoR106, IoR111, IoR113, IoR121, IoR202, IoR203, IoR207, IoR301 |
| Inhibit Response Function | IoR006, IoR017, IoR018, IoR107, IoR109, IoR113, IoR206, IoR207, IoR405, IoR406, IoR503, IoR504, IoR507, IoR508 |
| Impair Process Control | IoR017, IoR018, IoR107, IoR113, IoR206, IoR207, IoR301, IoR305, IoR309, IoR310, IoR311, IoR414 |

By linking all monitored IoRs directly to their corresponding Tactic node, it is possible to take into account unknown risks as well as known ones. Thus, when a set of IoRs is observed, this would not only result in an increased probability for the risks that are being monitored explicitly, but also for one or more of these Tactic nodes, providing indication that there is also a risk that an unknown or undefined threat is developing.

## 5.9. Non-security related triggers

Some IoRs provide information about conditions that increase the probability of a security incident, however their observation can also relate to non-security related causes. These causes can be divided into two types; one type consists of licit or normal operations, and the other of issues, which are not related to a security incident but can still be considered problems, such as malfunction of a component. A simple example of the first type is when a USB device is connected to a workstation. This by itself does not necessarily represent a threat, but it increases the probability of one, since a USB device can be used to transfer malicious files to the system either on purpose or accidentally, and also to steal sensitive information. If connection of USB devices is allowed, this IoR should not have much significance (low degree of influence) unless it is observed in conjunction with other IoRs. However, if a strict security policy against connecting USB drives is in place, then this IoR could be related with more certainty to a threat (high degree of influence), since it constitutes, at the very least, a violation of an explicit security policy.

As mentioned earlier, the selection of appropriate IoRs and definition of their degree of influence will be up to each organisation and depend on its own ICS implementation, operational and security rules and perception of risk. The general recommendation provided in this thesis is that IoRs that might also be observed due to licit or permitted activities should be assigned a low degree of influence and hence a low conditional probability related to adversary TTPs, especially, when they are observed alone. This probability can increase when it is observed together with other IoRs. Overall, generation of false positive alerts as a result of this type of IoR should be avoided. The following are more specific recommendations on how to deal with this type of IoR:

- Check under which licit conditions the IoR is observed and how often such conditions occur.
- Check whether it is possible to have an automated validation of the triggering action that would cancel the IoR or prevent it to be triggered (e.g. an IoR for connection of removable media would be dismissed if it had been scanned for malware). Rather than cancelling the IoR as a result of such a validation, it may be judged preferable to just reduce the probabilities (because e.g. the removable media can still be used to steal information or have malicious files that are not recognizable by an automated scan, and hence, there are still risks associated to it).
- Check whether this IoR can provide valuable information about the security posture and the related risks in real time. The following questions can help this evaluation:
    - Does it reveal conditions that make the system more vulnerable?
    - Can it reveal a possible threat?
    - Is there another IoR that can provide the same information that might not be observed or would be observed less often when licit activities are performed?
- If this IoR provides valuable information, consider the following rules for handling circumstances in which this IoR is active but no other associated IoR is triggered:
    a. Set conditional probabilities so that the probability increase for affected nodes when only this IoR is active is low enough that no risk is raised over the tolerance level as a result of the influence of this IoR.
    b. Set a warning associated with this IoR to advise ICS operators to validate whether there is a normal activity related to this IoR observation and to check if there is any other related issue.

In the case of IoRs that can be observed also due to a non-cyber-security related system anomaly or malfunction, the general recommendation is to share this information with the responsible for operations and maintenance, since, regardless the root cause of the issue, there is still on the best interest of the organisation to investigate these conditions.

## 5.10. Discussion

Chapter 5 describes the the IoR Library, which is a catalogue of different types IoRs, and how it can be applied to monitor risks in ICS and IIoT. At the time of writing, the IoR Library contains 95 IoRs grouped according to the levels of the Purdue architecture model. Each IoR has a general description that includes rationale, observations, and examples, plus explanations of how it relates to specific adversary techniques. Alignment with the MITRE ICS ATT&CK framework allows using IoRs to extend existing security monitoring approaches adding a more risk-oriented perspective. Additionally, a BN template was built to demonstrate how the IoR Library could be leveraged to help an organisation implement a particular risk calculation method. This was done also in alignment with the methodology described in Chapter 4. The main contributions of this chapter is to provide a resource for continuous risk monitoring that is straightforward to map to known ICS TTPs and that can be used independently of the risk assessment approach adopted. Overall, the IoR Library constitutes a valuable and logical extension of MITRE ATT&CK, and IoRs appear to be a natural step to progress beyond attack Techniques.

The IoR Library was built as a resource that systematizes IoRs and provides guidance on their implementation for continuous cyber-risk monitoring in ICS. Despite being developed with the purpose of aiding the implementation of a BN-based risk calculation method, it can be used with other risk calculation methods or algorithms and also for security risk monitoring. In the same way, it is independent of the Continuous Risk Assessment methodology described in Chapter 4. Hence, the IoR Library can both assist the Continuous Risk Assessment methodology proposed in this thesis and be used as stand-alone product that can help risk and

security analysts to identify useful means of monitoring risks to ICS. The main benefit of the IoR Library is that it helps to identify variables that can be used to monitor risk at all levels of the system. IoR types are not restricted to host-based, endpoint or network detection, nor to a particular level of the system, but apply throughout an ICS or IIoT system. This also includes behavioural anomalies in the OT network and sensors and actuators I/O data.

Additionally, the following potential applications and uses of IoRs, the IoR Library, and the BN template have been identified:

**Creation of templates and playbooks for security risk monitoring tools**

As the MITRE ATT&CK framework is already used in SIEMs and other security monitoring tools to map threats, the IoR Library can be used as an extension of the framework to be referenced in risk and security monitoring tools. IoRs in combination with the ATT&CK matrices provide a common language for describing observable events and their implications that can be used in defining automated and manual SOC procedures. An example of this is the BN template which was developed as an additional resource to assist the Continuous Risk Assessment methodology presented in Chapter 4. By pre-loading IoRs in a tool, as was done with the BN template, it is possible to address the challenges regarding scalability that implementing continuous risk management might present.

**Enable performing a risk-based security monitoring**

Detecting IoRs rather than just IoCs enables threats to be anticipated, enabling a more proactive approach to defence. Even if there is no continuous risk assessment paradigm in place, integrating IoRs with a security monitoring system, can allow security analysist to relate certain observations to the risks of adversary techniques being performed. This can help situational awareness, and improve defence and response strategies. For example, the presence of certain IoRs can reveal excessive risk exposure or possible malicious activities related to a reconnaissance phase. Investigating these issues can prevent an attack and help improve the security posture of the organisation.

**Use in Forensics**

IoRs can be used to assist in the forensics analysis and documentation of new ICS adversary TTPs. Correlation of observable events (IoRs) with novel adversary Techniques will allow anticipation and early detection of future similar occurrences.

**Use of the BN template for simulation and risk analysis**

The BN template that was created to complement of the IoR library and facilitate the implementation of continuous risk monitoring based on BN can also be used for simulation purposes and anticipating risk scenarios. Besides being used for actual IoR monitoring, the BN can be used to create “what if” scenarios in which it can be assessed how the observation of different IoRs can modify risk estimations. Simulation is a useful technique for risk analysis and decision making that can assist in identification of potential outcomes, and the performance of evaluations, such as sensitivity analyses to check which IoRs have the most influence on risk probabilities. Simulation can allow evaluation of how risks can change under certain given conditions even if continuous risk monitoring is not implemented. Simulation can also be used for calibrating and improving the BN by using the simulation results to adjust the conditional probabilities in combination with historical IoR data and expert judgement. New conditional probability values can first be tested in a simulation environment before actually being implemented in the continuous risk analysis and assessment system.

**Prioritisation of security investments**

By monitoring IoRs it is possible to assess which security investments would have a higher cost-benefit ratio. Identifying which IoRs are observed most frequently and how many and which Techniques they related to provides an organisation with quantitative data to prioritise the future security investments that address the most threats and have the highest cost-benefit ratios.

# 6. Physical-based anomaly detection applied to continuous risk monitoring

This chapter contains examples of collection and modelling of sensor data to build anomaly detection algorithms. The motivation of these practical experiments was to highlight the potential of physical based anomaly detection for cyber-physical systems security. This was done by exploring the type of models and rules can be used to define IoRs from the I/O data threat group of the IoR Library, presented in Chapter 5. Each one of the experiments described on this chapter was performed with the objective of providing a lower level view of this type of indicator, since it is the group less covered by more traditional anomaly detection use cases, such as the ones based on network and end-points data. As detailed in Chapter 3 (section 3.2.4) the idea of using physical-based (or sensor-based) anomaly detection for security monitoring has been proposed and demonstrated in a hand full of academic papers but it is not a common practice in cyber-security yet. Hence, it was considered important to provide examples on how sensors data can be used for cyber-security. The demonstrations performed are described in section 6.2 using three different data sets. These data sets were obtained from two experimental settings and one real system. At the end of each subsection, an explanation is provided on the applicability of each model on the definition of IoRs.

The main reason why the approach proposed in this thesis considers the monitoring of sensors data as an important aspect of ICS continuous risk management is that cyber-physical systems have “blind spots” in which traditional detection strategies are not useful. While typical detection mechanisms work for the traditional TCP/IP ICT networks (levels 2 to 5 in the Purdue model) and specific ICS tools can cover OT networks (levels 0 to 2 in the Purdue model), advanced threats and insider attacks could still remain undetected. To compensate for this, methods for physical-based anomaly and misbehaviour detection are key, since they reveal signs of compromise that can remain undetected by network and host based detection tools. Hence, physical-based anomaly detection can help provide a complete overview of the risk picture in ICS.

These examples described in this chapter are based on samples collected using home automation devices, and data collected from a BMS in a Data Centre, which was shared by the telecommunications company that owns the datacentre. The objective was to explore anomaly and misbehaviour detection as an approach to generate IoRs. This complements the answers given in Chapter 5 to the questions “what information must be known in order to monitor security risks in IIoT/ICS?” and “how can that information be derived from what can actually be measured?” In this case, the focus is on specific IoRs, mostly ones belonging to the I/O data threat group (IoR5XX), which are based on data from sensors and actuators.

To detect anomalies in physical or I/O data it is necessary first to create a behavioural model of the system that defines what is normal. This can be based on the system specification, and the expected output, should obey both the control rules and the laws of physics. As the laws of physics cannot be broken, given a reasonably reliable behavioural model, any deviations from it can be only caused by a violation of the control rules or by data errors. The latter could be either due to a sensor malfunction or due to data tampering. As it is in the interest of ICS owners and operators that the system works as intended, the utility of misbehaviour detection goes beyond cyber-security. Models for early detection and prediction of flaws are also used for predictive maintenance and safety. Thus, the examples developed in this chapter will be focused mainly on the predictability of the behaviour of the data under normal conditions, and hence how feasible is modelling it. Modelling the normal behaviour of sensor data should allow the detection of abnormal events whether they are caused by a cyber-attack or not.

Additionally, possible correlations between physical variables, such as temperature and pressure, which are a consequence of the laws of physics rather than programmed behaviour, can also be used to detect anomalies in the system. For example, in the case of Stuxnet, nuclear centrifuges were maliciously programmed to spin at a higher speed than their operational limit, while the monitor and control system displayed a spoofed I/O image that made them appear normal. However, ineluctably, an induction motor that spins faster will consume more electrical current. Hence, by monitoring the electrical power consumption, an anomalous increase could have been detected. Consideration of the electrical loads most likely to cause this, could have led to an investigation in which the consumption of the centrifuges motors was measured directly allowing to find the malfunction.

Normal behaviour of an ICS could simply be defined in terms of limits to the ranges of operational variables. However, in many cases an ICS will have more complex system dynamics, and so a more sophisticated model will be required for anomaly detection. In such cases in which explicit causal behavioural models are not available, but historic data is, it is also possible to use machine learning techniques. Selecting appropriate methods, rules, and training mechanisms to detect misbehaviour and avoid false positives is another challenge for which gaining good level of understanding about the expected normal behaviour of the system is required. This subject constitutes a research topic by itself. Thus, this aspect of the thesis is limited to simple demonstrations as a complement of the main topic, continuous risk management.

The next section of this chapter describes how anomaly and misbehaviour detection fits into the approach for ICS continuous risk management presented in this thesis. This will be followed, in the subsequent sections and sub-sections, by practical examples based on data samples obtained during the course of the research. Finally, the chapter is concluded with a discussion.

## 6.1. How anomaly detection fits into continuous risk management

Figure 1.2. describes an architecture with the different building blocks of the continuous risk assessment approach proposed in this thesis in which anomaly detection was included as one of the systems that provides data for the observation of IoRs. In Chapter 3, Section 3.2.4, a state-of-the-art review of behavioural anomaly detection in ICS was presented, with special focus on the NISTIR 8219 report [12]. Examples of academic and industry research work were gathered to demonstrate how several references regard behavioural-based detection as an important aspect for next generation ICS cyber-security. The techniques proposed in these works are based on comparing the real behaviour of the ICS with a model of its expected behaviour in order to improve detection security of security threats. In this Thesis, particular attention was given to anomaly detection techniques based on sensor data, since it is an approach that is not widely used in cyber-security applications, even though it could complement other methods. This is particularly important considering that the objectives of ICS cyber-security go beyond information security and also cover the protection of reliability and safety on industrial operations from being compromised by cyber-incidents. In Section 3.2.5, a state-of-the-art review of the use of real time indicators for risk monitoring was reported, which establishes a connection between anomaly detection and continuous risk monitoring.

Figure 6.1 illustrates on where anomaly detection fits into the continuous risk management approach presented in this thesis. Raw security and operational data, including data from sensors and actuators, is processed to generate IoRs which are used for continuous risk monitoring. In the lower layer the data is obtained directly from the system including variables such as logs, network traffic type and volume, file exchanges, configuration changes, software and firmware changes, sensor's measurements, and actuator states. In the second layer this data is processed by security and anomaly detection tools generating Indicators of Risk (IoRs), and in the higher layer, these indicators are mapped into different risks and used to recalculate risk scores. This diagram is also consistent with the architecture described in Figure 1.2.

When an attacker's main goal is to cause disturbance and damage in industrial operations, they would select TTPs that allow them to bypass host and network based security detection mechanisms. Sophisticated procedures to target ICS developed in the past years such as Stuxnet, Industroyer, Triton, and Ekans, constitute advanced persistent threats that are specially crafted to cause physical damage and to avoid detection. Added to this, malicious insider threat, such as in the Maroochy water case in Australia in the year 2000 can also be hard to detect in time since the adversary has legitimate access rights and a privileged level of knowledge of the system. Hence, cyber-security risk monitoring should adopt strategies that cover these threats. In this sense, as much as it is desirable to detect an attack in its early stages, when other detection mechanisms have been successfully bypassed by an adversary, the effect of adversary TTPs could be observed through data from inputs and outputs (I/O) of the ICS, such as data from sensors. In the case of the Continuous Risk Monitoring approach base on IoRs proposed in this thesis, a specific group of IoR was defined for this type of data, which is the I/O data threat IoR Group, which has the naming scheme IoR5XX. This type of indicator in combination with the other groups allows having a view of possible observations from all the levels of an ICS system, which can be related to risks.

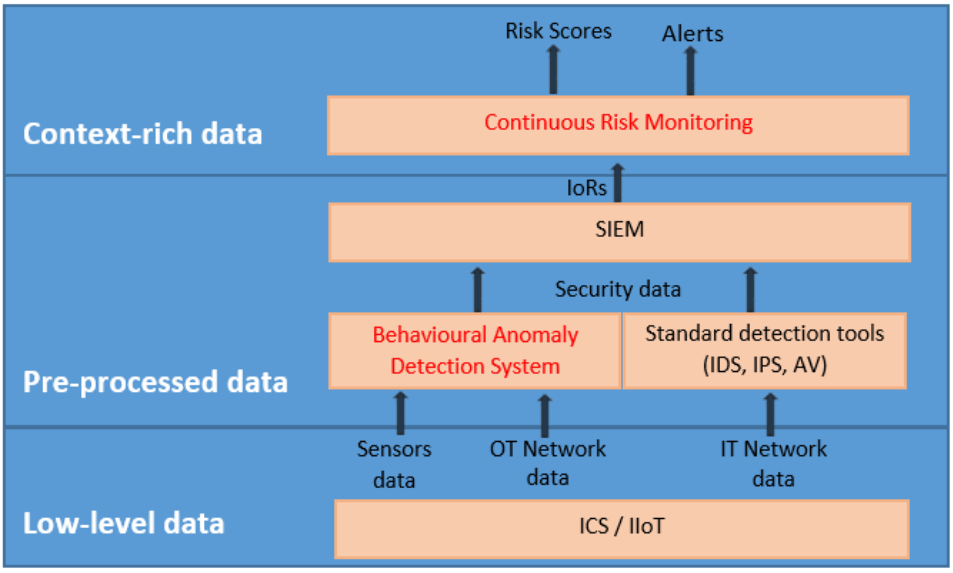



*Figure 6.1: behavioural anomaly detection in the context of continuous risk monitoring*

### 6.1.1. Anomaly detection principles

Figure 6.2. shows as an example, the architecture of a generic platform that enables a variety of anomaly detection models to be built based on different analytic and machine learning techniques. This image was extracted from [152]. Examples of models that learn the normal behaviour of time series data range from simple Moving Average, to ARIMA, Holt-Winters, Kalman Filters, and Fourier Transform of signals.

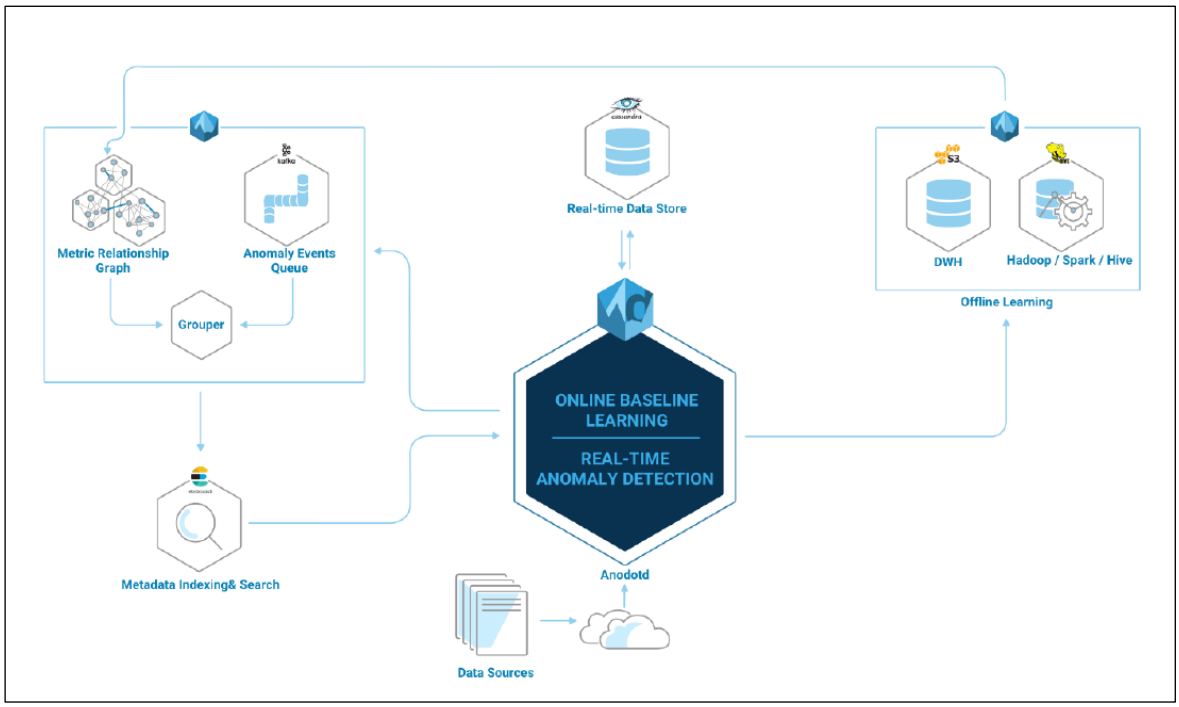



*Figure 6.2: Example architecture of an anomaly detection system*

Anomaly detection is based on creating a model that defines how the variables that are being monitored are expected to behave and compare the results of the calculations done using this model with the actual data values. This model could be based on offline learning from historical data, but it also can be improved and adjusted based on real time data, as in the example of Figure 6.2 for which machine learning techniques can be used. Independent on the tools, technology and mathematical approach and techniques used for anomaly

detection, its main principle remains the same: detecting significant differences between an expected and a real value. The measurement of what a "significant difference" means is given by the accuracy of the prediction model and the tolerance for variations, which would be based on the system's requirements. A good anomaly detection system should be able to detect variations that exceed the system's tolerance with a minimum of false positives. In the next section, practical demonstrations of behavioural anomaly detection models will be developed as a mean to provide some examples.

## 6.2. Practical demonstrations with data collected from sensors

This section describes practical experiments done with different data set to explore the implementation of an anomaly detection system based on data from sensors. Simple data models were build based on modelling and labelling the sampled data and using different techniques to create algorithms that define whether new incoming data presents anomalies. Examples of techniques used for anomaly detection are ARIMA (Auto-regressive Integrated Moving Average) for time series analysis and forecasting, definition of thresholds or control limits, and correlations between variables. These Proofs of Concept correspond to simplified scenarios and naive anomaly detection models, which have several opportunities to be improved further. As much as a deeper view on this topic would have been beneficial for this research, it was considered out of scope due to time and resources limitations. However, the following examples serve as reasonable demonstration of the potential of anomaly detection in continuous risk monitoring for ICS. This allows to make a case for the use of these techniques in ICS and IIoT cyber-security. Furthermore, the fact that as part of this work it was possible to build these models with simple tools such as MS Excel, shows that it is not unfeasible to build more sophisticated algorithms using more powerful tools such as R. To validate each model, the data samples used to test each anomaly detection algorithm were different from the ones used to build (aka train) the models and define the anomaly detection rules. This means that the behaviour across the data sets presented a defined behaviour, which could be observed in different samples.

### 6.2.1. Electrical consumption and use of computing resources

A setup was implemented using a Samsung Smart Things kit to monitor the electrical consumption of a laptop computer during different periods of time. The goal of this experiment was to check if it was possible to make a profile of the normal use of this computer according to its power consumption, and hence, identify any abnormal power usage, which could be associated to malicious activities. Figure 6.3 shows the experimental setup used to capture the data using a smart socket and a Samsung Smart Thing's hub. The smart socket monitors the power consumption of whatever is plugged to it, in this case, the computer, and sends the data by a wireless radio frequency protocol (in this case Zigbee) to the hub, which sends the data through the internet to the Smart Things platform. The data stored in this platform was collected and stored in a server in .cvs files format for its further analysis.

#### 6.2.1.1. Data analysis

The data analysis was based on reviewing the behaviour of the power consumption value, measured in Watts, in order to identify specific profiles that could be modelled for identifying the usage of the laptop computer based on its power consumption. The activities that were undertaken during the period of data collection were manually logged. The detailed logs of all the processes running at each moment was not captured, not making it possible to distinguish specific processes running, for which it was only possible to define two profiles: one for when the computer was in use and one for when the computer was not in use. Three labels were defined and manually assigned to the data, which were "Busy", "Idle", and "Transient", which corresponded to transitions between "Busy" and Idle".

The analysis was focused on trying to establish a mathematical model for the Idle and Busy states, for which the first step was doing a statistical analysis. The data corresponding to each one of these states was analysed separately to identify upper and lower thresholds, average values and standard deviations. Table 6.1 shows a

summary of the data from 20 samples corresponding to the Busy state in which the first columns represent the following values:

- **Period:** is an identifier of the sample
- **Duration:** the size of the sampling time window in hours, minutes, and seconds
- **Total points:** the amount of values captured during the time window
- **Mean, Min., and Max.:** the average, minimum and maximum values of the sample
- **Density:** the average amount of values captured per minute
- **Max. frequency:** the statistical frequency for the value range that corresponds to the mode
- **Range:** the difference between the maximum and the minimum values
- **Std. Dev:** the standard deviation of the sample

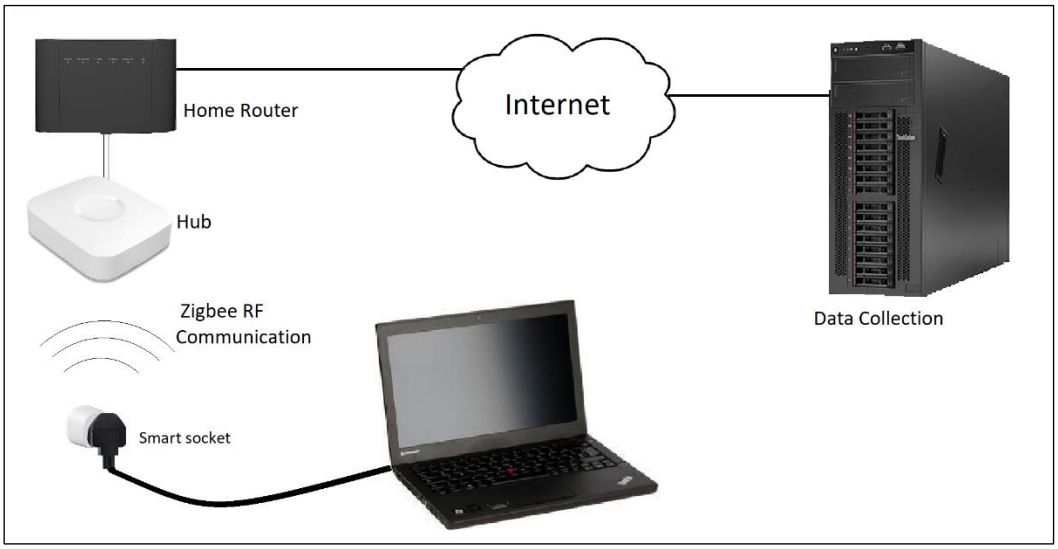



*Figure 6.3: Setup to capture power consumption data*

*Table 6.1: Summary of data samples for the "Busy" state*

| Period | Duration | Total Points | Mean | Min | Max | Density Av Points/ Minute | Max frequency | Range | Std. Dev |
|---|---|---|---|---|---|---|---|---|---|
| 1.1 | 00:20:48 | 1298 | 18.70 | 13.70 | 27.30 | 62 | 20.26% | 13.60 | 2.29 |
| 1.2 | 01:00:24 | 2087 | 11.82 | 0.60 | 25.40 | 35 | 21.27% | 24.80 | 2.81 |
| 2.1 | 03:43:00 | 8002 | 11.68 | 2.20 | 26.60 | 36 | 24.88% | 24.40 | 2.79 |
| 2.2 | 00:24:00 | 861 | 10.45 | 6.80 | 23.00 | 36 | 26.60% | 16.20 | 2.14 |
| 2.3 | 00:19:00 | 624 | 5.97 | 4.40 | 18.50 | 33 | 31.09% | 14.10 | 1.55 |
| 2.4 | 01:10:00 | 1033 | 10.32 | 0.40 | 25.20 | 15 | 28.36% | 24.80 | 2.65 |
| 2.5 | 00:21:00 | 599 | 5.90 | 0.20 | 17.20 | 29 | 32.22% | 17.00 | 1.69 |
| 2.6 | 01:20:00 | 985 | 9.78 | 0.30 | 25.00 | 12 | 35.03% | 24.70 | 2.64 |
| 2.7 | 01:01:00 | 2053 | 15.91 | 0.30 | 25.00 | 34 | 29.47% | 24.70 | 1.92 |
| 2.8 | 00:11:00 | 407 | 10.90 | 0.50 | 22.20 | 37 | 28.75% | 21.70 | 2.43 |
| 3.1 | 00:18:00 | 605 | 8.56 | 3.00 | 25.80 | 34 | 22.81% | 22.80 | 4.80 |
| 3.2 | 00:13:00 | 449 | 10.10 | 4.90 | 23.00 | 35 | 32.74% | 18.10 | 2.70 |
| 3.3 | 00:30:35 | 628 | 5.71 | 0.40 | 21.60 | 21 | 42.20% | 21.20 | 2.22 |
| 3.4 | 00:08:07 | 274 | 10.88 | 6.30 | 25.20 | 34 | 22.63% | 18.90 | 3.62 |
| 3.5 | 00:25:10 | 957 | 13.73 | 8.40 | 23.90 | 38 | 19.85% | 15.50 | 2.23 |
| 3.6 | 00:04:00 | 160 | 14.77 | 8.60 | 25.10 | 40 | 18.13% | 16.50 | 3.20 |
| 3.7 | 00:04:35 | 149 | 9.47 | 7.70 | 18.90 | 33 | 41.61% | 11.20 | 1.57 |
| 3.8 | 00:10:25 | 424 | 13.62 | 8.10 | 22.20 | 41 | 26.65% | 14.10 | 1.94 |
| 3.9 | 00:24:45 | 928 | 14.05 | 9.10 | 22.30 | 37 | 23.06% | 13.20 | 1.79 |
| 3.1 | 00:09:06 | 314 | 10.82 | 7.40 | 23.50 | 35 | 35.67% | 16.10 | 2.92 |
| | | **Mean** | **11.16** | **4.67** | **23.35** | **33.67** | **0.28** | **18.68** | **2.50** |
| | | **Min** | **5.71** | **0.20** | **17.20** | **12.31** | **0.18** | **11.20** | **1.55** |
| | | **Max** | **18.70** | **13.70** | **27.30** | **62.40** | **0.42** | **24.80** | **4.80** |

Table 6.2 shows a summary of the data from 2 samples corresponding to the Idle state.

*Table 6.2: Summary of data samples for the "Idle" state*

| Period | Duration | Total Points | Mean | Min | Max | Density Av Points/ Minute | Max frequency | Range | Std. Dev |
|---|---|---|---|---|---|---|---|---|---|
| 1.3 | 17:31:50 | 40 | 6.63 | 0.20 | 22.00 | 0.04 | 60.00% | 21.80 | 8.90 |
| 2.9 | 15:11:00 | 26 | 9.28 | 0.20 | 22.10 | 0.03 | 34.62% | 21.90 | 8.45 |
| | | **Mean** | **7.96** | **0.20** | **22.05** | **0.03** | **0.47** | **21.85** | **8.67** |
| | | **Min** | **6.63** | **0.20** | **22.00** | **0.03** | **0.35** | **21.80** | **8.45** |
| | | **Max** | **9.28** | **0.20** | **22.10** | **0.04** | **0.60** | **21.90** | **8.90** |

One important observation made from the statistical analysis was that while value ranges between samples from both states overlap, there is a distinctive difference between the frequency in which the values were captured in each case, which is reflected in the "Density" column. This is explained because the smart socket only forwards the data when there is a change in the value, for which in the "Idle" state, there could be no data captures for long periods of measurements. Hence, this metric was found to be useful for the state detection algorithm developed, which is explained in the next section.

Figure 6.4 shows examples of histograms showing the data distribution of samples for both states. It can be observed that while sample of the "Busy" state has a distribution skewed to the right, the sample of the "Idle" state presents a concentration of 60% of the points in the range between 0 and 0.9 Watts and distributed values. This correspond to particular points of time during the idle period in which a power consumption value is registered, which can be observed better through the time series analysis. It must be noted that, as observed in the comparison between the "Total Points" and "Density" columns from Table 6.1 and Table 6.2, in the Idle period there are very few value observations, which makes this random values of high power consumption to appear as more significant as they really are.

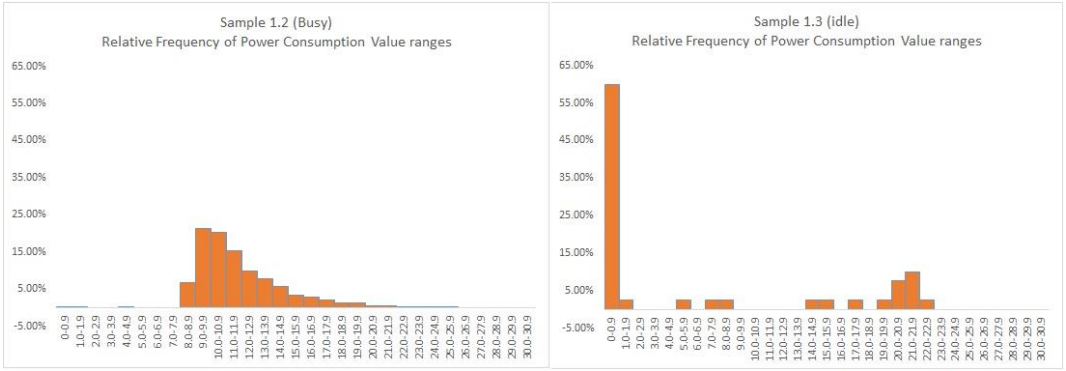



*Figure 6.4: histograms from "busy" and "idle" samples*

The second part of the data analysis was a time series analysis, which was done to observe the behaviour of the power consumption over time. While the statistical analysis was useful to check the distribution of the data, the time series analysis allowed observing the value changes over time and how stable and predictable they are. Figure 6.5 shows an example of a time series graph, in which the random power consumption values during the idle period can be observed. These values could have several explanations, and it has to be taken in account that during the idle period the laptop was on sleep mode and not off during this time, which means that it still consumes power. The state detection algorithm should take in account this behaviour not to confuse it with a "Busy" state.

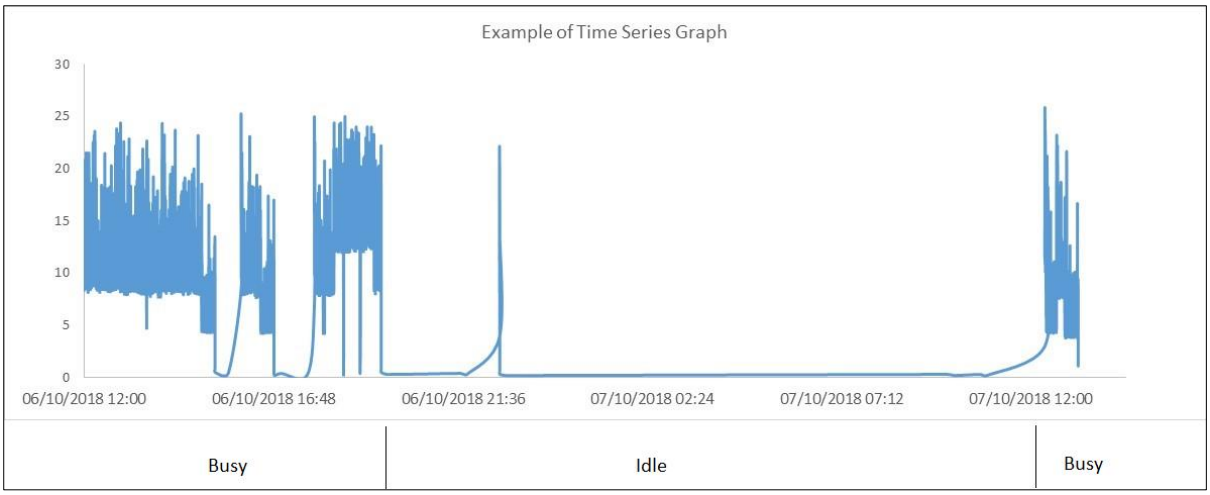



*Figure 6.5: Example of a time series graph*

As part of the time series analysis, the data was tested using the ARIMA (Auto-Regressive Moving Averages) forecasting algorithm. The purpose of this was to check if it is possible to predict the power consumption behaviour with enough level of accuracy, which could allow comparing the forecasted values with the actual data and spotting any anomalies. Figure 6.6 shows an example of the ARIMA forecasting results for a sample during the "Busy" state; the values represented with orange correspond to the forecasted values and the ones in blue to the actual data.

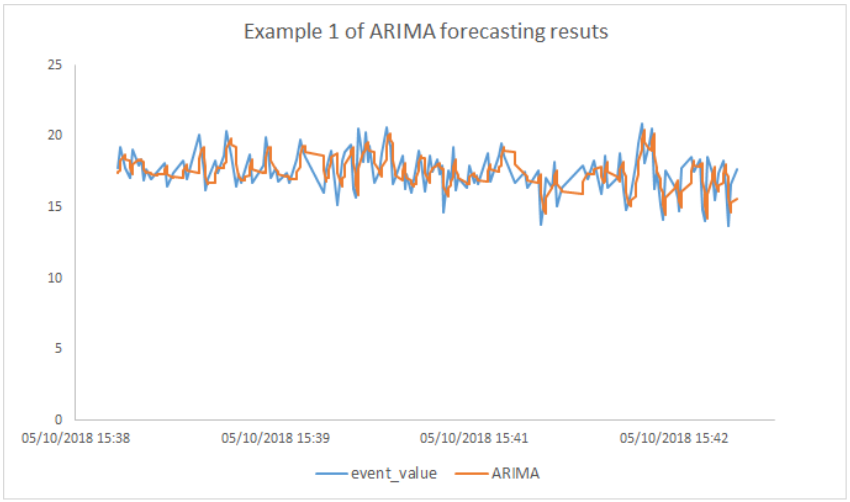



*Figure 6.6: Example of ARIMA in Busy state*

As can be observed in Figure 6.6 the ARIMA forecasting algorithm is reasonably accurate in predicting future values for the "Busy" state, which means that any unexpected value could be detected raising an alert. However, it is expected that during a transition period the forecasted values present a high error. As the ARIMA model is designed to reduce these errors by taking in account for the forecast not only a weighted average from the recent values (which corresponds to the Moving Average component) but also a weighted average of the errors, even in transition periods, the predictions catch up quickly with the real values reducing the errors as it can be observed in the example of Figure 6.7.

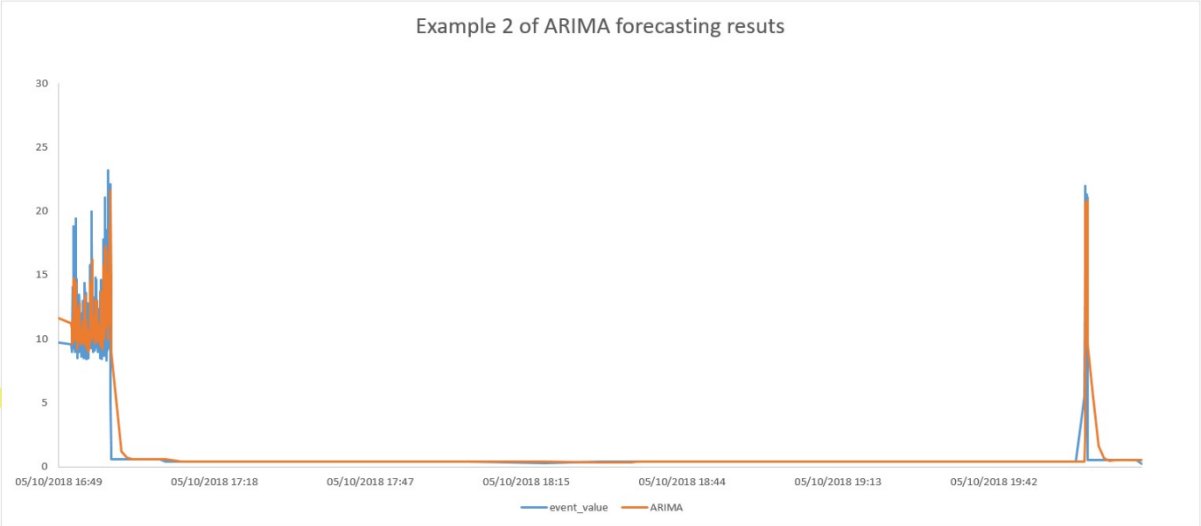



*Figure 6.7: Example of ARIMA in Transitions*

6.2.1.2. State detection algorithm

A state detection algorithm was built based on a combination of the results of the statistical and the time analysis described in section 6.2.1.2. After testing different approaches, two independent algorithms were chosen for identifying whether the laptop was in a "Busy" or "Idle" state, and a third one to check if the current behaviour of the data was stable. The final detection model was built based on the combination of the a results of these three algorithms.

The following algorithms were built for the identification of "Busy" or "Idle" states:

- Density of points: if the sensor readings captured were in a rate lower than 0.3 values per minute, the state was classified as "Idle" and if the rate was higher than 1 value per minute the state was classified as "Busy". Any intermediate rates were classified as "undefined".
- Power consumption value: the identification of "Busy" and "Idle" were based on whether the values where higher than 5 Watts or lower than 2 Watts, respectively. Intermediate values where classified as undefined. Two options where tested with similar results, the first one was based on the current measures values and the second on moving averages.

The following algorithm was built to define if the data was stable:

- ARIMA error: for differences between the forecasted and actual values higher than a 80% the data behaviour was considered as "unstable". This corresponded either to transition states or to abrupt changes in the power consumption.

The final state detection model was built based on comparing the three algorithms described above and applying the following rules:

- While the behaviour of the data is unstable, the state is recognized as "undefined"
- When the data presents stable values, the state is identified through comparing the results of the algorithm based on the density of points with the results of the algorithm based on the power consumption value.
- If the state identification obtained by both algorithms is the same, this state is defined as the current state. Otherwise, the state will be considered as "undefined".
- If the state is only identified by one algorithm but not by the other, the result from the algorithm that could identify the state is defined as the current state.
- If neither of the algorithms could identify the state, this is categorized as "undefined".

#### 6.2.1.3. Results

The results were measured in terms of the accuracy of the model considering both wrong predictions and instances in which the algorithm as unable to identify the state, which gave the results of "State Undefined". As the algorithm was not created to detect transient states, it was expected that at least the points labelled as "Transient" would be classified as state "Undefined". Figure 6.8 shows a summary of the results testing the state prediction model with two data sets, the first one was the one used to build the mathematical model used in the detection algorithm and the second a totally independent sample.

**Data set used to build Model**

| | |
|---|---|
| **Total Points:** | **49820** |
| **Busy:** | 49776 |
| **Idle:** | 33 |
| **Transient:** | 11 |

| Option 1 | Wrong Prediction | | State Undifined | |
|---|---|---|---|---|
| **Predicting Busy State** | 14 | 0.028% | 35 | 0.070% |
| **Predicting Idle State** | 2 | 6.061% | 13 | 39.39% |
| **Total** | 16 | 0.032% | 48 | 0.096% |
| **Total Accuracy** | **99.872%** | | | |

| Option 2 | Wrong Prediction | | State Undifined | |
|---|---|---|---|---|
| **Predicting Busy State** | 14 | 0.028% | 34 | 0.068% |
| **Predicting Idle State** | 2 | 6.061% | 13 | 39.39% |
| **Total** | **16** | **0.032%** | **47** | **0.094%** |
| **Total Accuracy** | **99.874%** | | | |

**Independent Data set**

| | |
|---|---|
| **Total Points:** | **18083** |
| **Busy:** | 18019 |
| **Idle:** | 62 |
| **Transient:** | 2 |

| Option 1 | Wrong Prediction | | State Undifined | |
|---|---|---|---|---|
| **Predicting Busy State** | 1 | 0.006% | 66 | 0.366% |
| **Predicting Idle State** | 7 | 11.290% | 36 | 58.06% |
| **Total** | 8 | 0.044% | 102 | 0.564% |
| **Total Accuracy** | **99.392%** | | | |

| Option 2 | Wrong Prediction | | State Undifined | |
|---|---|---|---|---|
| **Predicting Busy State** | 1 | 0.006% | 66 | 0.366% |
| **Predicting Idle State** | 7 | 11.290% | 36 | 58.06% |
| **Total** | **8** | **0.044%** | **102** | **0.564%** |
| **Total Accuracy** | **99.392%** | | | |

*Figure 6.8: Results of the state detection model*

As it can be observed in Figure 6.8., the total accuracy of the state detection model is high; however, it needs to be improved to reduce the prediction error of the Idle state. These errors appeared particularly on points in which spikes of power consumption appeared in the middle of an idle period, which can be observed in Figures 6.5 and 6.7.

#### 6.2.1.4. Applicability of the model related to defining IoRs

The experimental demonstration of building a data model to detect whether a computer is in use or idle had as an objective to illustrate possible IoRs such as IoR502-Misbehaviour in sensor data, and IoR504-Misbehaviour in data that can be correlated to a critical input misbehaviour. For example, industrial machines that have scheduled tasks will produce a bigger amount of data and hence, consume more computing resources in established periods of times. This would allow building profiles of which periods of time the ICS servers can be expected to be more busy. If servers are busier than expected, according to the defined profile, it could mean that malicious activity is consuming computing resources. However, very advanced targeted threats can be programed to disguise this action. Measuring the power consumptions of servers and knowing how this correlates with their expected behaviour, could allow to detect suspicious activity by identifying abnormal patterns of power consumption behaviour, even when an attacker is trying to hide their direct actions.

In cases in which the processes that run in the machine present clear patterns of data consumption different profiles of normal use can be modelled allowing also identifying anomalies in busy periods. Additionally, the combination and comparison of several algorithms can be useful to reduce errors or false positives, which can be applied on other types of anomaly detection models. This model, which allows inferring a misbehaviour in a system based on its electrical consumption can be improved and adapted to a variety of contexts and scenarios.

### 6.2.2. Temperature and humidity correlation based anomaly detection

This section describes the results of a correlation-based anomaly detection model based on the data gathered from an experimental set-up using environmental sensors. The main objective of this experiment is to show that anomalies in one variable can be spotted by monitoring the behaviour in a correlated variable. This approach can be useful to spot spoofed sensor data, for example, by replaying the image of recorded values of past data, such as it was done in the case of Stuxnet. Hence, in the IoR Library [144] described in Chapter 5, IoRs were defined related to this type of anomaly detection. This experiment also aims to illustrate how previous knowledge about the relationship between variables and their expected behaviour can help inferring possible data behaviour and correlations. In this case, the model inferred based on thermodynamic laws rather than only based on learning from the data itself. Once the model was built an anomaly was simulated to look into an attack scenario where a temperature control is manipulated to raise the temperature above the allowed threshold while data showing normal temperatures is used to disguise the attack. If the humidity data is gathered independently and can be trusted, then it is possible then, to infer a temperature anomaly from the humidity data.

To collect the data, a Bosch-XDK universal programmable multi-sensor was used, which is shown in Figure 6.9 and corresponds to a prototyping platform for developing IoT products and includes a variety of built-in sensors. For this experiment pressure, temperature, and humidity information were used. The resources available for this experiment only allowed an “open-loop” system, which means that there is only monitoring, but no control capabilities. Therefore, it was required to set the proper conditions to control the environment as much as possible so the variables were stable and the experiment was repeatable. After several trials, improvement in the stability of the data was achieved through placing the sensor in a plastic box, which was placed inside a bigger box. This set-up allowed air circulating in the space between the boxes providing a reasonable degree of physical isolation to maintain temperature levels within a range near the 20 ºC +/- 3 ºC, approximately at room temperature.

As the setup was made in a domestic rather than in a laboratory setting, an effort was made to constraint as much as possible the environmental set-up in order to simulate a “controlled environment” and collect meaningful data. Hence, the scope of the experiment and its results were limited and it mainly served illustrative purposes.


*Figure 6.9: Bosch-XDK platform*

#### 6.2.2.1. Data analysis

The data was collected by an application running in the Bosch-XDK controller which takes a sample every 1 second and writes the value in a text file, automatically generated and saved in a micro-SD card, every time the device is turned on. The power supply was given by a power bank connected to the device. While it would have been possible to collect the data directly into a computer through an USB cable, this setup has the advantage of allowing physical isolation of the sensors.

The data collection was done under two different scenarios. The first scenario was at room temperature, which was used as a benchmark to define the “normal” behaviour, and the second scenario was generating applying external heat to the space between the boxes to simulate an anomaly caused by maliciously changing the settings in a temperature control.

Before the data analysis, it was presumed that temperature values should de directly proportionality with pressure values, as well as an inverse proportional with humidity values. However, by just observing the data it became clear that within the given environmental conditions, the pressure levels were relatively stable, independent of the temperature and humidity levels. This was confirmed by building a mathematical model using temperature and pressure data which showed no strong relationship between temperature and pressure variations. On the other hand, the inverse proportionality between temperature and humidity became quite clear just from observing the data, which was also confirmed by building a mathematical model. Hence, the data analysis and anomaly detection model were built based only on temperature and humidity, leaving pressure as a constant variable.

The data used as reference for normal conditions was collected for periods of up to 90 minutes, obtaining 3 samples, which included a total of 11442 data vectors. The following are the ranges of data collected across the samples:

- Humidity range: 35-67%
- Temperature range: 16.8 to 28.3 ºC
- Pressure Range: 96.86 to 98.97 kPa (ref: 1 atm = 101.325 kPa)

Table 6.3 shows the linear model of temperature versus humidity for each individual sample and Figure 6.10 resents the model obtained by aggregating the data from all the samples. It can be observed that there is a strong inverse relationship, where the best fit is a linear model. "N" represents the number of data vectors collected, "m" represents the slope of the line, which is negative due to the inverse correlation, and "b" the intercept calculated between the line and the "y" axis. The Pearson correlation coefficient ($R^2$) is over 90% when modelling individual samples and it becomes weaker for the cross-sample model (82.9%).

*Table 6.3: Linear model for reference data*

| Reference environmental data | | | | |
|---|---|---|---|---|
| | N | M | B | $R^2$ |
| Sample 1 | 1399 | -0.357 | 38.84 | 97.03 |
| Sample 2 | 5610 | -0.578 | 48.59 | 95.87% |
| Sample 3 | 4433 | -0.394 | 42.21 | 92.04% |
| Cross-sample | 11442 | -0.333 | 38.49 | 82.93% |

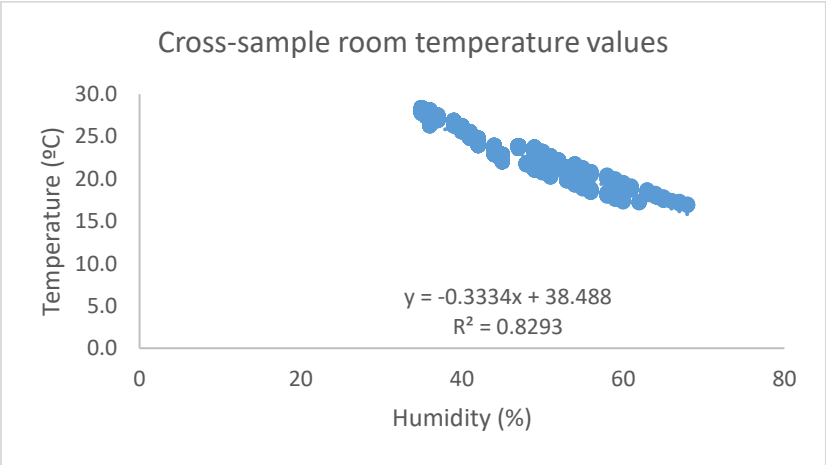



*Figure 6.10: Cross-sample linear model for reference data*

The data from the experimental set-up in which an anomaly was simulated by applying heat for normal conditions was collected for periods of up to 90 minutes, obtaining 3 samples, which included a total of 3967 data vectors. The following are the ranges of data collected across the samples:

- Humidity range: 33-61%
- Temperature range: 20.9 to 37.3 ºC
- Pressure Range: 96.70 to 98.66 kPa (ref: 1 atm = 101.325 kPa)

Table 6.4 and Figure 6.11 describe the linear model built representing the behaviour of temperature versus humidity when external heat is applied by aggregating data from all the samples. It can be observed that there is a strong inverse relationship, where the best fit is a linear model, presenting a Pearson coefficient ($R^2$) of 95.2%, and an even stronger correlation for individual samples.

*Table 6.4: Linear model applying heat*

| Reference environmental data | | | | |
|---|---|---|---|---|
| | N | m | b | $R^2$ |
| Sample H1 | 391 | -1.025 | 78.25 | 96.85% |
| Sample H2 | 1313 | -0.873 | 67.07 | 97.03% |
| Sample H3 | 2263 | -0.654 | 58.43 | 98.28% |
| Cross-sample | 3967 | -0.673 | 59.327 | 95.2% |

Although there was a variation in the linear equation between all samples, there was a noticeable difference between the equations for the samples that were taken at room temperature and the ones when external heat was applied. This condition allowed building the anomaly detection algorithm that will be presented in the next section.

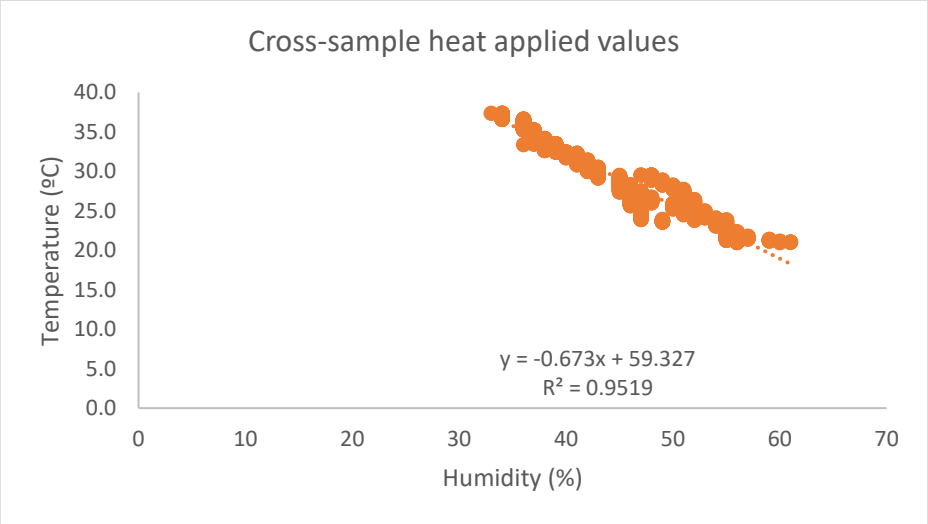



*Figure 6.11: Temperature and humidity correlation with external heat*

### 6.2.2.2. Anomaly detection algorithm

An anomaly detection model was built in order to identify any abnormal behaviour in environmental variables. The definition of normal behaviour was defined by the data collected at room temperature. The scope of the anomaly detection model was constrained to the range of temperature and humidity values obtained from the

samples. The detection algorithm was based in two techniques: threshold checking and correlation checking. In the first technique, it is checked that temperature, humidity, and pressure levels are within the permitted range, and in the second technique it is checked that the relationship between humidity and temperature corresponds to the one observed from the reference data.

To address the deviations from the linear equation presented in the samples, minimum and maximum values across the samples were used to define new equations which represent the frontiers between "normal" and "abnormal" parameters. As shown in Figure 6.12, rather than a single line, two lines were defined for each scenario to represent the minimum and maximum possible temperature for each humidity level between 48% and 56%, which was the scope defined for this exercise. The maximum and minimum temperatures where defined using the overall minimum and maximum values across all the samples for each humidity level and adding a tolerance range of - 5% and +5% respectively.

Defining thresholds is the simplest way to detect an anomaly since it would just compare a variable with a maximum and minimum value and check that it is within the permitted range. However, a variable can be within the permitted range and still there might be an anomaly in the system if the vector that conforms the different variables represents an invalid state. For example, in room temperature conditions, a temperature of 19 ºC and humidity is 48% are both within the valid range but both in combination would represent an invalid state. Therefore, adding correlations between variables allows finding anomalies that could not be detected by just checking thresholds.

The following algorithm was built for threshold checking:

- For temperatures outside the range between $T_{min}$ and $T_{max}$ an alert is raised
- For pressure outside the range between $P_{min}$ and $P_{max}$ an alert is raised
- For humidity outside the range between $H_{min}$ and $H_{max}$ an alert is raised

The threshold values were used as an adjustable parameter in the anomaly detection model, which for this experiment were set, as follows:

- $T_{min}$ = 17°C
- $T_{max}$ = 25°C
- $P_{min}$ = 96 (mPa)
- $P_{max}$ = 99 (mPa)
- $H_{min}$ = 48%
- $H_{max}$ = 56%

The following algorithm was built for correlation checking in which the first rule corresponds to the lower limit and the second rule to the upper limit in which the temperature-humidity relationship is considered as "normal":

- If T < -0.366 * H + 38.19, an alert is raised
- T > -0.353 * H + 41.54, an alert is raised

Figure 6.13 shows a graphical representation of the anomaly detection algorithm in which the area in green represents the permitted temperature-humidity coordinates.

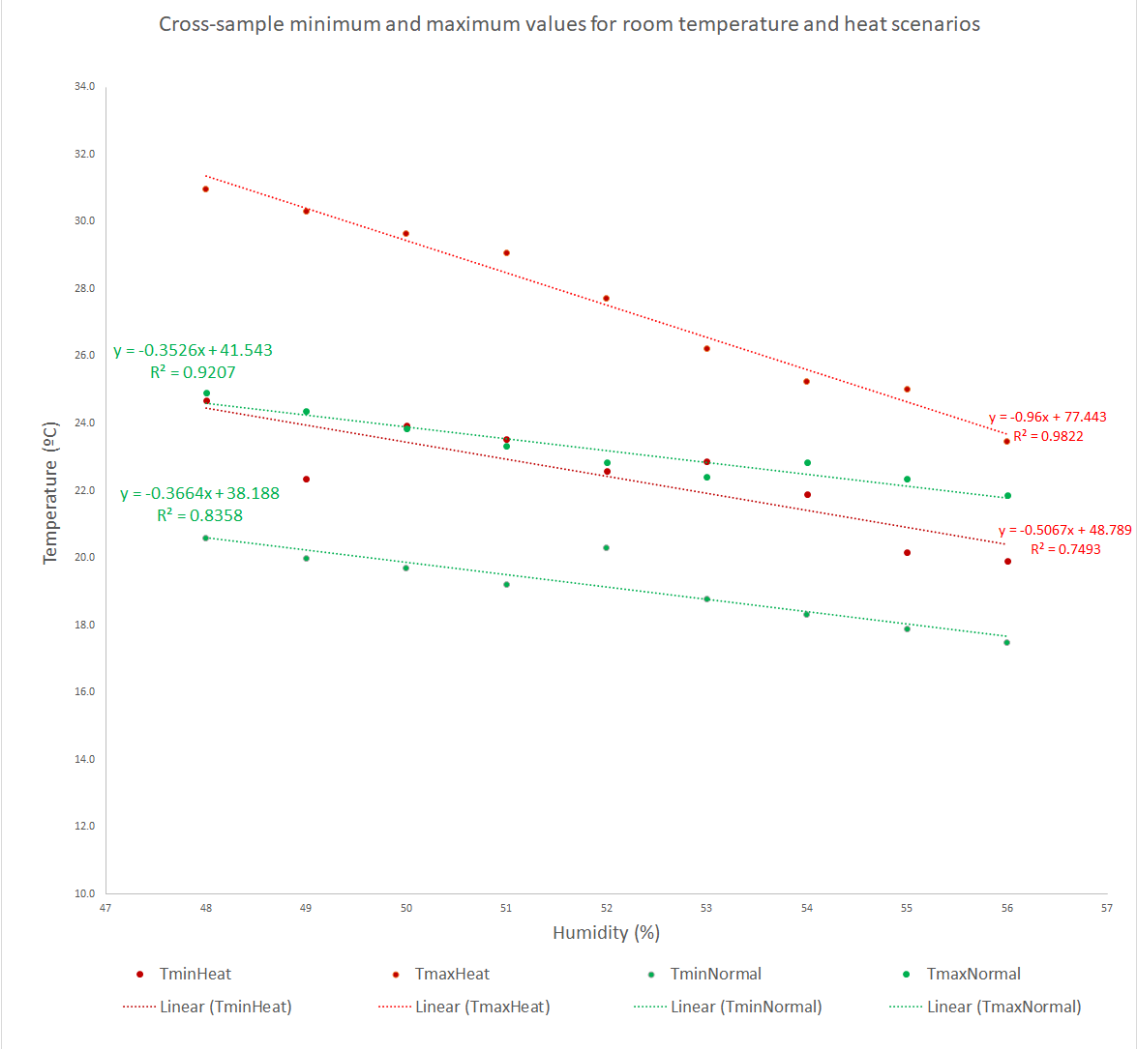



*Figure 6.12: Temperature Ranges for the 2 scenarios*

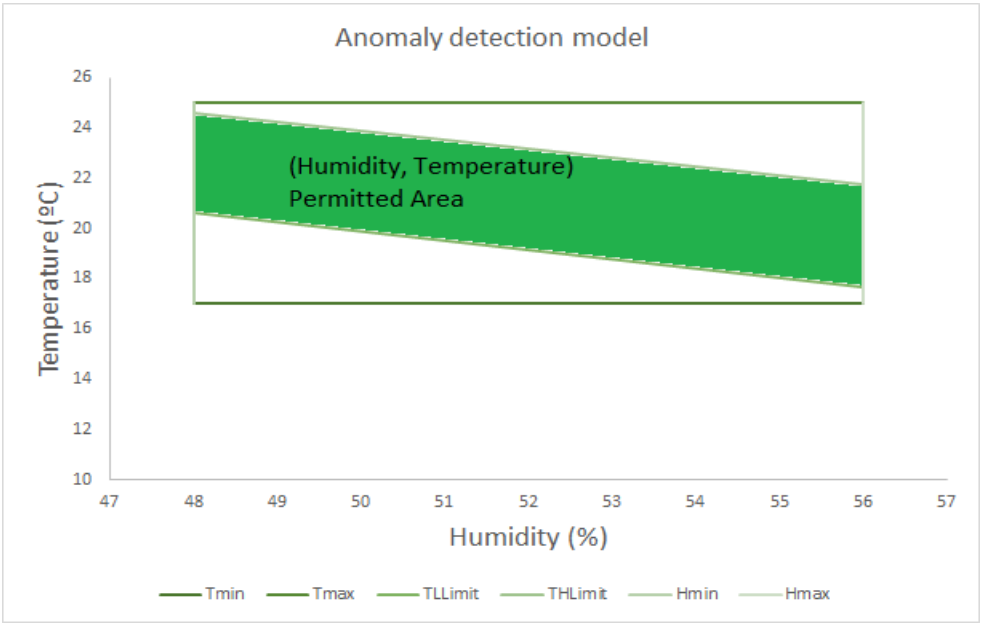



*Figure 6.13: Operational limits of the detection algorithm*

### 6.2.2.3. Results

The algorithm was checked using the different samples of data, in addition to which humidity values were modified to simulate a scenario in which the temperature values appear to be normal but do not correlate with the humidity values, as expected. Figure 6.14 shows an example in which humidity values are within the threshold but still do not correspond with the temperature level. This represents a case where temperature values might have been altered to appear normal while there are not which is detected through the humidity values.

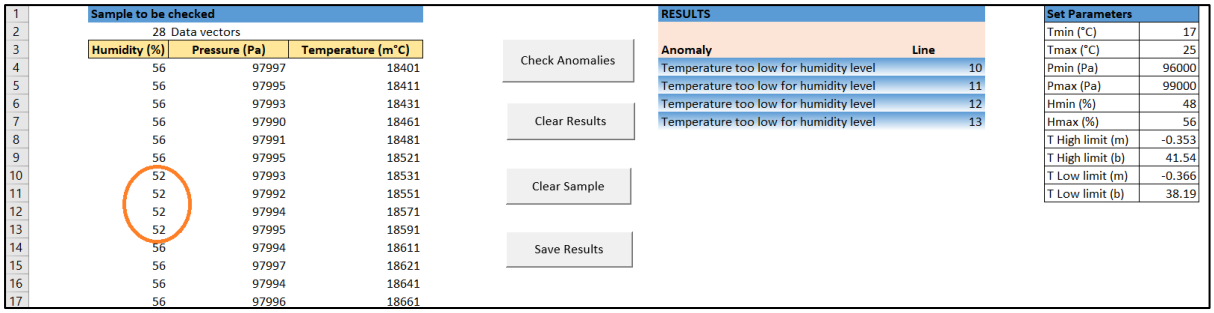



*Figure 6.14: Example of anomaly detection*

The results of this practical demonstration show that it is possible to build an anomaly detection model not only based on thresholds, but also in correlations between variables. Although the demonstration done was a simplified example, which had a limited scope it serves well as an illustration of what could be more advance techniques to detect attacks that target industrial operations by modifying control settings of physical variables.

### 6.2.2.4. Applicability of the model related to defining IoRs

The IoRs that can be related to this model are IoR501-Sensor data out of limits, IoR502-Misbehaviour in sensor data, IoR503-Indicator that can be correlated to a critical input out of limits, and IoR504-Misbehaviour in data that can be correlated to a critical input's misbehaviour. While alerts coming from a threshold check are likely to be programmed in the control system itself in order to maintain optimal operational levels, the correlation check might not be done. This means that the ICS might overlook states, which would be physically not possible, as long as they do not violate any threshold. For example, a humidity value that does not correspond to the temperature value, as shown in this example, or a power consumption level that does not correspond to the motor rotation speed.

Taking this example to a cyber-threat scenario, if an attacker manipulates the temperature control setting and disguises the action by re-playing valid temperature values, this will not be detected by the threshold check, but it can be detected by the anomaly in the expected correlation between humidity and temperature. A correlation-based anomaly detection model can be useful in cases of advanced threats in which the change of control settings is disguised by spoofing reporting messages and I/O images. This can be applied to multiple variables in an industrial process, for example, the flow and the pressure in a valve are expected to have a positive correlation, as well as the rotation speed of an electrical motor and its power consumption, and so on.

Hence, this approach can be used for I/O data integrity checks based on correlation checks, which would allow detecting a malicious alteration on the process variables even if this alteration is not directly visible. This works under the assumption that the attacker is not replaying also the values for the variable used for correlation check. To overcome the limitation of this approach for these cases, redundant measurements from independent sensors can be used, and several correlation check algorithms can be applied in parallel. This would make difficult for an attacker to evade detection since not only detailed knowledge about the industrial operations and their corresponding ICS would be required but also about the anomaly detection system. The difference between this model with the one in section 6.2.1, is that while the first one correlated the sensors data with states, this model correlates the sensors values with other sensors values.

## 6.3. Practical demonstrations with data from a Data Centre BMS

In Chapter 4, a worked use case was developed based on an environmental control of a Data Centre, which was integrated with a Building Management System (BMS). As part of the research work, data from a real BMS implemented in a Data Centre was obtained for analysis. The data was shared under a Non-Disclosure-Agreement for which the owner cannot be mentioned in this Thesis or any related publication. However, this is not relevant for the objectives of this data analysis. The data was shared in multiple .cvs files, which contained data collected between the 1st of November of 2018 and the 23rd of December of 2019. Temperature, humidity, and electrical current where gathered in independent files in which each data vector contained the time stamp, the tag of the corresponding sensor and the average measurements from intervals of 1 hour. The data was shared with a delay from over a year since the original request, for which there was not enough time available for a more thorough analysis due to the large amount of data and the need to pre-process the original files to do the analysis. Hence, there was an unused potential to make further use of this data for this research. This section describes the data analysis done and its potential use for developing anomaly detection algorithms.

The first two analysis and anomaly detection models were done using the data obtained from the data centre was based on analysing only room temperature values. The room had ten temperature sensors from which the data used to build the model, aka "training data", corresponded to the one collected during the first three months, leaving the rest of the data to be used to test the model. In order to make this exercise more realistic, the model was built without taking a close look to the remaining data. The third anomaly detection model was done using the data from the humidity sensors. The room has five humidity sensors and the periods of time chosen for building and testing the models were the same as in the case of the temperature data.

### 6.3.1. First data analysis for room temperature

The data analysis was done using 2,194 data vectors corresponding to the data gathered between the 1$^{st}$ of November of 2018 and the 30$^{th}$ of January of 2019. From this sample 10 data vectors were left aside as outliers and not considered in the model construction because one or more sensors presented odd values such as "0", "2", "3", "434", "490", "559", etc. These vectors, as explained by the engineer that provided the data, corresponded mostly to sensor malfunctions or periods in which part of the system was turned off for maintenance. These points corresponded to 0.46% of the data, for which it could be predicted that an anomaly detection model based only on the data supplied would at least detect a similar proportion of oddities. In a real case scenario, it would be expected that the results from the anomaly detection algorithms could be contrasted with the aforenamed situations allowing evaluating the causes and discarding a malicious origin of the anomaly.

Table 6.5 shows a summary of the remaining 2,184 data vectors used to build the model indicating the minimum, maximum and average values for each sensor. The original tags of the sensors were modified, leaving only the last digit as a mean of traceability with the original data, in order to preserve the confidentiality of the real data source.

*Table 6.5: Summary of temperature room model training data in data centre*

| Sensor | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9A | T9B | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 18 | 21 | 20 | 20 | 18 | 18 | 18 | 19 | 22 | 23 | 20 |
| Min | 17 | 19 | 19 | 19 | 18 | 17 | 18 | 18 | 21 | 22 | 17 |
| Max | 20 | 22 | 21 | 22 | 19 | 19 | 19 | 19 | 23 | 24 | 24 |

As it can be observed in Table 6.5 each sensors presented a slightly different value distribution, which can be also observed through a time series graph. In order to improve the image resolution only a sample of this data was included in the graph shown in Figure 6.14 in which each sensor is represented with a different colour. However, this behaviour is representative of the whole sample. Higher and lower control limits, which correspond to a degree over and under the maximum and minimum values, respectively is shown in the graph as a reference of the range in which the temperature values should be expected to be. It can be observed from both Table 6.5 and Figure 6.14 that each individual sensor presents a narrower range than the aggregated data

of all sensors. The variation between sensors' measurement ranges could be explained by temperature variations between the spots of the room in which individual sensor is located.

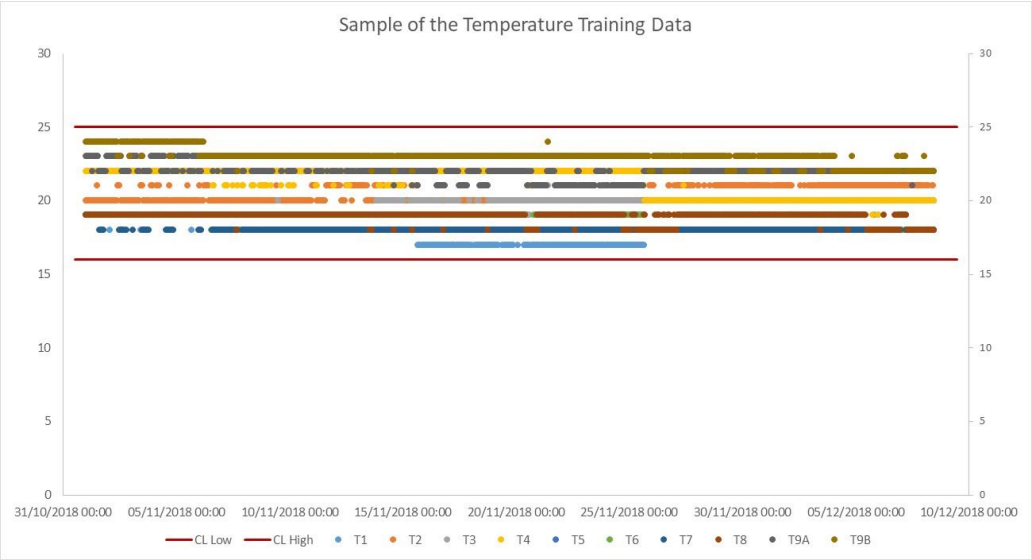



*Figure 6.15: Sample of training data of temperature sensors*

The results of this data analysis showed that it would be more accurate to model the behaviour of each individual sensor rather than the aggregated data from all of them. However, an initial anomaly detection model was built based on the aggregated data model to do a quick test using the data from the remaining eleven months, which corresponded to 7,825 data vectors.

### 6.3.2. First anomaly detection model for room temperature

As just explained, this first anomaly detection mode, aimed to test if the rest of the data would roughly fit into the behaviour of the data used to build the model. Hence, it had simple and naive rules, and two alert examples, and not all the detection possibilities, such as, for example, correlation between different temperature sensors. Two ranges were defined to check anomalies, the first one, which corresponds to a wider range, was given by $T_{min1}$ and $T_{max1}$, and the second and narrower was given by $T_{min2}$ and $T_{max2}$. The first range was used to check if there are significantly high deviations from the expected values which should raise an alert, rather because there is a sensor's malfunction, or because the temperature levels are not under control. If this condition appears on only one of the sensors it is likely that it is the first situation. The second range was used to check whether the temperatures values are within the expected threshold.

The following algorithm was built for the initial temperature anomaly detection model, which generates observations under the following conditions:

- For temperatures outside the range between $T_{min1}$ and $T_{max1}$ in one sensor an observation is made indicating "Prohibited Temperature values"
- For temperatures outside the range between $T_{min1}$ and $T_{max1}$ in more than one sensor an observation is made indicating "Prohibited Temperature values in several sensors"
- For temperatures outside the range between $T_{min2}$ and $T_{max2}$ in one sensor an observation is made indicating " Temperature out of range"
- For temperatures outside the range between $T_{min2}$ and $T_{max2}$ in more than one sensor an observation is made indicating "Temperature out of range in several sensors ".

Additionally, two alerts conditions were programmed based on the observations:

- If the temperature in one or more sensors showed values out of the rage between Tmin2 and Tmax2, the alert "Abnormal temperature values for extended period" was generated.

- For each "Prohibited Temperature values" observation an alert of “Possible sensor malfunction” is generated.

The threshold values were used as an adjustable parameter in the anomaly detection model, which for this experiment were set, as follows:

- Out of range tolerance period = 2 Hours
- Tmin1 = 6°C
- Tmax1 = 35°C
- Tmin2 = 16°C
- Tmax2 = 25°C

It must be noted that $T_{min1}$ = 6°C and $T_{max1}$ = 35°C corresponds to values that demark a range outside of which the temperature of the data centres room or any environmentally controlled building would be ever expected to be in the corresponding geographical location.

### 6.3.3. Results of the first anomaly detection model for room temperature

The data used to test the model corresponded to the period comprehended between the 31st of January and the 23rd of December of 2019, which was equal to 7,825 data vectors. A total of 296 alerts were generated by the anomaly detection algorithm based on this data, which are detailed in Table 6.BMS1. It can be observed that most alerts were concentrated in November and December and several of them obeyed the same cause that corresponded to sensor “T9B” repeatedly presenting values of 26°C, which is only 1°C over the limit defined for this model. Considering that an alert was registered every hour, consecutive alerts were grouped together in Table 6.6 showing a total of 9 periods in which alerts were raised. It is assumed that on a real scenario, rather than having an alert for several days, this should be resolved the first time the alert appears. Observations during this period were registered at 321 instances corresponding to a 4.1% of the total of data vectors. However, 312 of these observations were explained by sensor “T9B” having values only one degree above the threshold. Most of these occurrences were registered in November and December, which is the end of spring in the geographical area in which the data centre is located, in which outdoors temperatures close to 30°C could be registered.

The results of the first temperature anomaly detection model built with the data from the Data Centre BMS show that is possible to model the temperature data, within a range of tolerance and effectively detecting anomalies. It also was observed that there was a moderate number of alarms, all of which were justified. The number of observations was less moderate, however, most of them were attributed to the same anomaly of a single sensor. This means that in a real time context, it is likely that this issue could have been investigated and resolved with the first recurrent observations.

*Table 6.6: Alerts generated by the first anomaly detection model for room temperature*

| Date and time | Alert | Comments |
|---|---|---|
| 18/06/2019 13:00 | Abnormal temperature values for extended period<br>Possible sensor malfunction | Temperature out of range in several sensors registered at 12:00<br>Prohibited Temperature values (such as 0°C, 2°C, and 3°C) in several sensors at 13:00 |
| 11/07/2019 12:00 | Abnormal temperature values for extended period<br>Possible sensor malfunction | Temperature out of range in several sensors for 2 hours.<br>Sensor “T1” has a value of 0 °C |
| 01/11/2019 01:00 | Possible sensor malfunction | Sensor “T1” has a value of 0 °C |
| 25/11/2019 16:00 to 25/11/2019 17:00 | Abnormal temperature values for extended period<br>Possible sensor malfunction | Several sensors have prohibited values which are not possible (over 100 °C) |
| 27/11/2019 03:00 | Possible sensor malfunction | Sensor “T1” has a value of 0 °C |
| 07/12/2019 20:00 to 07/12/2019 23:00 | Abnormal temperature values for extended period | Sensor “T9B” showed a value of 26 °C |

| | | |
|---|---|---|
| 08/12/2019 00:00 to 12/12/2019 16:00 | Abnormal temperature values for extended period | Sensor “T9B” showed a value of 26 °C for several consecutive hours. |
| 18/12/2019 05:00 | Possible sensor malfunction | Sensor “T1” has a value of 0 °C |
| 13/12/2019 18:00 to 22/12/2019 05:00 | Abnormal temperature values for extended period | Sensor “T9B” showed a value of 26 °C for several consecutive hours. |

### 6.3.4. Second data analysis for room temperature

The second data analysis using the data centre room temperature was based on analysing the data of each individual sensor separately. The analysis was based on the same portion of the data that was used in the previous exercise, using this data to build the model and leaving the rest of the data to test it. Figure 6.16 shows and example with a sample of the time series data of sensors T1, T2, T3, and T4.

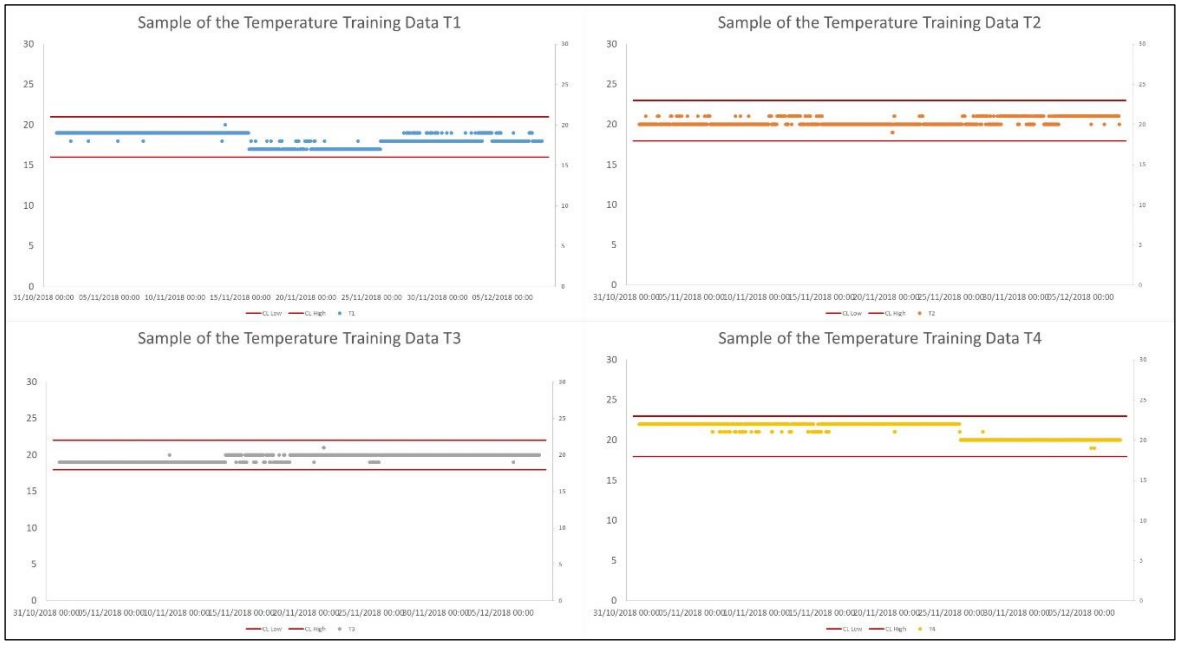



*Figure 6.16: Samples of time series data of individual temperature sensors*

It can be observed, as also described in Table 6.5 that each individual sensor presents a narrower range of values, for which a more sensitive anomaly detection algorithm can be developed if each sensor is tested against its own range rather than against a shared range. Additionally, a correlation check was done between the data of different pairs of sensors to explore the feasibility of building developing anomaly detection rules based on deviations on the data from a sensor respect to the data from other sensors. However, no significant correlations were found in order to use this strategy.

### 6.3.5. Second anomaly detection model for room temperature

The second anomaly detection algorithm had the same rules as the first one, based on threshold analysis, however, each sensor has its own range of permitted values. This also allowed narrowing the acceptable ranges while increasing the tolerance for each individual sensor. In the first model the range of acceptable temperature was set between 16°C and 25°C leaving a tolerance of one degree above and below the maximum and minimum across all sensors, respectively. This meant that there was an acceptable range of 9°C. In the second model, each individual range was set with two degrees below each individual sensor’s minimum and above each sensor’s maximum. Despite this increase in the tolerance, the individual sensors’ ranges varied from 5 to 7 °C. It was expected that by comparing each sensor with its own historical data the anomaly detection algorithm would both be more accurate and have a higher sensitivity.

### 6.3.6. Results of the second anomaly detection model for room temperature

The second anomaly detection model for the Data Centre room temperature was tested using the same data than in the first model. This time only 7 Alerts were generated, which are detailed in Table 6.7. The number of observations also was reduced considerably, registering only 10 of them, which included the alerts. The reason for this is that each sensor was compared with its own past behaviour rather than with the behaviour of the data from all sensors eliminating the high number of observations and alerts generated when sensor "T9B" was giving a value of 26 °C. As the training data provided values between 22°C and 24°C and this second model a tolerance of plus minus 2°C was defined for each individual sensor 26 °C were considered acceptable. Nevertheless, in a real case, the system owner and experts should validate this rule.

*Table 6.7: Alerts generated by the second anomaly detection model for room temperature*

| Date and time | Alert | Comments |
|---|---|---|
| 18/06/2019 13:00 | Abnormal temperature values for extended period<br>Possible sensor malfunction | Temperature out of range in several sensors registered at 12:00<br>Prohibited Temperature values (such as 0°C, 2°C, and 3°C) in several sensors at 13:00 |
| 11/07/2019 12:00 | Abnormal temperature values for extended period<br>Possible sensor malfunction | Temperature out of range in several sensors for 2 hours.<br>Sensor "T1" has a value of 0 °C |
| 01/11/2019 01:00 | Possible sensor malfunction | Sensor "T1" has a value of 0 °C |
| 25/11/2019 16:00 to 25/11/2019 17:00 | Abnormal temperature values for extended period<br>Possible sensor malfunction | Several sensors have prohibited values which are not possible (over 100 °C) |
| 27/11/2019 03:00 | Possible sensor malfunction | Sensor "T1" has a value of 0 °C |
| 18/12/2019 05:00 | Possible sensor malfunction | Sensor "T1" has a value of 0 °C |

### 6.3.7. Room humidity analysis

The second analysis done using the data obtained from the data centre was based on analysing the sensors that measure the room's relative humidity. The room had five temperature sensors from which the same as in the temperature analysis, the data used to build the model, corresponded to the one collected during the first three months, leaving the rest of the data to be used to test the model. In the same way as in the previous exercise, the model was built without taking a close look to the remaining data in order to make this exercise more realistic.

### 6.3.8. Data analysis for humidity

The data analysis was done using 2,194 data vectors corresponding to the data gathered between the 1st of November of 2018 and the 30th of January of 2019. As it was done in the temperature data analysis, the data from each humidity sensor was checked identifying minimum, maximum and average values. A time series graph was also plotted to check visually patterns of behaviour that could be confirmed mathematically a portion of which is shown on Figure 6.17. It can be observed that all sensors present different values at the same point in time, which could be explained by them been placed at different location of the room. However, there is a clear correlation between the behaviour of the sensors, which can be observed by the analogue shapes in the data plotted for each sensor. Hence, the following step consisted on doing a correlation analysis between sensors in order to evaluate the possibility to build an anomaly detection model based on checking correlation between sensors. The sensor tags in this case where also adjusted leaving just an "H" and the last digit of each sensor's original tag.

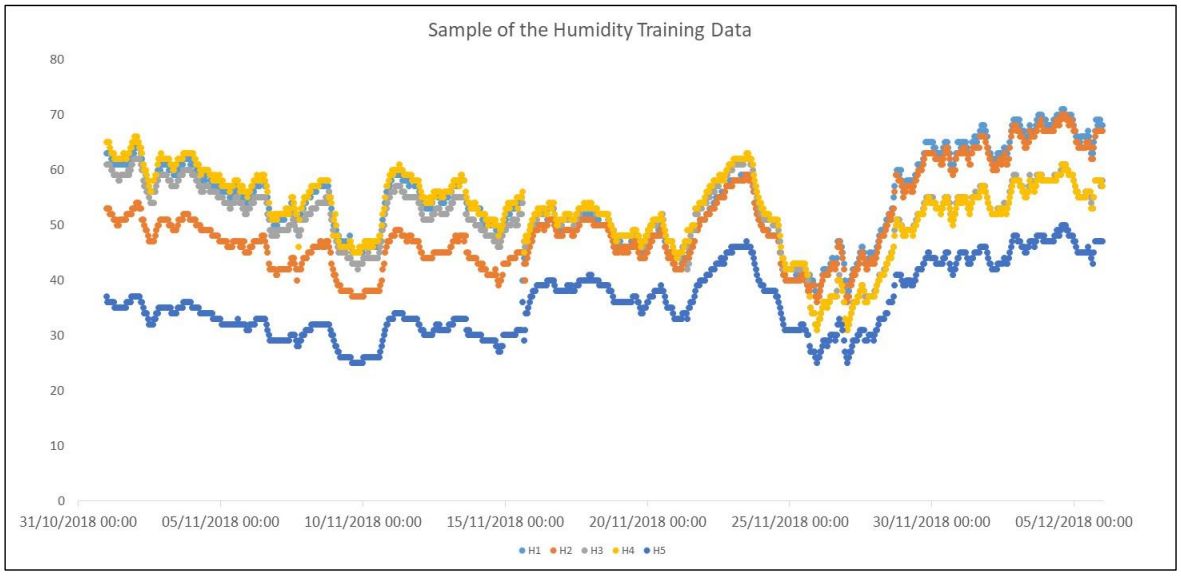



*Figure 6.17: Sample of the training data*

A total of 10 simultaneous correlation checks between sensors was done to cover all possible pairs selecting the four pairs with the best correlation to build a correlation-based anomaly detection model. Figure 6.18 shows the scatter plots of these four humidity sensor pairs.

The potential correlation between the data from temperature and humidity was considered, however, lack of knowledge of the specific locations of the sensors did not allow to infer which humidity sensor could be located close to which temperature sensor. However, a limited amount of correlation checks were done on the temperature and humidity data concluding that it was not possible to create a model.

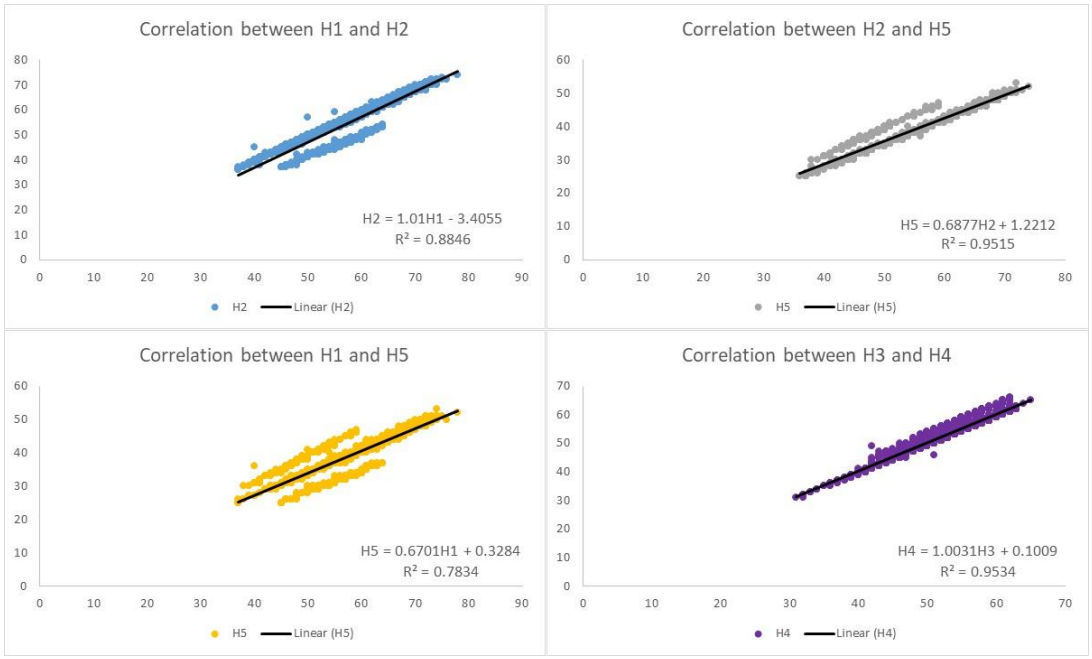



*Figure 6.18:Correlation between humidity sensors*

### 6.3.9. Anomaly detection model for humidity

A correlation-based anomaly detection model was built for humidity based on the four equations shown in Figure 6.18. The detection algorithm built which was design to be applied to the sensor pairs H1-H2, H1, H5, H2-H5, and H3, H4 and was defined considering a different linear equation model for each pair with a degree of tolerance based on the variations observed in the training data.

The following in the general linear equation used to build the detection model:

$$Hi = m \times Hj + b$$

The following is the meaning of each variable of the equation:

- Hj is the sensor data used as independent variable
- Hi is the sensor data used as dependent variable
- m is the slope value of the line
- b is the intercept of the line with the vertical axis

An alarm was triggered indicating an anomaly following the following anomaly detection algorithm in which "t" stands for the tolerance level:

- If $|Hi - (m \times Hj + b)| > t$

Additional, another alert was programmed for the case in which any sensor showed an odd value, such as over 100%, below 0% or too low.

### 6.3.10. Results for anomaly detection model for humidity

The data used to test the model corresponded to the period comprehended between the 31st of January and the 23rd of December of 2019, which was equal to 7,825 data vectors. Alerts were registered at eight points of time, six of which corresponded to instances in which all the sensors have a value of zero. Five of these occasions were at the same time in which alerts were shown in the temperature data, for which it is presumed that probably the system was under maintenance at those times. The correlation-based anomaly detection algorithm generated two alerts, one of which could not be explained just by looking at the data. The other presented a negative value, which was also spotted by the other detection rule. Table 6.19 shows a summary of the alerts registered by the humidity anomaly detection algorithm.

*Table 6.8: Summary of alerts from the humidity anomaly detection algorithm*

| Date and time | Alert | Comments |
|---|---|---|
| 18/06/2019 13:00 | Forbidden Value alert | All sensors display "0" as value |
| 11/07/2019 12:00 | Forbidden Value alert | All sensors display "0" as value |
| 19/09/2019 23:00 | Forbidden Value alert | All sensors display "0" as value |
| 01/11/2019 01:00 | Forbidden Value alert | All sensors display "0" as value |
| 27/11/2019 03:00 | Forbidden Value alert | All sensors display "0" as value |
| 04/12/2019 12:00 | Anomaly alert<br>Forbidden Value alert | Sensor "H4" has a negative value |
| 08/12/2019 01:00 | Anomaly alert | No explanation found |
| 18/12/2019 05:00 | Forbidden Value alert | All sensors display "0" as value |

### 6.3.11. Applicability of the models related to defining IoRs

These models have the potential to be used to generate the following IoRs: IoR501-Sensor data out of limits, IoR502-Misbehaviour in sensor data, IoR503-Indicator that can be correlated to a critical input out of limits,

IoR504-Misbehaviour in data that can be correlated to a critical input misbehaviour, IoR509-Sensor data unavailable, IoR511-Inconsistency between different sources of data. As in the previous example, this can be related to attacks targeting the industrial operations themselves and changing process variables, for example, by an alteration of the control settings, as in the example described in Chapter 4.

The temperature anomaly detection models developed in section 6.3. were based on simple rules comparing threshold values, which proved to be effective in detecting oddities without producing an excessive number of alerts. Especially in the case of the second anomaly detection model. Although these detection algorithms were simpler than the ones developed for the demonstrations of section 6.2, this demonstration had its main value on being based on data from a real BMS. It also gave the opportunity of doing an analysis with data collected over a long period of time, which allowed developing the model with the initial three months period and using the remaining eleven months to test the model.

The anomaly detection model based on correlation between data from different humidity sensors, as the other models, was validated by using different data to build an test the model, concluding that the test data fitted the model. This means that this algorithm has the potential to be capable to detect a misbehaviour in the data. Checking thresholds proved to be useful at a limited extent since the data had a broad distribution of values, however, checking very low and negative values, at least allowed identify evident oddities such as the sensors presented a value of zero or negative. Although it was not possible to correlate temperature and humidity data in this occasion, checking simultaneous presence of oddities in both temperature and humidity data can be useful to explain events. In this case the odd values probably could have been explained by normal or routine activities, such as maintenance tasks. It is expected that in a real setting this information can be known and contrasted with the data.

An important observation is that the low resolution of the data, which was one register per hour corresponding to the hourly average value, would made it not possible to detect anomalies in near real time. However, it was confirmed that it is technically possible to obtain and process this data in real time for which a real time anomaly detection model could be, in theory, developed. In that case, it needs to be checked whether the techniques used in this demonstration are still be valid.

### 6.4. Discussion

Chapter 6 provides practical demonstrations of physical anomaly detection using data collected from different types of sensor including both simple lab-based experiments and real industrial set-ups. This has the objective of illustrating physical anomaly detection for complementing industrial cyber-security detection techniques that has been proposed by several authors, however, not yet widely applied. In the context of continuous risk monitoring, this approach enables the generation of IoRs that can provide a comprehensive view of a cyber-physical system. The whole topic of anomaly and misbehaviour detection has great potential and allows the monitoring not only of cyber-security risks but also of operational risks in general. Since it allows to detect malfunctions, independent of their cause. Further work in more depth and complete practical demonstrations are required including tests with real-time implementations and using data sets that include data from compromised systems.

In [12] several demonstrations were done on tools that are capable of handling behavioural anomaly detection use cases for ICS. These solutions are relevant; however, they present a very limited number of use cases based on data from the physical layer of an ICS or IIoT system. Network and host based anomaly detection are quite relevant for industrial cyber-security and allow defining and observing several IoRs, however, they are not sufficient by themselves since security issues at one level or layer of a cyber-physical system cannot always be solved in another level or layer [13]. Thus, behavioural anomaly detection in industrial network can be enhanced by adding more use cases that are based on sensor data. While the examples developed in this chapter are simple demonstrations, they illustrate how physical variables can be modelled and used to build anomaly detection algorithms. This can be linked to the IoRs of the I/O data threat group (IoR5XX) which contains the IoRs based on data from sensors and actuators.

Anomalies in I/O data are not necessarily caused by a cyber-incident, which was observed also from the data samples obtained from the Data Centre BMS in which there were data vectors that were known to be showing sensor malfunctions. However, rather than considering this as a weakness of this approach and being concerned about a potential high amount of "false positives", this only shows that anomaly detection has a broader scope of utility. System operators can benefit from getting real time information about anomalies in data from sensors and actuators regardless of the causes of these anomalies. As mentioned in section 5.8 non-security related issues that can trigger the observation of an IoR can be divided into two types. The first type are those observations caused by a normal system condition, which would be indeed considered as "false positives". The second type are those that giving indication of a genuine operational risk, but not one that is security-related. By using a probabilistic approach to relate anomalies (expressed as IoRs) to cyber-risks and by using BNs to combine the observation of different IoRs during the same time frame "over reaction" to anomalies can be avoided. Furthermore, anomaly detection systems can "learn" through either manual or automatic feedback used to adjust parameters. The use of machine learning techniques can be highly beneficial in this regard.

Detection of anomalies or misbehaviour in the system's physical variables is an important part of the continuous risk management approach proposed in this thesis because it allows gaps left by typical intrusion detection methods to be filled. Based on the NIST IR 8219 [12] and on a review of capabilities of other commercial tools such as Splunk, finding available commercial tools to implement anomaly detection algorithms and techniques should not be an issue. However, there might be other challenges on integrating these techniques with a continuous risk monitoring system.

An important consideration that needs to be taken in account in anomaly detection, which was not addressed in this thesis, is building an appropriate trust model. Considering redundancy and stablishing roots of trust with a minimum number of components in order to ensure that results of anomaly detection are reliable is crucial. An attacker that wants to remain undetected might attempt to compromise or deceive security systems, too. For example, by remotely changing the configuration or disabling a firewall, or hiding or camouflaging malicious files. In ICS, sophisticated attacks, can also establish mechanisms to disguise their actions by replacing abnormal I/O data with data that emulates a normal behaviour. For example, Stuxnet replays historical values to make them look as current values. Having independent sources of data can help detecting an abnormal behaviour by identifying inconsistencies between different sensors since even indicators of risk based on correlation between operational variables would work only under the assumption that at least some of the variables can be trusted. As used already for safety, redundancy can also be a good security strategy for checking anomalies.

A limitation of the experimental analysis done in this chapter that requires attention is that all the anomaly detection models were tested based on off-line analysis. No real time implementation was done due to limitation of resources. Hence, practical challenges and particularities of real time data analysis were not addressed. Another limitation was the lack of data from a compromised system, which constrained the analysis to be based on modelling only "normal" data to demonstrate that certain deviations from a normal behaviour can actually be detected. Use of data sets that include scenarios such as a cyber-attacks would allow demonstrations of higher value to be performed.

To develop this approach further is important to have the necessary resources to do proofs of concept in a real industrial set-up. An important source of knowledge that can allow anomaly and misbehaviour detection to be implemented more efficiently, is to make use of the body of knowledge, techniques and methods used in operations management. An example of this is the use of statistical process control in predictive maintenance and safety. In general, industrial cyber-security would benefit from a higher level of integration with operations, quality, and safety management, which until now have been developed on totally separate tracks. Much of the knowledge in the fields of control engineering and safety can be useful in the implementation of misbehaviour and anomaly detection for cyber-risk management. The main message of this chapter is that the implementation of physical anomaly detection in ICS serves multiple purposes and should be considered as a useful approach for a comprehensive security and risk monitoring.

# 7. Conclusion

Industrial Control Systems have been crucial to the automation of industrial processes in recent decades allowing more effective, efficient and quality driven development of products. This has evolved through the interconnection of OT and IT systems, which has, on the other hand, made these systems highly vulnerable to cyber-attacks. With the rise of the Industrial Internet of Things, it is expected that industrial processes will become even more tightly integrated with IT. This means that they would also be more dependable on the reliability of these systems, which also includes preserving confidentiality, integrity and availability of data and processes. Hence, cyber-risk management becomes indispensable for ICS and IIoT.

This final chapter discusses how the applied research embodied in the thesis fulfils its original aim and contributes to cyber-risk management in industrial systems. The first section of this chapter describes how the research objectives and research questions posed in Chapter 1 were addressed. The second and third section discuss research contributions, and the fourth section comments on limitations and pending challenges, followed by the fifth and last section, which outlines opportunities for future research work.

## 7.1. Fulfilment of aim, objectives and research questions

The aim of this thesis is to define a methodology for monitoring cyber-risks during the operation of industrial systems so as to keep an organisation's security posture under constant review. Such a Continuous Risk Assessment methodology has been proposed in Chapter 4. It is based on extending the traditional risk management approach described in ISO/IEC 27005 and consists of three phases. The methodology addresses the objectives and implementation of the activities and processes associated with each phase, including workflows and the expected output of each activity. The generation of dynamic risk scores, as illustrated in Section 4.3, can allow risk analysts to base risk treatment plans and security recommendations on factual current information. At the same time, it provides security operations with an additional tool for detecting and evaluating potential security issues.

Table 7.1. lists the research objectives and provides an explanation of how each one of them has been achieved.

*Table 7.1: Implementation of research objectives*

| **Research objectives** | **Implementation** |
|---|---|
| To increase cyber situational awareness related to ICS and IIoT systems. | The Continuous Risk Assessment Methodology introduces a model that allows increasing security awareness by continuous monitoring of IoRs. The use of IoRs to observe conditions that can be associated with cyber-security risks facilitates increased awareness of the systems' exposure to cyber-risks. |
| To define a risk management approach to apply a continuous security risk assessment for ICS and IIoT through adapting and extending existing methods. | The Continuous Risk Assessment methodology is an adaptation of the ISO/IEC 27005 for risk management that enables security risk assessment for ICS and IIoT to be performed continuously. |
| To develop a risk assessment methodology and calculation algorithm to perform security risks calculations in ICS and IIoT during operation. | The risk calculation algorithm is based on Bayesian Networks and conditional probabilities. It makes use of existing tools for building BNs and performing Bayesian calculations. The calculation of baseline or initial risk scores can be done with an independent method based on calculating threat and vulnerability scores associated with different TTPs to estimate their likelihood and use this in computing a final risk likelihood. |
| To validate the methodology by developing worked use cases and validating assumptions with different sources of reference in the industrial sector. | The methodology is validated through a detailed worked example in Chapter 4 in which all the stages are illustrated. Additionally, the development of the IoR Library demonstrates the feasibility to monitor IoRs in real time by providing an extensive list of IoRs with their corresponding rationale and possible means of observation. This work was shared with the security community and reviewed by experts, and it was also presented in a poster competition (winning the first prize), in two MITRE ATT&CK workshops, and shared through three academic papers. |

Table 7.2. shows the list of research questions and summarises how each has been answered by this thesis.

*Table 7.2: Answers to research questions*

| Research questions | Answers |
|---|---|
| What information is needed in order to monitor security risks in IIoT/ICS? | The IoR Library constitutes a resource that allows an organisation to define what information is needed in order to monitor security risks in IIoT/ICS. This information relates to adversary Techniques from the ATT&CK framework and potentially to other TTPs. |
| How can that information be derived from what can actually be measured? | The specifics on how information for monitoring security risks can be derived from what you can measure depend on each system and its corresponding tooling. However, the IoR Library provides a number of "observations" for each IoR that illustrate how the IoR can be measured. |
| How can existing cyber-risk management frameworks be adapted for a more dynamic risk monitoring? | The Continuous Risk Assessment Methodology presented in this thesis does not consist of a re-invention of the risk management processes as conceived by most known frameworks. Rather, it adapts the ISO/IEC 27005 norm to allow the use of security operations information in dynamically monitoring and assessing risks. The adaptation proposed is based on adding additional activities that allow identifying IoRs, implementing a system to continuously monitor them and calculating their effect on residual risk. |
| How can these modifications be introduced? | The methodology proposed defines a Transition phase that allows implementing the continuous risk management approach. It also suggests additional activities as part of the initial or baseline risk assessment to prepare the organisation for the Transition and Continuous Risk Assessment phases. |

## 7.2. Addressing gaps and challenges

The main takeaway regarding continuous cyber-security risk management gathered from the state-of-the-art review is that there is a considerable amount of academic work pointing in that direction, but still the general practice in industry is to perform risk assessments as a static and manual task. While the most well-known and widely used risk management frameworks do highly recommend monitoring risks continuously, they do not give suggestions or guidelines to perform continuous risk assessments. This situation results in a gap between an increasing offer of security tools that claim to enable monitoring risks continuously and the capacity of an organisation to implement continuous risk monitoring in their ICS environment. The following is an evaluation of how this thesis addresses industry and research gaps and challenges.

**Availability and quality of information.**
Despite the availability of large amounts of security-relevant data in many organisations such as logs, network traffic data, malware detection, intrusion detection systems, and threat intelligence feeds, it is considered a challenge to transform these data into meaningful information. In [27] the lack of objective and accurate data for the calculation of risk probabilities is regarded as an obstacle for risk analysis methods. However, according to Douglas Hubbard, a well-known referent in the risk management field, risks are analysed as mean for uncertainty reduction in which designation of probabilities is done not in spite of the lack of data but because of it [58]. Hence systematic methods to make use of the available data, in combination with expert judgement, can provide an organisation with a much better characterisation of security risks compared to a mere qualitative risk assessment. The work developed in this thesis provides some of these systematic methods in which the security relevant data available can be used in near real-time to estimate risks. Expert opinion to assist risk quantification is used in the Baseline Risk Assessment phase, in the calculation of baseline risk scores and in the Transition phase in the definition of conditional probabilities. During the Continuous Risk Assessment phase experts should validate their previous judgements and continuously improve the continuous risk assessment system.

**Complexity of defining rules and identifying normal and abnormal states**
The IoR Library provides some guidance for observations in which the normal and abnormal state of variables at all levels of the system should be defined. As IoRs are represented at a high-level, the challenge of defining

specific rules to differentiate a normal behaviour from an abnormal one is not fully addressed in this thesis. Chapter 6 provides some examples of anomaly detection models based on sensor data, which demonstrate possible types of rules that can be used. Additionally, it is possible to make use of typical SIEM playbooks and use cases knowledge bases or use cases such as the ones published in [12] and [105].

**Limited scope of academic work**
The academic research performed in this work has a broad scope, which allowed addressing the problem of industrial security risk management as a whole rather than only focus on an individual aspect of it. Although this has the drawback of not allowing to develop an more in depth solution for any of the individual aspects explored or including more instantiations and worked examples, it also allowed keeping a holistic perspective.

**Lack of integration between risk management and cyber-security operations.**
This is addressed by the methodology itself and instantiated in the IoR Library, which utilised inputs from security operations and ICS operations to estimate risks. The methodology also defines instances of information sharing such as it can be observed in the continuous risk monitoring and continuous risk assessment workflows for security alerts and risk alerts.

**Cyber-security maturity level**
This work does not address maturity models or approaches for an organisation to develop more mature cyber-security operations and risk management processes. Furthermore, having reached a certain level of maturity will be a requirement for an organisation to adopt the approach proposed in this thesis. However, it was out of scope to provide guidance on how to achieve a more mature cyber-security level. Nevertheless, adopting certain aspects such as using the IoR Library could be done incrementally by any organisation in parallel with a process transformation project with the aim of defining and implementing more mature security processes.

## 7.3. Research Contributions

This research contributed to cyber-security risk management in industrial systems from different perspectives. These perspectives include proposing a Continuous Cyber-Risk Assessment Methodology, developing an Indicators of Risk Library and its instantiation in a Bayesian Network template, and giving examples of physical based behavioural anomaly detection to justify their inclusion as inputs for continuous risk monitoring. Each one of these contributions fulfils the objective of being applicable to any ICS or IIoT implementation, regardless of the industry sector or specific technologies used. Furthermore, they can either be implemented independently of each other, or as is proposed in the present thesis, as part of a logical structure to develop a continuous risk management system.

While this thesis does not, and cannot possibly, cover comprehensively all the aspects and complexity involved in adopting continuous cyber-security risk management in an industrial system, it addresses key points of it. It is in these key points in which the contributions of this thesis work lay. The following sub-sections highlight the research contributions of the work described in Chapters 4, 5, and 6 of this thesis, as well as how all of these aspects connect to each other and to the aim and objectives of this thesis.

### 7.3.1. The Continuous Risk Assessment Methodology

It has to be considered that certain cyber-attacks in an industrial environment could cause severe damages, such as the explosion of an oil pipeline or other serious hazards that can result in deaths, property destruction and environmental damage. This would imply high costs on reparation and compensation of the affected parties, plus temporal shutdowns and damages to their reputation that at the end could result on the business continuity not being viable anymore. Hence, risk Management is fundamental when it is necessary to make decisions that can affect an enterprise's performance or even subsistence. Since the information available to make answer questions such as how much security is enough security will always be incomplete, been able to communicate to decision makers about the possible outcomes and their chances the best as possible and based on factual data is crucial. The chances of occurrence of this will depend on both the organisation's exposure, and the

capabilities and motivations of the adversaries. The approach proposed in this thesis, in the form of a methodology, focuses on monitoring both threats and vulnerabilities. These threats and vulnerabilities are represented as observations of any event or condition that can give evidence of the increment of a cyber-risk, which were termed as IoRs.

The Continuous Risk Assessment Methodology was developed based on explanations of each process involved, its purpose, and its expected outcomes. Artefacts that were part of this methodology were workflow diagrams, and descriptions of activities with their inputs and outputs. A theoretical justification was done through a worked use case that allowed demonstrating how it would work. All of this was documented in Chapter 4 of this thesis and shared with the cyber-security research community through two published papers [130] [131].

The main contribution of the Continuous Risk Assessment Methodology presented in this thesis is to propose a shift of the classic risk management paradigm to become a more dynamic and interactive discipline. This is concretely put in evidence by the following aspects of the methodology:

- A blueprint to introduce continuous risk management
- Modification and extension of a known standard to be adapted for a continuous approach
- A method of automated recalculation of risk probabilities based on Bayesian Networks, which makes use of IoRs to provide dynamic updates of the risk landscape.
- Integration and information exchange with security operations
- Possible to integrate in a variety of industrial contexts
- Use of known standards, frameworks, and methods

Additionally, the concept of IoR, which is introduced in Chapter 4, represents an original contribution that allows the systematic use of a broad set of probabilistic indicators to assess risks continuously. This conceptualisation was developed as an extension of the general idea of IoC and connects all the outputs of this research by providing a common understanding and language for the design and implementation of continuous risk monitoring.

### 7.3.2. The IoR Library

To bring the implementation of continuous risk monitoring into practice it is necessary to have a model for identifying the information that needs to be gathered and how it is linked to different cyber-security risks. The IoR Library served as a means to answer two of the research questions which pointed out to the identification and processing of relevant information to be used for continuous risk monitoring. This artefact serves as both a model and an instrument to identify and use observable data that can allow inferring the presence of events and conditions that carry risk. This observable data was conceptualised under the term “Indicator of Risk”, which, as far as the author’s knowledge goes, has not been widely used in the context of cyber-security risk management. At least not as it is used in this thesis, neither from an operational risk perspective, nor in association with the concept of Indicator of Compromise. This constitutes an original concept, which has been already presented by the author in events such as the MITRE European workshop in May 2020 and June 2021, and through a workshop paper published in August 2021 [152]. Additionally, developing the IoR Library as an extension of a known framework such as the MITRE ATT&CK knowledge base allows increasing its chances of usability.

The IoR Library consists on an Excel spreadsheet providing a list of 95 IoRs with their rationale, observations, and examples, which gives a logical justification of each one of them. The IoR Library does not only allow identifying IoRs, but also links them to individual Techniques from the ICS ATT&CK knowledge base. Hence, it can be used to identify relevant IoRs using a threat model based on TTPs from ATT&CK, and also integrate IoRs with existing tools and processes that are also aligned with its taxonomy. The method for using this artifact was documented in Chapter 5 of this thesis. Once the IoR Library was built, the worked use case of Chapter 4 was re-done, this time using the IoR Library. It was observed that the development of the use case was more straight forward, structured, and faster when the IoR Library was available than in the first instance. This use case, plus the examples developed in Chapter 5, and the validation against ICS tools PoCs done by other researchers served

as justification of the method. Obtaining feedback from the research community served also as a means of validation of both the IoR concept and IoR Library as a product.

The implementation of the IoR Library in a BN template served as an instantiation of the IoR Library used to demonstrate its practical use, and as a useful resource by itself. This serves the purpose of both serving as guidance to build a BN based on the IoR Library and saving time in doing so by avoiding to create and link a high number of nodes from scratch. The BN Template was built to support the theoretical idea that the IoR Library can be pre-loaded in templates and playbooks in order to facilitate its use and thus make it more scalable. This artefact was developed using the software "Genie" and its main aspects together with the method to use it were documented in Chapter 5 of this thesis. Examples of the use of this template were developed through simulation.

The contributions of the IoR Library and the IoR Library implementation in a BN template are:

- Development of the concept of "IoR", which makes use of real time information to extend the concept of Indicator of Compromise (IoC) to "Indicator of Risk" (IoR) as a means to monitor cyber-risks in quasi-real-time.
- Providing a method and a tool to identify relevant information to be used for continuous risk monitoring based on IoRs and independent of the methodologies and technologies used.
- Guidelines on how to map indicators to different attack techniques, and consequently to risks.
- Demonstrating that it is possible to make this approach scalable by providing a template for the instantiation of continuous risk monitoring (in this case represented as a BN, which is the approach used in this thesis).
- Inclusion of a method for dealing with unknown risks.
- Promoting information sharing between risk management, security operations, cyber-security, perimeter security, and safety.

### 7.3.3. Physical based behavioural anomaly detection

The Continuous Risk Assessment Methodology and the IoR Library included the idea that using sensor data for behavioural anomaly detection as an input for continuous risk monitoring in cyber-physical systems could improve visibility of possible threats. This theory was mostly justified by references found in both the academic and industry literature, and further demonstrated with simplified examples in Chapter 6 in which sensor data was used to build anomaly detection models. The evaluation of this work was mostly based on the results of testing the detection algorithms, for which logical and mathematical proof was given. These tests had the limitation of not having access to data from a system under attack or presenting an abnormal behaviour.

As it has been stated earlier, the contributions on this area are limited, since this Chapter 6 was conceived with the purpose of developing practical examples of behavioural models, which was an aspect of the approach mentioned in both Chapter 4 and Chapter 5. The feasibility of developing anomaly detection models with sensor data was explored mostly through the state-of-the-art and literature review; however, some demonstrations were performed to complement this research. While no original outputs were generated in this particular chapter, this aspect of the research helped illustrating some of the ideas presented in the Continuous Risk Assessment Methodology and the IoR Library.

The contributions of Chapter 6 are:

- Inclusion of IoRs from different levels of an industrial system, including field devices and physical variables, such as input and output data from sensors and actuators which allows monitoring IIoT and ICS security considering all variables of the system beyond the ones that traditionally only refer to ICT.

### 7.3.4. How the contributions connect

The documentation of the proposition developed as a result of this research is mostly concentrated in Chapters 4, 5, and 6. These three chapters present different types of outputs which connect through a common thread

which the provision of resources to implement a continuous risk management approach for cyber-security in industrial networks. Chapter 4 was developed with the intention of providing a framework for continuous risk management in which the main processes, activities, and outputs for its implementation and execution were described. While developing the first version of the methodology, including the worked use case, it became evident the need to have a systematic and standardised way to identify IoRs and connect them with known threats. This led to elaborate more on the IoR concept and building the IoR Library, which is developed in Chapter 5. The usefulness of the IoR Library as both a tool and a method for identifying information relevant for continuous risk monitoring was demonstrated by re-doing the worked use case of Chapter 4 using the IoR Library. This example was previously developed without the IoR Library, which in comparison with the latest version, which corresponds to the one presented in this thesis, was more difficult and less structured. Additionally, the IoR Library could also be used as an independent resource which not necessarily requires following the Continuous Risk Assessment methodology. Finally, the work documented in Chapter 6 was developed in parallel with the rest of the research activities with the intention of demonstrating and highlighting the importance of the use of information from industrial operations data to have a complete overview of risks. This work is connected to the IoR Library through the I/O data threat IoR group (IoR5XX) which considers data from operations provided by sensors and actuators.

Overall, the outputs produced in this thesis are all directly or indirectly connected to the research aim and objectives. Regarding this, all these outputs orbit around one single notion, which is the concept of IoR. This leads to conclude that IoRs constitute the common thread that links all the contributions made in this thesis. In this sense, IoRs constitute a novel construct, which can be valuable for future research in terms of providing a common language for referring to indicators associated to probabilistic inference methods and risk-based detection. Even though this might appear as a small potential addition to the current body of knowledge of cyber-security and risk management, conceptualizations can become useful to define the common terms used when describing and thinking about tasks [155]. Hence, the adoption of the IoR concept could be extremely valuable to cyber-security and risk researchers and practitioners. Overall, this is a real novel approach that expands the common thinking and could serve as the basis for an important deal of new research into continuous (as opposed to static) risk assessment and management

Overall contribution of the thesis:

- To promote a shift for a new generation of risk assessments which should be fully integrated with operations management to allow a better informed strategic and tactical decision making.

## 7.4. Limitations

The work developed in this thesis presented several limitations that derived from having constraints on resources and from the broad scope of the topic. Resource constraints, which are typically present in PhD research, included time and access to systems, tools, and data. The broad scope of the topic, which included aspects of risk management, cyber-security operations and industrial systems, also demanded for a broad set of skills, knowledge, and additional resources.

One limitation of the Continuous Risk Assessment Methodology is that it would require considerable effort for its implementation, a fact that could be addressed by the development of automated tools. As the focus of this research was in the methodology and not in tool development. Thus, most of the work was not oriented to overcome this limitation.

The IoR Library presents also a number of limitations. One of these is that the prototype model presented in this thesis does not cover all the Techniques from the ICS ATT&CK knowledge base. It also was not possible to cover all the possible IoRs, for which just a selection that was considered reasonable was included, based on the 40 adversary Techniques of ICS ATT&CK that were reviewed. Two of these Techniques were deprecated by the time of the thesis submission to be merged with other Techniques, which also reveals that effort should be made in maintaining the IoR Library to keep it up-to-date in terms of consistency with the ATT&CK framework. It must also be considered that the information from ATT&CK is obtained from collaborative work and the participation of several members of the cyber-security community, whereas the prototype of the IoR Library was developed

as the work if an individual person. The IoR Library also requires a variety of in-depth technical knowledge from several areas, as well as knowledge related to the specifics of each TTP. Hence, the major contribution of the IoR Library lies in the proposal of the model, since the contribution of its content, although considerable, is still limited.

The work on behavioural anomaly detection using sensor data was also constrained by the data that was available and known to be available during the time of the research. Hence, this work was limited to illustrative examples to complement the main aspects of the thesis.

It is recognised by the researcher that the quality of the outputs of this research and their justification could have been substantially better if the scope of the work had been narrowed down. However, the motivation of exploring all the main aspects and particularities of cyber-security risk monitoring in ICS, IIoT, and cyber-physical systems in general also allowed approaching the research aim and objectives from a holistic perspective. Hence, several aspects proposed in this thesis have the potential of being explored further and more in-depth. This provides opportunities for further research, which will be discussed in the next section.

## 7.5. Real world considerations

**Computing elements for the model implementation**

As the technical implementation of the risk monitoring system was not in scope of this work, no practical evaluation was done regarding the use of computing resources. However, as the SIEM tool would do most of the computing tasks and state of the art SIEM tools, which run in on-premises servers or in the cloud, process thousands of events per second it is not expected that the continuous monitoring of IoRs and updating the dashboard would consume a considerable amount of additional computing resources. Regarding the implementation of likelihood recalculations in the BN, it has to be taken into account that only events that have certain conditions will trigger IoR observations and often an IoR will be observed as a consequence of a series of events, not single events. Each IoR observation will imply the re-calculation of the probabilities of the Techniques and consequently the risks that the IoR is linked to in the BN, only. Hence, it is not anticipated that there would be major issues regarding data processing from the perspective of the BN implementation either, when this approach is implemented in practice.

**Applicability to specific layers of the ICS model**

As remarked in the development of this thesis, there are several demonstrations and proofs of concept regarding the monitoring of system's variables at different layers of an ICS to be used in cyber-security. Some examples of this are the work in [60][106] [117] and [105]. The research published by the NIST [12] and in [53] also demonstrate the capabilities of off-the-shelf tools in this regard. It was also obtained information directly from interviews in the industry about the successful results of a proof of concept on gathering data from a manufacturing plant using the tool SilentDefense and Splunk, which is mentioned in Chapter 5. Hence, there is enough evidence to conclude that it is possible to apply the model based on data from different layers of an ICS system.

**Validation of the approach**

There were several limitations during the research process in regards to the validation of the approach due to its broad scope. Hence, different aspects of the model were validated separately based mostly on peer review, the use of known frameworks and reference to third party proofs of concept. The methodology was built based on the orchestration of different building blocks such as a process based on the ISO/IEC 27005 and an architecture based on existing tools and methods, which, together with the development of a peer-reviewed worked-example, allowed inferring its feasibility. The IoR Library was also based on variables that are known to be measurable, validating a percentage of the IoRs with available proofs of concept. The relationship between IoRs and adversary Techniques was based on the review of the ICS ATT&CK framework and their description and peer reviewed by five security professionals, as well as presented to the security community in three instances. It is recognised that there is room for improvement in terms of testing and validating the approach, which shall be considered in future work.

## 7.6. Future Research Directions

The work developed in this thesis opens several opportunities for further research that could be conducted in the future. Some of the future research directions that were identified are the following:

**Tooling support**
The Continuous Risk Assessment Methodology and the IoR Library would highly benefit from the development of a tooling set that makes their implementation more effective, efficient and repeatable.

**Standard language for IoR implementation**
IoR observation is based on several monitoring tools and systems that generate and process different sorts of data. The IoR Library provides an overview of this data at a high-level, however, it is important to develop forms of processing this data so it can be shared in a standard and machine-readable way. This would also enhance the potential of IoRs to be developed further and be used in templates or pre-loaded in existing tools.

**Adding temporal and logical considerations to IoRs**
The IoR definitions provided in this work do not take into consideration the temporal factor in IoR observations. For example, the BN approach used will always assume that two or more IoRs are observed simultaneously. However, there could be risk scenarios in which IoRs are observed sequentially within the execution of a related Technique. It's also possible that two IoRs need to be observed together for the observation to have a meaning, or that they could be alternative manifestations of the same threat and not been likely to be observed together. Future work should focus on defining a chronological order of IoRs as well as logical relationships between them.

**Formalisation of expert's opinion**
Assignation of conditional probabilities in the BN nodes under the observation of IoRs were at this instance suggested to be done merely based on expert's opinion. Some guidance on criteria that should be taken into account by experts when assigning conditional probabilities is the "degree of influence" which is part of the elements of the IoR Library presented in Chapter 5. An important recommendation is that experts should beware not to give conditional probabilities that are too high when a single IoR is observed, unless they have evidence to tell that this IoR has a reasonable degree of accuracy for detecting threats. This way frequent "false positive" risk alerts shall be avoided. Guidelines with this and other recommendations could be developed with rules and suggestions for a more systematic and repeatable way to assign conditional probabilities.

**Contributions from the community to the IoR Library**
The MITRE ATT&CK Framework is one of numerous examples for successful security community work where individual members have joined forces to create standards, concepts or frameworks. The IoR Library would highly benefit from this type of work model to grow in depth and breadth.

**Further testing and validation of the approach**
Further validation and testing of the approach is needed, since the validation done in this thesis was partial presented several limitations.

**Deeper and broader work in physical-based behavioural anomaly detection**
Several of the references cited during this research argue about the benefits of physical-based anomaly detection strategies for cyber-security risk monitoring but present a limited amount of demonstrations. In this research, these ideas were explored as a means to provide a more complete visibility of ICS security. However, the demonstrations on this thesis were, as well, limited. Towards the end of this PhD program knowledge emerged about the availability of publicly available data sets that can be used for this type of research [156] [157]. Future work should focus on making use of these data sets and creating new ones in order to provide a more substantial justification of the use of models to detect misbehaviour in sensor and actuator data for cyber-security.

**Collaboration with the OT community and safety professionals**

As described in Chapter 3, industrial systems present requirements and characteristics that differ from a typical enterprise ICT system. Control engineers have been working with these systems for decades and have a different perspective from IT security professionals from which ICS continuous risk assessment methods can benefit.

**Consideration on impact changes**

Dynamic risk updates should not only consider awareness on those conditions that can change the estimations of risk likelihood but also on those that can change impact estimations. Further research needs to be done on defining and detecting these conditions and on how this can be used for modifying risk scores.

The different outputs presented as part of the continuous risk management approach for cyber-security in industrial control systems developed in this research present contributions that enable developing more dynamic ways to monitor risks. These contributions are materialised mostly in the model of a Continuous Risk Assessment Methodology and by the proposition of the "Indicator of Risk" concept, and by the IoR Library. The instantiation of the IoR Library in a BN template, which was built as a complementary resource, also contributes in the sense that it serves as both a demonstration of the use of the IoR Library and as a tool by itself. Finally, the development of examples to demonstrate generation of IoRs based on physical anomaly detection allowed shedding some light on this aspect by showing some practical exercises rather than approach the idea just from a theoretical perspective. The limitations related to both, the research process and the research products, far from being an impediment to achieve the research goal, open an invitation for further research. These future directions present a rich and broad set of opportunities for the further development of continuous cyber-risk management in industrial environments.

## References


[1] A.-R. Sadeghi, C. Wachsmann and M. Waidner, “Security and privacy challenges in industrial internet of things,” in *52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*, 2015.

[2] L. J. Wells, J. A. Camelio, C. B. Williams and J. White, “Cyber-physical security challenges in manufacturing systems,” *Manufacturing Letters,* vol. 2, no. 2, pp. 74-77, 2014.

[3] World Economic Forum, “The Global Risks Report 2019,” Geneva, 2019.

[4] World Economic Forum, “The Global Risks Report 2020,” Geneva, 2020.

[5] J. Boehm, N. Curcio, P. Merrath and L. Shento, “The risk-based approach to cybersecurity,” McKinsey & Company, 2019.

[6] K. Dempsey, N. S. Chawla, A. Johnso, R. Johnston, A. C. Jones,, A. Orebaugh, M. Scholl, and K. Stine, “Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations: National Institute of Standards and Technology (NIST) Special Publication 800-137,” 2012.

[7] V. A. Desnitsky, I. V. Kotenko and S. B. Nogin, “Detection of anomalies in data for monitoring of security components in the Internet of Things,” in *XVIII International Conference on Soft Computing and Measurements (SCM)*, 2015.

[8] Cyber-X Labs, “2020 Global ICS & IIoT Risk. A data-driven analysis of vulnerabilities in our industrial and critical infrastructure,” 2020.

[9] European Union Agency for Network and Information Security - ENISA (2017), Baseline Security Recommendations for IoT, 2017.

[10] Gartner, “Gartner Newsroom,” 7 2 2017. [Online]. Available: http://www.gartner.com/newsroom/id/3598917. [Accessed 15 6 2017].

[11] “Research and Markets,” 03 2020. [Online]. Available: https://www.researchandmarkets.com/reports/5003969/industrial-iot-iiot-market-by-device-and?utm_source=MC&utm_medium=Email&utm_code=mzrzucgn3&utm_campaign=1363052+-+Industrial+IoT+(IIoT)+Market+-+Analysis+and+Global+Forecast+to+2025&utm_exec=lico287mtd. [Accessed 04 06 2020].

[12] J. McCarthy, M. Powell, K. Stouffer, C. Tang, T. Zimmerman, W. Barker, T. Ogunyale, D. Wynne and J. Wiltberge, “NISTIR 8219. Securing Manufacturing Industrial Control Systems: Behavioral Anomaly Detection,” McCarthy, J., Powell, M., Stouffer, K., Tang, C., Zimmerman, T., Barker, W., ... & Wiltberger, J. (2018). Securing Manufacturing Industrial Control Systems: Behavioral Anomaly Detection (No. NIST Internal or Interagency Report (NISTIR) 8219 (Draft)). Nati, 2018.

[13] Q. Jing, A. V. Vasilakos, J. Wan, L. Jingwei and D. Qiu, “Security of the Internet of Things: perspectives and challenges,” *Wireless Networks,* vol. 20, no. 8, pp. 2481-2501, 2014.

[14] Health and Safety Executive (HSE), “Cyber Security for Industrial Automation and Control Systems (IACS) OG86.,” Tech. rep., UK Government, 2018.

[15] S. McLaughlin, C. Konstantinou, X. Wang, L. Davi, L, . A. Sadeghi, M. Maniatakos and R. Karri, “The cybersecurity landscape in industrial control systems,” in *Proceedings of the IEEE*, 2016.

[16] Positive Technologies, “Industrial Companies Attack Vectors,” 2018.

[17] “Integrated Risk Management,” RSA, 2020. [Online]. Available: https://www.rsa.com/en-us/products/integrated-risk-management. [Accessed 05 10 2020].

[18] U.S. Department of Energy, “ Cybersecurity Capability Maturity Model,” 2014. [Online]. Available: https://energy.gov/oe/cybersecurity-critical-energy-. [Accessed 30 03 2017].

[19] US Department of Homeland Security, “Nuclear Sector Cyber Security Implementation Guidance,” 2015. [Online]. Available: https://www.dhs.gov/publication/nuclear-sector-cybersecurity-framework-implementation-guidance. [Accessed 20 06 2017].

[20] ISA, “International Society for Automation (ISA),” 04 09 2015. [Online]. Available: https://www.isa.org/isa99/. [Accessed 30 03 2017].

[21] M. Barrett, “Framework for improving critical infrastructure cybersecurity version 1.1 (No. NIST Cybersecurity Framework).,” National Institute of Standards and Technology (NIST), 2018.

[22] M. S. Lund, B. Solhaug and K. Stølen, Model-driven risk analysis: the CORAS approach, Springer Science & Business Media, 2010, p. 4.

[23] J. Freund and J. Jones, Measuring and managing information risk: a FAIR approach, Butterworth-Heinemann, 2014.

[24] International Organization for Standarization (ISO), ISO/IEC 27005:2018. Information technology -Security Techniques - Information security risk management, BSI Standards Publication, 2018.

[25] T. I. O. f. S. (ISO), ISO 31000:2018 - Risk management Guidelines, 2018.

[26] E. Humphreys, Information Security Risk Management Handbook for ISO/IEC 270001, London: BSI, 2010.

[27] Y. Cherdantseva, P. Burnap, A. Blyth, P. J. Eden, H. Soulsby, Stoddart and Kristan, “A review of cyber security risk assessment methods for SCADA systems,” *computers & security 56 (2016),* vol. 56, pp. 1-27, 2016.

[28] M. Henrie, “Cyber security risk management in the SCADA critical infrastructure environment,” *Engineering Management Journal,* vol. 25, no. 2, pp. 38-45, 2013.

[29] P. a. B. W. Saripalli, “Quirc: A quantitative impact and risk assessment framework for cloud security,” in *2010 IEEE 3rd International Conference on Cloud Computing (CLOUD)*, 2010.

[30] FIRST, “Common Vulnerability Scoring System v3.1: Specification Document,” [Online]. Available: https://www.first.org/cvss/v3.1/specification-document. [Accessed 30 3 2020].

[31] M. F. Ramadlan, “Introduction and implementation OWASP Risk Rating Management,” 2013. [Online]. Available: https://www.owasp.org/images/9/9c/Riskratingmanagement-170615172835.pdf. [Accessed 3 4 2020].

[32] ENISA, “Threat and Risk Management,” [Online]. Available: https://www.enisa.europa.eu/topics/threat-risk-management/risk-management. [Accessed 30 03 2020].

[33] British Standards Institution, PAS 555:2013 Cyber security risk - Governance and management Specification, BSi Standards Ltd, 2013.

[34] Joint Task Force, “Risk management framework for information systems and organizations. NIST Special Publication 800-37,” NIST, 2018.

[35] International Electrotechnical Commission, IEC 62443-3-2:2020. Security for industrial automation and control systems - Part 3-2: Security risk assessment for system design, IEC, 2020.

[36] M. Pareek, “What Is Your Risk Appetite?,” *ISACA JOURNAL,* vol. 4, 2013.

[37] P. L. Campbell and J. E. Stamp, “A classification scheme for risk assessment methods,” Sandia National Laboratories, 2004.

[38] The Open Group, FAIR – ISO/IEC 27005 Cookbook, The Open Group, 2010.

[39] The Open Group, Open Group Standard - Risk Analysis (O-RA), The Open Group, 2013.

[40] R. A. Caralli, J. F. Stevens, L. R. Young and W. R. Wilson, “Introducing OCTAVE Allegro: Improving the Information Security Risk Assessment Process,” Software Engineering Institute (SEI). Carnegie Mellon University, 2007.

[41] MITRE, Systems Engineering Guide, The MITRE Corporation, 2014.

[42] J. Wynn, J. Whitmore, G. Upton, L. Spriggs, D. McKinnon, R. McInnes, R. Graubart and L. Clausen, “Threat Assessment and Remediation Analysis Methodology Description,” MITRE, 9 2011. [Online]. Available: https://www.mitre.org/publications/technical-papers/threat-assessment--remediation-analysis-tara. [Accessed 30 3 2020].

[43] “Common Attack Pattern Enumeration and Classification,” The MITRE Corporation, 6 2019. [Online]. Available: https://capec.mitre.org/. [Accessed 3 4 2020].

[44] MITRE Corporation, “MITRE ATT&CK,” 2020. [Online]. Available: https://attack.mitre.org/. [Accessed 3 4 2020].

[45] M. Szpyrka, B. Jasiul, K. Wrona, K and F. Dziedzic, “Telecommunications networks risk assessment with Bayesian networks. (pp. 277-288). Springer,” in *IFIP International Conference on Computer Information Systems and Industrial Management*, 2013.

[46] SANS Institute, “A Qualitative Risk Analysis and Management Tool -,” 2002.

[47] R. S. Ross, “Guide for conducting risk assessments NIST special publication 800-30,” US Dept. Commerce, NIST, 2012.

[48] Joint Task Force, “Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy (Final Public Draft).,” National Institute of Standards and Technology, 2018.

[49] “Risk Manager for IRAM2 Application,” 4 11 2016. [Online]. Available: https://www.slideshare.net/AfrikaMorrisMScACIM/risk-risk-manager-for-iram2-application. [Accessed 3 4 2020].

[50] “The Next Generation of Assessing Information Risk,” 30 3 2015. [Online]. Available: https://www.csoonline.com/article/3504004/the-next-generation-of-assessing-information-risk.html. [Accessed 3 4 2020].

[51] OWASP, “OWASP Risk Rating Calculator,” [Online]. Available: https://www.owasp-risk-rating.com/. [Accessed 30 3 2020].

[52] Tenable Network Security, “Nessus Compliance Checks - Auditing System Configurations and Content,” 21 01 2017. [Online]. Available: https://static.tenable.com/documentation/nessus_compliance_checks.pdf. [Accessed 16 11 2020].

[53] C. M. Hurd and M. V. McCarty, “A survey of security tools for the industrial control system environment,” Idaho National Lab.(INL), Idaho Falls, ID (United States)., 2017.

[54] M. Mateski, C. Trevino, C. Veitch, J. Michalski, J. Harris, S. Maruoka and J. Frye, “ Cyber threat metrics,” Sandia National Laboratories, 2012.

[55] “Dragos Threat Intelligence: WorldView,” Dragos, [Online]. Available: https://dragos.com/dragos-threat-intelligence/. [Accessed 5 4 2020].

[56] M. Pendleton, R. M., Garcia-Lebron, J. H. Cho and S. Xu, “A survey on systems security metrics,” *ACM Computing Surveys (CSUR),* vol. 49, no. 4, pp. 1-35, 2016.

[57] S. S. Stevens, “On the theory of scales of measurement,” 1946.

[58] D. W. Hubbard and R. Seiersen, How to measure anything in cybersecurity risk, Hoboken: Wiley, 2016.

[59] J. Venable, J. Pries-Heje and R. Baskerv, “A Comprehensive Framework for Evaluation in Design Science Research,” in *International Conference on Design Science Research in Information Systems*, Berlin, Heidelberg, 2012.

[60] X. Ding, Y. Tian and Y. Yu, “A Real-Time Big Data Gathering Algorithm Based on Indoor Wireless Sensor Networks for Risk Analysis of Industrial Operations,” *IEEE transactions on industrial informatics 12,* vol. 12, no. 3, pp. 1232-1242, 2016.

[61] E. Chakir, M. Moughit and Y. Khamlichi, “A real-time risk assessment model for intrusion detection systems,” in *International Symposium on Networks, Computers and Communications (ISNCC)*, 2017.

[62] G. Gonzalez-Granadillo, S. Dubus, A. Motzek, J. Garcia-Alfaro, E. Alvarez, M. Merialdo and H. Debar, “Dynamic risk management response system to handle cyber threats,” *Future Generation Computer Systems,* vol. 83, pp. 535-552, 2018.

[63] A. Årnes, K. Sallhammar, H. T. Kjetil, B. Tønns, M. E. Gaup Moe and S. J. Knapskog, “Real-time risk assessment with network sensors and intrusion detection systems,” *Computational Intelligence and Security,* pp. 388-397, 2005.

[64] K. Haslum and A. Arnes, “Multisensor real-time risk assessment using continuous-time hidden markov models,” in *International Conference on Computational and Information Science*, 2006.

[65] K. Haslum, A. Abraham and S. Knapskog, “DIPS: A Framework for Distributed Intrusion Prediction and Prevention Using Hidden Markov Models and Online Fuzzy Risk Assessment,” in *Third International Symposium on Information Assurance and Security*, 2007.

[66] D. Yu-Ting, Q. Hai-Peng and T. Xi-Long, “Real-time risk assessment based on hidden markov model and security configuration,” in *International Conference on Information Science, Electronics and Electrical Engineering.*, 2014.

[67] A. Refsdal and K. Stølen, “Employing key indicators to provide a dynamic risk picture with a notion of confidence,” in *IFIP International Conference on Trust Management*, Heidelberg, 2009.

[68] Q. Zhang, C. Zhou, N. Xiong, Y. Qin, X. Li and S. Huang, “Multimodel-Based Incident Prediction and Risk Assessment in Dynamic Cybersecurity Protection for Industrial Control Systems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 46, no. 10, pp. 1429-1444, 2016.

[69] Q. Zhang, C. Zhou, Y. Tian, N. Xiong and Y. Qin, “A fuzzy probability bayesian network approach for dynamic cybersecurity risk assessment in industrial control systems,” *IEEE Transactions on Industrial Informatics,* vol. 14, no. 6, pp. 2497-2506, 2018.

[70] L. Chen, Y. Chao and Y. Ma, “Risk Warning System Based on Big Data Applied in the Power Informatization of State Grid,” in *3rd International Conference on Information Science and Control Engineering (ICISCE), 2016*, 2016.

[71] I. Kotenko, I. Saenko and S. Ageev, “Countermeasure Security Risks Management in the Internet of Things Based on Fuzzy Logic Inference,” *IEEE Trustcom/BigDataSE/ISPA,* vol. 1, 2015.

[72] J. Wang, K. Fan, W. Mo and D. Xu, “A Method for Information Security Risk Assessment Based on the Dynamic Bayesian Network,” in *2016 International Conference on Networking and Network Applications (NaNA)*, 2016.

[73] J. Wang, M. Neil and N. Fenton, “A Bayesian network approach for cybersecurity risk assessment implementing and extending the FAIR model.,” *Computers & Security,* vol. 89, 2020.

[74] H. Huang and D. Xie, “Real-time Network Risk Evaluation Paradigm-inspired by Immune,” in *2015 11th International Conference on Natural Computation (ICNC)*, 2015.

[75] C. Liu, Y. Zhang, J. Zeng, L. Peng and R. Chen, “Research on Dynamical Security Risk Assessment for the Internet of Things inspired by immunology,” in *2012 Eighth International Conference on Natural Computation (ICNC)*, 2012.

[76] R. Chen, C. M. Liu and L. X. Xiao, “A security situation sense model based on artificial immune system in the internet of things,” *Advanced Materials Research,* vol. 403, 2012.

[77] J. Greensmith, “Securing the Internet of Things with responsive artificial immune systems,” in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, 2015.

[78] X. Liu, J. Zhang, P. Zhu, Q. Tan and W. Yin, “Quantitative cyber-physical security analysis methodology for industrial control systems based on incomplete information Bayesian game,” *Computers & Security,* vol. 102, no. 102138, 2021.

[79] A. Álvarez, “WISER,” 27 07 2016. [Online]. Available: https://www.cyberwiser.eu/system/files/20160727_WISER_D5_2_v10%281%29_0.pdf. [Accessed 08 05 2020].

[80] CyberWiser, “About CyberWiser,” CyberWiser, 2018. [Online]. Available: https://www.cyberwiser.eu/about-cyberwisereu. [Accessed 08 05 2020].

[81] L. Obregon, “Secure architecture for industrial control systems,” SANS Institute InfoSec Reading Room., 2015.

[82] Industrial Internet Consortium, 4 6 2015. [Online]. Available: http://www.iiconsortium.org/IIRA-1-7-ajs.pdf. [Accessed 13 03 2017].

[83] Cyber-X Labs, “2019 Global ICS & IIoT Risk. A data-driven analysis of vulnerabilities in our industrial and critical infrastructure,” 2018.

[84] Claroty, “Claroty Biannual ICS Risk & Vulnerability Report: 2H 2020,” 2021.

[85] K. Stouffer, V. Pillitteri, S. Lightman, M. Abrams and A. Hahn, NIST Special Publication 800-82. Guide to Industrial Control Systems (ICS) Security, (NIST) National Institute of Standards and Technology, 2015.

[86] Fortinet, “Fortinet 2019 Operational Technology Security Trends Report. An Update on the Threat Landscape for ICS and SCADA Systems,” 2019.

[87] E. Byres and M. Fabro, “RISI Online Incident Database,” 28 01 2015. [Online]. Available: https://www.risidata.com/Database/event_date/asc. [Accessed 12 04 2020].

[88] MITRE Corporation, “ATT&CK for Industrial Control Systems,” 2020. [Online]. Available: https://collaborate.mitre.org/attackics/index.php/Main_Page. [Accessed 3 4 2020].

[89] B. Gregory-Brown, “Securing Industrial Control Systems - 2017,” SANS Institute, 2017.

[90] M. Lezzi, M. Lazoi and A. Corallo, “Cybersecurity for Industry 4.0 in the current literature: A reference,” *Computers in Industry,* vol. 103, p. 97–110, 2018.

[91] M. Latorre, “Fundamentals of real-time processing in automation and control,” 04 2014. [Online]. Available: https://www.controleng.com/articles/fundamentals-of-real-time-processing-in-automation-and-control/. [Accessed 12 04 2020].

[92] Hewlett Packard Enterprise, “What are Data Centre Tiers,” [Online]. Available: https://www.hpe.com/uk/en/what-is/data-center-tiers.html. [Accessed 12 04 2020].

[93] Industry Today, “Industrial Cyber Security Market to see Booming Business Sentiments | Honeywell, ABB, Schneider, McAfee,” 15 03 2021. [Online]. Available: https://industrytoday.co.uk/it/industrial-cyber-security-market-to-see-booming-business-sentiments---honeywell--abb--schneider--mcafee. [Accessed 31 03 2021].

[94] SIEMENS, “Siemens ProductCERT and Siemens CERT,” 2020. [Online]. Available: https://new.siemens.com/global/en/products/services/cert.html#SecurityPublications. [Accessed 13 04 2020].

[95] *Project Basecamp - PLC Hacking and Insecure By Design.* [Film]. 2017.

[96] Positive Technologies, “ICS vulnerabilities: 2018 in review,” 11 04 2019. [Online]. Available: https://www.ptsecurity.com/ww-en/analytics/ics-vulnerabilities-2019/. [Accessed 13 04 2020].

[97] O. Andreeva, S. Gordeychik, G. Gritsai and O. Kochetova, “Industrial control systems vulnerabilities statistics.,” Kaspersky Lab, Report., 2016.

[98] Wind River, “Security Vulnerability Response Information,” 07 2019. [Online]. Available: https://www.windriver.com/security/announcements/tcp-ip-network-stack-ipnet-urgent11/. [Accessed 16 04 2020].

[99] R. M. Lee, ICS Active defense and Incident Response 515.2 - Asset Identification and Network Security Monitoring, SANS Institute, 2018.

[100] Cybersecurity & Infrastructure Security Agency, “ICS Alert (ICS-ALERT-11-285-01). Open Automation Software OPC Systems.NET Vulnerability,” 23 08 2018. [Online]. Available: https://us-cert.cisa.gov/ics/alerts/ICS-ALERT-11-285-01 . [Accessed 08 02 2021].

[101] C. Bing, “Hackers try to contaminate Florida town's water supply through computer breach,” 08 02 2021. [Online]. Available: https://www.reuters.com/article/usa-cyber-florida/update-1-hackers-broke-into-florida-towns-water-treatment-plant-attempted-poisoning-sheriff-says-idUSL1N2KE2UE. [Accessed 09 02 2021].

[102] J. Montecinos, Interviewee, *Observations about insecure practices in environmental controls, BMS, and other ICS from an experienced ICS engineer.* [Interview]. 01 03 2018.

[103] A. Cardenas and S. Cruz, “Cyber-physical systems security knowledge area. The Cyber Security Body Of Knowledge (cybok).,” 2019.

[104] S. Hilt, F. Maggi, C. Perine, L. Remorin, M. Rösler and R. Vosseler, “Caught in the Act: Running a Realistic Factory Honeypot to Capture Real Threats,” Trend Micro Research, 2020.

[105] C. Vargas Martiınez and B. Vogel-Heuser, “Towards industrial intrusion preventionsystems: A concept and implementation for reactive protection,” *Applied Sciences,* vol. 8, no. 12, 2018.

[106] V. Desnitsky and I. Kotenko, “Event analysis for security incident management on a perimeter access control system,” in *XIX IEEE International Conference on Soft Computing and Measurements (SCM)*, 2016.

[107] MITRE Corporation, “Monitor Process State,” 13 01 2020. [Online]. Available: https://collaborate.mitre.org/attackics/index.php/Technique/T801. [Accessed 26 04 2020].

[108] C. Ten, J. Hong and C. Liu, “Anomaly detection for cybersecurity of the substations,” *IEEE Transactions on Smart Grid,* vol. 2, no. 4, pp. 865-873, 2011.

[109] O. Linda, M. Manic, T. Vollmer and J. Wright, “Fuzzy logic based anomaly detection for embedded network security cyber sensor. In 2011,” in *IEEE Symposium on Computational Intelligence in Cyber Security (CICS) (pp. )*, 2011.

[110] O. Linda, M. Manic, J. Alves-Foss and T. Vollmer, “Towards resilient critical infrastructures: Application of type-2 fuzzy logic in embedded network security cyber sensor. In 2011,” in *4th International Symposium on Resilient Control Systems*, 2011.

[111] O. Linda, M. Manic and T. McJunkin, “Anomaly detection for resilient control systems using fuzzy-neural data fusion engine. In 2011,” in *4th International Symposium on Resilient Control Systems*, 2011.

[112] F. Sicard, E. Zamai and J. Flaus, “Filters based approach with temporal and combinational constraints for cybersecurity of industrial control systems,” *IFAC-PapersOnLine,* vol. 51, no. 24, pp. 96-103, 2018.

[113] S. Kim, Y. Eun and K.-J. Park, “Stealthy Sensor Attack Detection and Real-Time Performance Recovery for Resilient CPS,” *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS,* 2020.

[114] J. Hong and C. Liu, “Intelligent electronic devices with collaborative intrusion detection systems,” *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 271-281, 2017.

[115] S. Pan, T. Morris and U. Adhikari, “Developing a hybrid intrusion detection system using data mining for power systems.,” *IEEE Transactions on Smart Grid,* vol. 6, no. 6, pp. 3104-3113, 2015.

[116] I. Ullah and Q. Mahmoud, “A hybrid model for anomaly-based intrusion detection in SCADA networks,” in *IEEE International Conference on Big Data*, 2017.

[117] I. Kiss, B. Genge, P. Haller and G. Sebestyén, “Data clustering-based anomaly detection in industrial control systems.,” in *IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2014.

[118] I. Kiss, B. Genge and P. Haller, “A clustering-based approach to detect cyber attacks in process control systems,” in *IEEE 13th international conference on industrial informatics (INDIN)*, 2015.

[119] K. Manandhar, X. Cao, F. Hu and Y. Liu, “Detection of Faults and Attacks Including False Data Injection Attack in Smart Grid Using Kalman Filter,” *EEE transactions on control of network systems,* vol. 1, no. 4, pp. 370-379, 2014.

[120] International Organization for Standardization (ISO), ISO/IEC 27000: 2018. Information technology-security techniques-information security management systems-overview and vocabulary.", 2018.

[121] K. Paine and O. Whitehouse, “Indicators of Compromise (IoCs) and Their Role in Attack Defence,” 6 3 2020. [Online]. Available: https://tools.ietf.org/html/draft-paine-smart-indicators-of-compromise-00. [Accessed 7 5 2020].

[122] D. Rhoades, “Machine actionable indicators of compromise,” in *International Carnahan Conference on Security Technology (ICCST)*, 2014.

[123] A. Rodriguez and V. Chadha, Key Risk Indicators, Risk Books, 2016.

[124] J. Andress, “Working with indicators of compromise,” *Journal Information Systems Security Association (ISSA),* vol. 5, pp. 14-20, 2015.

[125] S. Han, M. Xie, H. Chen and Y. Ling, “Intrusion detection in cyber-physical systems: Techniques and challenges,” *IEEE systems journal,* vol. 8, no. 4, pp. 1052-1062, 2014.

[126] D. Ding, Q. Han, Y. Xiang, X. Ge and X. Zhang, “A survey on security control and attack detection for industrial cyber-physical systems,” *Neurocomputing,* vol. 275, pp. 1674-1683, 2018.

[127] M. Abomhara and G. M. Køien., “Security and privacy in the Internet of Things: Current status and open issues,” in *IEEE 2014 International Conference on Privacy and Security in Mobile Systems (PRISMS)*, 2014.

[128] E. Cook and P. Kearney, “Security Chalenges and Cybercrime. Securing the Internet of Things,” *The Journal,* vol. 29, pp. 22-25, 2015.

[129] Joint task force, “Managing Information Security Risk. NIST Special Publication 800-39,” NIST, 2011.

[130] C. Adaros Boye, P. Kearney and M. Josephs, “Cyber-Risks in the Industrial Internet of Things (IIoT): Towards a Method for Continuous Assessment.,” in *In International Conference on Information Security*, 2018.

[131] C. Adaros-Boye, P. Kearney and M. Josephs, “Continuous Risk Management for Industrial IoT: A Methodological View,” in *International Conference on Risks and Security of Internet and Systems (CRiSIS)*, 2019.

[132] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington and C. B. Thomas, “MITRE ATT&CK: Design and Philosophy,” 03 2020. [Online]. Available: https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf. [Accessed 06 09 2020].

[133] Splunk, “10 Ways to Take the MITRE ATT&CK Framework From Plan to Action - A guide to creating a threat-informed defense for your organization,” 2020. [Online]. Available: https://www.splunk.com/pdfs/ebooks/10-ways-to-take-the-mitre-att-and-ck-framework-from-plan-to-action.pdf. [Accessed 13 07 2020].

[134] Securonix, “How MITRE ATT&CK Alignment Supercharges Your SIEM,” 14 10 2019. [Online]. Available: https://www.securonix.com/how-mitre-attack-alignment-supercharges-your-siem/?gclid=CjwKCAjwjLD4BRAiEiwAg5NBFg9jmX0QZNkL05ovPuFwsrMSuKCEMsTqUVBBVwEW-Xywd3MS53VoQxoCeegQAvD_BwE. [Accessed 13 07 2020].

[135] Exabeam, “Using the MITRE ATT&CK knowledge base to improve Threat Huntng and Incident Response,” 2019. [Online]. Available: https://www.wit.co.th/datasheet/exabeam/Exabeam_Whitepaper_MitreAttack.pdf. [Accessed 13 07 2020].

[136] D. Sony, “Introduction to Bayesian Networks,” Towards Data Science, 08 06 2018. [Online]. Available: https://towardsdatascience.com/introduction-to-bayesian-networks-81031eeed94e#:~:text=Bayesian%20networks%20are%20a%20type,edges%20in%20a%20directed%20graph.. [Accessed 11 06 2020].

[137] J. Wang, K. Fan, W. Mo and D. Xu, “A method for information security risk assessment based on the dynamic bayesian network.,” in *International Conference on Networking and Network Applications (NaNA)*, 2016.

[138] A. Twin and G. Scott, “Delphi Method,” Investopedia, 06 03 2020. [Online]. Available: https://www.investopedia.com/terms/d/delphi-method.asp. [Accessed 06 09 2020].

[139] M. Krisper, J. Dobaj and G. Macher, “Assessing Risk Estimations for Cyber-Security Using Expert Judgment,” in *European Conference on Software Process Improvement*, 2020.

[140] S. M. Rajasooriya, C. P. Tsokos and P. K. Kaluarachchi , “Cyber security: Nonlinear stochastic models for predicting the exploitability,” *Journal of information Security,* vol. 8, no. 2, 2017.

[141] D. Gollmann, Computer security, Wiley, 2011.

[142] H. Foulon and M. Van Den Berghe, “Security Navigator 2021. Research-driven insights to build a safer digital society,” Orange Cyberdefense, 2020.

[143] Forescout, “Cybersecurity in Building Automation Systems (BAS),” Forescout Technologies, Inc., Delaware, 2019.

[144] C. Adaros-Boye, “IoR Library,” 01 05 2021. [Online]. Available: https://tinyurl.com/7hthzpc5.

[145] The MITRE Corporation, “Levels,” 24 12 2019. [Online]. Available: https://collaborate.mitre.org/attackics/index.php/All_Levels. [Accessed 08 11 2020].

[146] The MITRE Corporation, “Location Identification,” 29 09 2020. [Online]. Available: https://collaborate.mitre.org/attackics/index.php/Technique/T0825. [Accessed 26 11 2020].

[147] ICS-CERT, “ICS Advisory (ICSA-14-178-01),” 22 08 2018. [Online]. Available: https://us-cert.cisa.gov/ics/advisories/ICSA-14-178-01. [Accessed 26 11 2020].

[148] Paladion, “SIEM Use Cases - 45 use cases for Security Monitoring,” 20 08 2020. [Online]. Available: https://securereading.com/downloads/45-siem-use-cases-for-security-monitoring-paladion/. [Accessed 29 10 2020].

[149] Dragos, “ICS CYBERSECURITY YEAR IN REVIEW 2020,” Dragos.inc, 2020.

[150] C. Adaros-Boye, “Mapping indicators of risk with ICS ATT&CK TTPs,” [Online]. Available: https://web.tresorit.com/l/lN841uqbRHdXCFzVVX8obs1OEUw.

[151] Anodot, Ultimate Guide to Building a Machine Learning Anomaly Detection System, 2017.

[152] C. Adaros-Boye, P. Kearney, M. Josephs and H. Ulmer, “An Indicators of Risk Library for Industrial Network Security,” in *The 16th International Conference on Availability, Reliability and Security*, 2021.

[153] S. T. March and G. F. Smith, “Design and natural science research on information technology,” *Decision Support Systems,* vol. 15, pp. 251-266, 1995.

[154] T. Morris, “Industrial control system (ics) cyber attack datasets - tommy morris.,” [Online]. Available: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets. [Accessed 26 02 2021].

[155] J. Goh, S. Adepu and K. N. Junejo, “A dataset to support research in the design of secure water treatment systems,” in *International conference on critical information infrastructures security*, 2016.

[156] V. Chandola, B. Arindam and V. Kumar, “Anomaly detection: A survey,” *ACM computing surveys (CSUR),* vol. 3, no. 15, (2009):.

[157] E. Fernandez, “Threat modeling in cyber-physical systems,” in *IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress* , 2016.

[158] I. Alqassem, “Privacy and security requirements framework for the internet of things (IoT),” in *Companion Proceedings of the 36th International Conference on Software Engineering*, 2014.

[159] D. Miorandi, S. Sicari and F. a. C. De Pellegrini, “Internet of things: Vision, applications and research challenges,” *Ad Hoc Networks,* vol. 10, no. 7, pp. 1497-1516, 2012.

[160] BMBF, “Industrie 4.0,” [Online]. Available: https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html. [Accessed 6 9 2019].

[161] Gartner, “IT Glossary,” 2019. [Online]. Available: https://www.gartner.com/it-glossary/operational-technology-ot/. [Accessed 6 9 2019].

[162] T. I. O. f. S. (ISO), ISO 31000:2018 - Risk management Guidelines, 2018.

[163] International Organization for Standardization (ISO), ISO/IEC 29147:2018. Information Technology - Security Techniques - Vulnerability Disclosure, 2018.

[164] ICS-CERT, “ICS Advisory (ICSA-18-107-02),” 17 04 2018. [Online]. Available: https://us-cert.cisa.gov/ics/advisories/ICSA-18-107-02. [Accessed 22 05 2021].

## Glossary of terms

The following set of definitions corresponds to terms and concepts, which are mentioned throughout this thesis and has the objective to make clear how these terms should be understood in this context.

| | |
|---|---|
| **Anomaly** | Is when the value of a variable significantly deviates from its expected or accepted value or behaviour. These non-conforming patterns can be referred to as anomalies, outliers, discordant observations, exceptions, aberrations, surprises, peculiarities or contaminants in different application domains [158]. |
| **Anomaly detection** | Is using different techniques to find anomalies or data patterns, which do not correspond to the expected behaviour. The simplest technique is establishing thresholds or valid ranges of operation. More complex techniques can include building predictive models and comparing the real values with the expected outcomes of the model. |
| **Asset** | In economy, assets are considered in as resources that allow an organisation obtaining either present or future benefits. On other words, an asset can be anything that has value to the organization [120]. The ISA/IEC 62443 standard agrees specifying that an asset is a physical or logical object owned by or under the custody of an organization, which has either a perceived or actual value to it [35]. The Open Group understands an asset in information security as any data, device, or component that supports information-related activities, which can be illicitly accessed, used, disclosed, altered, destroyed, and/or stolen, resulting in loss [38]. |
| **Attack** | A deliberate attempt to misuse a system in which its confidentiality, availability or integrity get compromised. According to ISO/IEC 2700:2018 is an “attempt to destroy, expose, alter, disable, steal or gain unauthorized access to or make unauthorized use of an asset" [120]. |
| **Availability** | Refers to legitimate users of a service having timely and correct access to it. According to ISO/IEC 27000:2018 it is “the property of being accessible and usable upon demand by an authorized entity business” [120]. |
| **Confidentiality** | Means that information is accessed only by authorised entities. According to ISO/IEC 27000 it is the “property that information is not made available or disclosed to unauthorized individuals, entities, or processes” [120]. |
| **Cyber-Physical System (CPS)** | A system that uses computational capabilities to monitor, control, and automate physical processes. These systems are usually distributed, work close to real-time, and include embedded devices, sensors, and wireless connections [159]. With the current development of technology, it is common for Cyber-Physical Systems to be connected with IT systems, which has made them become tightly close to the concept of “The Internet of Things.” Hence, through the course of this Thesis this term will be used as an umbrella to refer to any system that has computational and electromechanical components, independent of the use case. |

| | |
|---|---|
| **Cyber-risk** | Is a possible event which can impact the confidentiality, availability or integrity of a system. A more general definition is the one given by ISO/IEC 27000:2018 which states that a risk is the effect of uncertainty on objectives [24]. From the point of view of doing a risk estimation, risk is defined as the combination of the probability of and the consequences of a potential event [120]. The probability of a cyber-risk is often defined as a combination between threat and vulnerability [38]. |
| **Cyber-security** | It can be defined as the discipline that protects the confidentiality, availability and integrity of systems through prevention, detection, defence and recovery measures, which extends the ISO/IEC 27000:2018 definition of information security described as "the preservation of confidentiality, integrity and availability of information" [120]. |
| **DCS** | Stands for distributed control systems, which is an industrial control system whose components are deployed among a perimeter, for example, in different rooms of a manufacturing plant. |
| **Event (cyber-security event)** | Is a generic term to define something that occurs in a particular point of time, which might potentially have cyber-security implications. According to ISO/IEC 27000:2018, an event is the occurrence or change of a particular set of circumstances [120]. |
| **Incident (cyber-security incident)** | Is an event which can directly or indirectly have a chance of compromising the confidentiality, availability or integrity of a system. An incident can occur through an attack or an unintended error. According to ISO/IEC 27000:2018 terms and definitions, an event corresponds to a single or a series of unwanted or unexpected events that have a significant probability of compromising business operations and threatening information security [120]. |
| **Indicator** | Is a measurement, which provides an estimate or evaluation [120]. This measurement comes from the observation of a symptom which can conduct to a diagnose a condition, but it not necessarily determines this condition with full accuracy, for which it might need to be complemented with additional information. |
| **Indicator of Compromise (IoC)** | It can be defined as "one or more artefacts that relate to a particular security incident or attack" [124] or "artefacts observed on a network or in an operating system that indicate a computer intrusion with a high degree of confidence" [135] |
| **Indicator of Risk (IoR)** | A condition, the observation of which increases the estimated probability of one or more possible threat events. |
| **Industrial Control System (ICS)** | Is a set of interconnected devices that allow monitoring and controlling different variables of a process that is focused on production or maintenance of goods for which interaction with the physical environment is necessary. According to the United States National Institute of Standards and Technology (NIST) an ICS "consists of combinations of control components (e.g., electrical, mechanical, hydraulic, pneumatic) that act together to achieve an industrial objective (e.g., manufacturing, transportation of matter or energy)" [85]. Examples of different types of control systems that can be considered as an ICS are supervisory control and data acquisition (SCADA) systems, distributed control systems (DCS), and |

| | |
|---|---|
| | other configurations based on industrial computers such as Programmable Logic Controllers (PLC). |
| **Integrity** | Means that a system is not corrupted or suffers unauthorised changes which can be either physical, logical or information related. According to ISO/IEC 27000 this is defined as the property of accuracy and completeness [120]. |
| **Internet of Things (IoT)** | Is a general term to describe a set of interconnected devices and services that allow the interaction between IT networks and the physical environment through objects that have integrated sensors or actuators. A wide ecosystem of interconnected services and devices, such as sensors, consumer products and everyday smart home objects, cars, and industrial and health components that technologies collect, exchange and process data in order to dynamically adapt to a specific context" [9]. IoT has many different applications and could encompass a variety of technologies [160] [161]. |
| **Industrial Internet of Things (IIoT)** | Is a set of interconnected devices and services that facilitate monitoring and controlling industrial processes through computing devices, sensors, actuators, IT systems, and communication networks. According to the Industrial Internet Consortium, "the Industrial Internet is an internet of things, machines, computers and people, enabling intelligent industrial operations using advanced data analytics for transformational business outcomes. It embodies the convergence of the global industrial ecosystem, advanced computing and manufacturing, pervasive sensing and ubiquitous network connectivity" [82]. This concept is understood as a new generation of Industrial Control Systems (ICS) under the IoT paradigm, which also relates with the Industry 4.0 concept. |
| **Industry 4.0** | It is based on the concept of "*Industrie 4.0*", which comes from a German project that promotes manufacturing automation [162]. This term is also associated with the "4th Industrial Revolution" and the idea of using information technologies and connectivity to make processes more efficient, which is directly related with the concepts of IoT and IIoT. |
| **Misbehaviour** | When a system presents anomalies by a sustained period of time which leads to conclude that modifications have been directly or indirectly introduce to its behaviour. |
| **Operation Technologies** | Technologies developed to execute, monitor and control industrial processes. They are mostly based on electro-mechanical systems, which are controlled by devices with computational capabilities. According to Gartner, it corresponds to hardware and software that detects or causes changes, through the direct monitoring and control of industrial equipment, assets, processes and events [163]. On other words, the technologies used in ICS and IIoT systems. |
| **PLC** | Stands for Programmable Logic Controllers, which are a type if industrial computer with inputs and outputs to communicate with field devices such as sensors and actuators, and can be programmed with rules in order to control a process. |
| **Risk** | The ISO 31000 standard defines risks as "the effects of uncertainty in objectives" [164] In simpler words a risk is an event which is possible to occur and cause and have significant consequences. The level of risk is measured considering both how likely is the event to occur and how critical are the impacts. |

| | |
|---|---|
| **Risk Analysis** | Is the use of formal qualitative or quantitative methods to estimate the severity of a risk in a given context. It implies the understanding of the nature of risk and its level [120]. |
| **Risk Assessment** | Is the overall process of identifying, analysing, and evaluating risks [120]. |
| **Risk Evaluation** | Is a decision making process, which establishes whether an organisation can accept or not a risk. According to the ISO/IEC 2700 terms and definitions it is the process of comparing the estimated risk against given risk criteria to determine its significance [120]. |
| **Risk Identification** | The recognition and discovery of the most relevant risks within a certain scope. Set of activities for finding, recognising, and describing risks [120]. |
| **Risk Management** | Corresponds to a set of defined processes that are organised with the goal of been well prepared for facing risks. According to the ISO/IEC 27000 terms and definitions it corresponds to coordinated activities to direct and control an organization with regard to risk [120]. |
| **Risk Treatment** | Process of selection and implementation of measures to modify risk [120]. |
| **SCADA** | Stands for “Supervisory control and data acquisition system” and refers either to a software application which offers a graphical user interface for monitoring and controlling industrial systems or to the whole architecture which also comprehends field devices such as PLCs and other controllers. |
| **SIEM** | A Security Information and Event Management (SIEM) tool, is a software with capabilities of real-time data collection and processing of security information in order to perform data analytics. A SIEM tool allows defining rules to generate security alerts. |
| **Threat** | A condition or action external to the system itself that can result on a vulnerability been exploited. The probability of a threat becoming an incident depends on the attacker's motivation and capacities (knowledge, time, and tools), as well as on the level of vulnerability. According to the ISO/IEC 27000 terms and definitions a threat is the potential cause of an unwanted incident, which may result in harm to a system or organization [120]. |
| **Vulnerability** | A weakness in a system that can facilitate the success of an attack. A vulnerability can exist by design, development, configuration or operational flaws. Some accepted definitions of vulnerability are: “weakness of an asset or control that can be exploited by one or more threats” [120], “a behaviour or set of conditions that allows the violation of an explicit or implicit security policy” [165], and “the probability that an asset will be unable to resist actions of a threat agent” [38]. |

# Appendix A: Details of the context and risk landscape of the worked example of chapter 4

## A.1. Role of a building management system in a data centre

A data centre must satisfy certain environmental requirements in order to preserve business objectives such as performance, availability, and safety. These requirements are usually based on standards such as TIA-924 and the 4 Tier model defined by the Uptime Institute, which describes measures such as redundancy, environmental control, fire and flood protection, and physical access security. To ensure that these requirements are met, various monitoring and control systems need to be deployed, which in a typical case will be administrated by one or more Building Management Systems (BMS). As a BMS runs in an IT infrastructure that needs to communicate with field devices, it meets the general characteristics of an ICS or IIoT system.

The main systems that are monitored and controlled in a data centre building are the following: climate control (temperature and humidity), power supply, electrical distribution and consumption, illumination, access control, CCTV system, and fire alarm system. The scope of this use case is environmental control, and for reasons of simplification the focus will be centered specifically on temperature control. However, other variables monitored by a BMS will be considered as possible sources of data to monitor IoRs. The monitoring and control systems of a data centre focus mainly on assuring continuous supply of power server connectivity and optimal environmental conditions, and on disaster prevention.

Depending on how the control systems are implemented in a data centre, all these variables could be controlled by centralised software such as a BMS or a SCADA system or by independent ones. Typically, the CCTV system runs in a dedicated server and uses separate monitoring software due to the high demand on computational resources from image processing. Control systems could also be implemented separately, in which case, all these subsystems will often be monitored by centralised SCADA software. Possible reasons for control systems being separated include historical reasons such as maintaining systems that have been used in the past while subsequently integrating new systems. From a cyber-security perspective, having independent monitoring and control systems can avoid there being a single point of failure that can be exploited in a cyber-incident.

**Temperature and Climate control**
Although there is no consensus on an ideal temperature for data centre operation, it will be assumed that the operational temperature is 23 °C. Servers typically can tolerate up to 30 °C, or even higher, however, at higher temperatures, servers do not achieve their best performance.

**Electrical power consumption**
Continuity and stability of electrical power is critical in a data centre, and so its availability is assured by the use of uninterrupted power supplies (UPS) and generators. Electrical power consumption is also monitored to assure its stability and quality (e.g. control of power factor and harmonics) and to control operational costs. Typically, in addition to the total power consumption, the power consumption of individual servers and IT equipment is measured separately. The ratio of the total power consumption to the IT power consumption is known as the power usage efficiency or effectiveness (PUE) and it is expected to be near 1.5. Changes in temperature and in server utilisation can affect the power consumption, as well as the PUE.

## A.2. Risk landscape of the Data Centre BMS

The following are elements and characteristics of the system described in Chapter 4 (Section) were taken into account for the risk assessment:

**Field devices**
Devices distributed around the server rooms of the data centre include direct digital controllers (DDCs), temperature and humidity sensors, and electronic valves (actuators) that cause an increase or decrease in the

temperature levels. The DDCs send control signals to actuators according to the temperature measured by the sensors in order to maintain the temperature in the desired range. If the temperature control has a failure which can result on the temperature values not been within the accepted range or in other undesirable condition, an alarm is triggered. The sensors and actuators are hard wired to the DDCs assuming inherent trust. The temperature values are sent to the DDC by analogue signals and the DDC sends also analogue signals to the valves. Thus, the sensors and actuators are not what are considered to be smart devices.

**Communication networks and protocols**

The DDCs connect to a Modbus gateway, which uses standard internet communications (TCP/IP) to connect to servers and other equipment over a Local Area Network (LAN). Other protocols commonly used by controllers such as DDCs and PLCs are Bacnet, Profinet, and Lonworks, but for present purpose the specific protocol is not relevant as the IoRs will not be specified at a low level. A firewall is in place between the LAN and the external networks. It is configured to allow only specific required connections (inbound connections to a VPN server and outbound connections to company email and SMS servers/gateways to allow alerts to be sent to operators in case of critical events). A network server runs VPN server software to provide secure remote connections for occasional remote administration purposes. VPN access may be disabled when not required. No network monitoring and detection mechanisms (e.g. intrusion detection or prevention systems) are in place since in normal circumstances there should no communications to and from external devices, except for a limited e-mail service. The firewall and VPN are configured to log access events, which are stored in a server but these logs are not monitored continuously.

**Control and monitoring Software**

BMS back-end software is installed on an application server that processes operational data. The main user access to this server is through an on-site workstation, which has a PC terminal running BMS client software that connects to the back-end over the LAN. The Workstation PC also runs a VNC (or similar) desktop server so that the BMS front end can be accessed from other machines. Since the temperature is controlled automatically by the DDCs, the main purpose of this software is for monitoring and to access the controllers' configuration for maintenance purposes. However, it is possible to send commands to the controllers, such as, for example, to change the temperature settings.

Besides climate (temperature and humidity), the BMS software is in charge of monitoring other variables which are not visualised in the diagram of Figure 4.11 for simplification matters and because the risk analysis example will be focused on the temperature control. Such variables are electrical power supply and consumption, fire alarms, and illumination. The security cameras (CCTV) and access control are managed by different software.

**Control rules and alerts**

The DDCs are programmed to maintain temperatures at 21ºC including a tolerance margin of plus/minus 5ºC. The recommended humidity is between 40% and 60%. An alert is generated by the BMS whenever the temperature is under 16ºC or exceeds 26ºC. For safety reasons, visual and audible alarms are also automatically triggered by relays on-site if the temperature exceeds 30ºC.

An important part of assessing the cyber-security risks of the system is to identify how cyber-security operations are handled and which are the existing security controls and measures. This includes the cyber security strategy and management approach, if there is one, and the security related processes, protocols, tactics, and procedures. In the case of this data centre, as occurs in many air-gapped or partially isolated systems, the main security strategy is based on physical access control. Therefore, it is assumed that anyone who has gained access within the physical perimeter is trusted. However, it is possible to identify other cyber-security measures even if they are isolated and not part of an overall strategy. The following are the main aspects reviewed to complete the picture of the security controls in place:

**Cyber-security management**

Cyber-security management in the data centre is focused on IT infrastructure and on the various services that run as part of the data centre business processes and operations. The data centre infrastructure, including the building management and control systems are outside the scope of the organisation's cyber-security management system. Therefore, cyber-security controls that are in place that relate to the BMS and the temperature control are mostly ad-hoc and do not form part of a defined cyber-security strategy. Furthermore, data centre standards and regulations do not specifically refer to the cyber-security of the systems that are related to the building infrastructure, leaving a gap where no compliance is enforced in this regard.

**Physical access and authentication**

Physical access to the data centre's premises is restricted to authorised personnel, who are authenticated by means of an ID card, a 4 digit password, and their finger print. Special authorisation is required for visitors and contractors, who need to register. Although they should be accompanied by an authorised member of staff at all times, some contractors might be left alone in the data centre for small periods of time, as sometimes they require to work there for several hours. Personnel only visit the data centre when it is necessary; there is nobody permanently in the area.

**Logical access and authentication**

Logical access to the temperature control system can be obtained by accessing the DDCs directly or through the BMS. A DDC can be configured directly through its own key panel, by connecting a laptop through a serial port, or over the LAN using the BMS software. The BMS can be accessed from the workstation in the premises or remotely, but the remote access would be typically disabled. To configure the DDC using the keyboard a 4 digit security code is required. However, to connect a laptop through a physical port, such as a USB port, no authentication is necessary since, as in most industrial equipment, a physical connection assumes inherent trust. As only authorised personnel enter the area, it is not common practice to disable these ports by default.

**Defence against Brute Force attacks**

The BMS requires authentication through a user ID and password to connect to the application server as a user or as an administrator. There is no particular requirement to use a secure password. There are no defences against brute force attacks in place so an unlimited number of failed login attempts is allowed before a successful one.

**Separation of privileges**

Types of personnel requiring access to the BMS client include users, administrators and engineers. There is no clear differentiation of roles and privileges and most users just share credentials. This includes contractors.

**Event logs**

Events such as access, failed authentication attempts, configuration changes, and software updates are registered and stored in the application server, but they are not monitored. Regarding the configuration of the temperature control settings, there is no logging of any changes or events and there are no configuration management policies in place.

**Software security and updates**

An outdated OS runs in both the application server and the work station. Application of security patches is not performed regularly and is avoided because it is believed that it could affect the functionality of the system.

**Network security**

Network security is based only on isolation, and for this reason almost no network security measures other than the firewall and VPN are in place. The reason for this is that remote connections are supposed to be an exception, and it is expected that the software will mostly be accessed within the perimeter. Although remote connection

is possible using the VPN, policy says that it should be disabled by default and enabled only when it is necessary. However, in practice this is not always verified.

**Malware defence**
Malware detection is not considered important. Anti-malware software runs on the workstation PC, but it is not frequently updated. Laptops that are connected to the system are not required to be checked through a malware scan.

**Backups**
Backups of the system are performed every six months, but there are no assurance processes to audit this or any other cyber-security practice.

The description of the worked example provided in this section, including an overview of its cyber-security risk landscape will serve as a basis to be used in next section to illustrate the Continuous Risk Assessment Methodology.

# Appendix B: Baseline risk analysis method

This Appendix describes the risk analysis method used for the Baseline Risk Assessment in the worked example of Chapter 4, section 4.3 and provides details of the threat and vulnerability score calculations done in the example.

## B.1. Description of the method

The following method is provided just as an example in the context of this thesis, since the main contribution lays in the Continuous Risk Assessment methodology, which should be agnostic to the method used in the phase defined as Baseline Risk Assessment. The main requisite for a baseline risk analysis method for it to be appropriate for the Continuous Risk Assessment methodology is that it should be able to provide a quantitative estimation of probabilities for both technical risks (e.g. risk of a TTP to be executed) and business risks.

**Threat scoring**

Threat scoring is done based on an adaptation of the the CTSA method proposed in the TARA methodology [42]. Table B.1. shows the criteria to evaluate threats based on 5 factors. Equal weights were given to each factor so for each TTP or event, the resulting threat score will be the simple average of the scores obtained in the 5 factors. For each threat, the scoring of factors should be define using the knowledge available at the given moment. As this information might be incomplete, it is expected that there would the scoring process will have a component of subjectivity based on assumptions and "educated guesses".

*Table B.1: Threat scoring matrix*

| Factor | 5 (Very High) | 4 (High) | 3 (Moderate) | 2 (Low) | 1 (Very Low) |
|---|---|---|---|---|---|
| Prior use | Widespread use of TTP reported | Frequent use of TTP reported | Confirmed evidence of TTP use | Evidence of TTP use possible | No evidence of TTP used |
| Required skills | No specific skills required | Generic technical skills | Some knowledge of targeted system | Detailed knowledge of targeted system | Detail of both mission and targeted system |
| Required resources | No resources required | Minimal resources required | Some resources required | Significant resources required | Resources required and consumed |
| Stealth | Not detectable | Detection possible with specialized monitoring | Detection possible with routine monitoring | Detection likely with routine monitoring | TTP obvious without monitoring |
| Attribution | No residual evidence | Some residual evidence, attribution unlikely | Attribution possible from characteristics of the TTP | Same or similar TTPs previously attributed | Signature attack TTP used by adversary |

**Vulnerability Scoring**

Vulnerability scores are based on the CVSS calculator [30]. It is understood that CVSS is meant to be used for rating software, hardware, and firmware vulnerabilities and it is not commonly used for configuration or process related weaknesses. However, by using it to rate the overall vulnerability level it is meant to have a standard and repeatable way of measurement for the vulnerability level related to each TTP or threat event.

**Impact Scoring**

The impact levels were divided in six categories, which could be represented both quantitatively or qualitatively as shown in Table B.2.

*Table B.2: Impact level definition*

| **Impact** | **Probable loss Magnitude** |
|---|---|
| Severe | Costs would lead overall losses in the organisation for at least 2 years or possible lead to bankruptcy |
| High | Extra costs can lead to negative annual results (overall losses) at the end of the current year |
| Significant | Extra costs would imply significant reduction of the expected annual profit (over 20%) |
| Moderate | Extra cost would imply a reduction of the expected annual profit between a 5% and a 20% |
| Low | Extra cost would imply a reduction of the expected annual profit in less than a 5% |
| Very Low | Costs can be covered with the current cybersecurity budget not affecting the organisation's profit |

The impact is not expressed in absolute values but in value ranges that are relative to the organisation's capability to absorb the impact. The qualitative representation of these ranges is used to define the overall risk scoring using a risk calculation matrix, which is presented later in Figure B.2.

**Probability Scoring**

The probability scoring is based on the likelihood calculation matrix shown in Figure B.1. that shows a qualitative estimation of likelihood based on the vulnerability and the threat scores. This is translated to probabilities expressed in percentage ranges using Table B.3.

| **Likelihood** | | | | | |
|---|---|---|---|---|---|
| **Threat Score** | | | | | |
| 5 (Very High) | Moderate | High | Very High | Very High | Very High |
| 4 (High) | Low | Moderate | High | High | Very High |
| 3 (Moderate) | Low | Low | Moderate | Moderate | High |
| 2 (Low) | Very Low | Low | Low | Low | Moderate |
| 1 (Very low) | Very Low | Very Low | Very Low | Low | Low |
| **Vulnerability** | VLow (CVSS 0.1-1.9) | Low (CVSS 2.0-3.9) | Medium (CVSS 4.0-6.9) | High (CVSS 7.0-8.9) | Critical (CVSS 9.0-10.0) |

*Figure B.1: Likelihood calculation matrix*

Using threat and vulnerability scores to estimate a probability range allows using a systematic and repeatable method based on several factors that can influence the probability of a system to suffer certain type of attack. However, it must be kept in mind that as a risk analysis is always based on incomplete information and uncertainty, for which documenting the assumptions that led to the calculation of certain score is important.

*Table B.3: probability equivalence*

| Likelihood rating | Probability range |
|---|---|
| Very Low | Higher or equal than 0% and less than 0.01% |
| Low | Higher or equal than 0.01% and less than 0.1% |
| Moderate | Higher or equal than 0.1% and less than 1% |
| High | Higher or equal than 1% and less than 10% |
| Very High | Higher or equal than 10% and less than 100% |

**Risk scoring**

Risk scores are obtained based on the qualitative impacts and likelihood levels using the matrix shown in Figure B.2.

| **Impact** | **Risk Calculation Matrix** | | | | |
|---|---|---|---|---|---|
| Severe | High | High | Critical | Critical | Critical |
| High | Moderate | High | High | Critical | Critical |
| Significant | Moderate | Moderate | High | High | Critical |
| Moderate | Moderate | Moderate | Moderate | High | High |
| Low | Low | Moderate | Moderate | Moderate | Moderate |
| Very Low | Low | Low | Moderate | Moderate | Moderate |
| **Likelihood** | Very Low | Low | Moderate | High | Very High |

*Figure B.2: Risk calculation matrix*

It is important to keep transparent the traceability of what these qualitative scores mean in quantitative terms, especially when risks are presented to management and decision makers who would understand better risks that are expressed in terms of probabilities and monetary values. It is also valid to use ranges or probability

distributions rather than single estimates in order to represent uncertainty and give a more complete overview of the possible outcomes of a threat scenario. For example, a low risk can be also expressed as having a probability of less than 0.01% of a loss that can in the worst case reduce in a 5% the expected annual profit of the company. It is also considered as a low risk if there is a probability of occurrence of less than 0.1% and the expected impact can be absorbed with activities that are considered within the cyber-security budget.

## B.2. Development of the method for example of section 4.3

The following section of this appendix describes the rationale through which the threat, vulnerability, and probability scores were obtained for the baseline risk assessment of the worked example of Chapter 4, which is described in section 4.3

**T1078 - Valid Account (VPN)**

Moderate Likelihood (0.1% to 1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 4 |
| Required resources | 4 |
| Stealth | 3 |
| Attribution | 3 |
| Total Score | 3 |
| Threat | Moderate |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| **It is frequent for users to share credentials** | 5 |
| **Attack Vector** | N |
| **Attack Complexity** | L |
| **Privileges Required** | L |
| **User Interaction** | N |
| **Scope** | C |
| **Confidentiality** | L |
| **Integrity** | N |
| **Availability** | N |

**T0865 – Spearphishing Attachment**

Moderate Likelihood (0.1% to 1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 5 |
| Required skills | 5 |
| Required resources | 3 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 3 |
| Threat | Moderate |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| **Lack of cybersecurity awareness and training to personnel** | 7.4 |
| **Attack Vector** | N |
| **Attack Complexity** | L |
| **Privileges Required** | N |
| **User Interaction** | R |
| **Scope** | C |
| **Confidentiality** | H |
| **Integrity** | N |
| **Availability** | N |

**T1110 - Brute force attack**

Low Likelihood (0.01% to 0.1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 3 |
| Required resources | 3 |
| Stealth | 2 |
| Attribution | 3 |
| Total Score | 2 |
| Threat | Low |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| **Weak authentication to access applications** | 5.9 |
| **Attack Vector** | N |
| **Attack Complexity** | H |
| **Privileges Required** | N |
| **User Interaction** | N |
| **Scope** | U |
| **Confidentiality** | H |
| **Integrity** | N |
| **Availability** | N |

**T0822-External Remote Services**

Moderate Likelihood (0.1% to 1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 4 |
| Required resources | 4 |
| Stealth | 3 |
| Attribution | 3 |
| Total Score | 3 |
| Threat | Moderate |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| **The same network server will manage other networks giving the possibility to jump from one network to another** | 5 |
| **Attack Vector** | N |
| **Attack Complexity** | L |
| **Privileges Required** | L |
| **User Interaction** | N |
| **Scope** | C |
| **Confidentiality** | L |
| **Integrity** | N |
| **Availability** | N |

**T1078 - Valid Account (BMS)**

Moderate Likelihood (0.1% to 1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 4 |
| Required resources | 4 |
| Stealth | 3 |
| Attribution | 3 |
| Total Score | 3 |
| Threat | Moderate |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| **It is frequent for users to share credentials** | 5 |
| **Attack Vector** | N |
| **Attack Complexity** | L |
| **Privileges Required** | L |
| **User Interaction** | N |
| **Scope** | C |
| **Confidentiality** | L |
| **Integrity** | N |
| **Availability** | N |

**T0847-Replication through removable media (local)**

Moderate Likelihood (0.1% to 1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 4 |
| Attribution | 4 |
| Total Score | 3 |
| Threat | Moderate |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| **Open ports** | 4.3 |
| **Attack Vector** | P |
| **Attack Complexity** | L |
| **Privileges Required** | L |
| **User Interaction** | N |
| **Scope** | U |
| **Confidentiality** | N |
| **Integrity** | H |
| **Availability** | N |

**T1078 - Valid Account (credentials for physical access)**

Moderate Likelihood (0.1% to 1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 3 |
| Required resources | 3 |
| Stealth | 3 |
| Attribution | 3 |
| Total Score | 3 |
| Threat | Moderate |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| It is frequent for users to share credentials | 5 |
| Attack Vector | N |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | C |
| Confidentiality | L |
| Integrity | N |
| Availability | N |

**T0875-Change Program State**

Moderate Likelihood (0.1% to 1%)

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 2 |
| Required resources | 3 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

| Vulnerability | Score |
|---|---|
| Insufficient privilege separation | 9.9 |
| Attack Vector | N |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | C |
| Confidentiality | H |
| Integrity | H |
| Availability | H |

**T0814-Denial of Service**

Moderate Likelihood (0.1% to 1%)

| Factor | Score |
|---|---|
| Prior use | 5 |
| Required skills | 3 |
| Required resources | 3 |
| Stealth | 2 |
| Attribution | 4 |
| Total Score | 3 |
| Threat | Moderate |

| Vulnerability | Score |
|---|---|
| Field devices with known vulnerabilities | 5.5 |
| Attack Vector | L |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | U |
| Confidentiality | N |
| Integrity | N |
| Availability | H |

**T0816-Device Restart/Shutdown**

Moderate Likelihood (0.1% to 1%)

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 3 |
| Required resources | 3 |
| Stealth | 2 |
| Attribution | 3 |
| Total Score | 3 |
| Threat | Moderate |

| Vulnerability | Score |
|---|---|
| No clear segregation of duties and permissions for different users | 5.5 |
| Attack Vector | L |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | U |
| Confidentiality | N |
| Integrity | N |
| Availability | H |

**T0833-Modify Control Logic**

High Likelihood (1% to 10%)

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 4 |
| Attribution | 2 |
| Total Score | 3 |
| Threat | Moderate |

| Vulnerability | Score |
|---|---|
| Weak access control to program source code | 9.9 |
| Attack Vector | N |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | C |
| Confidentiality | H |
| Integrity | H |
| Availability | H |

**T0823-Graphical User Interface**

Moderate Likelihood (0.1% to 1%)

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 4 |
| Required resources | 3 |
| Stealth | 3 |
| Attribution | 3 |
| Total Score | 3 |
| Threat | Moderate |

| Vulnerability | Score |
|---|---|
| Insufficient privilege separation | 6.1 |
| Attack Vector | L |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | U |
| Confidentiality | L |
| Integrity | H |
| Availability | N |

**T0838-Modify Alarm settings**

Low Likelihood (0.01% to 0.1%)

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 2 |
| Required resources | 1 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

| Vulnerability | Score |
|---|---|
| Insufficient privilege separation | 6.5 |
| Attack Vector | L |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | C |
| Confidentiality | N |
| Integrity | H |
| Availability | N |

**T0835-Manipulate I/O image**

Low Likelihood (0.01% to 0.1%)

| Factor | Score |
|---|---|
| Prior use | 2 |
| Required skills | 1 |
| Required resources | 1 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

| Vulnerability | Score |
|---|---|
| Weak access control to program source code | 5.2 |
| Attack Vector | L |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | N |
| Scope | U |
| Confidentiality | L |
| Integrity | H |
| Availability | L |

B.3. Adjustments to the vulnerability scores for baseline risk scores, section 4.2.7

The following section shows the adjustments to vulnerability scores done after the risk treatment plan described in section 4.3.1.6 is applied. These score adjustments result in the baseline risk scores presented in section 4.3.1.7.

**T1078 - Valid Account (VPN)**

Mitigations:

- Credentials are unique per user
- Remote connections are disabled by default and if they need to be temporarily enabled for specific tasks this should be registered
- Network monitoring

Low Likelihood (0.01% to 0.1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| | 3.4 |
| Attack Vector | N |
| Attack Complexity | L |
| Privileges Required | H |
| User Interaction | R |
| Scope | C |
| Confidentiality | L |
| Integrity | N |
| Availability | N |

**T0865 – Spearphishing Attachment**

Mitigations:

- Personnel is trained to increase cyber-security awareness
- Remote connections are disabled by default and if they need to be temporarily enabled for specific tasks this should be registered
- Network monitoring

Low Likelihood (0.01% to 0.1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 5 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 1 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| | 5.4 |
| Attack Vector | N |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | R |
| Scope | C |
| Confidentiality | H |
| Integrity | N |
| Availability | N |

**T0822-External Remote Services**

Mitigations:

- Remote connections are disabled by default and if they need to be temporarily enabled for specific tasks this should be registered
- Passive network monitoring is implemented

Low Likelihood (0.01% to 0.1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| | 2.6 |
| Attack Vector | N |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | R |
| Scope | C |
| Confidentiality | L |
| Integrity | N |
| Availability | N |

**T1078 - Valid Account (BMS)**

Mitigations:

- Credentials are unique per user
- Network and log monitoring

Low Likelihood (0.01% to 0.1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 1 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| | 4.1 |
| Attack Vector | L |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | N |
| Scope | U |
| Confidentiality | H |
| Integrity | N |
| Availability | N |

**T0847-Replication through removable media (local)**

Mitigations:

- Malware scans for devices that are connected to the network, periodic scans to workstations
- Unused ports are disabled by default
- Storage devices should be authorised and checked for malware before connecting to any system

Low Likelihood (0.01% to 0.1%)

Threat Score

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

Vulnerability Score

| Vulnerability | Score |
|---|---|
| | 3.8 |
| Attack Vector | P |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | R |
| Scope | U |
| Confidentiality | N |
| Integrity | H |
| Availability | N |

**T0875-Change Program State**

Mitigations:

- Logs for changes and updates are registered and monitored
- Cyber-security policies specifically applied to the BMS were defined
- Policies of work procedures in secure areas were established
- Responsibility for cyber-security in the BMS is allocated in a SOC
- The Continuous Risk Assessment  method is implemented

Very Low Likelihood (less than 0.01%)

| Factor | Score |
|---|---|
| Prior use | 2 |
| Required skills | 1 |
| Required resources | 1 |
| Stealth | 1 |
| Attribution | 1 |
| Total Score | 1 |
| Threat | Very Low |

| Vulnerability | Score |
|---|---|
| | 2.9 |
| Attack Vector | L |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | R |
| Scope | U |
| Confidentiality | L |
| Integrity | L |
| Availability | N |

**T0814-Denial of Service**

Mitigations:

- Cyber-security policies specifically applied to the BMS were defined
- Policies of work procedures in secure areas were established
- Responsibility for cyber-security in the BMS is allocated in a SOC
- The Continuous Risk Assessment  method is implemented
- Passive network monitoring is implemented
- Malware scans for devices that are connected to the network, periodic scans to workstations

Low Likelihood (0.01% to 0.1%)

| Factor | Score |
|---|---|
| Prior use | 4 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

| Vulnerability | Score |
|---|---|
| Field devices with known vulnerabilities | 4.7 |
| Attack Vector | L |
| Attack Complexity | H |
| Privileges Required | L |
| User Interaction | N |
| Scope | U |
| Confidentiality | N |
| Integrity | N |
| Availability | H |

**T0816-Device Restart/Shutdown**

Mitigations:

- Cyber-security policies specifically applied to the BMS were defined
- Policies of work procedures in secure areas were established
- Responsibility for cyber-security in the BMS is allocated in a SOC
- The Continuous Risk Assessment  method is implemented
- Passive network monitoring is implemented
- Malware scans for devices that are connected to the network, periodic scans to workstations

Low Likelihood (0.01% to 0.1%)

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 2 |
| Required resources | 2 |
| Stealth | 2 |
| Attribution | 2 |
| Total Score | 2 |
| Threat | Low |

| Vulnerability | Score |
|---|---|
| No clear segregation of duties and permissions for different users | 5.5 |
| Attack Vector | L |
| Attack Complexity | L |
| Privileges Required | L |
| User Interaction | N |
| Scope | U |
| Confidentiality | N |
| Integrity | N |
| Availability | H |

**T0833-Modify Control Logic**

Mitigations:

- Logs for changes and updates are registered and monitored
- Cyber-security policies specifically applied to the BMS were defined
- Policies of work procedures in secure areas were established
- Responsibility for cyber-security in the BMS is allocated in a SOC
- The Continuous Risk Assessment method is implemented

Very Low Likelihood (less than 0.01%)

| Factor | Score |
|---|---|
| Prior use | 2 |
| Required skills | 1 |
| Required resources | 1 |
| Stealth | 1 |
| Attribution | 1 |
| Total Score | 1 |
| Threat | Very Low |

| Vulnerability | Score |
|---|---|
| | 2.9 |
| Attack Vector | L |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | R |
| Scope | U |
| Confidentiality | L |
| Integrity | L |
| Availability | N |

**T0823-Graphical User Interface**

Mitigations:

- Logs for changes and updates are registered and monitored
- Cyber-security policies specifically applied to the BMS were defined
- Policies of work procedures in secure areas were established
- Responsibility for cyber-security in the BMS is allocated in a SOC
- The Continuous Risk Assessment method is implemented

Low Likelihood (0.01% to 0.1%)

| Factor | Score |
|---|---|
| Prior use | 3 |
| Required skills | 2 |
| Required resources | 3 |
| Stealth | 1 |
| Attribution | 1 |
| Total Score | 2 |
| Threat | Low |

| Vulnerability | Score |
|---|---|
| | 2.9 |
| Attack Vector | L |
| Attack Complexity | H |
| Privileges Required | H |
| User Interaction | R |
| Scope | U |
| Confidentiality | L |
| Integrity | L |
| Availability | N |

## Appendix C: Extract from the IoR Library

The present extract from the IoR Library contains the IoR definitions, which consists of the full list of IoRs (Indicators-of-Risk) including their rationale, observations and examples. The full IoR Library can be found in [144], including an individual page for each one of the Techniques within its scope.

### Vulnerabilities

| **IoR001-Remote accesses enabled** | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | When remote access is enabled it might allow an adversary to attempt a remote attack to ICS infrastructure. Means of remote access can be enabled by system administrators for maintenance or data collection purposes [6]. However, they should enable this capability only when it actually needs to be used. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network security monitoring tools alert the presence of devices with remote connections [7].<br>Remote access system status indicates it is enabled.<br>Logs indicate that remote access was enabled and not subsequently disabled. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Remote access via SSH is provided to a gateway server, from which a remote worker can access other resources to perform duties normally carried out on premises. This capability should only be enabled when needed, but sometimes the system administrator forgets to disable it. |
| **Related Techniques** | T0807, T0818, T0822, T0866, T0869, T0883, T0882, T0884, T0885 |

| **IoR002-Unnecesary open ports** | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Unauthorised connections might be enabled to industrial network hosts or gateways via HTTP or non-standard ports which are not disabled or blocked by the firewall. Unnecessary ports should be disabled by default. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network security monitoring or port scan tools indicate that a typically unused port is open.<br>Logs indicate that a typically unused port has been enabled and not subsequently disabled. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Port 53 TCP/UDP is assigned to the DNS protocol and, hence, it is not needed in a premise in which connections to the public internet are forbidden. However, it might be necessary to connect to the website of an ICS vendor to downloads updates, for which this port could be temporarily enabled. |
| **Related Techniques** | T0803, T0804, T0805, T0807, T0808, T0847, T0882, T0883, T0884 |

| **IoR003-Unnecesary devices connected** | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Unnecessary devices connected can increase the amount of attack vectors. Identifying connected assets and their dependencies is a crucial part of ICS security monitoring. |
| **Observations**<br>*Examples of how this IoR can be observed* | A network mapping tool identifies a connected device that is not part of the inventory.<br>Logs indicated that a new device was connected to the network. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | During the initial asset inventory, a device already connected to the network might be identified whose purpose is unclear, but it cannot be disconnected until its purpose is validated. |
| **Related Techniques** | T0805, T0847, T0869, T0882, T0883, T0884, T0885 |

| IoR004-Internet access | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Direct connections to the internet, make systems potential targets of malware, targeted attacks, and phishing. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network security monitoring detects internet traffic.<br>Logs indicated that a connection to the public internet has been enabled and not subsequently disabled. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A temporal connection to the internet is enabled through a DMZ proxy in order to perform a specific task. Sometimes system administrators can forget to disable it when it is not needed anymore. |
| **Related Techniques** | T0807, T0818, T0825, T0866, T0869, T0882, T0883, T0884, T0885 |

| IoR005-Wireless Access Points | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Wireless access points and wireless sensors and instruments can be used in ICS when the system is distributed in a wide perimeter. This increases the risk from adversaries who can get physical proximity but do not have direct physical access to perform MITM attacks or connection of rogue devices. |
| **Observations**<br>*Examples of how this IoR can be observed* | The system has one or several wireless access points. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses a wireless device to gain unauthorised access to the network through a wireless access point.<br>An adversary takes advantage of a wireless access point using an old WLAN security protocols which has weak encryption and allow them listening to the traffic. |
| **Related Techniques** | T8060 |

| IoR006-Unnecesary privileges | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Giving unnecessary privileges to user accounts and processes could allow an adversary that gains access to them to have more possibilities to perform malicious actions on critical system. |
| **Observations**<br>*Examples of how this IoR can be observed* | Not having enough roles defined to give just the privileges of access and actions needed to each user.<br>Insufficient restrictions for users. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Users with too many privileges might be able to change settings and control programs, install unauthorised software or download malicious files just by having access to a system. It is important that privileges for each user are just restricted to the ones they need to perform their regular tasks. |
| **Related Techniques** | T0800, T0807, T0809, T0810, T0811, T0816, T0818, T0823, T0833, T0847, T0873, T0882 |

| IoR007-Inapropriate network segmentation | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Inappropriate network segmentation can facilitate an adversary lateral movement and jumping from enterprise IT systems to OT networks. |
| **Observations**<br>*Examples of how this IoR can be observed* | Possibility to access the industrial network from the enterprise environment. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary with access to the corporate network or to other segments of the enterprise network uses this access to jump to the industrial environment and compromise a workstation. |
| **Related Techniques** | T0818, T0860, T0869, T0884, T0885 |

| IoR008-Outdated OS | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Outdate OS can be missing security patches against known vulnerabilities. Also an OS that has passed the EOL will have no support and no further security updates. |
| **Observations**<br>*Examples of how this IoR can be observed* | Not having the latest version of the OS. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Adversaries might use known means of exploit of OS vulnerabilities, which can allow actions such as arbitrary code execution or a DoS attack. |
| **Related Techniques** | T0810, T0818, T0866 |

| IoR009-Default credentials | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | ICS and IoT devices could not enforce or even not allow to change default credentials, which allows an attacker to easily gain access to the device, and manipulate it or use it for a botnet. |
| **Observations**<br>*Examples of how this IoR can be observed* | Credentials for a device are the ones given by default by the manufacturer.<br>Default credentials of an asset cannot be changed.<br>Credentials used in a software are typical default ones such as "admin/admin" |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Adversaries might know or search in a device manual or in the internet for default credentials of field devices.<br>Adversaries might try typical dictionary passwords such as "admin", "12345", etc. |
| **Related Techniques** | T0812, T0818, T0859 |

| IoR010-Known hardware vulnerabilities | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Hardware vulnerabilities cannot always be fixed by firmware or software patches and might need to be mitigated with network security measures. Hardware vulnerabilities can allow an attacker to bypass authentication, extract access keys from a device, to execute arbitrary firmware code, or create DoS conditions. |
| **Observations**<br>*Examples of how this IoR can be observed* | Hardware that is used in the system has CVEs assigned to identify one or more vulnerabilities or vulnerabilities are listed in the NVD or other public databases.<br>The firmware developer informs of vulnerabilities through a security advisory. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Several vulnerabilities of a specific model of PLC used in the organisation are made public and a firmware upgrade available to fix them, however, there is a risk that this new firmware version will not be compatible with other components of the system, hence, the risk of not upgrading the firmware is accepted. |
| **Related Techniques** | T0814 |

| IoR011-Known software vulnerabilities | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Software vulnerabilities can allow an adversary to bypass authentication, execute arbitrary code, expose sensitive data, and access administrator or higher privilege functions, among other actions. |
| **Observations**<br>*Examples of how this IoR can be observed* | Software that is used in the system has CVEs assigned to identify one or more vulnerabilities or vulnerabilities are listed in the NVD or other public databases.<br>The firmware developer informs of vulnerabilities through a security advisory. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A vulnerability in a software used in the ICS system is published, and patches are made available. However, the software cannot be patched until there is a scheduled maintenance. Hence, this IoR should be observed and remain active until the vulnerability is patched. |
| **Related Techniques** | T0818, T0866 |

| IoR012-Known firmware vulnerabilities | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Firmware vulnerabilities can allow an adversary to bypass authentication, execute arbitrary code, expose sensitive data, and access administrator or higher privilege functions, among other actions. |
| **Observations**<br>*Examples of how this IoR can be observed* | Firmware that is used in the system has CVEs assigned to identify one or more vulnerabilities or vulnerabilities are listed in the NVD or other public databases.<br>The firmware developer informs of vulnerabilities through a security advisory. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A firmware vulnerability is published for a device that is at the end-of-life only workarounds and mitigations are recommended. The organisation decides to retain the risk since there is a plan to upgrade the hardware. This IoR should be observed and remain active while the affected devices are connected to the ICS network. |
| **Related Techniques** | T0800, T0814, T0818 |

| IoR013-Known OS vulnerabilities | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | OS vulnerabilities can allow execution of arbitrary code, execute shell commands, install rootkit malware, lead to a DoS condition. |
| **Observations**<br>*Examples of how this IoR can be observed* | The OS used in servers, workstations or other units has CVEs assigned related to one or more vulnerabilities or vulnerabilities listed in the NVD or other public databases.<br>The OS vendor informs of vulnerabilities through a security advisory. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Adversaries might use known means of exploit of OS vulnerabilities, which can allow actions such as arbitrary code execution or a DoS attack. For example, the "Eternal Blue" exploits the CVE-2017-0144 Remote Code Execution Windows Server Vulnerability. |
| **Related Techniques** | T0818, T0866 |

| IoR014-Lack of authentication | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Some legacy industrial control and network devices lack of an authentication mechanism to be accessed, since they had been designed to work in a local network and in a trusted environment. However, nowadays these devices might be connected to computer networks in which they get exposed to cyber threats |
| **Observations**<br>*Examples of how this IoR can be observed* | One or more devices or programs do not have an authentication mechanism. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A legacy industrial device can be accessed without authentication, allowing its reconfiguration, turning on and off, and restarting. |
| **Related Techniques** | T0833, T0860 |

| IoR015-Insecure authentication | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Some legacy industrial control and network devices have an insecure authentication mechanism or allow authentication to be bypassed |
| **Observations**<br>*Examples of how this IoR can be observed* | One or more devices or programs have an insecure authentication mechanism. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A legacy industrial device has a default password that cannot be changed, which is shared between all the devices of that brand and model.<br>An industrial device has hard-coded credentials. |
| **Related Techniques** | T0833, T0859, T0860 |

| IoR016-Shared accounts | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Accounts for ICS applications might be shared within several users that have the same role. Shared accounts often work on other endpoints in the same network, too. |
| **Observations**<br>*Examples of how this IoR can be observed* | More than one user has access to the same authentication credentials. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might use a shared account to perform a malicious action, in which case, there would not be accountability on which user performed the action. |
| **Related Techniques** | T0859, T0882 |

| IoR017-Anti-malware is disabled | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Disabling the anti-malware software reduces the capability of detecting known malware. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that the anti-malware software was disabled.<br>This IoR will not be observed at the same time as IoR113. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A system administrator temporarily disables the anti-malware software in a workstation because operators complain about this software interference with legitimate processes.<br>An adversary disables the anti-malware software as part of a malicious procedure. |
| **Related Techniques** | T0800, T0801, T0802, T0803, T0804, T0805, T0806, T0807, T0808, T0809, T0810, T0811, T0813, T0814, T0815, T0816, T0824, T0825, T0829, T0833, T0835, T0847, T0865, T0866, T0868, T0870, T0872, T0873, T0875, T0877, T0878, T0880, T0885 |

| IoR018- Firewall is disabled | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Disabling the anti-malware software reduces the capability of filtering potentially malicious traffic. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that the firewall was disabled. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A system administrator temporarily disables the firewall because operators complain about this software interference with legitimate processes.<br>An adversary disables the firewall as part of a malicious procedure. |
| **Related Techniques** | T0865, T0866, T0868, T0870, T0875, T0877, T0878, T0880, T0884, T0885 |

IT system threat

| IoR101-Suspicious account behaviour | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | An unusual account behaviour can be a symptom of an attempt to compromise a system or that the system has already been compromised and the adversary is trying to gain more privileges. |
| **Observations**<br>*Examples of how this IoR can be observed* | Login attempt to a disabled or expired account. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might attempt to create a new account to gain access to a system or to modify the privileges of an account in which they already have access in order to increase their capacity to perform damage. |
| **Related Techniques** | T0807, T0809, T0811, T0812, T0818, T0822, T0823, T0859 |

| IoR102-Failed login attempts | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | This is an indicator of brute force attempts. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate multiple login failures. Examples of observations are multiple login failures for single username, multiple login failures from the same source, and multiple login failures to the same destination. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary attempts to get access to a server or application by trying different credential combinations. |
| **Related Techniques** | T0812 |

| IoR103-Failed login attempts followed by a successful one | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | This is an indicator of a successful brute force attempt. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate multiple login failures followed by a successful one. Examples of observations are multiple login failures followed by a success from the same source, and multiple login failures followed by a success to the same destination. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary attempts to get access to a server or application by trying different credential combinations and succeeds. |
| **Related Techniques** | T0812 |

| IoR104-Files deletion | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Deletion of malicious files can be done by an adversary or by malware to destroy evidence of a compromise. Also legitimate files can be deleted to cause harm to industrial processes. |
| **Observations**<br>*Examples of how this IoR can be observed* | A command to erase files is observed.<br>There is a decrease on the number of files in a given directory. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious procedure deletes files after its execution to remove traces and avoid detection.<br>A malicious procedure deletes system or project files to inflict damage to the industrial operations. |
| **Related Techniques** | T0872 |

| IoR105-Control and Monitoring application suspicious access log | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A suspicious access log can indicate possible use of a control and monitoring application for malicious activity. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logins from an unusual IP or MAC addresses, at unusual times or with unusual frequency. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary gains access to a SCADA user interface to be able to have access to control and monitoring data and potentially execute malicious actions. |
| **Related Techniques** | T0801, T0811, T0823, T0831, T0833, T0859 |

| IoR106-Control and Monitoring application suspicious change logs | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Changes done through an application for monitoring and control can reveal an increased risk of malicious activity, specially changes that can affect operational processes such as changing control settings or disabling control functions. |
| **Observations**<br>*Examples of how this IoR can be observed* | Control and monitoring application logs indicate that control settings were modified. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses a SCADA user interface to modify control settings (e.g. Maroochy case). |
| **Related Techniques** | T0800, T0809, T0823, T0831, T0833, T0859 |

| IoR107-Unknown programs | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Unknown or unauthorised software can increase cybersecurity risks either by introducing vulnerabilities to the system or themselves are malicious programs by (e.g. zero day or not yet identifiable by detection systems malware) |
| **Observations**<br>*Examples of how this IoR can be observed* | Process monitoring detect unknown program running.<br>Logs indicate an attempt to install unknown software.<br>Logs indicate that an unknown program was installed. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A program is installed as part of a malicious procedure. |
| **Related Techniques** | T0800, T0801, T0802, T0803, T0804, T0805, T0806, T0807, T0808, T0809, T0810, T0811, T0814, T0815, T0816, T0824, T0825, T0829, T0833, T0835, T0865, T0868, T0870, T0872, T0873, T0875, T0877, T0878, T0880, T0882, T0885 |

| IoR108-Unknown APIs | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | APIs can increase cybersecurity risks in a system, for example, if it allows sharing too much data between two systems. |
| **Observations**<br>*Examples of how this IoR can be observed* | Process monitoring detect unknown API running. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary develops an specially crafted API to retrieve information from a control and monitoring system. |
| **Related Techniques** | T0874 |

| IoR109-Unknown files | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Examples of files related to malicious activities are those used to install malware, or that are used to execute a malicious process. |
| **Observations**<br>*Examples of how this IoR can be observed* | The amount of files in a directory has increased.<br>Logs indicate that a new file was created. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A file containing a malicious payload<br>A file with information to be referred to by a malicious process, such as DLLs, or port, device, or commands enumeration, among others.<br>A file used to collect data. |
| **Related Techniques** | T0801, T0802, T0803, T0804, T0805, T0806, T0808, T0809, T0811, T0825, T0835, T0847, T0865, T0868, T0870, T0872, T0873, T0874, T0877, T0882 |

| IoR110-File suspicious Change logs | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Changes on a file such as modification of its content or its size could reveal data manipulation or deletion. To define what changes are suspicious a profile of what changes are expected and not in certain files should be defined. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that a file from the historian data base was modified.<br>Logs indicate that a file automatically generated by the ICS process was modified. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might attempt to modify I/O or other data from the process as part of a malicious procedure.<br>An adversary might attempt to modify files with tables or references to process parameters, commands, or communication ports and addresses in order to achieve a goal. |
| **Related Techniques** | T0809, T0811, T0829, T0872 |

| IoR111-Server or workstation suspicious access log | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A suspicious access log can indicate possible malicious activity. |
| **Observations**<br>*Examples of how this IoR can be observed* | Remote access logs from unusual IPs or MAC addresses.<br>Access logs at unusual times or with unusual frequency. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary gains network access to an ICS server as part of an initial access tactic.<br>An adversary gains network access to an ICS server or workstation to steal intellectual property data. |
| **Related Techniques** | T0818, T0825, T0859 |

| IoR112-User unable to access monitor and control application | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a user cannot access applications normally used for monitoring and control this could be caused by a DoS or another malicious activity that is interfering with normal operations. Even if this condition is not caused by an adversary it would require attention. |
| **Observations**<br>*Examples of how this IoR can be observed* | A user reports issue to IT support or directly in the continuous risk monitoring system. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An malicious procedure creates a DoS condition in an application server to interrupt the normal functioning of a monitoring and control application. |
| **Related Techniques** | T0813, T0814, T0815, T0829 |

| IoR113-Malware detected | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Malware can be spread either intentionally or accidentally on ICS servers and workstations making use of unpatched OS and software vulnerabilities. Trojans, Ransomware, as well as specific ICS malware can cause important damages to an ICS environment. |
| **Observations**<br>*Examples of how this IoR can be observed* | An anti-malware software detects malware in an ICS asset such as a server, an HMI, or a workstation.<br>(This IoR will not be observed at the same time as IoR017). |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Typical sources of malware infection are USB drives, connected laptops, and spear-phishing attachments. |
| **Related Techniques** | T0800, T0801, T0802, T0803, T0804, T0805, T0806, T0807, T0808, T0809, T0810, T0811, T0813, T0814, T0815, T0816, T0824, T0825, T0829, T0833, T0835, T0847, T0865, T0866, T0868, T0870, T0872, T0873, T0875, T0877, T0878, T0880, T0882, T0885 |

| IoR114-Suspicious command line parameters | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | CLI parameters related to functions or processes that are not expected to be executed, either because they are not used, expected with less frequency, or expected only under certain conditions can reveal an increased risk of malicious activity. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate the use of unusual CLI commands.<br>Logs indicate the use of blacklisted CLI commands.<br>Logs indicate the use of not whitelisted CLI commands. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious procedure uses CLI to install a Remote Access Trojan (RAT). |
| **Related Techniques** | T0802, T0807, T0809, T0882 |

| IoR115-Data Historian suspicious access log | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A suspicious access log can indicate possible malicious activity. In the case of the Data Historian, data might be accessed to learn about the system's behaviour or tampered with to cause damage. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate access to the Data Historian from an unusual IP or MAC address. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary gains network access to a data historian server as part of an initial access tactic.<br>An adversary gains network access to a data historian server to collect historical data from the process and spoof I/O image to disguise their malicious activities. |
| **Related Techniques** | T0809, T0810, T0818, T0882 |

| IoR116-E-mail server suspicious activities | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Suspicious activities in e-mail servers can be related to spear-phishing or data exfiltration via e-mail. |
| **Observations**<br>*Examples of how this IoR can be observed* | E-mails from suspicious IP addresses are received (e.g. from abroad)<br>E-mails from suspicious senders are received.<br>Large amounts of data are sent or received via email. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary sends an spear-phishing e-mail to employees with access to the ICS environment. |
| **Related Techniques** | T0865, T0882 |

| IoR117-Suspicious OPC commands | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | OPC is a vendor-agnostic protocol used for IT systems to communicate with ICS and access data. For this IoR it is necessary to define a baseline of normally used commands in order to identify suspicious behaviour. |
| **Observations**<br>*Examples of how this IoR can be observed* | Use of OPC commands blacklisted or not whitelisted.<br>Use of certain OPC commands with an unusually high frequency or under unexpected conditions. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses an OPC command to reset all active alarms.<br>An adversary uses OPC commands to establish communication with another server in the network and make queries about the state of the system. |
| **Related Techniques** | T0801, T0802, T0808,T0825, T0868, T0870, T0877 |

| IoR118-Unknown processes running on server or workstation | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | The execution of certain TTPs will imply running processes that can be recognised as unusual or suspicious. For this IoR it might be necessary to define a baseline of processes normally running in a server or workstation in order to identify suspicious behaviour. |
| **Observations**<br>*Examples of how this IoR can be observed* | A process that is not whitelisted is running in the CPU. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary runs code specially crafted to target a particular operation.<br>An adversary uses port scanning tools to collect information about the system, which is not recognised as a whitelisted process.<br>An adversary runs Zero day malware and the processes executed by this malware are not recognised as whitelisted processes. |
| **Related Techniques** | T0801, T0802, T0806, T0807, T0808, T0810, T0814, T0816, T0818, T0825, T0835, T0882 |

| IoR119-Poor performance of processes or high CPU resource consumption | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Some TTPs might be intensive in the use of computing resources usual consumption of computing resources evidencing a higher CPU utilisation and making other processes to underperform. |
| **Observations**<br>*Examples of how this IoR can be observed* | Process monitoring indicates that the % of CPU utilisation is higher than usual.<br>Graphical CPU or Memory usage can help to identify differeces compared to normal usage. e.g. changes on the rate at which memory or CPU usage grows over time |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious procedure is executing malicious code an ICS server or workstation consuming computing resources on the machine where its running.<br>An adversary uses the device memory for crypto currency mining. |
| **Related Techniques** | T0801, T0802, T0810, T0814, T0818 |

| IoR120-Suspicious change in memory usage (server or workstation) | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Malicious activities taking place or malicious code is running on a system will consume resources of the system such as disk space or RAM. If there is an unexpected increase on the consumption of these resources this can reveal a hidden threat. |
| **Observations**<br>*Examples of how this IoR can be observed* | Process monitoring indicates an increase on RAM usage is detected.<br>Process monitoring indicates an increase on disk space usage is detected.<br>Process monitoring indicates a decrease on disk space usage is detected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious procedure is executing malicious code an ICS server or workstation consuming computing resources on the machine where it is running. |
| **Related Techniques** | T0801, T0802, T0809, T0810, T0814, T0818 |

| IoR121-OS suspicious event logs | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Windows or Linux event logs can reveal malicious activity. For this IoR it might be necessary to define a baseline profile of normal OS event logs in order to identify suspicious behaviour. |
| **Observations**<br>*Examples of how this IoR can be observed* | An event ID or Type is used with unusual frequency.<br>An unusual sequence of events is detected.<br>An event log that cab be related to accessing to critical system resources is detected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Adversaries might try to do queries or change settings at an OS level in order to get control of a system. |
| **Related Techniques** | T0818, T0866, T0872 |

| IoR122-Unauthorised changes in project files | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Unauthorised or suspicious changes in project files can reveal an increased risk of these files being compromised. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs reveal changes in project files. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Adversaries may attempt to infect project files (files that contain different types of objects and documentation needed for PLC programs to function) with malicious code. |
| **Related Techniques** | T0873 |

| IoR123-Unknown USB device plugged | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | An USB device can be used to spread malware and steal data. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that an USB device was plugged. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | It is believed that Stuxnet spread into an Iranian nuclear plant through an infected USB drive.<br>An operator plugs an USB device to a workstation computer and infects it with malware, either unintentionally or on purpose. |
| **Related Techniques** | T0847, T0882 |

IT Network threat

| IoR201-VPN suspicious Access log | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A suspicious access log to a VPN can indicate that an unauthorised party gained remote access to the system. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logins from an unusual IP or MAC addresses, at unusual times or with unusual frequency. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might steal credentials or bypass authentication to access a VPN.<br>A malicious insider uses the VPN connection to access the ICS environment and misuse their privileges.<br>A former employee has their access rights not revoked and accesses the ICS environment to steal data or cause damage. |
| **Related Techniques** | T0822 |

| IoR202-High volume of network traffic | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | An unusually high level of network traffic can reveal different sorts of malicious activities such as data exfiltration, execution of malicious commands, port probes, or a DoS attempt. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network monitoring tools detect an unusually high volume of messages been shared in the ICS local network during normal operation (this must exclude backup, maintenance or other batch processes that might explain this). |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is exfiltrating data using standard IP communication networks.<br>A malicious program is trying to establish communication with a C&C unit.<br>A malicious program is successfully communicating with a C&C unit. |
| **Related Techniques** | T0822, T0866, T0869, T0883, T0884, T0885 |

| IoR203-Unusually large inbound/outbound packets | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Network packets that are unusually large could contain malicious payload. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network monitoring tools detect packets that are larger than the usual been shared in the ICS local network. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is exfiltrating data using standard IP communication networks.<br>A malicious program is successfully communicating with a C&C unit.<br>Malicious code is been downloaded through standard IP communication networks. |
| **Related Techniques** | T0813, T0814, T0822, T0866, T0869, T0884, T0882, T0885 |

| IoR204-Network commands and responses do not match | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a command does not get a successful or matching response, this could be sign of commands being blocked, intercepted or tampered with. Examples of this are unexpected or out of time responses. |
| **Observations**<br>*Examples of how this IoR can be observed* | A process sends an error message or reports a malformed or unusual response.<br>A user reports a malformed or unusual response. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is performing a Man in the Middle attack to modify commands.<br>Command messages have been blocked and the system is giving random responses. |
| **Related Techniques** | T0803, T0805, T0813, T0830 |

| IoR205-Unresponded connection requests (port probes) | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | An adversary might send incomplete connection requests for reconnaissance purposes, for example port probes to identified ports used and devices connected to the ICS network. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network monitoring tools detects unresponded connection requests. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is doing port probes techniques to collect information about the ports that are open in the network. |
| **Related Techniques** | T0883 |

| IoR206-Unusual or unexpected commands in network packets | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Some malicious actions might generate unusual patterns of communication. For this IoR it might be necessary to define a baseline profile of normal commands and their frequency of use in the system in order to identify suspicious commands. |
| **Observations**<br>*Examples of how this IoR can be observed* | A command that is expected to be send once a day has been detected to be sent frequently within short period of time.<br>A command that is sent only within exceptional conditions is detected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might send a command to put a field device in stop mode. Commands to put a field device in stop mode should be only sent during maintenance windows.<br>An adversary might send commands to make queries about the system configuration which might not be done in a regular basis. |
| **Related Techniques** | T0803, T0804, T0805, T0806, T0808, T0814, T0824, T0825, T0830, T0831, T0868, T0869, T0870, T0875, T0884, T0885 |

| IoR207-Traffic with malicious signature detected | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Known malware could also be identified while it is transmitted by the IT network or while it uses the IT network to perform a procedure. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network monitoring tools detects traffic with a malicious signature such as known malicious hashes, IP addresses or byte sequence. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious procedure sends messages over a network which correspond to a known malware. |
| **Related Techniques** | T0800, T0801, T0802, T0803, T0804, T0805, T0806, T0808, T0809, T0810, T0811, T0813, T0814, T0815, T0816, T0824, T0825, T0829, T0833, T0835, T0866, T0872, T0873, T0875, T0878, T0880, T0882, T0885 |

| IoR208-Suspicious communication between devices | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | For this IoR it is necessary to define a baseline profile of communication between devices including which devices are expected to establish communication, how often and which are the characteristics to their communication traffic in order to identify anomalies. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network monitoring tools detect communication between two devices which are not meant to communicate.<br>Network monitoring tools detect communication between two devices with a higher frequency or sharing a higher amount of data than expected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses a device from the ICS network as a proxy to access or attack other assets. |
| **Related Techniques** | T0810, T0869, T0884, T0885 |

| IoR209-Unknown device connected to the IT network | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Unidentified or unauthorised hardware connected to the IT network may indicate anomalous activity [10] increasing the risk of an adversary trying to collect information, access other devices, or install malware. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network logs indicate that an unidentified device was connected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A contractor connects their laptop to the ICS network for a specific task, this device does not belong to the asset inventory.<br>An adversary connects a device to eavesdrop or interfere with network communications. |
| **Related Techniques** | T0801, T0802, T0830, T0847, T0860, T0882 |

Field devices threat

| IoR301-Controller suspicious access log | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a controller is directly accessed under unexpected circumstances (e.g. outside an scheduled maintenance period) this might be revealing malicious activities such as an attempt to manipulate the control of an industrial process. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that one or more controllers were accessed in a period of time or under circumstances in which they were not expected to. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary gains physical or remote access to a controller. |
| **Related Techniques** | T0800, T0801, T0809, T0812, T0824, T0831, T0833, T0875, T0877 |

| IoR302-Controller program change logs | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A controller's program should be only changed during scheduled maintenance periods or under other pre-defined conditions, for which any changes under other circumstances should be considered to have a risk potential. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate a change in a controller's program outside the maintenance period. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary changes a controller's program to change operational rules and disrupt industrial operations. For example, changing the doses of a component on a chemical process. |
| **Related Techniques** | T0831 |

| IoR303-Controller settings change logs | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Some basic controllers usually have already programmed functions and the control rules are assigned by modifying settings. These settings should only be only changed during scheduled maintenance periods or under authorised conditions. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate an unauthorised change in a controller's settings. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary could change the settings of a motor drive, which is used to regulate the speed of an induction motor.<br>An adversary can change the set point temperature on a direct digital controller. |
| **Related Techniques** | T0800, T0805, T0831, T0833 |

| IoR304-Controller/device in firmware update mode | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A device should only be in firmware update mode during scheduled maintenance periods or under other pre-defined conditions, for which any changes under other circumstances should be considered to have a risk potential. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that the mode of a device was changed. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious procedure turns a device into firmware update mode to make it unavailable (e.g. Industroyer)<br>A malicious procedure turns a device into firmware update mode to upload a specially crafted firmware to compromise the device. |
| **Related Techniques** | T0800, T0813 |

| IoR305-Controller in stop mode | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A controller should only be in stop mode during scheduled maintenance periods or under other pre-defined conditions, for which any changes under other circumstances should be considered to have a risk potential. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that the mode of a device was changed. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious procedure turns a device into stop mode to make it unavailable<br>A malicious procedure turns a device into stop mode to upload a specially crafted controller program to compromise the device. |
| **Related Techniques** | T0809, T0813, T0833, T0875 |

| IoR306-Suspicious change in memory usage (field device) | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | |
| **Observations**<br>*Examples of how this IoR can be observed* | Field device process monitoring indicates change in memory usage. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary downloads a malicious program in a controller which consumes more memory than the legitimate program.<br>An adversarial procedure is sending messages or requests to a field device causing it to have an abnormal memory consumption. |
| **Related Techniques** | T0809, T0824, T0877 |

| IoR307-Process state information unavailable | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If it is not possible to retrieve information from the process state such as progress of processes, measurements from sensors, status from actuators, modes from control devices, among others, there is already an issue plus a risk that this situation persists longer or escalates. |
| **Observations**<br>*Examples of how this IoR can be observed* | Process monitoring application sends an error message or reports inability to obtain and display process state information.<br>Users report that they are unable to access information about the process state. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary performs a DoS attack in a field device, which does not allow status updates to be sent. |
| **Related Techniques** | T0813, T0814, T0815, T0826, T0829 |

| IoR308-Firmware update | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Firmware updates are usually done during scheduled maintenance windows. Some devices do not check that the firmware that is installed is signed, for which it is possible to install malicious firmware. Also, newer firmware version could not be compatible with other devices connected in the same environment causing issues in the performance of the system processes. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicate that the firmware was updated. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary installs a firmware version in a controller that does not work well with other devices connected to it and causes a DoS condition. |
| **Related Techniques** | T0814, T0816 |

| IoR309-Frequency Increase of Trouble Calls from a Machine | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a machine sends more troubleshooting alerts than usual it could mean that something in the machine or its environment might not be working under the expected conditions, which could be caused by adversary actions |
| **Observations**<br>*Examples of how this IoR can be observed* | Alerts are trouble calls from a machine are registered in a higher frequency than expected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses a malicious procedure to send control signals for a machine to work outside its operational boundaries (for example, it makes a motor spin faster than allowed). This action is disguised by spoofing the I/O image, however, the machine's has an independent alert system, which is triggered by this anomaly. |
| **Related Techniques** | T0833, T0835, T0875, T0878, T0880 |

| IoR310-Machine Shuts Down During Normal Operations | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a machine shuts down during normal operations it could mean that something in the machine or its environment might not be working under the expected conditions, activating an automatic shut-down (e.g. triggered by a mechanical safety system). It also could mean that an adversary is shutting it down as part of their attack strategy. |
| **Observations**<br>*Examples of how this IoR can be observed* | A SCADA system or other monitoring application indicates a machine shut down.<br>An alarm is triggered due to a machine shut down.<br>Operators notice a machine shut down and log the incident. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses a malicious procedure to interrupt normal operations by shutting down a machine during normal operations. |
| **Related Techniques** | T0800, T0833, T0875, T0878, T0880 |

| IoR311- Abnormal Process Variable Data Is Transmitted to the PLC | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a PLC receives abnormal data from process variables it could be part of a procedure to manipulate the PLC behaviour such as generate a random because the variable value is not recognised or expected. |
| **Observations**<br>*Examples of how this IoR can be observed* | Process variable data do not match the expected or whitelisted values. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary takes advantage of a known vulnerability of a PLC in which a DoS condition is generated by introducing random inputs. |
| **Related Techniques** | T0801, T0813, T0814, T0818, T0827, T0831, T0833, T0868, T0870, T0875 |

OT Network threat

| IoR401-Unknown device connected to the OT network | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | There should be a proper inventory of authorised devices that can connect to the OT network and the detection of an unauthorised or unknown device should be consider an element of risk. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network logs indicate that an unidentified device was connected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious insider connects a hardware device to the OT network to interfere with industrial operations. |
| **Related Techniques** | T0801, T0805, T0830, T0860, T0882 |

| IoR402-High volume of network traffic | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Communication traffic in the OT network should follow a stable pattern of behaviour. An unusually high level of network traffic can reveal malicious activities such as execution of malicious commands, port probes requests, or a DoS attempt. |
| **Observations**<br>*Examples of how this IoR can be observed* | OT network monitoring tools detect a higher volume of data been shared in the ICS industrial protocol network. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary can be sending random traffic to perform a DoS attack in the OT network.<br>An adversary can be making a high amount of requests to collect information of the field devices or sending commands to them. |
| **Related Techniques** | T0813, T0814, T0815, T0869, T0883, T0884, T0882, T0885 |

| IoR403-Commands and responses do not match | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a command does not get a successful or matching response, this could be sign of commands being blocked, intercepted or tampered with. Examples of this are unexpected or out of time responses. |
| **Observations**<br>*Examples of how this IoR can be observed* | A process sends an error message or reports a malformed or unusual response.<br>A user reports a malformed or unusual response. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is performing a Man in the Middle attack to modify commands.<br>Command messages have been blocked and the system is giving random responses. |
| **Related Techniques** | T0803, T0805, T0813, T0827, T0829, T0830, T0831 |

| IoR404-Unresponded connection requests (port probes) | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | An adversary might send incomplete connection requests for reconnaissance purposes, for example port probes to identified ports used and devices connected to the ICS network. |
| **Observations**<br>*Examples of how this IoR can be observed* | OT network monitoring tools detects unresponded connection requests. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is doing port probes techniques to collect information about the ports that are open in the OT network. |
| **Related Techniques** | T0806, T0808, T0883 |

| IoR405-Unresponded commands | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a command does not get a successful response, this could be sign of a DoS or that commands being blocked, intercepted or tampered with. |
| **Observations**<br>*Examples of how this IoR can be observed* | A process sends an error message due to unresponded commands.<br>A user reports unresponded commands. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is performing a DoS attack in the OT network.<br>Backchannel network traffic is been rerouted to an adversary by spoofing the IP address as part of a Man in the Middle attack. |
| **Related Techniques** | T0800, T0803, T0805, T0813, T0814, T0816, T0826, T0827, T0830, T0831 |

| IoR406-Unexpected command sequence over network | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Communication traffic in the OT network should follow a stable pattern of behaviour in which a baseline profile can be built of what commands are expected to be used and at what frequency, for which deviations from this baseline can indicate a risk. |
| **Observations**<br>*Examples of how this IoR can be observed* | A command that is expected to be send once a day has been detected to be sent frequently within short period of time.<br>A command that is sent only within exceptional conditions is detected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary attempts to manipulate the operation workflows by injecting commands that generate disturbances to the process (e.g. starting or stopping a machine at undue time, adding more quantities of material, opening a valve and causing a flood, etc.) |
| **Related Techniques** | T0800, T0801, T0803, T0804, T0805, T0806, T0808, T0814, T0816, T0824, T0830, T0831, T0868, T0869, T0870, T0883, T0884, T0885 |

| IoR407-Communication port blocked | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a communication port that is expected to be used to established a connection appears to be blocked or used by an unknown process this can indicate a risk. |
| **Observations**<br>*Examples of how this IoR can be observed* | OT network security monitoring or port scan tools indicate that a port in a device using OT protocols is blocked.<br>Logs indicate that a port in a device using OT protocols has been blocked. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An attacker tries to reuse ports, which are already allowed through firewalls by sending malicious payload from or to it. |
| **Related Techniques** | T0803, T0805, T0815, T0829 |

| IoR408-Unusual connection between devices | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | In OT networks it should be possible to establish normal communication profile, such as which devices normally communicate with another, protocols they use and the type and amount of traffic. Deviations from this pattern are anomalies that could reveal risk scenario. |
| **Observations**<br>*Examples of how this IoR can be observed* | OT network monitoring tools detect communication between two devices which are not meant to communicate.<br>OT network monitoring tools detect communication between two devices with a higher frequency or sharing a higher amount of data than expected. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary is using one device to compromise others in the same network (Lateral Movement). |
| **Related Techniques** | T0805, T0830, T0869, T0883, T0884, T0882, T0885 |

| IoR409-Delay or timeout between connections | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A delay in communications can mean that an adversary can be interfering with the communication process and that messages could have been tampered with. |
| **Observations**<br>*Examples of how this IoR can be observed* | Time stamps reveal an unusual latency between the emission of a command and the reception of a response. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary performs an active Man in the Middle attack and tries to modify messages.<br>An adversary attempts to jam wireless communication to slow down the traffic |
| **Related Techniques** | T0830 |

| IoR410-Maximum of connections exceeded | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Connections in many industrial network tend to be stable since they followed established communication patterns based on the requirements of the underlying processes. Hence, it many cases it might be possible to establish a maximum number of connections between devices in a network. |
| **Observations**<br>*Examples of how this IoR can be observed* | Network monitoring tool detects that the number of simultaneous connections exceeds the maximum limit. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary tries to use a server or workstation as a proxy to jump to other areas of the system. |
| **Related Techniques** | T0802,T0806, T0814, T0884, T0882, T0885 |

| IoR411-Use of unusual communication protocol | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A malicious procedure might attempt to establish communication between devices using a communication protocol that is supported by the industrial network but not commonly used in a particular context. |
| **Observations**<br>*Examples of how this IoR can be observed* | Communication is attempted using a protocol that is not whitelisted. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Malware might have been programed to attempt establishing communication via a protocol that is commonly used in industrial network (e.g. Profibus). However, if an ICS that does not use this particular protocol, this communication attempt would be consider an anomaly. |
| **Related Techniques** | T0802, T0814, T0822, T0825, T0868, T0885 |

| IoR412-Communication through unused ports | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A malicious procedure might attempt to establish communication through ports that are not currently been used in devices from the ICS. |
| **Observations**<br>*Examples of how this IoR can be observed* | Communication is attempted from or to a port that is not whitelisted. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary my do probes by attempting connections to different port addresses to identify which ports are been used. |
| **Related Techniques** | T0802, T0806, T0818, T0822, T0868, T0882 |

| IoR413-File transfers between devices | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | A malicious procedure might use devices to infect other devices with malware. Also devices can be used to collect data from other devices. |
| **Observations**<br>*Examples of how this IoR can be observed* | Examples of observations of file transfers between devices:<br>Data Exfiltration Between ICS Devices via Server Message Block<br>Data Exfiltration Between ICS Devices via User Datagram Protocol |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary tries to use a device to send malicious payload to other devices.<br>An adversary uses as a device as a proxy to steal operational data. |
| **Related Techniques** | T0802, T0882 |

| IoR414-Abnormal OT communication | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Abnormal patterns of OT communications such as increase on frequency of connections or malformed traffic, among others, can be a sign of malicious activities. |
| **Observations**<br>*Examples of how this IoR can be observed* | Examples of observations of abnormal OT communication:<br>Loss of Communications with Modbus TCP Device<br>ICS Device Receives Diagnostic Modbus TCP Function Codes<br>ICS Device Receives Undefined Modbus TCP Function Codes<br>ICS Device Receives Malformed Modbus TCP Traffic |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary can send random communication messages to cause a Denial of Service condition. |
| **Related Techniques** | T0802, T0806, T0801, T0813, T0814, T0818, T0831, T0833, T0868 |

I/O data threat

| IoR501-Sensor data out of limits | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a physical variable that is critical for the process is out of limits, and there is not known cause, there could be a malfunction in the system whose cause might or might not be security related, but it any case requires attention. |
| **Observations**<br>*Examples of how this IoR can be observed* | I/O data indicates that a sensor provided an input measurement which is out of the control limits.<br>I/O data indicates that a sensor provided an input measurement which was out of expected limits. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary changes control parameters to make physical variables to exceed their operational limit (e.g. Stuxnet). |
| **Related Techniques** | T0804, T0831, T0833, T0878 |

| IoR502-Misbehaviour in sensor data | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a physical variable that is critical for the process presents trends or variations that are not common, even when within the control limits, and there is not known cause, this behaviour can be revealing an anomaly which might constitute a risk. |
| **Observations**<br>*Examples of how this IoR can be observed* | I/O data indicates that the behaviour of an input measurement provided by a sensor has unexpected trends, shows abrupt changes, oscillates, or present another unusual behaviour. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might alter the behaviour of a control system causing some variables to be outside the operational limits. |
| **Related Techniques** | T0804, T0831, T0832, T0833, T0878 |

| IoR503-Indicator that can be correlated to a critical input out of limits | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a physical variable that is not critical for the process is out of limits, and there is not known cause, this could indicate a malfunction in the system and that data of critical inputs could have been tampered with to give an fake sense of normality. |
| **Observations**<br>*Examples of how this IoR can be observed* | Data that can be correlated with a critical input is out of the expected limits. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might alter the behaviour of a control system causing critical variables to be outside the operational limits. However I/O images of critical variables might be spoofed to show normal values (e.g. Stuxnet). In this case, other correlated variables could reveal the anomaly (e.g., a higher temperature can reduce humidity levels) |
| **Related Techniques** | T0804, T0831, T0832, T0833, T0835, T0878 |

| IoR504-Misbehaviour in data that can be correlated to a critical input misbehaviour | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a physical variable that is not critical for the process presents trends or variations that are not common, without known cause, there could be a malfunction in the system and that data of critical inputs could have been tampered with to give an fake sense of normality. |
| **Observations**<br>*Examples of how this IoR can be observed* | Data that can be correlated with a critical input has unexpected trends, shows abrupt changes, oscillates, or presents another unusual behaviour. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might alter the behaviour of a control system causing misbehaviour on critical variables. However I/O images of critical variables might be spoofed to show normal values (e.g. Stuxnet). In this case, other correlated variables could reveal the anomaly (e.g., a higher temperature can reduce humidity levels) |
| **Related Techniques** | T0804, T0831, T0833, T0835, T0878 |

| IoR505-Actuator data out of limits | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If an actuator is operating out of limits, there could be a malfunction in the system whose cause might or might not be security related, but it any case requires attention. |
| **Observations**<br>*Examples of how this IoR can be observed* | I/O data indicates that output data sent to an actuator or the operation parameters of an actuator present operating parameters which are out of limits. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary manipulates control variables for an actuator to operate outside of the permitted limits (e.g. a motor operating at a faster speed). |
| **Related Techniques** | T0804, T0831, T0833, T0878 |

| IoR506-Misbehaviour in actuator data | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If an actuator presents a behaviour that is not the expected, even if this is not by itself a forbidden or out of bounds state, it could by a symptom of other malfunctions of the system. |
| **Observations**<br>*Examples of how this IoR can be observed* | I/O data indicates that the behaviour of output data or the operation parameters of an actuator present unexpected trends, show abrupt changes, oscillate, or present another unusual behaviour. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary manipulates control variables for an actuator to behave randomly (e.g. a motor changing its speed continuously, a valve been opened and closed randomly). |
| **Related Techniques** | T0804, T0831, T0832, T0833, T0878 |

| IoR507-Data that can be correlated to a critical output out of limits | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a physical variable that is not critical for the process is out of limits, and there is not known cause, this could indicate a malfunction in the system and that data of critical inputs could have been tampered with to give a fake sense of normality. |
| **Observations**<br>*Examples of how this IoR can be observed* | Data that can be correlated with a critical output is out of the expected limits. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might alter the behaviour of a control system causing misbehaviour on critical variables. However, I/O images of critical variables might be spoofed to show normal values (e.g. Stuxnet). In this case, other correlated variables could reveal the anomaly (e.g., an increased rotation speed in a motor increases also the electrical power consumption) |
| **Related Techniques** | T0804, T0831, T0832, T0833, T0835, T0878 |

| IoR508-Misbehaviour in data that can be correlated to a critical output | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a physical variable that is not critical for the process presents trends or variations that are not common, without known cause, there could be a malfunction in the system and that data of critical inputs could have been tampered with to give an fake sense of normality. |
| **Observations**<br>*Examples of how this IoR can be observed* | Data that can be correlated with a critical output has unexpected trends, shows abrupt changes, oscillates, or presents another unusual behaviour. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary might alter the behaviour of a control system causing misbehaviour on critical variables. However I/O images of critical variables might be spoofed to show normal values (e.g. Stuxnet). In this case, other correlated variables could reveal the anomaly (e.g., an increased rotation speed in a motor increases also the electrical power consumption) |
| **Related Techniques** | T0804, T0831, T0832, T0833, T0835, T0878 |

| IoR509-Sensor data unavailable | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If it is not possible to retrieve information from sensors it means that there is already an issue, and the risk that the causes are security related and that the situation persists longer or escalates should be considered. |
| **Observations**<br>*Examples of how this IoR can be observed* | I/O image is not available<br>Data from a sensor is not captured by an I/O image<br>Data from a sensor is not available in a Control and Monitoring application. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Adversary floods the network through a DDoS not allowing to retrieve operational data.<br>Sensor data is redirected to a different destination by an adversary. |
| **Related Techniques** | T0804, T0814, T0815, T0826, T0829 |

| IoR510-Actuator data unavailable | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If it is not possible to retrieve information from actuators it means that there is already an issue, and the risk that the causes are security related and that the situation persists longer or escalates should be considered. |
| **Observations**<br>*Examples of how this IoR can be observed* | I/O image is not available<br>Data from an actuator is not captured by an I/O image<br>Data from an actuator is not available in a Control and Monitoring application. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | Adversary floods the network through a DDoS not allowing to retrieve operational data.<br>Actuator data is redirected to a different destination by an adversary. |
| **Related Techniques** | T0804, T0814, T0815, T0826, T0829 |

| IoR511-Inconsistency between different sources of data | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If I/O data is inconsistent between different sources or between redundant measurements, or at different levels of the system it can mean that I/O data have been tampered with. |
| **Observations**<br>*Examples of how this IoR can be observed* | The I/O value given by a redundant sensor differs significantly from the main sensor.<br>I/O value displayed in an HMI close to the physical operations is different to the one displayed in SCADA application in the engineering workstation. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary spoofs the I/O image in the SCADA application running in the workstation. |
| **Related Techniques** | T0832, T0835, T0878 |

Physical security threat

| IoR601-Suspicious Physical Access log | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Access control systems are usually independent from the main ICS. If it is possible to identify unusual patterns of behaviour in the physical access critical ICS areas, this could constitute an indicator of risk of an insider's threat. |
| **Observations**<br>*Examples of how this IoR can be observed* | Access logs at an unusual days or times.<br>Access logs or signals of occupancy on restricted areas by unnauthorised personnel. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious insider with access privileges enters the premises outside their shift or outside working hours.<br>An adversary steals an ID card and uses it to get into the perimeter. |
| **Related Techniques** | T0818, T0847 |

| IoR602-Physical Intrusion alert | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Access control systems are usually independent from the main ICS. An intrusion alert could constitute a risk of systems and devices been compromised. |
| **Observations**<br>*Examples of how this IoR can be observed* | Alert in access control and perimeter security system. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary attempts to break perimeter security.<br>An adversary attempts to get close proximity to intercept wireless devices.<br>An adversary gains access to the perimeter by tailgating during a busy period (e.g. time workers are accessing the plant) and this is spotted by the security cameras. |
| **Related Techniques** | T0818, T0847 |

| IoR603-Misbehaviour or malfunction in Access Control System | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If an Access Control System presents and abnormal behaviour such as an abrupt re-start of the system or it is unresponsive, this can present a risk of malicious actors trying to gain physical access. |
| **Observations**<br>*Examples of how this IoR can be observed* | Access control system network issues.<br>Access control system malfunction.<br>Access control system restart. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses a vulnerability in an access control system to create a DoS attack and create a distraction to break into the perimeter. |
| **Related Techniques** | T0818, T0847 |

| IoR604-Signal from security cameras is interrupted or disabled. | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If security cameras are not working properly, surveillance activities would be less effective allowing an adversary to perform malicious actions unnoticed. |
| **Observations**<br>*Examples of how this IoR can be observed* | CCTV footage becomes unavailable.<br>CCTV footage presents interruptions or performs badly. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary uses a vulnerability in security cameras to create a DoS attack and interrupt the surveillance system to create a distraction. |
| **Related Techniques** | T0818, T0847 |

Safety threat

| IoR701-Alarm Historian anomaly | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If there is a deviation from the alarms usual behaviour, there is a risk of an unexpected or undesired behaviour of the system or alarm parameters or thresholds could have been changed. |
| **Observations**<br>*Examples of how this IoR can be observed* | An alarm is triggered without the system showing any alarm triggering condition.<br>An alarm is not triggered while an alarm triggering condition is observed. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary modifies alarm settings or disables an alarm.<br>An adversary triggers false alerts to create a distraction or to force operators to disable the alarm due to the increase of false positives. |
| **Related Techniques** | T0878, T0880 |

| IoR702-Anomaly in Protection Relay | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | If a protection relay does not react when expected conditions there can be a risk that trigger conditions are been forced or inhibited by a malicious procedure. |
| **Observations**<br>*Examples of how this IoR can be observed* | Sensors detect an electrical current rise that should have triggered the protection relays but they are not triggered. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | A malicious insider might create a short-cut to bypass a protection relay.<br>A malicious insider might disable a smart relay or change the triggering conditions. |
| **Related Techniques** | T0878, T0880 |

| IoR703-Change logs in Safety Instrumented System | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Safety Instrumented systems are usually independent from the main ICS. Any unauthorised changes or changes outside of maintenance windows should be considered as a risk. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs from a safety instrumented system indicate that safety settings have been changed. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary physically manipulates safety protections so they are not triggered in case of safety threats. |
| **Related Techniques** | T0878, T0880 |

| IoR704-Anomaly in Safety Instrumented System | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Safety instrumented systems are usually independent from the main ICS so they could provide redundant data which, in case is not consistent with the ICS data can reveal a risk. |
| **Observations**<br>*Examples of how this IoR can be observed* | A physical variable (e.g. temperature, valve pressure) has a different value in the Safety Instrumented than the one provided by the main control an automation system. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary spoofs the I/O image in the SCADA application running in the workstation. However, the safety instrumented system presents the legitimate values. |
| **Related Techniques** | T0878, T0880 |

| IoR705-Change in alarm thresholds | |
|---|---|
| **Rationale**<br>*Why this is consider an IoR?* | Changing alarms thresholds can allow the system to reach a forbidden state without alerts been triggered on time. Any changes in alarms thresholds should be authorised, otherwise the risk of something malicious going on increases. |
| **Observations**<br>*Examples of how this IoR can be observed* | Logs indicated that a threshold in an alarm settings was changed. |
| **Examples**<br>*Scenarios in which the IoR might be observed* | An adversary changes alarm thresholds to avoid the alarm to be triggered when a malicious procedure is performed. |
| **Related Techniques** | T0878, T0880 |