Linköping University | Department of computer science
Master thesis, 30 ECTS | Information Technology
Spring 2017 | LIU-IDA/LITH-EX-A--17/051—SE

Mapping out dependencies in network components in critical infrastructure

Karl Andersson

Supervisor: Klervie Toczé & Jonas Olsson
Examinator: Simin Nadjm-Tehrani

LINKÖPINGS
UNIVERSITET

Abstract

Companies that operate with critical infrastructure face a growing threat from cyber-attacks while at the same time the development in the business is rapidly moving towards a higher level of digitalization. A common type of system in critical infrastructure is supervisory control and data acquisition systems, these systems have properties that can affect their security and will therefore serve as the basis for this thesis work. To stay protected despite systems changes, companies need to make risk assessments in order to analyze how changes will affect the overall system. One thing that is important to focus on is dependencies within the system, this means that not only interaction among computers and networks are concerned but instead a more holistic view of the system need to be considered.

This thesis aims to aid the process of a future risk assessment by providing a methodology to be used as a preparatory step before a risk assessment by describing the current situation of the system. This is done by evaluating two system modeling approaches, and also by proposing a number of perspectives that each provides different kind of information about the system's dependencies. These perspectives are then evaluated by creating system models and dependency graphs, and discussing the outcomes with experts in a utility company to find out their applicability.

According to the experts, the proposed perspectives have promising properties that can be useful in future risk assessments as well as in other scenarios. Moreover, the evaluated modeling approaches got positive comments during evaluation and are considered to serve their purpose.

Acknowledgement

First of all, I would like to thank my supervisor at Tekniska verken Jonas Olsson for his valuable input and for all the encouragement during the work. Further, I would also like to thank my examiner and supervisor at Linköping University, Simin Nadjm-Tehrani and Klervie Toczé for giving great support and guidance during the work. Finally, I would like to thank everyone at Tekniska verken that in some way contributed to the making of this thesis.

Karl Andersson
Linköping, October 2017

Table of Contents

List of Figures

1 Introduction

Tekniska verken is a company that among other things is responsible for the delivery of electricity, water, broadband and direct heating i.e. critical infrastructure. Companies in this line of business have in recent years become more concerned with their Information Technology (IT) security, this is due to the increased risks of being the target of a cyber-attack. Critical infrastructure needs to be well protected since a successful attack could have great impact on today's society e.g. affect the delivery of electricity or water.

The fact that a successful attack on a critical infrastructure could have huge impact on society can attract attackers with great resources (money, manpower etc.) e.g. terrorists, adversary nations or those who seek economic gain. To be able to mitigate vulnerabilities you need to have a good idea of how a network is structured and what connections there are in and out of the network in order to be able to locate vulnerable parts. One important aspect is the ability to share this kind of information with people without a deeper technical knowledge. This is because the person who is in charge of funding the IT-infrastructure does not necessarily have technical knowledge. If company management has the wrong perception or lack of knowledge about their IT security they cannot take the necessary precautions [1].

At Tekniska verken, as for all companies in this line of business, there are always situations where actions that also affects the IT-security will be performed. To be able to make good decisions there is a need to have a solid understanding of how a decision will impact the overall situation. To get a solid understanding you first need to have a good overview of the network i.e. what devices exist, what dependencies exist and how devices communicate with each other. One aspect that has gotten more attention lately is dependencies within a network. Since the definition of dependencies can vary, this can cause problems when information about a system's dependencies is shared. The result of an unclear definition may lead to ambiguous information which in turn can cause decisions regarding a company's IT security to be taken based on inaccurate information. Dependencies may not always be obvious but at the same time if a subsystem stops functioning, it can impact the network and its ability to function as normal. There can be different kinds of dependencies within a system which also can make it hard to know how to find them and how to describe them.

According to the National Institute of Standards and Technology (NIST) specification 800-30 [2] regarding risk assessments, the first component of risk management is to establish a risk context i.e. describe the environment and the systems concerned by the assessment. This is the foundation for the future risk assessment and therefore needs to be solid and holistic. There are some common ways to map out networks but they most often also rely on technical tools such as scanning to map out network devices. Such techniques may lack the ability to capture a system with a broad enough perspective [3]. With a too narrow perspective, important aspects and dependencies within a system might be overlooked. For example, a system may appear well secured with good firewalls, but what if an attacker is given physical access to a computer inside the system? Or if someone with access to a system just remembers information and then speaks with a potential attacker? These are only two examples of when there is a need for a methodology with a broader scope. Figure 1 illustrates three different kinds of attacks, (1) a cyber-attack, (2) target people working with the targeted system or (3) to physically break in into the system. This shows the importance of knowing every aspect of your system in order to make correct decisions.

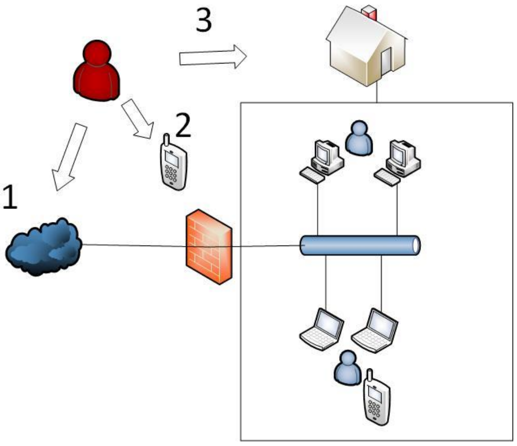


Figure 1 Illustration of different attack ways

This thesis will therefore concern the analysis of system dependencies and what methodology is needed to have a wider scope that can capture more than just a system's physical properties. The importance of analyzing dependencies is motivated by the need to know what happens if there is failure of a node in the system and what then happens to the rest of the system. All nodes that are dependent on a failed node will be affected in some way, how they are affected depends on e.g. the coupling, the kind of dependency and the state of operation.

A common type of network in this business is the so called Supervisory Control and Data Acquisition (SCADA) network which is a set of connected devices designed to monitor and control cyber-physical systems. SCADA networks were introduced long before cyber-attacks were a real threat and therefore they do not have the appropriate security mechanisms that are needed in today's society. Further, a SCADA network is typically in use for a long time and during its lifetime it can be changed or extended with new devices that can increase the complexity of the network. Today there exists a great number of threats against a SCADA network and it can be hard to know all the different kinds of possible attacks and how to mitigate them. A typical SCADA network has a specific architecture and is built on a number of core devices that each has its own unique role in the network. When SCADA networks were introduced they were often separated from other networks and most often isolated from any outside communication, but nowadays they are instead most often connected to the corporate network and also to the Internet and thereby SCADA networks face the same threats as any other network, which is problematic due to the lack of appropriate security mechanisms [4]. Due to these circumstances, a SCADA network will be used as subject of the analysis.

1.1 Motivation and problem statement

To be able to make well founded risk assessments and to take well-grounded decisions there is a need to know the system's structure, what kind of dependencies exists among the system's components and also how people interact with the system.

To capture a system with a wider scope, different kinds of approaches can be used. One approach is to consider that a system consists of different layers where each layer consists of nodes and connections. To view a system as a set of layers can be done by using an eight-layer model where each layer describes different properties of a system [5]. This model can also be simplified into a three-layer model to make the model more applicable [6]. The relation between the models can be viewed in Figure 2 which illustrates how the layers of the different models relate to each other. Since the layers of the three-layer model are more general but still able to capture a system's characteristics, this model will be used in this thesis.

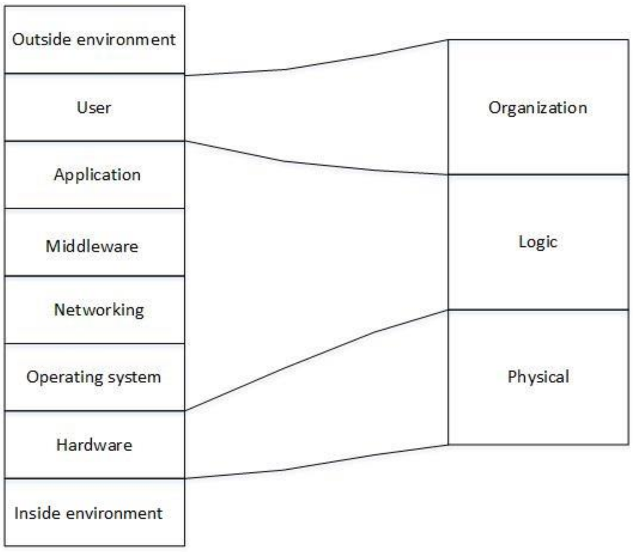


Figure 2 Relationship between eight-layer [5] and three-layer [6] model.

By using a layered model you have the ability to model a system with a high level of abstraction combined with a wide range of different type of nodes. By using the layered model, a top down approach is used when mapping out a system.

The Predictive Probabilistic Cyber Security Modeling Language (P^2CySeMoL) is a probabilistic relational model initially developed to model SCADA systems [7] . SecuriCad[1] is a commercial tool that is grown out of an earlier research prototype of P^2CySeMoL and has extended that language. SecuriCad is a tool primarily for performing vulnerability analysis of a system. A feature in SecuriCad is the possibility of performing simulation of the systems security and provide information about Time To Compromise (TTC), which can be valuable for future risk assessments. The fact that P^2CySeMoL was originally focused on critical infrastructure and that the language then has been extended in SecuriCad can make it a useful tool in this context. Therefore a part of this thesis will evaluate how well SecuriCad can provide system owners with more information about their systems. However, the focus of SecuriCad will be to use its functionality for system mapping. In contrast to the layered model, the system mapping in SecuriCad uses a bottom-up approach.

[1] https://www.foreseeti.com/products

As previously stated, SCADA networks can be complex and have many different dependencies of different kinds. Dependencies are important to keep track of since they can have great impact on a system's performance and as part of a risk assessment there is a need to know what happens to connected nodes when other nodes fail. The aim of this thesis is to investigate what is needed by a methodology to capture a system with a wide scope and also capture the system's dependencies. The goal is that the methodology can be used as preparatory step to a risk analysis. This can be concretized into the following three research questions:

1. Can the top-down approach of the three-layered model be combined with the bottom-up approach of SecuriCad?
2. What suitable methods exist for dependency analysis and how can they be combined with different system models?
3. Is there any evidence that knowledge about system dependencies are helpful for future risk assessments?

To answer the first question each of the models will be implemented on the same system and then evaluated to determine how well they can map out a system. To answer the second question a review of literature on system dependencies and system modeling will be performed. Based on the studied literature a number of methods for dependency analysis will be selected and combined with different system models. To answer the third question, a method for providing information about system dependencies will be put together. The method will then be implemented and evaluated based on how helpful it can be to a future risk assessment.
By answering these questions, a methodology will be developed that is intended to be used as preparatory step to a risk assessment. This methodology will henceforth be referred to as, the mapping method.

1.2 Goals

By investigating the earlier stated problems, the work in this thesis aims to make the following contributions:

- An evaluation of the three-layer model and SecuriCad based on how well they can model a SCADA network.
- A number of perspectives that can be used to provide information about a system's dependencies.
- A structure that can be used when sharing information about a system's current security posture.

The underlying goal of the work is to provide different views to describe a system, where the result does not necessarily have to provide any new information but aims to make it easier to understand the system's properties and can therefore be used when sharing this kind of information among people with different levels of knowledge. This means that the work should not be seen as an analysis in itself but should be seen as a way to make room for a following analysis.

1.3 Limitations

This thesis only focuses on the part of modeling systems and then mapping their dependencies. Any further analysis like vulnerability analysis, mitigations etc. will not be considered, but left for future work.

A system can be analyzed and modeled from a number of perspectives that can provide different kind of information and thereby aid a future risk assessment. However, the work in this thesis will focus on the three-layer model, SecuriCad, dependency analysis, and their relation to risk assessments. This means that other perspectives will be left for future work.

The work will be carried out at the company Tekniska verken which means that that some parts may not be widely generalizable, although the aim is to keep the work as general as possible. Also, no specific information about Tekniska verken can be included due to confidentiality reasons.

There is also a time limit that needs to be kept in mind, the time frame is about twenty weeks and the extent of the work will be done to fit within this time frame.

1.4 Thesis outline

The remainder of this thesis is structured as follows:

- Chapter 2 will present theory about the main topics, SCADA networks, system dependencies and system modeling. This will provide a basic understanding of the researched area and the challenges that exist and provide a foundation for the rest of the thesis.
- In Chapter 3 the used research methodology will be described. Further, different perspectives of dependency analysis will be presented and what kind of information they can provide. The chapter also includes a description of the test system that will be used in the thesis.
- Chapter 4 will first show how the three-layer model and SecuriCad can be used on the test system. The next part focuses on dependencies and how to analyze dependencies based on the test system. The results will consist of images describing the system along with an explanation of the presented data.
- Chapter 5 will include a discussion, the conclusions and possible ways of future work.

1.5 Ethical considerations

There are some important ethical aspects that need to be considered in this thesis work. First and foremost, nothing regarding how Tekniska verken works and operates will be included in the report. If specific information about Tekniska verken is published, it can affect their operations and harm the integrity of specific information. Also, everyone giving their input about the methodology will be kept anonymous.

2 Dependencies and SCADA systems

This chapter will provide some background knowledge of the researched area and lay a foundation for the remainder of the thesis.

2.1 Terminology

This section will highlight important terms that are frequently occurring in the thesis and it is therefore good to have a basic understanding of their meaning.

2.1.1 Dependency terminology

The term *dependency* can have a somewhat broad meaning but the work in this thesis will use the following definition: a dependency is a connection between two components where the state of one component influences the state of a connected component [8]. This essentially means that if there is a connection of any kind between two nodes, there is also a dependency between them. This definition means that in a network topology any kind of connection will be modeled as a dependency since a connection between two nodes means that they can affect each other's behavior is some way.

The term *interdependency* can have slightly different definitions in research, the definition in this thesis will be the one proposed by Rinaldi et al. [8] which is that an interdependency is a two-way dependency i.e. if there are two components A and B, then there is an interdependency if A depends on B and B also depends on A.

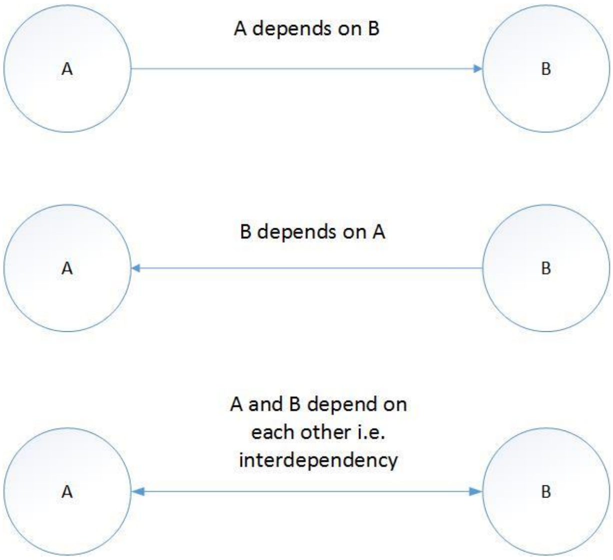


Figure 3 The different kinds of dependencies.

Further it is also important to know that there can also be indirect dependencies of n:th-order e.g. a dependency where n = 2, would be if component A depends on B which in turn depends on component C, then there is a dependency of second order between A and C.

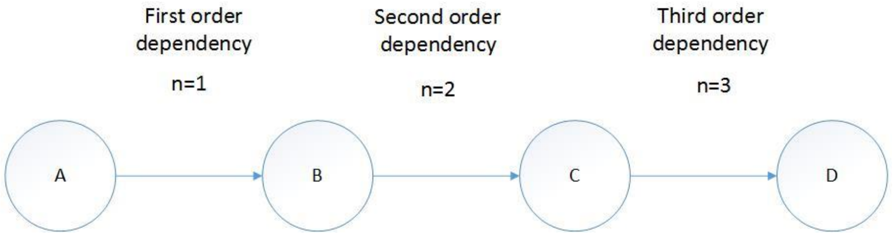


Figure 4 n:th-order dependencies

Dependencies can also be analyzed in different dimensions. To understand these dimensions, there is a model that can be used to get an understanding of dependencies and problems that might follow [8]. The model is built on six dimensions that describe a dependency, shown in Figure 5.

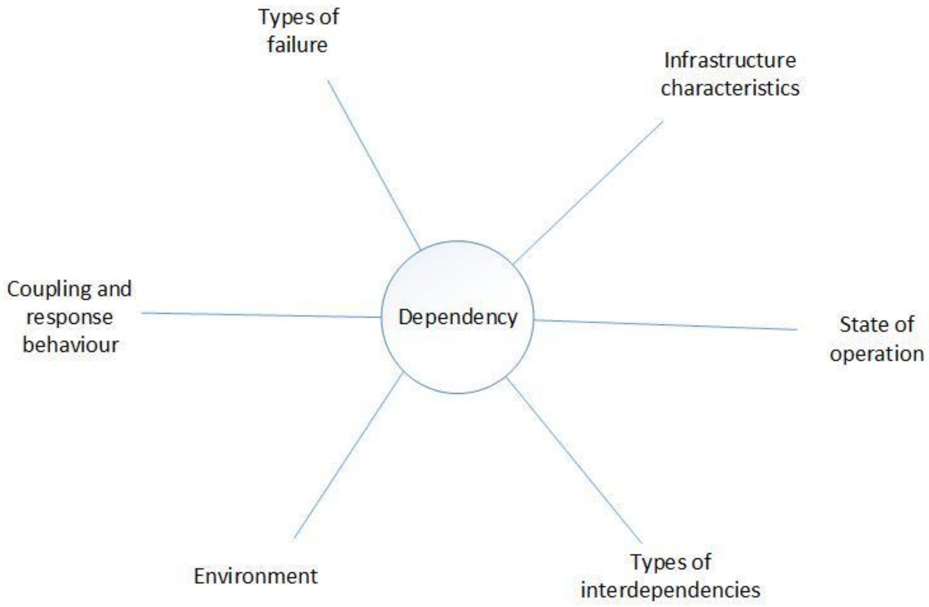

Types of failure
Infrastructure characteristics
Coupling and response behaviour
Dependency
State of operation
Environment
Types of interdependencies

Figure 5 Dependency dimensions [8]

The following as a description of how each dimension can be used. In order to be able to handle failures within a system it is important to know what *types of failures* there can be. In case of failure at one node within a system, the failure can have different characteristics, Rinaldi et al. [8] present the following types of failure:

- Cascading failure
- Escalating failure
- Common cause failure.

A cascading failure is a failure that starts somewhere in the systems which in turn causes a new failure in a different place. An escalating failure is a failure that starts somewhere in the system and then creates new failures of greater magnitude in the system. Finally, a common cause failure is when several failures occur, caused by the same event. As a way to analyze cascading failure's effects on interdependent systems there has been an experiment where each node is assigned an initial load and a limit. The limit describes the maximum load a node can handle. If the maximum load level is passed the node's load is transferred to its neighbors. To start the simulation a randomized disturbance is introduced to random nodes which will start a chain of reactions through the system. If a node fails but has no connections the cascading effect will be stopped. The conclusions are, that there are two parameters that can reveal a lot about the system's vulnerability to cascading failures, the critical load and the average cascade size. By doing a number of simulations reflecting different operational conditions, it is concluded that the system load is crucial to control since after a certain point failures start to cascade in a nonlinear manner [9].

System dependencies can lead to two different kinds of *coupling* among nodes, tight or loose [10]. If the coupling is loose it means that the components are less dependent on each other and if the coupling is tight it means that they are highly dependent on each other. With tight coupling, interference in one component will heavily affect the dependent functions. This does not only apply to direct dependencies but also applies to dependencies of higher order through cascading effects.

The *environment* in which a dependency exists can also play an important role since it highly affects how to handle a potential failure. Examples of different environments can be technical, political and economical [8].

There are different *types of interdependencies*, the type describes the primary characteristics of a dependency. There exists several different set of categories that aims to capture different kinds of dependencies and then classify them according to pre-defined categories. The different set of categories will be presented and discussed later.

The *state of operation* in which an infrastructure is currently located can affect the possibility to handle a failed dependency. If the current state can be considered normal then a failure can be handled according to a pre-defined plan. If the current status is for example stressed, resources might have to be focused on other things than a failed dependency, depending on the severity of the failure.

Different parts of an infrastructure have different *characteristics* affecting its ability to handle dependencies. Characteristics such as organizational, operational and temporal are presented by Rinaldi et al. [8] as important to analyze.

By analyzing these dimensions together, it is possible to have well founded understanding of the analyzed dependencies. In terms of a risk analysis some dimensions are more applicable then others. Further, some dimensions have a wider scope which leads to more possible ways to work with the dimension.

2.1.2 System terminology

In this thesis, the terms *network* and *system* will be frequently used and therefore it is important to define the difference between them. The difference is that a network is based on physical components i.e. computers, cables, switches etc., while a system has a broader definition and includes for example people, the organization and environments, and also includes networks i.e. a network is a part of a system.

There is also a need to define different levels of granularity in a system description. The following terminology is defined by Perrow [10] and among others adapted by Rinaldi et al. [8] and will therefore also be used in this thesis.

- **Part**: smallest component that can be identified in an analysis.
- **Unit**: a functionally related collection of parts.
- **Subsystem**: an array of units.
- **System**: a grouping of subsystems.
- **Infrastructure**: a complete collection of like systems.
- **Interdependent Infrastructures**: the interconnected web of infrastructures and environment.

The term *holistic* can also be a bit abstract, the intended meaning in this thesis is to include the human role in a system analysis to provide a new dimension of information about the systems properties. In most other cases, a system is considered to be just physical hardware and their interconnections.

2.2 SCADA architecture

The SCADA network is a type of network used in critical infrastructure. A SCADA network typically has a set of devices that includes:

- Sensors
- Remote Terminal Unit (RTU)
- Programmable Logic Controller (PLC)
- Master Terminal Unit (MTU)
- A historian
- A control room with a Human Machine Interface (HMI)

The flow of a typical SCADA system is that the sensors gather data that is then collected by RTUs and PLCs and is then sent to the control room where it can be supervised by operators, the information is also stored in the historian where it can be analyzed in the future. In the control room there is also a MTU that operators use to send commands to units in the field. When SCADA networks were introduced they were used in an isolated environment where availability was very important and therefore security was not an important aspect of the networks since it often reduces availability to some degree. This fact makes SCADA networks a so called legacy network, meaning that it has legacy properties that cannot be overlooked and therefore affect its current state. In Figure 6 a general SCADA architecture is shown where the system is segmented by using Demilitarized Zones (DMZ) prior to a Local Area Network (LAN).

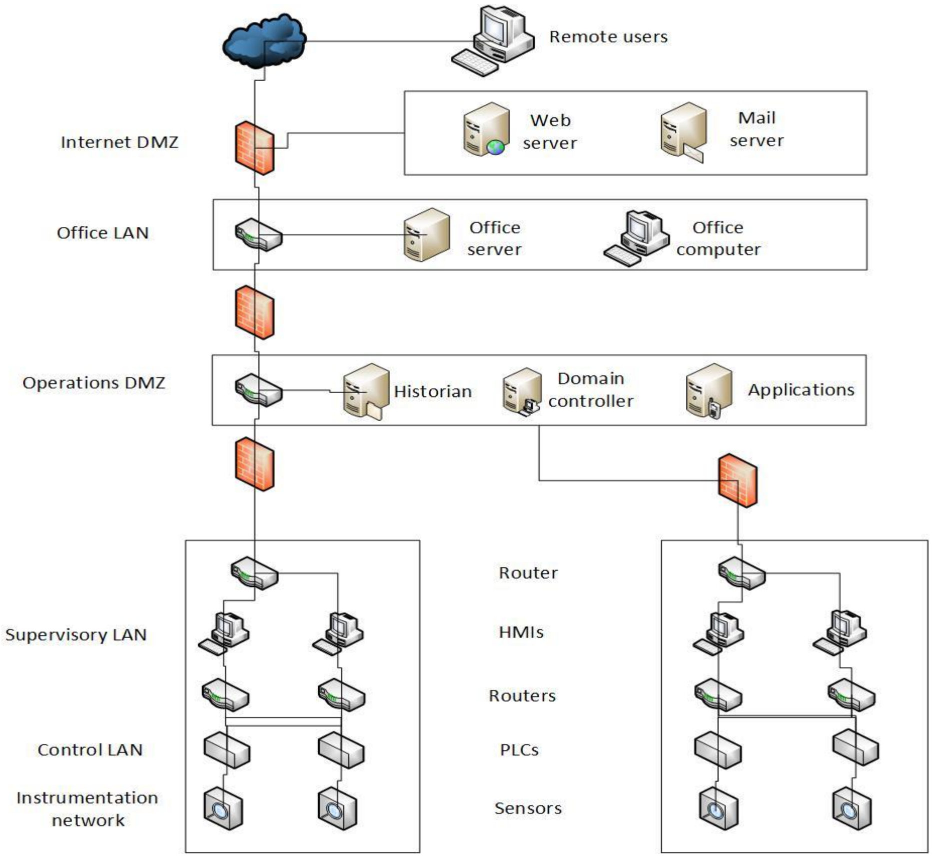


Figure 6 General SCADA and corporate network architecture.

2.2.1 Security issues in SCADA networks

Nowadays the SCADA networks are most often connected to both the Internet and the corporate network so that employees do not have to be on-site to be able to work. This can increase productivity and efficiency but also demands that the network is accessible from remote locations. This means that the environment has drastically changed and security now needs to be a top priority, otherwise the system is an easy target for cyber-attacks.

When a SCADA network gets connected to the Internet and the corporate network there are a number of consequences that needs to be handled. One important difference to keep in mind is that a SCADA network and a corporate network have very different purposes and therefore need different security mechanisms. SCADA networks aims to provide process control and management while the corporate network is more focused on processing, storing and retrieving corporate data. The key aspects of the SCADA network is that it is reliable and controllable. However, these can be hard to coordinate with the security goals of a corporate network that is to guarantee confidentiality, integrity and availability. One of the biggest problems is that with these new interconnections, the SCADA network now faces the same threats as the corporate network but lacks the appropriate security mechanisms that the corporate network has [4].

There exists much research on the topic of SCADA networks and its security issues. One important factor is the legacy property of SCADA networks. During the years, SCADA has moved from using its own protocols to using standardized open protocols, which means that it is fairly easy to find out how these protocols operate. As in any other business you want to save money when possible and therefore it is often cheaper and more time-efficient to buy products than to develop them yourself. In terms of SCADA networks, commercial off-the-shelf (COTS) products, i.e. a generic product that is developed to fit in a lot of different scenarios, have been very common and these products are rarely developed with security in mind. With the use of COTS products there is a need to be extra careful, as hardware used by SCADA networks can be very sensitive to disruptions. If COTS software is used and the manufacturer releases a patch, can the SCADA controlled system handle that several units will restart to install the patch? If they cannot, will they instead keep the old software and thereby not be recipient of security updates and such.

Further, SCADA protocols usually do not have support for cryptography which means that attackers can listen to data being sent and then use this information to carry out a range of different attacks. The future research challenges in this area are among others access control, firewalls, intrusion detection systems, operating systems security and device security, which all needs special concern when it comes to SCADA networks due to its special properties [11].

2.2.2 Modelling for security analysis

The complexity of an infrastructure can be divided into two general categories: structural and dynamical. The structural complexity is a result of heterogeneity of components and the large number of connected components connected by dependencies. Dynamical complexity refers to how the system reacts to changes in the operational environment. The dynamical complexity can increase when the system grows and evolves. The result can be that the system's properties may change and thereby creating a new operational environment [12]. When it comes to vulnerability and risk analysis of critical infrastructures there are a number of challenges that need to be handled. One potential problem is that classical methods do not have the capabilities to capture the structural and dynamical complexity of critical infrastructures and therefore there is a need for new methods that better can capture the infrastructures characteristics.

There are some great challenges within modelling and simulation of critical infrastructure that needs to be addressed in the future. The most significant problem is how to obtain the necessary data to accurately populate the model without at the same time creating a security vulnerability by showing the systems and its properties. Some data is regulated by laws to protect functions that is important to the society. Another challenge is how to verify and validate a model, since models aim to provide insight into how a system would react to scenarios that have not yet occurred, it can be hard to determine the accuracy of a model. The most common ways to validate and verify is to use historical data or to gather the opinions of field experts [13].

The CORAS method is used to conduct security risk analyses. As a part of the method so called threat diagrams are created. These threat diagrams can also be used to analyze mutual dependencies. The approach is to, after a security analysis of a system, reduce the scope to just concern individual components. By having a comprehensive security analysis, it can be assumed that dependencies will arise on a component level. To exemplify the approach, consider an example where two sticks lean on each other, which means that if one stick falls the other stick will also fall. Based on different factors each stick has a probability to fall which then can be inserted into a mathematical formula to calculate the risk that the sticks fall [14]. One negative aspect of using CORAS in this situation is that the language is not detailed enough and therefore might fail to capture important aspects.

2.3 System modelling

A part of the work in this thesis aims to investigate what steps are necessary to provide a holistic view of a system and its dependencies. There exists earlier work that have had partially the same focus and can therefore benefit the work in this thesis. The following section will describe different kinds of method that can be used to model a system.

2.3.1 System modelling methodologies

When modelling a system there exist a number of different approaches and methods that each have their strengths and weaknesses. One major challenge when it comes to dependencies within complex systems is how they can be modelled to capture their characteristics. To add further complexity, systems are often dependent on external systems, i.e. systems that are managed by others, which means that data about the systems and its current state may be hard to obtain.

An infrastructure should have a protection plan. When developing such a plan, there is a methodology called "Criticality assessment methodology" that can be used. This methodology is aimed at the development of an infrastructure protection plan. The methodology uses a three-layer model: operator layer, sector layer and intra sector layer. The operator layer can for instance be an infrastructure, then the sector layer defines the infrastructure that the first infra structure is directly connected to. The intra sector layer then includes infrastructures the is not directly connected. The layers are used to define different characteristics and interdependencies and by using a layered model they can define different goals and requirements for each of the levels. The method is built on a dependency tree that the Chief Information Officer (CIO) of the company should assemble, since the CIO is assumed to know or to get knowledge about previous risk assessments, what dependencies that exists and how they affect the own organization. With this information, a formula is suggested to calculate how risks can travel through the organization and what the impact can be at different stages [15]. This method will have reduced applicability since it requires the CIO to participate.

To combine different perspectives is important in order to get a holistic view of a system. One way to this is to use a method called Mixed Holistic Reductionistic (MHR). The method is intended to use the benefits of both the reductionistic and holistic perspective. The benefits of the holistic perspective are that it has clear boundaries which makes it easier to define communication between entities and which functionality that belongs to which entity. The reductionistic is a more low-level perspective where the focus is more on individual components. With the reductionistic view there is a risk of getting a very complex system and with the holistic view the risk is to miss important aspects by using a too high-level analysis. Further, the importance of including the human role in a system is highlighted. The suggested method is to add a fifth dependency category called social dependency as an extension to the four earlier presented categories. To illustrate the difference between reductionistic and holistic, using a holistic approach a control room would be modeled as one entity. With the reductionistic approach all the underlying components that make a control room would be modeled instead. To create a bridge between the perspectives they add a new middle layer called the service layer. The service layer is intended to specify what service is provided by the entity in the holistic view and what components that are needed in the reductionistic view [16].This definition applies well to the goal of this work, although the definitions are rather loose and open for interpretation.

2.3.2 Modelling perspectives

When modelling a critical infrastructure, a combination of perspectives can be used. To fully have the possibility to capture a system's full characteristics, a combination of perspectives needs to be used. Four different modelling approaches that are considered as necessary parts of a complete framework are the following [12]:

- Topological
- Flow
- Logical
- Phenomenological

The topological perspective is used to describe the connectivity of the system and how cascading failures can travel through the system. Further it can also be used to analyze which components that are the most essential to the system. The flow perspective should describe the system's physics and how it is monitored and controlled. The logical perspective should capture the functional properties of the system and how failures of hardware, software and humans will affect the systems functionality. The phenomenological will capture the dynamics of different parts of the system and how their respective operations are related to each other. It has been concluded that there is a great need for new modelling and simulation techniques. These new techniques should have the ability to fully capture the complex properties of critical infrastructures to lay the foundation for risk assessments that are well grounded [12].

By creating a topology map, important indicators about the network can be gathered. For example, indicators such as node degrees clustering and the centrality of the network can be revealed. The node degree represents the number of connections there is to a node which in turn can be seen as an indicator of which nodes that are the most essential and important for the system's functionality [17].

A system can also be shown from a functional view i.e. what is the systems input, how is it traveling through the system and what is the systems output. This is suitable when the goal is to understand the route through the system and how different parts work together.

Another possible way to model a network is to model it from a hierarchal point of view. With this approach, the focus is on what entities that exist on different levels and not on how many of each kind there are in the network. A hierarchal approach is good when e.g. segmentation needs to be analyzed since it presents what entities that are available at each level and what security mechanism need to be passed in order to get to another level.

The functional view and the hierarchal view is covered with the three-layer model and with SecuriCad, which is why a topology map is well suited to complement these models. The topological view is commonly used and therefore more familiar to people. Therefore there is a higher probability that people can understand the topological map.

2.3.3 Modelling techniques

A critical infrastructure is an example of a system-of-systems, which means that the infrastructure has a number of underlying subsystems that each also has underlying units and so on. The theories of systems-of-systems have been applied to SCADA networks to analyze what kind of dependencies exist between the SCADA network and its underlying units. One thing that can be concluded from this kind of analysis is that, in order to provide accurate models and useful results, it is necessary to analyze the system and its subsystems in depth and thereby cover all components. When working with a system-of-systems there is a need to handle the underlying systems to reduce the complexity of the analysis. Three approaches for reducing the complexity are Agent based modeling, High level architecture and Hybrid systems which all are suitable when there are large networks with a high number of nodes, connections and different capabilities among the nodes [18].

Rinaldi [13] presents an overview of modelling and simulation techniques within critical infrastructures. To start with, some factors that are important to consider when analyzing dependencies are pointed out. These include among others, time scale, cascading effects and business policy. The next part concerns different types of dependency modelling and simulation techniques and their different characteristics:

- Aggregate supply and demand tools, evaluate the need for a critical infrastructure service in a region and how well the demand can be met.
- Dynamic simulations are used to investigate the characteristics of an infrastructure, how it reacts to disruptions and what the consequences can be.
- Agent-based models are used to model the operational properties of an infrastructure and its different physical states.
- Physics-based models are used to analyze the physical aspects using physical measurements.
- Population mobility models are used to investigate how entities move within a region.
- Leontief input-output models can provide analysis of commodities life cycle in an area.

Agent-based modelling means that so called agents are created to represent a set of devices that together realize a functionality. By doing this the structural complexity can significantly be reduced and make the system easier to comprehend. Agent-based models can be used in system overviews in a way where agents are created for parts of the system that have similar characteristics. Further, agent-based models are very common when modeling critical infrastructure [19] due to the fact that an infrastructure is a very complex system. While at the same time having many nodes with the same properties. The use of agents addresses the problem of the high complexity of a critical infrastructure. With the use of well-defined agents the complexity can be reduced while at the same time keeping the characteristics of the system.

With dynamic simulations, predictions about how the system reacts to e.g. disruptions or disturbance can be made. The simulation model will be built on nodes and their connections, where nodes can either be agents or a physical entity. These simulations can provide information about nodes that are important for the system’s functionality and which connections that are the most important.

The Leontief input-output model is mainly aimed to study the economic flow, but it has also been used to study infrastructures. The model can provide a linear analysis of how commodities are used among infrastructure sectors. The model has also been extended to include an analysis of how risk can spread among interdependent infrastructures [13].

The interference level provides information about how much interference that a part is subject to, the interference can both be internal and external. Haimes et al. [20] have, based on the theory of Leontief, developed a mathematical expression for calculating the interoperability level *x*. The formula is as follows where A is a matrix consisting of the internal interference among parts and C is a matrix consisting of the external interference at each part.

$$x = Ax + c$$

By calculating x you get a number that represents the interoperability level of each part. Therefore, the formula can also be a tool when there is a need to translate information from system users into an actual number that can be used by decision makers.

2.4 Dependencies in SCADA systems

The research on dependencies among critical infrastructures is quiet rich, several studies have focused on how different kind of infrastructures depend on each other and what happens if one infrastructure fails. The work in this thesis is instead focused on dependencies of one single infrastructure and its underlying dependencies. This can be seen as making the analysis at a lower level of abstraction. Based on the studied literature, much of the theory that is used to analyze dependencies among infrastructures can be generalized to a lower level of abstraction. Therefore the same kind of theory is applicable to this thesis.

One modelling approach that has got some attention is a modelling approach which focuses on what the system's goal is and what dependencies exist to achieve this goal [21]. This is then repeated for the identified dependencies to identify the next level of dependencies. This procedure is repeated a number of times, creating a tree structure of what dependencies exist to achieve the goal. The approach also includes a process called *zooming* which aims to solve the problem of sharing data between external systems in order to make the individual risk analysis more comprehensive. The idea is to create a central storage where system owners can upload their dependency model and then other system owners can request parts of the model in order to extend their own dependency model. With this approach, all contributing system owners will have the opportunity to get information about how their external dependencies can influence their own system and thereby be more prepared when something happens. This approach is well suited for analyzing dependencies between infrastructures due to its extensive workflow. The higher amount of work needed to set up the workflow is also why it is less suited in this thesis.

2.4.1 Dependency categorization

As earlier stated a dependency between two nodes is possible to model by modelling the connection between two nodes. According to a review by Ouyang [19] there exists several different proposals of dependency categories, some are similar while some have completely different perspectives. Ouyang have gathered five different sets of dependency categories that is evaluated in terms of which set that has the best coverage. The number of categories in each set differs from two to five and some categories are very broad e.g. logical and some very narrow e.g. budgetary. The complete sets are defined as follows:

- Physical, Cyber, Geographic and Logical [8]
- Functional and Spatial [22]
- Physical, Geospatial, Policy and Informational [23]
- Input, Mutual, Shared, Exclusive or and Co-located [24]
- Functional, Physical, Budgetary and Market and economic [25]

To evaluate which set of categories that have the best coverage, Ouyang uses ten different scenarios that each describes different kinds of failure scenarios affecting infrastructures. Then, Ouyang uses each set of dependencies and tries to cover as many scenarios as possible. The results show that only one set of categories has the capability to cover all the scenarios and this set of categories is proposed by Rinaldi et al. [8] and will therefore be the categories used in this thesis. The categories are described as follows:

- **Physical**: a physical dependency means that the state of one node is dependent of a material output from another node.
- **Cyber**: a cyber dependency means that one node is dependent on information that is sent through the information infrastructure from another node.
- **Geographical**: if there is a geographical dependency it means that an event in the environment can affect several nodes at the same time due to closeness in space e.g. flooding of a building.
- **Logical**: there is a logical dependency if the state of one node depends on another node and it is not a physical, cyber or geographical dependency.

It is worth noting the definition of the logical category, which is that anything that does not fit in any of the other of the three categories will end up in the logical category. With this definition, the logical category will have very broad scope which both can cause it to appear more frequently and also affect the possibility to perform precise analyses. Although it is by far the set of categories that is the most common in this field of research which further motivates the use of these categories despite its shortcomings.

2.4.2 Human dependency

To involve the human factor in a system's performance can be done in different ways. One way is to extend the four previous dependency categories by adding a new kind of dependency, called the social category [26]. The social category is described as "an infrastructure has a sociological dependency when its operativeness is affected by the spreading of disorder related to human activities".

The importance of including the human role in a system has been pointed out by e.g. Zio [12] and Gheorghe et al. [27]. Two methods that are considered to be a part of the second-generation methods for analyzing human behavior are the Cognitive Reliability and Error Analysis Method (CREAM) and A Technique for Human Error Analysis (ATHEANA). Both methods are based on the principle that the context which human tasks are performed in, is the most influential property. Therefore, it should be an important aspect when modeling the human error probability. Despite being part of the second generation, these methods are quite old and outdated, making them a less suitable choice.

2.4.3 Dependency risk assessment

When assessing risks that comes from dependencies it is important to include dependencies of higher order. One way to do such analysis is to start by choosing a root node, which is the node that the analysis will focus on, and then go through the following steps in order to assess the risks. The steps also include changing the root node until the whole system is analyzed [28].

1. Identification of the first order dependencies
2. Identification of n-order dependencies
3. Evaluation of the cumulative dependency risk
4. Change root node
5. Rank cascading risks
6. Mitigate cascading risks

To assess the risk of each chain, a formula is used to calculate the dependency risk, the formula for one single dependency is first defined as follows:

$$R_{Y_{O},Y_1} = L_{Y_{O},Y_1} \cdot I_{Y_{O},Y_1}$$

Where $R_{Y_{O,}Y_1}$is the risk value of a dependency between node Y_0 and Y_1, $L_{Y_{O,}Y_1}$ represents the likelihood of an event and $I_{Y_{O,}Y_1}$ represents the impact an event will have. The likelihood and the impact for each dependency is defined on a scale that includes the following steps along with a numerical representation to be used in the calculations:

- Very low (0-0.05)
- Low (0.05-0.25)
- Medium (0.25-0.5)
- High (0.5-0.75)
- Very high (0.75-1)

To assess the likelihood and impact of a potential event at each node it is claimed that qualitative data should be used i.e. that each factor should be determined by system experts. Although this can cause the work to be time consuming depending on the size of the system. By covering all dependency chains and thereby getting the cumulative dependency risk ($DR_{Y_o,\ldots,Y_n}$), the previous formula is extended into the following for the same root node:

$$DR_{Y_o,\ldots,Y_n} = \sum_{i=1}^{n} R_{Y_0,\ldots,Y_i} \equiv \sum_{i=1}^{n} \left(\prod_{j=1}^{i} L_{Y_{j-1}\ ,Y_j} \right) \cdot I_{Y_{i-1}\ ,Yi}$$

The dependency risk is calculated by first summarizing the risk for each underlying node pair which is according to the previous formula the likelihood times the impact a failure will have on each connected node.

3 Methodology

This chapter will first describe the research methodology that was used to answer the research questions. Then, there will be a description of the work that was performed when putting together and implementing the mapping method.

3.1 Methodology overview

The work is divided into four steps, the first is to gather data about system modeling, system dependencies etc. Then, based on the gathered data the mapping method will be created. After creating the mapping method it will be implemented on a test system. Finally, based on the implementation, there will be an evaluation of the method. This workflow is illustrated in Figure 7.

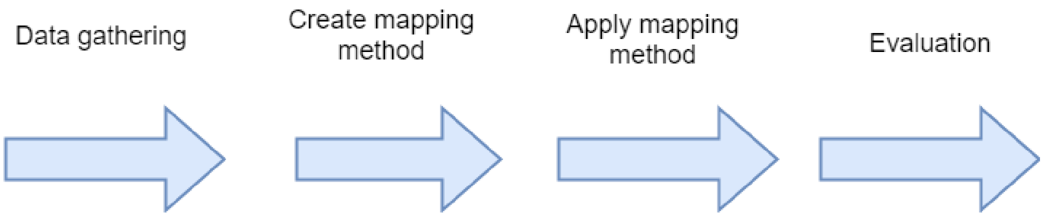


Figure 7 Thesis workflow

The first part will be to gather data that is needed for the following work. The data gathering will lay the foundation for how the mapping method will be constructed, and what aspects the method should focus on. The next part will be to put together the mapping method, this will involve choosing what parts to include and how information should be presented. After creating the method, it will be applied to a system to enable an evaluation of the method.

3.2 Data gathering

To determine how a dependency analysis can be useful for a future risk assessment, there are two important aspects. The first aspect is: how can dependencies be analyzed and how can they be presented. The second aspect is: what properties about dependencies are useful for a future risk assessment.

To gather information about how dependencies can be analyzed, a literature review was conducted. The review was focused on literature that concerns system dependencies, SCADA systems and system modelling. To find useful literature, the following aspects were considered when searching papers.

- Frequently sited
- Published in well-known journals
- Presented at well-known conferences

A mix of old and new literature was used in order to both get information about well-established concepts and theories, as well as new perspectives.

To further gather data, discussions with employees at Tekniska verken were held to get a better understanding of how a SCADA system can work and what kind of dependencies there can be. These discussions were very informal and served the purpose of getting a better understanding of how SCADA systems work. The employees had different work assignments and could therefore provide different perspective and information about SCADA systems.

3.2.1 Interviews

Assessment of the current security situation can be done in several different ways. One study has focused on the process of assessing the security situation and present seven steps that should be involved in the process [29]. The first step is to perform interviews with employees that have knowledge about different parts of the systems in question. In this thesis, interviews will be the main contributor to the resulting maps and will be used for guidance. An important part is to identify which employees that will be interviewed based on who has knowledge about the system. One benefit of doing interviews is that you can get different perspectives of the same system and identify key areas [30].

Therefore the chosen method for data gathering within the organization is to use interviews with several benefits. Since the purpose is to create a map of the system, one other option would have been to scan the network for devices, but scanning has some major drawbacks which makes it a less good choice in this situation. When performing scanning within a SCADA network it can affect the performance of some devices and thereby interfere with the system’s capability to deliver information. Often it is also harder to get the system owners approval for a scan as opposed to doing interviews with the users. On the other hand, there are drawbacks with interviews as well, for example interview relies on identifying the right people to talk to and also that these people can provide all their information. Although these are most often problems that are easier to handle than the ones with scanning. By using interviews, there is also a possibility to include another aspect that is not possible with scanning, which is the human role within the system. The human role can be very significant for a system and should therefore be included in the analysis, especially with the increasing number of social engineering attacks.

To use interviews is to use a so-called knowledge-based approach. This means that the results relies on the knowledge of the people that participate in the interviews. Interviews also have the benefit that you get the opportunity to aske in-depth questions and really focus on certain areas that might be of special interest. With interviews, the possibility to find previously unknown dependencies increase since the questions aim to gather new knowledge and not only confirming already known information.

3.2.2 Interview guide

The first part is to identify people to interview that have some knowledge about the system. It can also be useful if they know who else has knowledge about the system and thereby can provide more information. This part is crucial, if not done in a good way there might be time loss and if information is gathered in a suboptimal order this can make the assembling of knowledge much harder.

The interview questions were developed through an iterative process, where the questions were subject to feedback both from an interview technical and topic-related point of view. The topic related part gave feedback in terms of domain specific questions, using the right terminology and questions that potentially could be hard to answer. The interview technical part gave feedback in terms of how to structure and formulate the questions and also how to ask questions that can provide interesting results that can be analyzed. The intention of the interview is to start by letting the interview subject explain his/her general picture of the system and then ask a number of follow-up questions that are supposed to add more detailed information to the picture which clearly relates to the earlier described top-down approach. The picture is supposed to be a working product through the interview and by asking questions about it, make the interview subject think of new perspectives that they have not previously considered. This means that the interviews will be semistructured and the questions are only means to make it easier for the interview subject to describe the system and its parts. Therefore, the interview guide should be more seen as a backup tool during the interview, and more focus will be on the talking and discussing with the interview subject. The interview guide that was used during data gathering can be found in Appendix A.

3.2.3 System modeling

To be able to evaluate the three-layered model and SecuriCad, an understanding of both tools was needed. Both to be able to apply them in a correct way, but also to understand the results they can produce. To gather knowledge about SecuriCad, the documentation served as a good basis. There are also some research papers that was conducted when developing SecuriCad, these papers were also read to further understand how to work with SecuriCad. When choosing to work with SecuriCad the following aspects was identified as a beneficial and should therefore be utilized.

- Good structure
- Well-known system view
- Focused on critical infrastructure

To gather information about the three-layered model, previous applications were studied to understand how to work with the model. The three-layered model allows much freedom to apply it in a way that is best suited in specific situations. Therefore, there was a need to understand how it best can be used in the scenario of this thesis. The important factors of the three-layered model that was identified were the following and should therefore be taken advantage of:

- High level of abstraction
- Easy to understand
- Model complex systems in an understandable way

Finally, an approach to reduce the complexity of a system model were needed to be chosen. When choosing the approach, it was important that the approach could reduce the complexity of a system while at the same time do not loose tom much of the system's characteristics. As mentioned in Section 2.3.3 there are different techniques that can be used in system modelling. When reducing the complexity, it is important that the characteristics of the system still can be kept, otherwise important information can be lost. If information is lost, a following risk assessment cannot be done in a proper way.

3.3 Creating the mapping method

This section will describe each of the individual parts that together are intended to make the whole mapping method. Each section will include a motivation of why each part was chosen.

3.3.1 Method overview

As stated in the problem statement, one goal is to investigate if system dependencies can be modeled to contribute to future risk assessments. The chosen approach to investigate this is to base the mapping method on the dimensional model presented in Figure 5. The reason for choosing the dimensional model, is that it provides a wide spectrum of perspectives to focus on, while at the same time is has clear definitions of each dimension and what should be incorporated and what should not. Among the six dimensions, three dimensions are selected to be concerned in this thesis. The three dimensions were chosen since they are closely related to the concept of dependencies but also since there are several different ways to work with and model them. The selected dimensions are *types of interdependency, type of failure, coupling and response behavior.*

As a first step of the mapping method, a topological map will be used together with the dependency categorization to show an overview of the system. From a topological map, it is possible to derive information about the connectivity of the system's nodes. The second part will be to illustrate higher order dependencies by using a tree structure. The final step will concern coupling among system functions.

In section 2.3.2 four perspectives were presented as needed in order to fully capture a system's characteristics. The steps of the methodology are intended to apply to these perspectives. The three-layer model can be seen as a combination of the topological and logical perspective since any kind of connection is relevant in the model. By creating a topology map, the topological perspective gets covered. By creating a tree with dependencies of higher order the logical perspective is also included. The purpose of the logical perspective is to understand how failure of hardware, software and human affects the system, which applies well to the analysis that can be made from the dependency tree. The step that concerns coupling relates well to the phenomenological perspective since it aims to describe how different parts with different functionalities relate and depend on each other.

3.3.2 Analysis approach

To model a complex network, you need to have a good structure when it comes to where to start and where to end the analysis. Basically there are two major approaches to choose from, either you start from the bottom and work your way up or you start from the top and work your way down. They have different pros and cons and are suitable in different systems. A top-down approach tends to be more focused on what dependencies there are to achieve a system's goals while a bottom-up approach focuses more on what can go wrong and what implications that might follow.

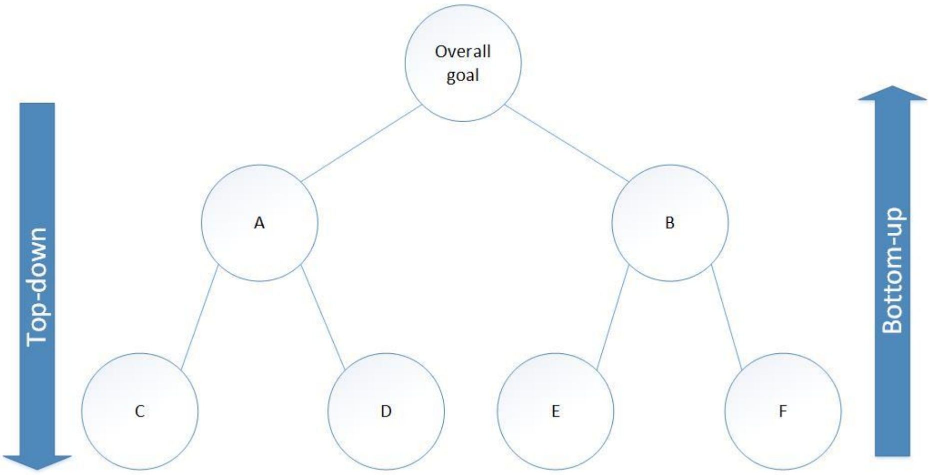


Figure 8 Bottom- up and Top-down approach

With the bottom-up approach the analysis would focus on either component C, D, E or F and then go through the tree to find what consequence there might be from a failure in any of these components and then how it would affect the overall goal. Some of the components C, D, E or F might just have minor impact and therefore do not need much attention.

With the top-down approach the analysis would instead start with the overall goal and see which components that are the most essential. With this approach, you ensure that the primary focus is on the essential components and how they contribute to the system's functionality.

Since SecuriCad uses a bottom-up approach, the top-down approach will be used in the layered model. A top-down approach is useful to a critical infrastructure since critical infrastructure needs to be able to deliver its service at all times. This makes it natural to start the analysis by identifying key components that are needed for the infrastructure to function. This will highly affect the approach of data gathering, instead of asking "*which components can fail in the system*?" the main question will be "*what components are essential for the system's functionality*?" i.e. what are the most important dependencies. With a bottom-up approach there is a great risk of wasting time since there is often a huge number of things that can fail and some of them may just have minor impact on the system. With the top-down approach you instead focus on the components that really are essential and worth spending time on. The analysis approach will be used in all the presented models and perspectives.

3.3.3 System overview

With reference to the first research question stated in Section 1.1, this section will how the three-layered model and SecuriCad were used to map out the test system.

As previously stated SCADA systems have an increased structural complexity due to the high number of nodes where many nodes have very similar functionality. Therefore agent-based modeling will be used in both cases to reduce the complexity.

3.3.3.1 Three-layer model

To be able to map out how different systems and networks interact and depend on each other it is important to keep in mind that there can be different kinds of interaction. Therefore, to provide a useful result, different kinds of interaction needs to be individually analyzed. When using the three-layer model, the final description of the system will be a holistic view of the system including physical connections and human interactions.

The result will be presented with the help of a map consisting of three horizontal levels describing the three levels of interaction. At each level, there will be a number of nodes that represent the components within each layer. The nodes will be connected with edges that represents dependencies between the components, the edges can be both horizontal and vertical. In Figure 9 this is exemplified, the model is inspired by Nan et al. [31] and Johansson et al. [32] who uses similar models to visualize dependencies among critical infrastructures.

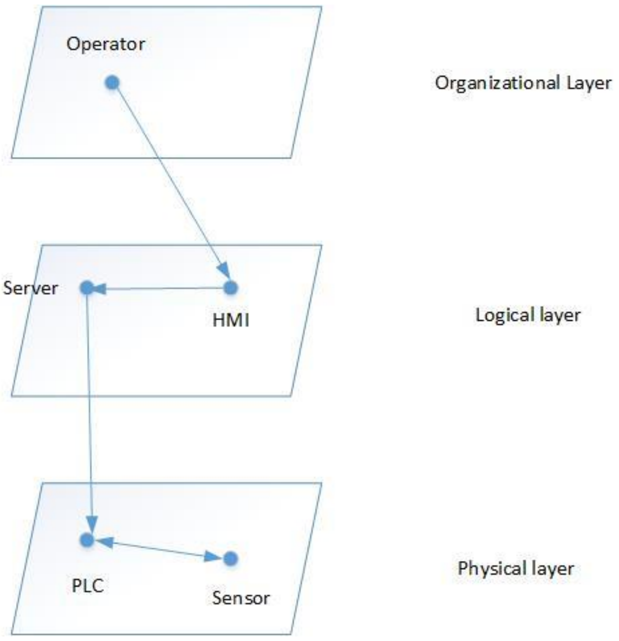


Figure 9 Exemplified dependencies in layered model

The first layer is called the physical layer, essentially this layer consists of things that you can physically grab, such as cables, routers, sensors, etc. The second layer is called the logical layer, this layer consists of firewalls, data, 3G communication etc., i.e. some sort of software that can be programmed to act in different ways. The logical layer will therefore include nodes that are used to realise functions, such as communication and control software. The final layer is the organizational layer, this layer focuses on people and organizational behavior and how they interact with each other. When these three layers are analyzed together they can provide an extensive overview of the organization and what connections and interactions exist. This framework can provide a system description with a wide scope and will therefore be evaluated based on its ability to capture a SCADA system.

By using the three-layered model to provide a high-level view of the system is to use a top-down approach. With a top-down approach the focus is on high-level dependencies that are essential for the whole system’s functionality. This means that the information is focused on high-level components that depends on other high-level components.

3.3.3.2 SecuriCad

The other model that is used, is to use the commercial tool SecuriCad. SecuriCad is based on a research prototype of the modelling language P^{2}CySeMoL and has extended the language.

SecuriCad is available with different amounts of included functions, where the one with the least number of functions is available for free. The free version was the one used in this thesis. The limited number of functions will not have an impact since the left out functions does not concern drawing a system.

The reason for choosing SecuriCad is that it can provide good structure when modeling a system and can therefore be a good fit to be able to model a SCADA system. With respect to the property of laying the foundation for a future risk assessment, SecuriCad with its focus on SCADA and security, is a good fit. It is also claimed that the tool does not require the user to be an expert of cyber security or system modeling. With the possibility to exactly define the configuration of the system, the model can give valuable input on potential vulnerabilities in the system and which the weak points are. One great strength with SecuriCad is the ability to model parts of the system in a very detailed manner. Consider Figure 10 which shows the interactions of just one host. With this level of detail, it is possible to analyze dependencies that would have been missed in a less detailed model.

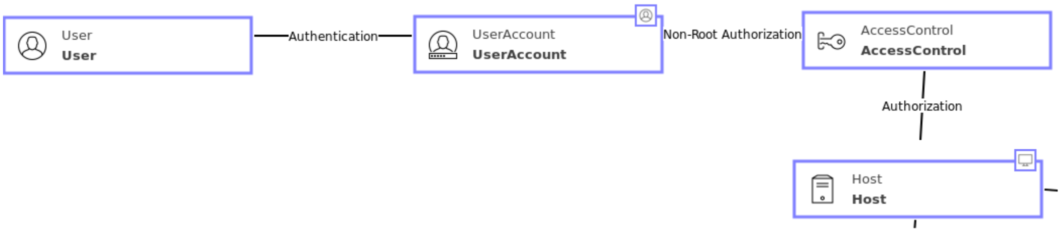


Figure 10 Interactions with a host

By using SecuriCad to model a system, a logical perspective of the system is provided. Further, by using SecuriCad you get a bottom-up approach perspective. This means that the model allows an analysis of individual components what then together makes the whole system. By using the bottom-up approach the focus is more on low-level dependencies among system components. With the logical perspective, the functional properties of a system can be analyzed, this includes how failures in the system will affect the overall system. To analyze the effects of failures is important in terms of a risk analysis which makes this a suitable choice in this scenario.

3.3.4 Dependency categorization

To analyze the first perspective, *types of interdependencies,* the categorization in Section 2.4.1 will be used in this thesis since it has the best coverage. The categories are:

- Physical
- Cyber
- Geographical
- Logical

As a way to analyze dependencies these categories will be visualized in the topology map to show where in the network different kinds of dependencies appear. This is intended to make room for an analysis of where certain categories of dependencies appear and if there is some category that appears more frequently or if there is some category that appears very frequently in a specific part of the network.

To include the human role within a system, the chosen way is to work with nodes that represents roles performed by humans. In cases where humans instead are the linkage between hardware, the logical category will be used to visualize the human role. The reason for not using the social dependency category is that it is not very commonly used and also its definition is a bit vague.

3.3.5 Topology

Topology-based method belongs to the category of network-based approaches [19]. This means that the modeling is based on the system's topologies where components are represented as nodes and connections between nodes are represented as links. This will give a map that provides a good overview that is also easy to understand. According to Ouyang [19] the topology-based method needs rather low quantity of data and the needed data is considered to have relatively good availability compared to other kind of data. Further it can also model all of the interdependencies of the set used in this thesis. No other model has the ability to model all kinds of dependencies and therefore cannot give a full coverage. Different methods can provide different kinds of information, and according to Ouyang the topology-based methods can be used when e.g. the goal is to "harden and protect key components". This is because a topology can provide good overview and thereby makes it easier to identify what components that are of extra importance and need extra protection.

The topology map was created using the tool Microsoft Visio[2] which has templates for network drawings and therefore is suitable in this situation. Visio has also been used for similar purposes earlier at Tekniska verken and is therefore a natural solution that can present a result that can easily be understood at the company.

Following the creation of a topological map there are a number of characteristics that can be analyzed. The mapping method will include the degree of connectivity of each node. This information will be presented to show which nodes are the most essential and thereby play a central part in the system.

3.3.6 Higher order dependencies

The next dimension is *type of failure*, which originates from a topology map. If e.g. there are a lot of geographical dependencies located at some area, all of the interconnected nodes are exposed to a common cause failure if there is an event that affects the environment e.g. flooding. Cascading and escalating failures follow the same pattern, the only difference is that cascading failure increase in severity, hence they can be viewed in the same way. To find these risks, focus needs to be placed on dependencies of higher order. To analyze higher order dependencies, a tree structure will be used as done by e.g. Kotzanikolau et al. [28] where one node (physical, logical or organizational) will be chosen as root and then the root's first order dependencies will be investigated and mapped as new nodes into the tree. This procedure is then repeated for every

[2] https://products.office.com/sv-se/visio/flowchart-software

node until no more sub-dependencies can be found and the tree is completed. With this tree structure, ways that cascading/escalating failures can travel can be identified. If it is cascading or escalating needs to be assessed by system experts and is therefore outside of the scope of this thesis.

To further add information about the root nodes current situation the algorithm by Kotzanikolau et al. [28] presented in Section 2.4.3 will be used. This algorithm can be used to translate subjective opinions from system users into an actual number that represents the calculated dependency risk and can be easily comprehended by decision makers. The usage of the formula is a straightforward process, the data that is needed is an assessment of the likelihood and impact that each dependency in the tree structure will have on the system. These assessments are then inserted in the formula and are used to calculate the final dependency risk.

3.3.7 Coupling

The remaining dimension is *coupling and response behavior*. To illustrate this the systems will be split into its underlying units. Then, based on the information from the interviews, each part is assessed on its influence on the connected units and then added to a map. The map includes connections that represents the influence among units and thereby create a map showing how the system's units are coupled. In terms of response behavior, it may be very specific for individual dependencies, but by describing a node's properties the response behavior can be assessed by system experts.

The analysis of coupling will consist of two steps, the first will be to use the formula by Haimes et al. [20] presented in section 2.3.3. The formula provides information about how a part's interoperability level is affected by the interference from its connected parts.

The second step is to analyze each part individually. There are two variables that together fully can describe accidents within a system and these are *Interactions* and *Coupling*, which will therefore be used. The interactions between system components can either be tightly coupled or loosely coupled: with tightly coupled components there is no tolerance of delay and loosely coupled components are the opposite. These interactions can either be linear or complex, linear is the same as simple and complex is the opposite [10].

3.4 Application of the mapping method

To evaluate the mapping method, it was applied to a test system. The data needed to apply the method is based on the answers from the interviews during data gathering. The test system was created together with Tekniska verken in order to get a system the as far as possible have the characteristics of a real-world system.

Based on the description in Section 2.2 the test system will consist both of SCADA specific parts but also other parts e.g. an office network. The system will have the characteristics and properties of a real-world system to make the analysis more reliable. Some of the most important parts are:

- Remote sensors
- Office workstations
- Operations center
- External connections

As a result of using a system similar to a real-world system, it will have the different kinds of dependencies that are important to analyze and will be an important part for proving the mapping method's usefulness.

Each of system's locations serves different purposes in the system and people working at the different locations are focused on different parts. As an example, at the office there are engineers working with the system's overall performance and planning future development. At the operations center, there are operators that have responsibility for monitoring the current status. External connections can be used by e.g. people working from home or third-party suppliers that are responsible for the support of some parts in the system. If something goes wrong with entities in the field, a technician must find that location to fix, repair or solve the problem. This map is only focused on showing logical connections in the system which means that geographical location of the nodes is not accounted for i.e. entities that are located close to each other logically may be separated geographically. Figure 11 is an overview of the system and therefore shows the system with a higher level of abstraction, a legend can be found in Appendix B.

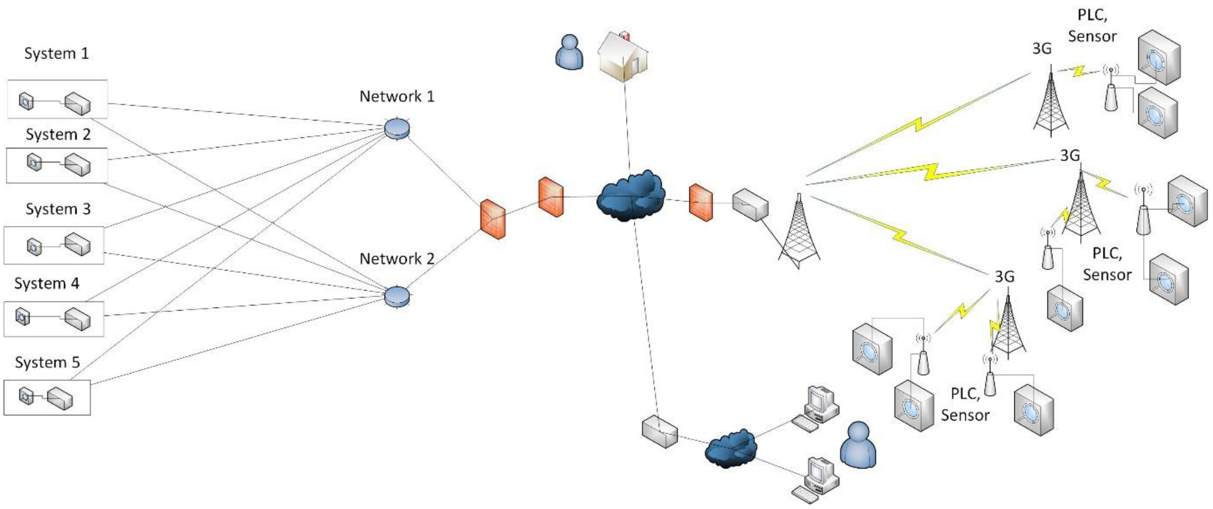


Figure 11 Topology of the test system with a high level of abstraction

To make the analysis less complex and to make it easier to highlight important aspects the focus will be placed on how System 1 and System 2 are connected to each other. This is to limit the system to a reasonable size. Therefore Figure 12 is focused on Network 1, Network 2, System 1 and System 2 from Figure 11.

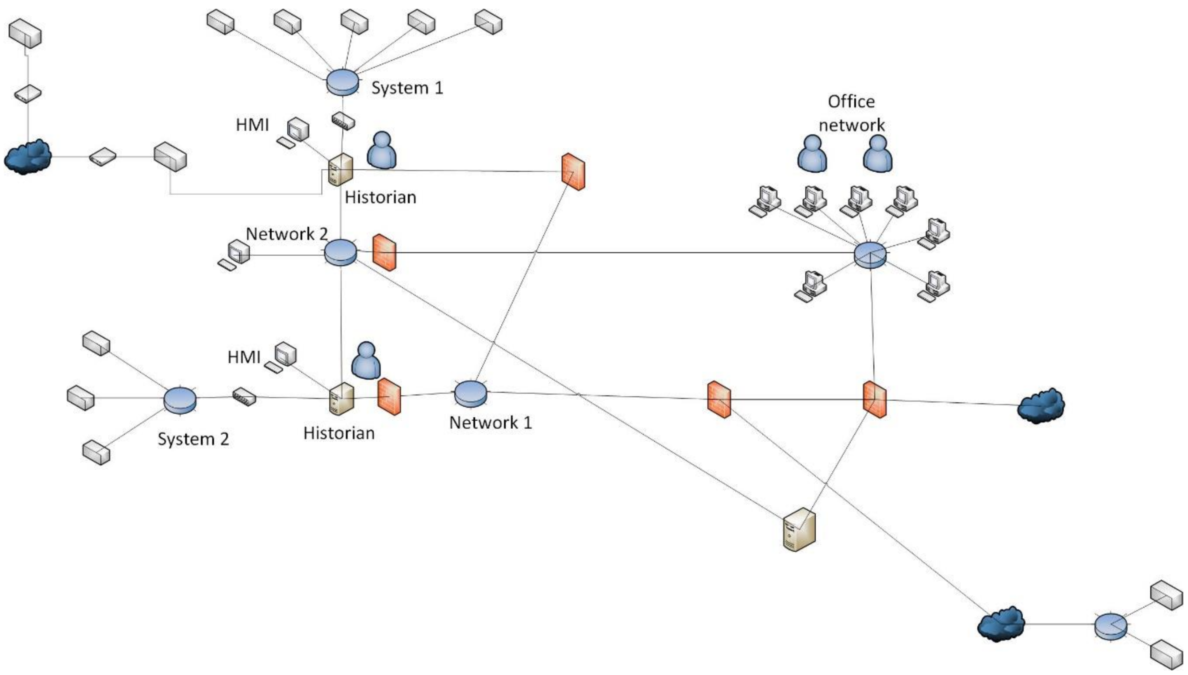


Figure 12 Topology of the test system with lower abstraction

Since geographical dependencies are important to analyze the map can be restructured to instead show the entities geographical location relative to each other. The black frames in Figure 13 indicates that the entities inside are placed geographically close to each other e.g. in the same building while the other entities are located at individual locations.

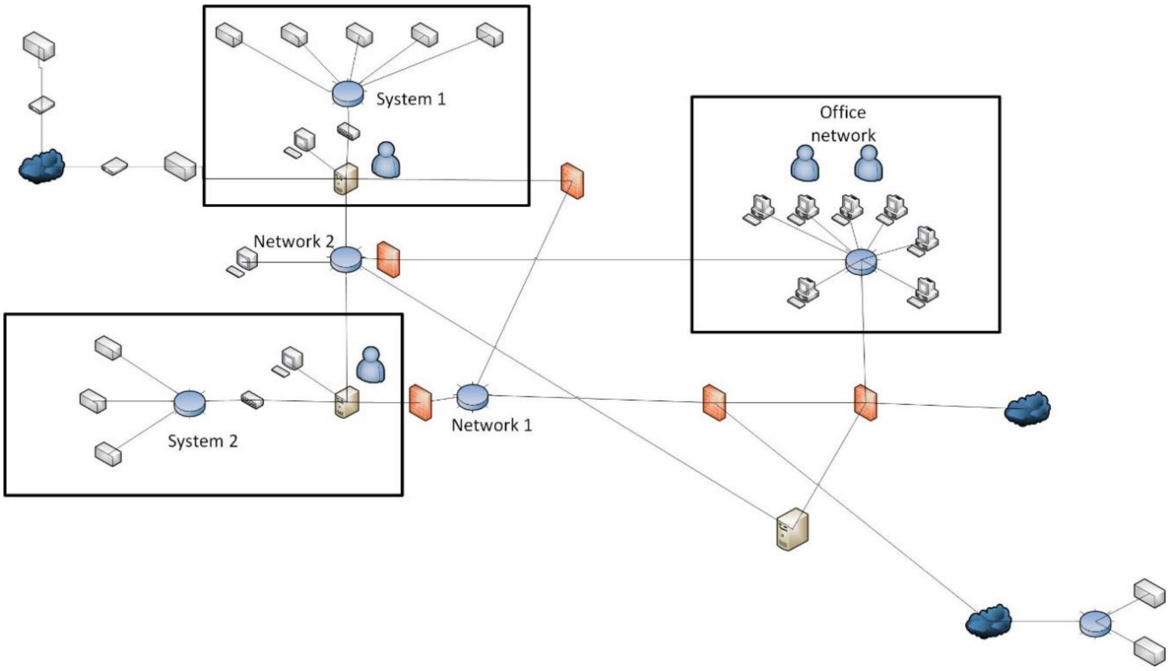


Figure 13 Topology with geographical indicators

The whole system is distributed over a wide geographical area, with certain parts that are located on a smaller geographical area. It makes use of a number of different communication techniques, such as cables, 3G, VPN, and Wi-Fi. It is segmented with firewalls that separates different parts of the system from each other. The picture also shows where users normally interact with the system. It has connections to other infrastructures such as telecom, broadband, electricity, and finally there are also dependencies to other networks. All these parameters together make a complex system. In the next chapter where the test system is subject to an analysis some parts are simplified in order to facilitate the work, although the intention is not to change important properties but instead to reduce the number of nodes to a certain extent.

As earlier stated there are several services that belong to the category critical infrastructure and in most cases the infrastructures depend on each other. From the information of the test system dependencies to telecom, broadband and electricity can be found. These kinds of dependencies are important to identify since they play a crucial role for the systems functionality.

3.5 Evaluation

When evaluating the mapping method, two key factors were identified, namely how easy the information was to understand and how informative the method is. These two are key aspects when the information should work as a preparatory step to a future risk analysis, as the knowledge is key to make correct decisions [1]. If the information cannot be understood it cannot be properly analyzed. Also, if the method is informative i.e. can provide much information, it increases the possibility to make a better risk assessment.

To evaluate the mapping method, interviews with key persons were held, these key persons were picked in collaboration with Tekniska verken and the most important aspect is that they have some kind of interest in the method and the results it can produce. Further they also have prior knowledge about risk assessments and can therefore assess the mapping method in this context.

When deciding on how many people to interview, the most important aspect was that they had some basic knowledge about the researched field. This was considered a higher priority than to have a higher number of interview subjects. This was because, without a basic knowledge, they cannot asses how well the different models can provide information. Further there is a risk that the interview would mostly consist of explaining the models and therefore take too much time. The result of this decision was to interview four people with the purpose of evaluating the mapping method. Three people belong to the IT department and the fourth person were a SCADA system operator. Each of these people have work assignments related to this work and therefore can provide valuable feedback.

The reason for choosing to do interviews instead of e.g. sending out a questionnaire along with the models is that with interviews you can have more of a discussion about the results and thereby pick up on opinions that would not have been possible to collect with a questionnaire. Further there is risk of getting very few responses by just sending out a questionnaire, by scheduling interviews there is a better chance of getting more commitment from the subject and thereby getting more valuable input. The interviews were about 45 minutes each to allow the subject to take as much time as they needed in order to understand and comprehend the material.

The interview was specifically concern their ability to understand and analyze the results, since this is a very important aspect to make correct risk assessments. One other important concern is how they feel about reusing the methodology in similar scenarios. If it is not likely that they will use it in the future, what aspects do they feel can be improved in order to make the methodology better. During the evaluation interviews, they were asked about their general opinion of the method and the results it can produce. Continuing, they will be asked about what general pros and cons they think the method has, what they would like to improve, and if they would reuse the method. The interview guide used in the evaluations is found in Appendix C. Much of the evaluation consists of letting the subject assess the material from different perspectives on a scale 1-5. This is a way to make room for comparison between answers and also make the subject to provide a precise answer. These assessment questions are accompanied by a question that concerns general strength and drawbacks of the material to make the evaluation more in-depth and thereby making room for further work.

The first part of the interviews concerned the first research question, regarding the applicability of SecuriCad and the three-layered model to model the test system. During the evaluation interviews, the interview subject was first given an introduction to the intention of the work and then a short presentation of the two different models and what they were supposed to show. Then they were given some time to process the information before the questions were asked. Important to note is that all subjects chose by themselves when they felt that they understood the models.

4 Analysis of the test system

This chapter will present the analysis of the test system. The first part is focused on the comparison of system modeling approaches and the second part is focused on how dependencies can be modeled and analyzed.

4.1 Three-layer model

Figure 14 shows a layered model of the test system, where some of the most common and important nodes have been included. Important to note here is the difference to a topology map in the sense that nodes that are not physical or logical are included, for example company policy and confidentiality of information can be included. These factors are crucial for a system and it is therefore important that they can be mapped out to show the role they are playing in the system. The selected nodes can occur several times but to reduce complexity, agents are used to highlight dependencies among different kinds of nodes. To further reduce complexity nodes of minor importance have been left out to make the information easier to handle. Also, some dependencies among the presented nodes have been left out that would be included in a real analysis. The reason for this to highlight the functionality of the mapping method and how it can be applied to a system.

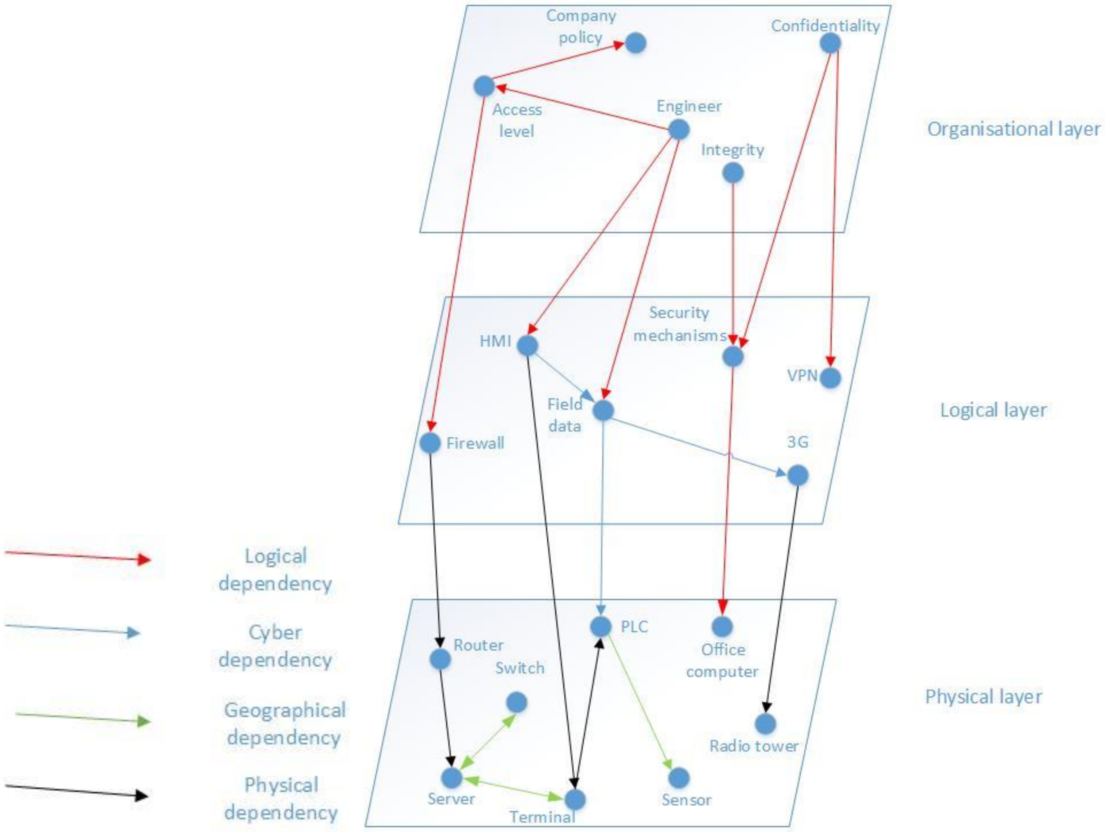


Figure 14 Three-layer model of the test system

The dependencies in Figure 14 are only of the first order, this is to reduce the complexity of the figure. Following the earlier definition, if two components C and Y fit in the sentence “X interacts with Y” or “X requires Y” then there is a dependency should be modeled.

4.2 SecuriCad

The map that is created with SecuriCad aims to illustrate the test system from a logical perspective and the flow perspective as described in section 2.3.2 since physical entities can be identified. Figure 15 shows the test system modelled in SecuriCad. As with the three-layered model, agents are used to represent parts where there are a high number of similar devices. This is the case in e.g. the office network where there are a lot of similar workstations connected. Further it is worth noting that it can be shown if a router executes a firewall and if a client on a host is a root user or not. The high level of detail is also apparent since it is possible to model how a user interacts with a host with his/her user account and the following access control.

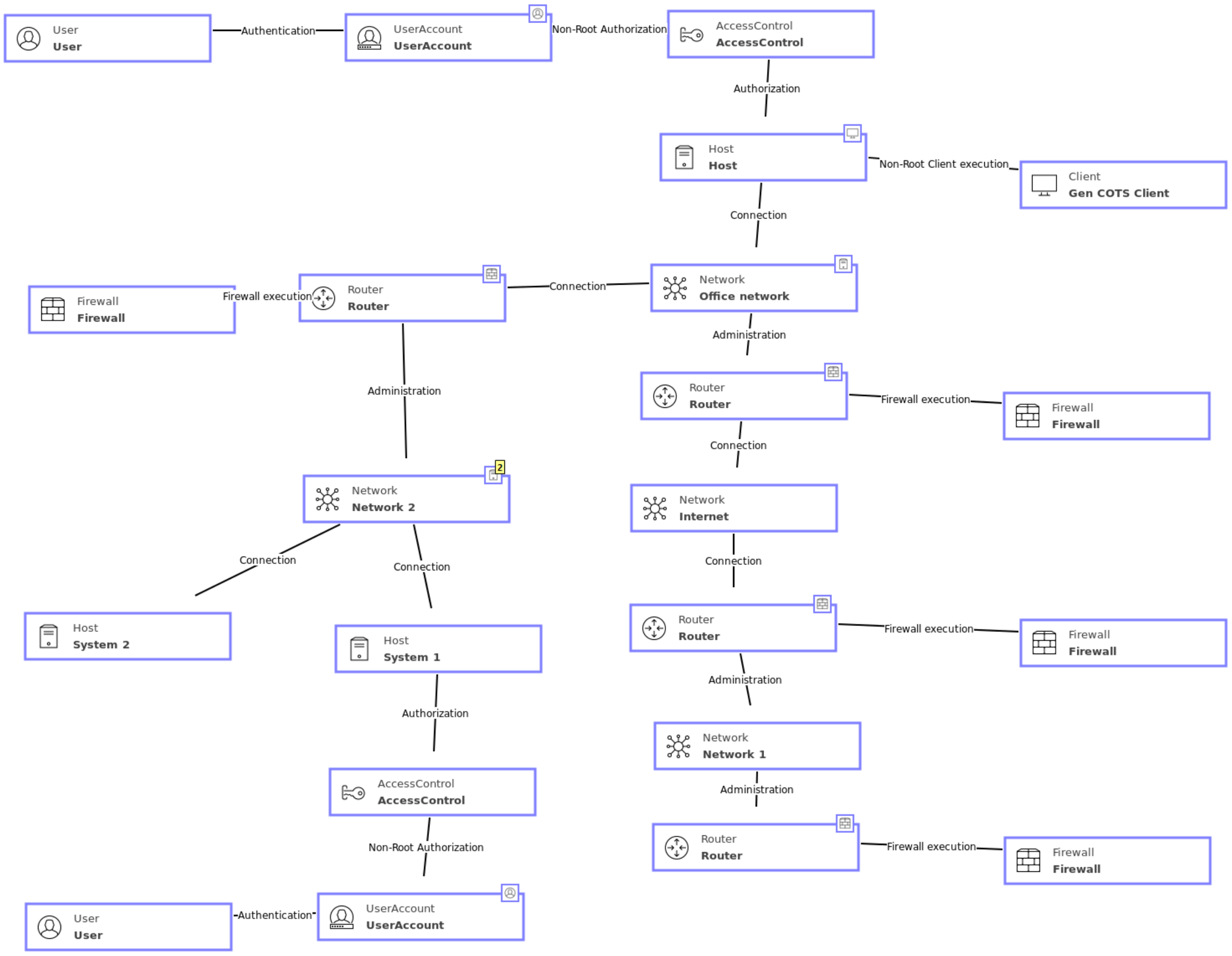


Figure 15 The test system with SecuriCad

What is harder to see in SecuriCad compared to the three-layered model is the kind of dependency between nodes. When working with SecuriCad you are limited to the kinds of nodes and connections that are available. This means that some nodes in the three-layered model cannot be added to this model.

4.3 Dependency analysis

This section will describe the selected dimensions (types of interdependency, type of failure, coupling) can provide information about system dependencies and how this information can be presented. The work starts with the system's topology and then the next steps are presented in chronological order.

4.3.1 Topology dependencies

The first step is to create a topology map of the system that is going to be analyzed, the topology map includes entities and connections to show how the system is connected. In the topology map nodes with similar functionality are represented as one node to make the map less complex, although the characteristics of the system are still kept. For example, the parts that are geographically wide spread and communicating with the system using 3G are modeled as a single node to illustrate how the system is structured the same applies for office workstations.

The connections in Figure 16 are based on connections that transfer information e.g. cables and wireless networks. Anything that is not a pure communication medium is considered to be a node in the network which means that anything with some kind of built-in logic will be a modeled as a node. The topology is based on the test system, both Figure 11 and Figure 12 to give a view of how a whole system can look like. The left parts are from Figure 12 and the right parts of the figure is from Figure 11. Since the functionality of each node is not concerned in this view, an undirected graph is used.

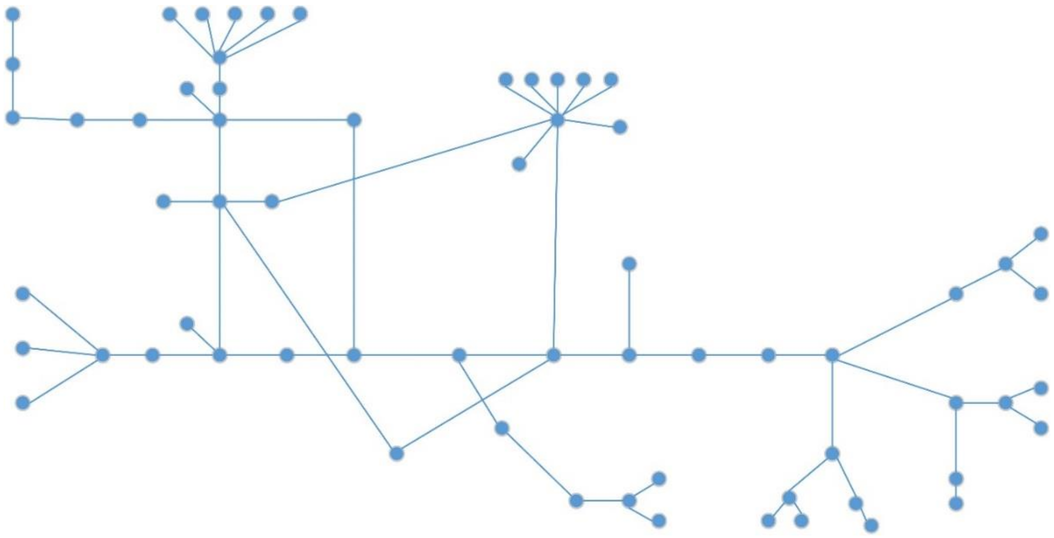


Figure 16 Test system topology

The next step is to calculate the node degree for each of the nodes in the network to see how many dependencies each node handles. Figure 17 shows that the number of dependencies per node is rather similar in most cases with just a few exceptions where a node handles more connections.

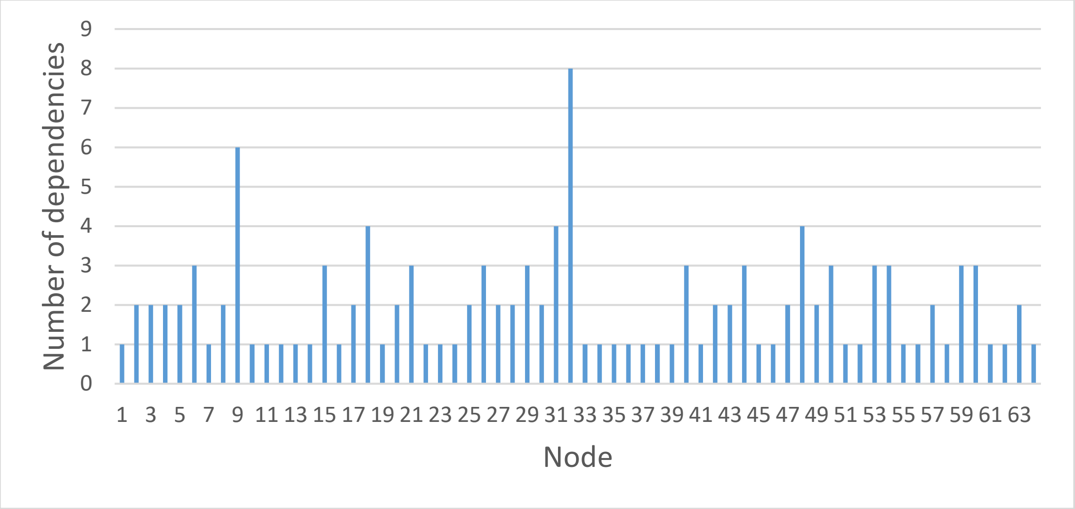


Figure 17 Number of dependencies per node

The next step is to show how the number of dependencies are distributed among the nodes. Figure 18 shows that the majority of the nodes have one or two dependencies and that there are just two nodes that have five dependencies or more. Nodes that have a lot of dependencies are often essential to the system and if they fail there will be severe consequences. However, it is hard to only have nodes with few connections which means that as long that there is knowledge about which nodes that have many dependencies and plans on how to handle failure the system can still have good and long-lasting availability.

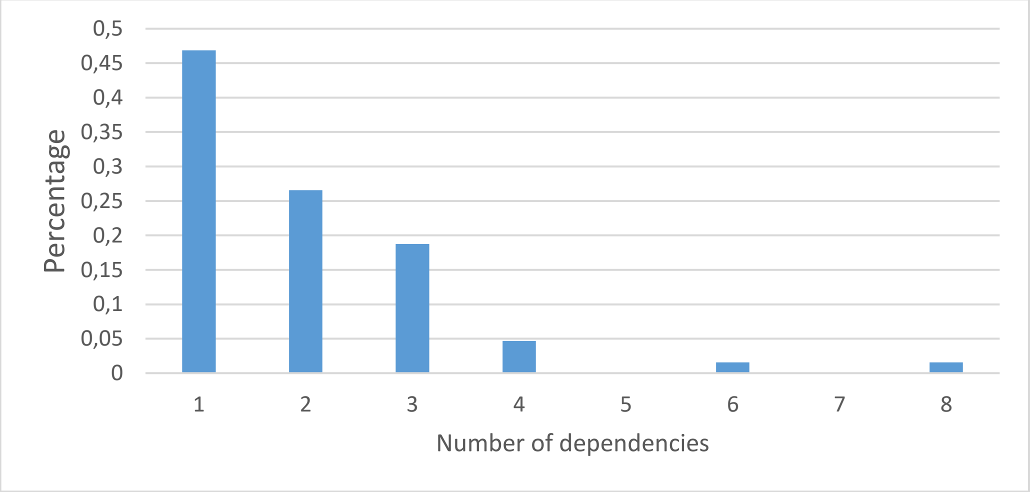


Figure 18 The distribution of number of dependencies

The topology map in Figure 16, can be used as a basis for identifying which category that each dependency belongs to. Each of the connections in the map can be transferred into one of the earlier presented dependency categories to show how different nodes depend on each other. To illustrate the different dependencies a color representation will be used for each kind of dependency as shown in Figure 19. In the figure, it can be seen that similar dependencies have a tendency to gather together e.g. in the office network there are mostly logical and cyber dependencies while among nodes on remote locations dependencies of physical and geographical kind are more common.

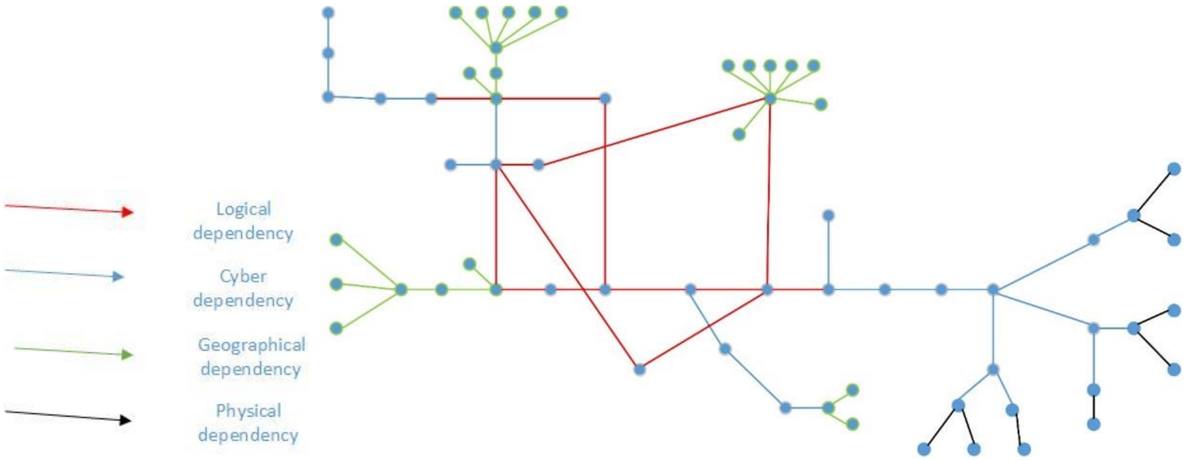


Figure 19 Topology with different dependency categories

4.3.2 Higher order dependencies

So far only dependencies of first order have been in focus, the next step is to focus more on dependencies of higher order. Dependencies of higher order can essentially be seen as chains of first order dependencies which means that you have to select one node at a time for which the analysis will be made. In Figure 20 the node *Engineer* is chosen as start node to exemplify what dependencies of higher order are related to this node. The node is located at the organizational layer in the three-layered model and depends on several system parts to be able to function as normal. The dependency categorization is the same as in previous pictures i.e. a red arrow is a logical dependency, a blue arrow is a cyber dependency, a green arrow is a geographical dependency and a black arrow is physical dependency.

By using the connections from the three-layered model, several chains starting from *Engineer* can be found and following these chains several nodes that the engineer is dependent on can be found. The color of the arrow still represents the kind of dependency that is between the nodes. In this perspective interdependencies among nodes are not included, this is because the purpose is to find all sub-nodes that the root node is dependent on and the sub-nodes characteristics. The approach when developing this kind of tree structure should be, as described in Section 3.2.1, to start with the question, what does the engineer needs in order to perform his/her assignment, and then ask this question again for the next level of nodes. The process continues until the desired level of abstraction is reached.

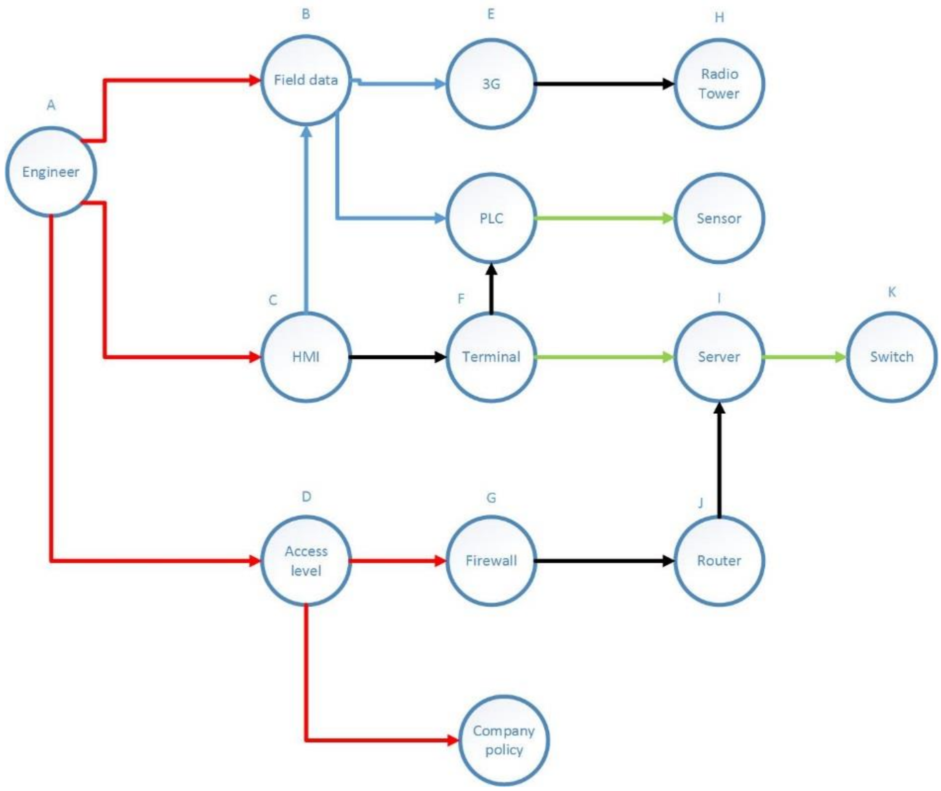


Figure 20 Higher order dependencies

In Figure 20 it can be seen that the engineer node depends on several other nodes, both direct and indirect, to function as intended. Further it can be seen that the nodes *Switch* and *Radio Tower* are both physical entities and are so called sink nodes, which means that they have no outgoing dependencies and thereby will be endpoints for cascading and escalating failures within the system.

The next step is to calculate the cumulative dependency risk for the engineer node by analyzing the three main dependency chains. To show how the formula can be used, it is applied to some of the nodes from Figure 19. The resulting formula is the following, the subscripts correspond to node sequences in Figure 19:

$$DR_A = L_{A,B} \cdot I_{A,B} + L_{A,C} \cdot I_{A,C} + L_{A,D} \cdot I_{A,D} + L_{A,B,E} \cdot I_{B,E} + L_{A,C,F} \cdot I_{C,F} + L_{A,D,G} \cdot I_{D,G} + L_{A,B,E,H} \cdot I_{E,H} + L_{A,C,F,I} \cdot I_{F,I} + L_{A,D,G,J} \cdot I_{G,J} + L_{A,C,F,I,K} \cdot I_{I,K}$$

L is the likelihood that a node will failure due to its dependency and I is the impact that the failure will have. The likelihood and impact are defined in percentage of how much a failed node will affect its connected nodes i.e. on a scale from 0-1.

By inserting the percentage into the formula, the outcome will be a number between 0 and 1 representing the cumulative dependency risk for the engineer node. Given a predefined impact, a higher value means that there is a greater risk and a lower means a minor risk. As claimed in the description of the formula, what is to be considered as the acceptable threshold should be determined by security experts. When used in a real-world scenario, the impact and likelihood of each node needs to be assessed by system experts.

4.3.3 Coupling

To make the results less complex, a higher level of abstraction is used when analyzing coupling. The system is spilt up into its different underlying units represented as agents, that each have its own special purpose to the system's and its overall functionality. The following are the identified units that the system mainly is composed of:

- Sensors in the field
- Mobile communication e.g. 3G
- External connections
- Office IT-infrastructure
- Operations infrastructure
- Users

Figure 21 shows the identified main units within the system, which all are crucial for the system's performance. The coupling is performed at unit level since the characteristics of the different units are easier to identify and related to their influence on other units.

The numbers attached to each arrow in Figure 21 represent the influence that one unit has on the pointed unit. The influence is graded on a scale from 1-5 where 1 is less influence and 5 is much influence. Influence is the same as coupling, if two units have high influence on each other they are tightly coupled. The influence values were gathered during the interviews.

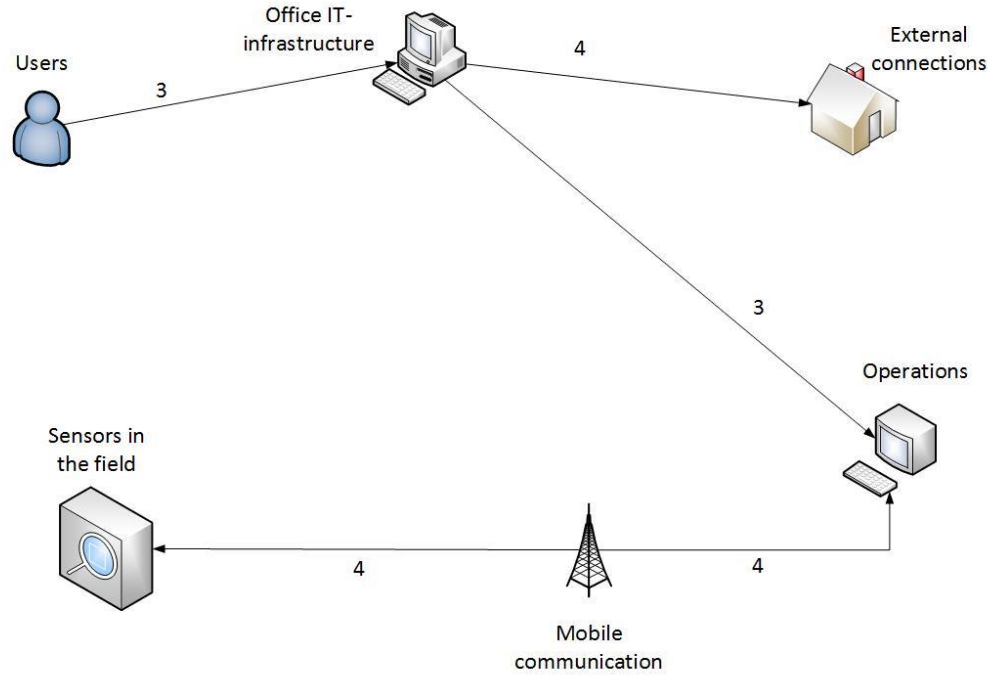


Figure 21 Influence among system parts

The matrix A represents the internal influences among the components while matrix C is the external perturbation. This means that values in A are considered as your own domain and can be changed by internal changes. Values in C on the other hand comes from external actors and can therefore not be controlled. With these values the independence level *x,* of each unit can be calculated. The letter T on the final matrix show that the matrix is transposed.

$$x = Ax + c = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 3 & 4 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 3 \\ 1 \\ 1 \\ 4 \\ 2 \\ 2 \end{bmatrix}$$

$$x = Sc = (I - A)^{-1}\,\mathrm{c}$$

$$x = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 4 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 12 & 0 & 4 & 0 & 1 & 0 \\ 9 & 0 & 3 & 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 1 \\ 4 \\ 2 \\ 2 \end{bmatrix}$$

$$x = [3 \quad 17 \quad 10 \quad 4 \quad 42 \quad 48]^{T}$$

The matrix displays that external connections and mobile communication are the units that are subject to external interference and the final value of x say that operation infrastructure are the nodes that have the highest interoperability level since it has the highest number.

The earlier identified main units are added to the model in Figure 22 to show where they are on the scale of coupling and kind of interaction.

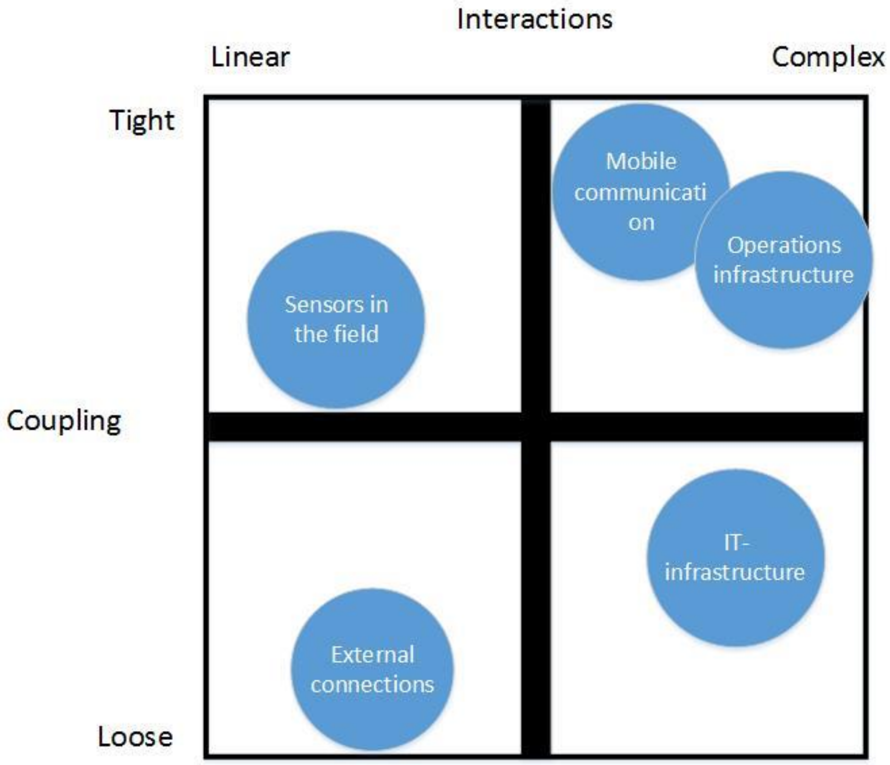


Figure 22 coupling and interactions within units

In the diagram, each of the main units that together make the whole system has been added based on their inner interactions, either complex or linear and whether their underlying coupling is either tight or loose. Consider the IT-infrastructure node as an example. The IT-infrastructure consists of computers, Wi-Fi, printers etc. that are most often used at an office. The coupling among these different kinds of components are quite loose since most components' functionality is not crucial to the system. However, the IT-infrastructure often consist of a high number of devices, systems, connections etc. which can make the total picture about interactions within the system rather complex.

4.4 Evaluation

Since the interviewed experts have different perspectives of the system they can evaluate the material from different point of views. The following diagrams show how the subjects responded to questions where they were asked to map their opinion on a scale from 1 to 5, in all cases a 1 was the worst and a 5 was the best.

4.4.1 Evaluation of system overview

The first question was to rank how informative they thought that each model was from their perspective. The answers were distributed as follows:

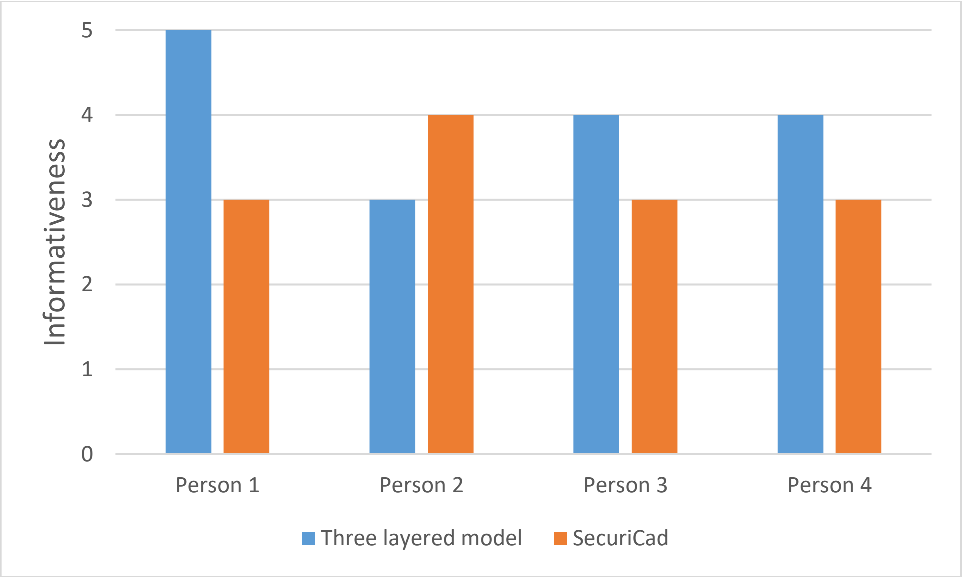


Figure 23 A ranking on how informative each model is.

Figure 23 shows that both models in general got good scores. In three out of four people thought that the three-layer model were more informative. In three out of four cases the difference was only one step between the models.

The next question was regarding how easy they felt it was to understand the models. Each subject was given a short period of time to process each model before they needed to answer. Also in this case the interview subject was told to rank the understandability on a scale from 1 to 5 and the answers were distributed as follows.

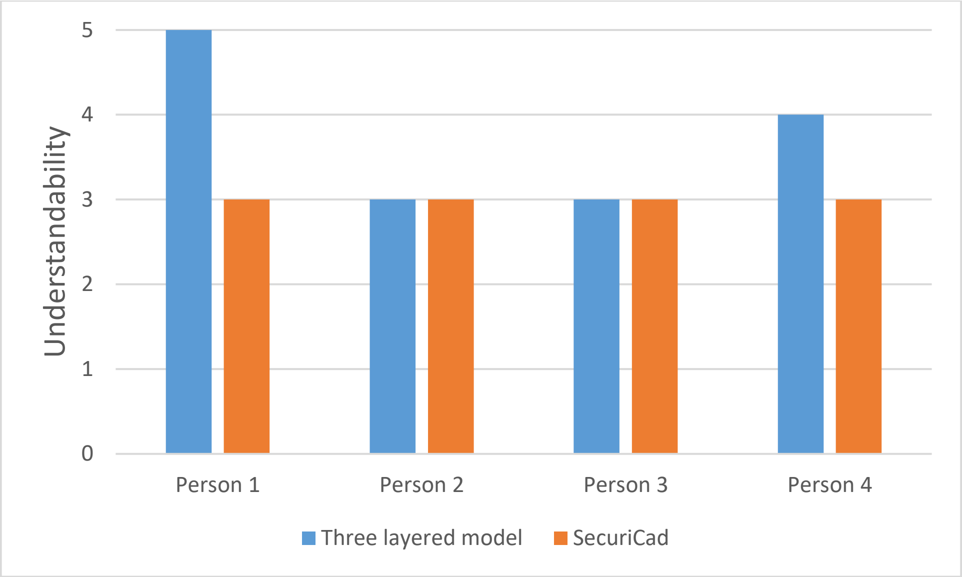


Figure 24 A ranking on how easy it is to understand the models.

Also in this case, the three-layer model gets a better result but the total difference is marginal. Two subjects felt that the two models were equally easy to understand.

The third question aimed to investigate how likely it was that they in some way would use any of the models, either themselves or to suggest it to others, in a future scenario. The question concerns both models since they both can be a part of the complete methodology. The answers were as follows:

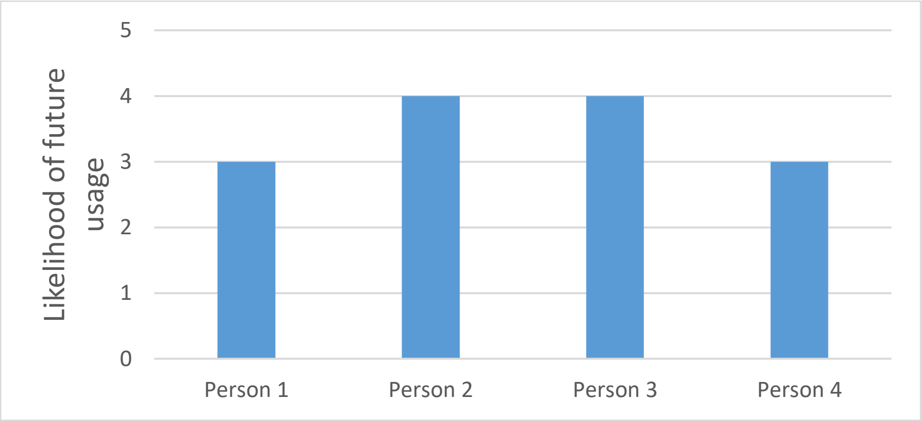


Figure 25 A ranking on how likely it is that they use any model in the future

Finally, they had a chance to bring up general comments that they wanted to share about the material, both positive and negative. The most common comments were that the three-layer model gave a good overview of the system and that it was fairly easy to understand what the models showed despite that most of the interview subjects were not familiar with the model before the interview. Other comments were that SecuriCad felt more familiar from the start since it is more closely related to previous ways to model a system and that it could incorporate users in a good way.

4.4.2 Evaluation of dependency methods

The next part concerned dependencies within a system, where the aim was to give some structure to the term dependency as well as showing how dependencies can be analyzed. The ability to understand the material was highly correlated to how much prior knowledge they had about the different perspectives. The same procedure as in the first part was also applied to this part i.e. a short presentation followed by some time to let the information sink in. Prior to this part, the categorization of dependencies was presented along with a short description about what each category was intended to capture and why some dependencies were categorized in a particular way.

The first question concerned how informative they felt that each dependency perspective was (topology, dependencies of higher order and coupling within the system). The interview subjects were asked to rank each perspective on a scale from 1 to 5 representing their opinion. The answers were distributed as follows:

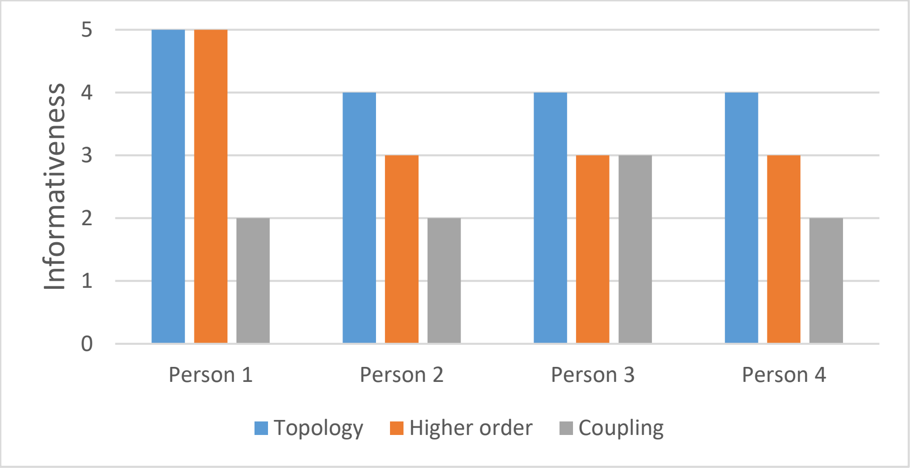


Figure 26 A ranking on how informative each perspective is.

In general, the topology is ranked as the most informative followed by higher order and then coupling. All subjects have a very similar view of which perspective they think is the most and least informative, they only differ on at what level they rank them on.

The second question was regarding how easy they felt that it was to understand the content of the data that were presented within each perspective. Also in this case the subjects were asked to rank the understandability of each question on a scale from 1 to 5. The answers were distributed as follows:

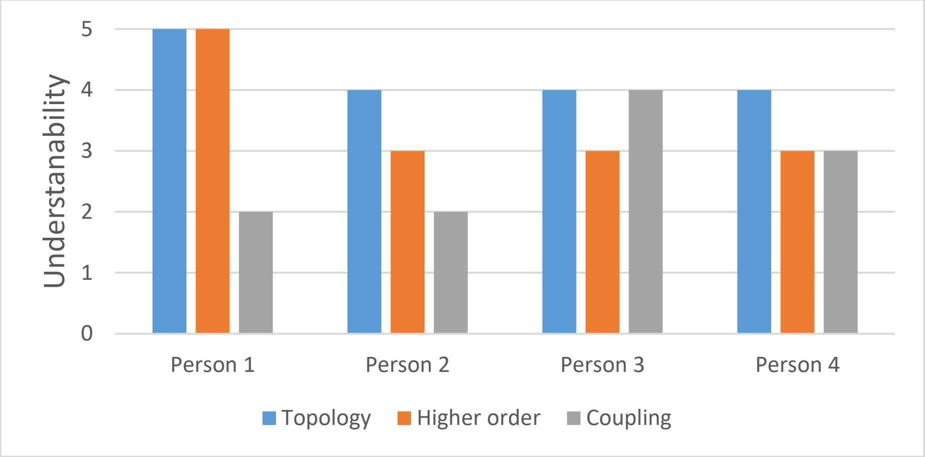


Figure 27 A ranking on how easy it is to understand each perspective

Once again topology has in general the highest score, the other two perspectives vary more from subject to subject.

The third question was regarding the dependency categorization and how they felt about the relevance of the dependency categorization in this type of context.

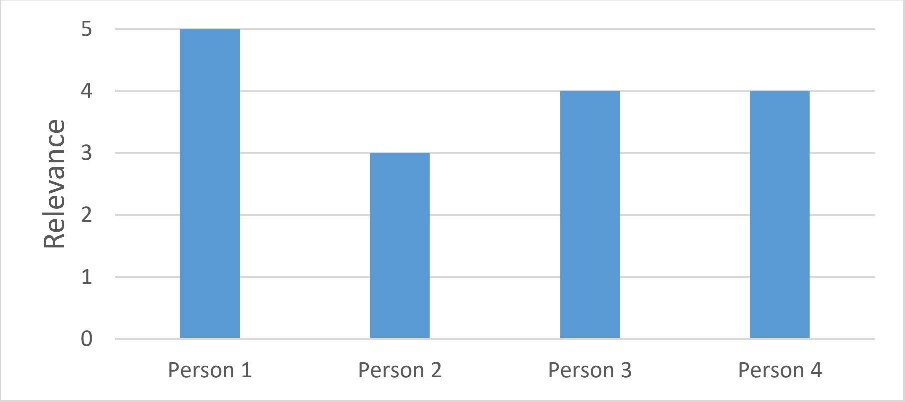


Figure 28 A ranking on how relevant the dependency categorization

Figure 28 shows that the categorization seems to be relevant, three out of four subjects give it a four or higher and no one gives it lower than a three. The overall results were that the dependency categorization got approved by all the subjects. The general opinion was that the geographical and physical category were fairly easy to understand while it was harder to separate cyber from logical dependencies. One other comment was that the geographical category felt less interesting than the others in terms of a risk analysis.

The fourth question was the same as in the system overview evaluation i.e. if they could see themselves in any way make use of the material in the future and if they could how likely it would be that they would use it.

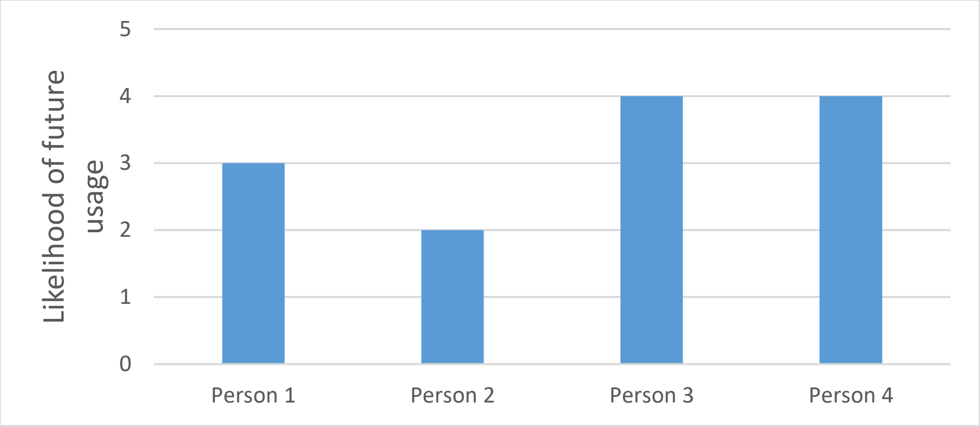


Figure 29 A ranking on how likely it is that they use material in the future

The results show that the likelihood of future use varies a bit, two subjects rank it as a four on a scale to five that they would use it while another subject rank it as a two on the same scale.

The final question concerned any other general comments that they would like to share about the material, any kind of feedback was greatly appreciated. One comment that was made by all subjects was that the combination of all three perspectives together provided a very good overview of the system and therefore could be part of a future methodology to map out systems.

4.4.3 General opinions

The third part of the interview was based on three questions that concerned all of the earlier material and was intended to collect the general opinion about the material. The first question was if they could see that any of the presented material could be used in the context of a risk analysis and if so, what part and at what stage of the risk analysis could the material be helpful. All subjects answered yes to this question and the general motivation was that it was a good way to provide a structure on how to present systems and that it therefore made it easier to present information about the system to people in a position to make decisions that often do not have the same level of technical knowledge as the people that works with the system. For example, the dependency risk formula in the higher order perspective was pointed out as a good way to translate a subjective opinion to an actual number that is easier to understand and compare.

The second question was if they, based on all the material, felt that they got a good view of the test system and its high-level characteristics. Most subjects answered that they got a very good overview of the system but some comments were made saying that a more detailed level of information would have been useful. Everyone pointed out that structure was very important and that this material provided structure in a good way.

And finally, there was an opportunity for general comments about the work in general for example if they missed anything or if they felt that some parts were more interesting than others. The comments in general were that the material could be good as a framework for mapping out a system and that it could ease the information sharing within a company where people have different levels of technical knowledge. The overall impression from the interview subjects were very positive and they felt that the structure provided by the work in this thesis could be very useful in the future.

5 Discussion and conclusion

This section will provide a discussion of the work carried out to make this thesis, this includes the methodological choices, the results and the validity of the results that are provided. Then, the final conclusions of this thesis are presented and how the results relate to the research questions and the intended scope of the thesis. Finally, some ways of future work will be presented.

5.1 Discussion

During the different parts of conducting this thesis a number of choices had to be made that potentially could affect the results. This section will bring up some aspects and discuss them in relation to the thesis work.

5.1.1 Data gathering

The data gathering about the system was carried out through interviews which opens up the possibility that the data reflects the opinions of the interviewed subjects, although this do not really affect the results in this thesis since the goal was not to evaluate the actual system. The main point was to show how and why interviews are a suitable method for data gathering in this scenario. The test system was simply used to illustrate how the different perspectives could be used following the interviews. Further, all interview subjects stated that this work could be of use to their department and therefore the subject has no reason to give false data.

The system is presented on a high level which makes interviews suitable as the method for data gathering. If more detailed properties would have been needed, interviews would be less suitable since it is very time consuming to go through the details of different system components.

5.1.2 System modelling

The system modeling consisted of two different models: the three-layer model and SecuriCad. To show a system based on three layers have been proposed several times before and visualization of the layers with a map consisting of the three layers has been used before but in other situations. Therefore there exists literature that can provide guidelines on how to work with these layers, but since the models in the literature are implemented in different scenarios the guidelines needed to be translated to fit the purpose of this thesis. Since the three-layer model got a majority of positive comments, the translation and modeling can be considered correct enough.

SecuriCad was originally created to make risk analyses of SCADA systems and the tool has been updated with new and improved features along the way. SecuriCad was chosen because its original focus on SCADA systems and that it therefore has capabilities to model these kinds of systems in an accurate way. During the interviews, all subjects stated that they understood the model and how users could be incorporated which provides validity to the tool's ability to model a system and that it can fit as a complement to the three-layer model.

The first choice in dependency modeling was to define what is meant by a dependency and what kinds of dependencies there are. From a literature review the conclusions were that any kind of interaction should be seen as a dependency. This was approved during evaluation in most cases, one counter opinion was that the term dependency should have a more functional focus and only be used in terms of what core parts the system’s functionality depends on. Everything else is not dependencies but extra functionality to make operating the system easier.

With the basis that any kind of interaction is a dependency there is a need to separate different kinds of dependencies. The chosen set of dependencies is proved to have the best coverage, although this is the result of a very broad definition of the logical category. The creators of the chosen set have specified what should be in the physical, geographical and cyber category and then states that everything else that does not fit in any of the first categories should be a logical dependency. In a future scenario this may lead to confusion about what to call a logical dependency but this was not commented during the evaluation. To reduce the scope of the logical dependency, social dependency could be added aimed at representing human dependencies. This alternative was excluded since the social category did not seemed to be widely used.

Out of the six dimensions describing a dependency, three of these were chosen to get extra focus in this thesis. During the evaluation all of these perspectives were described as interesting, to analyze the number of connections at different nodes were often commented as the most interesting perspective. This was because the number of connections can show which nodes that might be essential to the system and it was also commented that you might set a maximum number of connections that you allow system nodes to have. The importance of modeling dependencies of higher order is often stated in literature and agreed by interview subjects. The tree that was created to illustrate dependencies of higher order were, after some time to process it, a good way to show these kinds of dependencies. The formula to translate subjective opinions to an actual number got mostly positive reactions. The final part was to analyze coupling among system units, all of the interview subjects were unfamiliar with the term coupling and therefore they were all provided with a short explanation and opportunity to ask questions about it to make sure they understood the intended goal with the material. This part was the part they felt the least informative to the subjects which might be a result of them being unfamiliar with coupling while at the same time it can be rather complex to analyze. However, all subjects stated that they understood the intended scope of coupling but that they just not felt that it was as useful as the other perspectives.

5.1.3 Results

The evaluation interviews were very informative, the interviews provided a lot of valuable knowledge. All of the interviewed subjects had some kind of relation to the work in this thesis and therefore some prior knowledge about the domain, risk analysis, SCADA systems or system dependencies and therefore could provide feedback with good validity. This was also the reason for only doing four evaluation interviews as, mentioned in Section 3.1.1, since a decision was made that prior knowledge was more important than to get a higher number of interview subjects.

From the evaluation of SecuriCad and the three-layered model it can be seen that no one gave a score lower than three. There can be several reasons for this, for example that they did not think that any model deserved less than three. Another option is that they felt uncertain about the model and then just said three as a middle way.

One thing that was noted during the evaluation interviews was that the questions regarding how informative and how easy a view was to understand, seemed to be strongly connected. The reason might be that it can be hard to assess how informative something is if it is hard to understand. Therefore, a case where something is very informative but requires more time to get comfortable with might not appear in this evaluation since evaluation interview had to be kept at reasonable length.

Topology is a very common way to show a system and therefore people have easier to understand the information it presents. As stated by Fioriti et al. [17] the topology alone can provide much information about a system's characteristics. This fact was confirmed during the interviews were the topology map was often highlighted as the most informative view and the one that was easiest to understand.

A topic that was brought up during evaluation was how to move forward after completing this methodology? As stated in the introduction the aim of this thesis is to provide a view of the system and thereby ease future risk assessments. Therefore the concern on how to move forward is outside the scope of this thesis, but it can still be said that the presented steps just show the current situation and therefore need to be updated when the system changes in order to still be useful. The problem of keeping the information updated is not within the scope of this thesis but is an important step following this work.

Something that was pointed out by all interview subjects was that the methodology could be useful to make information understandable by people in different positions with in a company that also have different levels of technical knowledge. Both the formula used to calculate dependency risk from higher order dependencies and the formula used to calculate interference from coupling can be used to provide a magnitude that can be easily comprehended by people without expert knowledge of the system. This is an important aspect in terms of risk analysis since people in management positions rarely have expert knowledge of the system but still need to know the systems situation.

In terms of the methodology being holistic, all interview subjects said that they thought they got a good overview of the system.

5.1.4 Validity and reliability

The results are very coherent and therefore the results and conclusions will have good reliability. Although there is a weakness in having too few evaluation interviews who thereby do not providing accurate feedback or missing important aspects. The interview subjects were chosen based on two criteria, their knowledge about networks and their involvement in risk analyses. If other subjects, with less knowledge, would have been chosen that have less knowledge there is a risk that either the interview will have too much focus on just understanding the material and thereby take too much time, or that the subject understands the material but do not have knowledge about how to apply it in a work situation. An alternative to interviews could have been to do a survey but since the material cannot be considered as common knowledge it might need some explanation which cannot be provided by a survey, this alternative was rejected.

To reduce the bias from the interviewer during the evaluation all interview subjects got the same introduction to the work followed by an explanation that aimed to be as similar as possible. However, there is still a risk that the answers are biased to some extent since there was only one person doing the interviews which makes it hard to detect biased answers. Although the answers are still comparable to each other since the interviews were held individually and it was the same person doing all of the interviews.

Important to note is that only people working at Tekniska verken were interviewed during evaluation which can affect the generalizability of this conclusion, although Tekniska verken has no special properties that should affect the interviewed people and their opinions any significant way.

5.2 Conclusion

The aim of this thesis has been to investigate how systems can be modeled and how dependencies between nodes in critical infrastructure can be mapped out. The work is intended to lay the ground for future risk assessments in the sense that to be able to make correct assessments you need to have a good overview of the system. To capture the system's characteristics, not only computers and network need to be considered, human and organizational factors also play an important role. As a way to include human and organizational factors the three-layer model and SecuriCad was evaluated on their applicability to model SCADA systems, combined with the fact that human interaction was added to the logical dependency category.

The results show that the three-layer model is considered as being both the very informative and rather easy to understand, the most appreciated aspect is its ability to model a system on a high level that does not require deep technical knowledge. Regarding SecuriCad, the most appreciated aspect was that the model felt familiar and was therefore easy to understand. Something that also was positive with SecuriCad was the structure it could provide when modeling a system. The suggested perspectives of dependency analysis all have properties that would be useful when mapping out a system. Furthermore, the perspectives are different when it comes to how well known they are, which can affect how appreciated they are. However, all perspectives were considered to be informative to some extent and can serve different purposes in the future.

Even though dependency is a very broad term, the results show that the structure and guidelines provided in this work are much appreciated since there is no common way to work with dependencies at this level.

The results can be related to the research questions presented in Chapter 1 as follows:

Can the top-down approach of the layered model be combined with the bottom-up approach of SecuriCad?
Yes, overall both models got good scores on both informativeness and understandability. It was pointed out during evaluation, that thanks to their different approach to system mapping, they can complement each other in a useful way. For example, one scenario could be to first use the three-layer model to provide a high-level view and then use SecuriCad to model specific areas on a more detailed level. By doing this, both the top-down perspective the three-layered model and the bottom-up perspective of SecuriCad is included.

One a more detailed level, the most appreciated aspect of the three-layer model that was pointed out during evaluation, was its ability to model system dependencies on a high level. The high level makes it a suitable tool to share information among persons with different levels of technical knowledge. Further, it can be concluded that all interview subjects stated that they could see themselves use it or promote the use of the three-layer model in the future, which further contributes to the fact that the three-layer model is suitable for mapping SCADA systems.

Regarding SecuriCad, the most appreciated aspect was the structure that you get and also that it shows a system in way that is rather common and therefore easier to understand. By including SecuriCad you also have the possibility to use it for simulating the security level of the system.

What suitable methods exist for dependency analysis and how can they be combined with different system models?
The suggested mapping method consist of a number of different methods for dependency analysis. The methods include using the dependency categorization, to investigate dependencies of higher order and to analyze coupling among system functions. The information from these methods are presented with different system models. The overall reception among the interview subjects was positive.

The dependency categorization felt useful and provided good structure. The downside was the logical category, due to its broad definition. To analyze dependencies of higher order was considered to provide a lot of valuable information. This information was presented using a tree-structure that was considered useful in this scenario. The tree-structure also felt intuitive to use in this scenario and was therefore easy to understand. The information on coupling among system functions was considered to be the information that was the hardest to understand. However, after some time to process it, the subjects stated that this also was very valuable information. The downside was that, since this was an unfamiliar expression it was hard to understand the presented information.

Is there any evidence that knowledge about system dependencies are helpful for future risk assessments?
Yes, based on the evaluation of the suggested perspectives with their respective tools on how they could aid a future risk assessment. It was pointed out several times that each perspective had interesting properties that in some way each could be useful prior to a risk assessment. Therefore, it can be concluded that an analysis of system dependencies can aid a future risk assessment. However, nothing can be said about to what extent they can aid the process, this would require a comparison of risk analyses performed both with and without a prior dependency analysis.

Something that was stressed during the evaluation was that all three selected perspectives together gave a good understanding of the system on a high level and thereby are well suited as a preparatory step of a risk assessment. This agrees well with the goal that all perspectives together can make a whole methodology for dependency analysis.

It can also be concluded that the three dimensions that were chosen (type of interdependency, type of failure and coupling and response behavior) based on their applicability to risk assessments served their purpose well. All interview subjects stated that they thought each perspective was useful in terms of a risk analysis. However, they were not introduced to three perspectives that had been excluded (infrastructure characteristics, state of operation and environment). Still at least no one mentioned that they missed any kind of information.

5.3 Future work

The intention with using the three-layer model and SecuriCad is that they have capabilities to include users in the system description. Both these models add users as a node that interacts with other nodes. Another option is to use the social dependency categorization presented by De Porcellinis et al. [26] which instead keeps nodes to just being physical entities and can show users impact by modeling a social dependency between nodes. By implementing this approach instead, the benefits could be maps that have a fewer number of nodes and also nodes that are more similar in their capabilities. These properties can in turn make the final result easier to comprehend.

Since only three perspectives were used to investigate if dependency analysis can aid risk assessments nothing can be said about how well they work. In order to address this there is a need to compare the perspectives to other approaches and evaluate their usefulness.

This work has only been evaluated with people from Tekniska verken, to further evaluate its applicability and to gather more input on the proposed methodology similar work needs to be created and evaluated at other companies. To further understand how to aid risk assessments, companies that are more specialized on risk assessments can be consulted to gather input from another perspective.

6 References

[1] M. Naedele, "Addressing IT Security for Critical Control Systems," in *40th Annual Hawaii International Conference on System Sciences, 2007*, Waikoloa, 2007. IEEE.

[2] ”NIST 800-30 Guide for conducting Risk assessment,” National institue of standards and technology, 2012.

[3] A. Laugé, J. Hernantes och J. M. Sarriegi, ”Critical infrastructure dependencies: A holistic, dynamic and quantitative approach,” *International Journal of Critical Infrastructure Protection,* vol. 8, pp. 16-23, 2015. Elsevier.

[4] P. S. M. Pires and L. A. H. g. Oliveira, "Security Aspects of SCADA and Corporate Network Interconnection: An Overview," in *IEEE International Conference on Dependability of Computer Systems*, Szklarska Poreba, 2006. IEEE.

[5] P. G. Neumann, ”Practical architectures for survivable systems and networks,” http://www.csl.sri.com/users/neumann/survivability.pdf, 2000.

[6] C. Blackwell, "A multi-layered security architecture for modelling complex systems," in *Proceedings of the 4th Annual Workshop on Cyber Security and Information Intelligence Research: Developing Strategies to Meet the Cyber Security and Information Intelligence Challenges Ahead*, Oak Ridge, Tennessee, USA, ACM, 2008, pp. 35:1--35:4.

[7] H. Holm, K. Shahzad, M. Buschle och M. Ekstedt, ”P^2CySeMoL: Predictive, Probabilistic Cyber Security Modeling Language,” *IEEE Transactions on Dependable and Secure Computing,* vol. 12, nr 6, pp. 626-639, 2015. IEEE.

[8] S. M. Rinaldi, J. P. Peerenboom och T. K. Kelly, ”Identifying, understanding, and analyzing critical infrastructure interdependencies,” *IEEE Control Systems,* vol. 21, nr 6, pp. 11-25, Dec 2001. IEEE.

[9] E. Zio och G. Sansavini, ”Modeling Interdependent Network Systems for Identifying Cascade-Safe Operating Margins,” *IEEE Transactions on Reliability,* vol. 60, nr 1, pp. 94-101, 2011. IEEE.

[10] C. Perrow, Normal accidents: Living with high risk technologies, Princeton University Press, 2011.

[11] V. Igure, S. Laughter och R. Williams, ”Security issues in SCADA networks,” *Computers & Security,* vol. 25, nr 7, pp. 498-506, 2006. Elsevier.

[12] E. Zio, ”Challenges in the vulnerability and risk analysis of critical infrastructures,” *Reliability Engineering & System Safety,* vol. 152, pp. 137-150, 2016. Elsevier.

[13] S. Rinaldi, "Modeling and simulating critical infrastructures and their interdependencies," in *IEEE 37th Annual Hawaii International Conference on System Sciences*, 2004. IEEE.

[14] G. Brændeland, H. E. I. Dahl, I. Engan and K. Stølen, "Using Dependent CORAS Diagrams to Analyse Mutual Dependency," in *Critical Information Infrastructures Security*, Berlin, Springer, 2007, pp. 135-148.

[15] M. Theoharidou, P. Kotzanikolaou och D. Gritzalis, ”A multi-layer Criticality Assessment methodology based on interdependencies,” *Computers & Security,* vol. 29, nr 6, pp. 643-658, 2010. Elsevier.

[16] S. D. Porcellinis, G. Oliva, S. Panzieri and R. Setola, "A Holistic-Reductionistic Approach for Modeling Interdependencies," in *IFIP Advances in Information and Communication Technology, Springer, Berlin, Heidelberg*, Berlin, 2009. Springer.

[17] V. Fioriti, G. D'Agostino och S. Bologna, ”On Modeling and Measuring Inter-dependencies among Critical Infrastructures,” *Complexity in Engineering,* pp. 85-87, 2010. IEEE.

[18] I. Eusgeld, C. Nan och S. Dietz, ”System-of-systems” approach for interdependent critical infrastructures,” *Reliability Engineering & System Safety,* vol. 96, nr 6, pp. 679-686, 2011. Elsevier.

[19] M. Ouyang, ”Review on modeling and simulation of interdependent critical infrastructure systems,” *Reliability Engineering & System Safety,* vol. 121, pp. 43-60, 2014. Elsevier.

[20] Y. Haimes, B. Horowitz, J. Lambert, J. Santos, C. LIan and K. Crowther, "Inoperability Input-Output Model for Interdependent Infrastructure Sectors. I: Theory and Methodology," *Journal of Infrastructure Systems,* vol. 11, no. 2, pp. 67-79, 2005. ASCE Library.

[21] P. Burnap, Y. Cherdantseva, A. Blyth, P. Eden, K. Jones, H. Soulsby och K. Stoddart, ”Determining and Sharing Risk Data in Distributed Interdependent Systems,” *Computer,* vol. 50, nr 4, pp. 72-79, 2017. IEEE.

[22] R. Zimmerman, ”Social implications of infrastructure network interactions,” *Journal of Urban Technology,* vol. 8, nr 3, pp. 97-119, 2001. Routledge.

[23] D. D. Dudenhoeffer, M. R. Perman and M. Milos, "CIMS: A framework for infrastructure Interdependency Modeling and analysis," in *IEEE Winter Simulation Conference*, 2006. IEEE.

[24] E. E. Lee, J. E. Mitchell och W. A. Wallace, ”Restoration of services in interdependent infrastructure systems: A network flows approach,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 37, nr 6, pp. 1303-1317, 2007. IEEE.

[25] P. Zhang och S. Peeta, ”A generalized modeling framework to analyze interdependencies among infrastructure systems,” *Transportation Research Part B: Methodological,* vol. 45, nr 3, pp. 553-579, 2011. Elsevier.

[26] S. D. Porcellinis, R. Setola, S. Panzieri och G. Ulivi, ”Simulation of heterogeneous and interdependent critical infrastructures,” *International Journal of Critical Infrastructures,* vol. 4, nr 1-2, pp. 110-128, 2008. Inderscience Publishers.

[27] A. V. Gheorghe and M. Schlapfer, "Ubiquity of Digitalization and Risks of Interdependent Critical Infrastructures," in *IEEE International Conference on Systems, Man and Cybernetics*, Taipei, 2006. IEEE.

[28] P. Kotzanikolaou, M. Theoharidou och D. Gritzalis, ”Assessing n-order dependencies between critical infrastructures,” *International Journal of Critical Infrastructures,* vol. 9, nr 1-2, pp. 93-110, 2013. Inderscience Publishers.

[29] A. Creery och E. J. Byres, ”Industrial cybersecurity for power system and SCADA networks,” i *Record of Conference Papers Industry Applications Society 52nd Annual Petroleum and Chemical Industry Conference*, 2005. IEEE.

[30] A. D. Nicola, M. L. Villani, M. C. Brugnoli och G. D'Agostino, ”A methodology for modeling and measuring interdependencies of information and communications systems used for public administration and eGovernment services,” *International Journal of Critical Infrastructure Protection,* vol. 14, pp. 18-27, 2016. Elsevier.

[31] C. Nan och G. Sansavini, ”Multilayer hybrid modeling framework for the performance assessment of interdependent critical infrastructures,” *International Journal of Critical Infrastructure Protection,* vol. 10, pp. 18-33, 2015. Elsevier.

[32] J. Johansson och H. Hassel, ”An approach for modelling interdependent infrastructures in the context of vulnerability analysis,” *Reliability Engineering & System Safety,* vol. 95, nr 12, pp. 1335-1344, 2010. Elsevier.

7 Appendix A

Interview guide – Data gathering

1. Visa din bild av nätverket med en skiss, så detaljerat som möjligt.
   a. Beskriv topologi
   b. Beskriv geografisk plats
   c. Vilka delar har ni själva ansvar för?
2. För alla kopplingar, vilken typ av koppling är det?
   a. Fiber
   b. 3g/4g
   c. Gsm/gprs
   d. Wi-Fi
   e. Annat
3. Finns det andra typer av kopplingar?
   - □ VPN
   - □ SSH
   - □ Leverantörsingångar
   - □ Andra liknande typer

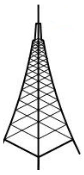
4. Beskriv systemets funktion och vad olika delars uppgift är.

5. Kategorisera vilka delar av systemet som är mest kritiska på en skala 1-5 (1 – mindre viktigt, kan hantera ett längre avbrott, 5 – direkt avgörande, kan inte leverera tjänsten)

6. Uppskatta hur känsligt för störningar systemets olika delar är på en skala 1-5 5 (1 – mindre känsligt, 5 – mycket känsligt)

Övriga Kommentarer/Information som kan bidra till kartläggningen?

8 Appendix B

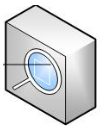


Radio tower

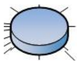
Firewall

Hub

PLC



Sensor



Network

User

Internet

Workstation

9 Appendix C

Interview guide – Evaluation

Utvärdering PCySeMoL och ”tre lager modellen”

Börja med en kort genomgång av materialet

1. På en skala 1-5 hur informativ är respektive modell?
2. På en skala 1-5 hur lätt är det att förstå resultatet?
3. På en skala 1-5 hur troligt är det att du skulle använda detta material i framtiden?
4. Övriga kommentarer?

a. Fördelar?

b. Nackdelar?

Utvärdering beroenden

Börja med en kort genomgång av materialet

1. På en skala 1-5 hur informativ är varje del?
2. På en skala 1-5 hur lätt är det att förstå resultatet?
3. På en skala 1-5 hur relevant känns kategoriseringen av beroenden?
4. På en skala 1-5 hur troligt är det att du skulle använda detta material i framtiden?
5. Övriga kommentarer?

a. Fördelar?

b. Nackdelar?

Slutgiltiga frågor

1. Kan detta vara användbart som en början på en framtida riskanalys?

a. Om ja, på vilket sätt?

b. Om nej, varför inte?

2. Allt material sammantaget, ger det en bra helhetsbild av systemet?
3. Övriga kommentarer?