Distinguishing Attacks and Failures in Industrial Control Systems
Knowledge-based Design of Bayesian Networks for Water Management Infrastructures

Chockalingam, S.

DOI
10.4233/uuid:17da1df4-3295-45d3-9119-9f92a547e7c6

Publication date
2020

Document Version
Final published version

Citation (APA)
Chockalingam, S. (2020). *Distinguishing Attacks and Failures in Industrial Control Systems: Knowledge-based Design of Bayesian Networks for Water Management Infrastructures*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:17da1df4-3295-45d3-9119-9f92a547e7c6

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

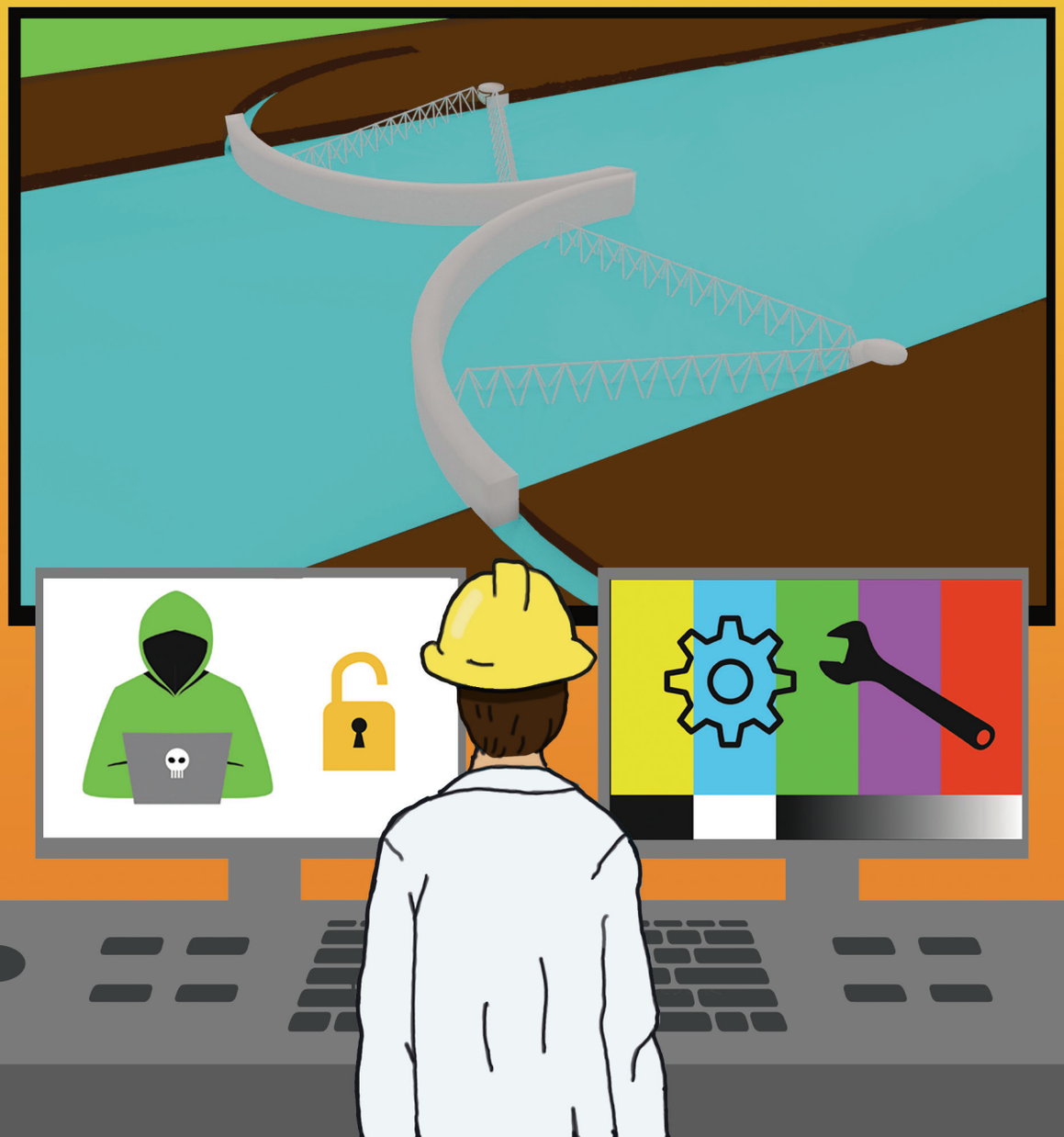
Distinguishing Attacks and Failures in Industrial Control Systems

Sabarathinam Chockalingam

Distinguishing Attacks and Failures in Industrial Control Systems:

Knowledge-based Design of Bayesian Networks for Water Management Infrastructures

Sabarathinam Chockalingam



DISTINGUISHING ATTACKS AND FAILURES IN INDUSTRIAL CONTROL SYSTEMS

KNOWLEDGE-BASED DESIGN OF BAYESIAN NETWORKS FOR WATER MANAGEMENT INFRASTRUCTURES

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Tuesday, 15 December 2020 at 10:00 o'clock

by

Sabarathinam CHOCKALINGAM

Master of Science in Cyber Security and Management, University of Warwick, England
born in Karaikudi, India.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	Chairperson
Prof. dr. ir. P.H.A.J.M. van Gelder	Delft University of Technology, Promotor
Dr. ir. W. Pieters	Delft University of Technology, Promotor
Dr. A. Teixeira	Uppsala University, Copromotor

Independent Members:

Prof. dr. ir. M. Kok	Delft University of Technology
Prof. dr. ir. J. van den Berg	Delft University of Technology
Prof. dr. M.I.A. Stoelinga	University of Twente and Radboud University
Dr. P. Smith	Austrian Institute of Technology

This research received funding from the Netherlands Organisation for Scientific Research (NWO) in the framework of the Cyber Security research program under the project *"Secure Our Safety: Building Cyber Security for Flood Management (SOS4Flood)"*.

Keywords: Bayesian network, Cyber security, Intentional attack, Knowledge elicitation, Risk assessment, Safety, Technical failure, Water management.
Printed by: Gildeprint
Cover Design: Remco Wetzels

ISBN 978-94-6384-178-8

An electronic version of this thesis is available at
`http://repository.tudelft.nl/`.

Dedicated to Meena and Evian

ACKNOWLEDGEMENTS

"Anything is possible when you have the right people there to support you." Likewise, I had the support from so many wonderful people during my PhD journey. I would like to thank and acknowledge them here.

I feel lucky to have a committed, enthusiastic and supportive supervisory team in this memorable PhD journey: Dr. Wolter Pieters, Dr. André Teixeira and Prof. Pieter van Gelder. Thanks for reviewing and providing constructive feedback on innumerable draft papers and presentation slides. This not only improved the quality of those papers and presentation slides, but also helped me to learn writing easy-to-read papers and structuring presentation slides in an effective way.

I am extremely grateful to Wolter for providing me the freedom and opportunity to explore and learn different academic skills apart from research. Under your wings, I learnt to be critical and conduct high-quality research. I deeply appreciate André for your commitment. Even after moving to Uppsala in the mid of my PhD, you always had the time for our regular meetings and provided crucial inputs to various issues which we encountered at different stages like data collection, effectively addressing reviewers' feedback. Your attention to detail skill inspired me to strive for thoroughness and accuracy. I am very thankful to Pieter for being optimistic about my research and supportive throughout the process. Especially, I am very grateful to your support when we encountered bumpy paths during the data collection phase for the final study of my PhD. Finally, even though I moved to Norway to start my full-time job during the final phase of my PhD, you were all committed to see this through for which I am very thankful.

I am really grateful to Dr. Dina Hadžiosmanović, from being a supervisor to consortium partner, you were always enthusiastic and chipped in with ideas. I enjoyed working with you closely on the first publication of my PhD. Thanks for your constructive feedback which enhanced the work and eventually won an award.

A big thanks to Dr. Nima Khakzad for sharing your knowledge on Bayesian Networks (BNs) which is an integral part of this study. You always answered questions related to BNs with clarity which furthered my understanding on this topic.

I am deeply grateful to our SOS4Flood project consortium partners: Adviescentrum BVI, Deloitte, Forescout (previously known as Security Matters B.V.), HKV consultants, Kuipers Electronic Engineering B.V., Rijkswaterstaat and Royal HaskoningDHV for providing relevant and useful inputs during our annual consortium meetings. My special thanks go to Hellen Havinga for your support during the early years of my PhD, especially in facilitating site visit to a flood defence system and also providing relevant information which helped me shape the research problem that has practical relevance. I am also deeply grateful to Peter-Paul Kuipers who was very supportive and enthusiastic about my research right from the start. In particular, thanks for taking your time to organize a visit to Kuipers Electronic Engineering B.V., pumping station and water board at the start of my PhD which helped me gain an understanding of the domain.

I want to thank all the members of my doctoral defence committee, for taking their time to read my thesis and provide constructive feedback: thank you Prof. Jan van den Berg, Prof. Marielle Stoelinga, Prof. Matthijs Kok and Dr. Paul Smith.

I am extremely thankful to my MSc thesis supervisor, more importantly my mentor, Dr. Harjinder Singh Lallie. You were one of the main reasons why I embarked on this PhD journey as your passion for academic research rubbed off on me. Also, you had the belief more than myself that I can successfully complete a PhD during the start of this journey which always encouraged me.

A huge thanks to Dr. Leon Hermans, my mentor at TUDelft. During the early years of my PhD, we have had many informal but informative meetings. You were always open to discuss and shed light on different topics ranging from teaching to culture.

I also want to thank my amazing PhD peer group friends with diverse backgrounds: Clara, Diego, Fernando, Laura, Sander and Yi for many fruitful discussions related to our research and providing valuable feedback. In addition, we also had various social events like Mexican dinner which helped me survive the early years of my PhD.

Thanks to my wonderful colleagues at the Safety and Security Science (3S) Section: Anca, Behnaz, Chao, Dick, Eelco, Eleonora, Federica, Floor, Frank, Genserik, Jie, Johan, Laobing, Maju, Paul, Pengfei, Peter, Simone, Xin, Yamin, Yuling and Zahra. You have motivated me with your inspiring research that solve interesting real-world problems in diverse domains. Moreover, we have had numerous interesting conversations, memorable social events like excursion to Zeeland, social dinners and annual Secret Santa. I would also like to show my appreciation to the amazing support staff at the 3S Section who were always ready to help. My special thanks go to Astrid and Monique for all the support.

Over the years, I have shared the office with brilliant researchers who always maintained a pleasant working environment: Bas, Eileen, Ivy, Martijn, Paolo and Shenae.

I am extremely grateful to the Institute for Energy Technology (IFE) in general, and to Bjørn Axel Gran specifically, who gave me an opportunity to work full-time and simultaneously complete my PhD. Special thanks to my colleagues in the RSS department at IFE: André, Bjørn Axel, Coralie, Fabien, John Eidar, Per-Arne, Silje, Sizarta, Vikash and Xueli for recharging me with your encouragements, numerous cakes and exciting lunch-time chess games that helped me to work on my PhD even in late evenings/early mornings.

A special thanks to Shiva-Kiran and Priya for making Delft feel like home especially with regular meetups, delicious food and movies. A big thanks to my SRM buddies, even though we were in different time zones, our WhatsApp conversations kept me energized even during challenging times.

Last but not least, I will be forever grateful to for all the love and support that my family has always been giving me at various stages of my life: Chockalingam, Thenmozhi, Kathiresan, Shanthi and Ramu. A special thanks to Senthil, Kamal and Rahul for interesting conversations on topics ranging from sports to movies which kept me refreshed even after a tiring day at work.

Meenakshi, I am very thankful to your relentless support and encouragement. It was definitely a roller coaster ride with highs and lows. However, we stuck together and made it, having you by my side was the best. Evian, thank you for providing the final push to complete my PhD. I am looking forward to the exciting journey ahead.

Halden, November 2020

CONTENTS

SUMMARY

Incidents in critical infrastructures would have a negative effect on the well-being of people and the economy of the country. In countries like The Netherlands, the proper operation of water management infrastructures is essential, as around one third of the country is below sea level. In addition to dikes and dunes, The Netherlands also relies on floodgates to protect the land against flooding. These floodgates are primarily operated via Industrial Control Systems (ICS), which integrate hardware and software with network connectivity to monitor and steer critical processes like closure/opening of floodgates.

Incidents in water management infrastructures, such as unexpected closure/opening of floodgates, could be initiated by problems that are not dealt with appropriately. These problems could be caused by (accidental) technical failures and (intentional) attacks. A typical example of a problem caused by a technical failure is a water-level sensor sending incorrect water-level measurements due to a misconfigured water-level sensor. In contrast, a typical example of a problem caused by an attack is a water-level sensor sending incorrect water-level measurements due to the manipulation of water-level measurements sent to the Programmable Logic Controller (PLC) in the communication channel between the sensor and PLC. A problem that could be caused by both technical failure and attack in case not addressed appropriately as soon as it occurs, can lead to the unexpected closure of a floodgate, causing economic damage.

When the operators observe such problems in infrastructures operated by ICS in practice, they typically assume that the problem is due to a technical failure and initiate corresponding response strategies. In case the problem is caused by an attack, the response strategy initiated towards a technical failure might not be effective. For instance, an effective response strategy for a misconfigured water-level sensor that sends incorrect water level measurements would be to repair the water-level sensor. However, this would not be appropriate for a sensor that sends incorrect water level measurements due to the manipulation of water-level measurements, as it would not block the corresponding attack vector. Determining whether a problem is caused by attacks or technical failures is important for choosing the appropriate response strategy. Therefore, this thesis aims to tackle the problem of distinguishing attacks and technical failures by addressing the following research question:

- **RQ. How to develop decision support to distinguish between intentional attacks and accidental technical failures for problems in water management infrastructures operated by Industrial Control Systems (ICS)?**

We use the Design Science Research (DSR) method to tackle the above-mentioned RQ as it is widely used to create artefacts that solve practical problems. The DSR process consists of four main phases which include: (i) problem identification, (ii) solution design, (iii) evaluation and (iv) communication. In the problem identification phase, we gather

constraints and high-level requirements using semi-structured interviews and focus group sessions with experts in safety and/or security of ICS in the water management sector in the Netherlands. These constraints and high-level requirements are mainly for developing a framework which would then be used as a means to develop decision support for distinguishing attacks and technical failures and their evaluation. This phase results in a set of high-level requirements and constraints based on the responses from experts, which are used as an input for the solution design and evaluation phase of the DSR process. Furthermore, this thesis utilises five different studies to tackle the RQ. The first four studies correspond to the solution design phase, whereas the final study corresponds to the evaluation phase of the DSR process.

In this thesis, the first study systematically reviews and identifies integrated safety and security risk assessment methods in scientific literature, which is a part of the literature research step in the solution design phase of the DSR process. We identify seven integrated safety and security risk assessment methods on the basis of the review methodology employed. Furthermore, the analyses of the identified methods are performed using five different criteria: (i) citations in scientific literature, (ii) steps involved, (iii) stage(s) of risk assessment process addressed, (iv) integration methodology and (v) application(s) and application domain. Sequential and non-sequential are the two classes of integrated safety and security risk assessment methods based on the steps involved in the identified methods. Transportation, power and utilities, and chemical domain are the application domains of the identified methods. There is no specific integrated safety and security risk assessment methods for domains such as water management. A major limitation of the identified methods is that they did not have the capability to consider real-time system information for the analysis.

This study contributes to the scientific community by presenting key characteristics and limitations of integrated safety and security risk assessment methods, which pave the way for improvements and advances in such methods. There is a need for integrated safety and security methods that could consider real-time system information especially to deal with the problem of distinguishing attacks and technical failures, which is where this thesis provides a contribution.

One possibility to include real-time information is through Bayesian Networks (BNs). BNs are part of the family of probabilistic graphical models. BNs are composed of two different components: qualitative and quantitative as shown in Figure 1. The qualitative part is a Directed Acyclic Graph (DAG) that includes nodes representing random variables and directed edges representing cause-effect relationships between these nodes. The quantitative part is the Conditional Probability Tables (CPTs) corresponding to each node, representing a priori marginal and conditional probabilities. BNs have the ability to compute posterior probabilities of target variable(s) as new information (or evidences) for other variables in the BN are available, which is termed as belief updating or probability propagation. Specifically, BNs support four different types of reasoning: (i) predictive reasoning, which is reasoning from cause (Example: pollution) to effects (Example: lung cancer), (ii) diagnostic reasoning, which is reasoning from effect (Example: dyspnea) to cause (Example: lung cancer), (iii) intercausal reasoning, which is reasoning about mutual causes (Example: pollution, smoking) of a common effect (Example: lung cancer) and (iv) combined reasoning, which is the combination of different types of reasoning.

BNs possess the potential to tackle the problem of distinguishing attacks and technical failures considering real-time system information, as they support belief updating based on the combination of predictive and diagnostic reasoning. This is promising based on applications of BNs in medical diagnosis and fault diagnosis. Therefore, the second study is associated with a part of the literature research step in the solution design phase of the DSR process, which systematically reviews and identifies BN models in cyber security in scientific literature.

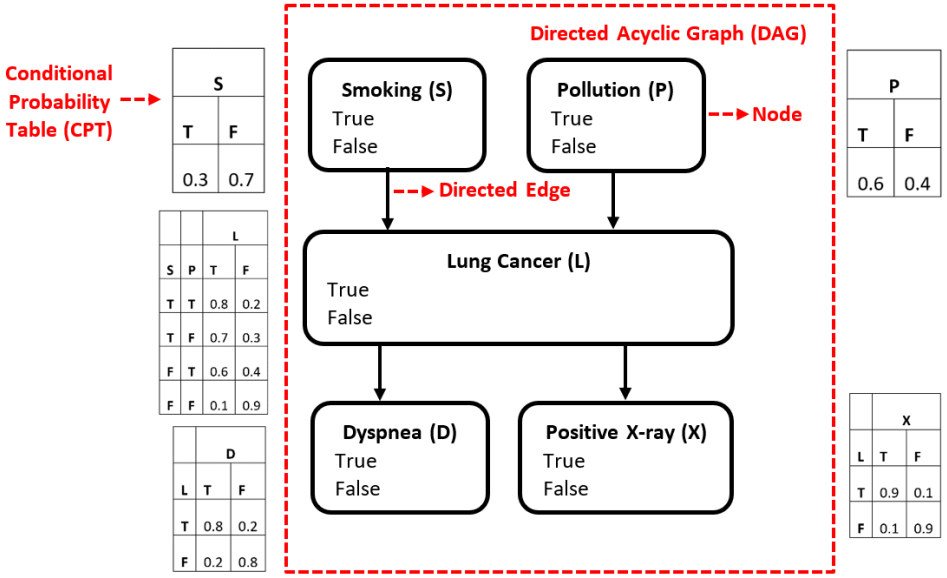


Figure 1: Bayesian Network – Example

We identify 17 BN models in cyber security based on the review methodology employed. Furthermore, the analysis of the identified BN models is performed using eight different criteria: (i) citation details, (ii) data sources used to construct Directed Acyclic Graphs (DAGs) and populate Conditional Probability Tables (CPTs), (iii) the number of nodes used in the model, (iv) type of threat actor, (v) application and application sector, (vi) scope of variables, (vii) the approach(es) used to validate models, and (viii) model purpose and type of purpose. Expert knowledge and empirical data mainly from cyber security reports such as Verizon data breach investigations report are the data sources used to construct Directed Acyclic Graphs (DAGs) and populate Conditional Probability Tables (CPTs) in the identified BN models. In addition, the identified BN models are mainly used for problems that correspond to Information Technology (IT) environments rather than ICS environments.

This study contributes to the scientific community mainly by identifying important usage patterns of BN models in cyber security, which guide new applications. For instance, expert knowledge is an alternate data source to tackle problems associated with ICS environments, which is a useful pattern to develop BN models for our application of

distinguishing attacks and technical failures.

For the problem addressed in this thesis, this study implies that a framework with appropriate types of variables is essential to construct BN models for our application. Moreover, this framework should also include methods to effectively elicit knowledge from experts to construct DAGs and populate CPTs as expert knowledge is an alternate data source for our application. This is the aim of the third study, which corresponds to the artefact design step in the solution design phase of the DSR process. The attack-failure distinguisher framework that we developed is a combination of three different types of variables adapted from existing BNs: (i) contributory factors, (ii) problem and (iii) observations (or test results). The contributory factors are factors that could contribute to the considered problem due to an attack or technical failure, whereas the observations (or test results) provide real-time system information based on the outcome of tests conducted as soon as an operator notices the problem. The attack-failure distinguisher framework would help to construct BN models with appropriate type of variables for diagnosing attacks and technical failures as shown in Figure 2.

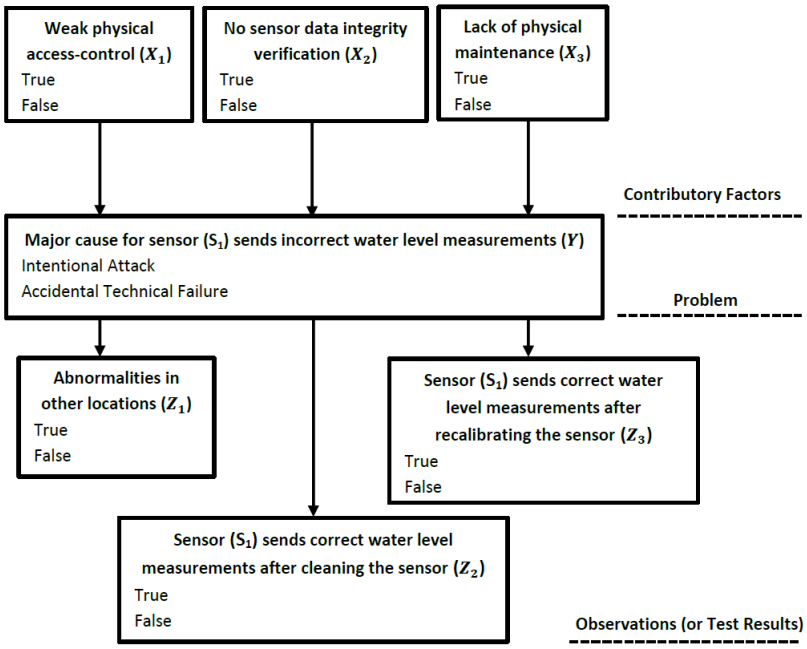


Figure 2: Attack-Failure Distinguisher Framework (Type of Variables): Example

As the BNs themselves are not suitable to effectively elicit knowledge from experts to construct DAGs of BN models, we rely on fishbone diagrams for this purpose. Fishbone diagrams help to systematically identify and organise contributory factors (or sub causes) of a problem related under different categories. The structure of a typical fishbone diagram is shown in Figure 3. However, the typical fishbone diagrams are not directly applicable as they do not include observations (or test results). Therefore, we extend fishbone diagrams to suit our purpose by incorporating observations (or test results) as

shown in Figure 3. The proposed knowledge elicitation method also includes a mapping scheme which would help to translate the extended fishbone diagram to corresponding BN model. We show how the qualitative part of a BN model for our application could be constructed based on the attack-failure distinguisher framework developed in this study with an example problem in the water management domain.

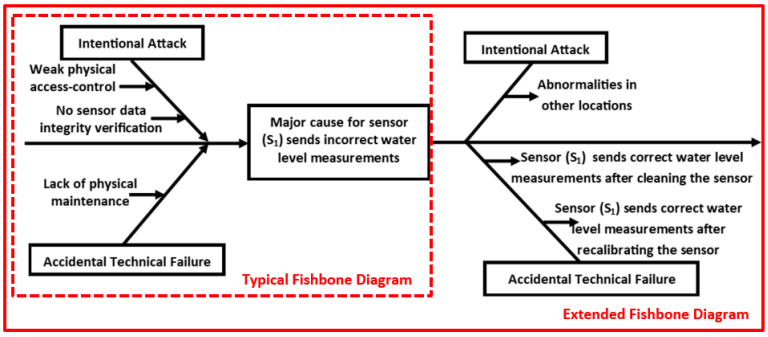


Figure 3: Extended Fishbone Diagram: Structure

This study contributes to the scientific community through the development of the attack-failure distinguisher framework for ICS based on BNs. In particular, it proposes extended fishbone diagrams as a knowledge elicitation method for developing the qualitative part of BN models, which can be used for different problems in different domains.

The fourth study develops a knowledge elicitation method to elicit reliable probabilities from experts to populate CPTs, which provides the input for the quantitative part of the BN models. In order to elicit reliable probabilities, the method should reduce the workload of experts in probability elicitation by reducing the number of conditional probabilities to elicit and facilitate the individual probability entry. We analyse well-known techniques and chose the DeMorgan model to reduce the number of conditional probabilities to elicit, as it helps to deal with opposing influences i.e., contributory factors that would mainly influence one major cause of the problem (attack) as well as contributory factors that would mainly influence the other major cause of the problem (technical failure). The CPT of a child variable can be computed using the DeMorgan model with only n+1 entries elicited from experts instead of $2^{(n+1)}$ entries, where n is the number of parent variables corresponding to the child variable.

In addition, we utilise probability scales with numerical and verbal anchors as shown in Figure 4 to facilitate the individual probability entry. They are effective and practicable as they provide numerical anchors for experts who prefer numbers and verbal anchors for experts who prefer words. Furthermore, there is also a provision to provide precise probabilities using the probability scale with numerical and verbal anchors as an aid. This completes the holistic attack-failure distinguisher framework that would help to construct BN models from expert knowledge for determining the major cause (attack/technical failure) of problems. We demonstrate how the CPTs of a BN model for our application could be populated based on the proposed knowledge elicitation method, with an example problem in the water management domain.

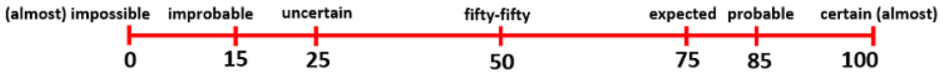


Figure 4: Probability Scale with Numerical and Verbal Anchors

This study contributes to the scientific community by proposing a knowledge elicitation method, which reduces the workload of experts, to develop the quantitative part of BN models. This method is a combination of the DeMorgan model and probability scales with numerical and verbal anchors, which can be used for different problems in different domains.

The developed attack-failure distinguisher framework needs to be evaluated to determine its suitability or utility in practice. Therefore, we evaluate the developed attack-failure distinguisher framework in the fifth study, which is associated with the evaluation phase in the DSR process. Due to the unavailability of real water management infrastructure for evaluation, we rely on artificial evaluation. However, we involve real-users and realistic problems to relate the results to real use.

Firstly, we develop a BN model for determining the major cause of the problem "sensor sends incorrect water level measurements" using the attack-failure distinguisher framework. Knowledge to construct DAG for the considered problem is gathered from experts in safety and/or security of ICS via a focus group workshop and questionnaire. Once the DAG of the BN model is fully constructed, this is validated with experts in safety and/or security of ICS in the water management sector in the Netherlands through a focus group workshop. Furthermore, we gather knowledge from experts in safety and/or security of ICS in the water management sector in the Netherlands to populate CPTs via a focus group workshop and questionnaire. In particular, we utilise the DeMorgan model to reduce the number of conditional probabilities to elicit and probability scales with numerical and verbal anchors to facilitate the individual entry. The CPT size of the problem in the constructed BN model is 512 ($2^{(8+1)}$) entries. However, we elicit only nine entries for the CPT corresponding to the problem from the experts and compute the other probabilities to completely define the CPT of the problem using the DeMorgan model, which notably reduces the workload of experts during probability elicitation. In addition, we used two different illustrative scenarios to demonstrate the suitability or utility of the constructed BN model. The most likely cause for the considered problem in the first illustrative scenario is technical failure, whereas the most likely cause for the considered problem in the second illustrative scenario is attack based on the evidences provided.

This final study provides a full case study of attack-failure distinguisher framework by developing a BN model for the problem of incorrect sensor measurements in floodgates. Furthermore, this study motivates the scientific community focussed on cyber security to investigate other knowledge-based approaches to model cyber security for ICS when there is an unavailability of empirical data by presenting suitability or utility of the developed method.

The results from most of these studies are communicated individually through peer-reviewed publications and conference/workshop presentations. Furthermore, the results from these studies as a whole are communicated through this thesis.

We tackled the RQ of this thesis using the DSR method, especially by following four main phases of the DSR process: (i) problem identification, (ii) solution design, (iii) evaluation and (iv) communication. This study conducted based on the above-mentioned process mainly resulted in the following artefacts: (i) a holistic attack-failure distinguisher framework which also includes methods to effectively elicit knowledge from experts to construct DAGs and populate CPTs of BN models for our application and (ii) decision support to distinguish between intentional attacks and accidental technical failures for a problem in a floodgate operated by ICS. The attack-failure distinguisher framework is used as a means to develop the above-mentioned decision support as a part of their evaluation.

This study has the following limitations: (i) historical data on attacks and technical failures in the water management sector is not available for research, which creates dependence on experts to construct the BN model, (ii) the limited number of experts on safety and/or security of ICS in the water management sector in the Netherlands leads to fewer respondents for the questionnaire to gather knowledge from experts to populate CPTs and (iii) the naturalistic evaluation of the developed artefact with real users in a real setting is not possible as the real system is unavailable. Even though further evaluation of the developed artefact is needed, this thesis shows that distinguishing attacks and failures in ICS is feasible in principle and can be accomplished based on expert knowledge with a manageable workload for the experts.

Furthermore, this study has the following societal benefits: (i) this study contributes to society by motivating the need for methods that integrate safety and security. Specifically, extended fishbone diagrams facilitate experts from the safety and security community to work together and tackle the common problem of distinguishing attacks and technical failures, (ii) this study also contributes to society by developing a method that enables operators to be more proactive about reactive safety and security, which would also help to minimise negative consequences in case of an attack or technical failure by taking informed decisions.

The attack-distinguisher framework determines the major cause (attack/technical failure) of a problem. What is still needed is a complete root cause analysis framework, which would determine the attack vector (in case of an attack) or failure mode (in case of a failure) of a problem to inform appropriate response strategies. Furthermore, the structure of a decision tree could help to visualise and choose effective response strategies for each attack vector and failure mode. It would be intriguing to investigate the use of alternate data sources for our application which might create opportunities to use data-driven approaches in modelling cyber security for ICS.

1

INTRODUCTION

1.1. MOTIVATION

Critical Infrastructures (CIs) are essential to ensure smooth functioning of contemporary society. Disasters in such infrastructures would have direct impact on the well-being of people and a country's economy. CIs are divided into different sectors in different countries. However, the sectors usually in such a list typically include banking and finance, emergency services, energy, environmental protection, food, government services, health, information and communication technologies (ICT), transportation, and water [1].

The proper functioning of water management infrastructures is vital in countries like The Netherlands, as about one third of the country lies below sea level [2]. In addition to dikes and dunes, The Netherlands also relies on floodgates and pumping stations to protect the land against flooding. The unexpected opening of floodgates could lead to flooding. On the other hand, the unexpected closure of floodgates could lead to severe economic damage, for instance, by delaying cargo ships. These problems could be caused by (accidental) technical failures and (intentional) attacks.

In the Netherlands, researchers showed that the password of the systems that could help to control pumping stations in Veere was "veere", which is easy to guess for an adversary [3]. In case an adversary exploits this vulnerability, they could control the water pump and cause flooding. On the other hand, there are problems initiated by technical failures in infrastructures operated by control systems, such as the chemical spill at Haviland enterprises, which is caused by faulty sensor [4]. Similarly, a misconfigured water-level sensor could initiate unexpected closure or opening of the floodgate.

There is a need to effectively respond to such problems that could be observed by operators in infrastructures operated by control systems to recover the system from adversaries in a timely manner and limit negative consequences. It is important to determine whether a problem is caused by (accidental) technical failures or (intentional) attacks for choosing the appropriate response strategy. When the operators notice such problems in infrastructures operated by control systems in practice, they predetermine that the problem is due to an (accidental) technical failure and initiate corresponding

response strategies [5]. This is prevalent in practice based on the discussion with experts in the water management sector. This is because they assume that their infrastructure is not an attractive target for adversaries [6] and the frequency of successful attacks is low compared to technical failures in such infrastructures. However, these problems could also be caused by (intentional) attacks.

In case the problem is caused by an attack, the response strategy initiated towards a technical failure might not be effective. This is because the effective response to an attack would be to block the corresponding attack vector used to cause the problem, whereas an effective response to a technical failure would be to repair or replace the component that caused the problem. For instance, an effective response strategy for a misconfigured water-level sensor that initiates unexpected opening of the floodgate would be to repair the water-level sensor. However, this would not be appropriate for a data manipulation attack on the water-level sensor that initiates unexpected opening of the floodgate, as it would not block the corresponding attack vector. The correct diagnosis is important to choose appropriate response strategies to the observed problem. The major motivation of this research is to develop an effective method that would help to distinguish between attacks and technical failures in the water management domain.

Therefore, the major aim of this thesis is to address the following research question to tackle the practical problem of diagnosing attacks and technical failures:

- **RQ. How to develop decision support to distinguish between intentional attacks and accidental technical failures for problems in water management infrastructures operated by Industrial Control Systems (ICS)?**

The following section provides domain background which is essential for clear understanding of the above-mentioned problem. This is followed by a methodological background which explains the research method which we use to tackle the above-mentioned problem. Furthermore, the method which we use as a basis to develop decision support for distinguishing attacks and technical failures is presented.

1.2. DOMAIN BACKGROUND

This section begins by explaining safety and security, followed by the differences and interdependencies between safety and security. Furthermore, this section explains Industrial Control Systems (ICS), followed by the general architecture and components of the ICS which we deal within this thesis is described. In addition, the differences between ICS and Information Technology (IT) systems are listed. Finally, factors that could help in the diagnosis of attacks and technical failures are highlighted.

1.2.1. SAFETY VS. SECURITY

Safety and security are defined as attributes of dependability [7–9]. Dependability is defined as the ability of a system to provide required services that can justifiably be trusted [7–9]. Safety implies dependability with respect to non-occurrence of catastrophic failures, whereas security implies dependability with respect to non-occurrence of unauthorised action or information handling [7].

The same word is used for both safety and security in different languages such as Danish (Sikkerhed) and Spanish (Seguridad) [10, 11]. However, Pietre-Cambacedes et al.

distinguished safety and security using intentionality [11]. Safety deals with problems caused by accidental acts, whereas security deals with problems caused by malicious acts [11]. The blackout in the Canadian province of Ontario and the North-eastern and Mid-western United States is an example of a problem caused by an accidental act, which is associated with safety [12]. On the other hand, the cyber-attack on a German steel mill is an example of a problem caused by a malicious act, which is associated with security [13].

Furthermore, Pietre-Cambacedes et al. also distinguished safety and security based on the origin and target of the problem [11]. Problems that originate from the system and potentially impact the environment are addressed by safety [11]. In the Northeast blackout, the problem originated from the system via the software bug in the alarm system and impacted the environment by affecting approximately 55 million people with power outage [12]. On the other hand, problems that originate from the environment and potentially impact the system are addressed by security [11]. In the German steel mill cyber-attack, the problem originated from the environment through phishing and impacted the system by damaging the blast furnace, which is associated with security [13]. This distinction may not completely work for security especially for problems that originate from the system and potentially impact the system, which is the case with insider attacks. The above-mentioned difference of intentionality in terms of security and safety is used as a basis in this thesis.

Safety and security are also interdependent, which needs to be considered for effective risk management. Pietre-Cambacedes et al. identified four different types of interdependencies between safety and security: (i) conditional dependency, (ii) mutual reinforcement, (iii) antagonism, and (iv) no interaction [14, 15]. Conditional dependency refers to safety as a condition to security or vice versa [14]. For instance, the faulty installation of a burglary alarm could lead to opportunistic malicious acts, which is an example of a conditional dependency in which safety is a condition for security. Mutual reinforcement refers to the measures that strengthen both safety and security [14]. For instance, activity and event logging could strengthen both safety and security via accident anticipation and attack detection respectively. On the other hand, antagonism refers to safety measures that weaken security or vice versa [14]. For instance, the entry door of a prison should be designed to automatically open in case of fire from the safety viewpoint. However, this could weaken security as it provides an opportunity for the prisoners to escape. Finally, no interaction refers to the cases in which there is no interaction between safety and security [14]. For instance, enhancing physical access-control of an organisation is a security-related measure and there is no interaction between safety and security in this case.

SAFETY AND SECURITY RISK ASSESSMENT

In both safety and security, risk assessment plays an important role to deal with corresponding risks as it is the basis for choosing appropriate risk treatment measures. Risk is defined as the potential for harm due to the likelihood of a problem and its adverse consequences [16]. There are three different phases in a typical risk assessment process as shown in Figure 1.1 which includes: (i) risk identification, (ii) risk analysis and (iii) risk evaluation [17]. There are conventional risk assessment methods in safety such as Failure

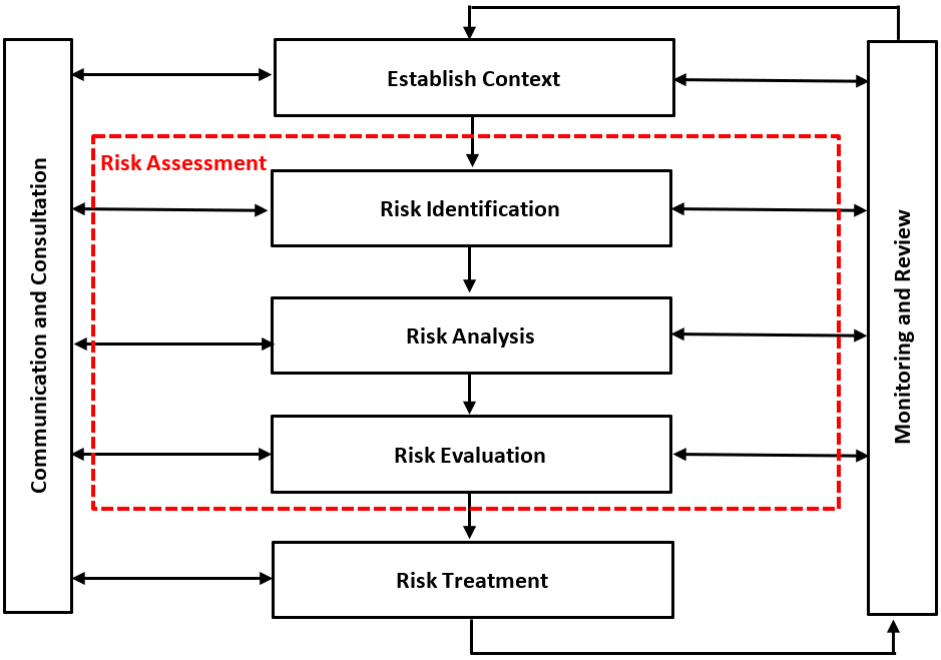


Figure 1.1: Phases of Risk Assessment in ISO 31000

Mode and Effects Analysis (FMEA) [18], Fault Tree Analysis (FTA) [19]. On the other hand, there are conventional risk assessment methods in security such as Attack Trees (ATs) [20], CORAS [21]. These methods only deal with the last type of interdependency (i.e.,) no interaction. However, there are recent developments of integrated safety and security risk assessment methods [22–28] which also deal with other types of interdependencies.

These methods are appropriate for the design phase in the system development lifecycle, but not for the operational phase. For instance, Sabaliauskaite et al. identified safety and security risks that could lead to overpressure condition in the vessel using the Failure-Attack-CounTermeasure (FACT) graph [24]. Furthermore, they also identified appropriate safety and security risk treatment measures, which need to be implemented during the development of such systems to avoid overpressure condition in the vessel during the operation. These risk treatment measures are proactive measures which would help to prevent/reduce the probability of occurrence of the problem (overpressure condition in the vessel) [29].

These proactive measures alone are not enough to protect the system against safety and security risks. This is because we might have overlooked proactive risk treatment measures for some safety and security risks as the threat landscape changes or not implemented it as it is not cost-effective. This could lead to problems which could be observed. Reactive measures would help to prevent/reduce the impact/consequence of the problem [29]. Once the problem is identified, we need to distinguish between technical failures and attacks to put in place effective reactive measures to minimise the negative consequences. This is because the effective reactive measure in case of an attack would be to block the

corresponding attack vector used by an adversary to cause the identified problem. In contrast, the effective reactive measure in case of a technical failure would be to repair or replace the component that caused the identified problem. The method which we develop would act as decision support to operators by providing the most likely cause for the observed problem to put in place reactive measures that strengthen both safety and security.

1.2.2. INDUSTRIAL CONTROL SYSTEMS IN WATER MANAGEMENT

An Industrial Control System (ICS) is defined as an information system used to monitor and steer industrial processes like flood control, gas distribution, power generation and water treatment [30, 31]. An ICS is a common term which includes several type of control systems such as Supervisory Control and Data Acquisition (SCADA) systems, Distributed Control Systems (DCS) [30]. SCADA systems are highly distributed systems used to control geographically distributed assets, whereas DCS is usually located in one plant area [31]. We use the general term ICS in the rest of this thesis.

The fundamental differences between traditional Information Technology (IT) systems and ICS need to be understood by security experts to develop effective and feasible solutions for ICS as not all IT security solutions are suitable for ICS [32]. Some of these differences includes: (i) the cyber security objective of IT systems is to protect data (confidentiality), whereas the cyber security objective of ICS is to protect the integrity of its production process and availability of its system components, (ii) availability deficiencies can often be tolerated in IT systems, whereas availability deficiencies cannot be tolerated in ICS, (iii) IT systems manage data, whereas ICS control physical world, (iv) IT system components have a lifetime of 3 – 5 years, whereas the ICS components have a lifetime of 10 – 15 years, (v) Patching is much easier in IT systems compared to ICS, and (vi) IT systems may support additional security capabilities, whereas ICS may not support additional security capabilities such as encryption due to the computing resource constraints [30, 32].

A typical ICS consists of three layers: (i) field instrumentation, (ii) process control, and (iii) supervisory control [33], bound together by network infrastructure, as shown in Figure 1.2.

The field instrumentation layer consists of field devices which includes sensors (S_i) and actuators (A_i). The sensor is a device which detect environmental changes/events and send the measurement data to the Programmable Logic Controllers (PLCs)/Remote Terminal Units (RTUs) in the process control layer. There are different types of sensor such as proximity sensor, pressure sensor and water-level sensor. The process control layer consists of Programmable Logic Controllers (PLCs)/Remote Terminal Units (RTUs). Typically, PLCs have wired communication capabilities whereas RTUs have wired or wireless communication capabilities. The PLC/RTU receives measurement data from sensors, executes the program logic and controls the physical systems through actuators [34]. There are different types of actuator like electric motor actuator, pneumatic control valve actuator, and solenoid actuator. The supervisory control layer consists of historian databases, software application servers, the Human-Machine Interface (HMI), and the workstation. The historian databases logs production and process data timewise which can be extracted whenever needed. For instance, logs in the historian database could

help to answer the question like how much time was this equipment running in the last 24 hours. The historian databases and software application servers enable the efficient operation of the ICS. The low-level components are configured and monitored with the help of the workstation and the HMI, respectively [34]. The operators are also a part of the supervisory control layer who monitors the system status through HMI and respond to problems.

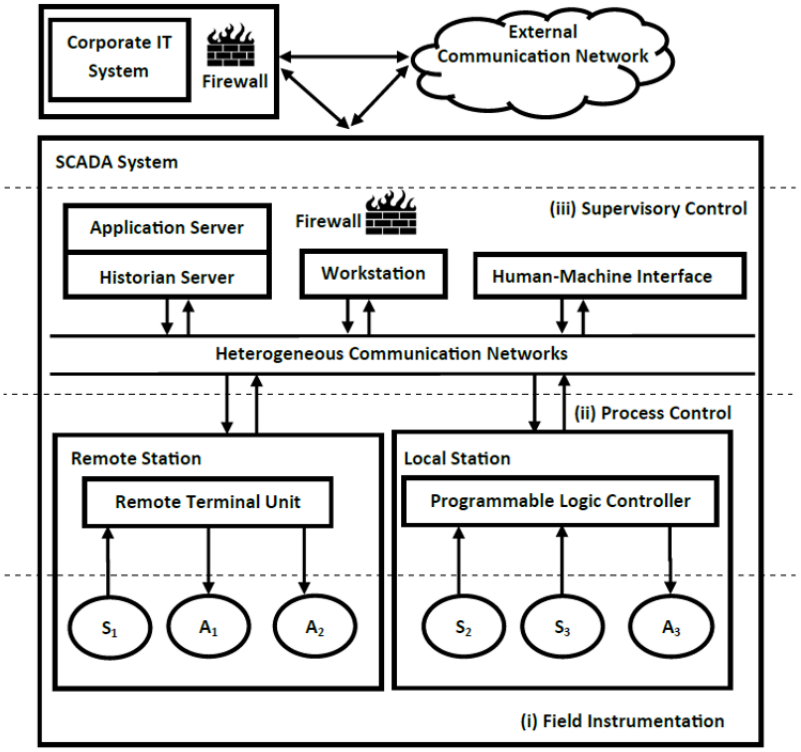


Figure 1.2: Typical ICS Architecture and Layers

Over the last years, floodgates are automated with ICS to reduce the chances of human error during operation [35]. Furthermore, this could provide real-time information about the status of floodgates which are significantly related and connected to each other in a region [36]. This is especially important for decision makers. In addition to dikes, dunes, and pumping stations, The Netherlands also relies on five different storm surge barriers, which is a specific type of floodgate designed to prevent a storm surge from flooding the protected area behind the barrier: (i) Maeslantkering, (ii) Hollandse IJsselkering, (iii) Oosterscheldekering, (iv) Ramspolkering, and (v) Hartelkering [35]. The location of these storm-surge barriers is shown in Figure 1.3. Most of these storm surge barriers are primarily operated with ICS with the manual operation as a backup option [35]. Furthermore, some of the water management infrastructures in the Netherlands solely relies on ICS without any other backup option. In case of a floodgate operated by an ICS, the water-level sensor detects the water level and sends the corresponding water level

measurements to the PLC [37, 38]. Once the PLC receives the water level measurements, this is now compared with the threshold value. In case the water level measurements are above/below the threshold, the PLC closes/opens the floodgate through the actuator which could be an electric motor actuator.

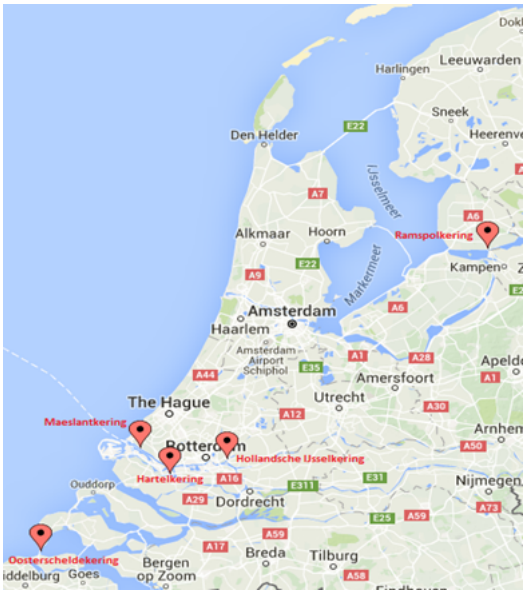


Figure 1.3: Location of Storm-surge Barriers

The automation of floodgates with ICS makes it susceptible to both attacks and technical failures. The factors which includes vulnerabilities and Indicators of Compromise (IoC) might help to determine attacks. A vulnerability is a security weakness which permits unauthorised actions or information handling [39]. Stouffer et al. grouped vulnerabilities of ICS into six different categories which includes: (i) policy and procedure, (ii) architecture and design, (iii) configuration and maintenance, (iv) physical, (v) software development, and (vi) communication and network [39]. For instance, no formal ICS security training and awareness program is a vulnerability under the category policy and procedure. These vulnerabilities can be observed in ICS during the operation. Furthermore, Robinson provided a list of vulnerabilities for ICS and mapped each vulnerability to potential type of attacks which could make use of it [40]. This includes: (i) lack of physical security, which is mapped to unauthorised local access to ICS components, (ii) lack of protocol security, (iii) erosion of isolation, (iv) configurations: convenience over security, (v) weak audit trails, (vi) vendor backdoors, (vii) interconnectedness, (viii) unpatched systems, (ix) malware, and (x) adoption of standards and common technologies. This information might play an important role in the diagnosis of attacks and technical failures as the existence of well-known vulnerabilities in an ICS could increase the likelihood that an observed problem is caused by attacks.

Indicators of Compromise (IoC) are types of evidence which suggest that an unauthorised action or information handling may have happened [41]. Log-in anomaly is an example IoC. Such IoCs could help in the diagnosis of attacks and technical failures.

For instance, the presence of log-in anomaly increases the likelihood that an observed problem is caused by attacks. Furthermore, Hadziosmanovic et al. proposed an approach which could help to identify anomalies based on process logs [42]. They validated their approach using data from a real facility. During this phase, they identified an anomalous event in which an engineering workstation worked during night shifts which is expected to work only during day shifts. This is one of the existing approaches which could provide input for IoCs in ICS. For instance, this method could provide input to the IoC log-in anomaly.

1.3. PROBLEM IDENTIFICATION

In this thesis, we aim to solve the practical problem of distinguishing attacks and technical failures by addressing the above-mentioned research question. We tackled our research question using the Design Science Research (DSR) methodology, which is widely used to create artefacts [43]. The practical problems can be solved using artefacts in numerous cases. The phases of DSR method is detailed in the next section 1.4.1. This chapter corresponds to the problem identification phase in the DSR process, which also includes requirements elicitation for the artefact.

Requirements elicitation is the process of seeking, uncovering, acquiring and elaborating requirements for the artefact [44]. There are different requirements elicitation techniques such as focus groups, interviews, questionnaires and brainstorming [44, 45]. In this thesis, we utilised interviews as it helps to collect data quickly and also it provides an opportunity for probing to get detailed information compared to questionnaires [44, 46]. Furthermore, this is probably the most traditional and commonly used technique for requirements elicitation [44]. There are three different types of interviews which includes: (i) unstructured interviews, (ii) semi-structured interviews, and (iii) structured interviews. In this thesis, we used semi-structured interviews as it is flexible and helps to delve deep into issues. In addition, we also used focus groups as it is an effective way of tapping the views of a number of experts at a time [47].

In the Netherlands, Rijkswaterstaat is responsible for the construction and maintenance of waterways and roads, and flood protection and prevention. Rijkswaterstaat is a part of the ministry of infrastructure and water management. We conducted one-to-one interviews with experts in Rijkswaterstaat who have either of the following roles: (i) technical managers in industrial automation, (ii) security architects. The experts in these roles have a lot of experience working with safety and/or security of ICS in the water management sector. Furthermore, we conducted two focus group sessions during the solution design phase of the DSR process as it is an iterative process. Both the focus group sessions had five participants who have a lot of experience working with safety and/or security of ICS in the water management sector in the Netherlands. We gathered the requirements and constraints using the above-mentioned methods. The list of questions which we asked the experts is provided in Appendix A.

Based on the responses which we received from the experts to those questions, the following set of constraints and high-level requirements is extracted by manually analysing the interview notes and summarising the essence of the responses:

- C1. When the operators notice an abnormal behaviour in a component of the

ICS, they presume that this is due to a technical failure and initiate corresponding response procedures. The response strategy initiated towards a technical failure is not effective in case of an attack.

- C2. There is a lack of real data regarding cyber-attacks as they claim that there are no/limited cyber-attacks on their infrastructures. Furthermore, this is not shareable due to the sensitivity of data.
- C3. Technical failures occur in their infrastructures which are documented as technical failure reports. However, they are also not shareable due to the sensitivity of data.
- C4. The automation department deals with the technical failures, whereas the security department deals with cyber-attacks in the water management infrastructure. There are experts who have expertise in dealing with both technical failures and cyber-attacks.
- C5. Experts are limited in this domain with limited time availability.
- C6. The real water management infrastructure like a floodgate is not available for the evaluation of the developed method due to availability and criticality issues.
- C7. There are system architectures with specific components which are not shareable due to the sensitivity issues. However, there is a possibility to arrange a visit to a water management infrastructure which could help to understand the system architecture on a high-level. Furthermore, the system architecture needs to be anonymised when publishing it.
- C8. There is a need for decision support that would help operators to distinguish between intentional attacks and accidental technical failures as it provides input to the decision-makers to choose appropriate response strategy. However, the selection of these response strategies also depends on cost-benefit and feasibility. Therefore, the focus of this research is to distinguish between attacks and failures which then could be used as an input to choose appropriate response strategy if used in a real infrastructure.
- R1. An effective and practical alternative to data-driven approaches for developing decision support to distinguish between attacks and technical failures is required.
- R2. Decision support should help operators to distinguish between attacks and technical failures by taking into account real-time system information.
- R3. The method for developing decision support should facilitate to involve experts from the department that deals with technical failures and the department that deals with cyber-attacks including experts who have expertise in dealing with both technical failures and cyber-attacks.
- R4. The workload of experts during the knowledge elicitation process for developing decision support to distinguish between attacks and technical failures should be limited.

- R5. The reliability of knowledge elicited for developing decision support to distinguish between attacks and technical failures should be ensured.

- R6. The developed decision support should be scalable to different problems in the real environment.

The above-mentioned set of constraints and high-level requirements plays an important role in structuring the problem space and deriving design decisions systematically. This is used as a basis for the solution design and evaluation phase of the DSR process. This is highlighted as a part of corresponding studies in Section 1.5. Furthermore, we reflect on how these requirements are met in Section 7.1.

1.4. METHODOLOGICAL BACKGROUND

This section explains the Design Science Research (DSR) process, which is widely used to create artefacts [43]. An artefact is defined as an object made by humans for the purpose of solving practical problems [48]. An artefact could be a construct (or concept), a model, a method or an instantiation [49]. The practical problems can be solved using artefacts in numerous cases. In this thesis, we use DSR to create artefacts for solving the practical problem of distinguishing attacks and technical failures. Furthermore, a brief introduction to Bayesian Networks (BNs) is provided with an example and also a few applications of BNs is highlighted. In this thesis, we use BNs as the basis to develop a decision support based on the real-world applications in different domains.

1.4.1. DESIGN SCIENCE RESEARCH

The DSR process consists of four main phases as shown in Figure 1.4: (i) problem identification, (ii) solution design, (iii) evaluation, and (iv) communication [50, 51]. In the first phase of the DSR process, a practical problem is identified [50]. Literature research and expert interviews are predominantly used techniques in this phase to identify a practical problem [49]. The solution is designed in the second phase of the DSR process [50]. Artefact design and supporting literature research are the two steps involved in this phase of the DSR process. The role of literature research in this phase is to ensure research rigour by considering state-of-the-art and existing solutions. In the next phase of the DSR process, the evaluation of the solution is carried out [50]. Finally, the results of the research are communicated through scholarly and/or professional publications.

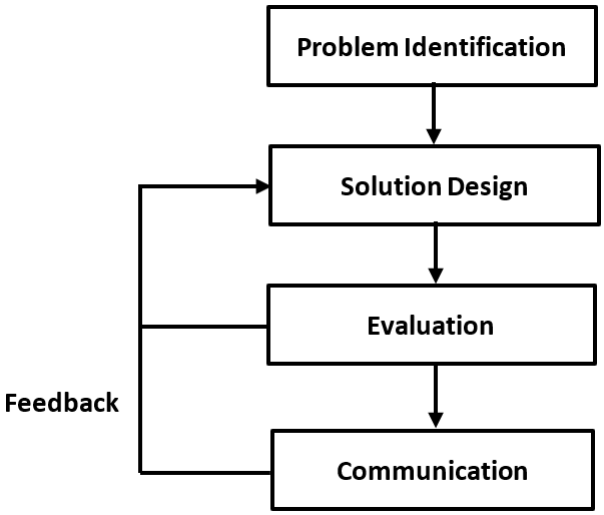


Figure 1.4: Phases of DSR Process

Pries-Heje et al. developed a strategic DSR evaluation framework that could help researchers build evaluation strategies [52]. This framework distinguishes evaluation strategies along three dimensions as shown in Figure 1.5: (i) what to evaluate? ('Design Process' or 'Design Product'), (ii) how to evaluate? ('Naturalistic' or 'Artificial'), and (iii) when to evaluate? ('Ex Ante' or 'Ex Post'). Building an appropriate evaluation strategy would guide researchers in choosing appropriate evaluation method based on the evaluation patterns identified by Sonnenberg et al. from existing DSR literature related to evaluation [53]. The evaluation methods used in the DSR scientific literature includes action research, case study, controlled experiment, field experiment, formal proof, illustrative scenario, logical argument, prototype, and survey [54, 55].

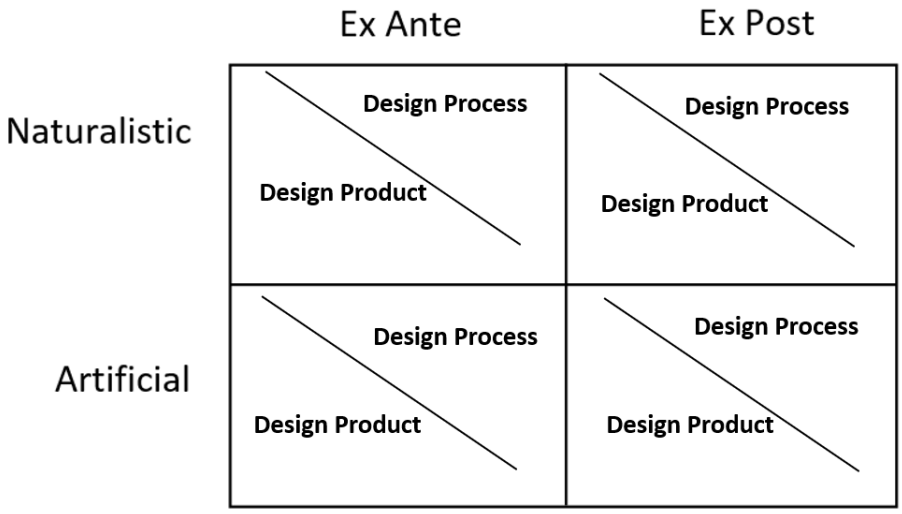


Figure 1.5: A Strategic DSR Evaluation Framework

1

Artificial evaluation evaluates the artefacts in a contrived and non-realistic way [52, 56]. Naturalistic evaluation evaluates the artefacts in its real environment, i.e., within the real system [52, 56]. The 'ex ante' refers to the evaluation prior to artefact construction, whereas 'ex post' refers to the evaluation after artefact construction [56]. Choosing between ex ante or ex post evaluation (or both) in DSR depends on the scope of the research project [52].

1.4.2. BAYESIAN NETWORKS

BNs belong to the family of probabilistic graphical models [57]. BNs consist of both qualitative and quantitative components [58, 59]. The qualitative component is a Directed Acyclic Graph (DAG) as shown in Figure 1.6, which is a combination of a set of variables and directed edges between these variables. The directed edges in DAGs represent the cause-effect relationship between the corresponding variables. Furthermore, each of these variables has a finite set of mutually exclusive states. The quantitative component is the Conditional Probability Tables (CPTs) as shown in Figure 1.6. The CPT includes conditional probabilities for all possible combinations of the child and parent variable states. For the variable without any parent, the CPT includes prior probabilities of the corresponding variable.

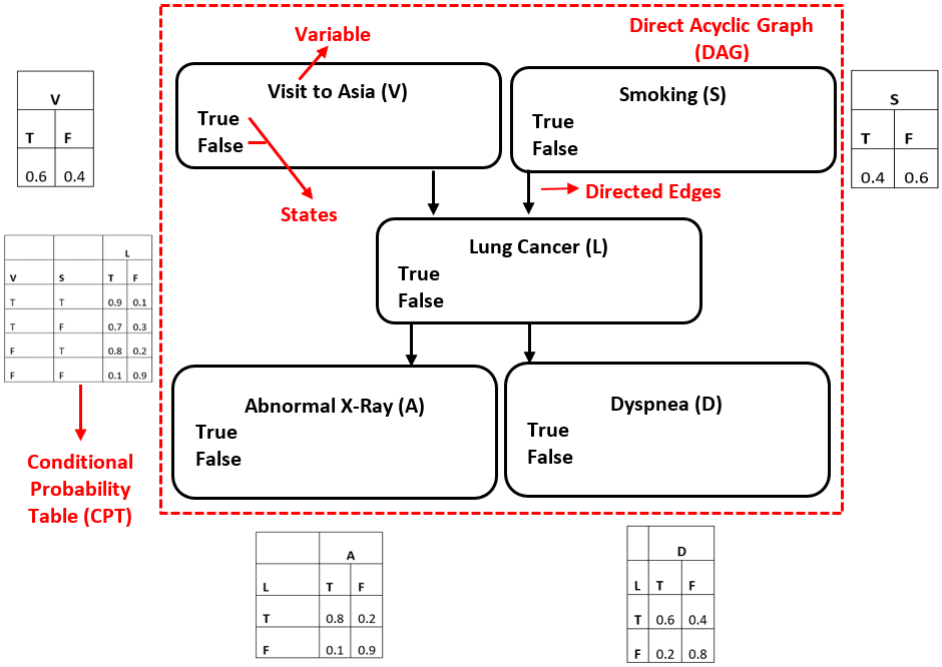


Figure 1.6: BN Example

The BN example provided in Figure 1.6 consists of three layers. The upper layer consists of factors ("visit to Asia", "smoking") that would increase the likelihood of a patient having lung cancer. Furthermore, the middle layer consists of the variable which

we want to query ("lung cancer") with evidences for variables in the other layers of the BN. Finally, the lower layer includes symptoms or test results ("abnormal X-ray", "dyspnea"). This BN would help to diagnose whether the patient have lung cancer or not, given evidence for variables in other layer(s) of the BN. When the evidence for variables in the BN is obtained, the posterior probabilities of non-evidenced variables would be updated. This process is termed as belief updating or inference or probability propagation.

BNs support four different types of reasoning which includes: (i) predictive reasoning, (ii) diagnostic reasoning, (iii) intercausal reasoning, and (iv) combined reasoning. Predictive reasoning is the reasoning from cause (Example: smoking) to effects (Example: lung cancer) as shown in Figure 1.7. In the example shown in Figure 1.7, the evidence for a variable in the upper layer (cause) is provided which in turn would help to query the posterior probabilities of non-evidenced variable(s) in the middle and lower layer (effects) of the BN.

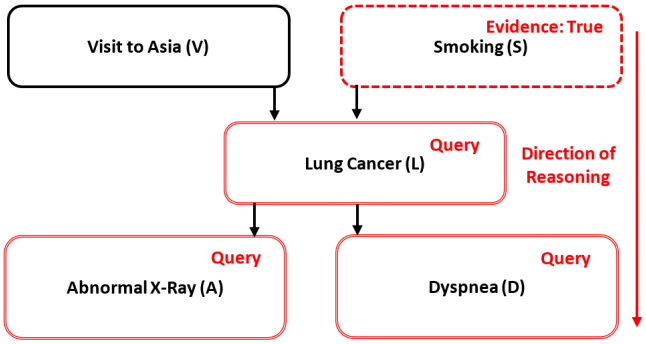


Figure 1.7: BN Example - Predictive Reasoning

Diagnostic reasoning is the reasoning from effect (Example: abnormal X-ray) to cause (Example: lung cancer) as shown in Figure 1.8. In the example shown in Figure 1.8, the evidence for a variable in the lower layer (effect) is provided which in turn would help to query the posterior probabilities of non-evidenced variable(s) in the middle and upper layer (causes) of the BN.

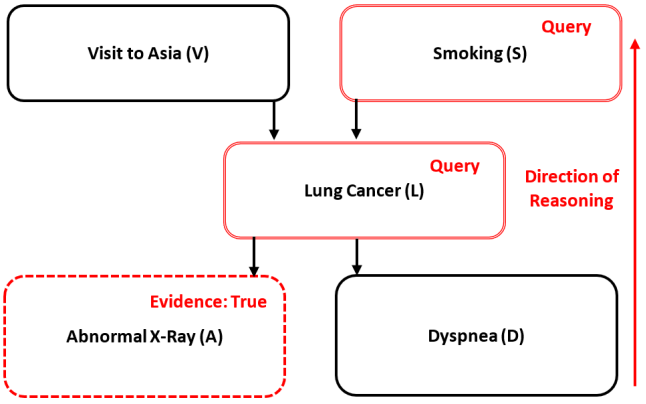


Figure 1.8: BN Example - Diagnostic Reasoning

Intercausal reasoning is the reasoning about mutual causes (Example: visit to Asia, smoking) of a common effect (Example: lung cancer) as shown in Figure 1.9. For instance, the lung cancer could be caused by visit to Asia or smoking. Initially, these causes are independent. Suppose that we find evidence that the patient smokes. This new information explains lung cancer, which in turn lowers the probability that the lung cancer was caused by visit to Asia. Even though these causes are initially independent, the alternative cause is explained away with the evidence of another cause.

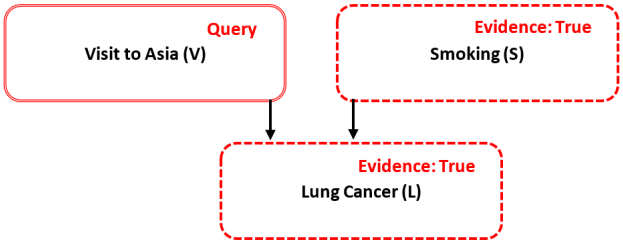


Figure 1.9: BN Example - Intercausal Reasoning

Combined reasoning is the combination of different types of reasoning. In the example shown in Figure 1.10, the evidence for a variable in the upper and lower layer is provided which would help to query the posterior probabilities of non-evidenced variables in the middle and lower layer of the BN. This is a combination of predictive and diagnostic reasoning. Unlike FTAs and ATs which are well established in safety and security domain respectively, BNs support diagnostic (or backward) reasoning [60].

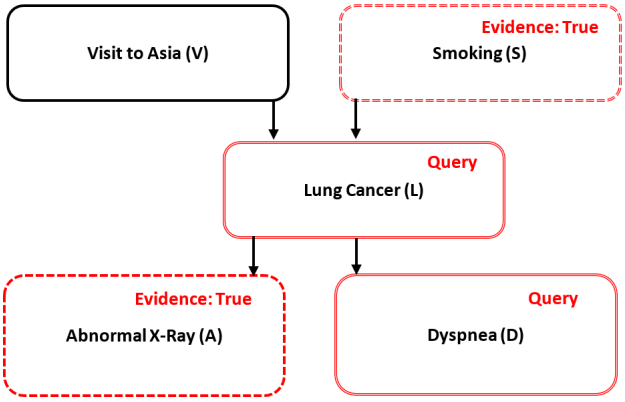


Figure 1.10: BN Example - Combined Reasoning

BNs are used for developing medical decision support systems [61–66]. Furthermore, BNs are also used in fault diagnosis [67–69], cyber security [70–83]. Kahn Jr et al. developed and evaluated a BN (MammoNet) that supports diagnosis of breast cancer [61]. MammoNet is a three-layer BN. The upper layer consists of five patient-history features. Furthermore, the middle layer is a target variable (breast cancer) which we intend to query with evidences for variables in other layer(s) of the BN. Finally, the lower layer includes two physical findings and 15 mammographic features. The data sources which were

used to populate CPTs includes peer-reviewed medical literature, census data, health statistics reports and an expert mammographer. Once the evidence for variables in the upper and/or lower layer is provided, the posterior probability of the target variable is updated. MammoNet is evaluated using 77 cases with known outcomes, in which 23 of the 25 positive cases were identified correctly by MammoNet.

Curiac et al. developed and evaluated a BN that assists in diagnosis of psychiatric disease [62]. This is a three-layer BN. The upper layer consists of risk factors such as recent birth, unwanted incident. Furthermore, the middle layer includes four psychiatric diseases which are the target variables in this BN which we intend to query with evidences for variables in other layer(s) of the BN. Finally, the lower layer consists of symptoms such as personality/emotional life deterioration, social life deterioration. The data sources which were used to populate CPTs includes medical statistics from the psychiatric division in the Lugoj municipal hospital, and physicians. Once the evidence for variables in the upper and /or lower layer is provided, the posterior probabilities of the target variables are updated. This BN model is evaluated using four imaginary case studies. This BN identify which of the considered diseases is more likely in a particular patient after providing evidences for some variables in the BN, whereas the BN model developed by Kahn Jr et al [61] identify the likelihood of breast cancer in a particular patient after providing evidences for some variables in the BN.

1.5. RESEARCH APPROACH

This thesis tackles the practical problem of diagnosing attacks and technical failures by addressing the following research question:

- **RQ. How to develop decision support to distinguish between intentional attacks and accidental technical failures for problems in water management infrastructures operated by Industrial Control Systems (ICS)?**

This chapter corresponds to the problem identification phase in the DSR process. We identified the practical problem based on literature research and expert interviews. The main research question is divided into several sub-questions. These sub-questions are explored in subsequent chapters through five separate studies that addresses main phases of the DSR process as shown in Table 1.1. The sections below introduce these studies and their corresponding sub-questions in detail.

1.5.1. STUDY 1: INTEGRATED SAFETY AND SECURITY RISK ASSESSMENT METHODS: A SURVEY OF KEY CHARACTERISTICS AND APPLICATIONS (CHAPTER 2)

The first study is a survey of integrated safety and security risk assessment methods, which corresponds to the literature research step in the solution design phase of the DSR process. As a part of this phase in the DSR process, the state-of-the-art methods need to be considered to ensure research rigour [50], which is the aim of this study. There are recent developments of integrated safety and security risk assessment methods to facilitate the safety and security community working together in risk management considering the interdependencies between safety and security. However, a comprehensive review of such

methods is missing. The objectives of this study are: (a) to identify integrated safety and security risk assessment methods in scientific literature, and (b) to analyse the identified methods to pinpoint key characteristics and applications.

In short, the study aims to address the following research question:

- **SQ1.** What are the key characteristics of integrated safety and security risk assessment methods, and their applications?

1.5.2. STUDY 2: BAYESIAN NETWORK MODELS IN CYBER SECURITY: A SYSTEMATIC REVIEW (CHAPTER 3)

In study 1, we concluded that the identified integrated safety and security risk assessment methods did not consider real-time system information. Therefore, these methods are not appropriate for the practical problem of diagnosing attacks and technical failures in the operational phase. BNs possess the potential to address this challenge based on real-world applications in medical diagnosis and fault diagnosis. BNs have also been used in cyber security. However, the comprehensive review of BN models in cyber security is missing. The objectives of this study are: (a) to identify standard BN models in cyber security literature, and (b) to pinpoint patterns in the use of such models in cyber security based on the analysis of identified models. This study corresponds to the literature research step in the solution design phase of the DSR process in which we identify state-of-the-art BNs in cyber security. This study would subsequently help to design artefact for our practical problem considering important usage patterns and challenges of the method which we chose. Furthermore, this study intends to fulfil R1 on the need for an effective and practical alternative to data-driven approaches by systematically reviewing the use of BNs in cyber security.

In short, this study aims to address the following research question:

- **SQ2.** What are the important patterns in the use of standard Bayesian Network (BN) models in cyber security?

1.5.3. STUDY 3: COMBINING BAYESIAN NETWORKS AND FISHBONE DIAGRAMS TO DISTINGUISH BETWEEN INTENTIONAL ATTACKS AND ACCIDENTAL TECHNICAL FAILURES (CHAPTER 4)

In study 2, we concluded that BNs possess the potential to develop a BN model that would help to distinguish between intentional attacks and accidental technical failures. However, a framework that would help to build BN models for distinguishing attacks and technical failures is missing. Furthermore, we concluded that expert knowledge is one of the predominant data sources utilised to build BN models in cyber security due to lack of data. However, BNs themselves are not suitable for knowledge elicitation. The objectives of this study are: (a) to develop the attack-failure distinguisher framework for constructing BN models for determining the major cause of an abnormal behaviour in a component of the ICS, (b) to leverage fishbone diagrams for knowledge elicitation within our framework to construct Directed Acyclic Graphs (DAGs) of such BN models, and (c) to demonstrate the application of the developed methodology via a case study. This study corresponds to the artefact design step in the solution design phase of the DSR process in

which an artefact is developed to the problem which we considered.The attack-failure distinguisher framework aims to realize R2 on the capability of decision support to take into account real-time system information. Furthermore, using fishbone diagrams for knowledge elicitation as a part of the attack-failure distinguisher framework intends to fulfil R3 on involving experts from both the departments that deal with technical failures and cyber-attacks and also experts who have expertise in both. Furthermore, this intends to achieve R4 on reducing workload of experts during the knowledge elicitation process and also to realize R5 on ensuring the reliability of knowledge elicited.

In short, this study aims to address the following research question:

- **SQ3.** How could we combine Bayesian Networks and Fishbone Diagrams to find out whether an abnormal behaviour in a component of the ICS is due to (intentional) attack or accidental technical failure or neither?

1.5.4. STUDY 4: PROBABILITY ELICITATION FOR BAYESIAN NETWORKS TO DISTINGUISH BETWEEN INTENTIONAL ATTACKS AND ACCIDENTAL TECHNICAL FAILURES (CHAPTER 5)

The attack-failure distinguisher framework which we developed in study 3 would be incomplete without the method that would help to elicit prior/conditional probabilities from experts to construct Conditional Probability Tables (CPTs). The objectives of this study are: (a) to propose a method that would reduce the workload of experts in probability elicitation, and (b) to demonstrate the application of the proposed method via a case study. This method should reduce the number of conditional probabilities to elicit from experts and provide visual aid that could help experts to answer in terms of probabilities without much difficulty. This will help to elicit reliable probabilities from experts. This study corresponds to the artefact design step in the solution design phase of the DSR process. This study also aims to achieve R4 on reducing workload of experts during the knowledge elicitation process and R5 on ensuring the reliability of knowledge elicited. The development of an artefact is an iterative process. This study addresses the major limitation of the artefact designed in the previous study, which is the attack-failure distinguisher framework did not include a method to effectively elicit probabilities from experts to construct the CPTs.

In short, this study aims to address the following research question:

- **SQ4.** How could we elicit expert knowledge to effectively construct Conditional Probability Tables (CPTs) of Bayesian Network models for distinguishing attacks and technical failures?

1.5.5. STUDY 5: BAYESIAN NETWORK MODEL TO DISTINGUISH BETWEEN INTENTIONAL ATTACKS AND ACCIDENTAL TECHNICAL FAILURES: A CASE STUDY OF FLOODGATES (CHAPTER 6)

The artefact (attack-failure distinguisher framework) is developed in studies 3 and 4. However, this needs to be evaluated to assess the utility or suitability of the artefact. The objectives of this study are: (a) to develop a BN model for the problem related to incorrect water level measurements using the developed attack-failure distinguisher framework,

and (b) to demonstrate the developed BN model using 2 different scenarios. Due to the lack of data in cyber security for ICS, we would rely on expert knowledge to develop the BN model. Furthermore, expert knowledge is substantive information on a specific domain based on the system knowledge that is not commonly known by others [84]. This study corresponds to the evaluation phase in the DSR process, in which we evaluate the artefact (attack-failure distinguisher framework). We utilise artificial evaluation strategy due to the unavailability of real environment for this evaluation. However, we intend to make it more realistic by involving real-users, and realistic problems. Therefore, the results from the artificial evaluation could correspond to real use and fulfil R6. There are different above-mentioned evaluation methods used in practise, but not all can be used for ex-post, artificial evaluation. For instance, field experiment can be used for ex-post, natural evaluation, but not appropriate for artificial evaluation as the real environment is not available. Therefore, we utilise appropriate evaluation methods for ex-post, artificial evaluation to assess the utility or suitability of the artefact.

In short, this study aims to address the following research question:

- **SQ5.** How could we develop Bayesian Network (BN) models for distinguishing attacks and technical failures in floodgates using the attack-failure distinguisher framework?

The corresponding output of the main phases of DSR process is also shown in Table 1.1. For instance, we identified the main problem in the first phase of the DSR process: lack of decision support to distinguish between attacks and technical failures. A practical problem in this thesis is identified using literature research and expert interviews. Furthermore, the solution design phase of the DSR process utilise state-of-the-art and existing solutions as the base for the creative process. An output of the solution design phase is the design artefact which we developed includes attack-failure distinguisher framework with appropriate methods to effectively elicit expert knowledge to construct DAGs and CPTs for our application as shown in Figure 1.11. Finally, the evaluation phase of the DSR process rely on experts via focus groups and questionnaires to gather data to build the prototype (i.e., the BN model to distinguish between attacks and technical failures for a problem). This prototype helps to perform illustrative scenarios and demonstrate the utility or suitability of the developed artefact. In this phase, we conducted two focus group sessions to build the qualitative and quantitative component of the prototype. The focus group session to build the qualitative component of the prototype had five participants who have a lot of experience working with safety and/or security of ICS in the water management sector in the Netherlands. We complemented it with a questionnaire to build the qualitative component of the prototype, which had nine respondents who have at least a year of experience working with safety and/or security of ICS. Furthermore, we conducted another focus group session to review the completed qualitative component of the prototype and build the quantitative component of the prototype, which had five participants who have a lot of experience working with safety and/or security of ICS in the water management sector in the Netherlands. We complemented it with a questionnaire to build the quantitative component of the prototype, which had five respondents who have at least a year of experience working with safety and/or security of ICS in the water management sector in the Netherlands.

	Problem Identification	Solution Design		Evaluation
		Literature Research	Artefact Design	
Introduction	Lack of decision support to distinguish attacks and technical failures			
Study 1		Integrated safety and security risk assessment methods		
Study 2		Bayesian Network models in cyber security		
Study 3			Attack-failure distinguisher framework (First iteration)	
Study 4			Attack-failure distinguisher framework (Second iteration)	
Study 5				Prototype and Illustrative scenarios based on a case study in floodgates

Table 1.1: Main Phases of DSR Process and Corresponding Chapters in this Thesis

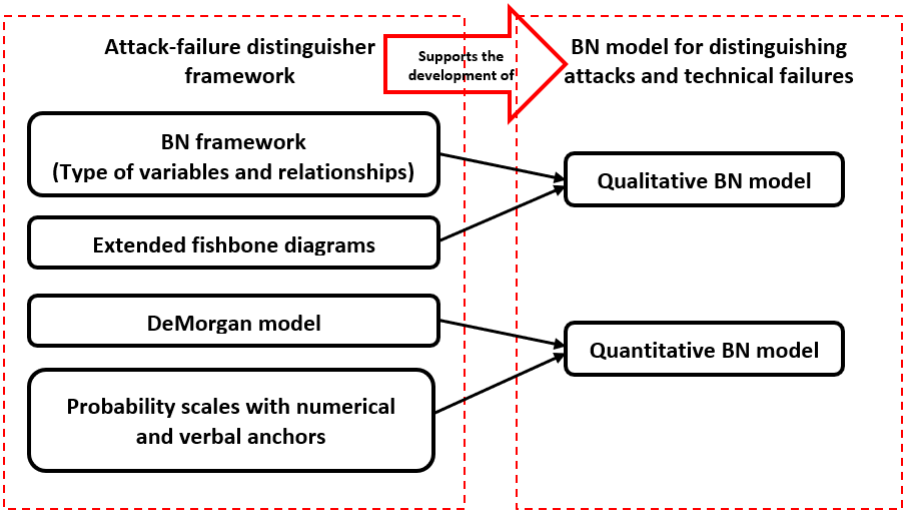


Figure 1.11: Attack-Failure Distinguisher Framework Components and their Role on the BN Model Development

1

1.6. THESIS OVERVIEW

The remainder of this thesis is organised in six chapters as shown in Figure 1.12. Chapter 2 analyses state-of-the-art integrated safety and security risk assessment methods. Chapter 3 analyses BN models in cyber security to identify important patterns that could be used to develop the BN model for our application. Chapter 4 describes the attack-failure distinguisher framework that could help to construct BN models for diagnosing attacks and technical failures. This framework includes a method to elicit expert knowledge to construct DAGs of such BN models. Chapter 5 includes a method to elicit expert knowledge to populate CPTs of such BN models. The developed attack-failure distinguisher framework is applied to a case study in floodgates using the problem of incorrect water level measurements in Chapter 6. Finally, the concluding remarks of this research is provided in Chapter 7.

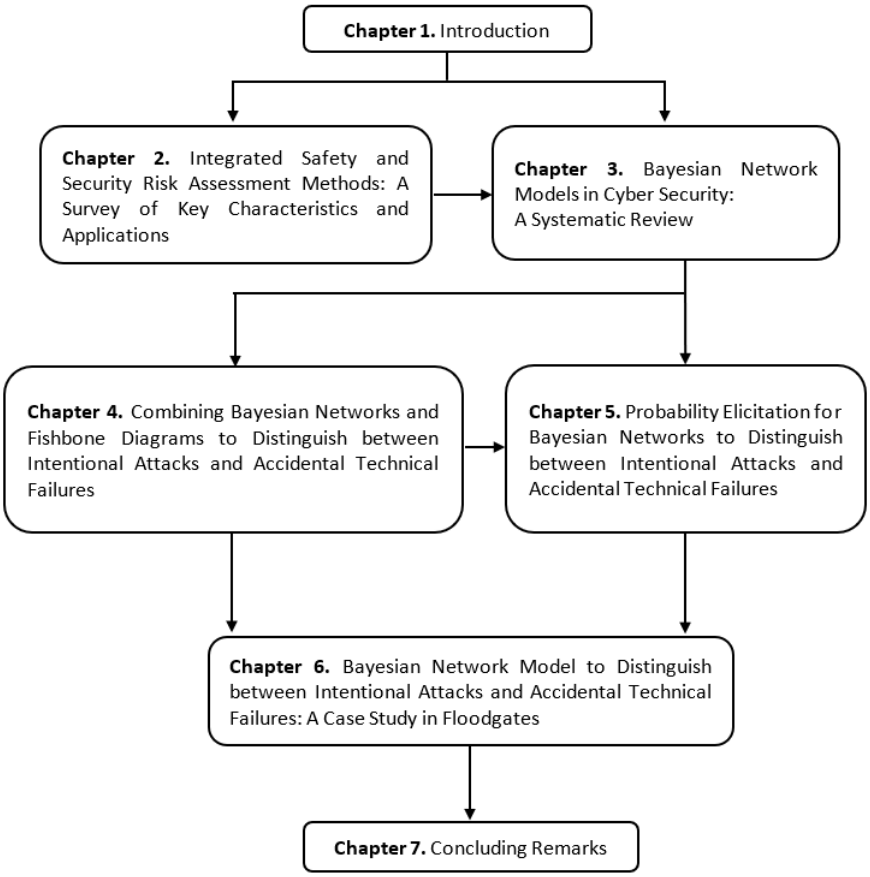


Figure 1.12: Thesis Overview

REFERENCES

[1] CIPedia.: Critical Infrastructure Sector. Available: https://publicwiki-01.fraunhofer.de/CIPedia/index.php/Critical_Infrastructure_Sector (2019)

[2] Hekstra, G.: Will Climatic Changes Flood the Netherlands? Effects on Agriculture, Land use and Well-being, Ambio, pp. 316 - 326. (1986)

[3] Tofino Security.: Cyber Security Nightmare in the Netherlands. Available: https://www.tofinosecurity.com/blog/cyber-security-nightmare-netherlands (2012)

[4] mLIVE.: Faulty Sensor Causes Chemical Spill at Haviland Enterprises. Available: https://www.mlive.com/news/grand-rapids/2015/08/tank_level_sensor_cause_of_che.html (2015)

[5] Macaulay, T., Singer, B. L.: Cybersecurity for Industrial Control Systems: SCADA, DCS, PLC, HMI, and SIS, Auerbach Publications. (2016)

[6] Kaspersky Lab.: Five Myths of Industrial Control Systems Security, Available: https://media.kaspersky.com/pdf/DataSheet_KESB_5Myths-ICSS_Eng_WEB.pdf (2014)

[7] Schoitsch, E.: Design for Safety and Security of Complex Embedded Systems: A Unified Approach, In Cyberspace Security and Defense: Research Issues, pp. 161 - 174, Springer. (2005)

[8] Avižienis, A., Laprie, J.-C., Randell, B.: Dependability and its Threats: A Taxonomy, In Building the Information Society, pp. 91 - 120, Springer. (2004)

[9] Al-Kuwaiti, M., Kyriakopoulos, N., Hussein, S.: A Comparative Analysis of Network Dependability, Fault-tolerance, Reliability, Security, and Survivability, IEEE Communications Surveys & Tutorials, vol. 11, no. 2, pp. 106 - 124. (2009)

[10] Burns, A., McDermid, J., Dobson, J.: On the Meaning of Safety and Security, The Computer Journal, vol. 35, no. 1, pp. 3 - 15. (1992)

[11] Piètre-Cambacédès, L., Chaudet, C.: The SEMA Referential Framework: Avoiding Ambiguities in the Terms “Security” and “Safety”, International Journal of Critical Infrastructure Protection, vol. 3, no. 2, pp. 55 - 66. (2010)

[12] Zhivich, M., Cunningham, R. K.: The Real Cost of Software Errors, IEEE Security & Privacy, vol. 7, no. 2, pp. 87 - 90. (2009)

[13] RISI.: German Steel Mill Cyber Attack. Available: http://www.risidata.com/database/detail/german-steelmill-cyber-attack (2014)

[14] Piètre-Cambacédès, L., Bouissou, M.: "Modeling Safety and Security Interdependencies with BDMP (Boolean Logic Driven Markov Processes)," In 2010 IEEE International Conference on Systems, Man and Cybernetics, 2010, pp. 2852 - 2861: IEEE.

[15] Sadvandi, S., Chapon, N., Pièrre-Cambacédès, L.: Safety and Security Interdependencies in Complex Systems and SoS: Challenges and Perspectives, In Complex Systems Design & Management, Springer, pp. 229 - 241. (2012)

[16] Cox, Jr., Anthony, L.: Some Limitations of "Risk= Threat× Vulnerability× Consequence" for Risk Analysis of Terrorist Attacks, Risk Analysis, vol. 28, no. 6, pp. 1749 - 1761. (2008)

[17] European Union Agency for Network and Information Security (ENISA).: The Risk Management Process, Available: https://www.enisa.europa.eu/activities/risk-management/current-risk/risk-management-inventory/rm-process (2019)

[18] Stamatis, D.H.: Failure Mode and Effect Analysis: FMEA from Theory to Execution, ASQ Quality Press. (2003)

[19] Lee,W.-S., Grosh, D. L., Tillman, F. A., Lie, C. H.: Fault Tree Analysis, Methods, and Applications: A Review, IEEE Transactions on Reliability, vol. 34, no. 3, pp. 194 - 203. (1985)

[20] Schneier, B.: Attack Trees, Dr. Dobb's journal, vol. 24, no. 12, pp. 21 - 29. (1999)

[21] Den Braber, F., Hogganvik, I., Lund, M. S., Støolen, K., Vraalsen, F.: Model-based Security Analysis in Seven Steps — A Guided Tour to the CORAS Method, BT Technology Journal, vol. 25, no. 1, pp. 101 - 117. (2007)

[22] Macher, G., Höller, A., Sporer, H., Armengaud, E., Kreiner, C.: A Combined Safety-hazards and Security-threat Analysis Method for Automotive Systems, In International Conference on Computer Safety, Reliability, and Security, pp. 237 - 250, Springer. (2014)

[23] Schmittner, C., Ma, Z., Schoitsch, E., Gruber, T.: A Case Study of FMVEA and CHASSIS as Safety and Security Co-analysis Method for Automotive Cyber-Physical Systems, In Proceedings of the 1st ACM Workshop on Cyber-Physical System Security, pp. 69 - 80, ACM. (2015)

[24] Sabaliauskaite, G., Mathur, A. P.: Aligning Cyber-Physical System Safety and Security, In Complex Systems Design & Management Asia, Springer, pp. 41 - 53. (2015)

[25] Schmittner, C., Ma, Z., Smith, P.: FMVEA for Safety and Security Analysis of Intelligent and Cooperative Vehicles, In International Conference on Computer Safety, Reliability, and Security, pp. 282 - 288, Springer. (2014)

[26] Chen, Y.-R., Chen, S.-J., Hsiung, P.-A., Chou, I.-H.: Unified Security and Safety Risk Assessment - A Case Study on Nuclear Power Plant, In 2014 International Conference on Trustworthy Systems and Their Applications, pp. 22 - 28, IEEE. (2014)

[27] Steiner, M., Liggesmeyer, P.: Combination of Safety and Security Analysis - Finding Security Problems that Threaten the Safety of a System, In 2013 International Conference on Computer Safety, Reliability, and Security Workshops, pp. 233 - 240. (2013)

[28] Fovino, I. N., Masera, M., De Cian, A.: Integrating Cyber Attacks within Fault Trees, Reliability Engineering & System Safety, vol. 94, no. 9, pp. 1394 - 1402. (2009)

[29] Aqlan, F., Lam, S.S.: A Fuzzy-based Integrated Framework for Supply Chain Risk Assessment, International Journal of Production Economics, vol. 161, pp. 54-63. (2015)

[30] NIST 800-30.: Guide for Conducting Risk Assessments (Revision 1). (2012)

[31] Hadziosmanovic, D.: The Process Matters: Cyber Security in Industrial Control Systems, University of Twente. (2014)

[32] Neitzel, L., Huba, B.: Top Ten Differences between ICS and IT Cybersecurity, InTech, vol. 61, no. 3, pp. 12 - 18. (2014)

[33] Endi, M., Elhalwagy, Y., Attalla, H.: Three-layer PLC/SCADA System Architecture in Process Automation and Data Monitoring, The 2nd International Conference on Computer and Automation Engineering (ICCAE), vol. 2, pp. 774 - 779, IEEE. (2010)

[34] Skopik, F., Smith, P.D.: Smart Grid Security: Innovative Solutions for a Modernized Grid, Syngress. (2015)

[35] Nogueira, H. I. S., Walraven, M.: Overview of Storm Surge Barriers, Rijskwaterstaat & Deltares. (2018)

[36] Hajjaj, H., Salama, S., Hameed Sultan, M. T., Moktar, M. H., Lee, S. H.: Utilizing the Internet of Things (IoT) to Develop a Remotely Monitored Autonomous Floodgate for Water Management and Control, Water, vol. 12, no. 2. (2020)

[37] Park, S., Kim, B., Won, T., Heo, J.: IoT-based Floodgate Control System, In Proceedings of the Conference on Research in Adaptive and Convergent Systems, pp. 61-62. (2019)

[38] Sahu, V., Tripathi, N.: Automation of Gates of Water Reservoir Using Programmable Logic Controller (PLC), International Journal of Research in Applied Science & Engineering Technology, vol. 6, no. 4. (2018)

[39] Stouffer, K., Falco, J., Scarfone, K.: Guide to Industrial Control Systems (ICS) Security, NIST Special Publication, vol. 800, no. 82. (2011).

[40] Robinson, M.: The SCADA Threat Landscape, In 1st International Symposium for ICS & SCADA Cyber Security Research 2013 (ICS-CSR 2013), pp. 30 - 41.(2013)

[41] Panwar, A.: iGen: Toward Automatic Generation and Analysis of Indicators of Compromise (IoCs) using Convolutional Neural Network,Arizona State University. (2017)

[42] Hadžiosmanović, D., Bolzoni, D., Hartel, P. H.: A Log Mining Approach for Process Monitoring in SCADA, International Journal of Information Security, vol. 11, no. 4, pp. 231 - 251. (2012)

[43] Hevner, A., Chatterjee, S.: Design Science Research in Information Systems, In Design Research in Information Systems, pp. 9 - 22, Springer. (2010)

[44] Zowghi, D., Coulin, C.: Requirements Elicitation: A Survey of Techniques, Approaches, and Tools, In Engineering and Managing Software Requirements, pp. 19 - 46, Springer. (2005)

[45] Zhang, Z.: Effective Requirements Development - A Comparison of Requirements Elicitation Techniques, Software Quality Management XV: Software Quality in the Knowledge Society, pp. 225 - 240. (2007)

[46] Kajornboon, A. B.: Using Interviews as Research Instruments, E-journal for Research Teachers, vol. 2, no. 1, pp. 1 - 9. (2005)

[47] López, X. S. P.-S.: Advantages of Focus Group, Available: `https://xaperezsindin.com/2013/05/16/four-essentials-of-focus-group` (2014)

[48] Johannesson, P., Perjons, E.: An Introduction to Design Science, Springer. (2014)

[49] March, S. T., Smith, G. F.: Design and Natural Science Research on Information Technology, Decision Support Systems, vol. 15, no. 4, pp. 251 - 266. (1995)

[50] Offermann, P., Levina, O., Schönherr, M., Bub, U.: Outline of a Design Science Research Process, In Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, pp. 1 - 11, ACM. (2009)

[51] Vaishnavi, V., Kuechler, W.: Design Science Research in Information Systems, vol. 20, pp. 1 - 62. (2004)

[52] Pries-Heje, J., Baskerville, R., Venable, J. R.: Strategies for Design Science Research Evaluation, European Conference on Information Systems (ECIS) 2008 Proceedings. (2008)

[53] Sonnenberg, C., Vom Brocke, J.: Evaluation Patterns for Design Science Research Artefacts, In European Design Science Symposium, pp. 71 - 83, Springer. (2011)

[54] Cleven, A., Gubler, P., Hüner, K. M.: Design Alternatives for the Evaluation of Design Science Research Artifacts, In Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, pp. 1 - 8, ACM. (2009)

[55] Peffers, K., Rothenberger, M., Tuunanen, T., Vaezi, R.: Design Science Research Evaluation, In International Conference on Design Science Research in Information Systems, pp. 398 - 410, Springer. (2012)

[56] Venable, J., Pries-Heje, J., Baskerville, R.: A Comprehensive Framework for Evaluation in Design Science Research, In International Conference on Design Science Research in Information Systems, pp. 423 - 438, Springer. (2012)

[57] Ben-Gal, I.: Bayesian Networks. Encyclopedia of Statistics in Quality and Reliability. John Wiley & Sons, Ltd. (2008)

[58] Bhandari, J., Abbassi, A., Garaniya, V., Khan, F.: Risk Analysis of Deepwater Drilling Operations Using Bayesian Network, Journal of Loss Prevention in the Process Industries, vol. 38, pp. 11 - 23. (2015)

[59] Darwiche, A.: Chapter 11 - Bayesian Networks. In: Foundations of Artificial Intelligence, vol. 3, pp. 467 - 509. (2008)

[60] Khakzad, N., Khan, F., Amyotte, P.: Safety Analysis in Process Facilities: Comparison of Fault Tree and Bayesian Network Approaches, Reliability Engineering & System Safety, vol. 96, no. 8, pp. 925 - 932. (2011)

[61] Kahn Jr, C. E., Roberts, L. M., Shaffer, K. A., Haddawy, P.: Construction of a Bayesian Network for Mammographic Diagnosis of Breast Cancer, Computers in Biology and Medicine, vol. 27, no. 1, pp. 19 - 29. (1997)

[62] Curiac, D., Vasile, G., Banias, O., Volosencu, C., Albu, A.: Bayesian Network Model for Diagnosis of Psychiatric Diseases, Information Technology Interfaces, 2009. ITI'09. Proceedings of the ITI 2009 31st International Conference on, pp. 61 - 66, IEEE. (2009)

[63] Oniśko, A., Druzdzel, M. J., Wasyluk, H.: A Bayesian Network Model for Diagnosis of Liver Disorders, In Proceedings of the Eleventh Conference on Biocybernetics and Biomedical Engineering, vol. 2, pp. 842 - 846, Citeseer. (1999)

[64] Milho, I., Fred, A.: A User-friendly Development Tool for Medical Diagnosis based on Bayesian Networks, In Enterprise Information Systems II, pp. 113 - 118, Springer. (2001)

[65] Kahn, C. E., Laur, J. J., Carrera, G.: A Bayesian Network for Diagnosis of Primary Bone Tumors, Journal of Digital Imaging, vol. 14, no. 1, pp. 56 - 57. (2001)

[66] Luciani, D., Marchesi, M., Bertolini, G.: The Role of Bayesian Networks in the Diagnosis of Pulmonary Embolism, Journal of Thrombosis and Haemostasis, vol. 1, no. 4, pp. 698 - 707. (2003)

[67] Huang, Y., McMurran, R., Dhadyalla, G., Jones, R.P.: Probability Based Vehicle Fault Diagnosis: Bayesian Network Method, Journal of Intelligent Manufacturing, vol. 19, no. 3, pp. 301 - 311, Springer. (2008)

[68] Zhao, Y., Xiao, F., Wang, S.: An Intelligent Chiller Fault Detection and Diagnosis Methodology using Bayesian Belief Network, Energy and Buildings, vol. 57, pp. 278 - 288. (2013)

[69] Cai, B., et al.: Multi-source Information Fusion based Fault Diagnosis of Ground-source Heat Pump using Bayesian Network, Applied Energy, vol. 114, pp. 1 - 9. (2014)

[70] Kwan, M., Chow, K.-P., Law, F., Lai, P.: Reasoning about Evidence using Bayesian Networks. In: IFIP International Conference on Digital Forensics, pp. 275 - 289, Springer. (2008)

[71] Axelrad, E.T., Sticha, P.J., Brdiczka, O., Shen, J.: A Bayesian Network Model for Predicting Insider Threats. In: Security and Privacy Workshops, pp. 82 - 89. (2013)

[72] Greitzer, F.L., et al.: Identifying at-risk Employees: Modeling Psychosocial Precursors of Potential Insider Threats. In: System Science (HICSS), Hawaii International Conference on, pp. 2392 - 2401. (2012)

[73] Greitzer, F.L., et al.: Identifying at-risk Employees: A Behavioral Model for Predicting Potential Insider Threats. Pacific Northwest National Laboratory. (2010)

[74] Pecchia, A., et al.: Identifying Compromised Users in Shared Computing Infrastructures: A Data-driven Bayesian Network Approach. In: Reliable Distributed Systems (SRDS), 2011 30th IEEE Symposium on, pp. 127 - 136. IEEE. (2011)

[75] Shin, J., Son, H., Heo, G.: Development of a Cyber Security Risk Model using Bayesian Networks. Reliability Engineering & System Safety, vol. 134, pp. 208 - 217. (2015)

[76] Kornecki, A.J., Subramanian, N., Zalewski, J.: Studying Interrelationships of Safety and Security for Software Assurance in Cyber-Physical Systems: Approach based on Bayesian Belief Networks. In: Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on, pp. 1393 - 1399, IEEE. (2013)

[77] Wang, J.A., Guo, M.: Vulnerability Categorization using Bayesian Networks. In: Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research, pp. 1 - 4, ACM. (2010)

[78] Mo, S.Y.K., Beling, P.A., Crowther, K.G.: Quantitative Assessment of Cyber Security Risk using Bayesian Network-based Model. In: Systems and Information Engineering Design Symposium, 2009. SIEDS'09., pp. 183 - 187, IEEE. (2009)

[79] Holm, H., Korman, M., Ekstedt, M.: A Bayesian Network Model for Likelihood Estimations of Acquirement of Critical Software Vulnerabilities and Exploits. Information and Software Technology, vol. 58, pp. 304 - 318. (2015)

[80] Kwan, M., et al.: Analysis of the Digital Evidence Presented in the Yahoo! Case. In: IFIP International Conference on Digital Forensics, pp. 241 - 252, Springer. (2009)

[81] Ibrahimović, S., Bajgorić, N.: Modeling Information System Availability by using Bayesian Belief Network Approach. Interdisciplinary Description of Complex Systems, vol. 14, pp. 125 - 138. (2016)

[82] Herland, K., Hammainen, H., Kekolahti, P.: Information Security Risk Assessment of Smartphones using Bayesian Networks. Journal of Cyber Security and Mobility, vol. 4, pp. 65 - 85. (2016)

[83] Apukhtin, V.: Bayesian Network Modeling for Analysis of Data Breach in a Bank. University of Stavanger, Norway. (2011)

[84] Martin, T.G., et al.: Eliciting Expert Knowledge in Conservation Science, Conservation Biology, vol. 26, no. 1, pp. 29 - 38. (2012)

2

INTEGRATED SAFETY AND SECURITY RISK ASSESSMENT METHODS: A SURVEY OF KEY CHARACTERISTICS AND APPLICATIONS*

2.1. INTRODUCTION

Information technologies and communication devices are increasingly being integrated into modern control systems [1]. These modern control systems are used to operate life-critical systems where the human lives are at stake in case of failure. At the same time, they are often vulnerable to cyber-attacks, which may cause physical impact. An incident in Lodz is a typical example where a cyber-attack resulted in the derailment of 4 trams, and the injury of 12 people [2]. It is therefore becoming increasingly important to address the combination of safety and security in modern control systems.

However, safety and security have been represented by separate communities in both academia and industry [3]. In our context, we think of the safety community as dealing with unintentional/non-malicious threats caused by natural disasters, technical failures, and human error. On the other hand, we think of the security community as dealing with intentional/malicious threats caused by intentional human behavior.

Risk management plays a major role in dealing with both unintentional/non-malicious, and intentional/malicious threats. In the recent years, we have seen a trans-

*This chapter has been published as Chockalingam, S., Hadžiosmanović, D., Pieters, W., Teixeira, A., and van Gelder, P.: *"Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications,"* International Conference on Critical Information Infrastructures Security, pp. 50 – 62, 2016. Springer, Cham. https://doi.org/10.1007/978-3-319-71368-7_5

formation among the researchers of safety and security community to work together especially in risk management. As an example, there are developments of integrated safety and security risk assessment methods [4–10]. Risk assessment is one of the most crucial parts of the risk management process as it is the basis for making risk treatment decisions [11]. The integrated safety and security risk assessment method helps to improve the completeness of risk assessment conducted by covering the interactions between malicious and non-malicious risks. However, a comprehensive review of integrated safety and security risk assessment methods which could help to identify their key characteristics and applications is lacking. Therefore, this research aims to fill this gap by addressing the research question: "What are the key characteristics of integrated safety and security risk assessment methods, and their applications?". The research objectives are:

- **RO 1.** To identify integrated safety and security risk assessment methods.
- **RO 2.** To identify key characteristics and applications of integrated safety and security risk assessment methods based on the analysis of identified methods.

The scope of this analysis covers important features of identified integrated safety and security risk assessment methods mainly, in terms of how these methods are created, and what the existing applications of these methods are. The analysis of identified methods is performed based on the following criteria: I. Citations in the Scientific Literature, II. Steps Involved, III. Stage(s) of Risk Assessment Process Addressed, IV. Integration Methodology, and V. Application(s) and Application Domain. The motivations for selecting these criteria are described in Section 2.5.

The remainder of this chapter is structured as follows: Section 2.2 describes the related work, followed by the review methodology in Section 2.3. In Section 2.4, we present the identified integrated safety and security risk assessment methods, and describe the steps involved in these methods. In Section 2.5, we perform the analysis of identified methods based on the criteria that we defined above. Finally, we highlight key characteristics and applications of integrated safety and security risk assessment methods followed by a discussion of future work directions in Section 2.6.

2.2. RELATED WORK

Cherdantseva et al. presented 24 cybersecurity risk assessment methods for Supervisory Control and Data Acquisition (SCADA) systems [12]. In addition, they analyzed the presented methods based on the following criteria: I. Aim, II. Application domain, III. Stages of risk management addressed, IV. Key concepts of risk management covered, V. Impact measurement, VI. Sources of data for deriving probabilities, VII. Evaluation method, and VIII. Tool support. Based on the analysis, they suggested the following categorization schemes: I. Level of detail and coverage, II. Formula-based vs. Model-based, III. Qualitative vs. Quantitative, and IV. Source of probabilistic data. However, Cherdantseva et al. did not present integrated safety and security risk assessment methods. We used and complemented some of the criteria provided by Cherdantseva et al. to perform the analysis of integrated safety and security risk assessment methods as described in Section 2.5.

Risk assessment methods like Failure Mode and Effects Analysis (FMEA) [13], Fault Tree Analysis (FTA) [14], Component Fault Tree (CFT) [15] have been used by safety community whereas the risk assessment methods like Attack Trees [16], Attack-Countermeasure Trees (ACT) [17], National Institute of Standards and Technology (NIST) 800-30 Risk Assessment [18] have been used by security community. Several authors used these methods as a starting point for the development of integrated safety and security risk assessment methods.

Kriaa et al. highlighted standard initiatives such as ISA-99 (Working Group 7), IEC TC65 (Ad Hoc Group 1), IEC 62859, DO-326/ED-202 that consider safety and security co-ordination for Industrial Control Systems (ICS) [1]. They described various generic approaches that considered safety and security at a macroscopic level of system design or risk evaluation, and also model-based approaches that rely on a formal or semi-formal representation of the functional/non-functional aspects of system. They classified the identified approaches based on the following criteria: I. Unification vs. Integration, II. Development vs. Operational, and III. Qualitative vs. Quantitative. However, Kriaa et al. did not primarily focus on integrated safety and security risk assessment methods that have been already applied in at least one real-case/example involving control system. Also, Kriaa et al. did not identify key characteristics and applications of integrated safety and security risk assessment methods. We included methods such as Failure Mode, Vulnerabilities, and Effect Analysis (FMVEA) [7], Extended Component Fault Tree (CFT) [9], and Extended Fault Tree (EFT) [10] from Kriaa et al. in our work as they satisfy our selection criteria. In addition, we included other methods that satisfy our selection criteria, such as Security-Aware Hazard Analysis and Risk Assessment (SAHARA) [4], Combined Harm Assessment of Safety and Security for Information Systems (CHASSIS) [5], Failure-Attack-CountTermeasure (FACT) Graph [6], and Unified Security and Safety Risk Assessment [8].

2.3. REVIEW METHODOLOGY

This section describes the methodology for selecting the integrated safety and security risk assessment methods. The selection of these methods mainly consists of two stages:

- Searches were performed on IEEE Xplore Digital Library, ACM Digital Library, Scopus, DBLP, and Web of Science – All Databases. The search-strings were constructed from keywords “Attack”, “Failure”, “Hazard”, “Integration”, “Risk”, “Safety”, “Security”, and “Threat”. DBLP provided a good coverage of relevant journals and conferences.

- Methods were selected from the search results according to the following criteria:

  I. The method should address any or all of the following risk assessment stages: risk identification, risk analysis, and/or risk evaluation.

  II. The method should consider both unintentional and intentional threats.

  III. The method should have been already applied in at least one real-case/example involving control system.

  IV. The literature should be in English language.

Once an integrated safety and security risk assessment method was selected, the scientific literature that cited it was also traced.

2.4. INTEGRATED SAFETY AND SECURITY RISK ASSESSMENT METHODS

This section presents the identified integrated safety and security risk assessment methods, and describes the steps involved in these methods. This section aims to address the RO 1. Based on the review methodology described in Section 2.3, we have identified 7 integrated safety and security risk assessment methods: I. SAHARA [4], II. CHASSIS [5], III. FACT Graph [6], IV. FMVEA [7], V. Unified Security and Safety Risk Assessment [8], VI. Extended CFT [9], and VII. EFT [10].

2.4.1. SAHARA METHOD

The steps involved in the SAHARA method [4] are as follows: I. The ISO 26262 – Hazard Analysis and Risk Assessment (HARA) approach is used in a conventional manner to classify the safety hazards according to the Automotive Safety Integrity Level (ASIL), and to identify the safety goal and safe state for each identified potential hazard; II. The attack vectors of the system are modelled. The STRIDE method is used to model the attack vectors of the system [4, 19]; III. The security threats are quantified according to the Required Resources (R), Required Know-how (K), and Threat Criticality (T); IV. The security threats are classified according to the Security Level (SecL). SecL is determined based on the level of R, K, and T; V. Finally, the security threats that may violate the safety goals (T>2) are considered for the further safety analysis.

2.4.2. CHASSIS METHOD

The steps involved in the CHASSIS method [5] are as follows: I. The elicitation of functional requirements which involve creating the use-case diagrams that incorporates the users, system functions and services; II. The elicitation of safety and security requirements which involve creating misuse case diagram based on the identified scenarios for safety and security involving faulty-systems and attackers respectively; III. Trade-off discussions are used to support the resolution of conflict between the safety, and security mitigations.

2.4.3. FACT GRAPH METHOD

The steps involved in the FACT Graph method [6] are as follows: I. The fault trees of the system analyzed are imported to start the construction of FACT graph; II. The safety countermeasures are attached to the failure nodes in the FACT graph; III. The attack trees of the system analyzed are imported to the FACT graph in construction. This is done by adding an attack-tree to the failure node in the FACT graph with the help of OR gate, if the particular failure may also be caused by an attack; IV. The security countermeasures are attached to the attack nodes in the FACT graph. This could be done based on the ACT technique [17].

2.4.4. FMVEA METHOD

The steps involved in the FMVEA method [7] are as follows: I. A functional analysis at the system level is performed to get the list of system components and functions of each component; II. A component that needs to be analyzed from the list of system components is selected; III. The failure/threat modes for the selected component are identified; IV. The failure/threat effect for each identified failure/threat mode is identified; V. The severity for the identified failure/threat effect is determined; VI. The potential failure causes/vulnerabilities/threat agents are identified; VII. The failure/attack probability is determined. Schmittner et al. described the attack probability as the sum of threat properties and system susceptibility ratings. The threat properties is the sum of motivation and capabilities ratings, whereas the system susceptibility is the sum of reachability and unusualness of the system ratings; VIII. Finally, the risk number is determined, which is the product of severity rating and failure/attack probability.

2.4.5. UNIFIED SECURITY AND SAFETY RISK ASSESSMENT METHOD

The steps involved in the Unified Security and Safety Risk Assessment method [8] are as follows: I. The system boundary, system functions, system and data criticality, system and data sensitivity are identified; II. The threats, hazards, vulnerabilities, and hazard-initiating events are identified; III. The current and planned controls are identified; IV. The threat likelihood is determined; V. The hazard likelihood is determined; VI. The asset impact value is determined; VII. The combined safety-security risk level is determined; VIII. The control recommendations are provided; IX. The risk assessment reports are provided.

2.4.6. EXTENDED CFT METHOD

The steps involved in the extended CFT method [9] are as follows: I. The CFT for the system analyzed is developed. This could be done based on [15]; II. The CFT is extended by adding an attack tree to the failure node with the help of OR gate, if the particular event may also be caused by an attack; III. The qualitative analysis is conducted by calculating Minimal Cut Sets (MCSs) per top level event. MCSs containing only one event would be single point of failure which should be avoided; IV. The quantitative analysis is conducted by assigning values to the basic events. Therefore, MCSs containing only safety events would have a probability P, MCSs containing only security events would have a rating R, MCSs containing both safety and security events would have a tuple of probability and rating (P, R).

2.4.7. EFT METHOD

The steps involved in the EFT method [10] are as follows: I. The fault tree for the system analyzed is developed by taking into account the random faults; II. The developed fault tree is extended by adding an attack tree to the basic or intermediate event in the fault tree, if the particular event in the fault tree may also be caused by malicious actions. The attack tree concept used in the development of EFT is based on [20]; III. The quantitative analysis is performed based on the formulae defined in [10] which help to calculate the top event probability.

2.5. Analysis of Integrated Safety and Security Risk Assessment Methods

This section performs the analysis of integrated safety and security risk assessment methods based on the criteria: I. Citations in the Scientific Literature, II. Steps Involved, III. Stage(s) of Risk Assessment Process Addressed, IV. Integration Methodology, and V. Application(s) and Application Domain. This allows us to identify key characteristics and applications of integrated safety and security risk assessment methods. This section aims to address the RO 2.

The integrated safety and security risk assessment methods described in the previous section are listed in Table 2.1. In Table 2.1, country is the country of the first author of the paper and citations is the number of citations of the paper according to Google Scholar Citation Index as on 31st August 2016.

From Table 2.1, we observe that the researchers started to recognize the importance of integrated safety and security risk assessment methods which resulted in the increase in number of papers produced especially during 2014, and 2015. The largest number of citations (63) is acquired by the EFT method published in 2009. The second most cited paper, among analyzed, with 17 citations, is the Extended CFT method published in 2013. However, it is understandable that the methods published during the last few years received lower number of citations ranging from 1 to 5.

Table 2.1: List of Integrated Safety and Security Risk Assessment Methods (Ordered by the number of citations)

Integrated Safety and Security Risk Assessment Method	**Year**	**Country**	**Citations**
EFT [10]	2009	Italy	63
Extended CFT [9]	2013	Germany	17
FACT Graph [6]	2015	Singapore	5
CHASSIS [5]	2015	Austria	4
FMVEA [7]	2014	Austria	4
SAHARA [4]	2015	Austria	2
Unified Security and Safety Risk Assessment [8]	2014	Taiwan	1

Based on the steps involved in each method as described in Section 2.4, we conclude that there are two types of integrated safety and security risk assessment methods:

- Sequential Integrated Safety and Security Risk Assessment Method: In this type of method, the safety risk assessment, and security risk assessment are performed in a particular sequence. For instance, the Extended CFT method starts with the development of CFT for the system analyzed. Later, the attack tree is added to extend the developed CFT. This method starts with the safety risk assessment followed by the security risk assessment. Methods such as SAHARA, FACT Graph, Unified Security and Safety Risk Assessment, Extended CFT, and EFT come under the sequential type.

- Non-sequential Integrated Safety and Security Risk Assessment Method: In this type of method, the safety risk assessment, and security risk assessment are performed

without any particular sequence. For instance, in the FMVEA method, the results of safety risk assessment and security risk assessment are tabulated in the same table without any particular sequence. Methods such as FMVEA and CHASSIS come under the non-sequential type.

Cherdantseva et al. used 'stage(s) of risk management process addressed' as a criteria to analyze the identified cybersecurity risk assessment methods for SCADA systems [12]. We adapted and used this criteria as 'stage(s) of risk assessment process addressed' because the major focus of our research is on risk assessment. This criteria will allow us to identify the predominant stage(s) of risk assessment process addressed by the integrated safety and security risk assessment methods.

A risk assessment process consists of typically three stages:

- Risk Identification: This is the process of finding, recognizing and describing the risks [21].
- Risk Analysis: This is the process of understanding the nature, sources, and causes of the risks that have been identified and to estimate the level of risk [21].
- Risk Evaluation: This is the process of comparing risk analysis results with risk criteria to make risk treatment decisions [21].

Table 2.2 highlights the integrated safety and security risk assessment method and the corresponding stage(s) of the risk assessment process addressed. This is done based on the definitions of risk identification, risk analysis, and risk evaluation. We also take into account the safety risk assessment method, and security risk assessment method that were combined in the integrated safety and security risk assessment method.

Table 2.2: Stage(s) of Risk Assessment Process Addressed

Integrated Safety and Security Risk Assessment Method	**Risk Identification**	**Risk Analysis**	**Risk Evaluation**
SAHARA	✓	✓	×
CHASSIS	✓	×	×
FACT Graph	✓	×	×
FMVEA	✓	✓	×
Unified Security and Safety Risk Assessment	✓	✓	✓
Extended CFT	✓	✓	×
EFT	✓	✓	×

In Table 2.2, ✓(×) indicates that the particular method addressed (did not address) the corresponding risk assessment stage.

From Table 2.2, we understand that all methods addressed the risk identification, 5 out of 7 methods addressed the risk analysis, whereas only 1 out of 7 methods addressed the risk evaluation stage of the risk assessment process. This implies that the risk evaluation stage is not given much attention compared to the other stages of the risk assessment process in the integrated safety and security risk assessment methods. Cherdantseva et

al. also highlighted that the majority of the cybersecurity risk assessment methods for SCADA systems concentrates on the risk identification and risk analysis stages of the risk assessment process [12]. The risk evaluation phase in the Unified Security and Safety Risk Assessment method starts by comparing the risk analysis result with the suggested four levels of risk to determine the appropriate level of risk. Once the level of risk is determined, the risk treatment decision is made accordingly.

We used the criteria 'Integration methodology' because this will allow us to understand which combination of safety, and security risk assessment methods are being used in the integrated safety and security risk assessment methods as summarized in Table 2.3.

From Table 2.3, we observe that there are four ways in which the integrated safety and security risk assessment methods have been developed:

- Integration through the combination of a conventional safety risk assessment method and a variation of the conventional safety risk assessment method for security risk assessment. The methods SAHARA and FMVEA come under this category.
- Integration through the combination of a conventional security risk assessment method and a variation of the conventional security risk assessment method for safety risk assessment. The Unified Security and Safety Risk Assessment method come under this category.
- Integration through the combination of a conventional safety risk assessment method and a conventional security risk assessment method. The methods FACT Graph, Extended CFT, and EFT come under this category.
- Others - There is no conventional safety risk assessment, and conventional security risk assessment method used in the integration. The CHASSIS method come under this category. The CHASSIS method used a variation of Unified Modeling Language (UML)-based models for both the safety and security risk assessment.

Table 2.3: Integration Methodology

Integrated Safety and Security Risk Assessment Method	**Safety Risk Assessment Method**	**Security Risk Assessment Method**
SAHARA	ISO 26262: HARA	Variation of ISO 26262: HARA
CHASSIS	Safety Misuse Case (Involving Faulty-systems)	Security Misuse Case (Involving Attackers)
FACT Graph	Fault Tree	Attack Tree
FMVEA	FMEA	Variation of FMEA
Unified Security and Safety Risk Assessment	Variation of NIST 800-30 Security Risk Estimation	NIST 800-30 Security Risk Estimation
Extended CFT	CFT	Attack Tree
EFT	Fault Tree	Attack Tree

We used the criteria 'Application(s) and Application domain' because this will allow us to understand the type of application(s), and the corresponding application domain of integrated safety and security risk assessment methods. Table 2.4 highlights the integrated safety and security risk assessment method and the corresponding application(s) and application domain.

Table 2.4: Application(s) and Application Domain

Integrated Safety and Security Risk Assessment Method	**Application(s)**	**Application Domain**
SAHARA	Battery Management System use-case [4]	Transportation
CHASSIS	Over The Air (OTA) system [5], Air traffic management remote tower example [23].	Transportation
FACT Graph	Over-pressurization of a vessel example [6]	Power and Utilities
FMVEA	OTA system [5], Telematics control unit [7], Engine test-stand [24], Communications-based train control system [25].	Transportation
Unified Security and Safety Risk Assessment	High pressure core flooder case-study [8]	Power and Utilities
Extended CFT	Adaptive cruise control system [9]	Transportation
EFT	Release of toxic substance into the environment example [10]	Chemical

From Table 2.4, we observe that 4 methods were applied in the transportation domain, 2 methods were applied in the power and utilities domain, and 1 method was applied in the chemical domain. The major development, and application of integrated safety and security risk assessment methods, is in the transportation domain. The Threat Horizon 2017 listed "death from disruption to digital services" as one of the threats especially in the transportation and medical domain [22]. In the transportation domain, there is a potential for cyber-attacks which compromises system safety and result in the injury/death of people which was illustrated by a tram incident in Lodz [2].

2.6. CONCLUSIONS AND FUTURE WORK

In this study, we have identified 7 integrated safety and security risk assessment methods. Although we cannot completely rule out the existence of other unobserved integrated safety and security risk assessment methods that fulfil our selection criteria, the review methodology that we adopted helped to ensure the acceptable level of completeness in the selection of these methods. Based on the analysis, we identified key characteristics and applications of integrated safety and security risk assessment methods.

- There are two types of integrated safety and security risk assessment methods

based on the steps involved in each method. They are: a. Sequential, and b. Non-sequential.

- There are four ways in which the integrated safety and security risk assessment methods have been developed. They are: a. The conventional safety risk assessment method as the base and a variation of the safety risk assessment method for security risk assessment, b. The conventional security risk assessment method as the base and a variation of the security risk assessment method for safety risk assessment, c. A combination of a conventional safety risk assessment method, and a conventional security risk assessment method, d. Others.

- Risk identification and risk analysis stages were given much attention compared to the risk evaluation stage of the risk assessment process in the integrated safety and security risk assessment methods.

- Transportation, power and utilities, and chemical were the three domains of application for integrated safety and security risk assessment methods.

The identified integrated safety and security risk assessment methods did not take into account real-time system information to perform dynamic risk assessment which needs to be addressed to make it more effective in the future. This study provided the list of combinations of safety, and security risk assessment methods used in the identified integrated safety and security risk assessment methods. In the future, this would act as a base to investigate the other combinations of safety, and security risk assessment methods that could be used in the development of more effective integrated safety and security risk assessment methods. Furthermore, this study provided the type of applications and application domains of the identified integrated safety and security risk assessment methods. In the future, this would act as a starting point to evaluate the applicability of these methods in the other domains besides transportation, power and utilities, and chemical.

REFERENCES

[1] Kriaa, S., Pietre-Cambacedes, L., Bouissou, M., Halgand, Y.: A Survey of Approaches Combining Safety and Security for Industrial Control Systems. Reliability Engineering & System Safety. vol. 139, pp. 156 – 178. (2015)

[2] RISI Database.: Schoolboy Hacks into Polish Tram System (2016).http://www.risidata.com/Database/Detail/schoolboy_hacks_into_polish_tram_system

[3] Stoneburner, G.: Toward a Unified Security-Safety Model. Computer. vol. 39, no. 8, pp. 96 – 97. (2006)

[4] Macher, G., Höller, A., Sporer, H., Armengaud, E., Kreiner, C.: A Combined Safety – Hazards and Security - Threat Analysis Method for Automotive Systems. Koornneef, F., van Gulijk, C. (eds.) SAFECOMP 2015 Workshops. LNCS, vol. 9338, pp. 237 – 250. Springer, Heidelberg (2015)

[5] Schmittner, C., Ma, Z., Schoitsch, E., Gruber, T.: A Case Study of FMVEA and CHASSIS as Safety and Security Co-Analysis Method for Automotive Cyber Physical Systems. In: Proceedings of the 1st ACM Workshop on Cyber Physical System Security (CPSS), pp. 69 – 80. (2015)

[6] Sabaliauskaite, G., Mathur, A.P.: Aligning Cyber-physical System Safety and Security. Cardin, M.A., Krob, D., Cheun, L.P., Tan, Y.H., Wood, K. (eds.) Complex Systems Design & Management Asia 2014. LNCS, pp. 41 – 53. (2015)

[7] Schmittner, C., Ma, Z., Smith, P.: FMVEA for Safety and Security Analysis of Intelligent and Cooperative Vehicles. Bondavalli, A., Ceccarelli, A., Ortmeier, F. (eds.) SAFECOMP 2014 Workshops. LNCS, vol. 8696, pp. 282 – 288. Springer, Heidelberg (2014)

[8] Chen, Y., Chen, S., Hsiung, P., Chou, I.: Unified Security and Safety Risk Assessment – A Case Study on Nuclear Power Plant. In: Proceedings of the International Conference on Trusted Systems and their Applications (TSA), pp. 22 – 28. (2014)

[9] Steiner, M., Liggesmeyer, P., Combination of Safety and Security Analysis – Finding Security Problems that Threaten the Safety of a System. In: Workshop on Dependable Embedded and Cyber-physical Systems (DECS), pp. 1 – 8. (2013)

[10] Fovino, I.N., Masera, M., De Cian, A., Integrating Cyber Attacks within Fault Trees. Reliability Engineering and System Safety. vol. 94, no. 9, pp. 1394 – 1402. (2009)

[11] European Union Agency for Network and Information Security (ENISA). The Risk Management Process (2016). https://www.enisa.europa.eu/activities/risk-management/current-risk/risk-management-inventory/rm-process

[12] Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., Stoddart, K.: A Review of Cyber Security Risk Assessment Methods for SCADA Systems. Computers & Security. vol. 56, pp. 1 – 27. (2016)

[13] International Electrotechnical Commission (IEC).: IEC 60812: Analysis Techniques for System Reliability – Procedures for Failure Mode and Effects Analysis. (2006)

[14] Lee, W.S., Grosh, D.L., Tillman, F.A., Lie, C.H.: Fault Tree Analysis, Methods, and Applications – A Review. IEEE Transactions on Reliability. vol. 34, no. 3, pp. 194 – 203. (1985)

[15] Kaiser, B., Liggesmeyer, P., Mackel, O.: A New Component Concept for Fault Trees. In: Proceedings of the 8th Australian Workshop on Safety Critical Systems and Software (SCS), vol. 33, pp. 37 – 46. (2003)

[16] Schneier, B.: Attack Trees. Dr. Dobb's Journal. vol. 24, no. 12, pp. 21 – 29. (1999)

[17] Roy, A., Kim, D.S., Trivedi, K.S.: Scalable Optimal Countermeasure Selection Using Implicit Enumeration on Attack Countermeasure Trees. In: Proceedings of the 42nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 1 – 12. (2012)

[18] National Institute of Standards and Technology (NIST).: Risk Management Guide for Information Technology Systems. (2002)

[19] Scandariato, R., Wuyts, K., Joosen, W.: A Descriptive Study of Microsoft's Threat Modeling Technique. Requirements Engineering. vol. 20, no. 2, pp. 163-180. (2015)

[20] Fovino, I.N., Masera, M.: Through the Description of Attacks: A Multi-Dimensional View. Gorski, J. (eds.) SAFECOMP 2006. LNCS, vol. 4166, pp. 15 – 28. Springer, Heidelberg (2006)

[21] International Organisation for Standardization (ISO).: ISO 31000: 2009 - Risk Management – Principles and Guidelines. (2009)

[22] Information Security Forum.: Threat Horizon 2017: Dangers Accelerate (2015). https://www.securityforum.org/uploads/2015/03/Threat-Horizon_2017_Executive-Summary.pdf

[23] Raspotnig, C., Karpati, P., Katta, V.: A Combined Process for Elicitation and Analysis of Safety and Security Requirements. Bider, I., Halpin, T., Krogstie, J., Nurcan, S., Proper, E., Schmidt, R., Soffer, P., Wrycza, S. (eds.) BPMDS and EMMSAD 2012. LNBIP, vol. 113, pp. 347 – 361. Springer, Heidelberg (2012)

[24] Schmittner, C., Gruber, T., Puschner, P., Schoitsch, E.: Security Application of Failure Mode and Effect Analysis (FMEA). Bondavalli, A., Di Giandomenico, F. (eds.) SAFECOMP 2014. LNCS, vol. 8666, pp. 310 – 325. (2014)

[25] Chen, B., Schmittner, C., Ma, Z., Temple, W.G., Dong, X., Jones, D.L., Sanders, W.H.: Security Analysis of Urban Railway Systems: The Need for a Cyber-Physical Perspective. Koornneef, F., van Gulijk, C. (eds.) SAFECOMP 2015 Workshops. LNCS, vol. 9338, pp. 277 – 290. Springer, Heidelberg (2015)

3

BAYESIAN NETWORK MODELS IN CYBER SECURITY: A SYSTEMATIC REVIEW*

3.1. INTRODUCTION

The lack of data, especially historical data on cyber security breaches, incidents, and threats, hinders the development of realistic models in cyber security [1, 2]. However, standard (or classical) Bayesian Networks (BNs) possess the potential to address this challenge. In particular, the capability to combine different sources of knowledge would help to overcome the scarcity of historical data in cyber security modeling.

Standard BNs belong to the family of probabilistic graphical models [3]. A standard BN consists of two components: qualitative, and quantitative [4]. The qualitative part is a Directed Acyclic Graph (DAG) consisting of nodes and edges. Specifically, each node represents a random variable, whereas the edges between the nodes represent the conditional dependencies among the corresponding random variables. The quantitative part takes the form of conditional probabilities, which quantify the dependencies between connected nodes in the DAG by specifying a conditional probability distribution for each node. A toy example of a standard BN model, representing the probabilistic relationships between cyber-attacks ("Denial of Service Attack" and "Malware Attack") and symptoms ("Internet Connection" and "Pop-ups"), is shown in Figure 3.1. Given symptom(s), the BN can be used to compute the posterior probabilities of various cyber-attacks as shown in Figure 3.1. In this case, the user sets evidence for the "Pop-ups" node as "True", and "Internet Connection" node as "Normal" in the BN model based on his/her observations. Based on these evidences, the BN computes the posterior probabilities of the other nodes "Denial of Service Attack" and "Malware Attack" using Bayes rule. The BN model shown

*This chapter has been published as Chockalingam, S., Pieters, W., Teixeira, A., and van Gelder, P.: *"Bayesian Network Models in Cyber Security: A Systematic Review,"* Nordic Conference on Secure IT Systems, pp. 105 – 122, 2017. Springer, Cham. https://doi.org/10.1007/978-3-319-70290-2_7

in Figure 3.1 determines that the presence of pop-ups and normal internet connection are more likely due to a Denial of Service attack rather than to a Malware attack.

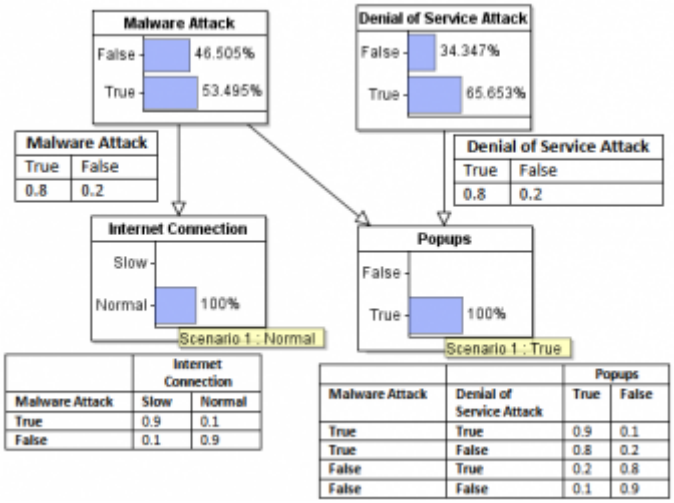


Figure 3.1: Standard BN Model - Example

The major advantages of standard BNs include: the ability to combine different sources of knowledge, the capacity to handle small and incomplete datasets, and the availability of a broad range of validation approaches apart from data-driven validation approaches [5, 6]. Some notable real-world applications of standard BNs include medical diagnosis [7] and fault diagnosis [8]. In addition, the advantages lead to the predominant use of standard BNs in domains where there is a limited availability of data, notably in Ecosystem Services (ESS)[5], water resource management [9], and security [10]. Similarly, we have seen the use of standard BNs in cyber (or information) security in recent years [11–29]. However, an overarching comparison and analysis of standard BN models in cyber security which could help to identify important usage patterns is currently lacking. Kordy et al. give a broader overview of modeling approaches based on DAGs, and thus only briefly mention BNs [10]. In contrast to Kordy et al., we specifically focus on BN models with the aim of performing comparison and analysis of these models to identify important usage patterns and key research gaps. This review would benefit the practical application of BN models in cyber security by providing important usage patterns and key research gaps. Therefore, this research aims to fill this gap by addressing the research question: "What are the important patterns in the use of standard Bayesian Network (BN) models in cyber security?". The research objectives are:

- **RO 1.** To identify standard BN models in cyber security literature.
- **RO 2.** To identify the important patterns in the use of standard BN models in cyber security based on the analysis of identified models.

In this study, we focus on comparison and analysis of standard BN models [11–29] which also include Bayesian Attack Graphs (BAGs) [11–13] as they possess more comparable features. This would help to identify consistent patterns in the use of standard BN

models in cyber security. However, the approaches in cyber security modeling that extend BN such as Bayesian Decision Network (BDN) [30], Causal event graph [31], Dynamic BN [32–34], Extended influence diagram [35, 36], and Multi-entity BN [37, 38] are beyond the scope of this study as they are incomparable especially based on their structure development. For instance, decision and utility nodes are specific to BDN/Influence Diagram which would allow decision making under uncertainty. In contrast, these types of nodes are not applicable to standard BN.

The scope of this comparison and analysis is the structured development, application and validation of the existing standard BN models in cyber security. The comparison and analysis of identified models is performed using the characteristics that were chosen based on related literature and domain-specific objectives as described in Section 3.2. The key contributions of this work are: important patterns in the use of standard BN models in cyber security, and key research gaps in the use of standard BN models in cyber security.

The remainder of this chapter is structured as follows. Section 3.2 describes the review methodology. In Section 3.3, we perform the comparison and analysis of identified BN models using the characteristics that we chose, followed by a discussion on the key findings in Section 3.4. Finally, we highlight important patterns in the use of standard BN models in cyber security followed by future work directions in Section 3.5.

3.2. REVIEW METHODOLOGY

We perform the systematic literature review based on the guidelines provided by Okoli et al. [39]. The methodology which we used to select the standard BN models in cyber security literature and the appropriate characteristics to perform the comparison and analysis of the selected BN models is described below.

The selection of standard BN models in cyber security literature consists of two stages:

- Searches were performed on ACM Digital Library, DBLP, Google Scholar, IEEE Xplore Digital Library, Scopus, and Web of Science – All Databases. Search-strings were constructed from keywords “Bayesian”, “Bayesian Belief Network”, “Bayesian Network”, “BBN”, “BN”, “Cyber*”, “Information*”, and “Security”. The wildcard “*” was used for “Cyber” and “Information” to match all words around these two keywords.

- Models were selected from the search results according to the listed criteria:

  I. The model should employ standard BN.

  II. The model should address problem(s) associated with cyber (or information) security.

  III.The literature should have basic information about both DAGs and Conditional Probability Tables (CPTs). This criterion is important taking into account the scope of our comparison and analysis which is the structured development, application and validation of the existing standard BN models in cyber security.

  IV. The literature should be in English language.

Once a standard BN model in cyber security was selected, the scientific literature that cited it was also traced.

Table 3.1: Adopted Characteristics from Landuyt et al. and Phan et al.

Characteristics used in our Analysis	**Adopted from Landuyt et al.**	**Adopted from Phan et al.**
I. Citation details		✓
II. Data sources used to construct DAGs and populate CPTs	✓	✓
III. The number of nodes used in the model	✓	
IV. Type of threat actor		
V. Application and Application sector		
VI. Scope of variables		
VII. The approach(es) used to validate models	✓	✓
VIII. Model purpose and Type of purpose		

The characteristics used to perform the analysis of the selected BN models were chosen based on related literature and domain-specific objectives as described in Section 3.2 and 3.3. Landuyt et al. presented 47 BN models in ESS published from 2000 to 2012 [5]. In addition, they analysed these models based on 9 characteristics. Similarly, Phan et al. presented 111 BN models in water resource management [9]. Moreover, they analysed these models based on 10 characteristics. We adopted the characteristics from Landuyt et al. and Phan et al. that are generic and relevant to the scope of our analysis, as shown in Table 3.1. Also, we adapted and used the characteristic: *Citation details* provided by Phan et al. to perform the analysis of BN models in cyber security as described in Section 3.3.

3.3. ANALYSIS OF STANDARD BAYESIAN NETWORK MODELS IN CYBER SECURITY

This section aims to address *RO 1. To identify standard BN models in cyber security literature*, and *RO 2. To identify the important patterns in the use of standard BN models in cyber security based on the analysis of identified models.* Based on the methodology described in Section 3.2, we identified 17 standard BN models in cyber security. The corresponding article titles are listed in Table 3.2. Furthermore, this section performs the analysis of identified BN models based on the following characteristics.

- Citation details
- Data sources used to construct DAGs and populate CPTs
- The number of nodes used in the model
- Type of threat actor
- Application and Application Sector
- Scope of variables
- The approach(es) used to validate models
- Model purpose and Type of purpose

3.3.1. CITATION DETAILS

We adapted and used the components of the characteristic *"Citation details"* provided by Phan et al. Specifically, we used an additional component citations in our definition of *"Citation details"* because this will help us to assess the research impact/quality of each BN model [40]. In Table 3.2, citations is the number of citations of the article according to Google Scholar Citation Index as on 15th September 2017. The number of articles covering standard BN model in cyber security varies between 0 and 3 per year. No noticeable increase in the number of papers over time is encountered. The largest number of citations (247) is acquired by Poolsappasit et al. [11] published in 2012. The second most cited paper, among analysed, with 136 citations, is Frigault et al. [12] which is published in 2008. Interestingly, BAG-based standard BN models [11–13] are extensively used compared to the other standard BN models [14–29] in cyber security based on the number of citations.

3.3.2. DATA SOURCES USED TO CONSTRUCT DAGS AND POPULATE CPTS

We used the characteristic *"Data sources used to construct DAGs and populate CPTs"* to identify the type of data sources utilised in the reviewed BN models. We employed the coding scheme provided by Phan et al. as shown in Table 3.2 [9]. From Table 3.2, we observe that 5 out of 17 BN models used only expert knowledge to construct DAGs, whereas 5 out of 17 BN models employed only empirical data to construct DAGs. 7 out 17 BN models made use of both expert knowledge and empirical data to construct DAGs. In particular, 10 out of 12 BN models which utilised empirical data to construct DAGs relied on the literature. In contrast, 2 out of 12 BN models which utilised empirical data to construct DAGs relied on the inputs from vulnerability scanner [11] and incidents data [18].

From Table 3.2, we infer that 11 out of 17 BN models utilised only expert knowledge to populate CPTs, whereas 3 out of 17 BN models used only empirical data to populate CPTs. On the other hand, there were 3 out of 17 BN models which employed both expert knowledge and empirical data to populate CPTs. Specifically, the sources of empirical data includes literature, incidents data, National Vulnerability Database (NVD), Open Source Vulnerability Database (OSVDB), and Exploithub to populate CPTs. Notably, the review of BN models in water resource management and ESS pointed out *model simulations* as another data source used to construct DAGs and populate CPTs [5, 9]. *Model simulations* refers to outputs of other empirical, deterministic or stochastic models [5]. Interestingly, there was no standard BN model in cyber security that used *model simulation* as the data source to construct DAGs and populate CPTs.

3.3.3. THE NUMBER OF NODES USED IN THE MODEL

The number of nodes can be used to describe the model complexity [5]. A high number of nodes often lead to a lot of intermediary layers between the layer of input nodes and the layer of output nodes. This could weaken the relation between input and output nodes. Marcot et al. recommended to limit the number of node layers or sequential relationships to less than five to prevent this dilution of interactions [41].

Landuyt et al. indicate that BN models with nodes lower than 40 can safeguard the functionalities of BNs [5]. Based on our analysis, we conclude that the amount of nodes is

Table 3.2: List of Bayesian Network Models in Cyber Security (Ordered by the number of citations)

Article Title (Year)	Citations	Data Source (DAG)	Data Source (CPT)	Application	Application Sector
Dynamic Security Risk Management Using Bayesian Attack Graphs [11] (2012)	247	D	K	Risk Management	Non-specific
Measuring Network Security Using Bayesian Network-Based Attack Graphs [12] (2008)	136	K	K	Risk Management	Non-specific
Network Vulnerability Assessment Using Bayesian Networks [13] (2005)	106	K	K	Risk Management	Non-specific
Reasoning about Evidence using Bayesian Networks [14] (2008)	39	K	K	Forensic Investigation	Law Enforcement
A Bayesian Network Model for Predicting Insider Threats [15] (2013)	35	D,K	D,K	Threat Hunting (Insider Threat)	Non-specific
Identifying at-risk Employees: Modeling Psychosocial Precursors of Potential Insider Threats [16, 17] (2012,2010)	31,24	D,K	K	Threat Hunting (Insider Threat)	Non-specific
Identifying Compromised Users in Shared Computing Infrastructures: A Data-Driven Bayesian Network Approach [18] (2011)	23	D	D	Forensic Investigation	University
Development of Cyber Security Risk Model using Bayesian Networks [19] (2015)	21	D,K	K	Risk Management	Nuclear
Studying Interrelationships of Safety and Security for Software Assurance in Cyber Physical Systems: Approach Based on Bayesian Belief Networks [20] (2013)	20	K	K	Risk Management	Petroleum (Oil)
Vulnerability Categorization using Bayesian Networks [21] (2009)	10	D	D	Vulnerability Management (Classification)	Software
Quantitative Assessment of Cyber Security Risk using Bayesian Network-based Model [22] (2009)	8	D	D,K	Risk Management	Non-specific
A Bayesian Network Model for Likelihood Estimations of Acquirement of Critical Software Vulnerabilities and Exploits [23] (2015)	7	D,K	D,K	Governance	Software
Analysis of the Digital Evidence Presented in the Yahoo! Case [24] (2009)	2	K	K	Forensic Investigation	Law Enforcement
Modeling Information System Availability by using Bayesian Belief Network Approach [25] (2016)	1	D,K	K	Risk Management	Non-specific
A Bayesian Network Model for Predicting Data Breaches [26] (2016)	0	D,K	K	Risk Management	Health Care
Information Security Risk Assessment of Smartphones using Bayesian Networks [27, 28] (2016,2015)	0,0	D,K	K	Risk Management	Smartphone (In Finland)
Bayesian Network Modelling for Analysis of Data Breach in a Bank [29] (2011)	0	D	D	Risk Management	Banking

In Table 3.2, "Expert Knowledge (K)" refers to domain expert(s) and/or article's author(s) knowledge and "Empirical Data (D)" refers to observational or experimental evidence or data, either available directly to the authors or derived from the literature.

relatively kept low in the identified BN models in cyber security as 16 out of 17 BN models have a node number lower than 40. On the other hand, the BN model developed by Shin et al. exceeds the node number 40 [19]. However, the BN model developed by Shin et al. is a combination of two networks. If it is not possible to keep the model structure shallow, Marcot et al. suggested to break up the model into two or more networks [41]. Shin et al. utilised this idea to prevent the dilution of interactions between the input and output nodes.

3.3.4. TYPE OF THREAT ACTOR

We used the characteristic *"Type of threat actor"* because this will allow us to understand whether the BN model in cyber security was developed with a focus on particular type of threat actor(s). We classified threat actors as insider versus outsider [42]. Furthermore, we also considered their intentions, which could be either malicious/deliberate or accidental [42]. Figure 3.2 shows the general distribution of the BN models reviewed according to the type of threat actors and their intent.

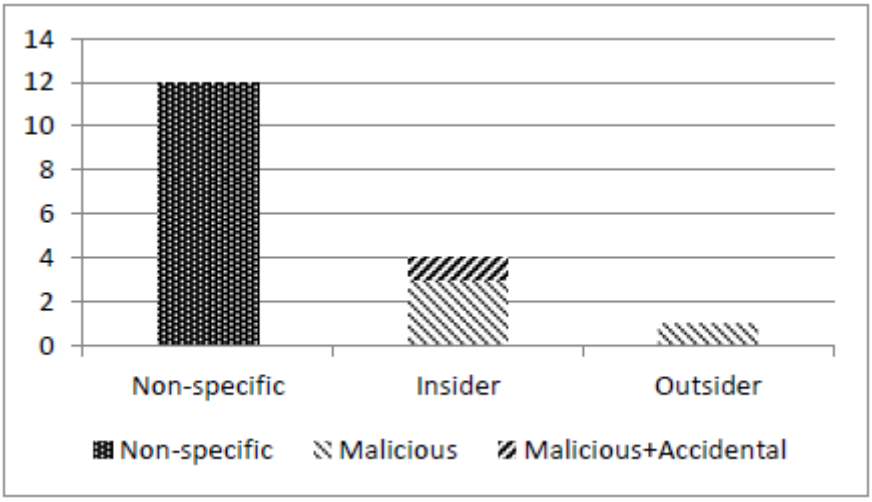


Figure 3.2: Characterization of Threat Actors in the BN Models Reviewed

From Figure 3.2, we infer that 4 out of 17 BN models are used only for problems associated with insiders [15, 16, 26, 29]. In particular, we observe that 4 out of these 4 BN models are appropriate for malicious insiders [15, 16, 26, 29], and only 1 out of these 4 BN models is relevant for accidental insiders in addition to malicious insiders [26]. Holm et al. developed a BN model with a focus on malicious outsider (professional penetration tester) [23].

Importantly, there was no integrated BN model that considers problem(s) associated with both insider and outsider type of threat actors, and their interactions. This type of BN models would help to combat especially social engineering attacks, and outsider collusion attacks [43]. Finally, there were 12 out of 17 BN models which did not focus on any specific type of threat actor [11–14, 18–22, 24, 25, 28]. For instance, the BN model developed by Pecchia et al. is used to identify compromised users in shared computing

infrastructures based on alerts [18]. This model did not focus on any specific type of threat actor, but rather focused on alerts which could be appropriate to any type of threat actor. Therefore, we categorized it as 'non-specific'.

3.3.5. APPLICATION AND APPLICATION SECTOR

We used the characteristic *"Application"* to understand the type of applications that partially or completely benefit from these BN models. We used the Chief Information Security Officer (CISO) mind map with CISO professional responsibilities as the basis to classify the reviewed BN models based on their application [44]. In addition, we used the characteristic *"Application Sector"* to identify the type of application sectors in which these BN models were demonstrated. From Table 3.2, we infer that 10 out of 17 BN models in cyber security completely or partially benefit Risk management. In addition, Forensic investigation, Governance, Threat hunting, and Vulnerability management were the other applications which completely or partially benefit from these BN models. From Table 3.2, we observe that the application sectors were quite diverse. However, 15 out of 17 BN models focused on the cyber security of Information Technology (IT) environment. In contrast, 2 out of 17 BN models focused on the cyber security of Industrial Control Systems (ICS) environment [19, 20].

Table 3.3: Scope of Variables used in the BN Models Reviewed

Authors	**Variables - Entities**	**Variables - Key Element(s) of Cyber Security**
Poolsappasit et al. [11]	Mail server, DNS server, SQL server, NAT Gateway server, Web server, Administrator machine, Local desktops	Technology
Frigault, Wang [12]	N/A	N/A
Liu, Man [13]	Network hosts	Technology
Kwan et al. [14]	Seized computer	Technology
Axelrad et al. [15]	Employee	People
Grietzer et al. [16, 17]	Employee	People
Pecchia et al. [18]	User profile, Shared computing infrastructure	People, Technology
Shin et al. [19]	Organization (Management) checklist, Reactor Protector System (RPS) components	Process, Technology
Kornecki et al. [20]	Components of ICS used to control oil pipeline flow	Technology
Wang, Guo [21]	Software	Technology
Mo et al. [22]	Organization (Management), Attack pathway	Process, Technology
Holm et al. [23]	Software	Technology
Kwan et al. [24]	Suspect, Seized computer, Yahoo! email account, Internet service provider	People, Technology
Ibrahimovic, Bajgoric [25]	Organization (Management)	Process
Wilde [26]	Employee, Organization (Management), Mobile Device	People, Process, Technology
Herland et al. [27, 28]	Smartphone	Technology
Apukhtin [29]	Employee, Organization (Management), Security controls	People, Process, Technology

3.3.6. SCOPE OF VARIABLES

We used the characteristic *"Scope of Variables"* to identify the entities to which the variables used in the reviewed BN models are related. In addition, we classify the variables used in the reviewed BN models based on the key elements of cyber security. Cyber security is a combination of three key elements: People, Process and Technology [45].

From Table 3.3, we observe that the variables used in the BN models that focus on the cyber security of ICS environment did not consider the 'people' element of cyber security

[19, 20]. Importantly, the variables used in these BN models are mainly related to the technological components of ICS ('Technology' focussed) [19, 20]. In addition, we infer that the variables used in 2 out of 4 BN models employed for the problems associated with insiders consider the three key elements of cyber security [26, 29] which are application-specific, whereas the variables used in 2 out of 4 BN models employed for the problems associated with insiders take into account only the 'people' element of cyber security [15, 16] which might be applicable to different organizations.

3.3.7. THE APPROACH(ES) USED TO VALIDATE MODELS

We used the characteristic *"The approach(es) used to validate models"* to identify the type of validation approaches used in the reviewed BN models. Based on our analysis, we observe that real-world case study [14, 24], cross-validation [15, 18], goodness of fit [16], Monte-Carlo simulation [25], expert evaluation [26, 27], and sensitivity analysis [26, 29] were the approaches used to validate the reviewed BN models. Importantly, there was no validation performed in 8 out 17 BN models [11–13, 19–23]. Finally, there was only one BN model which utilized several approaches such as sensitivity analysis, and expert evaluation to perform the validation [26]. However, the reviewed BN models validated different aspects depending on their objectives. For instance, Wilde [26] validated the usefulness of their model in practice, whereas Herland et al.[27, 28] validated the accuracy and completeness of the qualitative BN model.

Table 3.4: BN Model Purpose and Type of Purpose

Authors	**Model Purpose**	**Type of Purpose**
Poolsappasit et al. [11]	To quantify the chances of network compromise at various levels	Predictive
Frigault, Wang [12]	To determine the likelihood of attaining the goal state by exploiting vulnerabilities in a network	Predictive
Liu, Man [13]	To perform quantitative vulnerability assessment of a network of hosts	Predictive
Kwan et al. [14]	To reason about digital evidence in the BitTorrent case	Diagnostic
Axelrad et al. [15]	To predict degree of interest in a potentially malicious insider	Predictive
Greitzer et al. [16, 17]	To predict the psychosocial risk level of an individual	Predictive
Pecchia et al. [18]	To detect compromised users in shared computing infrastructures	Diagnostic
Shin et al. [19]	To evaluate the cyber security risk of the reactor protection system	Predictive
Kornecki et al. [20]	To jointly assess safety and security of a SCADA system used to control oil pipeline flow	Predictive
Wang, Guo [21]	To categorise software security vulnerabilities	Diagnostic
Mo et al. [22]	To evaluate the security readiness of organizations	Predictive
Holm et al. [23]	To estimate the likelihood that a penetration tester is able to obtain information about critical vulnerabilities and exploits for these vulnerabilities corresponding to a desired software and under different circumstances	Predictive
Kwan et al. [24]	To reason about digital evidence in the Yahoo! Case	Diagnostic
Ibrahimovic, Bajgoric [25]	To predict information system availability	Predictive
Wilde [26]	I. To predict the probability of a data breach caused by a group of insiders who lose employee- and employer-owned mobile devices or misuse the employer-owned mobile devices, II. To help health care organizations determine which additional measures they should take to protect themselves against data breaches caused by insiders.	Predictive, Diagnostic
Herland et al. [27, 28]	To assess information security risks related to smartphone use in Finland	Predictive
Apukhtin [29]	To predict the probability of a data breach in a bank caused by a malicious insider	Predictive

3.3.8. MODEL PURPOSE AND TYPE OF PURPOSE

We used the characteristic *"Model Purpose"* to point out the problems that were tackled using BN models in cyber security. In addition, we used the characteristic *"Type of Purpose"* to identify the corresponding category of model purpose. Table 3.4 highlights the authors of the BN model, the corresponding purpose of the BN model, and the corresponding type based on the model purpose.

From Table 3.4, we observe that the reviewed BN models in cyber security were mainly used for two types of purposes based on their model purpose: I. Diagnostic: To reason from effects to causes, and II. Predictive: To reason from causes to effects. Importantly, 13 out of 17 BN models in cyber security were used for predictive purposes.

3.4. DISCUSSION

In the previous section, we identified key usage patterns of BNs in cyber security. This section discusses potential reasons for the key findings and suggests future research directions.

There is an emphasis on problems associated with insiders compared to outsiders in the use of standard BN models in cyber security. In general, this emphasis could be due to the most significant threat posed by insiders. This was elucidated by IBM's cyber security intelligence index which concluded that 60% of all attacks were carried out by insiders [46]. In connection with the use of standard BNs, the availability of characteristics associated with insiders in the literature provided a good starting point to determine appropriate variables and their relationships which form an integral part of a standard BN. In addition, the variables and their relationships determined from the literature were fine-tuned and/or complemented with other suitable variables based on expert knowledge in a few instances. This is one of the major advantages of standard BNs described in Section 3.1 which is the ability to combine different sources of knowledge. This could be the rationale behind the predominant use of standard BNs for problems associated with the insiders.

Special importance is given to problems associated with malicious insiders compared to accidental insiders in the use of standard BN models in cyber security. In general, this could be due to the fact that malicious insiders are more natural than accidental insiders in security contexts, as malicious insiders have a clear intent of compromising security, while accidental insiders do not. Moreover, malicious insiders have been shown to be the cause of more incidents than accidental insiders, as it was demonstrated by IBM's cyber security intelligence index which concluded that 44.5% of attacks were carried out by malicious insiders, and accidental insiders were responsible for 15.5% of attacks [46]. In order to use standard BNs for problems associated with accidental insiders compared to malicious insiders, it is important to identify features associated with accidental insiders in the literature to determine appropriate variables and their relationships, which form an essential part of a standard BN. There are studies which identify features associated with accidental insiders in the literature [43, 47]. Once the appropriate variables and their relationships are determined for problems associated with accidental insiders, this could always be updated based on expert knowledge. It would also be useful to explore variables and their relationships in the reviewed BN models that focus on problems associated with malicious insiders, as some of the indicators might also apply for problems associated

with accidental insiders [43].

The focus on insiders may also explain why there is little research on applications in the ICS domain. The reviewed BN models that focus on problems associated with the insiders might not be suitable for ICS environments, especially for control rooms with an operator. This is prevalent in control rooms that are used to operate sluices in the Netherlands. Not accepting feedback, Anger management issues, Confrontational issues, Counterproductive behaviour towards individuals (CPB-I), Counterproductive behaviour towards the organization (CPB-O) were some of the variables used in the reviewed BN models [15, 16]. Most of these variables might be measured/observed based on interactions of the particular individual with the co-workers. However, this would not be possible in the control rooms where there would be no co-worker. It would be interesting to explore in the future whether the variables and their relationships in the reviewed BN models focused on problems associated with the insiders are suitable for ICS environment, and also whether the size of the organization in which the BN model would be applied have an effect on these variables and their relationships. In general, the limited use of standard BN models in cyber security on problems associated with ICS environment could be due to the shortage of ICS security expertise [48] as majority of the reviewed BN models relied on expert knowledge especially to construct DAGs and populate CPTs.

There is no integrated BN model which takes into account the problem(s) associated with both insiders and outsiders, and their interactions. The German steel mill incident is a typical example of a cyber-attack which involves both accidental insiders and malicious outsiders, and their interactions [49]. As an initial step, the adversaries used both the targeted email and social engineering techniques to acquire credentials for the plant's office network. Later, once they acquired credentials for the plant's office network, they worked their way into the plant's control system network and caused damage to the blast furnace. Standard BNs would help to tackle problem(s) associated with both insiders and outsiders, and their interactions, for instance a standard BN model that could predict the probability of an individual being deceived by outsider(s) to cause a cyber-attack in an organization, given certain risk factors and symptoms. This BN model would especially help to identify vulnerable individuals in an organization against social engineering attacks, and effective measures which could reduce the likelihood of an individual deceived by outsiders to cause a cyber-attack in an organization.

It is evident that the initial attempts in the use of standard BN models in cyber security were using BAG-based standard BN models [11–13]. BAG-based standard BN model combines acyclic attack graph which acts as the DAG with computational procedures of BN. Attack graph is one of the extensively used approaches in security modeling which was introduced in 1998 [10, 50]. The use of BAG-based standard BN models in the intial attempts could be due to practicality. It could be practical to build attack graphs first which had been extensively studied in this domain and use BN computational procedures for quantification during the early stages in the use of standard BN models in cyber security. Similarly, there were attempts in the safety domain which mapped fault tree to BN [51, 52]. Importantly, BAG-based standard BN models model static systems. Therefore, they are not directly applicable to multi-step attacks.

Risk management, forensic investigation, governance, threat hunting, and vulnera-

bility management were the applications of standard BNs in cyber security. However, it would also be useful to investigate the potential of standard BNs to benefit other applications. Chockalingam et al. highlighted the importance of integrating safety and security especially in the context of modern ICS [53]. BNs possess the potential to develop an integrated BN model that could diagnose the root cause of an abnormal behavior in the ICS especially whether the abnormal behavior is caused by an attack (security-related) or technical failure (safety-related) by taking into account certain risk factors and symptoms. This would allow the operator(s) to point out the best possible response strategy. For instance, the process of routing traffic through a scrubbing center would be a suitable response strategy for a Distributed Denial of Service (DDoS) attack whereas this may not be an appropriate response strategy for a sensor failure.

The sources of empirical data used to construct DAGs and populate CPTs include: literature, incidents data, NVD, OSVDB, and exploithub. It is important to identify other domain-specific empirical data sources which would help to develop realistic models in cyber security. For instance, Capture-The-Flag (CTF) events like SWaT security showdown (S3) [48] could be a potential data source to construct DAGs and populate CPTs. CTF events could generate datasets that are realistic in nature [54]. However, this could have been overlooked because the data generated in these events would be in most cases specific to that particular system, and the quality of data generated could depend on the participants.

3.5. CONCLUSIONS AND FUTURE WORK

In this study, we have identified 17 standard BN models in cyber security. Based on the analysis, we identified important patterns in the use of standard BN models in cyber security.

- The standard BN models in cyber security were significantly used for problems associated with malicious insiders.
- There is an emphasis on the use of standard BN models in cyber security for problems associated with IT environment compared to ICS environment. In addition, the standard BN models that focus on the cyber security of ICS environment did not consider the 'people' element of cyber security. This implies that there is no standard BN model which deal with problem associated with insiders in ICS environment.
- There is a lack of standard BN models usage for problems associated with insiders and outsiders, and their interactions.
- Expert knowledge, and empirical data predominantly from literature were the data sources utilised to construct DAGs and populate CPTs.
- The standard BN models in cyber security completely or partially benefited risk management, forensic investigation, governance, threat hunting, and vulnerability management.

- The approaches used to validate standard BN models in cyber security were real-world case study, cross-validation, goodness of fit, monte-carlo simulation, expert evaluation, and sensitivity analysis.

These patterns in the use of standard BN models in cyber security would help to make full use of standard BNs in cyber security in the future especially by pointing out the current trends, limitations and future research gaps.

In the future, it is important to investigate whether the BN models used for problems associated with insiders are applicable for ICS environments, especially for a control room with an operator. It would be useful to demonstrate the capacity of standard BNs to tackle problems associated with both insiders and outsiders, and their interactions like social engineering attacks, collusion attacks. It would be intriguing to investigate how to deal with multi-step attacks using standard BNs. The potential of alternative data sources like model simulations, CTF events to construct DAGs and populate CPTs in cyber security also needs to be explored, as well as the capability of standard BNs to completely or partially benefit the other applications in cyber security.

REFERENCES

[1] WEF: Partnering for Cyber Resilience: Towards the Quantification of Cyber Threats. (2015)

[2] Yu, S., Wang, G., Zhou, W.: Modeling Malicious Activities in Cyber Space. IEEE Network, vol. 29, pp. 83 - 87. (2015)

[3] Ben-Gal, I.: Bayesian Networks. Encyclopedia of Statistics in Quality and Reliability. John Wiley & Sons, Ltd. (2008)

[4] Darwiche, A.: Chapter 11 - Bayesian Networks. In: Foundations of Artificial Intelligence, vol. 3, pp. 467 - 509. (2008)

[5] Landuyt, D., et al.: A Review of Bayesian Belief Networks in Ecosystem Service Modelling. Environmental Modelling & Software, vol. 46, pp. 1 - 11. (2013)

[6] Uusitalo, L.: Advantages and Challenges of Bayesian Networks in Environmental Modelling. Ecological Modelling, vol. 203, pp. 312 - 318. (2007)

[7] Nikovski, D.: Constructing Bayesian Networks for Medical Diagnosis from Incomplete and Partially Correct Statistics. IEEE Transactions on Knowledge and Data Engineering, vol. 12, no. 4, pp. 509 - 516. (2000)

[8] Nakatsu, R.T.: Reasoning with Diagrams: Decision-Making and Problem-Solving with Diagrams. John Wiley & Sons. (2009)

[9] Phan, T.D., et al.: Applications of Bayesian Belief Networks in Water Resource Management: A Systematic Review. Environmental Modelling & Software, vol. 85, pp. 98 - 111. (2016)

[10] Kordy, B., Pièrre-Cambacédès, L., Schweitzer, P.: DAG-based Attack and Defense Modeling: Don't Miss the Forest for the Attack Trees. Computer Science Review, vol. 13, pp. 1 - 38. (2014)

[11] Poolsappasit, N., Dewri, R., Ray, I.: Dynamic Security Risk Management using Bayesian Attack Graphs. IEEE Transactions on Dependable and Secure Computing, vol. 9, pp. 61 - 74. (2012)

[12] Frigault, M., Wang, L.: Measuring Network Security using Bayesian Network-based Attack Graphs. IEEE. (2008)

[13] Liu, Y., Man, H.: Network Vulnerability Assessment using Bayesian Networks. In: Proc. SPIE, pp. 61-71. (2005)

[14] Kwan, M., Chow, K.-P., Law, F., Lai, P.: Reasoning about Evidence using Bayesian Networks. In: IFIP International Conference on Digital Forensics, pp. 275-289. (2008)

[15] Axelrad, E.T., Sticha, P.J., Brdiczka, O., Shen, J.: A Bayesian Network Model for Predicting Insider Threats. In: Security and Privacy Workshops, pp. 82-89. (2013)

[16] Greitzer, F.L., et al.: Identifying at-risk Employees: Modeling Psychosocial Precursors of Potential Insider Threats. In: System Science (HICSS), Hawaii International Conference on, pp. 2392-2401. (2012)

[17] Greitzer, F.L., et al.: Identifying at-risk Employees: A Behavioral Model for Predicting Potential Insider Threats. Pacific Northwest National Laboratory. (2010)

[18] Pecchia, A., et al.: Identifying Compromised Users in Shared Computing Infrastructures: A Data-driven Bayesian Network Approach. In: Reliable Distributed Systems (SRDS), 2011 30th IEEE Symposium on, pp. 127-136, IEEE. (2011)

[19] Shin, J., Son, H., Heo, G.: Development of a Cyber Security Risk Model using Bayesian Networks. Reliability Engineering & System Safety, vol. 134, pp. 208 - 217. (2015)

[20] Kornecki, A.J., Subramanian, N., Zalewski, J.: Studying Interrelationships of Safety and Security for Software Assurance in Cyber-Physical Systems: Approach based on Bayesian Belief Networks. In: Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on, pp. 1393-1399, IEEE. (2013)

[21] Wang, J.A., Guo, M.: Vulnerability Categorization using Bayesian Networks. In: Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research, pp. 1 - 4, ACM. (2010)

[22] Mo, S.Y.K., Beling, P.A., Crowther, K.G.: Quantitative Assessment of Cyber Security Risk using Bayesian Network-based Model. In: Systems and Information Engineering Design Symposium, 2009. SIEDS'09., pp. 183-187, IEEE. (2009)

[23] Holm, H., Korman, M., Ekstedt, M.: A Bayesian Network Model for Likelihood Estimations of Acquirement of Critical Software Vulnerabilities and Exploits. Information and Software Technology, vol. 58, pp. 304 - 318. (2015)

[24] Kwan, M., et al.: Analysis of the Digital Evidence Presented in the Yahoo! Case. In: IFIP International Conference on Digital Forensics, pp. 241-252, Springer. (2009)

[25] Ibrahimović, S., Bajgorić, N.: Modeling Information System Availability by using Bayesian Belief Network Approach. Interdisciplinary Description of Complex Systems, vol. 14, pp. 125 - 138. (2016)

[26] Wilde, L.: A Bayesian Network Model for Predicting Data Breaches Caused by Insiders of a Health Care Organization. University of Twente. (2016)

[27] Herland, K., Hammainen, H., Kekolahti, P.: Information Security Risk Assessment of Smartphones using Bayesian Networks. Journal of Cyber Security and Mobility, vol. 4, pp. 65 - 85. (2016)

[28] Herland, K.: Information Security Risk Assessment of Smartphones using Bayesian Networks. Aalto University, Finland. (2015)

[29] Apukhtin, V.: Bayesian Network Modeling for Analysis of Data Breach in a Bank. University of Stavanger, Norway. (2011)

[30] Khosravi-Farmad, M., Rezaee, R., Harati, A., Bafghi, A.G.: Network Security Risk Mitigation using Bayesian Decision Networks. In: Computer and Knowledge Engineering (ICCKE), 4th International eConference on, pp. 267-272. IEEE. (2014)

[31] Pan, S., Morris, T.H., Adhikari, U., Madani, V.: Causal Event Graphs Cyber-Physical System Intrusion Detection System. In: Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop, pp. 40. ACM. (2013)

[32] Frigault, M., et al.: Measuring Network Security using Dynamic Bayesian Network. In: Proceedings of the 4th ACM Workshop on Quality of Protection, pp. 23 - 30. (2008)

[33] Sarala, R., Kayalvizhi, M., Zayaraz, G.: Information Security Risk Assessment under Uncertainty using Dynamic Bayesian Networks. International Journal of Research in Engineering and Technology, pp. 304 - 309. (2014)

[34] Tang, K., Zhou, M.-T., Wang, W.-Y.: Insider Cyber Threat Situational Awareness Framwork using Dynamic Bayesian Networks. In: Computer Science & Education, 2009. ICCSE'09. 4th International Conference on, pp. 1146-1150. IEEE. (2009)

[35] Sommestad, T., Ekstedt, M., Johnson, P.: Cyber Security Risks Assessment with Bayesian Defense Graphs and Architectural Models. In: System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on, pp. 1 - 10. IEEE. (2009)

[36] Ekstedt, M., Sommestad, T.: Enterprise Architecture Models for Cyber Security Analysis. In: Power Systems Conference and Exposition, pp. 1 - 6. IEEE. (2009)

[37] Laskey, K., et al.: Detecting Threatening Behavior using Bayesian Networks. In: Conference on Behavioral Representation in Modeling and Simulation, pp. 33. (2006)

[38] AlGhamdi, G., et al.: Modeling Insider Behavior using Multi-entity Bayesian Networks. (2006)

[39] Okoli, C., Schabram, K.: A Guide to Conducting a Systematic Literature Review of Information Systems Research. Sprouts: Working Papers on Information Systems, vol. 10. (2010)

[40] Meho, L.I.: The Rise and Rise of Citation Analysis. Physics World, vol. 20, pp. 32. (2007)

[41] Marcot, B.G., Steventon, J.D., Sutherland, G.D., McCann, R.K.: Guidelines for Developing and Updating Bayesian Belief Networks Applied to Ecological Modeling and Conservation. Canadian Journal of Forest Research, vol. 36, pp. 3063 - 3074. (2006)

[42] Alberts, C., Dorofee, A.: OCTAVESM Threat Profiles.

[43] Bureau, F.I.P.: Unintentional Insider Threats: A Foundational Study. (2013)

[44] Rehman, R.: CISO MindMap. http://rafeeqrehman.com/wp-content/uploads/2017/07/CISO_Job_MindMap_v9.png. (2017)

[45] Andress, A.: Surviving Security: How to Integrate People, Process, and Technology. CRC Press, Boca Raton. (2003)

[46] 2016 Cyber Security Intelligence Index. IBM Security. (2016)

[47] Greitzer, F.L., et al.: Unintentional Insider Threat: Contributing Factors, Observables, and Mitigation Strategies. In: System Sciences (HICSS), 2014 47th Hawaii International Conference on, pp. 2025 - 2034. IEEE. (2014)

[48] Antonioli, D., et al.: Gamifying Education and Research on ICS Security: Design, Implementation and Results of S3. arXiv preprint arXiv:1702.03067. (2017)

[49] RISI Database.: German Steel Mill Cyber Attack. http://www.risidata.com/database/detail/german-steel-mill-cyber-attack. (2017)

[50] Lippmann, R.P., Ingols, K.W.: An Annotated Review of Past Papers on Attack Graphs. Massachusetts Institute of Technology Lincoln Laboratory, Lexington. (2005)

[51] Bobbio, A., Portinale, L., Minichino, M., Ciancamerla, E.: Improving the Analysis of Dependable Systems by Mapping Fault Trees into Bayesian Networks. Reliability Engineering & System Safety, vol. 71, pp. 249 - 260. (2001)

[52] Khakzad, N., Khan, F., Amyotte, P.: Safety Analysis in Process Facilities: Comparison of Fault Tree and Bayesian Network Approaches. Reliability Engineering & System Safety, vol. 96, pp. 925 - 932. (2011)

[53] Chockalingam, S., et al.: Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications. In: International Conference on Critical Information Infrastructures Security (CRITIS). Paris. (2016)

[54] Salem, M.B., Hershkop, S., Stolfo, S.J.: A Survey of Insider Attack Detection Research. Insider Attack and Cyber Security, pp. 69 - 90. (2008)

4

COMBINING BAYESIAN NETWORKS AND FISHBONE DIAGRAMS TO DISTINGUISH BETWEEN INTENTIONAL ATTACKS AND ACCIDENTAL TECHNICAL FAILURES*

4.1. INTRODUCTION

Today's society depends on the seamless operation of Critical Infrastructures (CIs) in different sectors such as energy, transportation, and water management, which is essential to the success of modern economies. Over the years, CIs have heavily relied on Industrial Control Systems (ICS) to ensure efficient operations, which are responsible for monitoring and steering industrial processes as, among others, water treatment and distribution, and flood control.

Modern ICS no longer operates in isolation, but uses other networks to facilitate and improve business processes [1]. For instance, ICS uses internet to facilitate remote access to vendors and support personnel. This increased connectivity, however, makes ICS more vulnerable to cyber-attacks. The German steel mill incident is a typical example of a cyber-attack in which adversaries made use of corporate network to enter into the

*This chapter has been published as Chockalingam, S., Pieters, W., Teixeira, A., Khakzad, N., and van Gelder, P.: *"Combining Bayesian Networks and Fishbone Diagrams to Distinguish between Intentional Attacks and Accidental Technical Failures,"* International Workshop on Graphical Models for Security, pp. 31 – 50, 2018. Springer, Cham. https://doi.org/10.1007/978-3-030-15465-3_3

ICS network [2]. As an initial step, the adversaries used both the targeted email and social engineering techniques to acquire credentials for the corporate network. Once they acquired credentials for the corporate network, they worked their way into the plant's control system network and caused damage to the blast furnace.

It is essential to distinguish between (intentional) attacks and (accidental) technical failures that would lead to abnormal behavior in a component of the ICS and take suitable measures. However, there are challenges to achieve these goals. One particularly important challenge is that the abnormal behavior in a component of the ICS due to attacks is often initially diagnosed as a technical failure [3]. This could be due to the imbalance in the frequency of attacks and technical failures. On the other hand, this could be based on one of the myths of ICS security: *"our facility is not a target"* [4]. In most cases, the initiation of response strategy aimed at technical failures would be ineffective in case of a targeted attack, and may lead to further complications. For instance, replacing a sensor that is sending incorrect measurement data with a new sensor would be a suitable response strategy to technical failure of a sensor. However, this may not be an appropriate response strategy to an attack on the sensor as it would not block the corresponding attack vector. Furthermore, the initiation of inappropriate response strategies would delay the recovery of the system from adversaries and might lead to harmful consequences. Noticeably, there is a lack of decision support to distinguish between attacks and technical failures.

Bayesian Networks (BNs) can be potentially used to tackle the challenge of distinguishing attacks and technical failures as they enable diagnostic reasoning, which could help to identify the most likely cause of an event based on certain symptoms (or effects) [5]. The diagnostic inference capability of BN has been widely employed in real-world applications especially in medical diagnosis [6], and fault diagnosis [7]. However, BNs are difficult to interpret for ICS domain experts and are therefore unsuitable for extracting the necessary knowledge. Conversely, fishbone diagrams are easy-to-use for brainstorming with experts [8], but lack essential capacities for diagnostic inference. Therefore, fishbone diagrams can be potentially combined with BNs to suit the purposes of present challenge. This research aims to provide decision support for distinguishing between attacks and technical failures by addressing the research question: "How could we combine Bayesian Networks and Fishbone Diagrams to find out whether an abnormal behavior in a component of the ICS is due to (intentional) attack or (accidental) technical failure or neither?". The research objectives are:

- **RO1.** To develop a framework for constructing BN models for determining the major cause of an abnormal behavior in a component of the ICS.
- **RO2.** To leverage fishbone diagrams for knowledge elicitation within our BN framework, and demonstrate the application of the developed methodology via a case study.

The scope of our BN framework development is the choice of appropriate types of variables and relationships between the determined variables. Firstly, we identify appropriate types of variables from existing diagnostic BN models in other domains and adapt them to the purposes of the present study (i.e., distinguishing attacks and technical failures); accordingly, the relationships between the selected variables should

be established. Furthermore, we provide a systematic method for incorporating fishbone diagrams within our BN framework to effectively elicit knowledge from different sources.

The remainder of this study is structured as follows: Section 4.2 provides an essential foundation of diagnostic BNs and previous related work, followed by an overview of the state-of-the-art regarding fishbone diagrams in Section 4.3. In Section 4.4, we illustrate the different layers and components of ICS and describe the case study in the water management domain that is used to demonstrate our proposed methodology. In Section 4.5, our BN framework is developed with appropriate types of variables and the relationships between these variables are established. Furthermore, we demonstrate the application of the developed methodology to a case study in the water management domain in Section 4.5. Section 4.6 presents the conclusions and future work directions.

4.2. DIAGNOSTIC BAYESIAN NETWORKS

This section explains diagnostic BNs with an example, and reviews existing diagnostic BNs in different domains. BNs belong to the family of probabilistic graphical models [9]. BNs consist of a qualitative and a quantitative part [10]. The qualitative part is a directed acyclic graph consisting of nodes and edges. Each node represents a random variable, while the edges between the nodes represent the conditional dependencies among the random variables. The quantitative part takes the form of a priori marginal and conditional probabilities so as to quantify the dependencies between connected nodes. An example of a BN model, representing the causal relationships between the risk factor “Smoking", the diseases “Bronchitis" and “Lung Cancer", and the symptoms “Shortness of Breath" and “Fatigue", is shown in Figure 4.1(a).

When more evidence or information becomes available for some variables in the BN, the probabilities of other variables in the BN could be updated. This is called probability propagation, inference, or belief updating [5]. In the example shown in Figure 4.1(b), the physician provides the evidence (via observation or supposition) for the symptoms “Shortness of Breath = False" and “Fatigue = True". Based on such evidence, the BN computes the posterior (updated) probabilities of the other nodes using Bayes’ theorem. The BN in Figure 4.1(b) determines that the absence of shortness of breath and the presence of fatigue are more likely due to lung cancer than bronchitis. In this case, we had evidence for symptoms (or effects) and inferred the most likely cause. This is called diagnostic or bottom-up reasoning. BNs also support three other types of reasoning: (i) Predictive or top-down: reasoning from causes to symptoms, (ii) Intercausal: reasoning about mutual causes of a common effect, and (iii) Combined: combination of different types of the above-mentioned reasoning [5].

BN models have widely been used for diagnostic analysis in different domains including agriculture [11], cyber security [12–15], health care [16–22], and transportation [23–25]. Chen et al. [11] proposed a two-layer BN for maize disease diagnosis. In their model, the upper layer consists of diseases and the lower layer consists of symptoms. However, their BN model did not take into account other variables like risk factors. In this case, it could be difficult to diagnose a particular disease among other potential diseases with the same symptoms.

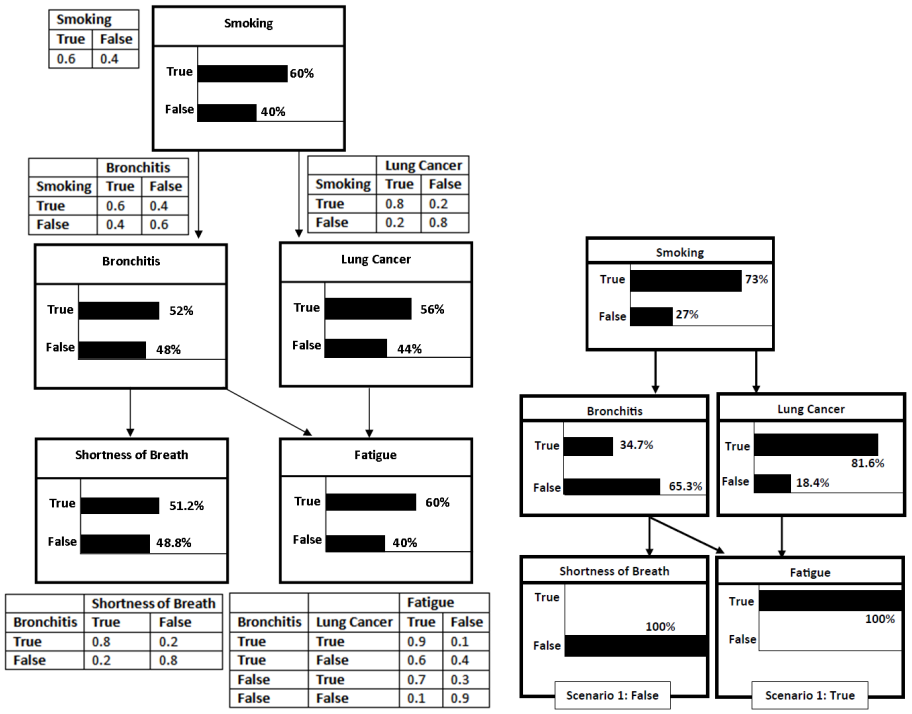


Figure 4.1: (a) A Typical BN Model for Disease Diagnosis. (b) Updated Probabilities Given Observed Symptoms (Evidence).

Pecchia et al. [12] developed a two-layer naïve BN model for detecting compromised users in shared computing infrastructures. In their model, the upper layer consists of a hypothesis variable “the user is compromised" while the lower layer consists of information variables. When more evidence or information becomes available for the information variables, this BN would help to diagnose whether the user has been compromised. In contrast to the BN model developed by Chen et al. [11], the upper layer consists of only one variable.

Oniśko et al. [16] proposed a three-layer BN for multiple-disorder diagnosis. In their model, the upper layer consists of risk factors, the middle layer consists of disorders, and the lower layer consists of symptoms and test results. In contrast to the BN models developed by Chen et al. [11] and Pecchia et al. [12], their BN model takes into account risk factors. Curiac et al. [17] also proposed a similar three-layer BN model for psychiatric disease diagnosis.

Huang et al. [23] proposed a four-layer BN for fault diagnosis of vehicle infotainment system. In their work, the upper layer consists of root causes, the middle layer consists of intermediate nodes which are usually the group or category of the root causes, and two lower layers being distinguished with different colours. One of the lower layers consists of observations (or test results) while the other consists of a symptom. In contrast to the BN models proposed by Oniśko et al. [16] and Curiac et al. [17], their BN model did not take into account risk factors. On the other hand, their BN model considered

observations (or test results) and symptom as separate layers. The observations (or test results) nodes could better help the diagnostic technicians who were not familiar with the list of diagnostic tests to be performed for diagnosing a particular root cause in the BN. The accuracy of posterior probabilities of non-evidenced variables in the BN would be improved as the observations (or test results) would make more evidence or information available based on the results of diagnostic tests performed.

Huang et al. [23] defined symptom as the failure symptom reported by the customer such as “no-sound", “no-display" in their vehicle infotainment system. In addition, they defined observations as any information useful for allocating the root causes such as those mentioned in the customer’s reports or the outcomes of tests performed by diagnostic technicians. However, there is no clear distinction between the information from customer’s reports that could be used to determine the observation nodes and a symptom node in the BN construction.

4.3. FISHBONE DIAGRAMS

This section explains fishbone diagrams, and highlights their application in both safety and security. Fishbone diagrams help to systematically identify and organise the possible contributing factors (or sub-causes) of a particular problem [8, 26–29]. Figure 4.2 shows the generic structure of a fishbone diagram, consisting of a problem and its possible contributing factors (or sub-causes) sorted and related under different categories. Each category represents the major cause of the problem. The categories used in the fishbone diagram depend on the classification scheme used for that application. In general, the arrows in the fishbone diagram represent the causal relation between the causes and the problem (effect). The major advantages of fishbone diagram include: (i) fishbone diagrams are easily adaptable based on the discussions during brainstorming sessions [8], (ii) fishbone diagram encourages and guides data collection by showing where knowledge is lacking [8, 26], (iii) fishbone diagram structure stimulates group participation [8, 26], (iv) fishbone diagram structure helps to stay focused on the content of the problem during brainstorming sessions [8].

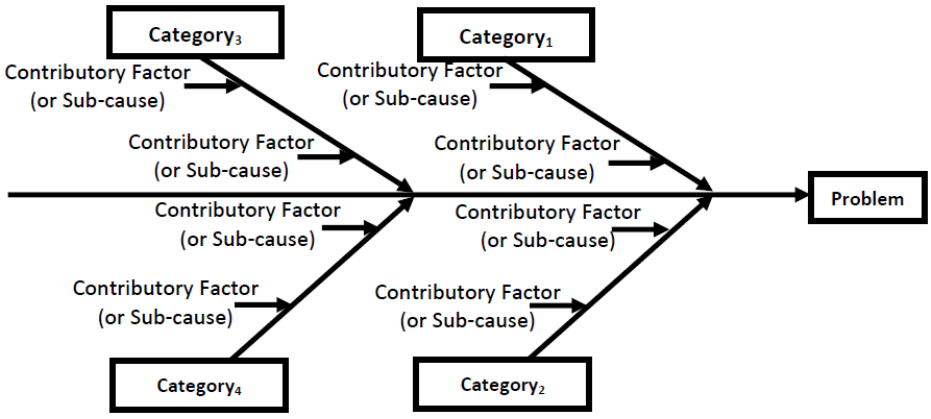


Figure 4.2: Generic Fishbone Diagram Structure

Fishbone diagrams are used in security and safety applications [30–33]. Asllani et al. [30] used fishbone diagrams to identify possible contributory factors of network failure/intrusions, and used six different categories to sort and relate contributory factors. For instance, they considered the problem as “Network Failure/Intrusions" and one of the potential contributory factors as “Antivirus Update" under the category “Processes". This implies that not updating antivirus could contribute to network failure/intrusions. Zhao et al. [31] used fishbone diagrams to illustrate possible contributory factors of tower crane accidents under five different categories. Luca et al. [32] used fishbone diagrams to illustrate possible contributory factors of noisy functioning of an automotive flue gas system under four different categories. Zhu et al. [33] used fishbone diagrams to illustrate possible contributory factors of crude oil vapors explosion in the drain under six different categories.

4.4. INDUSTRIAL CONTROL SYSTEMS

In this section, we illustrate the three different layers and major components in each layer of ICS. Furthermore, we provide an overview of a case study in the water management domain.

4.4.1. ICS ARCHITECTURE

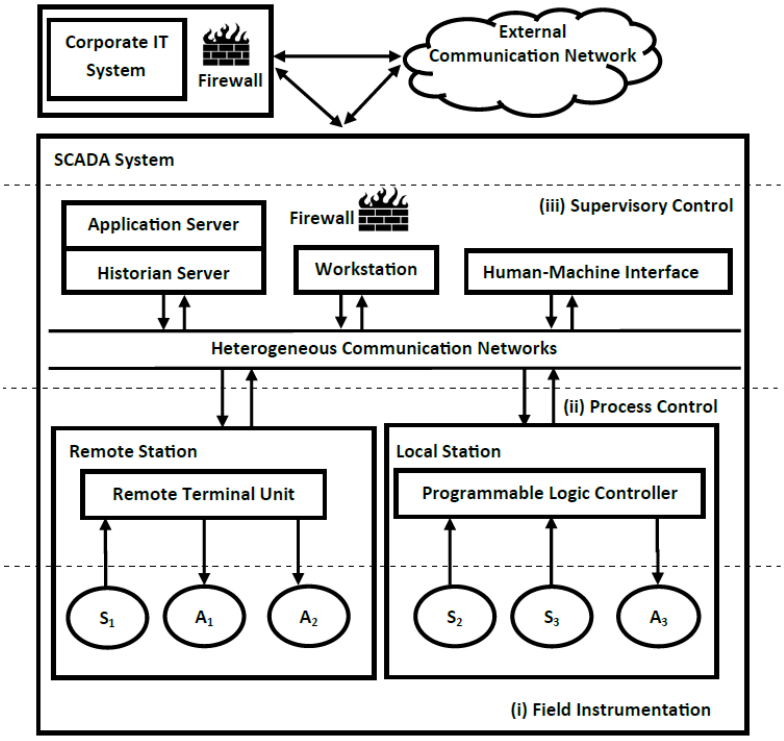


Figure 4.3: Typical ICS Architecture and Layers

Domain knowledge on ICS is the starting point for the development and application of our BN framework. A typical ICS consists of three layers: (i) Field instrumentation layer, (ii) Process control layer, and (iii) Supervisory control layer [34], bound together by network infrastructure, as shown in Figure 4.3.

The field instrumentation layer consists of sensors (S_i) and actuators (A_i), while the process control layer consists of Programmable Logic Controllers (PLCs)/Remote Terminal Units (RTUs). Typically, PLCs have wired communication capabilities whereas RTUs have wired or wireless communication capabilities. The PLC/RTU receives measurement data from sensors, and controls the physical systems through actuators [35]. The supervisory control layer consists of historian databases, software application servers, Human-Machine Interface (HMI), and workstation. The historian databases and software application servers enable the efficient operation of the ICS. The low-level components are configured and monitored with the help of workstation and HMI, respectively [35].

4.4.2. CASE STUDY OVERVIEW

This case study overview is based on a site visit to a floodgate in the Netherlands. Some critical information has purposely been anonymised for security concerns. Figure 4.4 schematises a floodgate being primarily operated by Supervisory Control and Data Acquisition (SCADA) system along with an operations centre.

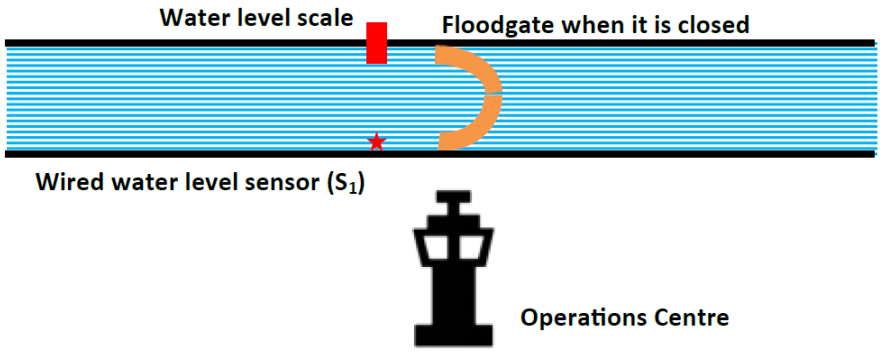


Figure 4.4: Physical Layout of the Floodgate

Figure 4.5 illustrates the SCADA architecture of the floodgate. The sensor (S_1) (which is located near the floodgate) is used to measure the water level. There is also a water level scale which is visible to the operator from the operations centre. The sensor measurements are then sent to the PLC. If the water level reaches the higher limit, PLC would send an alarm notification to the operator through the HMI, and the operator would need to close the floodgate in this case. The HMI would also provide information like the water level and the current state of the floodgate (open/close). The actuator opens/closes the floodgate.

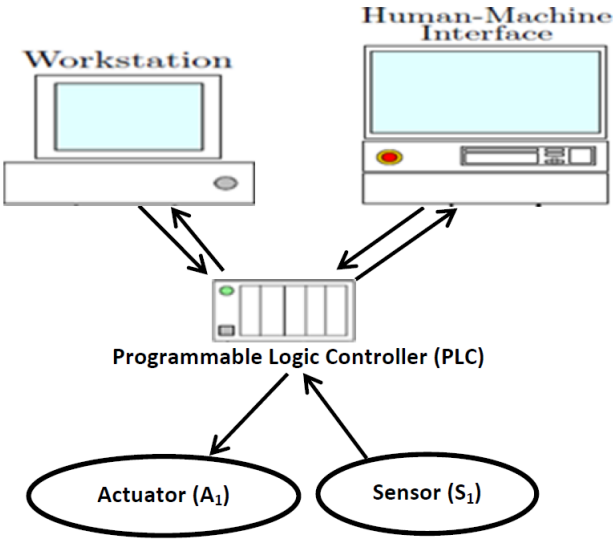


Figure 4.5: SCADA Architecture of the Floodgate

4.5. DEVELOPMENT AND APPLICATION OF THE METHODOLOGY

In this section, we describe our framework with the type of variables and their relationships. Furthermore, we use an illustrative case of a floodgate in the Netherlands to explain how we combine BN and fishbone diagram to distinguish between (intentional) attacks and (accidental) technical failures.

4.5.1. FRAMEWORK FOR DISTINGUISHING ATTACKS AND TECHNICAL FAILURES

The developed BN framework is grounded in BN models used for diagnostic purposes in different domains [12, 16, 17, 23]. Studying the aforementioned diagnostic BN models in Section 4.2, we adopted and customised a set of variables to develop our BN framework. The type of variables which we adopted are: (i) risk factors [16, 17], (ii) hypothesis [12], and (iii) observations (or test results) [23].

Pecchia et al. [12] used a hypothesis variable in their BN model as a classifier node to classify whether the user is compromised or not in shared computing infrastructures. We adopted the notion of a classifier node from Pecchia et al. [12] as it is the basis to the purposes of the present study. However, we defined it as the problem variable as it is an abnormal behavior in a component of the ICS (observable problem) in our work. For instance, the sensor (S_1) sends incorrect water level measurements. The purpose of the hypothesis variable in Pecchia et al. is to determine whether the user is compromised or not in sharing computing infrastructures, whereas in our work it is used to determine the major cause of the problem. An abnormal behavior in the technological components could be mainly caused by intentional attacks, accidental technical failures, human errors, or natural disasters [36]. However, the main objective of our study is to distinguish

between attacks and technical failures. Therefore, we considered intentional attack and accidental technical failure as major causes of the problem. In addition, we introduced a category “others" in case the major cause of the problem is neither intentional attack nor accidental technical failure. For instance, the sensor (S_1) is misplaced in a different location by an operator. In this case, the major cause of the problem is human error and would thus be determined as “others".

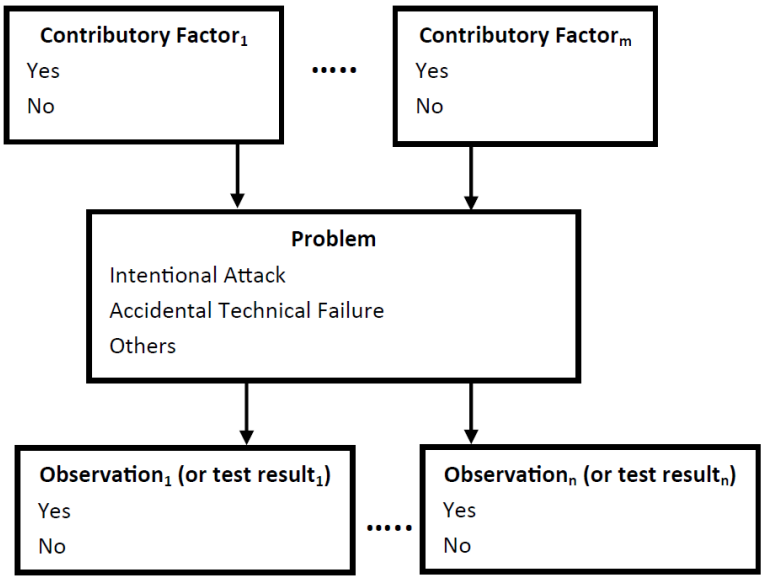


Figure 4.6: BN Structure to Determine the Major Cause of an Abnormal Behavior in a Component of the ICS

Oniśko et al. [16] and Curiac et al. [17] defined risk factors as the factors that would increase the likelihood of a disease. We, accordingly, adopted the term risk factors, and defined them as contributory factors since they contribute to the major cause of the problem in our work. For instance, “weak physical access-control" could contribute to the sensor (S_1) sending incorrect water level measurements due to an attack. Furthermore, there might be common contributory factors to different major causes of the problem. For instance, “outdated technology" could contribute to both the sensor (S_1) sending incorrect water level measurements due to an attack and a technical failure.

In general, observations (or test results) play an important role in diagnostics. Huang et al. [23] defined observations as any information useful for allocating the root causes such as those mentioned in the customer’s reports or the outcomes of tests performed by diagnostic technicians. We defined observations (or test results) as any information useful for determining the major cause of the problem based on the outcomes of tests. For instance, the outcome of the test “whether the sensor (S_1) sends correct water level measurements after cleaning the sensor (S_1)?" would provide an additional information to determine the major cause (accidental technical failure) of the problem accurately. The observation (or test results) variables can be elicited from different sources such as experts, product manuals, and previous incident reports. For instance, the global water level sensor WL400 product manual lists troubleshooting tests for incorrect water level

measurements due to (accidental) technical failures [37]. One of the troubleshooting tests listed in the product manual is to clean the sensor following the maintenance instructions and check whether the sensor sends correct water level measurements. Figure 4.6 shows the BN structure to build BN models for determining the major cause of an abnormal behavior in a component of the ICS, representing the causal relationship between the contributory factors, the problem, and the observations (or test results).

4.5.2. COMBINING BAYESIAN NETWORKS AND FISHBONE DIAGRAMS

Knowledge elicitation plays an important role to construct BN model especially with the appropriate variables for the considered problem [38, 39]. There are challenges to solely rely on BN for knowledge elicitation. For instance, BN is not easy-to-use for brainstorming with domain experts as it could be time-consuming to explain the notion of BN and also to change its structure instantly based on discussions during brainstorming sessions. Notably, expert knowledge is one of the predominant data sources utilised to build BN structure with appropriate variables especially in domains where there is a limited availability of data like cyber security [40]. Therefore, our framework would be incomplete without an effective method for knowledge elicitation.

In our work, fishbone diagram is used as the foundation to develop an effective method for knowledge elicitation especially based on their advantages stated in Section 4.3. Furthermore, there are additional benefits in the use of fishbone diagram in our work. We would mainly rely on experts from two different domains in addition to other sources for knowledge elicitation to construct BN models: (i) security, dealing with intentional attacks, and (ii) safety, dealing with accidental technical failures. In case we start building a BN model directly without utilising the fishbone diagram to elicit data from experts, it would be difficult to visualise which contributory factors and observations (or test results) corresponds to each major cause of the problem. This could make it difficult for the experts especially during brainstorming sessions. The fishbone diagram structure shows the potential to tackle this challenge. In some cases, there might be common contributory factors. For instance, “outdated technology" is a common contributory factor to two major causes of the problem (i.e., “outdated technology" could contribute to the sensor (S_1) sending incorrect water level measurements due to both “intentional attack" and “accidental technical failure"). If we start building a BN model directly without utilising the fishbone diagram to elicit data from experts, this could lead to duplication of common contributory factors using different terminologies in the BN.

In addition, BN structure is not easily changeable especially with a large number of contributory factors and observations (or test results) elicited from experts during brainstorming sessions. The fishbone diagram structure makes it easier to refine/update a large number of contributory factors and observations (or test results) instantly based on discussions during brainstorming sessions with experts. It would also help to visualise contributory factors and observations (or test results) from other sources such as literature and previous incidents. Finally, we can convert the constructed fishbone diagram into a corresponding BN model after the completion of knowledge elicitation to constitute the quantitative part of the corresponding BN model.

4.5.3. EXTENDED FISHBONE DIAGRAMS AND TRANSLATED BNS

We considered the example problem “sensor (S_1) sends incorrect water level measurements" as it could develop more complex situations in the case of floodgate. In case the floodgate closes when it should not based on the incorrect water level measurements sent by the sensor (S_1), it would lead to severe economic damage, for instance, by delaying cargo ships. On the other hand, in case the floodgate opens when it should not due to incorrect water level measurements sent by the sensor (S_1), it would lead to flooding.

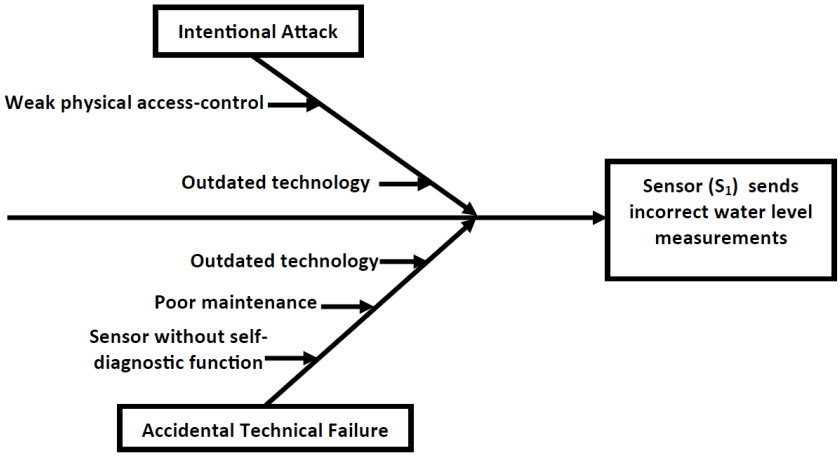


Figure 4.7: Fishbone Diagram Example

Figure 4.7 shows a fishbone diagram based on the example mentioned above. We considered “sensor (S_1) sends incorrect water level measurements" as the problem. Furthermore, we considered two major causes of the problem: intentional attack and accidental technical failure as mentioned earlier. These major causes of the problem would be the categories in our fishbone diagram. Finally, we mapped the appropriate contributory factors under each category. In this case, “outdated technology" is the common contributory factor that could contribute to sensor (S_1) sending incorrect water level measurements due to intentional attack and accidental technical failure. In this case, we listed “weak physical access-control" as one of the contributory factors in the category of intentional attack. This is because weak physical access-control could contribute to sensor (S_1) sending incorrect water level measurements due to an intentional attack.

Noticeably, fishbone diagrams do not consist of observations (or test results), which need to be elicited in our work. However, we could extend the fishbone diagram to incorporate observations (or test results) as shown in Figure 4.8. This would allow us to elicit complete information needed to construct BN models especially with the three different types of variables and cause-effect relationships in our BN framework. The extended fishbone diagram is shown in Figure 4.8 with an additional component: observations (or test results). The arrows in the fishbone diagram represent the causal relationship. The categories stated on the left side of the problem in the fishbone diagram are the major causes of the problem. Therefore, these categories has the arrows directing towards the problem which represent the causal relationship between the causes and the problem.

However, the categories stated on the right side of the problem are used for reference to elicit observations (or test results) that would be useful for determining the particular major cause of the problem. Figure 4.9 shows the extended version of our fishbone diagram example with observations (or test results).

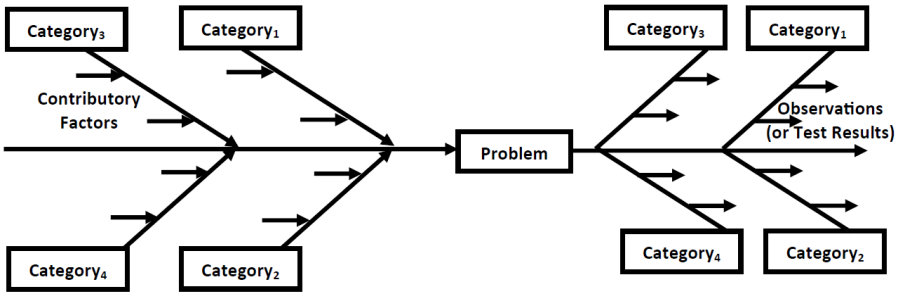


Figure 4.8: Extended Fishbone Diagram Structure

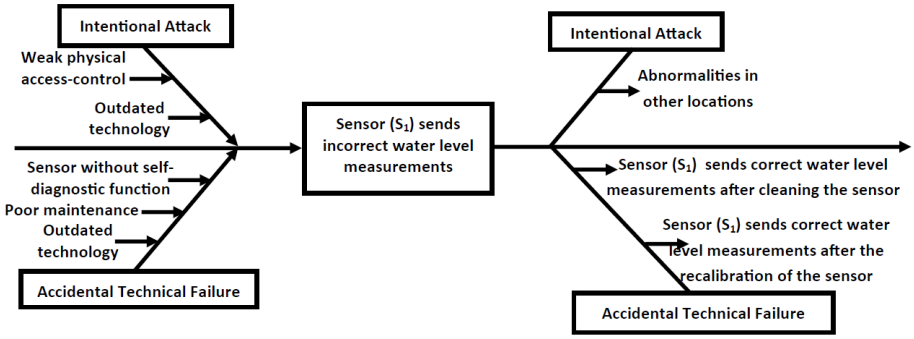


Figure 4.9: Extended Fishbone Diagram Example

Extended fishbone diagrams might look similar to qualitative bowtie diagrams, but, they are different. The observations (or test results) on the right side of the problem node in the extended fishbone diagram help distinguish between *different* events (intentional attack and accidental technical failure), Whereas bowtie diagrams are aimed at representing the possible consequences of a *fixed* event. Furthermore, qualitative bowties [41] consider recovery measures/reactive controls on the right side of the problem node. This is not relevant to our application because we focus on diagnostics. On the other hand, extended fishbone diagrams consider preventive controls/barriers implicitly on the left side of the problem node, as part of the contributory factors. For instance, “weak physical access-control for the sensor" is one of the contributory factors. The evidence supplied by the operator in the BN for this node would depend on the preventive controls/barriers that are in place. In case there are physical access-control measures implemented in that specific application, the operator would supply the evidence as ‘No’ for this node in the BN.

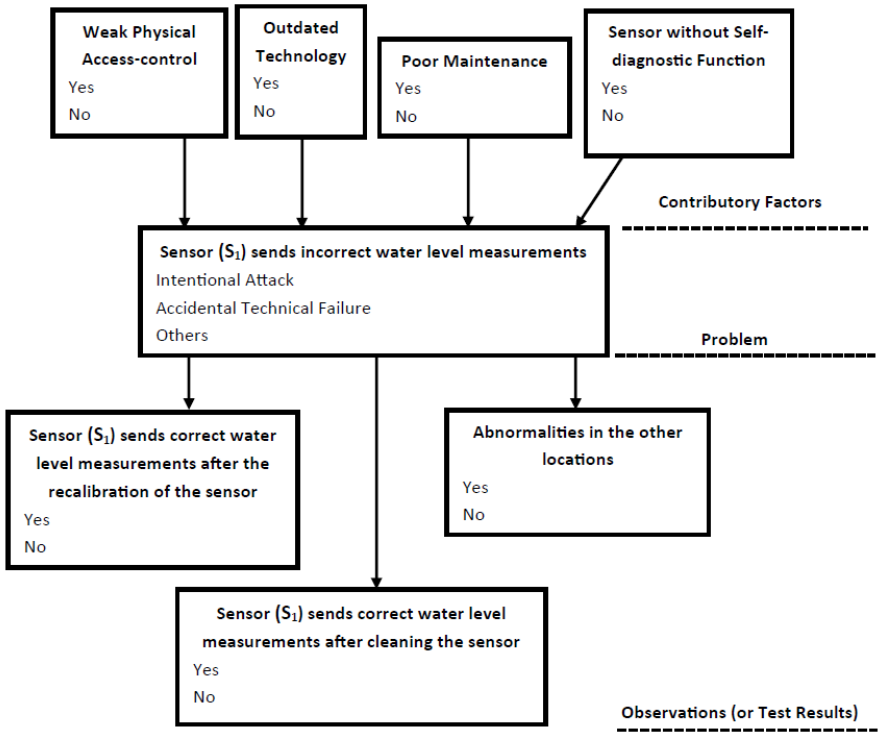


Figure 4.10: Translated BN from Fishbone Diagram Example

Once the fishbone diagram is developed, it should be translated to a BN based on the following steps:

1. The considered problem in the fishbone diagram is mapped to the problem variable in the middle layer of the BN as shown in Figure 4.10.

2. The categories used in the fishbone diagram would be states of the problem variable in our BN. In addition to these states, there would be an additional state “Others" in our BN. As mentioned in Section 4.5.1, this would be determined in case the major cause of the problem is neither intentional attack nor accidental technical failure.

3. The elicited contributory factors in the fishbone diagram are mapped to the contributory factor variables in the upper layer of the BN as shown in Figure 4.10. The contributory factors that correspond to both intentional attack and accidental technical failure in the fishbone diagram would be treated as a single contributory factor in the BN. For instance, “outdated technology" in our example would be treated as a single contributory factor in BN as shown in Figure 4.10. However, the contributory factors that correspond to both intentional attack and accidental technical failure would be reflected through the conditional probabilities of “sensor (S_1) sends incorrect water level measurements". We considered the contributory factors as binary discrete variables based on their features. However, continuous

variables could also have been used. We utilised the states *"Yes"* and *"No"* for our contributory factors as shown in Figure 4.10.

4. The elicited observations (or test results) in the fishbone diagram are mapped to the observations (or test results) in the lower layer of the BN as shown in Figure 4.10. We considered the observations (or test results) as binary discrete variables based on their characteristics. We employed the states *"Yes"* and *"No"* for our observations (or test results) as shown in Figure 4.10.

Once the fishbone diagram is translated to a corresponding BN model, the quantitative part of the BN should be populated. Due to limited data availability, expert knowledge is the predominant data source used to populate CPTs of BNs in cyber security [40]. In our work, we did not investigate whether fishbone diagrams could be used as a means to elicit probabilities from experts as our main objective is to elicit appropriate variables in the construction of the BN structure for the considered problem.

However, it is important to investigate whether fishbone diagrams could be used to elicit CPTs from experts in the future. The translated BN with illustrative priori marginal and conditional probabilities, representing the causal relationships between the contributory factors, the problem, and the observations (or test results), is shown in Figure 4.11.

Once the quantitative part of the BN is populated, the BN could be used in practice for different scenarios and their probabilities could be updated based on evidences obtained. In the example shown in Figure 4.11, we provided the evidence for the contributory factors "Weak Physical Access Control = Yes", "Outdated Technology = Yes", "Poor Maintenance = No" and "Sensor without Self-diagnostic Function = No", and observation (or test result) "Abnormalities in the other locations = Yes". Based on such evidence, the BN computes the posterior (updated) probabilities of the other nodes. The BN in Figure 11 determines that the problem "Sensor (S1) sends incorrect water level measurements" is most likely due to (intentional) attack based on the evidence provided.

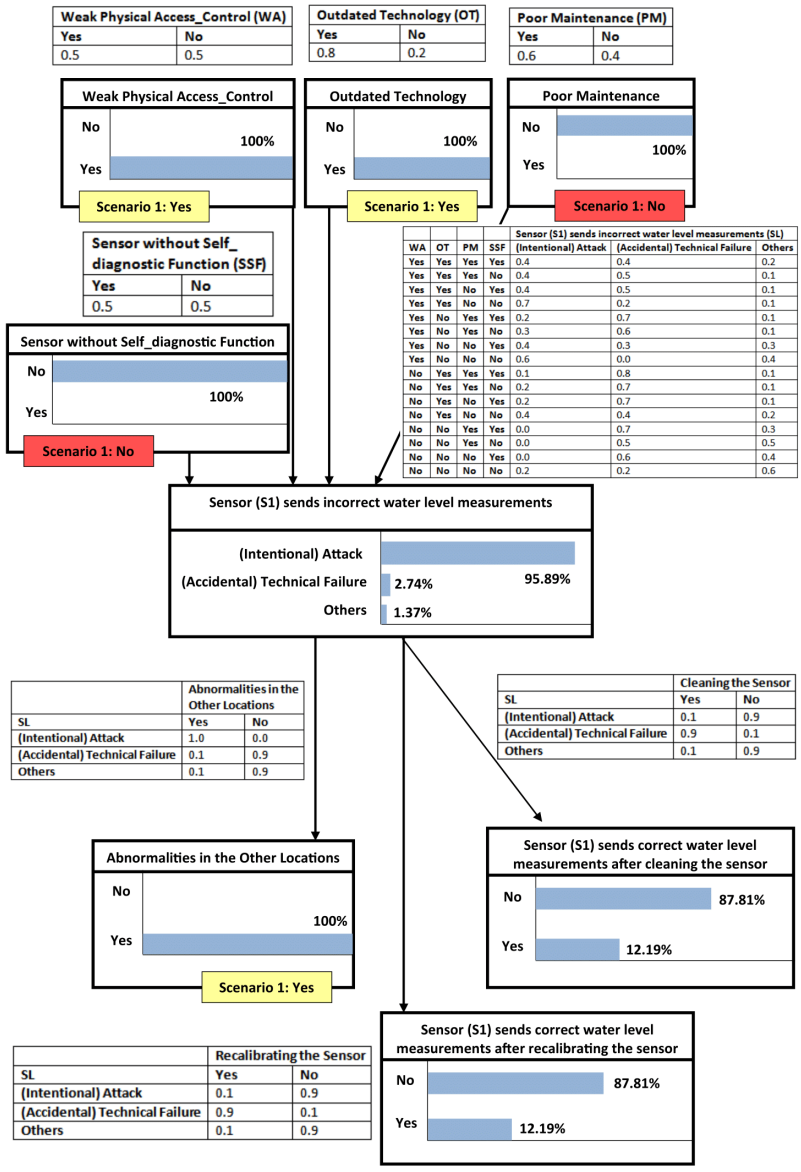


Figure 4.11: Translated BN with Updated Probabilities Based on the Evidence

4.6. CONCLUSIONS AND FUTURE WORK

Adequate decision support for distinguishing intentional attacks and accidental technical failures is missing. In this study, we customised and utilised three different types of variables from existing diagnostic BN models in a BN framework to construct BN models for distinguishing intentional attacks and accidental technical failures. In our BN framework, the upper layer consists of contributory factors, the middle layer consists of a problem variable and the lower layer consists of observations (or test results). Furthermore, we extended and combined fishbone diagram with our BN framework to support knowledge elicitation from different sources. The important characteristics of our framework include: (i) it serves as a basis to provide decision support for responding to safety and security problems arise in the components of ICS, (ii) While determining the most likely cause of an abnormal behavior in a component of the ICS, it helps to consider both the contributory factors and observations (or test results) associated with it, and (iii) it facilitates knowledge elicitation especially from experts and its integration in BNs. Finally, we demonstrated the use of the developed methodology with an example problem "sensor (S_1) sends incorrect water level measurements" based on a case study in water management domain.

This work belongs to the broader theme of "Integrated safety and security". There are several studies within the sub-theme of "Integrated safety and security risk assessment" [42]. However, this work is associated with the sub-theme of "Integrated safety and security diagnostics", which mainly deals with the problem of distinguishing intentional attacks and accidental technical failures.

In the future, it would be useful to investigate whether fishbone diagrams could be used to elicit CPTs. The developed methodology would not be directly applicable when several problems arise at the same time. Therefore, it is important to address how fishbone diagrams could be used to elicit knowledge for those cases in the future and how it could be translated to a corresponding BN. Furthermore, we aim to evaluate our methodology based on applications in the water management domain.

REFERENCES

[1] Knowles, W., Prince, D., Hutchison, D., Disso, J., Jones, K.: A Survey of Cyber Security Management in Industrial Control Systems, International Journal of Critical Infrastructure Protection, vol. 9, pp. 52 - 80, Elsevier. (2015)

[2] RISI Database.: German Steel Mill Cyber Attack, `http://www.risidata.com/database/detail/german-steel-mill-cyber-attack`. (2018)

[3] Macaulay, T., Singer, B.L.: Cybersecurity for Industrial Control Systems: SCADA, DCS, PLC, HMI, and SIS, Auerbach Publications. (2016)

[4] KasperskyLab.: Five Myths of Industrial Control Systems Security, `https://media.kaspersky.com/pdf/DataSheet_KESB_5Myths-ICSS_Eng_WEB.pdf`. (2014)

[5] Korb, K.B, Nicholson, A.E.: Bayesian Artificial Intelligence, CRC Press. (2010)

[6] Nikovski, D.: Constructing Bayesian Networks for Medical Diagnosis from Incomplete and Partially Correct Statistics, IEEE Transactions on Knowledge & Data Engineering, vol. 9, no. 4, pp. 509 - 516, IEEE. (2000)

[7] Nakatsu, R.T.: Reasoning with Diagrams: Decision-Making and Problem-Solving with Diagrams. John Wiley & Sons. (2009)

[8] Doggett, A.M.: Root Cause Analysis: A Framework for Tool Selection, Quality Management Journal, vol. 12, no. 4, pp. 34 - 45, Taylor & Francis. (2005)

[9] Ben-Gal, I.: Bayesian Networks. Encyclopedia of Statistics in Quality and Reliability. John Wiley & Sons, Ltd. (2008)

[10] Darwiche, A.: Chapter 11 - Bayesian Networks. In: Foundations of Artificial Intelligence, vol. 3, pp. 467 - 509. (2008)

[11] Chen, G., Yu, H.: Bayesian Network and Its Application in Maize Diseases Diagnosis, International Conference on Computer and Computing Technologies in Agriculture, pp. 917 - 924, Springer. (2007)

[12] Pecchia, A., et al.: Identifying Compromised Users in Shared Computing Infrastructures: A Data-driven Bayesian Network Approach. In: Reliable Distributed Systems (SRDS), 2011 30th IEEE Symposium on, pp. 127-136. IEEE. (2011)

[13] Kwan, M., Chow, K.-P., Law, F., Lai, P.: Reasoning about Evidence using Bayesian Networks. In: IFIP International Conference on Digital Forensics, pp. 275-289, Springer. (2008)

[14] Wang, J.A., Guo, M.: Vulnerability Categorization using Bayesian Networks. In: Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research, pp. 29. ACM. (2010)

[15] Kwan, M., et al.: Analysis of the Digital Evidence Presented in the Yahoo! Case. In: IFIP International Conference on Digital Forensics, pp. 241-252. Springer. (2009)

[16] Oniśko, A., Druzdzel, M., Wasyluk, H.: Extension of the Hepar II Model to Multiple-disorder Diagnosis, Intelligent Information Systems, pp. 303 - 313, Springer. (2000)

[17] Curiac, D., Vasile, G., Banias, O., Volosencu, C., Albu, A.: Bayesian Network Model for Diagnosis of Psychiatric Diseases, Information Technology Interfaces, 2009. ITI'09. Proceedings of the ITI 2009 31st International Conference on, pp. 61 - 66, IEEE. (2009)

[18] Estabragh, Z., Kashani, M., Moghaddam, F., Sari, S., Taherifar, Z., Moosavy, S., Oskoyee, K.: Bayesian Network Modeling for Diagnosis of Social Anxiety using Some Cognitive-behavioral Factors, Network Modeling Analysis in Health Informatics and Bioinformatics, vol. 2, no. 4, pp. 257 - 265, Springer. (2013)

[19] González-López, J., García-Aparicio, Á., Sánchez-Ponce, D., Muñoz-Sanz, N., Fernandez-Ledo, N., Beneyto, P., Westcott, MC.: Development and Validation of a Bayesian Network for the Differential Diagnosis of Anterior Uveitis, Eye, vol. 30, no. 6, pp. 865 - 872, Nature Publishing Group. (2016)

[20] Moreira, M.W., Rodrigues, J.J., Oliveira, A.M, Ramos, R.F., Saleem, K.: A Preeclampsia Diagnosis Approach using Bayesian Networks. 2016 IEEE Conference on Communications (ICC), pp. 1 - 5, IEEE. (2016)

[21] Kahn Jr, C.E., Roberts, L.M, Shaffer, K.A, Haddawy, P.: Construction of a Bayesian Network for Mammographic Diagnosis of Breast Cancer, Computers in Biology and Medicine, vol. 27, no. 1, pp. 19 - 29, Pergamon. (1997)

[22] Wang, X.H., Zheng, B., Good, W.F., King, J.L., Chang, Y.H.: Computer-assisted Diagnosis of Breast Cancer using a Data-driven Bayesian Belief Network, International Journal of Medical Informatics, vol. 54, no. 2, pp. 115 - 126, Elsevier. (1999)

[23] Huang, Y., McMurran, R., Dhadyalla, G., Jones, R.P.: Probability Based Vehicle Fault Diagnosis: Bayesian Network Method, Journal of Intelligent Manufacturing, vol. 19, no. 3, pp. 301 - 311, Springer. (2008)

[24] Jianhui, L., Zhang, J., Mingdi, J.: Application of BN in the Fault Diagnosis of Brake Failure System, Applied Mechanics & Materials. (2014)

[25] Kipersztok, O., Dildy, G.A.: Evidence-based Bayesian Networks Approach to Airplane Maintenance, Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN), vol. 3, pp. 2887 - 2892, IEEE. (2002)

[26] Ilie, G., Ciocoiu, C.N.: Application of Fishbone Diagram to Determine the Risk of an Event with Multiple Causes, Management Research and Practice, vol. 2, no. 1, pp. 1 - 20, Research Centre in Public Administration and Public Services, Bucharest, Romania. (2010)

[27] Ishikawa, K.: Guide to Quality Control, vol. 2, Asian Productivity Organization Tokyo. (1982)

[28] Desai, M.S., Johnson, R.A.: Using a Fishbone Diagram to Develop Change Management Strategies to Achieve First-year Student Persistence, SAM Advanced Management Journal, vol. 78, no. 2, pp. 51 - 63, Society for the Advancement of Management. (2013)

[29] White, A.A., Wright, S., Blanco, R., Lemonds, B., Sisco, J., Bledsoe, S., Irwin, C., Isenhour, J., Pichert., J.: Cause-and-Effect Analysis of Risk Management Files to Assess Patient Care in the Emergency Department, Academic Emergency Medicine, vol. 11, no. 10, pp. 1035 - 1041, Wiley Online Library. (2004)

[30] Asllani, A., Ali, A.: Securing Information Systems in Airports: A Practical Approach, 2011 International Conference for Internet Technology and Secured Transactions (ICITST), pp. 314 - 38, IEEE. (2011)

[31] Zhao, C.H., Zhang, J., Zhong, X.Y., Zeng, J., Chen, S.J.: Analysis of Accident Safety Risk of Tower Crane based on Fishbone Diagram and the Analytic Hierarchy Process, Applied Mechanics and Materials, vol. 127, pp. 139 - 143, Trans Tech Publ. (2012)

[32] Luca, L., Stancioiu, A.: The Study Applying a Quality Management Tool to Identify the Causes of a Defect in an Automotive, Proceedings of the 3rd International Conference on Automotive and Transport Systems. (2012)

[33] Zhu, Y., Qian, X.M., Liu, Z.Y., Huang, P., Yuan, M.Q.: Analysis and Assessment of the Qingdao Crude Oil Vapor Explosion Accident: Lessons Learnt, Journal of Loss Prevention in the Process Industries, vol. 33, pp. 289 - 303, Elsevier. (2015)

[34] Endi, M., Elhalwagy, Y., Attalla, H.: Three-layer PLC/SCADA System Architecture in Process Automation and Data Monitoring, The 2nd International Conference on Computer and Automation Engineering (ICCAE), vol. 2, pp. 774 - 779, IEEE. (2010)

[35] Skopik, F., Smith, P.D.: Smart Grid Security: Innovative Solutions for a Modernized Grid, Syngress. (2015)

[36] Grimvall, G., Holmgren, A., Jacobsson, P., Theden, T.: Risks in Technological Systems, Springer Science & Business Media. (2009)

[37] GlobalWater.: Global Water Level Sensor - WL400 Product Manual, `http://www.globalw.com/downloads/WL400/WL400manual.pdf` (2009)

[38] Przytula, K., Thompson, D.: Construction of Bayesian Networks for Diagnostics, 2000 IEEE Aerospace Conference Proceedings, vol. 5, pp. 193 - 200, IEEE. (2000)

[39] Henrion, M.: Practical Issues in Constructing a Bayes' Belief Network, arXiv preprint arXiv:1304.2725. (2013)

[40] Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P.: Bayesian Network Models in Cyber Security: A Systematic Review, Nordic Conference on Secure IT Systems, pp. 105 - 122, Springer. (2017)

[41] de Ruijter, A., Guldenmund, F.: The Bowtie Method: A Review, Safety Science, vol. 88, pp. 211 - 218, Elsevier. (2016)

[42] Chockalingam, S., Hadžiosmanović, D., Pieters, W., Teixeira, A., van Gelder, P.: Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications, International Conference on Critical Information Infrastructures Security, pp. 50 - 62, Springer. (2016)

5

PROBABILITY ELICITATION FOR BAYESIAN NETWORKS TO DISTINGUISH BETWEEN INTENTIONAL ATTACKS AND ACCIDENTAL TECHNICAL FAILURES*

5.1. INTRODUCTION

Modern societies rely on proper functioning of Critical Infrastructures (CIs) in different sectors such as energy, transportation, and water management which is vital for economic growth and societal wellbeing. Over the years, CIs have become over-dependent on Industrial Control Systems (ICSs) to ensure efficient operations, which are responsible for monitoring and steering industrial processes as, among others, electric power generation, automotive production, and flood control. ICSs were originally designed for isolated environments [1]. Such systems were mainly susceptible to technical failures. The blackout in the Canadian province of Ontario and the North-eastern and Mid-western United States is a typical example of a technical failure in which the absence of alarm due to software bug in the alarm system left operators unaware of the need to redistribute power [2]. However, modern ICSs no longer operate in isolation, but use other networks to facilitate and improve business processes [3]. This increased connectivity, however,

*This chapter is submitted to a Journal as Chockalingam, S., Pieters, W., Teixeira, A., and van Gelder, P.: *"Probability Elicitation for Bayesian Networks to Distinguish between Intentional Attacks and Accidental Technical Failures"*

makes ICSs more vulnerable to cyber-attacks apart from technical failures. A cyber-attack on a German steel mill is a typical example in which adversaries made use of corporate network to enter into the ICS network [4]. As an initial step, the adversaries used both the targeted email and social engineering techniques to acquire credentials for the corporate network. Once they acquired credentials for the corporate network, they worked their way into the plant's control system network and caused damage to the blast furnace.

It is essential to distinguish between attacks and technical failures that would lead to abnormal behavior in the components of ICSs and take suitable measures. In most cases, the initiation of response strategy presumably aimed at technical failures would be ineffective in the event of a targeted attack, and may lead to further complications. For instance, replacing a sensor that is sending incorrect measurement data with a new sensor would be a suitable response strategy to technical failure of a sensor. However, this may not be an appropriate response strategy to an attack on the sensor as it would not block the corresponding attack vector. Furthermore, the initiation of inappropriate response strategies would delay the recovery of the system from adversaries and might lead to harmful consequences. Noticeably, there is a lack of decision support to distinguish between attacks and technical failures.

Bayesian Networks (BNs) have the capacity to tackle this challenge especially based on their real-world applications in medical diagnosis [5] and fault diagnosis [6]. BNs belong to the family of probabilistic graphical models [7], consisting of a qualitative and a quantitative part [8]. The qualitative part is a directed acyclic graph of nodes and edges. Each node represents a random variable, while the edges between the nodes represent the conditional dependencies among the random variables. The quantitative part takes the form of a priori marginal and conditional probabilities so as to quantify the dependencies between connected nodes.

In order to address the above-mentioned research gap, we developed a framework in our previous work to help construct BN models for distinguishing attacks and technical failures [9]. Furthermore, we extended and combined fishbone diagrams within our framework for knowledge elicitation to construct the qualitative part of such BN models. However, our previous work lacks a systematic method for knowledge elicitation to construct the quantitative part of such BN models. This present study aims to provide a holistic framework to help construct BN models for distinguishing attacks and technical failures by addressing the research question: "How could we elicit expert knowledge to effectively construct Conditional Probability Tables of Bayesian Network models for distinguishing attacks and technical failures?". The research objectives are:

- **RO 1.** To propose an approach that would help to effectively construct Conditional Probability Tables (CPTs) for our application.
- **RO 2.** To demonstrate the proposed approach using an example in the water management domain.

Expert knowledge is one of the predominant data sources utilised to populate conditional probability tables (CPTs) especially in domains where there is a limited availability of data like cyber security [10]. Probability elicitation is the most challenging part of constructing BN models especially when it relies on expert knowledge as we need to elicit probability for every possible combination of parent variables state to complete the CPT

of a child variable from experts. The CPT size of a child variable grows exponentially with the number of parents. For instance, the CPT size of a binary child with 5 binary parents is 64 ($2^{(5+1)}$) entries. The burden of probability elicitation could be reduced by: (i) reducing the number of conditional probabilities to elicit by imposing structural assumptions, and (ii) facilitating individual probability entry by providing visual aids to help experts answer elicitation questions in terms of probabilities [11]. We evaluate several techniques for reducing the number of probabilities to elicit, and conclude that DeMorgan models is most suitable for our purpose [12]. Furthermore, we review several methods for facilitating individual probability entry and conclude that probability scales with numerical and verbal anchors is most appropriate for our application [13, 14].

The remainder of this study is structured as follows. In Section 5.2, we illustrate the different layers and the components of an ICS and describe a case study in the water management domain that is used to demonstrate our proposed approach. In Section 5.3, we describe our existing framework that would help to construct BN models for distinguishing attacks and technical failures in addition to a systematic method for knowledge elicitation to construct the qualitative part of such BN models. Section 5.4 provides an essential foundation of techniques to reduce the burden of probability elicitation. In Section 5.5, we demonstrate the proposed approach using an example in the water management domain. Section 5.6 presents the conclusions and future work directions.

5.2. INDUSTRIAL CONTROL SYSTEMS

5.2.1. ICS ARCHITECTURE

Domain knowledge on ICSs is the starting point for the development and application of our proposed approach. We illustrated the three different layers and major components in each layer of an ICS in Section 4.4.1.

5.2.2. CASE STUDY OVERVIEW

The case study overview provided in Section 4.4.2 would be used to demonstrate our proposed approach that would help to effectively construct CPTs involving domain experts.

5.3. FRAMEWORK FOR DISTINGUISHING ATTACKS AND TECHNICAL FAILURES

This section describes the framework proposed in our previous work to construct BN models for distinguishing attacks and technical failures with an example [9]. The framework consists of three layers as shown in Figure 5.1. The middle layer consists of a problem variable which is the major cause for an abnormal behaviour in a component of the ICS (observable problem). In the example shown in Figure 5.1, we considered "Sensor (S_1) sends incorrect water level measurements" as the problem, which is observable. For instance, this problem could be observed by comparing the water level measurements sent by the sensor (S_1) against the measurements in the water level scale. We considered the major causes of the problem (intentional attack and accidental technical failure) as the states of the problem variable.

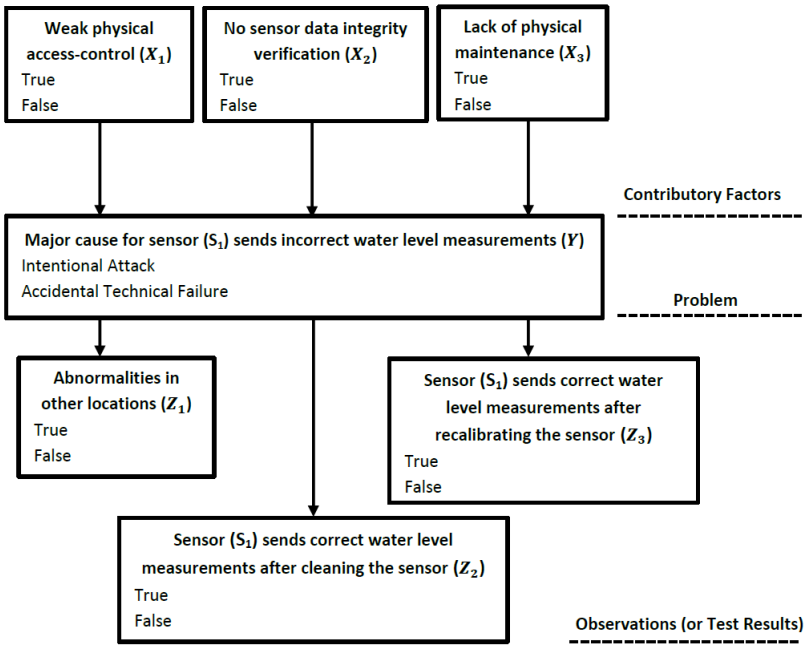


Figure 5.1: Framework for Distinguishing Attacks and Technical Failures: Example

The upper layer consists of factors contributing to the major causes of the problem. For instance, the factor "Weak physical access-control" contributes to "Sensor (S_1) sends incorrect water level measurements" due to intentional attack, whereas "Lack of physical maintenance" contributes to "Sensor (S_1) sends incorrect water level measurements" due to accidental technical failure. The lower layer consists of observations (or test results) which is defined as any information useful for determining the major cause of the problem based on the outcome of tests. For instance, the outcome of the test whether "Sensor (S_1) sends correct water level measurements after cleaning the sensor" would provide additional information to determine the major cause (accidental technical failure or intentional attack) of the problem accurately.

The framework which we proposed in our previous work includes a systematic method based on fishbone diagrams for knowledge elicitation to construct the qualitative part of BN models [9]. We adopted this approach because there are challenges to solely rely on BNs for knowledge elicitation to construct the qualitative part of BN models. It is not easy-to-use for knowledge elicitation involving domain experts as it could be time-consuming for elicitors to explain the notion of BNs [9]. Furthermore, it could not encourage and guide data collection by showing where knowledge is lacking as it is not well-structured. On the other hand, fishbone diagrams help to systematically identify and organise the possible contributing factors (or sub-causes) of a particular problem [15–19]. We extended fishbone diagrams to incorporate observations (or test results) in our previous work, which needs to be elicited for our application in addition to contributory factors (or sub-causes).

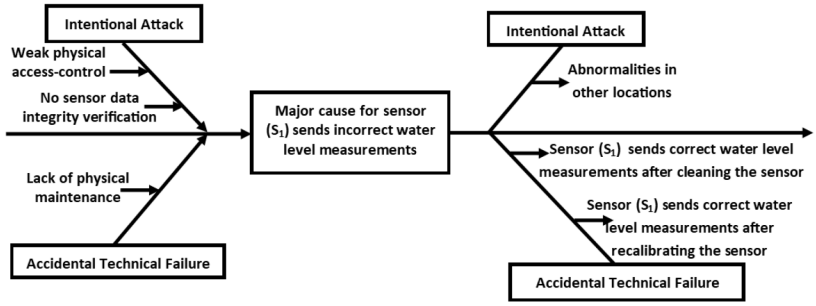


Figure 5.2: Extended Fishbone Diagram: Example

Figure 5.2 shows an example extended fishbone diagram which consists of a problem ("Major cause for sensor (S_1) sends incorrect water level measurements"), its possible contributing factors (or sub-causes) sorted and related under different categories on the left side of the problem. Each category on the left side of the problem represents the major causes of the problem (intentional attack and accidental technical failure). Our example shows that "Lack of physical maintenance" is the contributing factor to the problem ("Sensor (S_1) sends incorrect water level measurements") due to accidental technical failure. Furthermore, the observations (or test results) on the right side of the problem would provide additional information to determine the major cause of the problem accurately. Each category on the right side of the problem are used for reference to elicit observations (or test results) that would be useful for determining the particular major causes of the problem [9]. Our example shows that the observation "abnormalities in other locations" would increase the probability of the problem ("Sensor (S_1) sends incorrect water level measurements") due to intentional attack.

Once the extended fishbone diagram is developed, it would be translated into a corresponding qualitative BN model based on the mapping scheme in our previous work [9]. However, the proposed framework lacked a systematic method for knowledge elicitation to construct the quantitative part of BN models (the CPTs), which we address in the current work.

5.4. TECHNIQUES FOR REDUCING THE BURDEN OF PROBABILITY ELICITATION

Probability elicitation is a challenging task in building BNs, especially when it relies heavily on expert knowledge [11]. The extensive workload for experts in probability elicitation could affect the reliability of elicited probabilities. However, the workload for experts in probability elicitation could be reduced by reducing the number of conditional probabilities to elicit and facilitating individual probability entry.

5.4.1. TECHNIQUE FOR REDUCING THE NUMBER OF CONDITIONAL PROBABILITIES TO ELICIT

This section analyses well-known techniques and describes the most suitable technique for our application, which would help to reduce the number of conditional probabilities to elicit.

In order to reduce the number of conditional probabilities to elicit, we could exploit the causal independence models [11]. Causal independence refers to the situation where the contributory factors (causes) contribute independently to the problem (effect) [20]. By utilizing these models, only a number of parameters that is linear in the number of contributory factors is needed to be elicited to define a full CPT for the problem variable as the total influence on the problem is a combination of the individual contributions [21]. As an example, we shall consider the BN model depicted in Figure 5.1, where the problem variable (Y) is a binary discrete variable with the states "Intentional Attack" and "Accidental Technical Failure". In the CPT shown in Figure 5.3, Y = "Intentional Attack" denotes Y = "True", and Y = "Accidental Technical Failure" denotes Y = "False". We translated the states of Y into "True" and "False" to comply with the inherent assumptions of the noisy-OR model with regard to the states of variables. The typical state of each variable in the noisy-OR model is "False". For instance, the typical state of a child variable (Fever) in the noisy-OR model is "False" as it is normal. Therefore in our application, we assigned Y = "False" for Y = "Accidental Technical Failure" as this is the a priori expected major cause, based on the higher frequency of technical failures compared to the intentional attacks [9].

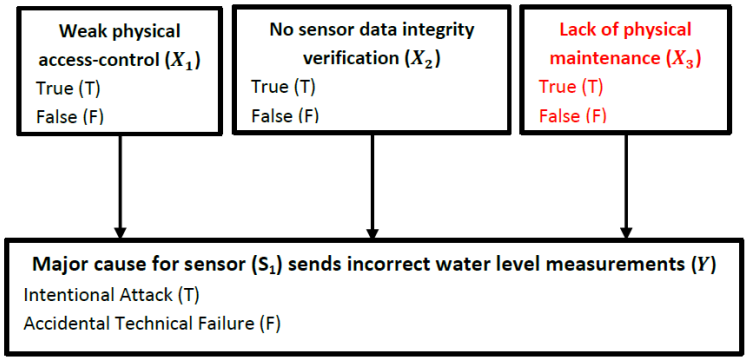
In our application, we are dealing with a combination of promoting and inhibiting influences. In case of a promoting influence, the presence (or absence) of the parent will result in the child event with a certain probability. When the parent is absent (or present), it is certain not to cause the child event. In other words, the presence (or absence) of the contributory factor will result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "intentional attack" with a certain probability as it denotes "True" state. For instance, the presence of "Weak physical access-control" will result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "intentional attack" with a certain probability, whereas the absence of "Weak physical access-control" will not certainly result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "intentional attack". This type of promoting influence is defined as a cause [12]. On the other hand, the absence of "Sensor data integrity verification" will result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "intentional attack" with a certain probability, whereas the presence of "Sensor data integrity verification" will not certainly result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "intentional attack". This type of promoting influence is defined as a barrier [12].

In case of an inhibiting influence, the presence (or absence) of the parent will inhibit the child event with a certain probability. When the parent is absent (or present), it is certain not to inhibit the child event. In other words, the presence (or absence) of the contributory factor will result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "accidental technical failure" with a certain probability as it denotes "False" state. For instance, the presence of "Lack of physical maintenance" will result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "accidental technical failure" with a certain probability, whereas the absence of "Lack of physical maintenance" will not certainly result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "accidental technical failure". This type of inhibiting influence is defined as an inhibitor [12]. On the other hand, the absence

of "Well-written maintenance procedure" will result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "accidental technical failure" with a certain probability, whereas the presence of "Well-written maintenance procedure" will not certainly result in the problem ("Sensor (S_1) sends incorrect water level measurements") due to "accidental technical failure". This type of inhibiting influence is defined as a requirement [12].

Our example BN model shows that it possesses a mixture of promoting and inhibiting influences (causes and inhibitors) especially with regard to the interaction between the contributory factors and the problem. Therefore, we need a technique that would help to model opposing influences as we deal with a mixture of promoting and inhibiting influences in our application, which would help to reduce the number of conditional probabilities to elicit.

We analysed several techniques and chose the most suitable technique for our application which would be described in next Section. The description of techniques that are unsuitable for our application can be found in Appendix B which includes the noisy-OR model and Causal Strength (CAST) logic. The noisy-OR model is one of the most commonly used causal independence models which helps to reduce the number of conditional probabilities to elicit [5, 22]. The noisy-OR model inherently assumes binary variables [23]. The noisy-MAX model is an extension of the noisy-OR model which is suitable for multi-valued variables [24]. We analysed the noisy-OR model as we deal with only binary variables in our application.



X_1	X_2	X_3	Y	
			Intentional Attack (Y = True)	Accidental Technical Failure (Y = False)
True	True	True	Calculated	$1 - P(Y = T \mid X_1 = T, X_2 = T, X_3 = T)$
True	True	False	Calculated	$1 - P(Y = T \mid X_1 = T, X_2 = T, X_3 = F)$
True	False	True	Calculated	$1 - P(Y = T \mid X_1 = T, X_2 = F, X_3 = T)$
True	False	False	Elicited	$1 - P(Y = T \mid X_1 = T, X_2 = F, X_3 = F)$
False	True	True	Calculated	$1 - P(Y = T \mid X_1 = F, X_2 = T, X_3 = T)$
False	True	False	Elicited	$1 - P(Y = T \mid X_1 = F, X_2 = T, X_3 = F)$
False	False	True	Elicited	$1 - P(Y = T \mid X_1 = F, X_2 = F, X_3 = T)$
False	False	False	0	1 (based on the property of accountability)

Figure 5.3: Application of Noisy-OR: Problem

The noisy-OR model assumes that the properties of exception independence and accountability hold true [25]. In case all the modelled contributory factors of the problem ("Sensor (S_1) sends incorrect water level measurements") are false, the property of accountability requires that the problem be presumed false ("Sensor (S_1) sends incorrect water level measurements" due to "accidental technical failure"). However, this would not work for inhibiting influences such as "Lack of physical maintenance" in the noisy-OR model as shown in Figure 5.3. In case "Lack of physical maintenance" is absent, it is certain not to inhibit the problem which is incompatible with the property of accountability. Therefore, we found that the noisy-OR model is unsuitable for the purposes of our application because the property of accountability does not hold true.

Alternatively, CAST logic is one of the techniques mainly developed for modelling opposing influences [26]. CAST logic assumes all the variables in the model are binary. The parameters which need to be elicited to completely define CPTs using CAST logic are: (i) causal strengths for each edge, and (ii) baseline probability for each variable. The baseline probability of a parent variable can be interpreted as the prior probability of the corresponding parent variable. However, it would not be appropriate to interpret the baseline probability of the child variable as a prior probability or a leak probability, as the parent variables have no state in which they are guaranteed to have no influence on the child variable [27]. As the definition of baseline probability of child variable is not clear, we cannot formulate appropriate question to elicit baseline probability of child variable. This is the major disadvantage of CAST logic which resulted in the lack of practical applications [12, 27]. We conclude that neither the noisy-OR model nor the CAST logic is suitable for the purposes of our application.

DEMORGAN MODEL

As an alternative to the previously discussed models, the DeMorgan model can potentially be used to tackle the challenge of modelling opposing influences, which would help to reduce the number of conditional probabilities to elicit. This section explains the DeMorgan model.

The DeMorgan model is a technique mainly developed for modelling opposing influences, which would help to reduce the number of conditional probabilities to elicit [12, 27]. The DeMorgan model is applicable when there are several parents and a common child. The DeMorgan model inherently assumes binary variables. The DeMorgan model assumes that one of the two states of each variable is always the distinguished state as shown in Figure 5.4. Usually such state of the child variable depends on the modelled domain [28]. This is a typical state of the corresponding child variable [29]. In case the child variable consists of two states ("disease", "no disease") in the medical domain, the distinguished state of the corresponding child variable is chosen as "no disease" as it is normal [28]. In our application, the distinguished state of the problem variable ("Major cause for sensor (S_1) sends incorrect water level measurements") is chosen as "accidental technical failure" as this is the a priori expected major cause, based on the higher frequency of technical failures compared to the intentional attacks [12]. The distinguished state of a parent variable is relative to the type of causal interaction with the child variable [12]. The same parent variable can have different distinguished states in different interactions that it participates in with the different child variables.

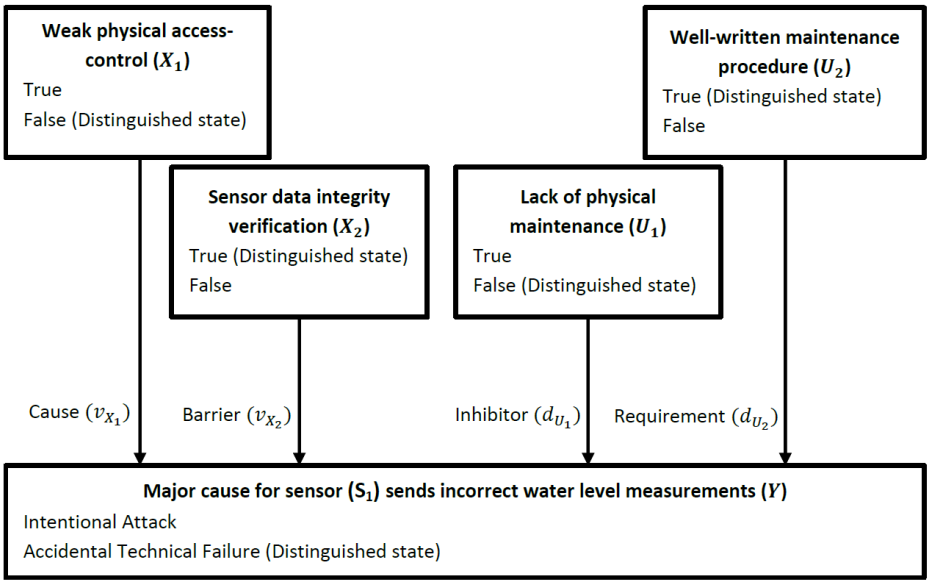


Figure 5.4: DeMorgan Model: Causal Interaction Types

There are 4 different types of causal interactions between an individual parent (X) and a child (Y) in the DeMorgan model: (i) cause, (ii) barrier, (iii) inhibitor, and (iv) requirement.

(i) Cause: X is a causal factor and has a positive influence on Y. In this type of causal interaction between an individual parent (X) and a child (Y), the distinguished state of the corresponding parent variable is "False" [12]. Consequently, when the parent variable is "False", it is certain not to trigger a change from the typical state of the child variable as shown in Table 5.1. When the parent variable is "True", it will trigger a change from the typical state of the child variable, with a certain probability (v_X) as shown in Table 5.1.

X	Y	
	Intentional Attack	**Accidental Technical Failure**
True	v_X	$1 - v_X$
False	0	1

Table 5.1: Type of Causal Interaction: Cause

(ii) Barrier: This is a negated counterpart of cause, i.e., $X^{'}$ is a causal factor and has a positive influence on Y. In this type of causal interaction between an individual parent X and a child Y, the distinguished state of the corresponding parent variable is "True" [12]. Accordingly, when the parent variable is "True", it is certain not to trigger a change from the typical state of the child variable as shown in Table 5.2. When the parent variable is "False", it will trigger a change from the typical state of the child variable, with a certain probability (v_X) as shown in Table 5.2.

X	Y	
	Intentional Attack	Accidental Technical Failure
True	0	1
False	v_X	$1 - v_X$

Table 5.2: Type of Causal Interaction: Barrier

(iii) Inhibitor: X inhibits Y. In this type of causal interaction between an individual parent X and a child Y, the distinguished state of the corresponding parent variable is "False" [12]. As a result, when the parent variable is "False", it is certain not to prevent a change from the typical state of the child variable as shown in Table 5.3. When the parent variable is "True", it will prevent a change from the typical state of the child variable, with a certain probability (d_X) as shown in Table 5.3.

X	Y	
	Intentional Attack	Accidental Technical Failure
True	$1 - d_X$	d_X
False	1	0

Table 5.3: Type of Causal Interaction: Inhibitor

(iv) Requirement: The relationship between an inhibitor and requirement is similar to the relationship between a cause and barrier. $X^{'}$ inhibits Y. In this type of causal interaction between an individual parent (X) and a child (Y), the distinguished state of the corresponding parent variable is "True" [12]. Hence, when the parent is "True", it is certain not to prevent a change from the typical state of the child variable as shown in Table 5.4. When the parent variable is "False", it will prevent a change from the typical state of the child variable, with a certain probability (d_X) as shown in Table 4.

X	Y	
	Intentional Attack	Accidental Technical Failure
True	1	0
False	$1 - d_X$	d_X

Table 5.4: Type of Causal Interaction: Requirement

The DeMorgan model is an extension and a combination of the noisy-OR and noisy-AND model which supports modelling the above-mentioned types of causal interactions [12]. Maaskant et al. modelled promoting influences which includes causes and barriers by mimicking the noisy-OR model [12]. Furthermore, Maaskant et al. modelled inhibiting influences which includes inhibitors and requirements by mimicking the noisy-AND model [12]. Finally, Maaskant et al. modelled the combination of promoting and inhibiting influences by combining the noisy-OR and noisy-AND model.

The property of accountability in the noisy-OR model is applicable to the DeMorgan model with a slight modification as it also exploits causal independence: In case all the modelled parents of the child are in their distinguished state, the property of accountability requires that the child be presumed their distinguished state. However, in many cases,

this is not a realistic assumption as it is difficult to capture all the possible parents of the child [23]. Specifically, this is not realistic in our example as it is difficult to capture all the possible contributory factors of the problem ("Sensor (S_1) sends incorrect water level measurements") due to "intentional attack". In the DeMorgan model, the leak parameter (υ_{XL}) deals with the possible parents of the child that are not previously known and explicitly modelled.

In general, the size of the CPT of a binary variable with n binary parents is $2^{(n+1)}$. However, only n+1 parameters are sufficient to completely define CPT using the DeMorgan model as it exploits causal independence. In the example shown in Figure 5.4, only 5 parameters are sufficient to completely define the CPT of child variable (Y) using the DeMorgan model instead of 64 entries. There are 2 different parameterisations for the Noisy-OR model with a leak parameter (the Leaky Noisy-OR model) proposed by Henrion [30] and Diez [24] which are mathematically equivalent. These 2 parameterisations differ only in the type of question that needs to be asked to the experts for knowledge elicitation. Henrion's parameterisation is supported by a question like: *"What is the probability that the effect is true given that a cause (X_i) is true and all the modelled causes are false?"*. On the other hand, Diez's parameterisation is supported by a question like: *"What is the probability that the effect is true given that a cause (X_i) is true and all other modelled and unmodelled causes are false?"*. The DeMorgan model utilised the Diez's parameterisation with a slight modification.

We could find the values for required parameters from the experts to completely define CPT using the DeMorgan model based on appropriate question for each type of causal interaction detailed below:

(i) The leak parameter: To find the value for the leak parameter, the elicitor could ask experts: *"What is the probability that the child is in their non-distinguished state given that the parents are in their distinguished states?"*. In our example shown in Figure 5.4, the elicitor could ask experts to find the value for parameter (υ_{XL}): *"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is intentional attack given that the physical access-control for sensor (S_1) is strong, data integrity verification is performed for sensor (S_1) data, sensor (S_1) is always physically maintained, maintenance procedure for sensor (S_1) is well-written?"*.

(ii) Cause: To find the value for parameter corresponding to a cause, the elicitor could ask experts: *"What is the probability that the child is in their non-distinguished state given that all the parents are in their distinguished states, except (X_i)* and no other unmodelled causal factors are present?". In our example shown in Figure 5.4, the elicitor could ask experts to find the value for parameter (υ_{X1}): *"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is intentional attack given that the physical access-control for sensor (S_1) is weak, data integrity verification is performed for sensor (S_1) data, sensor (S_1) is always physically maintained, maintenance procedure for sensor (S_1) is well-written, and no other unmodelled causal factors are present?"*.

(iii) Barrier: To find the value for parameter corresponding to a barrier, the elicitor could ask experts: *"What is the probability that the child is in their non-distinguished state given that all the parents are in their distinguished states, except (X_i) and no other unmodelled causal factors are present?"*. In our example shown in Figure 5.4, the elicitor

could ask experts to find the value for parameter (v_{X2}): *"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is intentional attack given that the physical access-control for sensor (S_1) is strong, data integrity verification is not performed for sensor (S_1) data, sensor (S_1) is always physically maintained, maintenance procedure for sensor (S_1) is well-written, and no other unmodelled causal factors are present?"*.

(iv) Inhibitor: Maaskant et al. did not directly determine the value for parameters corresponding to inhibitors similar to causes and barriers as it is not practical for the example which they considered [27]. Specifically, it makes less sense to ask for the effect of presence of parent ("Rain") on the child ("Bonfire"), when the child ("Bonfire") is "False". Therefore, they determined the value for parameter corresponding to each inhibitor by determining the negative influence relative to an arbitrary (non-empty) set of causes/barriers/leak parameter. However, in our application, it is possible to determine the value for parameter corresponding to inhibitors directly as we ask for the effect of presence of parent ("Lack of physical maintenance") on the child ("Major cause for sensor (S_1) sends incorrect water level measurements"), when the latter ("Major cause for sensor (S_1) sends incorrect water level measurements") is "Accidental technical failure". In order to find the value for parameter corresponding to an inhibitor directly, the elicitor could ask the experts: *"What is the probability that the child is in their distinguished state given that the parents are in their distinguished states, except (U_i) and no other unmodelled causal factors are present?"*. In our example shown in Figure 5.4, the elicitor could ask experts to find the value for parameter (d_{U1}): *"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is accidental technical failure given that the physical access-control for sensor (S_1) is strong, data integrity verification is performed for sensor (S_1) data, sensor (S_1) is not always physically maintained, maintenance procedure for sensor (S_1) is well-written and no other unmodelled causal factors are present?"*.

(v) Requirement: Maaskant et al. did not directly determine the value for parameters corresponding to requirements similar to causes and barriers as it is not practical for the example which they considered [27]. Specifically, it makes less sense to ask for the effect of absence of parent on the child, when the child is "False". Therefore, they determined the value for parameter corresponding to each requirement by determining the negative influence relative to an arbitrary (non-empty) set of causes/barriers/leak parameter. However in our application, it is possible to determine the value for parameter corresponding to requirements directly as we ask for the effect of absence of parent ("Well-written maintenance procedure") on the child ("Major cause for sensor (S_1) sends incorrect water level measurements"), when the latter ("Major cause for sensor (S_1) sends incorrect water level measurements") is "Accidental technical failure". In order to find the value for parameter corresponding to a requirement directly, the elicitor could ask the experts: *"What is the probability that the child is in their distinguished state given that the parents are in their distinguished states, except U_i and no other unmodelled causal factors are present?"*. In our example shown in Figure 5.4, the elicitor could ask experts to find the value for parameter (d_{U2}): *"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is accidental technical failure given that the physical access-control for sensor (S_1) is strong, data integrity*

verification is performed for sensor (S_1) data, sensor (S_1) is always physically maintained, maintenance procedure for sensor (S_1) is not well-written and no other unmodelled causal factors are present?".

Once we determine the required parameters based on appropriate elicitation questions, we can completely define the CPT of the child variable using (1):

$$P(y|X,U) = \left(1 - (1 - v_{X_L}) \prod_{X_i \in +X} (1 - v_{X_i})\right) \prod_{U_i \in +U} (1 - d_{U_i}) \tag{1}$$

In the equation (1), Y represents the effect variable which has values y for the effect being in the non-distinguished state ("Intentional attack") and y' for the effect being in the distinguished state ("Accidental technical failure"). X denotes the set of parents which interact with the effect variable as promoting influences, U denotes the set of parents which interact with the effect variable as inhibiting influences, $+X$ denotes the subset of X that contains all parents that are in their non-distinguished states, $+U$ denotes the subset of U that contains all parents that are in their non-distinguished states. v_{XL} denotes the leak parameter which expresses the probability of y ("Intentional attack") given all parents are in their distinguished states, (v_{Xi}) denotes the probability of y ("Intentional attack") given that the parent X_i is not in its distinguished state and all other parents are in their distinguished states, d_{Ui} denotes the probability of y' ("Accidental technical failure") given that the parent U_i is not in its distinguished state and all other parents are in their distinguished states.

We choose the DeMorgan model for our application to reduce the number of conditional probabilities to elicit as they support modeling opposing influences with clear parameterisations.

5.4.2. TECHNIQUE FOR FACILITATING INDIVIDUAL PROBABILITY ENTRY

This section explains our chosen technique for facilitating individual probability entry for our application.

Our systematic method for knowledge elicitation to construct CPTs of BN models would be incomplete without a technique that facilitates individual probability entry. The DeMorgan models would help to reduce the number of conditional probabilities to elicit and allow elicitors to ask appropriate questions during probability elicitation. In addition, there should be a suitable technique which would make it easy for experts to answer elicitation questions in terms of probabilities.

There are well-known methods such as probability scale [13, 31], and probability wheel [32] which would help to facilitate individual probability entry [11, 33]. The basis for choosing a particular method includes accuracy, less probability elicitation time, and usability [33]. Wang et al. compared three different methods: (i) direct estimation, (ii) probability wheel and (iii) probability scale in terms of their accuracy and time taken to elicit probabilities from experts [34]. They pointed out that probability scale is better in terms of both accuracy and probability elicitation time compared to the other two methods.

A probability scale can be a horizontal or vertical line with several anchors [33]. Figure 5.5 shows a probability scale with 7 numerical and verbal anchors [35]. However, there are several variants of probability scales which would help to facilitate individual probability entry. Witteman et al. compared 3 probability scales: (i) probability scale with numerical and verbal anchors, (ii) probability scale with only numerical anchors, and (iii) probability scale with only verbal anchors [36]. They compared 3 probability scales based on a study with general practitioners in the domain of medical decision making. They concluded that all 3 probability scales are equally suitable to facilitate individual probability entry. However, they recommended the probability scale with numerical and verbal anchors to facilitate individual probability entry as it provides numerical anchors for experts who prefer numbers and verbal anchors for experts who prefer words. Furthermore, Witteman et al. compared 2 different probability scales: (i) probability scale with numerical and verbal anchors, (ii) probability scale with only numerical anchors [37]. They compared 2 probability scales based on a study with arts and mathematics students. They concluded that the probability scale with numerical and verbal anchors is more comfortable to use compared to the probability scale with only numerical anchors.

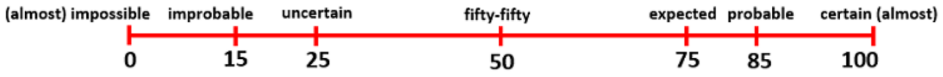


Figure 5.5: Probability Scale with Numerical and Verbal Anchors

There are real-world applications of the probability scale with numerical and verbal anchors in the elicitation of probabilities to construct the quantitative part of BN models [13, 31]. Van der Gaag et al. used the probability scale with numerical and verbal anchors for a case study in oesophageal cancer [13]. This study was conducted with two experts in gastrointestinal oncology. The experts found that this method is easier to use than any other method they used before. Van der Gaag et al. also highlighted that the large number of probabilities are elicited in a reasonable time using this method. Furthermore, Hanninen et al. used the probability scale with numerical and verbal anchors for the construction of quantitative part of collision and grounding BN model [31]. This study was conducted with 8 experts who possessed maritime working experience between 3 and 30 years. These studies show that the probability scale with numerical and verbal anchors can be used for facilitating individual probability entry involving experts with different background.

We choose probability scales for our application as they are better in terms of accuracy and probability elicitation time compared to other methods. In particular, we would employ the probability scale with numerical and verbal anchors to facilitate individual probability entry in our application as they are effective and practicable based on previous studies. We would utilise the probability scale with 7 numerical and verbal anchors to facilitate individual probability entry with a variation. In our application, the experts could mark the suitable probability among 7 anchors in the scale directly or express fine-grained probabilities using the probability scale with numerical and verbal anchors as a supporting aid to visualise the probability range. This is convenient when the experts would like to express fine-grained probabilities based on historical data which is realistic for accidental technical failures in our application.

5.5. APPLICATION OF THE METHODOLOGY

In this section, we use an illustrative case of a floodgate in the Netherlands to explain how we effectively construct CPTs of BN models for distinguishing attacks and technical failures.

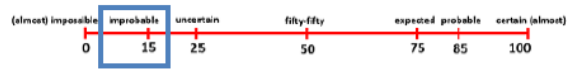
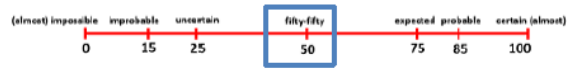
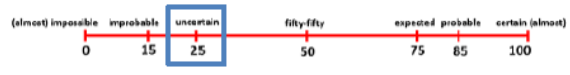
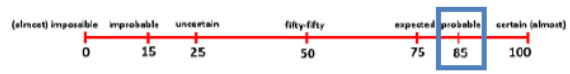
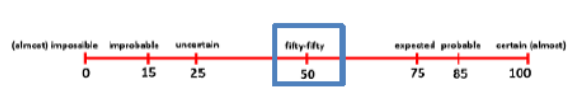
Major cause for sensor (S_1) sends incorrect water level measurements (Y)		
v_{X_L}	*"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is intentional attack ~~given that the physical access control for sensor (S_d) is strong, data integrity verification is performed for sensor (S_d) data, sensor (S_d) is always physically maintained, maintenance procedure for sensor (S_d) is well-written~~?"*	
v_{X_1}	*"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is intentional attack given that the physical access-control for sensor (S_1) is weak~~, data integrity verification is performed for sensor (S_d) data, sensor (S_d) is always physically maintained, maintenance procedure for sensor (S_d) is well written~~, and no other unmodelled causal factors are present?"*	
v_{X_2}	*"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is intentional attack given that the ~~physical access control for sensor (S_d) is strong~~, data integrity verification is not performed for sensor (S_1) data~~, sensor (S_d) is always physically maintained, maintenance procedure for sensor (S_d) is well written~~, and no other unmodelled causal factors are present?"*	
d_{U_1}	*"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is accidental technical failure given that the ~~physical access control for sensor (S_d) is strong, data integrity verification is performed for sensor (S_d) data~~, sensor (S_1) is not always physically maintained~~, maintenance procedure for sensor (S_d) is well-written~~, and no other unmodelled causal factors are present?"*	
d_{U_2}	*"What is the probability that the major cause for the observed problem (sensor (S_1) sends incorrect water level measurements) is accidental technical failure given that the ~~physical access control for sensor (S_d) is strong, data integrity verification is performed for sensor (S_d) data, sensor (S_d) is always physically maintained~~, maintenance procedure for sensor (S_1) is not well-written, and no other unmodelled causal factors are present?"*	

Table 5.5: Parameter Elicitation for the Problem Variable Y: Example

In this Table 5.5, the double strikethrough text denotes the child variables being in its distinguished state.

We considered the upper and middle layer of our framework for the application of our methodology. It is important to reduce the number of conditional probabilities to elicit for the problem variable as a considerable number of contributory factors (upper layer), corresponding to intentional attack and accidental technical failure, typically

interact with the problem variable (middle layer), which in turn increases the CPT size of the problem variable exponentially. On the other hand, the conditional probabilities for observations (or test results) (lower layer) would be easy to elicit directly as there is only one problem variable (middle layer) in our framework, which makes the CPT size of an observation (or test result) variable to 4 ($2^{(1+1)}$). We shall consider the BN model with the upper and middle layer of our framework depicted in Figure 5.4 for the application of our methodology. We considered the problem "Sensor (S_1) sends incorrect water level measurements" as it could develop more complex situations in the case of floodgate. In case the floodgate closes when it should not based on the incorrect water level measurements sent by the sensor (S_1), it would lead to severe economic damage, for instance, by delaying cargo ships. On the other hand, in case the floodgate opens when it should not due to incorrect water level measurements sent by the sensor (S_1), it would lead to flooding.

X_1	X_2	U_1	U_2	Y	
				Intentional Attack	Accidental Technical Failure
True	True	True	True	0.09	0.91
True	True	True	False	0.04	0.96
True	True	False	True	0.50 (v_{X_1})	0.50
True	True	False	False	0.29	0.71
True	False	True	True	0.10	0.90
True	False	True	False	0.05	0.95
True	False	False	True	0.68	0.32
True	False	False	False	0.34	0.66
False	True	True	True	0.15	0.85 (d_{U_1})
False	True	True	False	0.01	0.99
False	True	False	True	0.15 (v_{X_L})	0.85
False	True	False	False	0.50	0.50 (d_{U_2})
False	False	True	True	0.05	0.95
False	False	True	False	0.03	0.97
False	False	False	True	0.25 (v_{X_2})	0.75
False	False	False	False	0.18	0.82

Table 5.6: Application of the DeMorgan Model: CPT Example

We considered 4 contributory factors to the major causes (intentional attack or accidental technical failure) of the observed problem: (i) Weak physical access-control (X_1), (ii) Sensor data integrity verification (X_2), (iii) Lack of physical maintenance (U_1), and (iv) Well-written maintenance procedure (U_2) as shown in Figure 5.4 to depict each type of causal interaction. The type of causal interaction between individual parent X_1 and the child Y is cause. The type of causal interaction between individual parent X_2 and the child Y is barrier. The type of causal interaction between individual parent U_1 and the child Y is inhibitor. The type of causal interaction between individual parent X_2 and the child Y is requirement. In this example, we need to elicit only 5 (4+1) parameters instead of 32 ($2^{(4+1)}$) to completely define CPT for the problem variable. The 5 parameters which

we need to elicit are: (υ_{XL}), (υ_{X1}), (υ_{X2}), (d_{U1}), (d_{U2}).

The values for these 5 parameters could be elicited from experts by providing the appropriate elicitation questions based on the DeMorgan model and the probability scale with numerical and verbal anchors, which could help experts answer in terms of probabilities to elicitation questions as shown in Table 5.5. The double strikethrough text in Table 5.5 makes the probability elicitation process simple as they do not affect the corresponding probability based on our structural assumptions. The experts could directly read the remaining text and mark the answer for each question in Table 5.5 which could also reduce probability elicitation time. Suppose the expert marks the answer for (υ_{XL}) as 0.15, (υ_{X1}) as 0.50, (υ_{X2}) as 0.25, (d_{U1}) as 0.85, (d_{U2}) as 0.50. These probabilities are examples to demonstrate the application of the methodology.

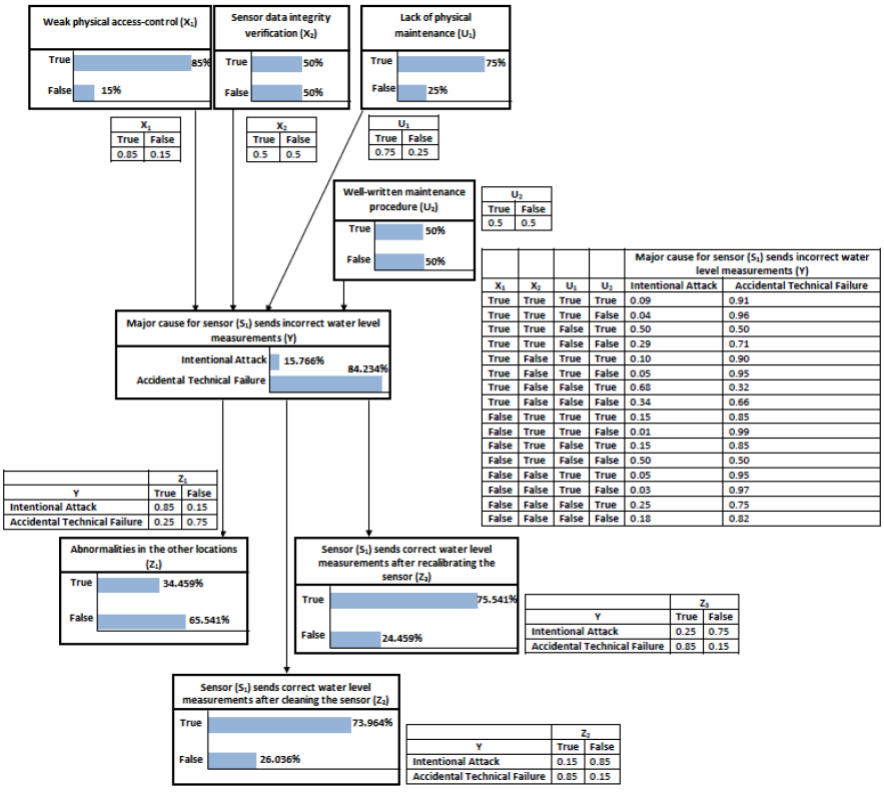


Figure 5.6: BN Model with CPTs Example

Once we elicit all the required parameters, we could use (1) to completely define CPT for our example BN model. For instance, we could use (1) to calculate: P($Y|X_1$',X_2',U_1, U_2')=(1-(1-0.15)(1-0.25))(1-0.85)(1-0.50) = 0.03. The red coloured text in Table 5.6 denotes this probability. The completed CPT for the problem variable (Y) is shown in Table 5.6.

Once we complete the CPT for the problem variable, we could define the a priori probabilities for each contributory factor and observation (or test result) by eliciting corresponding probabilities directly from the experts as they are not complicated. An

example BN model with corresponding CPTs for each variable is shown in Figure 5.6.

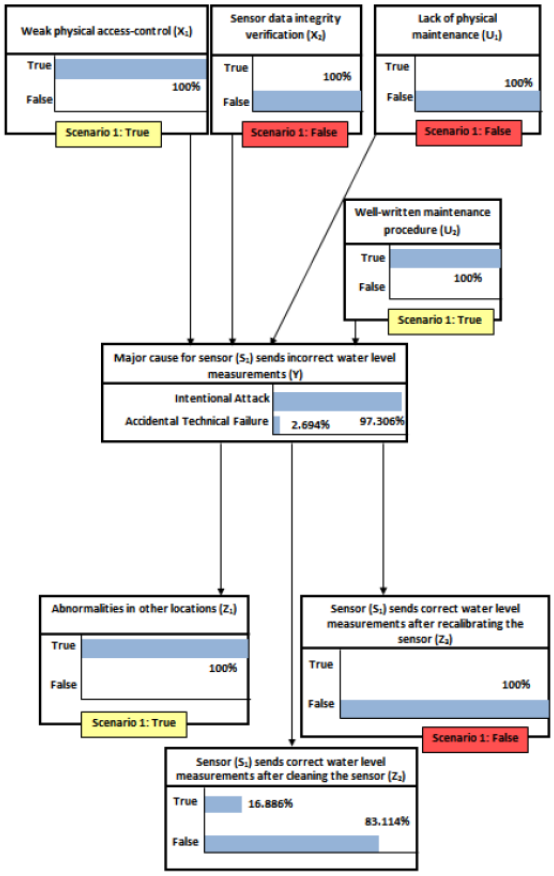


Figure 5.7: BN with Updated Probabilities Based on the Evidence

Once the problem ("Sensor (S_1) sends incorrect water level measurements") is observed in the floodgate, the evidence (True/False) contributory factors and observations (or test results) could be set by the operator (or end-user) to determine the major cause for the observed problem. Once the evidence for contributory factors and observations (or test results) is set, the posterior probability of the problem variable would be computed accordingly. Based on the computed posterior probability, the appropriate response strategy could be put in place for the most likely major cause (intentional attack/accidental technical failure) for the observed problem ("Sensor (S_1) sends incorrect water level measurements") thereby minimising negative consequences.

In the example shown in Figure 5.7, we provided the evidence for the contributory factors "Weak physical access-control (X_1) = True", "Sensor data integrity verification (X_2) = False", "Lack of physical maintenance (U_1) = False", "Well-written maintenance procedure (U_2) = True", and observation (or test result) "Abnormalities in other locations (Z_1) = True", "Sensor ($S1$) sends correct water level measurements after recalibrating the sensor

((Z_3)) = False". On the other hand, we did not provide the evidence for the problem "Major cause for sensor (S*1*) sends incorrect water level measurements (*Y*)" and observation (or test result) "Sensor (S_1) sends correct water level measurements after cleaning the sensor (Z_2)". The BN computes the posterior (updated) probabilities of the other nodes (*Y*, and (Z_2)) based on the provided evidence. The BN in Figure 5.7 determines that the major cause for the observed problem "Sensor (S_1) sends incorrect water level measurements" is most likely due to intentional attack as the corresponding posterior probability (0.97306) is higher compared to the posterior probability of accidental technical failure (0.02694).

5.6. CONCLUSIONS AND FUTURE WORK

Limited availability of data is one of the key challenges to construct BN models in domains like cyber security which result in modellers depending on expert knowledge. However, BNs are not suitable for knowledge elicitation involving domain experts. In our previous work, we developed a systematic method using fishbone diagrams for knowledge elicitation involving domain experts to construct the DAGs of BN models for distinguishing attacks and technical failures. Noticeably, the systematic method for knowledge elicitation involving domain experts to construct the CPTs of such BN models is missing in our previous work.

In this study, we utilised (a) DeMorgan models to reduce the number of conditional probabilities to elicit and (b) probability scales with numerical and verbal anchors to facilitate individual probability entry. We thereby reduce the burden of probability elicitation, which is critical for BN models that rely on expert knowledge. The proposed approach ensures the reliability of elicited probabilities by reducing the workload of experts in probability elicitation, especially DeMorgan models reduces the number of parameters that need to be elicited from exponential to linear in the number of parents to define a full CPT for the child variable. The proposed approach also completes a holistic framework to distinguish between attacks and technical failures by proposing a systematic method for probability elicitation involving domain experts.

Furthermore, we demonstrated the proposed approach with an example problem of incorrect sensor measurements in the water management domain. Our holistic framework is directly applicable to different domains for knowledge elicitation involving domain experts to construct BN models for distinguishing attacks and technical failures. The constructed BN models could be used by operators/end-users in different domains to determine the major cause (intentional attack or accidental technical failure) of an abnormal behavior in a component of the ICS, and initiate appropriate response strategies to minimise negative consequences.

In the future, we aim to evaluate our proposed framework by constructing BN models for observable problems in the water management domain involving domain experts. In addition, we aim at addressing the limitation that the DeMorgan model is suitable for binary variables only. In order to be able to reduce the number of conditional probabilities to elicit involving parents and/or child with more than two states, it is important to extend the DeMorgan model for multi-valued variables in the future.

REFERENCES

[1] Effendi, A., Davis, R.: ICS and IT: Managing Cyber Security Across the Enterprise, In: SPE Middle East Intelligent Oil and Gas Conference and Exhibition, Society of Petroleum Engineers. (2015)

[2] Zhivich, M., Cunningham, R. K.: The Real Cost of Software Errors, IEEE Security & Privacy, vol. 7, pp. 87 - 90. (2009)

[3] Knowles, W., Prince, D., Hutchison, D., Disso, J. F. P., Jones, K.: A Survey of Cyber Security Management in Industrial Control Systems, International Journal of Critical Infrastructure Protection, vol. 9, pp. 52 - 80. (2015)

[4] RISI.: German Steel Mill Cyber Attack, Available: http://www.risidata.com/database/detail/german-steel-mill-cyber-attack,2014

[5] Nikovski, D.: Constructing Bayesian Networks for Medical Diagnosis from Incomplete and Partially Correct Statistics, IEEE Transactions on Knowledge and Data Engineering, vol. 12, pp. 509 - 516. (2000)

[6] Nakatsu, R. T.: Reasoning with Diagrams: Decision-Making and Problem-Solving with Diagrams, John Wiley & Sons. (2009)

[7] Ben-Gal, I.: Bayesian Networks, In: Encyclopedia of Statistics in Quality and Reliability, John Wiley & Sons Ltd. (2008)

[8] Darwiche, A.: Bayesian Networks, Foundations of Artificial Intelligence, vol. 3, pp. 467 - 509. (2008)

[9] Chockalingam, S., Pieters, W., Teixeira, A., Khakzad, N., van Gelder, P.: Combining Bayesian Networks and Fishbone Diagrams to Distinguish between Intentional Attacks and Accidental Technical Failures, In: Graphical Models for Security, pp. 31 - 50, Springer. (2019)

[10] Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P.: Bayesian Network Models in Cyber Security: A Systematic Review, In: Nordic Conference on Secure IT Systems, pp. 105 - 122, Springer. (2017)

[11] Zhang, G., Thai, V. V.: Expert Elicitation and Bayesian Network Modeling for Shipping Accidents: A Literature Review, Safety Science, vol. 87, pp. 53-62. (2016)

[12] Maaskant, P. P., Druzdzel, M. J.: An Independence of Causal Interactions Model for Opposing Influences, In: 4th European Workshop on Probabilistic Graphical Models, pp. 185 - 192. (2008)

[13] van der Gaag, L. C., Renooij, S., Witteman, C., Aleman, B. M., Taal, B. G.: Probabilities for a Probabilistic Network: A Case Study in Oesophageal Cancer, Artificial Intelligence in Medicine, vol. 25, pp. 123 - 148. (2002)

[14] van der Gaag, L. C., Renooij, S., Witteman, C. L., Aleman, B. M., Taal, B. G.: How to Elicit Many Probabilities, In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc., pp. 647 - 654. (1999)

[15] Doggett, A. M.: Root Cause Analysis: A Framework for Tool Selection, The Quality Management Journal, vol. 12, pp. 34 - 45. (2005)

[16] Ilie, G., Ciocoiu, C. N.: Application of Fishbone Diagram to Determine the Risk of an Event with Multiple Causes, Management Research and Practice, vol. 2, pp. 1 - 20. (2010)

[17] Ishikawa, K.: Guide to Quality Control. (1982)

[18] Desai, M. S., Johnson, R. A.: Using a Fishbone Diagram to Develop Change Management Strategies to Achieve First-year Student Persistence, SAM Advanced Management Journal, vol. 78, pp. 51 - 63.(2013)

[19] White, A. A., Wright, S. W., Blanco, R., Lemonds, B., Sisco, J., Bledsoe, S., Irwin, C., Isenhour, J., Pichert, J. W.: Cause-and-Effect Analysis of Risk Management Files to Assess Patient Care in the Emergency Department, Academic Emergency Medicine, vol. 11, pp. 1035-1041. (2004)

[20] Zhang, N. L., Poole, D.: Exploiting Causal Independence in Bayesian Network Inference, Journal of Artificial Intelligence Research, vol. 5, pp. 301 - 328. (1996)

[21] Fallet-Fidry, G., Weber, P., Simon, C., Iung, B., Duval, C.: Evidential Network-based Extension of Leaky Noisy-OR Structure for Supporting Risks Analyses, In: Fault Detection, Supervision and Safety of Technical Processes, pp. 672-677. (2012)

[22] Bolt, J. H., van der Gaag, L. C.: An Empirical Study of the Use of the Noisy-OR Model in a Real-life Bayesian Network, In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 11 - 20, Springer. (2010)

[23] Anand, V., Downs, S. M.: Probabilistic Asthma Case Finding: A Noisy OR Reformulation, In: AMIA Annual Symposium Proceedings, American Medical Informatics Association, pp. 6 - 10. (2008)

[24] Diez, F. J.: Parameter Adjustment in Bayes Networks. The Generalized Noisy OR–Gate, In: Uncertainty in Artificial Intelligence, pp. 99 - 105, Elsevier. (1993)

[25] Woudenberg, S. P., Van Der Gaag, L. C.: Using the Noisy-or Model can be Harmful... But it often is not, In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, pp. 122-133, Springer. (2011)

[26] Rosen, J. A., Smith, W. L.: Influence Net Modeling with Causal Strengths: An Evolutionary Approach, In: Proceedings of the Command and Control Research and Technology Symposium, Citeseer, pp. 25 - 28. (1996)

[27] Maaskant, P.: A Causal Model for Qualitative Reasoning, Delft University of Technology. (2006)

[28] Zagorecki, A.: Local Probability Distributions in Bayesian Networks: Knowledge Elicitation and Inference, University of Pittsburgh. (2010)

[29] Kraaijeveld, P.: Genierate: An Interactive Generator of Diagnostic Bayesian Network Models, In: Citeseer. (2005)

[30] Henrion, M.: Practical Issues in Constructing a Bayes' Belief Network, arXiv preprint arXiv:1304.2725. (2013)

[31] Hänninen, M., Mazaheri, A., Kujala, P., Montewka, J., Laaksonen, P., Salmiovirta, M., Klang, M.: Expert Elicitation of a Navigation Service Implementation Effects on Ship Groundings and Collisions in the Gulf of Finland, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, vol. 228, pp. 19 - 28. (2014)

[32] Wang, H., Druzdzel, M. J.: User Interface Tools for Navigation in Conditional Probability Tables and Elicitation of Probabilities in Bayesian Networks, In: Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers Inc., pp. 617 - 625. (2000)

[33] Renooij, S.: Probability Elicitation for Belief Networks: Issues to Consider, The Knowledge Engineering Review, vol. 16, pp. 255 - 269. (2001)

[34] Wang, H., Dash, D., Druzdzel, M. J.: A Method for Evaluating Elicitation Schemes for Probabilistic Models, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 32, pp. 38 - 43. (2002)

[35] Renooij, S., Witteman, C. L. M.: Talking Probabilities: Communicating Probabilistic Information with Words and Numbers, Utrecht University: Information and Computing Sciences. (1999)

[36] Witteman, C. L., Renooij, S., Koele, P.: Medicine in Words and Numbers: A Cross-Sectional Survey Comparing Probability Assessment Scales, BMC Medical Informatics and Decision Making, vol. 7, pp. 1 - 8. (2007)

[37] Witteman, C., Renooij, S.: Evaluation of a Verbal–Numerical Probability Scale, International Journal of Approximate Reasoning, vol. 33, pp. 117 - 131. (2003)

6

BAYESIAN NETWORK MODEL TO DISTINGUISH BETWEEN INTENTIONAL ATTACKS AND ACCIDENTAL TECHNICAL FAILURES: A CASE STUDY OF FLOODGATES*

6.1. INTRODUCTION

Water management is one of the critical infrastructures in countries like the Netherlands [1]. The proper functioning of water management infrastructures is vital for economic growth and societal wellbeing. The unexpected closure of floodgates could lead to severe economic damage, for instance, by delaying cargo ships. Over the years, water management infrastructures have become dependent on Industrial Control Systems (ICSs) to ensure efficient operations of such infrastructures [2].

ICSs were originally designed for isolated environments [3]. Such systems were mainly susceptible to technical failures. The blackout in the Canadian province of Ontario and the North-eastern and Mid-western United States is a typical example of a technical failure in which the absence of alarm due to a software bug in the alarm system left operators unaware of the need to redistribute power [4]. However, modern ICSs no longer operate in isolation, but use other networks to facilitate and improve business processes

*This chapter is submitted to a Journal as Chockalingam, S., Pieters, W., Teixeira, A., and van Gelder, P.: *"Bayesian Network Model to Distinguish between Intentional Attacks and Accidental Technical Failures: A Case Study of Floodgates"*

[5]. This increased connectivity makes ICSs more vulnerable to cyber-attacks apart from technical failures. A cyber-attack on a German steel mill is a typical example in which adversaries made use of corporate network to enter the ICS network [6]. As an initial step, the adversaries used both the targeted email and social engineering techniques to acquire credentials for the corporate network. Once they acquired credentials for the corporate network, they worked their way into the plant's control system network and caused damage to the blast furnace.

It is essential to distinguish between attacks and technical failures that would lead to abnormal behavior in the components of ICSs and take suitable measures. In most cases, the initiation of response strategy presumably aimed at technical failures would be ineffective in the event of a targeted attack and may lead to further complications. For instance, replacing a water level sensor that is sending incorrect measurement data with a new water level sensor would be a suitable response strategy to technical failure of a water level sensor. However, this may not be an appropriate response strategy to an attack on the water level sensor as it would not block the corresponding attack vector. Furthermore, the initiation of inappropriate response strategies would delay the recovery of the system from adversaries and might lead to harmful consequences. Noticeably, there is a lack of decision support to distinguish between attacks and technical failures.

Bayesian Networks (BNs) have the capacity to tackle this challenge especially based on their real-world applications in medical diagnosis and fault diagnosis [7]. BNs belong to the family of probabilistic graphical models, consisting of a qualitative and a quantitative part [8]. The qualitative part is a directed acyclic graph of nodes and edges. Each node represents a random variable, while the edges between the nodes represent the conditional dependencies among the random variables. The quantitative part takes the form of a priori marginal and conditional probabilities so as to quantify the dependencies between connected nodes.

In order to address the above-mentioned research gap, we developed the attack-failure distinguisher framework in our previous work to help construct BN models for distinguishing attacks and technical failures [9, 10]. Furthermore, we extended and combined fishbone diagrams within our framework for knowledge elicitation to construct the qualitative part of such BN models. Finally, we integrated DeMorgan models and probability scales with numerical and verbal anchors within our framework for knowledge elicitation to construct the quantitative part of such BN models. The present study aims to construct a BN model based on the developed framework to distinguish between attacks and technical failures for an observable problem in floodgates, providing a full case study of the framework as well as addressing the problem of floodgate operators. This study addresses the research question: "How could we develop Bayesian Network (BN) models for distinguishing attacks and technical failures in Floodgates?". The research objectives are:

- **RO 1.** To develop a BN model for distinguishing attacks and technical failures in floodgates involving domain experts using the attack-failure distinguisher framework.

- **RO 2.** To demonstrate the developed BN model for distinguishing attacks and technical failures in floodgates.

Expert knowledge is one of the predominant data sources utilised to construct direct acyclic graphs (DAGs) and populate conditional probability tables (CPTs) especially in domains where there is a limited availability of data like cyber security [11]. Expert knowledge is the data source which we used to construct the DAGs and populate CPTs in our work due to the unavailability of other data sources. Specifically, we utilised experts who associate themselves with safety and/or security community as it is appropriate for our application which deals with distinguishing attacks and technical failures. In our context, we associate the security community as dealing with attacks. On the other hand, we associate the safety community as dealing with technical failures.

The remainder of this study is structured as follows. In Section 6.2, we illustrate the different layers and the components of an ICS. In Section 6.3, we describe our existing framework that would help to construct BN models for distinguishing attacks and technical failures in addition to the systematic methods for knowledge elicitation to construct the BN models. Section 6.4 demonstrates the constructed BN model. Section 6.5 presents the conclusions and future work directions.

6.2. ICS ARCHITECTURE

Domain knowledge on ICSs is the starting point for the application of our proposed approach. We illustrated the three different layers and major components in each layer of an ICS in Section 4.4.1.

6.3. FRAMEWORK FOR DISTINGUISHING ATTACKS AND TECHNICAL FAILURES

This section describes the attack-failure distinguisher framework proposed in our previous work to construct BN models for distinguishing attacks and technical failures [9].

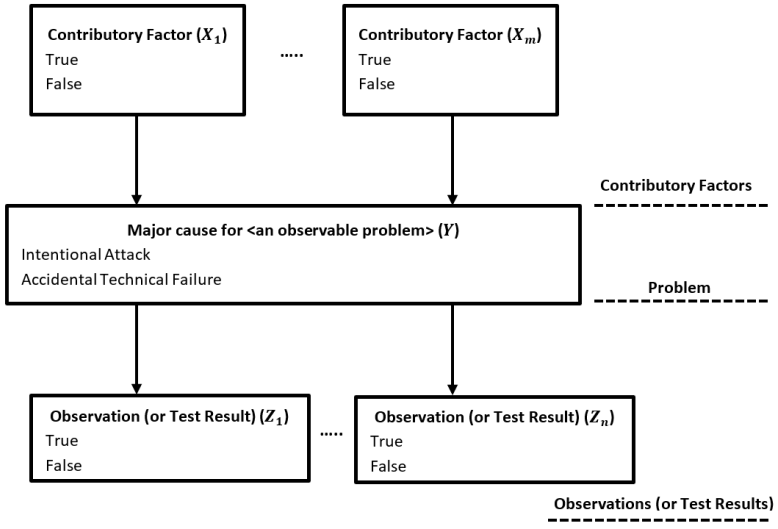


Figure 6.1: Framework for Distinguishing Attacks and Technical Failures

The framework consists of three layers as shown in Figure 6.1. The middle layer consists of a problem variable which is the major cause for an abnormal behaviour in a component of the ICS (observable problem). The states of the problem variable are the major causes of the observable problem (intentional attack and accidental technical failure). The upper layer consists of factors contributing to the major causes of the problem. The lower layer consists of observations (or test results) which is defined as any information useful for determining the major cause of the problem based on the outcome of tests conducted once the problem is observed by a floodgate operator.

The BN models would be incomplete without the quantitative part (CPTs for each variable). However, probability elicitation is a challenging task in building BNs, especially when it relies heavily on expert knowledge [12]. The extensive workload for experts in probability elicitation could affect the reliability of elicited probabilities. Therefore, the framework which we proposed in our previous work also includes DeMorgan models that reduces the number of conditional probabilities to elicit from domain experts in constructing the quantitative part of BN models, especially this technique reduces the number of parameters that need to be elicited from exponential to linear in the number of parents to define a full CPT for the child variable [9, 10]. We adopted DeMorgan models because it is the most suitable technique for our purpose [10]. Furthermore, we integrated probability scales with numerical and verbal anchors with DeMorgan models to facilitate individual probability entry by providing visual aids to help experts answer in terms of probabilities [10].

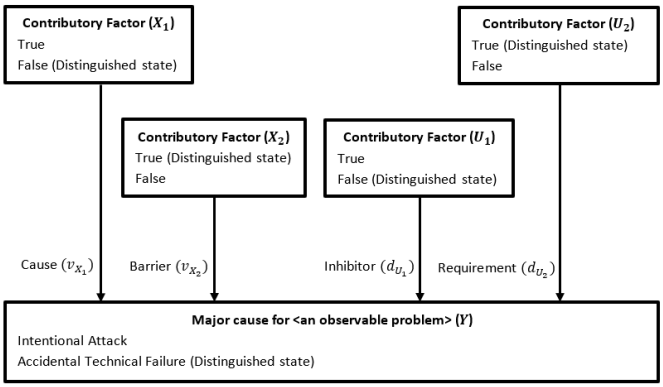


Figure 6.2: DeMorgan Model: Causal Interaction Types

The DeMorgan model is applicable when there are several parents and a common child. The DeMorgan model inherently assumes binary variables. In our application, the DeMorgan model could be used to elicit conditional probabilities for the problem variable as they have several contributory factors (parents). On the other hand, the CPTs of the contributory factors and observations (or test results) could be elicited directly from experts as they are straightforward when they do not have several parents. The DeMorgan model assumes that one of the two states of each variable is always the distinguished state as shown in Figure 6.2. Usually such state of the child variable depends on the modelled domain [13]. This is a typical state of the corresponding child variable [14]. In our application, the distinguished state of the problem variable ("Major cause for <an

observable problem>") is chosen as "accidental technical failure" as this is the a priori expected major cause, based on the higher frequency of technical failures compared to the attacks [9, 10]. The distinguished state of a parent variable is relative to the type of causal interaction with the child variable [15]. The same parent variable can have different distinguished states in different interactions that it participates in with the different child variables.

There are four different types of causal interactions between an individual parent (X) and a child (Y) in the DeMorgan model: (i) cause, (ii) barrier, (iii) inhibitor, and (iv) requirement. This is detailed in Section 5.4.1.

The DeMorgan model is an extension and a combination of the noisy-OR and noisy-AND model which supports modelling the above-mentioned types of causal interactions [15]. The property of accountability in the noisy-OR model is applicable to the DeMorgan model with a slight modification as it also exploits causal independence: In case all the modelled parents of the child are in their distinguished state, the property of accountability requires that the child be presumed their distinguished state. However, in many cases, this is not a realistic assumption as it is difficult to capture all the possible parents of the child [16]. Specifically, this is not realistic in our application as it is difficult to capture all the possible contributory factors of an observable problem due to "intentional attack". In the DeMorgan model, the leak parameter (ν_{XL}) deals with the possible parents of the child that are not previously known and explicitly modelled.

In general, the size of the CPT of a binary variable with n binary parents is $2^{(n+1)}$. However, only n+1 parameters are sufficient to completely define CPT using the DeMorgan model as it exploits causal independence. In the example shown in Figure 6.2, only five parameters are sufficient to completely define the CPT of child variable (Y) using the DeMorgan model instead of 64 entries. We could find the values for required parameters from the experts to completely define CPT using the DeMorgan model based on appropriate question for each type of causal interaction shown in Table 6.1.

Type of Causal Interaction	Elicitation Question
Leak	*"What is the probability that the child is in their non-distinguished state given that the parents are in their distinguished states?"*
Cause, Barrier Note: There is a difference between the non-distinguished state of a cause and barrier.	*"What is the probability that the child is in their non-distinguished state given that all the parents are in their distinguished states, except X_i and no other unmodelled causal factors are present?".*
Inhibitor, Requirement Note: There is a difference between the non-distinguished state of an inhibitor and requirement.	*"What is the probability that the child is in their distinguished state given that the parents are in their distinguished states, except U_i and no other unmodelled causal factors are present?".*

Table 6.1: Causal Interactions and their Corresponding Elicitation Questions in the DeMorgan Model

Once we determine the required parameters based on appropriate elicitation questions, we can completely define the CPT of the child variable using (1):

$$P(y|X,U) = \left(1 - (1 - v_{X_L}) \prod_{X_i \in +X} (1 - v_{X_i})\right) \prod_{U_i \in +U} (1 - d_{U_i}) \quad (1)$$

In the equation (1), Y represents the effect variable which has values y for the effect being in the non-distinguished state ("Intentional attack") and y' for the effect being in the distinguished state ("Accidental technical failure"). X denotes the set of parents which interact with the effect variable as promoting influences, U denotes the set of parents which interact with the effect variable as inhibiting influences, $+X$ denotes the subset of X that contains all parents that are in their non-distinguished states, $+U$ denotes the subset of U that contains all parents that are in their non-distinguished states. v_{XL} denotes the leak parameter which expresses the probability of y ("Intentional attack") given all parents are in their distinguished states, (v_{Xi}) denotes the probability of y ("Intentional attack") given that the parent X_i is not in its distinguished state and all other parents are in their distinguished states, d_{Ui} denotes the probability of y' ("Accidental technical failure") given that the parent U_i is not in its distinguished state and all other parents are in their distinguished states.

6.4. APPLYING BNS FOR DISTINGUISHING ATTACKS AND TECHNICAL FAILURES

This section describes how we constructed the BN model for distinguishing attacks and technical failures in floodgates.

We considered the observable problem for this application as "Sensor sends incorrect water level measurements" because it could lead to serious consequences in the case of floodgate. In case the floodgate closes when it should not, based on the incorrect water level measurements sent by the sensor, it would lead to severe economic damage, for instance, by delaying cargo ships. On the other hand, in case the floodgate opens when it should not, due to incorrect water level measurements sent by the sensor, it would lead to flooding.

6.4.1. CONSTRUCTION OF QUALITATIVE BN MODEL FOR DISTINGUISHING ATTACKS AND TECHNICAL FAILURES IN FLOODGATES

We have utilised a multimethodology approach for data collection. Multimethodology refers to using more than one method of data collection in a research study [17], providing more comprehensive data. In our study, we utilised a focus group workshop and a questionnaire to gather data for constructing the qualitative BN model. Firstly, we conducted a focus group workshop with five participants who have experience working with safety and/or security of water management infrastructures operated by ICS. The major objective of this focus group is to discuss and identify contributory factors and observations (or test results) for the problem which we considered. Each participant was provided with a set of questions as shown in Appendix C. Most of these questions are

open-ended that ask for factors that would contribute to the major cause of the considered problem (attack/technical failure) and tests that would provide additional information to distinguish between the major cause of the considered problem (attack/technical failure) after the problem is observed by the floodgate operator. For instance, we considered the problem *"the sensor sends incorrect water level measurements"* and asked the participants: *"Which contributory factors would increase the likelihood of the problem due to (accidental) sensor failure?"*. The moderator explained each question to the participants and facilitated the discussion among the participants to identify a set of contributory factors and observations (or test results) for the observable problem which we considered.

After the focus group workshop, we employed a questionnaire to gather data for constructing the qualitative BN model. We employed snowball sampling to recruit other participants for this study through initial participants. This sampling technique is useful as it helps to find experts in ICS safety and/or security quickly. The participants were provided with the same set of questions which we provided to focus group participants as shown in Appendix C to elicit contributory factors and observations (or test results) for the considered problem. We received 10 responses in total for the questionnaire. However, we excluded one response as the participant did not have any experience working with ICS. Importantly, seven out of nine respondents have five or more years working experience with ICS which helps to ensure reliability of data. In addition, we had a good mix of participants from safety and/or security community which is important for our application. Specifically, two out of nine respondents associate themselves with both safety and security, two out of nine respondents associate themselves with safety and five out of nine respondents associate themselves with security.

We combined the data gathered from the focus group and questionnaire for coding. We utilised thematic coding by grouping contributory factors which are similar under a category. For instance, there were nine responses such as "easy access to sensor", "attacker has physical access to the sensor", "free access to sensor" which we categorised into "easy physical access to sensor". On the other hand, we grouped and removed contributory factors which are not contributory factors based on our definition. For instance, "Man-in-the-Middle attack using the wired connection" is not a specific contributory factor but rather a type of attack that an attacker might employ. Once we categorised the contributory factors, there were 14 categories (parent nodes) in total. However, this would result in the CPT size of the problem variable as 16384, which makes it unmanageable. Therefore, we utilised parent node divorcing, which allows parent nodes to be grouped hierarchically to avoid excessive inbound links to the child node. By utilising parent node divorcing, we reduced the number of parent nodes to eight which in turn reduced the CPT size of the problem variable to 256. For instance, we grouped hierarchically three different parent nodes (location of sensor susceptible to severe weather, location of sensor susceptible to biological fouling, location of sensor susceptible to physical contact of marine vessel) into a single parent node (location of sensor susceptible to external factor) as shown in Figure 6.3, because they are of the same theme and no original interactions are lost in the process. Once the qualitative BN model shown in Figure 6.3 is constructed, we validated it through a focus group workshop with five experts who have experience working with safety and/or security of ICS in the water management sector in the Netherlands. We asked specifically whether anything is missing or not appropriate in

the qualitative BN model. However, the experts did not find any need to add or update anything in the constructed qualitative BN model.

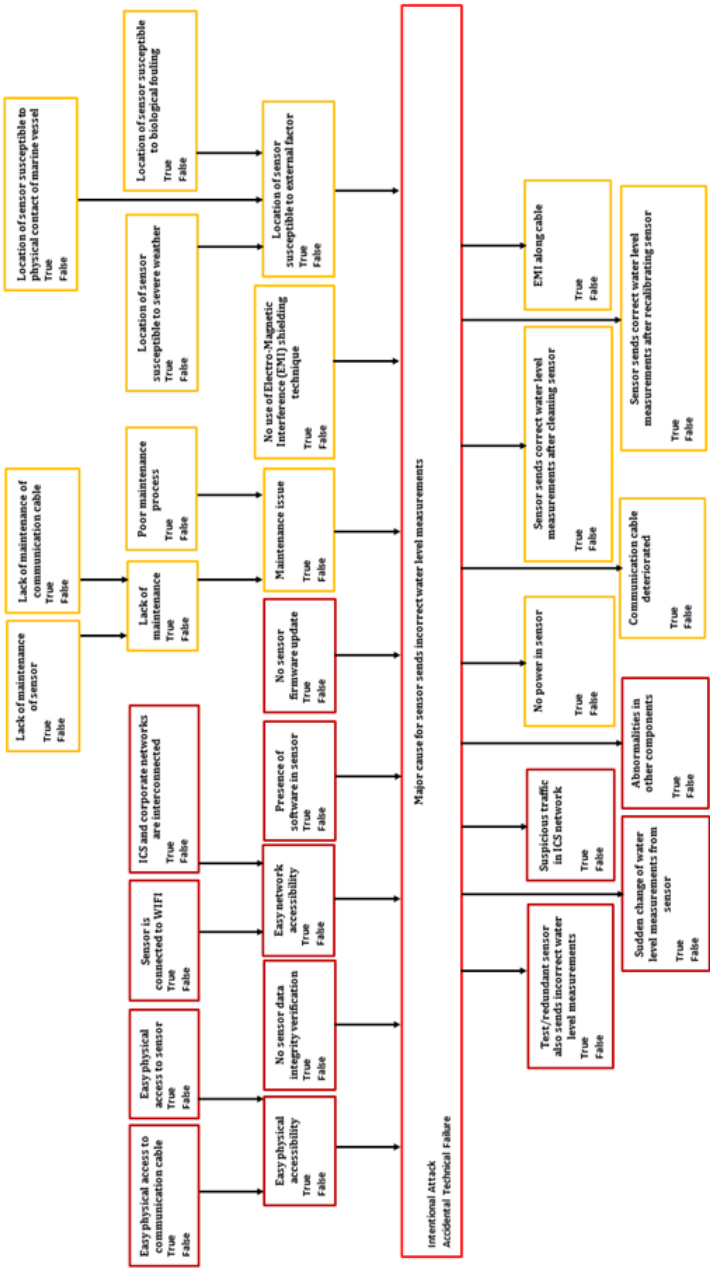


Figure 6.3: Constructed Qualitative BN Model

6.4.2. CONSTRUCTION OF QUANTITATIVE BN MODEL FOR DISTINGUISHING ATTACKS AND TECHNICAL FAILURES IN FLOODGATES

A multimethodology approach is used for quantitative data collection like we did for the construction of the qualitative BN model. In order to gather data for populating the BN model with probabilities, we utilised a focus group workshop and a questionnaire. Firstly, we conducted a focus group workshop with five participants who have experience working with safety and/or security of ICS in the water management sector in the Netherlands. The major objective of this focus group is to elicit probabilities corresponding to each variable in our qualitative BN model that could help to determine the major cause (intentional attack or accidental technical failure) of the problem (sensor sends incorrect water level measurements) when observed.

Appendix D shows a set of questions which we provided to each participant at the start of the focus group workshop. We asked each participant to answer the question using a probability scale with numerical and verbal anchors to elicit prior probabilities corresponding to the contributory factors and conditional probabilities corresponding to the problem and observations (or test results). For instance, we elicited the prior probability of the variable "Easy Physical Access to Sensor" and a conditional probability of the variable "Major cause for sensor sends incorrect water level measurements" as shown in Figure 6.4. The participants were asked to answer the questions individually to avoid bias in their responses. Furthermore, the moderator provided clarifications individually in case there are any questions from the participants. Once the participants answered the questions individually, the moderator facilitated a discussion on the reasoning behind the varied probabilities which they provided for some variables. However, the purpose of this discussion is not to make them reach a consensus as it could make the responses biased.

In addition to the focus group workshop, we utilised a questionnaire to gather data for populating the BN model with probabilities. We used snowball sampling to recruit other participants for this study through initial participants in the focus group workshop as the target group is limited and rare to find. This sampling technique makes it easier to find experts in safety and/or security of ICS in the water management sector in the Netherlands quickly. We provided a set of questions to the participants mainly to elicit probabilities corresponding to each variable in the constructed BN model as shown in Appendix D. For instance, we asked for the prior probability of the variable "Easy Physical Access to Sensor" and a conditional probability of the variable "Major cause for sensor sends incorrect water level measurements" as shown in Figure 6.5. The difference compared to the focus group workshop questions is that the probability scale with numerical and verbal anchors is not directly used as it is not practicable in the online questionnaire. However, we utilised the verbal and corresponding numerical anchors from the probability scale as answer choices for each question in the online questionnaire in addition to "others" option which could help participants to provide fine-grained probabilities as shown in Figure 6.5. We received five responses in total. Overall, seven out of 10 participants have more than five years work experience with safety and/or security of ICS in the water management sector in the Netherlands.

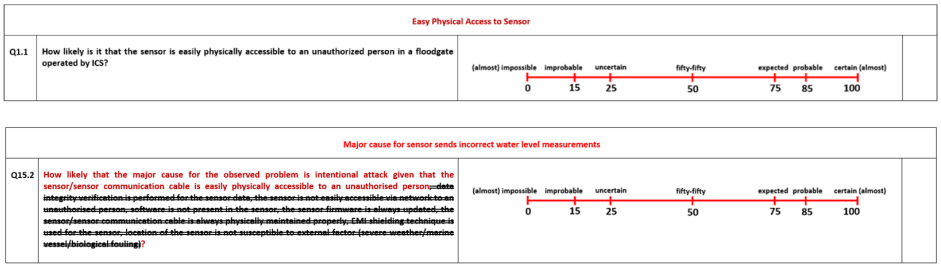


Figure 6.4: Focus Group Workshop - Example Questions

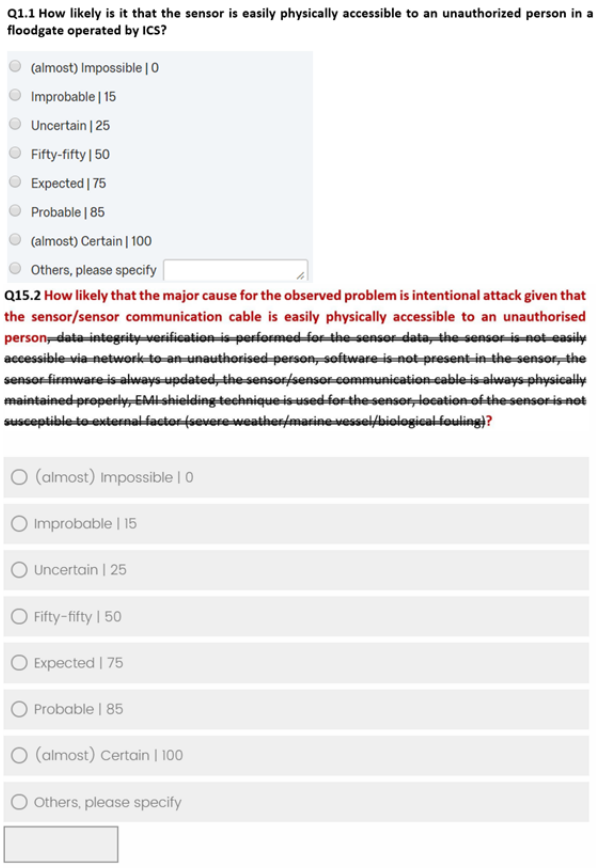


Figure 6.5: Questionnaire - Example Questions

Once we collected the responses from the participants in both the focus group workshop and questionnaire, we tabulated them together. Furthermore, we noticed that there were some missing data due to no or invalid response from some respondents. For instance, we considered responses like “others” without mentioning any specific likelihood value as an invalid response. Furthermore, it is also not possible to clarify with

the respondent as responses are anonymous. Ignoring or discarding missing data is one of the most common approaches used to deal with the missing data [18, 19]. Listwise deletion and pairwise deletion are the two different methods which could help to ignore or discard the missing data [18]. Pairwise deletion is appropriate for our application as it ignores or discards only the missing data and considers the other data provided by these experts. This is easy to implement. Therefore, we utilised pairwise deletion to ignore or discard the missing data in our application. Listwise deletion is not appropriate for our application as it leads to loss of data by completely ignoring or discarding data from four out of 10 experts since they have no or invalid response to a question.

Once the missing data is ignored or discarded, the probabilities $P_i(X)$ elicited from the experts need to be combined. One of the most widely used method to combine the probabilities elicited from the experts is linear pooling [20, 21]. Using the linear pooling method, the combined probabilities can be computed using (2):

$$P(X) = \sum_{i=1}^{n} w_i P_i(X) \tag{2}$$

Where w_i are positive weights given to each of the n experts with complete probabilities for the corresponding X and $\sum_{i=1}^{n} w_i = 1$.

C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	Y	
								Attack	Failure
True	True	True	True	True	True	True	True	0.02	0.98
True	True	True	True	True	True	True	False	0.09	0.91
True	True	True	True	True	True	False	True	0.06	0.94
True	True	True	True	True	True	False	False	0.24	0.76
True	True	True	True	True	False	True	True	0.09	0.91
True	True	True	True	True	False	True	False	0.38	0.62
True	True	True	True	True	False	False	True	0.24	0.76
True	True	True	True	True	False	False	False	0.97	0.03
True	True	True	True	False	True	True	True	0.02	0.98
True	True	True	True	False	True	True	False	0.09	0.91

Table 6.2: CPT Excerpt - Problem Variable

(In this table, C_1: Easy physical accessibility, C_2: No sensor data integrity verification, C_3: Easy network accessibility, C_4: Presence of software in sensor, C_5: No sensor firmware update, C_6: Maintenance issue, C_7: No use of EMI shielding technique, C_8: Location of sensor susceptible to external factor and Y: Major cause for sensor sends incorrect water level measurements.)

There are two different types of linear pooling method: (i) prior linear pooling, and (ii) posterior linear pooling [20]. Prior linear pooling combines elicited probabilities from experts corresponding to each variable in the BN model, which could then be used

to compute posterior probabilities of target variables by providing evidences to some variables. On the other hand, in posterior linear pooling, elicited probabilities from n experts are used to construct n distinct BNs. Once we construct the n distinct BNs, we run these BNs by providing same evidences to the same set of variables in these BNs and compute different posterior probabilities in each of these BNs. Finally, the posterior probabilities generated in n distinct BNs are combined. However, this is not appropriate for our application as it is not practicable for performing diagnostics in a timely way. Furthermore, this is not suitable for our application as we ignored or discarded missing data which could make it not possible to construct BNs with no probabilities for some variables.

In our application, we utilised prior linear pooling as it is appropriate based on its advantages [20]. Each of the 10 experts is given equal weighting as they all have experience working with safety and/or security of ICS in the water management sector in the Netherlands. Furthermore, we consider each respondent's experience to be equal in value. So, we combined the probabilities from n experts using (2).

The probabilities corresponding to contributory factors and observations (or test results) are now complete. However, we utilised DeMorgan model to reduce the number of CPT entries that needs to be elicited from experts to nine. Therefore, we computed the remaining CPT entries corresponding to the problem variable using (1). An excerpt of CPT entries corresponding to the problem variable is shown in Table 6.2. The complete BN model with both the qualitative and quantitative component is shown Figure 6.6.

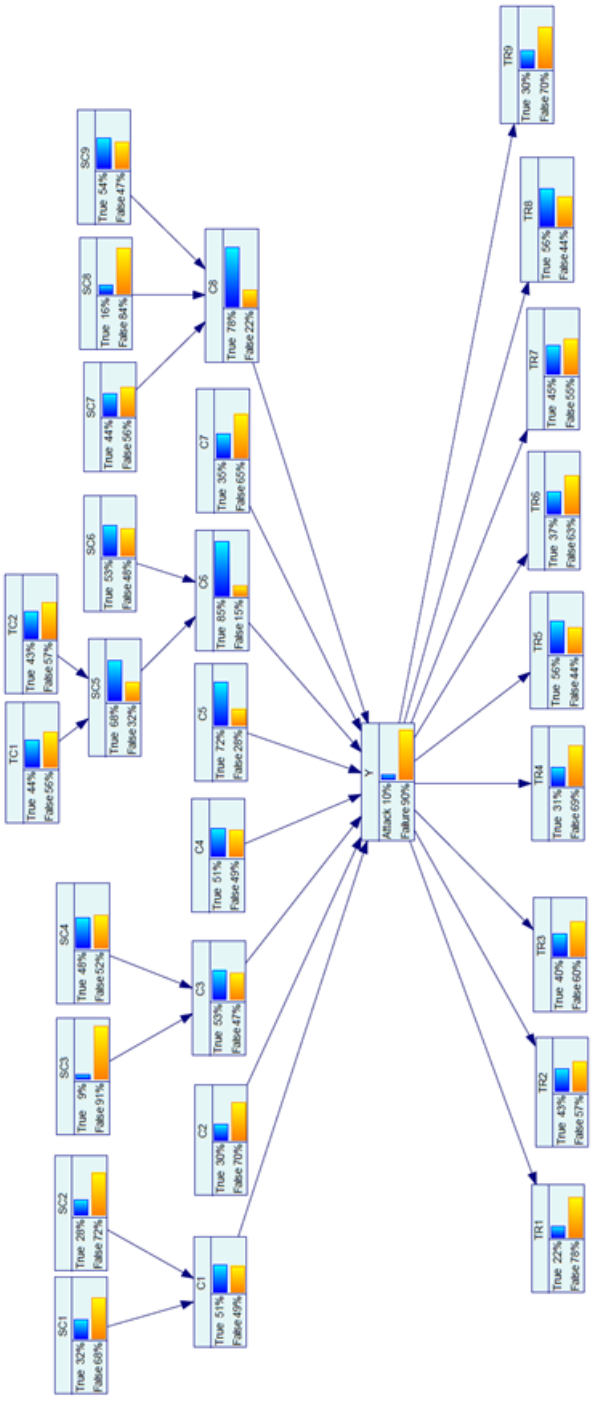


Figure 6.6: Constructed BN Model – No Evidence Provided

(In this figure, SC_1: Easy physical access to sensor, SC_2: Easy physical access to communication cable, C_1: Easy physical accessibility, C_2: No sensor data integrity verification, SC_3: Sensor is connected to WIFI, SC_4: ICS and corporate networks are interconnected, C_3: Easy network accessibility, C_4: Presence of software in sensor, C_5: No sensor firmware update, TC_1: Lack of maintenance of sensor, TC_2: Lack of maintenance of communication cable, SC_5: Lack of maintenance, SC_6: Poor maintenance process, C_6: Maintenance issue, C_7: No use of EMI shielding technique, SC_7: Location of sensor susceptible to severe weather, SC_8: Location of sensor susceptible to physical contact of marine vessel, SC_9: Location of sensor susceptible to biological fouling, C_8: Location of sensor susceptible to external factor, Y: Major cause for sensor sends incorrect water level measurements, TR_1: Test/redundant sensor also sends incorrect water level measurements, TR_2: Sudden change of water level measurements from sensor, TR_3: Suspicious traffic in ICS network, TR_4: Abnormalities in other components, TR_5: No power in sensor, TR_6: Communication cable deteriorated, TR_7: Sensor sends correct water level measurements after cleaning sensor, TR_8: Sensor sends correct water level measurements after recalibrating sensor, TR_9: EMI along cable.)

6.4.3. DEMONSTRATION OF THE CONSTRUCTED BN MODEL

In this section, we demonstrate the suitability or utility of the constructed BN model based on two different illustrative scenarios. It is not possible to utilise the real floodgate for demonstrating the suitability or utility of the constructed BN model by putting it into practice due to availability and criticality issues. Therefore, we relied on two different illustrative scenarios for this purpose. These two different illustrative scenarios help to show how and when the constructed BN model would be useful in practice. Firstly, we assume that the floodgate operator observed that a sensor sends incorrect water level measurements by noticing the mismatch between the measurements from physical water level scale and water level sensor. In order to choose the appropriate response strategy, the floodgate operator needs to determine the major cause of this problem (i.e., whether this problem is caused by an attack or technical failure), which is the aim of the constructed BN model.

Once the floodgate operator noticed the incorrect sensor measurements problem, they need to provide the evidence that is available for variables in the upper layer (contributory factors) and lower layer (test results). This could help the constructed BN model compute posterior probabilities of both the states in the problem variable (attack and technical failure) based on the provided evidences.

In the first illustrative scenario, the floodgate operator set evidence for variables based on the available information as shown in Table 6.3. Based on such evidence, the posterior probability is computed by the constructed BN model for other variables without any evidence. The BN model in Figure 6.7 shows that the incorrect water level measurement problem is most likely due to technical failure based on the provided evidences. This information would help to select the appropriate response strategy (i.e., to repair or replace the water level sensor).

Name of the Variable	Evidences (First Illustrative Scenario)	Evidences (Second Illustrative Scenario)
Easy physical access to sensor ($SC1$)	False	True
Easy physical access to communication cable ($SC2$)	False	True
No sensor data integrity verification ($C2$)	False	True
Sensor is connected to WIFI (SC3)	False	False
ICS and corporate networks are interconnected ($SC4$)	False	False
Presence of software in sensor ($C4$)	False	True
No sensor firmware update ($C5$)	False	True
Lack of maintenance of sensor ($TC1$)	True	False
Lack of maintenance of communication cable ($TC2$)	True	False
Poor maintenance process ($SC6$)	True	False
No use of EMI shielding technique ($C7$)	False	True
Location of sensor susceptible to severe weather ($SC7$)	True	False
Location of sensor susceptible to physical contact of marine vessel ($SC8$)	True	False
Location of sensor susceptible to biological fouling ($SC9$)	True	False
Test/Redundant sensor also sends incorrect water level measurements ($TR1$)	False	True
Sudden change of water level measurements from sensor ($TR2$)	False	False
Suspicious traffic in ICS network ($TR3$)	True	True
Abnormalities in other components ($TR4$)	True	True
Communication cable deteriorated ($TR6$)	True	True
Sensor sends correct water level measurements after cleaning sensor ($TR7$)	True	False

Table 6.3: Evidences Corresponding to both the Illustrative Scenarios

In the second illustrative scenario, the floodgate operator sets different evidence for variables in the constructed BN model based on the available information as shown in Table 6.3. Based on the provided evidences, the posterior probability is computed for other variables without any evidence in the constructed BN model. Figure 6.8 shows that the incorrect water level measurement problem is most likely due to attack based on the evidences provided by the floodgate operator. This information would help to choose the suitable response strategy (i.e., to block the corresponding attack vector).

The difference between the two scenarios can be explained as follows. In the first illustrative scenario, the sensor/sensor communication cable is not easily accessible to an unauthorised person, whereas there is a lack of maintenance of the sensor/sensor communication cable and the location of the sensor is susceptible to external factors such as biological fouling. In addition, the sensor communication cable is deteriorated, and the sensor sends correct water level measurements after cleaning the sensor. Typically, the above-mentioned factors increase the likelihood of the problem due to accidental technical failure, which is reflected in terms of the posterior probability of Y in Figure 6.7. In contrast, in the second illustrative scenario, the sensor/sensor communication cable is properly maintained, and the location of the sensor is not susceptible to external factors such as biological fouling, whereas the sensor/sensor communication cable is easily physically accessible to an unauthorised person. In addition, the test/redundant

sensor also sends incorrect water level measurements. Typically, the above-mentioned factors increase the likelihood of the problem due to intentional attack, which is reflected in terms of the posterior probability of Y in Figure 6.8.

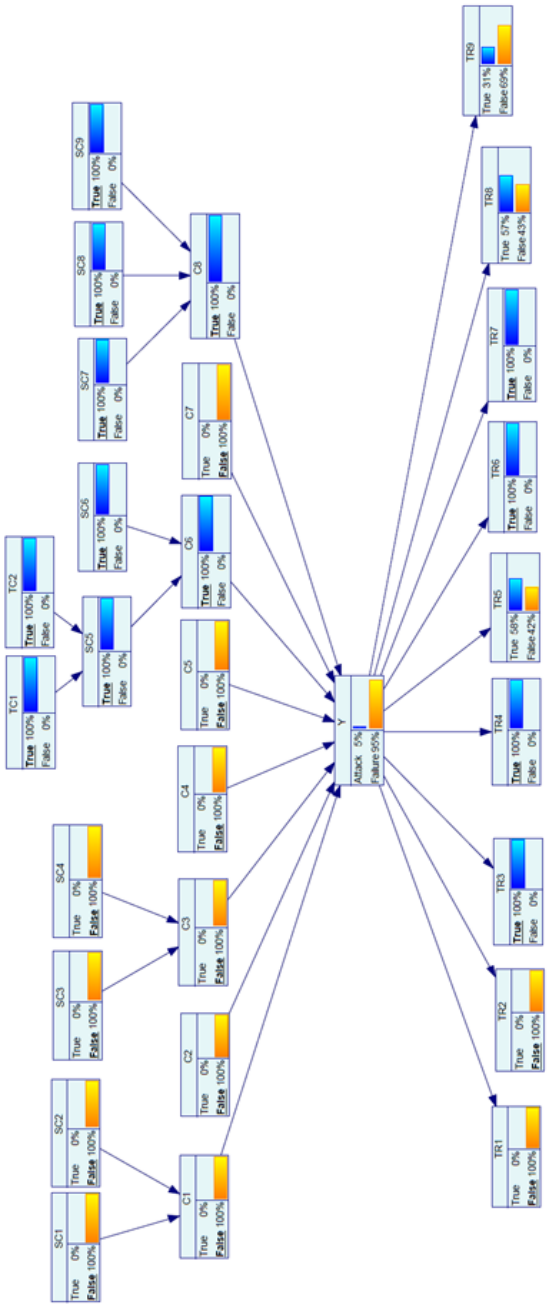


Figure 6.7: Constructed BN Model – First Illustrative Scenario

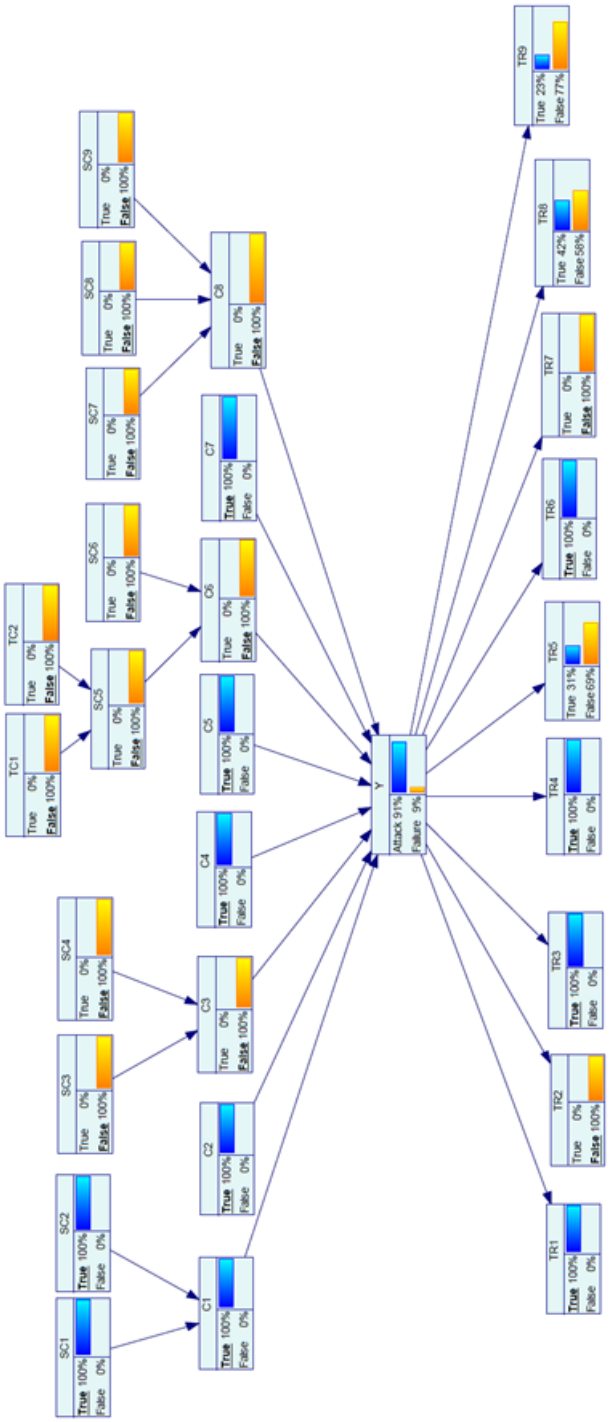


Figure 6.8: Constructed BN Model – Second Illustrative Scenario

6.5. CONCLUSIONS AND FUTURE WORK

Harmful consequences of a problem could be minimised by choosing the appropriate response strategy in a timely manner. However, this is not possible without determining the major cause of a problem. In our previous work, we developed the attack-failure distinguisher framework which could help to construct BN models that determine whether the problem is caused by an attack or technical failure. This framework also includes the knowledge elicitation methods such as the DeMorgan model, and probability scales with numerical and verbal anchors to effectively elicit expert knowledge to construct such BN models. This work mainly focused on providing a full case study of the framework on how to construct the BN model for a problem and demonstrate when and how this could be used in practice.

In this work, we developed a BN model for the problem of incorrect sensor measurements in floodgates in the Netherlands using the attack-failure distinguisher framework. Due to the lack of data, we relied on expert knowledge to construct the qualitative and quantitative part of the BN model for our problem. We elicited contributory factors and test results (or observations) through a focus group workshop and a questionnaire among respondents who have experience working with ICS. The data from both the focus group workshop and questionnaire were used to construct the qualitative BN model, which was also validated with five experts.

Once the qualitative BN model was constructed, we used the DeMorgan model to reduce the number of CPT entries that needs to be elicited for the problem variable to nine instead of 256. Firstly, we elicited probabilities corresponding to contributory factors, problem and test results (or observations) from experts who have experience working with safety and/or security of water management infrastructures operated by ICS in the Netherlands through a focus group workshop and questionnaire. During this elicitation, we employed probability scales with numerical and verbal anchors to facilitate individual probability entry by providing it as a visual aid. We computed the rest of the probabilities for the problem variable using the DeMorgan model. Finally, we demonstrated the suitability or utility of the constructed BN model using two different illustrative scenarios. The first illustrative scenario shows that the most likely cause for the considered problem is technical failure, whereas the second illustrative scenario shows that the most likely cause for the considered problem is attack based on the evidences provided.

The results of existing integrated safety and security risk assessment methods would help to choose suitable risk treatments during the design phase before an attack or technical failure occurs. On the other hand, our method involving the attack-failure distinguisher framework would help to choose appropriate response strategies during the operational phase when an attack or technical failure occurs. Furthermore, our method would help operators to think more proactively about reactive safety and security.

We provided a case study of attack-failure distinguisher framework by developing a BN model for the problem of incorrect sensor measurements in floodgates. In the future, this would help practitioners to develop BN models for different problems in different domains. Furthermore, we provided two different illustrative scenarios using the developed BN model to demonstrate the suitability and/or utility of such models.

Historical data on attacks and technical failures in the water management sector in

the Netherlands is unavailable for research due to sensitivity issues. Therefore, it would not be possible to develop models that could help to distinguish between attacks and technical failures for the problem of incorrect sensor measurements using a data-driven approach. However, in the future, the unavailability of historical data on attacks and technical failures would not deter modelling cyber security for ICS anymore as we utilised a knowledge-based approach to develop a model for distinguishing attacks and technical failures.

In addition, it was not possible to use real systems for evaluating the attack-failure distinguisher framework due to availability and criticality issues. However, we utilised real-users and realistic problems to evaluate the attack-failure distinguisher framework by developing a prototype and using the developed prototype for two different illustrative scenarios to relate the results to real use. Therefore, the developed BN model is usable in real settings in the future. However, this BN model can be further updated with appropriate contributory factors, test results and probabilities based on the performance measures in the confusion matrix, which includes four different combinations of diagnosed and actual classes. This is only possible when a dataset corresponding to the problem in the real setting is available for research.

In the future, it would be beneficial to put the constructed BN model into practice in a real floodgate in case it is available to showcase the value of the constructed BN model. Furthermore, we developed a root-cause analysis framework with the appropriate type of variables and relationships between them in our previous work, which would help to construct BN models to determine the attack-vector (in case of an attack) and failure mode (in case of a technical failure) [22]. However, the root-cause analysis framework needs to be applied and evaluated for a problem like incorrect sensor measurements in the future as it could complement the attack-failure distinguisher framework to determine the attack-vector (in case of an attack) and failure mode (in case of a technical failure). This could also help to choose the most effective response strategy between alternatives like repairing or replacing the sensor.

REFERENCES

[1] Castellon, N. Frinking, E.: Securing Critical Infrastructures in the Netherlands: Towards a National Testbed, The Hague Centre for Strategic Studies. (2015)

[2] Nogueira, H. I. S., Walraven, M.: Overview of Storm Surge Barriers, Rijskwaterstaat & Deltares, pp. 1 - 38. (2018)

[3] Effendi, A. Davis, R.: ICS and IT: Managing Cyber Security Across the Enterprise, In: SPE Middle East Intelligent Oil and Gas Conference and Exhibition, Society of Petroleum Engineers. (2015)

[4] Zhivich, M., Cunningham, R. K.: The Real Cost of Software Errors, IEEE Security & Privacy, vol. 7, no. 2, pp. 87 - 90. (2009)

[5] Knowles, W., et al.: A Survey of Cyber Security Management in Industrial Control Systems, International Journal of Critical Infrastructure Protection, vol. 9, pp. 52-80. (2015)

[6] RISI.: German Steel Mill Cyber Attack, Available: http://www.risidata.com/database/detail/german-steel-mill-cyber-attack (2014)

[7] Nakatsu, R. T.: Reasoning with Diagrams: Decision-Making and Problem-Solving with Diagrams, Wiley. (2009)

[8] Darwiche, A.: Bayesian Networks. Foundations of Artificial Intelligence, vol. 3, pp. 467 - 509. (2008)

[9] Chockalingam, S., Pieters, W., Teixeira, A., Khakzad, N., van Gelder, P.: Combining Bayesian Networks and Fishbone Diagrams to Distinguish between Intentional Attacks and Accidental Technical Failures, In: Graphical Models for Security, pp. 31 - 50, Springer. (2019)

[10] Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P.: Probability Elicitation for Bayesian Networks to Distinguish between Intentional Attacks and Accidental Technical Failures (Submitted to a Journal), pp. 1 - 24. (2020)

[11] Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P.: Bayesian Network Models in Cyber Security: A Systematic Review, In: Nordic Conference on Secure IT Systems, pp. 105 - 122, Springer. (2017)

[12] Zhang, G., Thai, V. V.: Expert Elicitation and Bayesian Network Modeling for Shipping Accidents: A Literature Review, Safety Science, vol. 87, pp. 53-62. (2016)

[13] Zagorecki, A.: Local Probability Distributions in Bayesian Networks: Knowledge Elicitation and Inference, University of Pittsburgh. (2010)

[14] Kraaijeveld, P.: Genierate: An Interactive Generator of Diagnostic Bayesian Network Models, Citeseer. (2005)

[15] Maaskant, P. P., Druzdzel, M. J.: An Independence of Causal Interactions Model for Opposing Influences, In: 4th European Workshop on Probabilistic Graphical Models, pp. 185 - 192. (2008)

[16] Fallet-Fidry, G., Weber, P., Simon, C., Iung, B., Duval, C.: Evidential Network-based Extension of Leaky Noisy-OR Structure for Supporting Risks Analyses, In: Fault Detection, Supervision and Safety of Technical Processes, pp. 672-677. (2012)

[17] Brewer, J., Hunter, A.: Multimethod Research: A Synthesis of Styles, Sage Publications Inc. (1989)

[18] Baraldi, A. N., Enders, C. K.: An Introduction to Modern Missing Data Analyses. Journal of School Psychology, vol. 48, no. 1, pp. 5 - 37. (2010)

[19] Twala, B.: An Empirical Comparison of Techniques for Handling Incomplete Data using Decision Trees, Applied Artificial Intelligence, vol. 23, no. 5, pp. 373 - 405. (2009)

[20] Farr, C., Ruggeri, F., Mengersen, K.: Prior and Posterior Linear Pooling for Combining Expert Opinions: Uses and Impact on Bayesian Networks — The Case of the Wayfinding Model, Entropy, vol. 20, no. 3, pp. 1 - 14. (2018)

[21] Ouchi, F.: A Literature Review on the Use of Expert Opinion in Probabilistic Risk Analysis, pp. 1 - 17. (2004)

[22] Chockalingam, S., Katta, V.: Developing a Bayesian Network Framework for Root Cause Analysis of Observable Problems in Cyber-Physical Systems. In: 2019 IEEE Conference on Information and Communication Technology, pp. 1 - 6, IEEE. (2019)

7

CONCLUDING REMARKS

In the Netherlands, water management infrastructures like floodgates are automated with Industrial Control Systems (ICS). The problems such as unexpected opening/closing of floodgates could be caused by (accidental) technical failures and (intentional) attacks. This thesis tackled a practical problem in the operational phase of water management infrastructures operated by Industrial Control Systems (ICS): when the operators notice such problems in infrastructures operated by control systems in practice, they predetermine that the problem is due to an (accidental) technical failure and initiate corresponding response strategies [1]. The wrong diagnosis could result in choosing ineffective response strategies. This has culminated into the following research question which this thesis set out to answer:

- **RQ. How to develop decision support to distinguish between intentional attacks and accidental technical failures for problems in water management infrastructures operated by Industrial Control Systems (ICS)?**

We tackled this research question using the Design Science Research (DSR) method as it is appropriate for the purpose of our study [2, 3].

This chapter summarises the main findings of this thesis. Furthermore, we discuss scientific and societal implications of this work. Finally, we present limitations of this work and point out future research directions.

7.1. SUMMARY OF THE FINDINGS

This thesis has first investigated state-of-the-art of integrated safety and security risk assessment methods. In particular, Chapter 2, "Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications" [4], addressed the following sub-question:

- **SQ 1.** What are the key characteristics of integrated safety and security risk assessment methods, and their applications?

In Chapter 2, we identified seven integrated safety and security risk assessment methods based on the review methodology we adopted: (i) Security-Aware Hazard Analysis and Risk Assessment (SAHARA) [5], (ii) Combined Harm Assessment of Safety and Security for Information Systems (CHASSIS) [6], (iii) Failure-Attack-CounTermeasure (FACT) Graph [7], (iv) Failure Mode, Vulnerabilities, and Effect Analysis (FMVEA) [8], (v) Unified Security and Safety Risk Assessment [9], (vi) Extended Component Fault Tree (CFT) [10], and (vii) Extended Fault Tree (EFT) [11]. The identified methods were analysed using five different criteria: (i) citations in scientific literature, (ii) steps involved, (iii) stage(s) of risk assessment process addressed, (iv) integration methodology, and (v) application(s) and application domain.

Based on the steps involved in each identified method, there are two types of integrated safety and security risk assessment methods: (i) sequential and (ii) non-sequential. Sequential methods including SAHARA, FACT Graph, Extended CFT, and EFT identify security risks that impact safety, with the intent to improve the completeness of safety risk assessment. However, these methods did not consider safety risks that impact security. Furthermore, the non-sequential methods and unified security and safety risk assessment method did not consider either safety risks that impact security or security risks that impact safety. Risk identification and risk analysis stages of the risk assessment process were given much attention compared to the risk evaluation stage in the identified methods. This could be because the interaction between safety and security risk assessments happens mainly during those phases of the risk assessment process.

There are four ways in which the integrated safety and security risk assessment methods are developed: (i) the conventional safety risk assessment method as the base and a variation of the safety risk assessment method for security risk assessment, (ii) the conventional security risk assessment method as the base and a variation of the security risk assessment method for safety risk assessment, (iii) a combination of a conventional safety risk assessment method, and a conventional security risk assessment method and (iv) others. This shows that the conventional safety risk assessment method can be adapted to perform security risk assessment and vice versa. The integrated safety and security risk assessment methods were applied in transportation, power and utilities, and chemical domain. Noticeably, there is a lack of methods that integrate safety and security in domains like water management.

Furthermore, the integrated safety and security risk assessment methods did not consider real-time system information. However, there is a need for integrated safety and security methods which consider real-time system information to be useful in the operational phase. This could help to tackle a practical problem in the operational phase: the abnormal behaviour in a component of the ICS due to attacks is initially diagnosed as a technical failure, which might result in choosing inappropriate response strategies [1].

Bayesian Networks (BNs) showed the potential to tackle this challenge especially based on their applications in medical diagnosis [12] and fault diagnosis [13]. BNs have been used in cyber security, but an overarching overview of BN models which identifies important usage patterns and key research gaps is missing. The identified usage patterns would be helpful in using BNs to tackle the practical problem of diagnosing attacks and technical failures.

This has led to Chapter 3, “Bayesian Network Models in Cyber Security: A Systematic

Review" [14], which addressed the following sub-question:

- **SQ 2.** What are the important patterns in the use of standard Bayesian Network (BN) models in cyber security?

In Chapter 3, we identified 17 standard BN models in cyber security based on the review methodology we adopted. The identified BN models were analysed using eight different criteria: (i) citation details, (ii) data sources used to construct Directed Acyclic Graphs (DAGs) and populate Conditional Probability Tables (CPTs), (iii) the number of nodes used in the model, (iv) type of threat actor, (v) application and application sector, (vi) scope of variables, (vii) the approach(es) used to validate models, and (viii) model purpose and type of purpose.

The data sources used to construct DAGs and populate CPTs in the identified BN models were expert knowledge and empirical data predominantly from cyber security reports. The identified BN models were significantly used for problems associated with malicious insiders. This could be based on the increasing percentage of attacks carried out by insiders compared to outsiders [15]. However, it would be difficult to obtain evidence in practise for some variables in BN models used for problems associated with malicious insiders as it could impact privacy.

The identified BN models were predominantly used to tackle problems associated with the Information Technology (IT) environment compared to the ICS environment. This could be because the availability of empirical data in the IT environment is better compared to the ICS environment as majority of the infrastructures operated by ICS are safety-critical. Therefore, the owners are reluctant to provide data for research from such infrastructures. The identified BN models completely or partially benefited risk management, forensic investigation, governance, threat hunting and vulnerability management in cyber security.

The identified BN models were considered as a starting point to develop a framework for constructing BN models that would help to distinguish between attacks and technical failures. For instance, the appropriate types of variables used in the identified BN models could be a basis to develop our framework. Furthermore, the identified patterns in the use of BN models in cyber security would be used as a basis to construct BN models for our application. For instance, expert knowledge is an alternate data source to tackle problems associated with ICS environment which could be a useful pattern to develop BN models for our application. This also shows that there is a need for methods in our framework that would help to effectively elicit knowledge from experts to construct DAGs and populate CPTs of BN models for our application.

Given the above conclusions and the potential of BNs to address the RQ while fulfilling the problem requirements and constraints, Chapter 4, "Combining Bayesian Networks and Fishbone Diagrams to Distinguish Between Intentional Attacks and Accidental Technical Failures" [16], addressed the following sub-question:

- **SQ 3.** How could we combine Bayesian Networks and Fishbone Diagrams to find out whether an abnormal behaviour in a component of the ICS is due to (intentional) attack or (accidental) failure or neither?

In Chapter 4, we developed attack-failure distinguisher framework to construct BN models for determining the major cause of an abnormal behaviour in a component of the ICS. This framework consists of three different types of variables adapted from existing diagnostic BNs: (i) contributory factors, (ii) problem, and (iii) observations (or test results). The contributory factors are factors that could lead to the considered problem due to an attack or technical failure, whereas the observations (or test results) provides information in the aftermath of the considered problem which could help to determine whether the problem is due to an attack or technical failure. This framework also embeds cause-effect relationship between the type of variables adapted from existing diagnostic BNs.

Because the BNs themselves are not suitable for knowledge elicitation, we extended fishbone diagrams to facilitate knowledge elicitation that would help to construct DAGs of BN models for our application. Furthermore, the typical fishbone diagrams are also not enough as they do not include observations (or test results) which needs to be elicited for our application. Therefore, we extended the typical fishbone diagrams to include observations (or test results). Extended fishbone diagrams facilitate brainstorming with experts to elicit knowledge from experts to construct DAGs of BN models for our application. Furthermore, extended fishbone diagrams allow safety and security community to work together during knowledge elicitation especially by showing the complete overview of contributory factors and observations (or test results) for the considered problem, which are categorised under intentional attack and accidental technical failure.

We demonstrated the developed methodology with an example problem "sensor (S_1) sends incorrect water level measurements" using a case study in the water management domain. We considered this example problem because this could lead to complex situations from flooding to severe economic damage. Finally, there is a lack of methods for knowledge elicitation to populate CPTs of BN models for determining the major cause of an abnormal behaviour in a component of the ICS.

In order to address the above-mentioned limitation, Chapter 5, "Probability Elicitation for Bayesian Networks to Distinguish between Intentional Attacks and Accidental Technical Failures" [17], dealt with the following sub-question:

- **SQ 4.** How could we elicit expert knowledge to effectively construct Conditional Probability Tables of Bayesian Network models for distinguishing attacks and technical failures?

In Chapter 5, we analysed state-of-the-art techniques and chose the most suitable technique to reduce the workload for experts in probability elicitation, which helps to elicit reliable probabilities from experts. Firstly, we chose the DeMorgan model to reduce the number of conditional probabilities to elicit as it could help to deal with a combination of promoting and inhibiting influences. The DeMorgan model would help to completely define the CPT of a child variable with only $(n+1)$ entries elicited from experts instead of $2^{(n+1)}$ entries, where n is the number of parent variables corresponding to the child variable.

Furthermore, we chose the probability scale with numerical and verbal anchors to facilitate the individual probability entry as they are effective and practicable based on previous studies. The probability scale with numerical and verbal anchors provides numerical anchors for experts who prefer numbers and verbal anchors for experts who

prefer words. There is also an option for experts to express fine-grained probabilities using the probability scale with numerical and verbal anchors as an aid. This completes the holistic attack-failure distinguisher framework for distinguishing attacks and technical failures.

We demonstrated the developed methodology using a case study in the water management domain with the example problem "sensor (S_1) sends incorrect water level measurements". Finally, the attack-failure distinguisher framework needs to be evaluated realistically.

This has prompted Chapter 6, "Bayesian Network Model to Distinguish between Intentional Attacks and Accidental Technical Failures: A Case Study in Floodgates" [18], which tackled the following sub-question:

- **SQ 5.** How could we develop Bayesian Network (BN) models for distinguishing attacks and technical failures in Floodgates?

In Chapter 6, we constructed a BN model for determining the major cause of a problem "sensor sends incorrect water level measurements" using the attack-failure distinguisher framework. We utilised the multi-methodology approach which includes focus groups and questionnaires to gather data from experts to construct DAG and populate CPTs. We conducted a focus group session to gather data from experts to construct DAG, which had five participants who have a lot of experience working with safety and/or security of ICS in the water management sector in the Netherlands. We complemented it with a questionnaire to gather data to construct DAG, which had nine respondents who have at least a year of experience working with safety and/or security of ICS. Furthermore, we conducted another focus group session to review the constructed DAG and populate CPTs, which had five participants who have a lot of experience working with safety and/or security of ICS in the water management sector in the Netherlands. We complemented it with a questionnaire to populate CPTs, which had five respondents who have at least a year of experience working with safety and/or security of ICS in the water management sector in the Netherlands. In the constructed DAG, there were eight contributory factors for the problem which we considered. The CPT size of the problem is 512 ($2^{(8+1)}$) entries. We elicited nine entries for the CPT corresponding to the problem from the experts and computed the other probabilities to completely define the CPT of the problem using the DeMorgan model. This significantly reduced the workload of experts in probability elicitation. Finally, we evaluated the constructed BN model using two different illustrative scenarios to demonstrate the utility or suitability of the developed artefact (attack-failure distinguisher framework). The first illustrative scenario shows that the most likely cause for the considered problem is technical failure, whereas the second illustrative scenario shows that the most likely cause for the considered problem is attack based on the evidences provided.

We tackled the main research question of this thesis by: (i) providing a holistic attack-failure distinguisher framework which also include methods to effectively elicit knowledge from experts to construct DAGs and populate CPTs of BN models for our application, (ii) developing decision support to distinguish between intentional attacks and accidental technical failures for a problem ("sensor sends incorrect water level measurements") in a floodgate operated by ICS.

Based on the above-mentioned studies, we reflect on how we fulfilled the requirements elicited from experts during the problem identification phase of the DSR process:

- The review of BNs in cyber security showed that BN is an effective and practical alternate to data-driven approaches. BN is a knowledge-based approach which we used as a basis for the decision support we developed. This helps to deal with unavailability of data by using expert knowledge, which is substantive information on a specific domain based on system knowledge that is not commonly known by others. This addressed R1.

- The decision support which we developed take into account real-time system information in terms of evidences from operators for contributory factors and observations (or test results) to determine the major cause of the considered problem. This fulfilled R2.

- The proposed extended fishbone diagrams would facilitate to involve experts from both the department that deals with technical failures and cyber-attacks and also experts with expertise in dealing with both technical failures and cyber-attacks during the knowledge elicitation process for the development of qualitative BN model. This partly addressed R3. However, this needs to be further evaluated with safety and security experts in the future.

- BNs are difficult to interpret for ICS domain experts and are therefore unsuitable for extracting the necessary knowledge. The use of proposed extended fishbone diagrams would reduce the workload of experts to extract necessary knowledge as it is easy to understand and guides data collection. This addressed R4.

- The workload of experts during knowledge elicitation of probabilities were limited by the use of DeMorgan model, which reduces the number of conditional probabilities to elicit from experts to *(n+1)* instead of $2^{(n+1)}$. Furthermore, we employed probability scales with numerical and verbal anchors to facilitate individual probability entry. This supported experts in visualising probability range and also allowed us to elicit probabilities in a reasonable time. For instance, we were able to elicit expert knowledge on 41 probabilities in about 30 - 45 minutes during the focus group workshop based on Appendix D. This fulfilled R4.

- The reliability of knowledge elicited for developing the decision support is ensured by reducing the workload of experts and also potentially facilitating discussions between experts from different departments using the extended fishbone diagrams. This fulfilled R5.

- We utilised artificial evaluation strategy due to the unavailability of real environment for this evaluation. However, we made it more realistic by involving real-users, and realistic problems. Therefore, the results from the artificial evaluation could correspond to real use. Furthermore, the develop decision support could be used for different problems in different domains which addressed R6.

At this point, it is important to answer the key question on the effectiveness of the attack-failure distinguisher framework: *"Is the attack-failure distinguisher framework effective?"*. We assessed the BN framework and also the knowledge elicitation method including DeMorgan model and probability scale with numerical and verbal anchors. This is done based on a case study in floodgates by developing a BN model involving experts and using the developed BN model for two illustrative scenarios. The BN framework helps to structure the BN model with appropriate variables and also causal relationships between the variables are maintained. Furthermore, the DeMorgan model significantly reduces the workload of experts by reducing the number of conditional probabilities to elicit from experts. The probability scale with numerical and verbal anchors makes it easier for experts to answer in terms of probabilities in a short time. These were evident from our case study and also make the attack-failure distinguisher framework feasible in practice. Finally, the extended fishbone diagram needs further evaluation in the future.

7.2. SCIENTIFIC AND SOCIETAL IMPLICATIONS

This thesis contributed towards the answer to the RQ corresponding to a practical problem in the operational phase of water management infrastructures operated by ICS, which in turn improves critical infrastructure protection especially by enabling operators to think more proactively about reactive safety and security. Furthermore, this thesis provides methods that allow the safety and security community to work together in order to tackle common problems more effectively. Finally, this thesis contributes to scientific community mainly on two themes: (i) integration of safety and security, (ii) reactive safety and security. This section reflects implications of this thesis towards science and society.

Characteristics and limitations pave the way for advancements in integrated safety and security risk assessment methods: An overarching overview of integrated safety and security risk assessment methods was missing. A part of this thesis contributes to the scientific community especially by addressing the above-mentioned research gap. After the review on existing integrated safety and security risk assessment methods was conducted as a part of this thesis, the researchers continued to recognise the importance of integrating safety and security risk assessments by developing integrated safety and security risk assessment methods [19–23].

This work would act as a starting point for developing more effective integrated safety and security risk assessment methods especially by considering key characteristics and limitations of existing methods, which is already evident from some of the scientific literature that cited our work [19, 20, 24]. We highlighted that this work would act as a base to investigate the combinations of safety, and security risk assessment methods that could be used in the future to develop integrated safety and security risk assessment methods. The scientific literature that cited this work investigated different combinations of safety, and security risk assessment methods. For instance, Temple et al. combined aspects of System-Theoretic Process Analysis for Security (STPA-Sec) and FMVEA to develop their integrated safety and security risk assessment method [19]. In addition, Bernsmed et al. used the combination of bow-tie diagrams to perform safety risk assessment and the variation of the bow-tie diagrams to perform security risk assessment [20].

Furthermore, this work would also help in practice as a guide to apply the appropriate integrated safety and security risk assessment method. The result from the application

of the relevant method would mainly help to choose appropriate risk treatments and prevent a problem occurring due to technical failure/attack. This in turn would help to ensure societal well-being and prevent economic impact. Based on this work, five out of seven methods are appropriate to perform risk analysis. Moreover, three out of these five methods were already applied in the transportation domain. This could provide a shortlist of the integrated safety and security risk assessment methods that are suitable for performing risk analysis in the transportation domain.

Patterns in the use of BN models in cyber security guide new applications: A comprehensive review of BN models in cyber security was missing. A part of this thesis contributes to the scientific community especially by addressing the above-mentioned research gap. This work would act as a knowledge base with important patterns in the use of BNs in cyber security and key research gaps that needs to be addressed in the future. This work would help in the practical application of BNs in cyber security and to investigate the use of BNs that could benefit other applications in cyber security. This is already evident from some of the scientific literature that cited our work [25–27]. Wang et al. proposed a two-layer framework to construct BN models that would help to determine the specific attack technique used to cause the identified type of attack [27]. For instance, we determined that the adversary performed Denial of Service (DoS) attack. However, DoS attack could be performed using different attack techniques such as teardrop, ping of death. The attack technique used by the adversary to perform the DoS attack could be identified using BN models constructed based on the developed framework [27]. We concluded that it would be intriguing to investigate how to deal with multi-step attacks using standard BNs, which is the basis for the recent scientific work conducted by Solomon on predicting multi-stage attack with normal IP addresses on a computer network using BNs [25].

An important pattern on the data sources used to construct DAGs and populate CPTs could help in practice as a guide to develop new BN models in cyber security. We concluded that the expert knowledge and empirical data predominantly from cyber security reports were the data sources utilised to construct DAGs and populate CPTs. Furthermore, the availability of empirical data in the IT environment is much better compared to the ICS environment. This pattern guided us to an alternate solution when the empirical data was not available in developing a BN model for our application, which also resulted in choosing expert knowledge as the alternate data source to construct DAGs and populate CPTs.

Attack-failure distinguisher framework can be used for different problems in different domains: As a part of this thesis, we developed an attack-failure distinguisher framework that would help to construct BN models for distinguishing attacks and technical failures, which also include methods that would help to effectively elicit knowledge from experts to construct DAGs and populate CPTs. This framework would support practical applications in the future, especially to construct BN models for distinguishing attacks and technical failures for different problems in different domains. This is also evident from the application of the developed framework for a problem in the water management domain in this thesis.

When we rely on expert knowledge as the data source, there need to be appropriate methods to effectively elicit knowledge from experts. The attack-failure distinguisher

framework includes extended fishbone diagrams to support brainstorming with experts in constructing the DAGs of BN models for our application. Furthermore, the attack-failure distinguisher framework includes DeMorgan model and probability scale with numerical and verbal anchors to effectively elicit probabilities from experts to completely define CPTs of BN models for our application. These methods would be used in practical applications for different problems in different domains in the future. Furthermore, extended fishbone diagrams would also help to elicit knowledge for similar problems. For instance, Jacobs used extended fishbone diagrams for an example ProRail case related to carriage registration [28]. They populated the contributory factors on the left side of the extended fishbone diagram for the problem ("Incorrect registration"). Furthermore, they populated with observations on the right side of the extended fishbone diagram.

Motivates the need for methods that integrate safety and security: This research motivates the need for integrated safety and security methods that would facilitate safety and security community to work together and share relevant information to tackle common problems in a more effective way. In this thesis, we tackled a common problem that needs safety and security community to work together and share relevant information. In order to bring the safety and security community together, there need to be appropriate methods that facilitate the safety and security community to work together and share relevant information. In our work, we developed extended fishbone diagrams that facilitate brainstorming with experts from the safety and security community to construct the DAGs of BN models.

Knowledge-based approaches are appropriate for modelling cyber security for ICS: Currently, there is a lack of empirical data for modelling cyber security in ICS environment as majority of the infrastructures operated by ICS are safety-critical. Therefore, the owners are reluctant to provide data for research from such infrastructures. A part of this thesis showed that expert knowledge is an alternate data source to model cyber security for ICS when the empirical data is unavailable. In the future, the unavailability of empirical data would not deter modelling cyber security for ICS anymore. In this thesis, we used a knowledge-based approach in constructing models for our application. Furthermore, this would motivate the scientific community to investigate other knowledge-based approaches to model cyber security for ICS when there is an unavailability of empirical data.

Developed method enables operators to be more proactive about reactive safety and security: The results of existing integrated safety and security risk assessment methods would mainly help to choose appropriate risk treatments during the design phase before an attack or technical failure occurs. These methods are associated with proactive safety and security. However, the results of our method would help to choose appropriate response strategies during the operational phase when an attack or technical failure occurs. This method is associated with reactive safety and security.

Currently, the abnormal behaviour in a component of the ICS due to attacks is initially diagnosed as a technical failure [1]. This leads to choosing ineffective response strategies. In the future, the method which we developed would help operators to be better prepared when they encounter a problem in a component of the ICS. The developed method would help to construct BN models for different problems that an operator could observe. In the constructed BN model, the operators would provide evidences for contributory factors

and observations (or test results), which would help to determine whether the abnormal behaviour in component of the ICS is due to an attack or technical failure based on posterior probabilities. This information would in turn help the operators to be better prepared as it enables them to know what they are dealing against. This information with root cause details would also help to reduce/prevent impact on societal well-being and/or economy by choosing appropriate response strategies once the problem has occurred. In order to choose appropriate response procedures, this information alone is not enough as choosing the effective response strategy also depends on the root cause (attack vector used to cause the problem in case of an attack or failure mode caused the problem in case of technical failure). BNs could also be used to determine the root cause once we determine whether it is due to an attack or technical failure, which would be detailed in future research directions.

This thesis proposed a solution to the practical problem which could impact societal well-being and economy. Furthermore, this research contributed to methods that would allow safety and security community to join forces and tackle a common problem. Finally, this study contributed to the scientific community on the integration of safety and security, reactive safety and security.

7.3. LIMITATIONS

This section explains the limitations of this thesis.

Historical data on attacks and technical failures in water management sector is not available for research: At the start of this research, we anticipated that we might get data from existing attacks and technical failures in the water management sector to gather appropriate contributory factors, test results (or observations) and probabilities. However, this is not available for research due to sensitivity issues. Therefore, we relied on expert knowledge which is an effective alternative when there is a lack of data based on the inputs which we received from the interviews with experts during requirements elicitation and related works which we reviewed. The other alternatives such as red team vs. blue team exercises were not possible due to practicalities, especially there is a lack of testbeds which could facilitate such exercises in the Netherlands. This could have improved the reliability of data used to construct DAG and populate CPTs for our application.

Lack of historical data on attacks and technical failures create dependence on experts: The lack of data excludes the use of data-driven approaches to develop decision support for distinguishing attacks and technical failures. However, we utilised BNs based on real-world applications in medical diagnosis and fault diagnosis, which is a knowledge-based approach. The dependence on experts could impact the reliability of elicited contributory factors, test results (or observations) and probabilities. Therefore, we developed a framework with different methods that would support knowledge elicitation from experts which includes: (i) extended fishbone diagrams to elicit contributory factors and test results (or observations), (ii) DeMorgan model to reduce the number of conditional probabilities to elicit from experts and (iii) probability scale with numerical and verbal anchors to facilitate individual probability entry. These methods would enhance the reliability of elicited contributory factors, test results (or observations) and probabilities by reducing the workload of experts during knowledge elicitation from experts.

Limited experts on safety and/or security of ICS in the water management sector

impacts sample size: We relied on experts who associate themselves with safety and/or security of ICS to elicit contributory factors and test results (or observations). We also relied on experts who associate themselves with safety and/or security of ICS in the water management sector in the Netherlands to elicit probabilities. This enhances the reliability of elicited contributory factors, test results (or observations) and probabilities as they have prior knowledge about the system. However, this leads to the limitation of fewer respondents. In the Netherlands, there is a limited group of safety and/or security experts in the water management sector. Therefore, we utilised snowball sampling as it helps to reach more number experts in that limited target group.

Limited time availability of experts impacts sample size: Initially, we employed focus groups as a technique to elicit contributory factors, test results (or observations) and probabilities. However, there were practical difficulties to gather a group of people at the same time due to the limited time availability of experts. This resulted in focus groups with a bit less number of experts (five). Therefore, we complemented focus groups with questionnaires to reach a bit more number of experts in that limited target group. Due to limited target group and time availability of experts, it was not possible to reach much more experts to elicit contributory factors, test results (or observations) and probabilities. In case we had a larger sample size, statistical tests would have been possible to identify significant relationships within the elicited data and generate more accurate results.

Naturalistic evaluation of the developed artefact is not possible as the real system is unavailable: The real water management infrastructure like a floodgate is not available for the evaluation of the developed artefact (attack-failure distinguisher framework) due to availability and criticality issues. Therefore, we could not perform naturalistic evaluation, which involves evaluating the developed artefact with real users and real systems in the real setting. Therefore, we relied on the artificial evaluation, which involves evaluating the developed artefact in a contrived and non-realistic way. However, we made it more realistic with real-users, and realistic problems to correspond the results to real use.

We evaluated the developed artefact based on a case study in floodgates by developing a prototype involving experts and using the developed prototype for two illustrative scenarios. However, the developed artefact including the methods to effectively elicit expert knowledge to construct DAGs and populate CPTs are generic, which could be applied in different domains. This is evident from the application of extended fishbone diagrams for an example ProRail case related to carriage registration [28].

7.4. FUTURE RESEARCH DIRECTIONS

In each chapter of this thesis, we provided future research directions specific to the study. This section covers future research directions in the broader context and how this could be addressed in the future.

Need for root cause analysis framework to choose appropriate response strategies: This thesis developed the attack-failure distinguisher framework to construct BN models that would help to distinguish between attacks and technical failures. In order to choose effective response strategies, the operators would also need to identify the root cause. In case of an attack, they would need to identify the attack vector used to cause the observed problem. On the other hand, they would need to identify the failure mode that caused the observed problem in case of technical failure. The attack-failure distinguisher framework

does not have this capability. However, the attack-failure distinguisher framework could provide input to the potential root cause analysis framework by determining whether the problem is due to an attack or technical failure.

In the future, the complete root cause analysis framework could be developed based on the BN model developed by Curiac et al. that assists in diagnosis of psychiatric disease [29]. In addition to disease specific risk factors and symptoms, the BN model developed by Curiac et al. consists of four different psychiatric diseases. Instead of the psychiatric diseases in the middle layer, the root cause analysis framework would have different attack vectors in case of an attack or failure modes in case of a technical failure. Once the attack-failure distinguisher framework provides input, the root cause analysis framework determines the attack vector in case of an attack or failure mode in case of a technical failure. It would be much more beneficial to develop a root cause analysis framework using BNs which determines the attack vector/failure mode as it would help operators to choose appropriate response strategies considering specific attack vector or failure mode.

Decision tree framework helps visualising and choosing effective response strategies: The attack-failure distinguisher framework is developed in this thesis which would help to construct BN models for distinguishing attacks and technical failures. The structure of a decision tree could be used to visualise the effective response strategies for each attack vector and failure mode. The basic structure of a decision tree involves three different type of nodes: (i) root node, (ii) internal node, and (iii) leaf node [30]. In our application, the root node could be a problem. Furthermore, the first layer of internal nodes could be the major causes of the problem (attack and technical failure). The second layer of internal nodes could be the attack vectors and failure modes. Finally, the leaf nodes could be effective response strategies corresponding to each attack vector and failure mode. Once the BN model developed using an attack-failure distinguisher framework determines whether the problem is caused by an attack or technical failure, this could be used as an input to the BN model developed using a root cause analysis framework in the future. Based on the input from the BN model developed using a root cause analysis framework regarding the specific attack vector or failure mode, the decision tree visualisation could support operators to choose effective response strategies. This could also help to consider safety and security interdependencies especially mutual reinforcement and antagonism as it visualises effective response strategies corresponding to each attack vector and failure mode.

Investigate methods to tackle multiple problems at the same time: The attack-failure distinguisher framework is applicable when there is a problem observed to determine whether the problem is due to an attack or technical failure. However, when there are multiple problems observed at the same time, it would be interesting to investigate whether we could still consider it as a separate problem and apply the developed attack-failure distinguisher framework or there needs to be an alternate framework that is applicable for such cases. Similarly, the extended fishbone diagrams can help to elicit expert knowledge for each problem separately. It would be useful to extend the extended fishbone diagrams to allow eliciting expert knowledge for multiple problems at the same time as it could provide a complete overview.

Important to predict the failure and attack probability in water management domain: The Repository of Industrial Security Incidents (RISI) database contains informa-

tion about cyber-attacks in different domains like water management, transportation [31]. However, this is not up-to-date and specific to the water management domain. With adequate historical data about cyber-attacks and technical failures in the water management domain, existing integrated safety and security risk assessment methods could help to predict the failure and attack probability in the future. This is appropriate as the result would mainly help to choose appropriate risk treatments and prevent a problem occurring due to technical failure/attack. These methods are associated with proactive safety and security.

Require further evaluation before use in real environment: A part of the attack-failure distinguisher framework which includes DeMorgan model and probability scales with numerical and verbal anchors was validated by developing a BN model for a problem in water management domain involving safety and security experts in water management in the Netherlands. However, extended fishbone diagram which is also a part of the attack-failure distinguisher framework needs further evaluation involving safety and security experts. Furthermore, the developed BN model is validated using expert evaluation and illustrative scenarios. However, this BN model needs to be further evaluated when the problem occurs in real environment. This would also help to answer the key question on the effectiveness of attack-failure distinguisher framework completely.

Explore use of alternate data sources for our application: This thesis used expert knowledge to develop decision support for the problem ("sensor sends incorrect water level measurements"). However, there are other potential data sources which need to be investigated to make it more objective. For instance, the technical failure report data could be a data source for eliciting information related to technical failures as it could provide contributory factors. Furthermore, the red team vs. blue team exercise could be a data source for eliciting information related to attacks as it could help to improve the reliability of elicited data. For instance, the Critical Infrastructure Security Showdown (CISS) is conducted by Singapore University of Technology and Design on their Secure Water Treatment (SWaT) testbed. Such type of events could provide information about contributory factors and observations (or test results) corresponding to attacks. For instance, we could interview members of the red team regarding which factors in the infrastructure contributed to the success of their attack. Furthermore, we could interview members of the blue team regarding tests (or observations) which help them to diagnose an attack.

Availability of data create opportunities for data-driven approaches in modelling cyber security for ICS: When there is an availability of historical data on attacks and technical failures in water management infrastructures, the machine learning algorithms such as logistic regression, artificial neural networks could be used to tackle our problem. Furthermore, the models developed using such machine learning algorithms could be evaluated based on their performance.

REFERENCES

[1] Macaulay, T., Singer, B. L.: Cybersecurity for Industrial Control Systems: SCADA, DCS, PLC, HMI, and SIS, Auerbach Publications. (2016)

[2] Peffers, K., Rothenberger, M., Tuunanen, T., Vaezi, R.: Design Science Research Evalua-

tion, In International Conference on Design Science Research in Information Systems, pp. 398 - 410, Springer. (2012)

[3] March, S. T., Story, V. C.: Design Science in the Information Systems Discipline: An Introduction to the Special Issue on Design Science Research, MIS Quarterly, vol. 32, no. 4, pp. 725 - 730. (2008)

[4] Chockalingam, S., Hadžiosmanović, D., Pieters, W., Teixeira, A., van Gelder, P.: Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications, International Conference on Critical Information Infrastructures Security, pp. 50 - 62, Springer. (2016)

[5] Macher, G., Höller, A., Sporer, H., Armengaud, E., Kreiner, C.: A Combined Safety – Hazards and Security - Threat Analysis Method for Automotive Systems. Koornneef, F., van Gulijk, C. (eds.) SAFECOMP 2015 Workshops. LNCS, vol. 9338, pp. 237 – 250. Springer, Heidelberg (2015)

[6] Schmittner, C., Ma, Z., Schoitsch, E., Gruber, T.: A Case Study of FMVEA and CHASSIS as Safety and Security Co-Analysis Method for Automotive Cyber Physical Systems. In: Proceedings of the 1st ACM Workshop on Cyber Physical System Security (CPSS), pp. 69 – 80. (2015)

[7] Sabaliauskaite, G., Mathur, A.P.: Aligning Cyber-physical System Safety and Security. Cardin, M.A., Krob, D., Cheun, L.P., Tan, Y.H., Wood, K. (eds.) Complex Systems Design & Management Asia 2014. LNCS, pp. 41 – 53. (2015)

[8] Schmittner, C., Ma, Z., Smith, P.: FMVEA for Safety and Security Analysis of Intelligent and Cooperative Vehicles. Bondavalli, A., Ceccarelli, A., Ortmeier, F. (eds.) SAFECOMP 2014 Workshops. LNCS, vol. 8696, pp. 282 – 288. Springer, Heidelberg (2014)

[9] Chen, Y., Chen, S., Hsiung, P., Chou, I.: Unified Security and Safety Risk Assessment – A Case Study on Nuclear Power Plant. In: Proceedings of the International Conference on Trusted Systems and their Applications (TSA), pp. 22 – 28. (2014)

[10] Steiner, M., Liggesmeyer, P., Combination of Safety and Security Analysis – Finding Security Problems that Threaten the Safety of a System. In: Workshop on Dependable Embedded and Cyber-physical Systems (DECS), pp. 1 – 8. (2013)

[11] Fovino, I.N., Masera, M., De Cian, A., Integrating Cyber Attacks within Fault Trees. Reliability Engineering and System Safety. vol. 94, no. 9, pp. 1394 – 1402. (2009)

[12] Nikovski, D.: Constructing Bayesian Networks for Medical Diagnosis from Incomplete and Partially Correct Statistics, IEEE Transactions on Knowledge & Data Engineering, vol. 9, no. 4, pp. 509 - 516, IEEE. (2000)

[13] Nakatsu, R.T.: Reasoning with Diagrams: Decision-Making and Problem-Solving with Diagrams. John Wiley & Sons. (2009)

[14] Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P.: Bayesian Network Models in Cyber Security: A Systematic Review, Nordic Conference on Secure IT Systems, pp. 105 - 122, Springer. (2017)

[15] Sohal, A. S., Sandhu, R., Sood, S. K., Chang, V.: A Cybersecurity Framework to Identify Malicious Edge Device in Fog Computing and Cloud-of-things Environments, Computers & Security, vol. 74, pp. 340-354. (2018)

[16] Chockalingam, S., Pieters, W., Teixeira, A., Khakzad, N., van Gelder, P.: Combining Bayesian Networks and Fishbone Diagrams to Distinguish between Intentional Attacks and Accidental Technical Failures, In: Graphical Models for Security, pp. 31 - 50, Springer. (2019)

[17] Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P.: Probability Elicitation for Bayesian Networks to Distinguish between Intentional Attacks and Accidental Technical Failures (Submitted to a Journal), pp. 1 - 24. (2020)

[18] Chockalingam, S., Pieters, W., Teixeira, A., van Gelder, P.: Bayesian Network Model to Distinguish between Intentional Attacks and Accidental Technical Failures: A Case Study in Floodgates (Submitted to a Journal). (2020)

[19] Temple, W. G., Wu, Y., Chen, B., Kalbarczyk, Z.: Systems-theoretic Likelihood and Severity Analysis for Safety and Security Co-engineering, In: International Conference on Reliability, Safety and Security of Railway Systems, pp. 51 - 67, Springer. (2017)

[20] Bernsmed, K., Frøystad, C., Meland, P. H., Nesheim, D. A., Rødseth, Ø., J.: Visualizing Cyber Security Risks with Bow-Tie Diagrams, In: International Workshop on Graphical Models for Security, pp. 38 - 56, Springer. (2017)

[21] Verma, S., Gruber, T., Schmittner, C., Puschner, P.: Combined Approach for Safety and Security, In: International Conference on Computer Safety, Reliability, and Security, pp. 87 - 101, Springer. (2019)

[22] Guzman, N. H. C., Kufoalor, D. K. M., Kozin, I., Lundteigen, M. A.: Combined Safety and Security Risk Analysis Using the UFoI-E Method: A Case Study of an Autonomous Surface Vessel, In: 29th European Safety and Reliability Conference, pp. 4099 - 4106. (2019)

[23] Torkildson, E. N., Li, J., Johnsen, S. O.: Improving Security and Safety Co-analysis of STPA, Proceedings of the 29th European Safety and Reliability (ESREL) Conference, Research Publishing Services. (2019)

[24] Dobaj, J., Iber, J., Krisper, M., Kreiner, C.: Towards Executable Dependability Properties, In: European Conference on Software Process Improvement, pp. 341 - 353, Springer. (2018)

[25] Alile, O. S.: Predicting Multi-stage Attack with Normal IP Addresses on a Computer Network Using Bayesian Belief Network, University of Benin. (2018)

[26] Pappaterra, M. J.: Implementing Bayesian Networks for Online Threat Detection, Linnaeus University. (2018)

[27] Zhou, Y., Zhu, C., Tang, L., Zhang, W., Wang, P.: Cyber Security Inference Based on a Two-Level Bayesian Network Framework, In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 3932 - 3937, IEEE. (2018)

[28] Jacobs, F.: Safety Through Machine Learning Applications: A Safety Case Analysis, Delft University of Technology. (2018)

[29] Curiac, D. -I., Vasile, G., Banias, O., Volosencu, C., Albu, A.: Bayesian Network Model for Diagnosis of Psychiatric Diseases, In: Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces, pp. 61 - 66, IEEE. (2009)

[30] Yu, Z., Haghighat, F., Fung, B. C., Yoshino, H.: A Decision Tree Method for Building Energy Demand Modeling, Energy Buildings, vol. 42, no. 10, pp. 1637 - 1646. (2010)

[31] RISI Database.: The Repository of Industrial Security Incidents.http://www.risidata.com

A

Requirements Elicitation – Discussion Guide

Q1. When the operator notices an abnormal behaviour in a component of the ICS, how do they respond to it?

Q2. Do you have a mechanism for the operator to determine whether an abnormal behaviour in a component of the ICS is due to attacks or technical failures?

Q3. Does the same department deal with the attacks and technical failures? If not, how?

Q4. Which functionalities do you think are important in a system which helps to distinguish between attacks and technical failures?

Q5. Are there any cyber-attacks reported in your infrastructure?

Q6. Are there any technical failures reported in your infrastructure?

Q7. Do you have a repository of technical failure reports?

Q8. If so, whether this repository of technical failure reports is available for research or not?

Q9. What do you think are the alternate data sources available for research?

Q10. What are the challenges you foresee in the alternate data sources you proposed?

Q11. In addition to risk factors and symptoms based on tests, what are other elements that you would take into account when you diagnose an (intentional) attack on a

component?

Q12. In addition to risk factors and symptoms based on tests, what are other elements that you would take into account when you diagnose (accidental) technical failure?

Q13. Is it possible to evaluate the developed method in the real water management infrastructure? If so, are there any challenges?

Q14. Whether do we have access to system architectures of any real water management infrastructure or not?

B

NOISY-OR MODEL AND CAUSAL STRENGTH (CAST) LOGIC

B.1. NOISY-OR MODEL

The noisy-OR model is applicable when there are several parents (causes) and a common child (effect) as shown in Figure B.1. In general, the CPT size of a binary variable with n binary parents is $2^{(n+1)}$. However, only n parameters are sufficient to completely define CPT using the noisy-OR model.

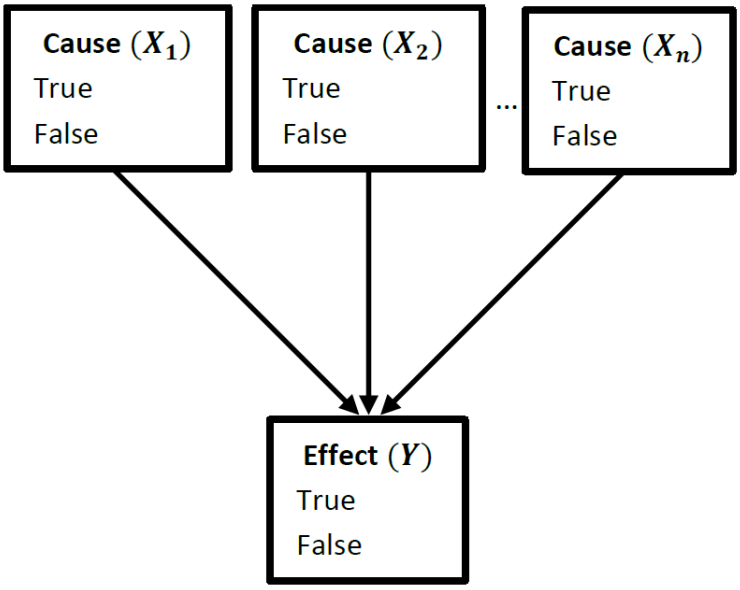


Figure B.1: Noisy-OR Model: Structure

B

In the noisy-OR model, each cause variable (X_i) has the values x_i and x_i' for the presence and absence of the cause respectively. Furthermore, the effect variable (Y) has values y for the effect being present and y' for the effect being absent. The noisy-OR model assumes that the properties of exception independence and accountability holds true [1]. The property of exception independence states that presence of any single cause is enough to produce the effect and that the hidden processes that may inhibit the occurrence of the effect are mutually independent [2]. In case all the modelled causes of the effect are false, the property of accountability requires that the effect be presumed false, i.e., $P(y'|x_1',x_2',...,x_n') = 1$.

In the noisy-OR model, the effect can be caused by any cause similar to a logical-OR. However, the relationship is not deterministic – each of the causes X_i alone can cause the effect with probability p_i, which is known as link probability [3].

$$p_i = P(y|only\ X_i \text{ is present}) = P(y|x_1',x_2',...,x_i,...,x_n')$$

Where $x_1',x_2',...,x_i,...,x_n'$ represents the absence of the other causes except X_i.
The probability of any combination of active causes can be calculated as:

$$P(y|X) = 1 - \Pi_{x_i \epsilon X}(1 - p_i)$$

Where X represents all active causes.

B.2. CAUSAL STRENGTH (CAST) LOGIC

CAST logic is applicable when there are several parents and a common child as shown in Figure B.2 [4]. CAST logic assumes all the variables in the model are binary. CAST logic is only applied in the international policy and crisis analysis domain [5]. The interaction between a parent and the common child can be either promoting or inhibiting. The promoting influence is depicted by an arrowhead, whereas the negative influence is illustrated by a filled circle as shown in Figure B.2.

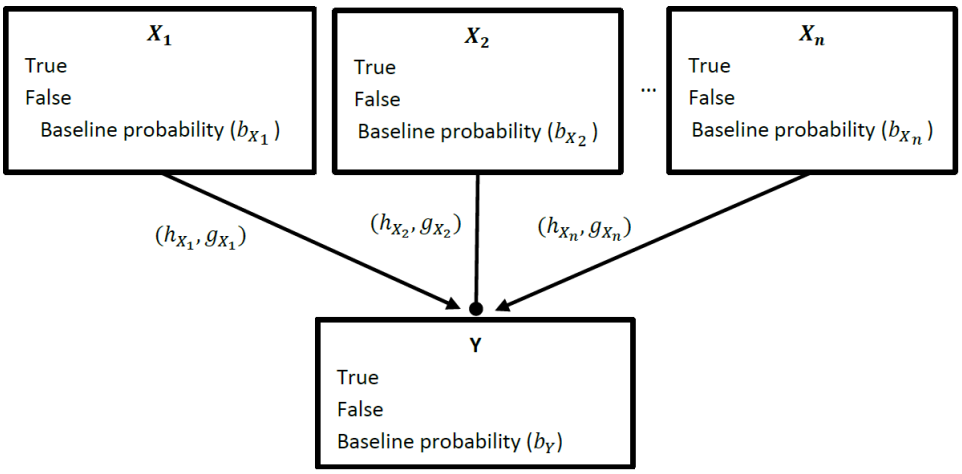


Figure B.2: CAST Parameters

The parameters which need to be elicited to completely define CPTs using CAST logic are: (i) causal strengths (g_{Xi},h_{Xi}) for each arc, and (ii) baseline probability (b) for each variable. The values of causal strengths (g_{X_i},h_{X_i}) are not probabilities and can take any arbitrary values from the range [-1, 1]. The value of causal strength (h_{Xi}) indicates the change in belief of Y relative to the baseline probability of Y (b_Y) under the assumption that X_i is in "True" state. For instance, h_{X1} indicates how much the presence of X_1 would change our belief of Y. On the other hand, the value of causal strength (g_{Xi}) indicates the change in belief of Y relative to the baseline probability of effect (b_Y) under the assumption that X_i is in "False" state. For instance, g_{X1} indicates how much the absence of X_1 would change our belief of Y.

Once we elicit the above-mentioned parameters, we could apply CAST algorithm for every combination of parent states to completely define the CPT of child variable. CAST algorithm consists of four steps: (i) aggregate positive causal strengths, (ii) aggregate negative causal strengths, (iii) combine the positive and negative causal strengths, and (iv) derive conditional probabilities.

In the first step, the positive causal strengths are aggregated using (B.1):

$$S_+ = 1 - \Pi_i (1 - s_{Xi}) \tag{B.1}$$

Where s_{Xi} can be g_{Xi} or h_{Xi} depending on the state of the parent.

In the second step, the negative causal strengths are aggregated using (B.2):

$$S_- = 1 - \Pi_i (1 - |s_{Xi}|) \tag{B.2}$$

Where s_{Xi} can be g_{Xi} or h_{Xi} depending on the state of the parent.

In the third step, the positive and negative causal strengths are combined. The overall influence (O) of all parents is determined using (B.3) if $S_+ >= S_-$ and using (B.4) if $S_- < S_+$:

$$O = 1 - ((1 - S_+)/(1 - S_-)) \tag{B.3}$$

$$|O| = 1 - ((1 - S_-)/(1 - S_+)) \tag{B.4}$$

In the final step, the conditional probabilities are derived using (B.5) if $O_j >= 0$ and using (B.6) if $O_j < 0$:

$$P(Y|X_j) = b_Y + (1 - b_Y) O_j \tag{B.5}$$

$$P(Y|X_j) = b_Y - b_Y |O_j| \tag{B.6}$$

Where O_j denotes the overall influence of j^{th} combination of parent states X_j.

REFERENCES

[1] Woudenberg, S. P., Van Der Gaag, L. C.: Using the Noisy-or Model can be Harmful... But it often is not, In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, pp. 122-133, Springer. (2011)

[2] Bolt, J. H., van der Gaag, L. C.: An Empirical Study of the Use of the Noisy-OR Model in a Real-life Bayesian Network, In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 11 - 20, Springer. (2010)

[3] Anand, V., Downs, S. M.: Probabilistic Asthma Case Finding: A Noisy OR Reformulation, In: AMIA Annual Symposium Proceedings, American Medical Informatics Association, pp. 6 - 10. (2008)

[4] Rosen, J. A., Smith, W. L.: Influence Net Modeling with Causal Strengths: An Evolutionary Approach, In: Proceedings of the Command and Control Research and Technology Symposium, Citeseer, pp. 25 - 28. (1996)

[5] Zagorecki, A.: Local Probability Distributions in Bayesian Networks: Knowledge Elicitation and Inference, In: University of Pittsburgh. (2010)

C

KNOWLEDGE ELICITATION METHOD TO DEVELOP QUALITATIVE BN MODEL

Root Cause Analysis in Industrial Control Systems – Differentiation of Cyber Attacks and Technical Failures based on Contributory Factors and Test Results (or Observations)

Objectives: To identify contributory factors (or risk factors) and tests that could help to differentiate between (accidental) component failure and an (intentional) attack on the component of Industrial Control System (ICS).

The results of this questionnaire would be used as a basis to develop a Bayesian Network model-based decision support system that could help to distinguish between (accidental) component failure and an (intentional) attack on the component of ICS in the water management sector.

This study is a first-of-its-kind. We will keep you up to date about the results of this study.

Estimated Time: 25 minutes

Examples: The examples provided below would help to clarify the terminologies used in this questionnaire. The *lung cancer example* is from medical domain which is not directly related to the questionnaire. However, this could help to easily understand the terminologies and translate it into our domain of interest. Furthermore, the *computer crash example* is from security domain which is closely related to the questionnaire. In general, the contributory factor (or risk factor) increases the likelihood of a disease or problem as shown in Figure C.1 and C.2. In addition, the test result (or observations) based on a test would help to diagnose a disease or problem after it occurred.

- **Contributory Factor:** Smoking
- **Disease:** Lung cancer
- **Test:** X-ray
- **Test Result:** Positive chest X-ray

Figure C.1: Lung Cancer - Example

- **Contributory Factor:** USB ports enabled in your computer
- **Problem:** Your computer crashes/restarts/shutdown due to an (intentional) attack
- **Test:** Run a malware scan in your computer
- **Test Result:** The malware scanner detects malware in your computer during the scan performed

Figure C.2: Computer Crash - Example

Case Outline

This is a hypothetical floodgate primarily operated by Supervisory Control and Data Acquisition (SCADA) system. Figure C.3 illustrates the physical layout of the floodgate and the view of operations centre.

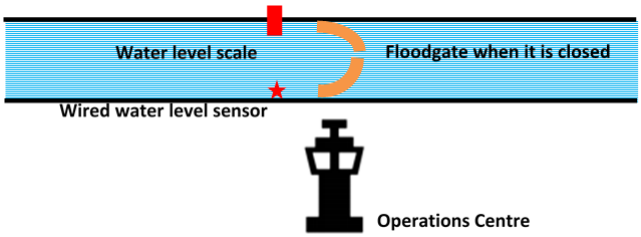


Figure C.3: Physical Layout of the Hypothetical Floodgate

The operator has a clear view of the floodgate from the operations centre. Figure C.4 illustrates the SCADA architecture of the hypothetical floodgate. The sensor (which is located near the floodgate) is used to measure the water level. There is also a water level scale which is visible to the operator from the operations centre. The sensor measurements are then sent to the PLC. If the water-level reaches the higher limit, PLC would send an alarm notification to the operator through Human Machine Interface (HMI), and the operator would need to close the floodgate through HMI. In addition, HMI would also provide information like the water-level and the current state of the floodgate (open/closed). The actuator opens/closes the floodgate.

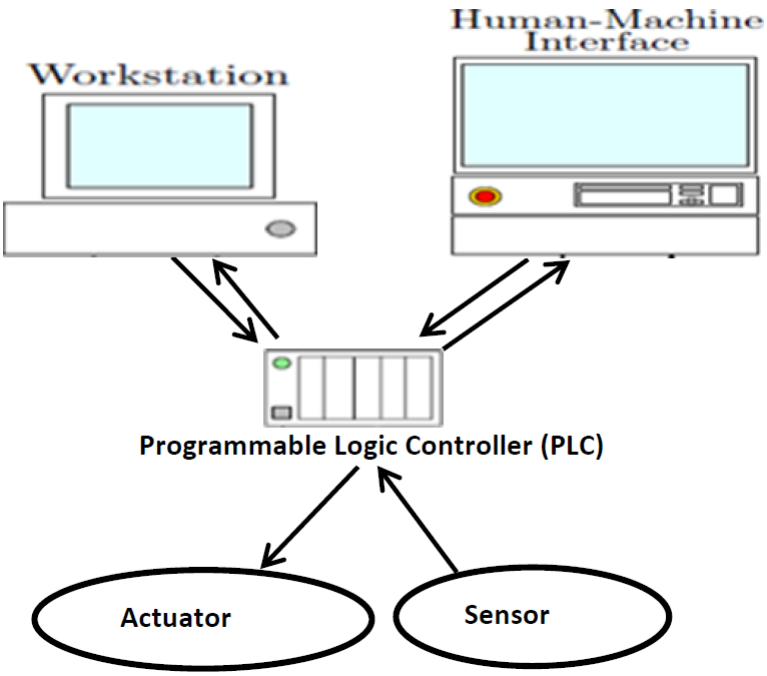


Figure C.4: SCADA Architecture of the Hypothetical Floodgate

Note: The case outline is provided to get you started. If you think anything is missing in the case outline, you could make your own assumptions, and explicitly mention it in your response.

Questions

Please answer the following questions to the best of your ability.

Background Information

1. How many years of experience do you have working with Industrial Control Systems (ICS)?

---

2. Which sector(s) do you work in?

☐ **Chemical**
☐ **Defence**
☐ **Energy**
☐ **Financial**
☐ **Nuclear**
☐ **Transport**

☐ **Water**
☐ **Others, please specify:** _____

3. Which community do you associate yourself with based on your experience?

☐ **Safety (dealing with accidental/non-malicious threats)**
☐ **Security (dealing with intentional/malicious threats)**
☐ **Both safety and security**
☐ **Others, please specify:** _____

Problem: The sensor sends incorrect water level measurements.

4. Which contributory factors would increase the likelihood of the problem due to (accidental) sensor failure?

_____

5. Which contributory factors would increase the likelihood of the problem due to an (intentional) attack?

_____

6. Which tests would you execute to distinguish between (accidental) sensor failure and an (intentional) attack on the sensor for the problem?

_____

7. If you have listed more than 1 test for 6., please rank the tests in the order of importance with "first rank" being the most significant test that would provide more clarity on the difference between (accidental) sensor failure and an (intentional) attack on the sensor, and "last rank" being the least significant test that would provide less clarity on the difference between (accidental) sensor failure and an (intentional) attack on the sensor.

_____

Miscellaneous

8. In addition to contributory factors and test results, what are other elements that you would take into account when you diagnose an (intentional) attack on a component?

_____

9. In addition to contributory factors and test results what are other elements that you would take into account when you diagnose (accidental) component failure?

---

10. What are the important elements that need to be included when you document an (intentional) cyber-attack?

---

11. What are the important elements that need to be included when you document an (accidental) technical failure?

D

KNOWLEDGE ELICITATION METHOD TO DEVELOP QUANTITATIVE BN MODEL

Probability Elicitation for the Bayesian Network (BN) Model to Distinguish Between Intentional Attacks and Accidental Technical Failures in Industrial Control Systems (ICS) based Floodgate

Objectives: To elicit probabilities corresponding to each variable in our BN model that could help to determine the major cause (intentional attack or accidental technical failure) of the problem (sensor sends incorrect water level measurements) when observed.

The results of this questionnaire would be used to complete a BN model-based decision support system for Rijkswaterstaat to determine the major cause of the problem when observed.

This study is a first-of-its-kind. We will keep you up to date about the results of this study.

Estimated Time: 40 minutes

Case Outline

Note: The case outline is provided to get you started and not completely depend on this for answering the questions.

This is a hypothetical floodgate primarily operated by Supervisory Control and Data Acquisition (SCADA) system. Figure D.1 schematises a floodgate being primarily operated by SCADA system along with an operation centre.

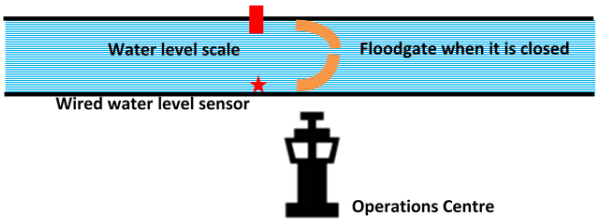


Figure D.1: Physical Layout of the Hypothetical Floodgate

Figure D.2 illustrates the SCADA architecture of the floodgate. The sensor, which is located near the floodgate, is used to measure the water level. There is also a water level scale which is visible to the operator from the operations centre. The sensor measurements are then sent to the PLC. If the water level reaches the higher limit, PLC would send an alarm notification to the operator through the Human-Machine Interface (HMI), and the operator would need to close the floodgate in this case. The HMI would also provide information such as the water level and the current state of the floodgate (open/close). The actuator opens/closes the floodgate.

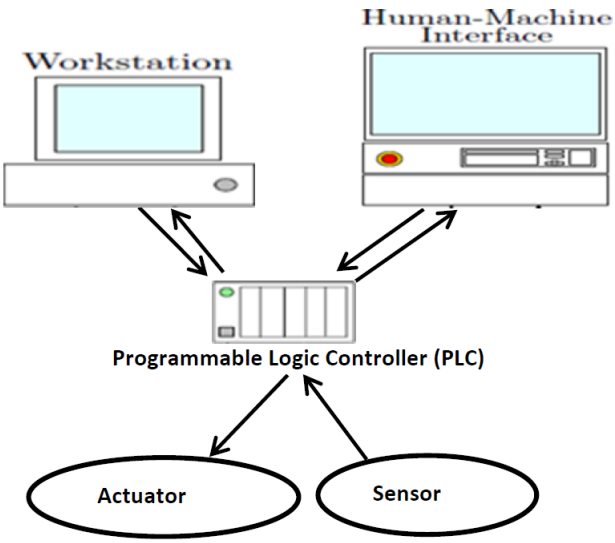
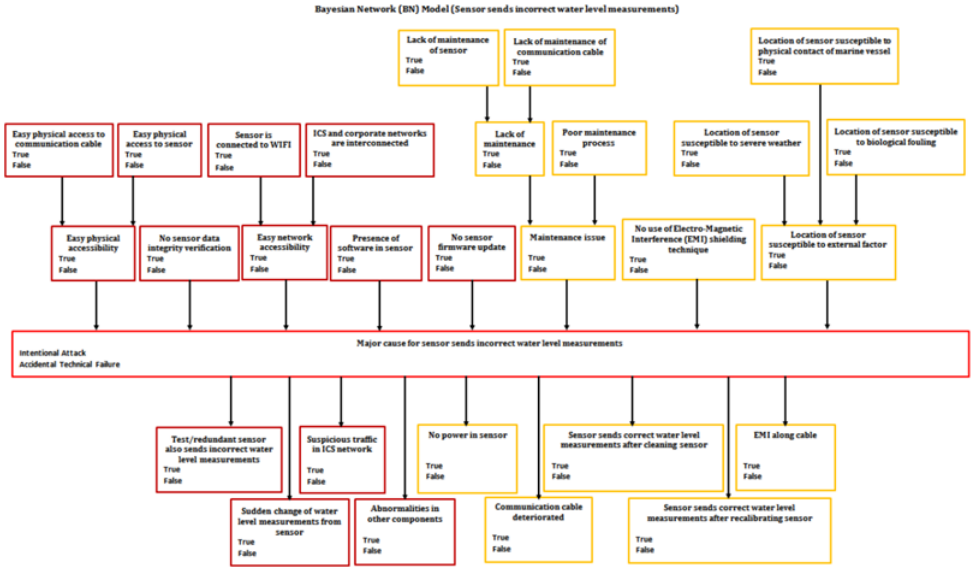


Figure D.2: SCADA Architecture of the Hypothetical Floodgate

BN Model: Please see below to know about the constructed qualitative BN model to determine the major cause (intentional attack or accidental technical failure) of the problem (sensor sends incorrect water level measurements) when observed. You will find the questions in next pages corresponding to each variable in our BN model.



Questions

Please answer the questions taking into account the type of floodgates that have the criticality rating as "very high" (on a 5-point scale: very low – low – medium – high – very high). Furthermore, please answer the questions by marking the suitable probability among 7 anchors ((almost) impossible (0) - Improbable (15) - Uncertain (25) - Fifty-fifty (50) - Expected (75) - Probable (85) - Certain (almost) (100)) directly or writing fine-grained probability (in the provided space) using the numerical and verbal anchors as a supporting aid.

Easy physical access to sensor													
Q1.1	How likely is it that the sensor is easily physically accessible to an unauthorized person in a floodgate operated by ICS?	(almost) impossible		improbable		uncertain		fifty-fifty		expected	probable	certain (almost)	
		0		15		25		50		75	85	100	

Easy physical access to communication cable													
Q2.1	How likely is it that the sensor communication cable is easily physically accessible to an unauthorized person in a floodgate operated by ICS?	(almost) impossible		improbable		uncertain		fifty-fifty		expected	probable	certain (almost)	
		0		15		25		50		75	85	100	

Sensor data integrity verification													
Q3.1	How likely is it that data integrity verification is performed for the sensor data in a floodgate operated by ICS?	(almost) impossible		improbable		uncertain		fifty-fifty		expected	probable	certain (almost)	
		0		15		25		50		75	85	100	

Sensor is connected to WIFI		
Q4.1	How likely is it that the sensor is connected to WIFI in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

ICS and corporate networks are connected		
Q5.1	How likely is it that the ICS and corporate networks are connected in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Presence of software in sensor		
Q6.1	How likely is it that software is present in the sensor in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

D

Sensor firmware update		
Q7.1	How likely is it that the sensor firmware is updated in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Maintenance of sensor		
Q8.1	How likely is it that the sensor is physically maintained in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Maintenance of communication cable		
Q9.1	How likely is it that the sensor communication cable is physically maintained in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Good maintenance process		
Q10.1	How likely is it that there is a good maintenance process for the sensor in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Use of Electro-Magnetic Interference (EMI) shielding technique		
Q11.1	How likely is it that EMI shielding technique is used for the sensor in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Location of sensor susceptible to severe weather		
Q12.1	How likely that location of sensor is susceptible to severe weather in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Location of sensor susceptible to physical contact of marine vessel		
Q13.1	How likely that location of sensor is susceptible to physical contact of marine vessel in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Location of sensor susceptible to biological fouling		
Q14.1	How likely that location of sensor is susceptible to biological fouling in a floodgate operated by ICS?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100

Please answer the questions (Q15.1 – Q15.9) taking into account the threat level as "substantial" which denotes there is a real chance of an attack (on a 5-point scale: minimal – limited – significant – substantial – critical). Furthermore, please answer the questions (Q15.2 – Q15.9) taking into account only the corresponding causal factor that is present (Example (Q15.2): "How likely that the major cause for the observed problem is intentional attack given that the sensor/sensor communication cable is easily physically accessible to an unauthorised person?") and assuming that other causal factors are absent. The double strikethrough text in the questions (Q15.1 – Q15.9) denotes the explicitly mentioned causal factors that are absent (Example: ~~the sensor/sensor communication cable is not easily physically accessible to an unauthorised person~~). Finally, please answer the question (Q15.1) taking into account the causal factors that are not explicitly mentioned (if any) as the explicitly mentioned causal factors are absent.

	Major cause for sensor sends incorrect water level measurements		
Q15.1	How likely that the major cause for the observed problem (sensor sends incorrect water level measurements) is intentional attack ~~given that the sensor/sensor communication cable is not easily physically accessible to an unauthorised person, data integrity verification is performed for the sensor data, the sensor is not easily accessible via network to an unauthorised person, software is not present in the sensor, the sensor firmware is always updated, the sensor/sensor communication cable is always physically maintained properly, EMI shielding technique is used for the sensor, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.2	How likely that the major cause for the observed problem is intentional attack given that the sensor/sensor communication cable is easily physically accessible to an unauthorised person~~, data integrity verification is performed for the sensor data, the sensor is not easily accessible via network to an unauthorised person, software is not present in the sensor, the sensor firmware is always updated, the sensor/sensor communication cable is always physically maintained properly, EMI shielding technique is used for the sensor, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.3	How likely that the major cause for the observed problem is intentional attack given that the ~~sensor/sensor communication cable is not easily physically accessible to an unauthorised person,~~ data integrity verification is not performed for the sensor data~~, the sensor is not easily accessible via network to an unauthorised person, software is not present in the sensor, the sensor firmware is always updated, the sensor/sensor communication cable is always physically maintained properly, EMI shielding technique is used for the sensor, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.4	How likely that the major cause for the observed problem is intentional attack given that ~~the sensor/sensor communication cable is not easily physically accessible to an unauthorised person, data integrity verification is performed for the sensor data,~~ the sensor is easily accessible via network to an unauthorised person~~, software is not present in the sensor, the sensor firmware is always updated, the sensor/sensor communication cable is always physically maintained properly, EMI shielding technique is used for the sensor, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.5	How likely that the major cause for the observed problem is intentional attack given that the ~~sensor/sensor communication cable is not easily physically accessible to an unauthorised person, data integrity verification is performed for the sensor data, the sensor is not easily accessible via network to an unauthorised person,~~ software is present in the sensor~~, the sensor firmware is always updated, the sensor/sensor communication cable is always physically maintained properly, EMI shielding technique is used for the sensor, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.6	How likely that the major cause for the observed problem is intentional attack given that ~~the sensor/sensor communication cable is not easily physically accessible to an unauthorised person, data integrity verification is performed for the sensor data, the sensor is not easily accessible via network to an unauthorised person, software is not present in the sensor,~~ the sensor firmware is not updated~~, the sensor/sensor communication cable is always physically maintained properly, EMI shielding technique is used for the sensor, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.7	How likely that the major cause for the observed problem is accidental technical failure given that ~~the sensor/sensor communication cable is not easily physically accessible to an unauthorised person, data integrity verification is performed for the sensor data, the sensor is not easily accessible via network to an unauthorised person, software is not present in the sensor, the sensor firmware is always updated,~~ the sensor/sensor communication cable is not physically maintained properly~~, EMI shielding technique is used for the sensor, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.8	How likely that the major cause for the observed problem is accidental technical failure given that ~~the sensor/sensor communication cable is not easily physically accessible to an unauthorised person, data integrity verification is performed for the sensor data, the sensor is not easily accessible via network to an unauthorised person, software is not present in the sensor, the sensor firmware is always updated, the sensor/sensor communication cable is always physically maintained properly,~~ EMI shielding technique is not used for the sensor~~, location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q15.9	How likely that the major cause for the observed problem is accidental technical failure given that the ~~sensor/sensor communication cable is not easily physically accessible to an unauthorised person, data integrity verification is performed for the sensor data, the sensor is not easily accessible via network to an unauthorised person, software is not present in the sensor, the sensor firmware is always updated, the sensor/sensor communication cable is always physically maintained properly, EMI shielding technique is used for the sensor,~~ location of the sensor is susceptible to external factor (severe weather/marine vessel/biological fouling)?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	

Please answer the questions (Q16.1 – Q24.2) taking into account the major cause ("*Intentional attack*" / "*Accidental technical failure*") of the observed problem ("*Sensor sends incorrect water level measurements*") is already known.

	Test/redundant sensor also sends incorrect water level measurements		
Q16.1	How likely is it that the test/redundant sensor also send incorrect water level measurements in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	
Q16.2	How likely is it that the test/redundant sensor also send incorrect water level measurements in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 / improbable 15 / uncertain 25 / fifty-fifty 50 / expected 75 / probable 85 / certain (almost) 100	

	Sudden change of water level measurements from sensor		
Q17.1	How likely is it that there is a sudden change of water level measurements from sensor in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q17.2	How likely is it that there is a sudden change of water level measurements from sensor in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

	Suspicious traffic in ICS network		
Q18.1	How likely is it that there is suspicious traffic in ICS network in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q18.2	How likely is it that there is suspicious traffic in ICS network in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

	Abnormalities in other components		
Q19.1	How likely is it that there are abnormalities in other the components in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q19.2	How likely is it that there are abnormalities in other the components in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

	No power in sensor		
Q20.1	How likely is it that there is no power in the sensor in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q20.2	How likely is it that there is no power in the sensor in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

	Communication cable deteriorated		
Q21.1	How likely is it that the sensor communication cable is deteriorated in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q21.2	How likely is it that the sensor communication cable is deteriorated in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

	Sensor sends correct water level measurements after cleaning sensor		
Q22.1	How likely is it that the sensor sends correct water level measurements after cleaning the sensor in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q22.2	How likely is it that the sensor sends correct water level measurements after cleaning the sensor in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

	Sensor sends correct water level measurements after recalibrating sensor		
Q23.1	How likely is it that the sensor sends correct water level measurements after recalibrating the sensor in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q23.2	How likely is it that the sensor sends correct water level measurements after recalibrating the sensor in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

	EMI along cable		
Q24.1	How likely is it that there is EMI along the sensor communication cable in a floodgate operated by ICS given that the major cause of the problem is intentional attack?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	
Q24.2	How likely is it that there is EMI along the sensor communication cable in a floodgate operated by ICS given that the major cause of the problem is accidental technical failure?	(almost) impossible 0 · improbable 15 · uncertain 25 · fifty-fifty 50 · expected 75 · probable 85 · certain (almost) 100	

Background Information

We will keep the background information anonymised for academic publishing.

Q25. Please write your name and email address (Optional).

_____

Q26. How many years of experience do you have working with Industrial Control Systems?

_____

Q27. Which sector(s) do you work in?

☐ **Chemical**
☐ **Defence**
☐ **Energy**
☐ **Financial**
☐ **Nuclear**
☐ **Transport**
☐ **Water**
☐ **Others, please specify:** _____

Q28. Which community do you associate yourself with based on your experience?

☐ **Safety (dealing with unintentional/non-malicious threats)**
☐ **Security (dealing with intentional/malicious threats)**
☐ **Both safety and security**
☐ **Others, please specify:** _____

Online Questions: Examples

Q1.1 How likely is it that the sensor is easily physically accessible to an unauthorized person in a floodgate operated by ICS?

- (almost) Impossible | 0
- Improbable | 15
- Uncertain | 25
- Fifty-fifty | 50
- Expected | 75
- Probable | 85
- (almost) Certain | 100
- Others, please specify

Q15.2 How likely that the major cause for the observed problem is intentional attack given that the sensor/sensor communication cable is easily physically accessible to an unauthorised person~~; data integrity verification is performed for the sensor data; the sensor is not easily accessible via network to an unauthorised person; software is not present in the sensor; the sensor firmware is always updated; the sensor/sensor communication cable is always physically maintained properly; EMI shielding technique is used for the sensor; location of the sensor is not susceptible to external factor (severe weather/marine vessel/biological fouling)~~?

- (almost) Impossible | 0
- Improbable | 15
- Uncertain | 25
- Fifty-fifty | 50
- Expected | 75
- Probable | 85
- (almost) Certain | 100
- Others, please specify

CURRICULUM VITÆ

Sabarathinam Chockalingam was born in Karaikudi, India, on 14th December 1990. In 2008, he obtained his high school certificate from Alagappa Matriculation Higher Secondary School, Karaikudi. In 2012, he received his Bachelor of Technology in Computer Science and Engineering from SRM University, Chennai. He became interested in cyber security while studying for his Bachelor of Technology in Computer Science and Engineering.

In 2012, Saba moved to the United Kingdom to pursue a master's degree in Cyber Security and Management at the University of Warwick. He presented his first scientific paper entitled "Cyber Security of a Wireless Vehicle" at the Kaspersky Academy – Cyber Security for the Next Generation Conference (European Round) in 2013. This furthered his interests for research in cyber security.

After graduating from the University of Warwick with distinction, Saba did his internship at the Satellite Applications Catapult Limited where his research mainly focused on Internet of Things security. In 2014, he moved to Malaysia for a short stint as a research assistant at the University of Malaya where his research mainly focused on big data problem in digital forensic investigations.

In 2015, Saba moved to the Netherlands to join the Safety and Security Science group at the Delft University of Technology as a PhD researcher. He worked for the project entitled "Secure Our Safety: Building Cyber Security for Flood Management (SOS4Flood)", funded by the Netherlands Organisation for Scientific Research (NWO). He received "CIPRNet Young CRITIS Finalist Award" for the paper entitled "Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications", which he presented at the 11th International Conference on Critical Information Infrastructures Security (CRITIS) in 2016.

In 2019, Saba moved to Norway to join the Department of Risk, Safety and Security at the Institute for Energy Technology, Halden as a research scientist. He has been involved in several projects which includes Cyber-Physical Security in Energy Infrastructure of Smart Cities (CPSEC), Developing a Serious Game Prototype for Cyber Security Training of ICT Users (Serious-Training), Safety Assessment Framework for Efficient Transport (SafeT).

LIST OF PUBLICATIONS

1. **Chockalingam, S.**, Lallie, H. S.: *"Alarming! Security Aspects of the Wireless Vehicle,"* International Journal of Cyber-Security and Digital Forensics, vol. 3, no. 4, pp. 200 – 208, 2014. https://doi.org/10.17781/P001344

2. **Chockalingam, S.**, Lallie, H. S.: *"The Conceptual Idea of Online Social Media Site (SMS) User Account Penetration Testing System,"* International Journal of Security, Privacy and Trust Management, vol. 3, no. 4, pp. 1 – 10, 2014. https://doi.org/10.5121/ijsptm.2014.3401

3. **Chockalingam, S.**, Hadžiosmanović, D., Pieters, W., Teixeira, A., and van Gelder, P.: *"Integrated Safety and Security Risk Assessment Methods: A Survey of Key Characteristics and Applications,"* International Conference on Critical Information Infrastructures Security, pp. 50 – 62, 2016. Springer, Cham.
https://doi.org/10.1007/978-3-319-71368-7_5

4. **Chockalingam, S.**, Pieters, W., Teixeira, A., and van Gelder, P.: *"Bayesian Network Models in Cyber Security: A Systematic Review,"* Nordic Conference on Secure IT Systems, pp. 105 – 122, 2017. Springer, Cham. https://doi.org/10.1007/978-3-319-70290-2_7

5. **Chockalingam, S.**, Pieters, W., Teixeira, A., Khakzad, N., and van Gelder, P.: *"Combining Bayesian Networks and Fishbone Diagrams to Distinguish between Intentional Attacks and Accidental Technical Failures,"* International Workshop on Graphical Models for Security, pp. 31 – 50, 2018. Springer, Cham.
https://doi.org/10.1007/978-3-030-15465-3_3

6. **Chockalingam, S.**, Pieters, W., Teixeira, A., Khakzad, N., and van Gelder, P.: *"Applying Bayesian Networks to Distinguish between Intentional Attacks and Accidental Technical Failures in Industrial Control Systems,"* The Fourth Annual Cyber Security Workshop in the Netherlands, 2018. [Abstract]

7. **Chockalingam, S.**, Katta, V.: *"Using Bayesian Networks for Root Cause Analysis of Observable Problems in Cyber-Physical Systems,"* 5th SRA Nordic Conference, 2019. [Abstract]

8. **Chockalingam, S.**, Katta, V.: *"Developing a Bayesian Network Framework for Root Cause Analysis of Observable Problems in Cyber-Physical Systems,"* IEEE Conference on Information and Communication Technology, pp. 1 – 6, 2019. IEEE. https://doi.org/10.1109/CICT48419.2019.9066167

9. Abbas, K., **Chockalingam, S.**, Dinh, T.T.N., and Katta, V.: *"A Prototype Tool for Distinguishing Attacks and Technical Failures in Industrial Control Systems,"* The 13th Norwegian Information Security Conference, 2020.

10. **Chockalingam, S.**, Pieters, W., Teixeira, A., and van Gelder, P.: *"Probability Elicitation for Bayesian Networks to Distinguish between Intentional Attacks and Accidental Technical Failures,"* Submitted to a Journal.

11. **Chockalingam, S.**, Pieters, W., Teixeira, A., and van Gelder, P.: *"Bayesian Network Model to Distinguish between Intentional Attacks and Accidental Technical Failures: A Case Study in Floodgates,"* Submitted to a Journal.