Combining Bayesian Networks and Fishbone Diagrams to Distinguish between Intentional Attacks and Accidental Technical Failures

Chockalingam, Saba; Pieters, Wolter; Teixeira, Andre M. H.; Khakzad, N.; van Gelder, Pieter

DOI
10.1007/978-3-030-15465-3_3

Publication date
2019

Document Version
Accepted author manuscript

Published in
Graphical Models for Security - 5th International Workshop, GraMSec 2018, Revised Selected Papers

Combining Bayesian Networks and Fishbone Diagrams to Distinguish between Intentional Attacks and Accidental Technical Failures

Sabarathinam Chockalingam[1](✉), Wolter Pieters[1], André Teixeira[2], Nima Khakzad[1], and Pieter van Gelder[1]

[1]Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands
{S.Chockalingam, W.Pieters, N.KhakzadRostami, P.H.A.J.M.vanGelder}@tudelft.nl
[2]Department of Engineering Sciences, Uppsala University, Uppsala, Sweden
Andre.Teixeira@angstrom.uu.se

Abstract. Because of modern societies' dependence on industrial control systems, adequate response to system failures is essential. In order to take appropriate measures, it is crucial for operators to be able to distinguish between intentional attacks and accidental technical failures. However, adequate decision support for this matter is lacking. In this paper, we use Bayesian Networks (BNs) to distinguish between intentional attacks and accidental technical failures, based on contributory factors and observations (or test results). To facilitate knowledge elicitation, we use extended fishbone diagrams for discussions with experts, and then translate those into the BN formalism. We demonstrate the methodology using an example in a case study from the water management domain.
Keywords: Bayesian network · Fishbone diagram · Intentional attack · Safety · Security · Technical failure

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 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 paper is structured as follows: Section 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 3. In Section 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 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 5. Section 6 presents the conclusions and future work directions.

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 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 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 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.

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.

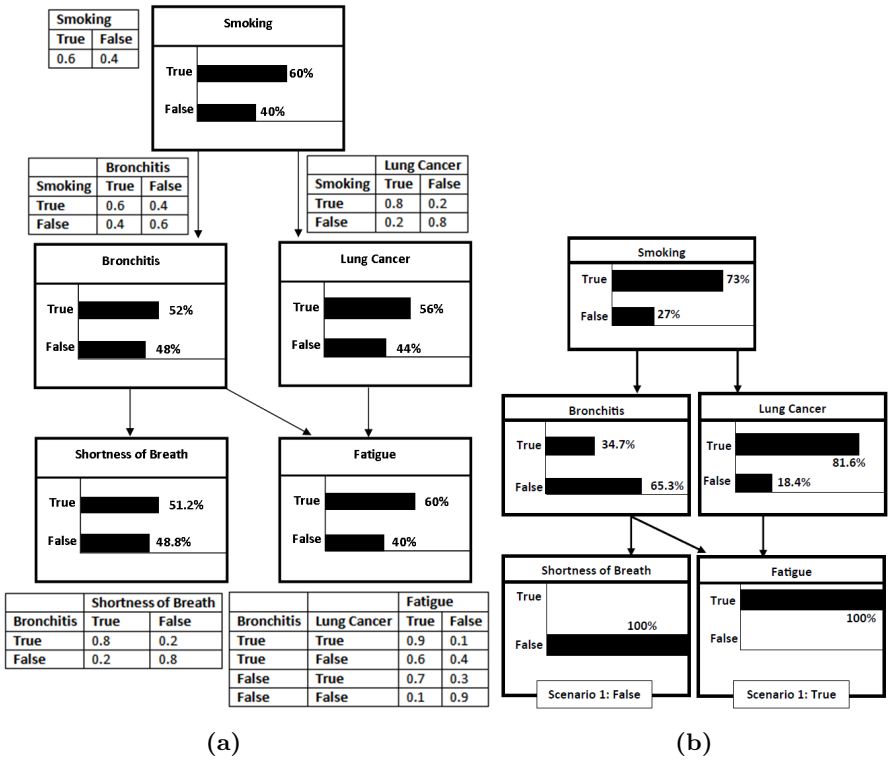

Fig. 1. (a) A Typical BN Model for Disease Diagnosis. (b) Updated Probabilities Given Observed Symptoms (Evidence).

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.

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 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].

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.

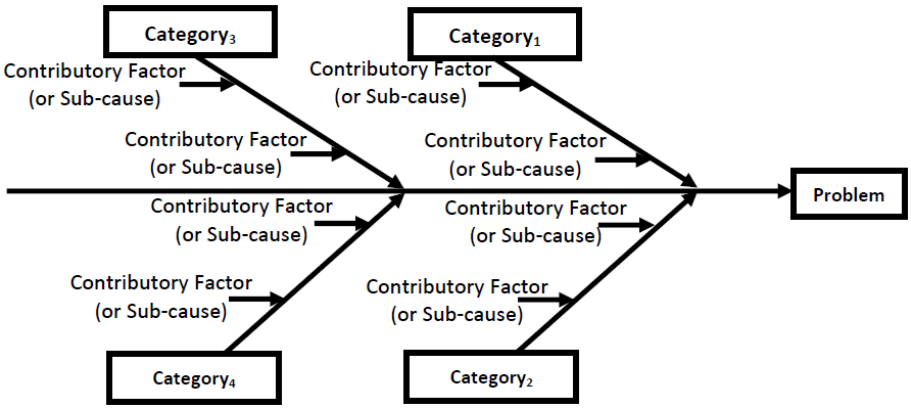

Fig. 2. Generic Fishbone Diagram Structure

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.

4.1 ICS Architecture

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 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].

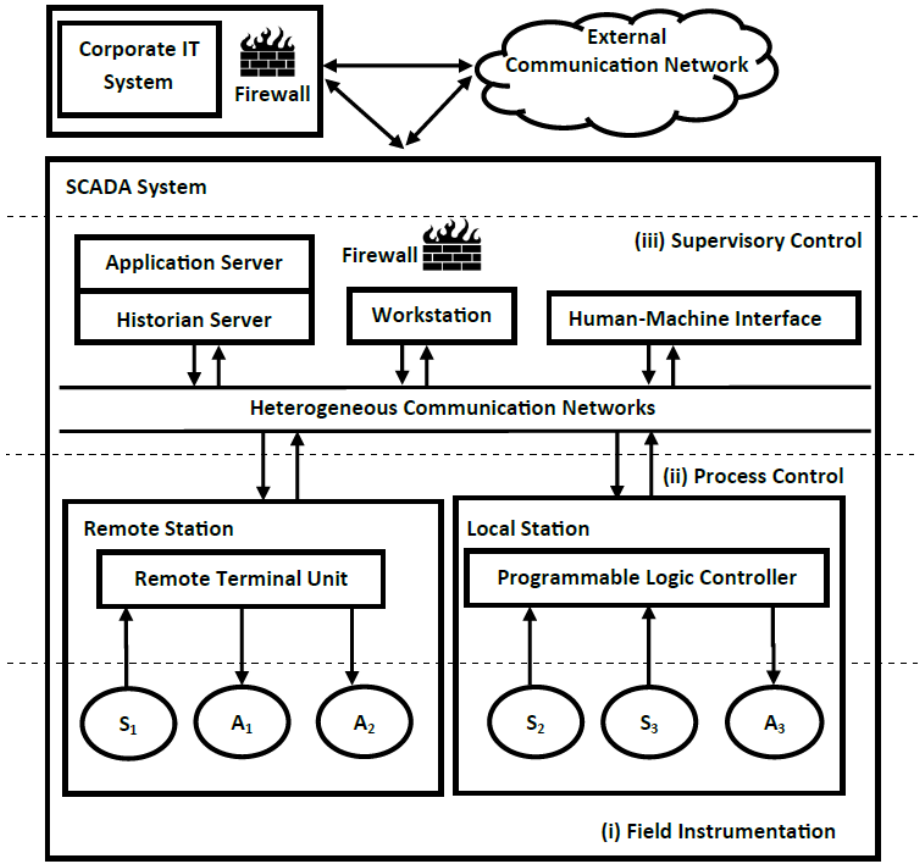

Fig. 3. Typical ICS Architecture and Layers

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 schematises a floodgate being primarily operated by Supervisory Control and Data Acquisition (SCADA) system along with an operations centre.

Figure 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. The data transmission used in this case is wired. Electricity is the only energy source in the operations centre.

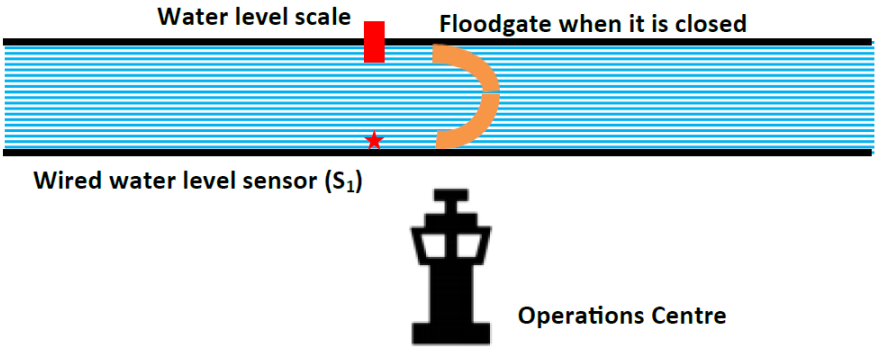

Fig. 4. Physical Layout of the Floodgate

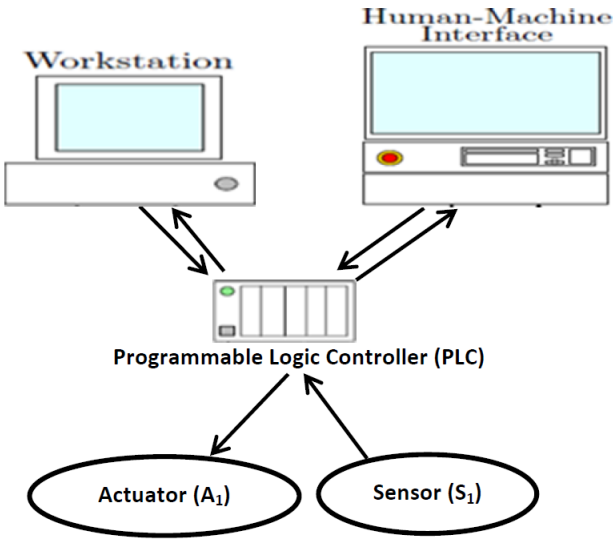

Fig. 5. SCADA Architecture of the Floodgate

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.

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 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".

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 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).

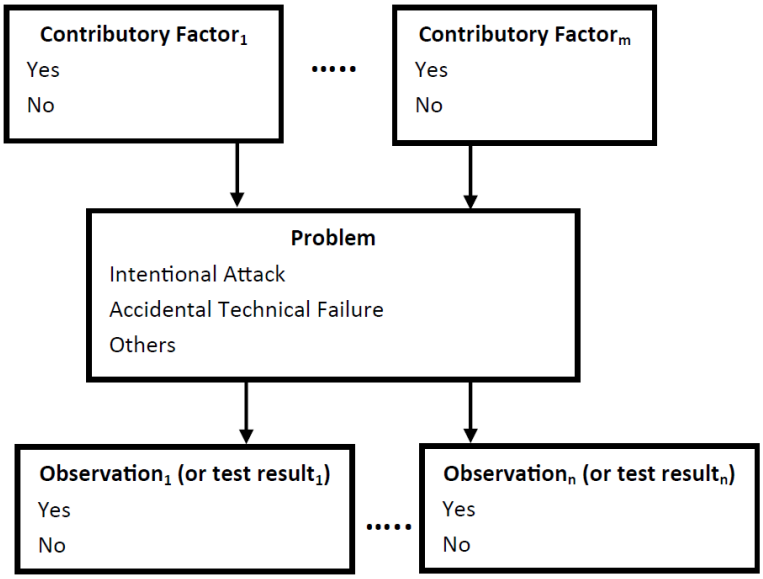

Fig. 6. BN Structure to Determine the Major Cause of an Abnormal Behavior in a Component of the ICS

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 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.

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 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.

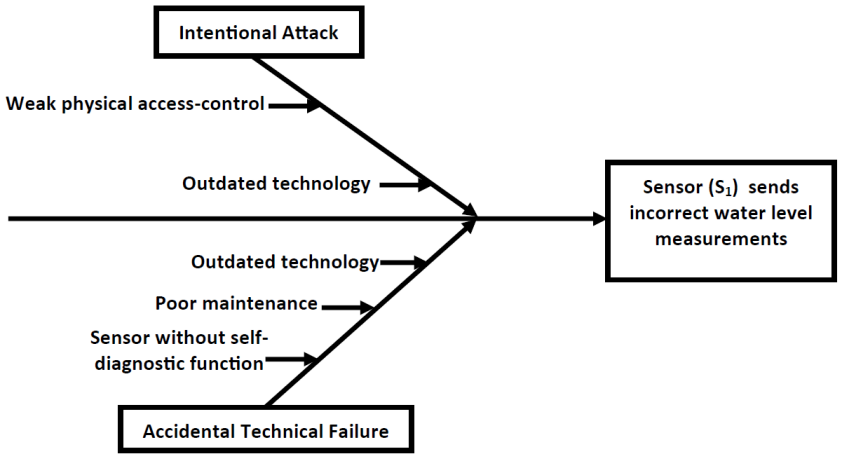

Fig. 7. Fishbone Diagram Example

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 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 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 9 shows the extended version of our fishbone diagram example with observations (or test results).

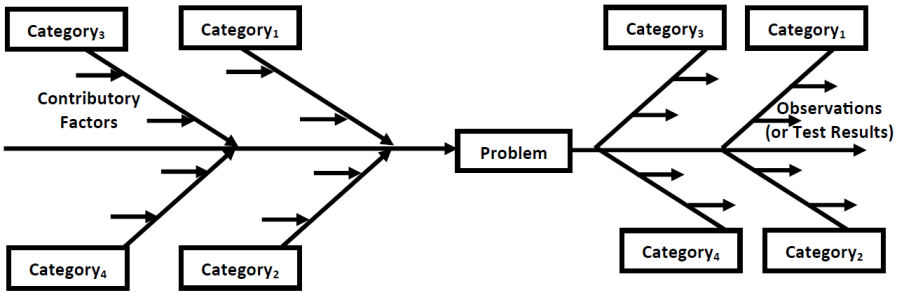

Fig. 8. Extended Fishbone Diagram Structure

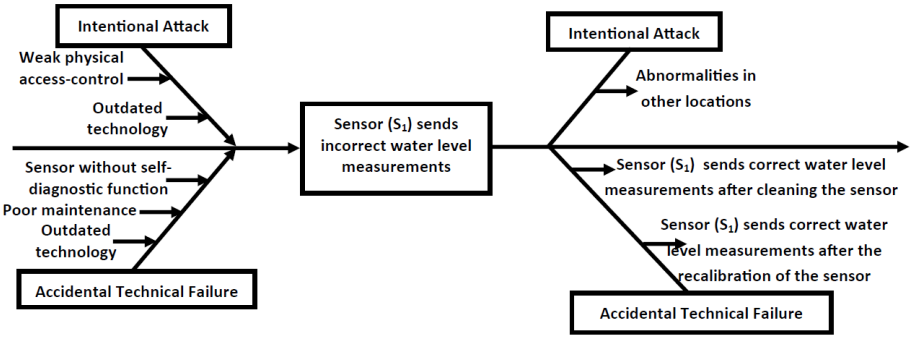

Fig. 9. Extended Fishbone Diagram Example

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, based 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.

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

i. The considered problem in the fishbone diagram is mapped to the problem variable in the middle layer of the BN as shown in Figure 10.
ii. 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 5.1, this would be determined in case the major cause of the problem is neither intentional attack nor accidental technical failure.
iii. 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 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 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 10.

iv. 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 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 10.

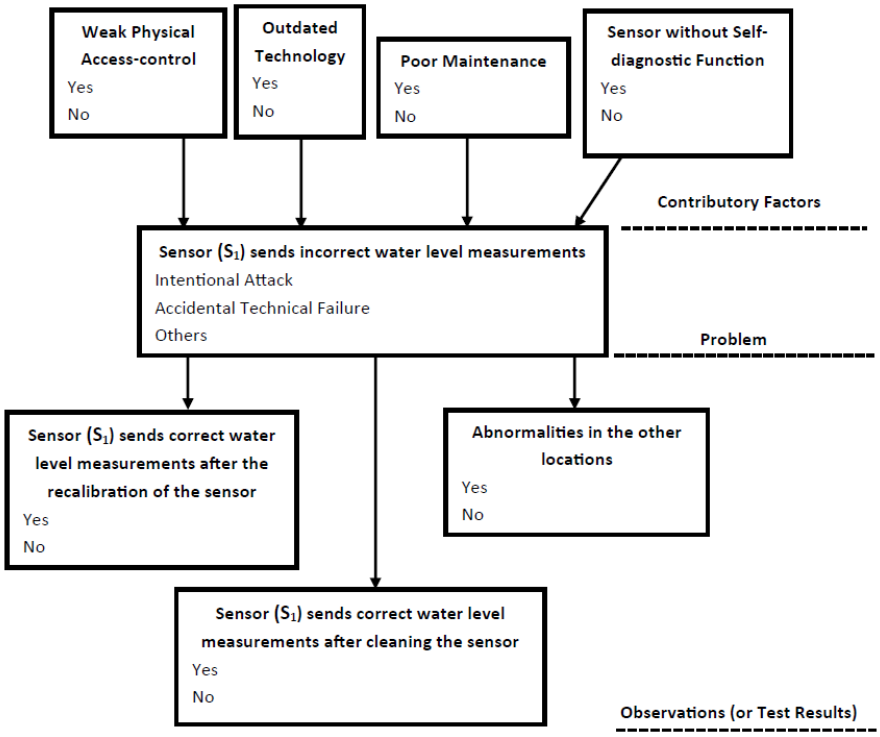

Fig. 10. Translated BN from Fishbone Diagram Example

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.

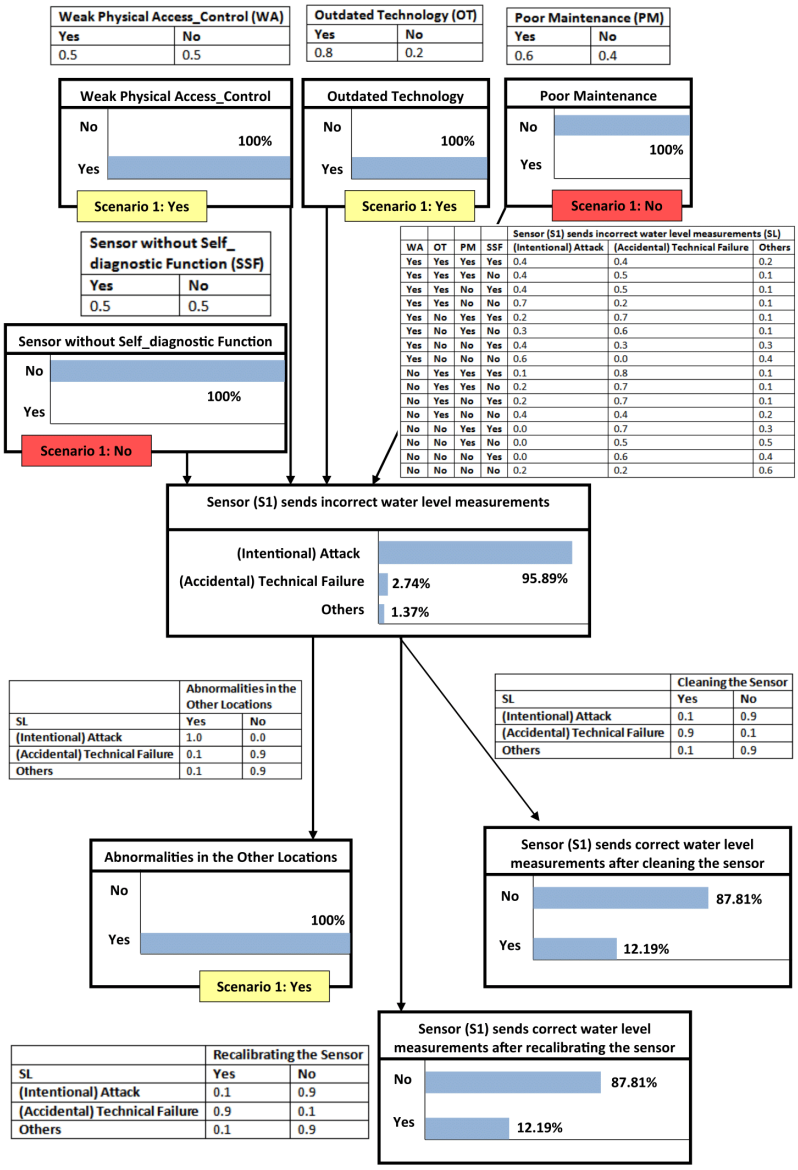

Fig. 11. Translated BN with Updated Probabilities Based on the Evidence

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 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 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.

6 Conclusions and Future Work

Adequate decision support for distinguishing intentional attacks and accidental technical failures is missing. In this paper, 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.

Acknowledgements

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)"*.

References

1. 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 9, 52-80 (2015)
2. RISI Database.: German Steel Mill Cyber Attack (2018). `http://www.risidata.com/database/detail/german-steel-mill-cyber-attack`
3. Macaulay, T., Singer, B.L.: Cybersecurity for Industrial Control Systems: SCADA, DCS, PLC, HMI, and SIS. CRC Press (2011)
4. Kaspersky Lab.: Five Myths of Industrial Control Systems Security (2014). `https://media.kaspersky.com/pdf/DataSheet_KESB_5Myths-ICSS_Eng_WEB.pdf`
5. Korb, K.B., Nicholson, A.E.: Bayesian Artificial Intelligence. 2003. Florida: Chapman & Hall/CRC
6. Nikovski, D.: Constructing Bayesian Networks for Medical Diagnosis from Incomplete and Partially Correct Statistics. IEEE Transactions on Knowledge and Data Engineering 12, 509-516 (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. The Quality Management Journal 12, 34 (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. In: International Conference on Computer and Computing Technologies in Agriculture, pp. 917-924. Springer, (2007)
12. Pecchia, A., Sharma, A., Kalbarczyk, Z., Cotroneo, D., Iyer, R.K.: Identifying Compromised Users in Shared Computing Infrastructures: A Data-driven Bayesian Network Approach. In: Reliable Distributed Systems (SRDS), 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, ACM, (2010)

15. Kwan, M., Chow, K.-P., Lai, P., Law, F., Tse, H.: 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.J., Wasyluk, H.: Extension of the Hepar II Model to Multiple-Disorder Diagnosis. Intelligent Information Systems, pp. 303-313. Springer (2000)
17. Curiac, D.-I., Vasile, G., Banias, O., Volosencu, C., Albu, A.: Bayesian network model for diagnosis of psychiatric diseases. In: Information Technology Interfaces, 31st International Conference on, pp. 61-66. IEEE, (2009)
18. Estabragh, Z.S., Kashani, M.M.R., Moghaddam, F.J., Sari, S., Taherifar, Z., Moosavy, S.M., Oskooyee, K.S.: Bayesian network modeling for diagnosis of social anxiety using some cognitive-behavioral factors. Network Modeling Analysis in Health Informatics and Bioinformatics 2, 257-265 (2013)
19. González-López, J., García-Aparicio, Á.M., Sánchez-Ponce, D., Muñoz-Sanz, N., Fernandez-Ledo, N., Beneyto, P., Westcott, M.: Development and validation of a Bayesian network for the differential diagnosis of anterior uveitis. Eye 30, 865-872 (2016)
20. Moreira, M.W.L., Rodrigues, J.J.P.C., Oliveira, A.M.B., Ramos, R.F., Saleem, K.: A preeclampsia diagnosis approach using Bayesian networks. In: Communications (ICC), IEEE International Conference on, 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 27, 19-29 (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 54, 115-126 (1999)
23. Huang, Y., McMurran, R., Dhadyalla, G., Jones, R.P.: Probability based vehicle fault diagnosis: Bayesian network method. Journal of Intelligent Manufacturing 19, 301-311 (2008)
24. Liu, J.H., Zhang, J.L., Jin, M.D.: Application of BN in the Fault Diagnosis of Brake Failure System. In: Applied Mechanics and Materials, pp. 1684-1688. Trans Tech Publications, (2014)
25. Kipersztok, O., Dildy, G.A.: Evidence-based Bayesian networks approach to airplane maintenance. In: Neural Networks, Proceedings of the International Joint Conference on, 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 2, pp. 1-20 (2010)
27. Ishikawa, K.: Guide to Quality Control (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 78, pp. 51-64 (2013)
29. 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 11, pp. 1035-1041 (2004)
30. Asllani, A., Ali, A.: Securing information systems in airports: A practical approach. In: Internet Technology and Secured Transactions (ICITST), International Conference for, pp. 314-318. 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. In: Applied Mechanics and Materials, pp. 139-143. Trans Tech Publications, (2012)
32. Luca, L., Stancioiu, A.: The study applying a quality management tool to identify the causes of a defect in an automotive. In: Proceedings of the 3^{rd} International Conference on Automotive and Transportation Systems. Montreux, Elvetia. (2015)
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 33, pp. 289-303 (2015)
34. Endi, M., Elhalwagy, Y., Hashad., A.: Three-layer PLC/SCADA system architecture in process automation and data monitoring. In: Computer and Automation Engineering (ICCAE), The 2nd International Conference on, 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, Å., Jacobsson, P., Thedéen, T.: Risks in Technological Systems. Springer Science & Business Media (2009)
37. Global Water Level Sensor: WL400 Product Manual. `http://www.globalw.com/downloads/WL400/WL400manual.pdf` (2009)
38. Przytula, K.W., Thompson, D.: Construction of Bayesian Networks for Diagnostics. In: Aerospace Conference Proceedings, pp. 193-200. IEEE, (2000)
39. Henrion, M.: Practical issues in constructing a Bayes' belief network. In: Proceedings of the Third Conference on Uncertainty in Artificial Intelligence, pp. 132-139 (1987)
40. 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)
41. de Ruijter, A., Guldenmund, F.: The Bowtie Method: A Review. Safety Science 88, pp. 211-218 (2016)
42. 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. In: International Conference on Critical Information Infrastructures Security, pp. 50-62. Springer, (2016)