IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 31, 2017, accepted June 5, 2017, date of publication June 22, 2017, date of current version September 19, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2718498

Evidential Network Modeling for Cyber-Physical System State Inference

IVO FRIEDBERG[1,2], XIN HONG[1], KIERAN MCLAUGHLIN[1], PAUL SMITH[2], AND PAUL C. MILLER[1]
[1]CSIT, Queen's University Belfast, Belfast BT3 9DT, U.K.
[2]AIT Austrian Institute of Technology, 1220 Vienna, Austria

Corresponding author: Ivo Friedberg (ifriedberg01@qub.ac.uk)

This work was supported in part by EPSRC under Grant EP/N022866/1 and in part by the EU FP7 SPARKS Project under Grant 608224.

ABSTRACT Cyber-physical systems (CPSs) have dependability requirements that are associated with controlling a physical process. Cyber-attacks can result in those requirements not being met. Consequently, it is important to monitor a CPS in order to identify deviations from normal operation. A major challenge is inferring the cause of these deviations in a trustworthy manner. This is necessary to support the implementation of correct and timely control decisions, in order to mitigate cyber-attacks and other causes of reduced dependability. This paper presents evidential networks as a solution to this problem. Through the evaluation of a representative use case for cyber-physical control systems, this paper shows novel approaches to integrate low-level sensors of different types, in particular those for cyber-attack detection, and reliabilities into evidential networks. The results presented indicate that evidential networks can identify system states with an accuracy that is comparable to approaches that use classical Bayesian probabilities to describe causality. However, in addition, evidential networks provide information about the uncertainty of a derived system state, which is a significant benefit, as it can be used to build trust in the results of automatic reasoning systems.

INDEX TERMS Communication technologies, computer networks, network security, data system, data processing, data integration, industry applications, security, security management, power engineering and energy, power systems, microgrid, attack causality.

I. INTRODUCTION

As cyber-physical systems (CPS), such as the smart grid, rely increasingly on information and communication technology (ICT) the effects of cyber-attacks can directly influence the operation of physical processes. One severe threat is the unauthorized and undetected manipulation of sensor measurements. A successful attack can cause incorrect decisions by operators or control algorithms, if the estimated system state is based on manipulated measurement values and is therefore incorrect [1]. Similarly, an attacker could also inject malicious control commands that cause unexpected behaviour in the physical part of the system [2]. Control decisions in CPS can be very time sensitive [3]. Therefore, monitoring data needs to be analyzed automatically to support the decision making process by providing state awareness [4]. State awareness requires a holistic approach, in the sense that malicious, erroneous and normal system states are considered of equal importance. This work shows that evidential networks are a solution that can infer system states accurately.

An evidential network (EN) [5] is a graph structure that encodes knowledge about variables in a system and the relationship between these variables. The information is encoded in belief structures based on Dempster-Shafer Theory [6]. Previous approaches in the cyber-security domain that leverage data from multiple sensors focus on the detection of malicious behavior. The correct classification of normal or erroneous behavior is of limited relevance. Alert correlation [7] is one such approach that focuses on the reduction of false positives and the correlation of multiple alerts about single attacks that form one multi-stage attack. In this work, it is proposed that the reasoning unit that provides state awareness needs to answer a question about causality. Given a set of (correlated or not) sensor information, the system needs to identify the system state that is caused by the underlying events. In return, this requires an approach that allows the integration of sensor evidence of various types of sensors. Intrusion detection systems (IDS) and other cyber-security sensors have to be analyzed and combined with sensor information from the physical domain, in order to distinguish different states.

Another problem is that different types of sensors not only provide different data, they also operate with different reliability. This introduces uncertainty to the data provided

by sensors; an aspect that is usually not considered in existing correlation approaches, but is critical for inference. Evidential networks are explicitly designed to handle uncertainty. Further, they have been successfully applied in the fields of threat assessment (see Benavoli *et al.* [8]), system reliability evaluation (see Simon and Weber [9]) or activity recognition in smart homes (see Hong *et al.* [10]). This use in different application areas shows the general applicability of evidential networks. However, only limited work was dedicated to the use of evidential networks in the field of cyber-security and state inference in cyber-physical systems [11]–[13].

This work discusses the use of evidential networks to accurately infer all types of system states (not only cyber-attacks) by reasoning about sensor evidence from the cyber-security domain and the physical system. This is a necessity for operators and control algorithms that ensure dependable operation under the threat of cyber-attacks in CPS. The contributions of this work are:

- The presentation of evidential networks as a solution to the problem of state inference. Compared to existing solutions, our approach does not prioritize the detection of malicious behavior, but considers normal and erroneous states of equal importance.
- A detailed analysis about the accurate integration of sensor evidence from different types of sensors, given a-priori knowledge about the sensor's reliability.
- The implementation of an evidential network for state inference in a real-world smart grid use-case that is representative of similar cyber-physical systems. This shows the applicability of evidential networks to real world problems more clearly than related work [11], [14], where abstract attack vectors are analyzed.
- A comparative evaluation of the proposed EN and considerations presented by Zomlot *et al.* [11] and Ou *et al.* [15]. The results show that some claims in the related work lead to high false positive rates and inaccurate detection when multiple system states are considered; something that this work improves upon.
- An evaluation of evidential networks compared to widely-used Bayesian network approaches, which shows that erroneous states can be detected much more accurately with the use of evidential networks. At the same time, evidential networks provide additional information about the level of trust that should be placed in the results through the remaining uncertainty.

II. RELATED WORK

To overcome the problems that arise with the number of low level alerts generated by traditional intrusion detection systems [16]–[18] alert correlation [7], [19], [20] is proposed. The goals of alert correlation are (i) to identify the alerts that can be filtered out, (ii) to group alerts to make them easier to analyze and (iii) to prioritize these groups of alerts to minimize the response time to the most critical issues [7]. A problem with correlation is that it cannot provide information about the causality between the sensor evidence (e.g. alerts, physical sensor readings, etc.) and the current system state. The question asked by operators however is, what each piece of evidence implies about the system state. Another shortcoming of many alert correlation approaches is that the type of information that is correlated is limited. Zhai *et al.* [21] identified this problem and presented a reasoning approach based on complementary intrusion evidence. Their work combines intrusion evidence from different sources (e.g. malware scanners, host information and network data), to reason about the progress of a cyber attack. The presented reasoning approach adopts the concepts of state-based and event-based evidence; an attack requires the system to be in a certain state which can be an indicator for previously missed event-based evidence and vice versa. For the security of cyber-physical systems, the concept of complementary evidence needs to be extended to include evidence from the physical domain. Another shortcoming of the approach is the focus on attack states. In cyber-physical systems, certain alerts can be caused by erroneous behavior (e.g. a component fault). Consequently, the exclusive consideration of intrusion detection systems and data about malicious behavior results in a limited ability to differentiate between system errors and malicious actions.

Squicciarini *et al.* [14] aim to achieve situational awareness by reasoning about network incidents with a tool called ReasONets. ReasONets aims to detect incidents with the use of machine learning and anomaly detection after which a case based reasoning unit tries to infer the indicated system state. The reasoning unit leverages Fuzzy Logic Theory to handle uncertainty in the system. The approach limits its potential with the focus on a self designed anomaly detection component as the only source of information about the monitored system. Anomaly detection approaches produce high false positive rates and cannot match the accuracy of signature based detection mechanisms when it comes to known attacks [16]. Given the amount intrusion detection solutions that are already deployed and used successfully, event inference needs to be able to integrate existing solutions to improve acceptance. Considerations about the need for uncertainty in the context of intrusion detection were also investigated by Ou *et al.* [15]. Their work was later picked up by Zomlot *et al.* [11] who apply Dempster-Schafer theory to alert correlation which is highly relevant in the context of this work. Their approach introduces many thoughts on the applicability of DS theory in the context of cyber-attacks. These concepts will be thoroughly discussed and evaluated in more detail throughout this work.

Dempster-Schafer theory in general, and evidential networks specifically are very suitable for reasoning about complementary evidence from sensors in different domains. This was recently shown in the context of smart homes [10], video surveillance [22], railway risk assessment [23], hazardous material transportation [24], and threat assessment [8], with promising results. This work shows that evidential networks can be used to reason about the causality between diverse

low level evidence and high level system states, with high accuracy. The novelty lies in the fact that uncertainty about the low level evidence is taken into account and that not only malicious but also erroneous and normal system states are considered.

III. PROBLEM STATEMENT

This section introduces a physical use-case around the remote control of photovoltaic inverters, followed by a description of the ICT network and the sensors that are the source of evidence. From this, a threat scenario is derived that is later used together with the use-case in the experimental evaluation of the proposed EN approach.

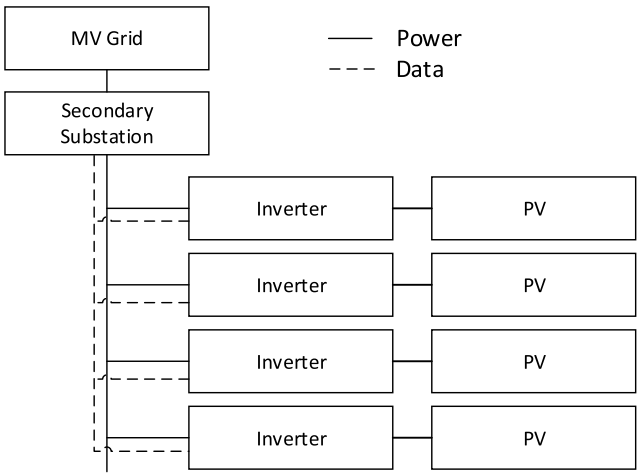

FIGURE 1. Overview of PV clusters on a distribution line.

A. PHOTOVOLTAIC USE-CASE

In this work a smart grid scenario based around the control of photovoltaic (PV) installations is considered. In a medium or low voltage distribution grid, PV clusters are placed along a distribution line. This line is usually connected to the main grid through a secondary substation that manages the line's voltage levels. Figure 1 shows the abstract system structure. Each PV cluster is connected to the distribution line by a PV inverter. A PV inverter is responsible for converting the direct current (DC) produced by the PV cells into alternating current (AC) that can be fed into the grid infrastructure. This conversion is performed based on a number of set-points that control different aspects of the outgoing AC. These set-points can be used to optimize the overall power output of the PV cluster but also to stabilize the grid. Since the PV inverters themselves are only aware of the local state of the grid and not the overall status, an external controller can be used to remotely change the set-points of the PV inverters based on its wider system view. This concept of a control-loop is one argument why the use-case at hand is representative for many cyber-physical systems which are operated with the same type of control. The integrity of this communication is critical as manipulated set-points can have undesired effects. One example of such an effect is the fact that safety restrictions imposed by the PV inverter force it to shutdown if the active power output is below 10%. In work by Kang *et al.* [2], [25] the authors analyzed how manipulations of these commands can be achieved and how they effect PV installations. This makes the use-case very appropriate for the evaluation of ENs because the case that set-points $< 10\%$ are sent to the controller can arise due to various reasons; both erroneous or malicious. This shows another aspect of the use-case that can be found in cyber-physical systems in general. In contrast to purely digital systems, physical components are more prone to error states that need to be handled differently than malicious actions. However, error states and malicious states can manifest in similar raw evidence which makes it challenging for a state inference system to identify the correct causal relationships. This complexity is usually not considered in related work where the focus is on the detection of a set of cyber-attacks.

In order to ensure stability on a specific distribution line, it is crucial to identify the current system state correctly. For this scenario, four high level states are identified which need to be accurately detected.

1) *Normal:* All components are working as expected.
2) *Controller Error:* The controller in the substation issues erroneous commands. This can be caused by human error or by an arbitrary fault in the control system. However, there is no malicious intent.
3) *Controller Malicious:* The controller is compromised and its behavior is part of a malicious agenda.
4) *Control Communication Manipulated:* The controller is behaving as expected, but the control commands to the inverter are manipulated in the communication network. This discrepancy is part of a malicious agenda.

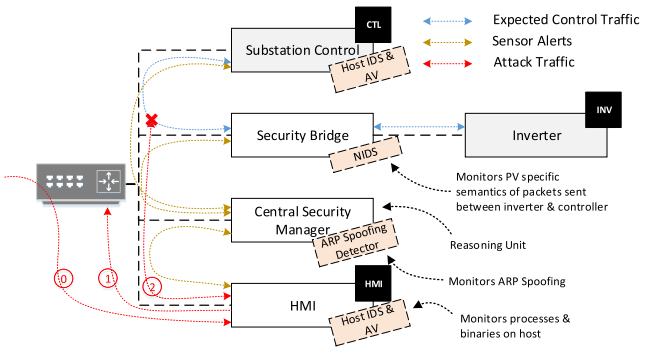

FIGURE 2. PV inverter control system. (*Grey boxes represent components from the physical domain; white boxes represent components in the ICT domain. Dashed boxes highlight the sensors in the network and their location. A legend about the different communication channels in the network is given in the top right corner.*)

B. SUBSTATION ICT NETWORK

The evaluation focuses on the substation controller and its communication to a single PV inverter. Figure 2 shows the network setup for the testcase. There are five network

nodes on the subnet. The two grey boxes represent systems that relate to the physical domain; namely the *Substation Control* (*CTL*) and the *PV Inverter* (*INV*). The inverter is connected to the network through a *Security Gateway*. By separating critical, and often legacy equipment from direct network access, the bridge can *(i)* monitor and potentially intercept all traffic from and to the device and *(ii)* improve the interfaces that the device offers by providing security features that cannot be introduced on the device itself [26].

Further, there is a *Central Security Manager* and a human machine interface in the network (*HMI*). The security manager node is usually connected to a mirroring port on the switch interfacing the local subnet with outside networks to monitor all incoming and outgoing traffic. For visibility of traffic within the subnet it relies on alerts triggered by other sensors installed on the subnet. In this work we assume that these aspects cannot be attacked directly.

C. THREAT SCENARIO

For the adversary model, it is assumed that the HMI exposes a vulnerability that is exploited by an attacker. As a consequence, the adversary gains a foothold in the network which is then used to perform reconnaissance. Subsequently *INV* and *CTL* are identified as PV inverter and controller respectively. The adversary then aims to shut down the PV inverter by sending it an erroneous packet with a control command to regulate the active power output to a value of $< 10\%$. Internal safety mechanisms will register this command as unsafe and perform a shutdown of the inverter.

To get the malicious control command accepted two options are available to the attacker. First, the weaponized code can aim to hijack the existing communication link between controller and PV inverter by performing an ARP spoofing attack to redirect the traffic. It will then send spoofed packets and block all original traffic from the controller. This is done to circumvent any IP based whitelisting approaches that prevent unauthorised control commands to reach the PV inverter [2].

Alternatively, the adversary can identify a further vulnerability in the substation controller. By infection of the controller, the adversary gains full control over the PV inverter.

D. LOW LEVEL SENSORS

In order to detect security incidents it is assumed that general purpose cyber-security solutions and mechanisms are deployed in the substation network. Hosts *HMI* and *CTL* can be considered general purpose machines. They are both monitored by a separate host based IDS (*HIDS*) and a malware detection unit (*AV*). Furthermore, system wide sensors comprise a sensor monitoring network traffic for potential ARP spoofing and a network IDS (*NIDS*) that monitors traffic from and to the inverter for unexpected set-points [25]. Each sensor can in general issue multiple types of alerts. To reason about the information provided, each potential alert needs to be handled separately. For example, if the installed network IDS monitors traffic using n rules, the derived reasoning network will contain n logical sensors where each sensor signals that a specific alert was raised (see also [11]).

IV. STATE INFERENCE METHODOLOGY

This work proposes to apply the concept of evidential networks to identify the causality between low level sensor alerts and higher level system states. Evidential networks are a special form of valuation algebra [5] based on the Dempster-Shafer (DS) theory of evidence [6]. This section will give a brief overview of the minimal mathematical concepts required to understand how state inference is performed.

A. DEMPSTER-SHAFER THEORY

Dempster-Shafer (DS) theory is a mathematical theory of evidence that constructs a coherent picture of reality through computing the degree of belief on an event, given evidence [6]. It does so by abstracting the represented system with information on a set of variables $V = \{x_1, \ldots, x_n\}$. Each variable represents either the possible states reported by a low level sensor or higher level information inferred about the system. It is further defined by the following concepts:

B. FRAME OF DISCERNMENT

The frame of discernment, denoted by Θ, is a set of mutually exclusive and exhaustive hypotheses about a problem domain (domain for the rest of this work). A domain in DS theory is described by a set of variables which in return describe aspects of the modeled system. Consider the problem described by variable x. $D_x = \{x\}$ represents the domain of this problem. If $\Theta_{D_x} = \{s : s \text{ is the value of } x\}$ is the frame of discernment for D_x (Θ_x in short), then the frame of discernment Θ_D of an arbitrary domain D is given as:

$$\Theta_D = \times\{\Theta_x : x \in D\} \tag{1}$$

C. MASS FUNCTION

A mapping $m : 2^\Theta \rightarrow [0, 1]$ is called mass function satisfying:

$$\sum_{A \subseteq \Theta} m(A) = 1 \quad \text{and } m(\emptyset) = 0. \tag{2}$$

A mass function is seen as a generalized probability function, defined on the power set of Θ rather than on Θ. Described over a specific domain, denoted as $d(m)$, it encodes arbitrary information about that domain. This information can be sensor evidence as well as information about the relationship between the variables in the domain. A mass function provides a richer description than a classic probability function as mass values are on subsets of Θ. Therefore, it has the capability to represent uncertainty, by assigning a part of the probability to a non-singular subset of Θ.

D. DEMPSTER'S RULE OF COMBINATION

Dempster's rule provides a mechanism to aggregate the evidence from multiple independent sources. Let m_i be the mass

function collected from the ith source over the frame Θ. Dempster's rule is given as:

$$m(C) = (m_1 \oplus m_2 \oplus \ldots \oplus m_n) = \frac{1}{K} \sum_{(C_1 \cap C_2 \cap \ldots \cap C_n) = C} m_1(C_1) \cdot m_2(C_2) \cdot \ldots \cdot m_n(C_n) \tag{3}$$

where K is the normalizing constant that is defined as:

$$K = 1 - \sum_{(C_1 \cap C_2 \cap \ldots \cap C_n) = \emptyset} m_1(C_1) \cdot m_2(C_2) \cdot \ldots \cdot m_n(C_n).$$

E. EVIDENTIAL NETWORKS

An evidential network (EN) [5] is a framework for knowledge representation and inference, by using DS theory. An EN models a real-world problem in a network of interlinked variables. An **evidential network** is represented by a tuple:

$$EN = \{V, \Theta_V, M_V, \oplus, \downarrow\} \tag{4}$$

where:

- $V = \{x_1, \ldots, x_n\}$ is the set of variables in the model;
- $\Theta_V = \{\Theta_x : x \in V\}$ is the set of frames of all variables;
- $M_V = \cup\{M_D : D \subseteq V\}$ is the set of all mass functions;
- $\oplus$ is the combination operator;
- $\downarrow$ is the marginalisation operator.

The joint mass function of an evidential network, denoted by $\oplus M$, is the combination of all mass functions in the network. It is computed to combine the low level evidence with the knowledge about the variable relationships to infer knowledge about higher level system states. The domain of $\oplus M$ is the union of the domains of all M_D, $d(\oplus M) = \cup d(M_D) = V$. The frame of discernment for $\oplus M_D$, denoted as Θ, is the Cartesian product of all Θ_D, $\Theta = \times \Theta_D$. Suppose $D^0 \subseteq V$ is the domain of our interest (i.e. the variables that denote the system states of interest). We extract the information of interest by computing $(\oplus M)^{\downarrow D^0}$ through the evidential operations described in the next subsection.

F. EVIDENTIAL OPERATIONS

To compute the joint mass function, two evidential operations: vacuous extension and marginalisation [27] are applied. Let D and D' be two domains, $D' \subseteq D$. Θ_D and $\Theta_{D'}$ represent the frame of discernment for D and D' respectively.

Vacuous extension of a mass function defined on domain D', $m_{D'}$, to domain D is defined as:

$$m_{D'}^{\uparrow D}(A) = \begin{cases} m_{D'}(B) & \text{if } A = B \times \Theta_{D \setminus D'}; \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

where $D \setminus D'$ represents the complement of D' in D.

Marginalisation is a projection of a mass function defined on D, m_D, into domain D':

$$m_D^{\downarrow D'}(B) = \sum_{A \subseteq B \times \Theta_{D \setminus D'}} m_D(A). \tag{6}$$

To describe causality between variables, Domain Experts usually give their subjective judgments in a form of IF-THEN rule, such as "if A then B". When uncertain knowledge is involved, degrees of confidence measuring uncertainty have to be attached to knowledge rules, such as "if A then B" with a certain degree of confidence $\rho \in [\alpha, \beta]$, $0 \leq \alpha \leq \beta \leq 1$. α and β represent the minimum and maximum degree to which a relationship is thought to hold. Such a rule is called **relation implication rule** since it represents a relation between conditions and a consequence. A relation implication rule with uncertainty measures can be conceptually represented within the framework of DS theory. Assume that D_A and D_B are two disjoint domains associated with frames Θ_{D_A} and Θ_{D_B} respectively, and $\subseteq \Theta_{D_A}$, $B \subseteq \Theta_{D_B}$. A relation implication rule can be then written as:

$$A \subseteq \Theta_{D_A} \Rightarrow B \subseteq \Theta_{D_B} \text{with } \rho \in [\alpha, \beta], 0 \leq \alpha \leq \beta \leq 1.$$

Using the principle of minimum commitment and the ballooning extension mechanism [28], [29] the above rule can be represented by a mass function over the product space $\Theta_{D_C} = \Theta_{D_B} \times \Theta_{D_A}$ on domain $D_C = D_B \cup D_A$:

$$m_{D_C}(C) = \begin{cases} \alpha & \text{if } C = (B \times A) \cup (\Theta_{D_B} \times A^c); \\ 1 - \beta & \text{if } C = (B^c \times A) \cup (\Theta_{D_B} \times A^c); \\ \beta - \alpha & \text{if } C = \Theta_{D_B} \times \Theta_{D_A}. \end{cases} \tag{7}$$

where A^c represents the complement of A in Θ_{D_A}, B^c is the complement of B in Θ_{D_B}.

G. DECISION MAKING

Belief functions cannot be directly used for decision making [30]. Mass functions have to be transformed into a pignistic probability distribution; its counterpart in the classical probability theory. Let m_D be a mass function defined on a subset of variables D with corresponding frame Θ_D. The pignistic transformation of m_D, called the pignistic probability, is defined for every element of the frame $\theta \in \Theta_D$ as follows [30]:

$$BetP(\theta) = \sum_{\theta \in A \subseteq \Theta_D} \frac{m_D(A)}{|A|}. \tag{8}$$

where $|A|$ stands for the total number of the elements in A. $BetP$ is the DS counterpart of the subjective probability that would quantify the human's beliefs in classical Bayesian probabilities.

V. EVIDENCE FUSION FOR STATE INFERENCE

Although the mathematical concepts of evidential networks and DS theory are well researched, limited work can be found that analyzes the challenges in applying the framework to concrete problems in cyber-physical systems. During this work two main challenges were identified and are subsequently highlighted. First, the design of relation implication rules needs to be supported to reduce the complexity and subsequently error probability in the design phase. Some approaches to this challenge exist to date and we will validate

their correctness through rigorous tests. The second challenge involves the interpretation of sensor evidence in this specific context. This work will show that different types of sensors need to be handled differently. Further, a-priori knowledge about the performance and trustworthiness of sensors needs to be handled with care and differently depending on the format of the knowledge.

TABLE 1. Linguistic scales for mapping design. (*A mapping in the 8-element scale can translated into the other scales by merging two or three of the dividing six scale elements.*).

8-Element Scale	5-Element Scale	4-Element Scale
Probable (99%)	Probable (99%)	Probable (99%)
Very Likely (85%)	Likely (74.5%)	Likely (0.67%)
Likely (71%)		
Possible (57%)	Possible (50%)	
Potentially (43%)		
Feasible (29%)	Feasible (25.5%)	Possible (0.33%)
Improbable (15%)		
Unlikely (1%)	Unlikely (1%)	Unlikely (1%)

A. RELATIONSHIP MODELING

Relation implication rules are designed based on expert knowledge. They represent the causal relationships between different sets of variables. Each variable either represents low level sensor evidence or the states of higher level system aspects. This rule design process involves two steps. First, the relevant relationships need to be identified (define A and B in Eq. IV-F); secondly, a level of belief into the represented causality needs to be specified (define α and β in Eq. IV-F). The first step is well explained in literature, and attack trees [31] or safety analysis techniques like FMEA [32] and STPA [33] provide repeatable processes that can be used. For the second step however, no well established approach exists. Expert knowledge needs to be formalized and Ou *et al.* [15] propose the use of a scale to support the human rational. They argue that α and β should be chosen from only four fixed elements. However, no evaluation exists to show whether the reduction of the probability range to four discrete steps limits the expressiveness of the rules and subsequently the performance of the evidential network. This work therefore aims to evaluate the impact of the size of the scale on the performance of the EN. Table 1 presents 3 different scales that divide the mapping space into a different number of equal portions. Rules that make use of 8 elements, can be transformed to smaller scales as seen in Tab. 1. For example, the certainty of *Improbable* is 15% in the 8-element scale, 25.5% when mapped to the 5-element scale or 33.3% in the 4-element scale. The edge cases (*Probable* and *Unlikely* stay constant as their impact was already evaluated by Zomlot *et al.* [11]. Later, Section VII will evaluate how different scales will effect the performance of the reasoning unit.

B. SENSOR EVIDENCE INTEGRATION

One critical task is the integration of sensor evidence into the evidential network. A sensor S in our problem space can be defined as a mapping from time t to the set of all possible mass functions m over a domain of a single variable $D = \{x\}$. At any given time, the sensor provides knowledge about a specific variable in the form of a mass function. During the design of the evidential network, three questions have to be answered about a sensor:

1) What type of information does the sensor provide?
2) How is the information about sensor reliability structured?
3) What is the *reliability* of the modeled sensor?

Reliability is defined by the IEEE 24765 Standard on Systems and Software Engineering Vocabulary [34] as *the ability of a system or component to perform its required functions under stated conditions for a specified period of time*. The reliability of a sensor is a measure of the sensor's ability to interpret the monitored aspect of the system correctly. A sensor is denoted as a single variable in DS theory. It therefore needs to provide information about a finite set of states. This means that sensors which measure physical quantities (e.g. an active power setpoint) need to be abstracted in a way that the physical measurements are interpreted under certain thresholds. In the presented use-case the sensor interprets whether the setpoint (which is initially an arbitrary number in the range of $0\% - 100\%$) is above or below the threshold of 10%; this results in two potential states.

To answer the three questions above, this work first classifies sensors into four types by differentiating them with regard to two aspects. With regard to the way they provide information about the monitored system and with regard to the type of sensor reliability. In general a sensor with a discrete number of potential states can represent knowledge about the system in two ways:

- **Single-State Sensor** The sensor provides a single state in the domain of interest that represents the current system state according to the sensor.
- **Probabilistic Sensor** Instead of a discrete state, the sensor provides a probability distribution over the domain of interest.

Further, sensors can be differentiated by the type of information that is provided about a sensor's reliability as follows.

- **Deterministic Sensor** Given a specific system state, the sensor will always produce the same result. It thus provides complete certainty that it interprets a system correctly.
- **Symmetric Sensor** The sensor has a certain probability of misinterpretation. This probability is also known as sensor *reliability* and it is independent of the system state. This means, that the sensor always has the same probability of incorrect detection.
- **Asymmetric Sensor** The sensor has a certain probability of misinterpretation which differs depending on the system state that is the input to the sensor. More specifically, the sensor might be more reliable in detecting one variable state than another. This might occur because

the sensor is designed in a way to optimize detection of one specific (probably more critical) state at the cost of higher error rates in detecting other states.

Each sensor can be classified by any combination of knowledge representation and reliability structure. Strictly deterministic sensors are effectively impossible in the real world. For very critical sensors, reliability can be very high in specific ranges of operation (e.g. within specific temperature ranges for physical sensors) but most likely, it will not be guaranteed. Therefore, deterministic sensors can be ignored for the work at hand. Consequently, the aforementioned four types of sensors are: single-state, symmetric (**SSS**) sensors, probabilistic, symmetric (**PS**) sensors, single-state, asymmetric (**SSA**) sensor and probabilistic asymmetric (**PA**) sensors.

In the presented use-case, an example for a SSS sensor is a signature in an IDS that detects packets that would change the active power setpoint to a value $< 10\%$. There could be a software bug or an encoding issue that limits the reliability, but the effects on the performance would be symmetrical (i.e. it is equally probable that a normal packet issues an alert than it is that a suspicious packet does not trigger an alert). In this case the reliability of the sensor can be used to describe the uncertainty in any sensor result. In contrast, a more complex cyber-security sensor like an antivirus scanner would be an SSA sensor; its reliability is traditionally provided in an asymmetric fashion. Here, true positive rate (TPR) and false positive rate (FPR) are a widely accepted metrics to measure sensor performance.

To use sensor evidence in evidential networks, a critical aspect is the representation of knowledge about sensor reliability with relation implication rules. In related work it is common practice to use reliability for discounting of sensor evidence [10], [22]. We argue, that such an approach is only valid for sensors with symmetric reliability and gives inaccurate results otherwise. More specifically, knowledge about the reliability of a specific state of a sensor variable x should be modeled by a set of relation implication rules from $D = \{x\}$ to $D' = \{x'\}$. Here, x' is a newly introduced variable that has the same states as x but represents the sensor evidence under consideration of the sensor's reliability. Each rule can then be written as

$$A \subseteq \Theta_D \Rightarrow A' \subseteq \Theta_{D'} \quad \text{with } \rho \in [\alpha, 1]$$

where α describes the sensor's reliability with respect to A. The challenge then becomes to identify α correctly. A-priori knowledge about sensor quality is most often given in classical probabilities. A valid use of this knowledge in DS belief structures calls for a transfer function f that fulfills the following conditions:

1) The fact that a sensor is in a certain state should not strengthen the belief in the opposite state in the resulting mass function (see Eq. 9).
2) Since we use the pignistic probability (see Eq. 8 in Sect. IV-A) to transfer the final derived belief structures back into the classical probability domain, any transformation (f) from the probability domain into a belief structure should have the pignistic probability as an inverse function (see Eq. 10).

$$m_D(A) = \gamma \implies m_{D'}(B) \leq 1 - \gamma \quad \text{with } A \in \Theta_D \text{ and } B \subseteq \Theta_{D'} \setminus A \tag{9}$$

$$BetP(f(S)) = S \tag{10}$$

Under these conditions, it is now possible to define a general rule stating how a-priori knowledge should be translated to relation implication rules. Consider a variable x with n states $\{x_1, \ldots, x_n\}$. Let's further assume that x represents the actual state of the monitored system and $\widetilde{x}$ represents the state detected by the sensor. Given is the conditional probability $P(x_i|\widetilde{x}_i) = \alpha$; the probability that the state of the system equals x_i if the sensor reported $\widetilde{x}_i$. The goal is to identify a relation implication rule

$$\widetilde{x}_i \subseteq \Theta_D \Rightarrow \widetilde{x}_i' \subseteq \Theta_{D'} \quad \text{with } \rho \in [\alpha', 1]$$

such that the conditions in Eq. 9 and Eq. 10 are fulfilled. The intuitive approach would be to set $\alpha' = \alpha$. This would fulfill Eq. 9 because $m_{D'}(B) = 0 \; \forall B \subseteq \Theta_{D'} \setminus \widetilde{x}_i'$. However, in this case $BetP(m_{D'}(\widetilde{x}_i')) = \alpha + \frac{1-\alpha}{n}$ (see Eq. 7 and Eq. 8). To counter this effect and ensure both conditions, the following rule can be applied.

$$\alpha' = \alpha - \frac{1-\alpha}{n-1} \tag{11}$$

This equation is valid if it is possible to ensure that $\alpha' \geq 0$ which in turn means that it is only applicable to sensors that have a reliability of $\alpha \geq \frac{1}{n}$ for all possible states the sensor can indicate. This happens to be the condition that a sensor performs not worse than simple guessing. A sensor which does not fulfill this requirement should not be used in the first place.

C. INTEGRATION OF CYBER-SECURITY SENSORS

Zomlot *et al.* [11] suggest that classical cyber security sensors are designed only with detection in mind. As a consequence, a sensor with two potential states x_a (alert) and x_n (normal), will only be mapped by one relation implication rule from x_a to x_a'. The sensors are considered completely unreliable when they indicates x_n. This approach has major implications for the suitability of DS theory in the context of this work. In fact, Sec. VII will show that this approach overfits towards malicious system states.

Instead this work presents a more generalized approach that considers evidence in all sensor states. To evaluate the quality of sensors, the receiver-operator characteristic (ROC) is well established. It describes the relationship between true positive rate (TPR) and false positive rate (FPR) of a sensor. TPR and FPR are conditional probabilities under empirical test results. Let X be the sensor that estimates the state of

variable x. Then the TPR and FPR are provided for each potential variable state and are given as follows

$$TPR(x_i) = P(\widetilde{x_i}|x_i) \tag{12}$$

$$FPR(x_i) = \sum_{j \in \{1,\ldots,n\} \setminus i} P(\widetilde{x_i}|x_j) \tag{13}$$

The ROC metric is very valuable for describing the quality of a sensor. However, to infer higher level information, the inverse conditional probability $P(x_i|\widetilde{x_i})$ is required. It is also known as positive predictive value (PPV) and it can easily be computed from TPR and FPR as shown in Eq. 14.

$$PPV(x_i) = P(x_i|\widetilde{x_i}) = \frac{TPR}{TPR + FPR} \tag{14}$$

By applying Eq. 14 and Eq. 11 it is possible to derive a set of rules that describe the trust in a sensor's reliability with respect to every variable state if TPR and FPR are given. In contrast to the mapping rule proposed in [11] this approach considers evidence for all states of the sensor which is important to accurately detect not only malicious but also erroneous or normal states.

VI. IMPLEMENTATION

This section will describe the implementation of the evidential network, based on the use-case described in Sect. III. First, the available sensors will be classified according to the sensor types presented in Sect. V-B. Based on each sensor's reliability, the relation implication rules for each sensor will be derived. Afterwards, the complete evidential network is developed, the knowledge base is defined, and the propagation process is described in an example.

A. SENSOR CLASSIFICATION

This subsection describes the sensors as they will be considered in the reasoning network, as well as their assumed reliability. According to the problem description in Sect. III, six different (logical) sensors are present in the system.

- An antivirus scanner that alerts if malware is detected at a specific host. (*Sensor Type: SSA*)
- A host-based intrusion detection system that triggers if privilege escalation is detected. (*Sensor Type: SSA*)
- A network IDS rule that sends alerts if a packet with an active power set-point of $\leq$ 10% is sent to the PV inverter from the monitored host (i.e., the source IP is the monitored host's IP). (*Sensor Type: SSS*)
- A network IDS rule that triggers an alert if a packet with an active power set-point $\leq$ 10% is received by the PV inverter. (*Sensor Type: SSS*)
- An ARP traffic monitor that alerts if the monitored host is the source of suspicious ARP traffic, or if its MAC address is provided in the ARP payload as the sender's MAC address. (*Sensor Type: PS*)
- An ARP table monitor that alerts if the IP/MAC mapping of the monitored host is changed or keeps changing. (*Sensor Type: SSA*)

To guide the design of relation implication rules that integrate the sensors in the evidential network, the following a-priori knowledge about sensor performance is provided. Where multiple reliability measures are given (namely for the antivirus sensor and the host IDS), the sensor mappings will be varied in Sect. VII to evaluate the performance of the inference approach under changing sensor reliability.

The performance of available antivirus solutions is regularly evaluated by independent organizations like AV-Comparatives. They issue regular reports about the performance of well-known antivirus solutions. Based on their October 2016 Real-World Protection Test,[1] as well as their March 2015 Heuristic / Behavior Test,[2] the following sensor performance can be considered.

Vendor	TPR (Real-World)	TPR (Heuristics)	FPR
Microsoft	95.6%	53%	3%
F − Secure	100%	93%	50%
Kaspersky	100%	92%	0%

For the reliability of host IDS implementations, results from work by Molina [35, p. 53] are adapted as follows.

HIDS Variant	TPR	FPR
Optimal	50%	2%
Realistic	30%	5%
Low	15%	10%

For the symmetric sensors, the level of uncertainty will be described with the use of the scales presented in Sect. V-A. The NIDS sender sensor is assumed to *likely* behave correctly, while confidence in the receivers sensor is *probable*. A confidence of *verylikely* is placed on the ARP victim detector. Finally, the ARP spoofing detector provides the probability that a given host is the source of a specific ARP spoofing attack. One implementation of ARP spoofing and ARP victim sensors is ARPwatch[3] – a linux command line utility. Two sources of information provided by this sensor can be used to identify the source of an ARP spoofing attack. The source (*src*) of the spoofed ARP packet (given in the packet header) and the value of the *sender* attribute in the packet payload. By combining these two values, we design the ARP spoofing sensor of type *PS*. The probability that a given host H is the source of the spoofing is defined as

State	Src $= H$	Src $\neq H$
Sender $= H$	95%	70%
Sender $\neq H$	30%	0%

If the values of *sender* and *src* are both the address of host H the probability that H is the source of the attack is quite high. However, if the header does not point to H the probability that H is the attacker is reduced, but it is very likely that the attacker only tries to masquerade. However, if the payload does not point to H the probability that H is the attacker

[1] https://www.av-comparatives.org/dynamic-tests/ (last accessed 25.11.2016)
[2] https://www.av-comparatives.org/retrospective-test/ (last accessed 25.11.2016)
[3] https://linux.die.net/man/8/arpwatch

is low. In this case the real attacker might put a random existing MAC address in the header to hide the real source. If both do not point to H there is no evidence for the fact that H is the source of the ARP spoofing traffic.

B. EVIDENTIAL NETWORK MODEL

In this section, we introduce an evidential network model that allows us to infer the state of the system, as described in Sect. III. Note that the node status for two different nodes (see Fig. 2) is constructed with the same subnetwork. For simplicity we describe the variables in each subnetwork only once. If the same variable occurs multiple times on different nodes, this is denoted by an @ sign followed by the respective node. The evidential network is described by the following tuple [8], [36]

$$EN = \{V,\ \Theta_V,\ M_V,\ \oplus,\ \downarrow\}$$

in which

$$\begin{aligned}
V_H &= \{NS,\ CM,\ EP,\ SP,\ MW,\ PE,\\
&\qquad AV,\ HI,\ NI,\ AR\};\\
V_S &= \{CS,\ MAN,\ MITM,\ EPR@INV,\ MAC@CTL,\\
&\quad MAC@INV,\ NI@INV,\ ARV@CTL,\ ARV@INV\};\\
V &= V_H@HMI \cup V_H@CTL \cup V_S;\\
\Theta_{v_H} &= \{\Theta_{NS},\ \Theta_{CM},\ \Theta_{EP},\ \Theta_{SP},\ \Theta_{MW},\ \Theta_{PE},\\
&\qquad \Theta_{AV},\ \Theta_{HI},\ \Theta_{NI},\ \Theta_{AR}\};\\
\Theta_{v_S} &= \{\Theta_{CS},\ \Theta_{MAN},\ \Theta_{MITM},\ \Theta_{EPR},\ \Theta_{MAC},\\
&\qquad \Theta_{NI},\ \Theta_{ARV}\};\\
\Theta_v &= \Theta_{v_H} \cup \Theta_{v_S};\\
M_V &= \{m_1,\ m_2,\ \ldots,\ m_{30},\ m_{31}\};\\
\oplus,\ \downarrow&\text{: evidential operations.}
\end{aligned}$$

The EN is illustrated in Figure 3, where each variable in the EN ($v \in V$) is represented by circular nodes, while the mass functions of M_V are indicated by the diamond shaped signs (note that the mass functions in the node status subtree occur twice; therefore, 31 mass functions are present in the EN). The node status is inferred for *HMI* and *CTL* (see Sect. III) with a subnetwork of the same structure. Only the sensor evidence varies. The variables with explanation and frame definitions are listed in Table 2. Each mass function is connected by edges to the subset of variables, which define its domain. Any pair of variables that are not directly connected are assumed to be conditionally independent. The domain of interest for the problem is the domain $D^0 = \{CS, MAN\}$. The two variables describe the control status (normal, erroneous or malicious) and the knowledge about manipulations of the communication between the controller (*CTL*) and the inverter (*INV*). This information is sufficient to identify the four states of interest in the use-case, as defined in Sect. III.

C. DOMAIN KNOWLEDGE

The EN represents the knowledge about the causal relationships between its variables that are described by relation

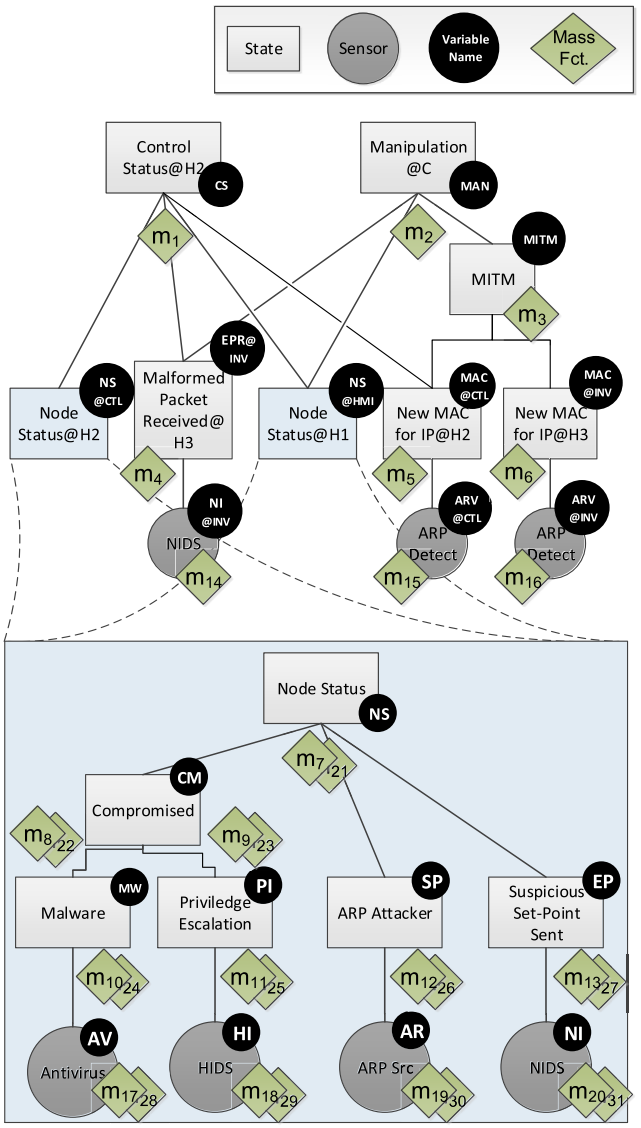

FIGURE 3. System wide evidential network.(*The node status of* CTL *and* INV *expands to the same subnetwork, which is initialized twice to form the complete network.*)

implication rules. The rules are defined based on expert knowledge and then transformed into the mass functions $m_1,\ m_2,\ \ldots,\ m_{13}$, according to Equation IV-F in Section IV-F. If more than one rule defines a relationship, the resulting mass functions are combined, as shown by Equation 3. Table 3 lists the mass functions and corresponding relation implication rules for sensor mappings. The rules are based on the sensor reliabilities discussed in Sect. VI-A. The remaining relation implication rules are provided in the Appendix.

In the following, the relationship between the HIDS sensor (variable HI) and the knowledge that privilege escalation was performed (PE) will be used as an example to demonstrate the complete inference process based on the presented concepts. In order to design the relation implication rule,

TABLE 2. Variables of the node status model.

Variables	Description	Frame	Explanation
CS	ControlStatus	$\{0,1,2\}$	0 normal, 1 erroneous, 2 malicious;
MAN	Manipulation	$\{0,1\}$	0 normal, 1 manipulat.
$MITM$	MITM	$\{0,1\}$	0 normal, 1 mitm
EPR	ErrPackRec	$\{0,1\}$	0 false, 1 true
MAC	MAC change	$\{0,1\}$	0 false, 1 true
ARV	ARP victim	$\{0,1\}$	0 false, 1 true
NS	NodeStatus	$\{0,1,2\}$	0 normal, 1 erroneous, 2 malicious;
CM	Compromised	$\{0,1\}$	0 false, 1 true;
EP	ErrPacket	$\{0,1\}$	0 false, 1 true;
SP	Spoofing	$\{0,1\}$	0 false, 1 true;
MW	Malware	$\{0,1\}$	0 false, 1 true;
PE	PrivEsc	$\{0,1\}$	0 false, 1 true;
AV	sensorAntivirus	$\{0,1\}$	0 inactive, 1 active;
HI	sensorHID	$\{0,1\}$	0 inactive, 1 active;
NI	sensorNID	$\{0,1\}$	0 inactive, 1 active;
AR	sensorARP	$\{0,1\}$	0 inactive, 1 active.

TABLE 3. Mass functions and corresponding relation implication rules for sensor mappings.

Mass functions	Relationships in rules
m_4	Relationships of EPR with NI@INV: $(NI@INV = 1) \Rightarrow (EPR = 1)$ `with confidence between` *probable* `and` 1; $(NI@INV = 0) \Rightarrow (EPR = 0)$ `with confidence between` *probable* `and` 1;
m_5, $m6$	Relationships of MAC with ARV: $(ARV = 1) \Rightarrow (MAC = 1)$ `with confidence between` *verylikely* `and` 1; $(ARV = 1) \Rightarrow (MAC = 1)$ `with confidence between` *verylikely* `and` 1;
m_{10}, m_{24}	Relationships of MW with AV: $(AV = 1) \Rightarrow (MW = 1)$ `with confidence` 1.0; $(AV = 0) \Rightarrow (MW = 0)$ `with confidence between` 0.92 `and` 1.
m_{11}, m_{25}	Relationships of PE with HI: $(HI = 1) \Rightarrow (PE = 1)$ `with confidence between` 0.72 `and` 1; $(HI = 0) \Rightarrow (PE = 0)$ `with confidence between` 0.16 `and` 1.
m_{12}, m_{26}	Relationships of SP with AR: $(AR = 1) \Rightarrow (SP = 1)$ `with confidence` 1.0; $(AR = 0) \Rightarrow (SP = 0)$ `with confidence` 1.0.
m_{13}, m_{27}	Relationships of EP with NI: $(NI = 1) \Rightarrow (EP = 1)$ `with confidence between` *likely* `and` 1; $(NI = 0) \Rightarrow (EP = 0)$ `with confidence between` *likely* `and` 1.

a-priori knowledge about the sensor reliability from Sect. VI-A is used. For a *realistic* sensor, a TPR of 30% and a FPR of 5% are given. Through Eq. 14 the PPV is computed as follows:

$$PPV(HI = 1) = \frac{0.3}{0.3 + 0.05} \approx 0.86$$
$$PPV(HI = 0) = \frac{0.95}{0.95 + 0.7} \approx 0.58$$

The discussion in Sect. V-B states that this information cannot be used directly for the confidence in the relation implication rule. Instead, Eq. 11 needs to be applied to retrieve the correct confidence.

$$\alpha(HI = 1) = PPV - \frac{1 - PPV}{2 - 1} \approx 0.86 - \frac{1 - 0.86}{1} = 0.72$$
$$\alpha(HI = 0) = PPV - \frac{1 - PPV}{2 - 1} \approx 0.58 - \frac{1 - 0.58}{1} = 0.16$$

These values are then used to form the relation implication rules for m_{11} (see also Tab. 3). Each of the two rules can be represented by a mass function over the domain {PE, HI}, namely m^a_{11} and m^b_{11}. Take the first implication rule as an example,

$$(HI = 1) \Rightarrow (PE = 1)$$
`with confidence between` 0.72 `and` 1.

To represent this rule in the format given in Equation IV-F, we have

$$D_A = \{HI\}, \quad \Theta_{D_A} = \{0, 1\}, \ A = 1, A^c = 0;$$
$$D_B = \{PE\}, \quad \Theta_{D_B} = \{0, 1\}, \ B = 1, B^c = 0;$$
$$\alpha = 0.72, \quad \beta = 1.$$

Applying Equation 7, m^a_{11} can be calculated as follows:

$$m^a_{11}(\{(1, 1), (1, 0), (0, 0)\}) = 0.72$$
$$m^a_{11}(\{(1, 1), (1, 0), (0, 1), (0, 0)\}) = 1 - 0.72 = 0.28$$

Similarly, m^b_{11} is calculated as follows:

$$m^b_{11}(\{(0, 0), (1, 1), (0, 1)\}) = 0.16$$
$$m^b_{11}(\{(1, 1), (1, 0), (0, 1), (0, 0)\}) = 1 - 0.16 = 0.84$$

When the two mass functions are combined using Dempster's rule in Equation 3, we obtain the mass function $m_{11} = m^a_{11} \oplus m^b_{11}$, which represents the domain knowledge and is in the product space $\Theta_{PE} \times \Theta_{HI}$:

$$m_{11}(\{(0, 0), (1, 1)\}) = 0.1152$$
$$m_{11}(\{(0, 0), (1, 1), (0, 1)\}) = 0.0448$$
$$m_{11}(\{(0, 0), (1, 1), (1, 0)\}) = 0.6048$$
$$m_{11}(\{(0, 0), (1, 1), (1, 0), (0, 1)\}) = 0.2352$$

VII. EVALUATION

The evaluation of the presented approach is based on 60 scenarios that represent different system states. A detailed description of the scenarios is given in Sect. VII-A. Based on these scenarios, three aspects of the proposed evidential network approach are evaluated. First, the accuracy of the presented evidential network is compared to a reasoning approach based on classical Bayesian probabilities (i.e., the alternative approach does not take uncertainty into account). Further the assumptions made by Ou *et al.* [15] regarding the use of discrete scales to describe the confidence in the relationships between variables is evaluated in Sect. VII-C. The final part of the evaluation concerns the correct integration of sensor evidence, based on a-priori knowledge about sensor reliability (see Sect. VII-D).

A. EVALUATION SETUP

To evaluate the performance of the presented evidential network (EN), the EN is executed over 60 different scenarios. The scenarios differ only in the states that each sensor variable is in. More specifically, the results are not retrieved by monitoring a real system under attack (e.g., no real traffic is monitored). Instead, the 60 specific scenarios (i.e., various sensor states) are derived manually from seven core scenarios. This approach is deliberate and necessary; the goal of the evaluation is to make an assessment of the abilities of the EN to correctly identify the system state based on sensor evidence. The accuracy of the EN – defined by the TPR and FPR of the detection of each state of interest – highly depends on the performance of the sensors. Therefore, it is essential to have full control over the sensors themselves, which is not possible with a real setup. Instead, a real setup would result in less expressive results. An unexpected error in the state detected by a sensor would influence the accuracy of the EN; the results would show the accuracy of the EN under a specific set of sensors, rather than the characteristic of the EN itself.

TABLE 4. Nine Different Configurations of Sensor Reliability. (*The values for each configuration specify the lower confidence bound for the respective relation implication rule specified in Tab. 3.* The rules that form m_{10} and m_{11} describe the reliability of HI and AV respectively.)

	1	2	3	4	5	6	7	8	9
m^a_{10}	0.72	0.2	0.92	0.72	0.2	0.92	0.72	0.2	0.92
m^b_{10}	0.16	0.02	0.32	0.16	0.02	0.32	0.16	0.02	0.32
m^a_{11}	0.92	0.92	0.92	0.3	0.3	0.3	1	1	1
m^b_{11}	0.58	0.58	0.58	0.86	0.86	0.86	0.92	0.92	0.92

However, it is of interest how the EN reacts to changes in sensor performance. The expected reliability of a sensor is encoded in the relation implication rules that specify the sensor mapping. The question to answer is, how the performance of the EN changes if the trust in a set of sensors changes. Therefore, Sect. VI-A introduced different a-priori probabilities of the sensor reliability for the antivirus scanner (AV) and the host IDS (HI). To evaluate the change in accuracy of the EN due to changes in sensor reliability the sensor reliabilities are varied in every evaluation. Section VI-A provided a-priori probabilities for antivirus scanners and host IDS systems. Table 4 shows the lower confidence level (i.e., α) of the respective relation implication rules in each of the nine configurations.

The seven core scenario descriptions are provided in the following. For each core scenario the number of derived scenarios and the expected state (i.e., ground truth) are given. *HMI*, *CTL* and *INV* refer to the nodes in the use-case, as defined in Sect. III. *CS* and *MAN* represent the two variables in the domain of interest (D_0), namely the control status and the manipulation of the network (see Tab. 2 in Sect. VI).

- **Normal Operation** Everything works as expected, but different sensors might issue false positives.

  Scenarios: 4
  Expected: CS = 0, MAN = 0

- **Attack HMI (I)** The attacker infects the *HMI* with malware. From there, a man-in-the-middle (MITM) attack is launched on the communication between the substation controller (*CTL*) and the PV inverter (*INV*). With the successful MITM attack, the active power setpoint is changed to a value $< 10\%$. To disguise the location of the attack, the attacker might spoof the source of the setpoint packet or the parameters of the ARP packets sent to perform the MITM attack. However, the host IDS and the malware scanner pick up on the infection of *HMI*.

  Scenarios: 14
  Expected: CS = 0, MAN = 0

- **Attack *HMI* (II)** The attack is performed like described in the previous scenario. However, the Host IDS is not able to detect any form of privilege escalation. Further, the antivirus scanner might be unable to detect the malware (False Negative). (The attacker still aims to disguise the location of the attack on a network level)

  Scenarios: 10
  Expected: CS = 0, MAN = 1

- **Attack *HMI* (III)** The attack is similar to the previous scenarios. However, the attacker successfully disguises the location of the attack completely (i.e., no alert indicates any malfunctioning of *HMI*). Only changes in the ARP table and the fact that erroneous packets are received are detected.

  Scenarios: 3
  Expected: CS = 0, MAN = 1

- **Masquerade as Controller** The attacker again infects *HMI* to manipulate the communication. In this scenario, to hide the real location of the attack, the attacker replaces *HMI*'s addresses with the controller's addresses to indicate that the controller is the source of the attack or at least the source of the erroneous packets received by the PV inverter. The goal for the EN is to detect that the controller is working as expected.

  Scenarios: 14
  Expected: CS = 0, MAN = 1

- **Controller Error** The controller acts erroneously and sends active power setpoints $< 10\%$ to the PV inverter. This might be caused by human error or a software bug. At the same time, false positives from other sensors might indicate attack behavior that should be correctly discarded.

  Scenarios: 4
  Expected: CS = 1, MAN = 0

- **Controller Attack** The attacker was able to infect the controller directly (either through lateral movement or directly from an external network). It is now possible to send erroneous packets directly without the use of a MITM attack.

  Scenarios: 10
  Expected: CS = 2, MAN = 0

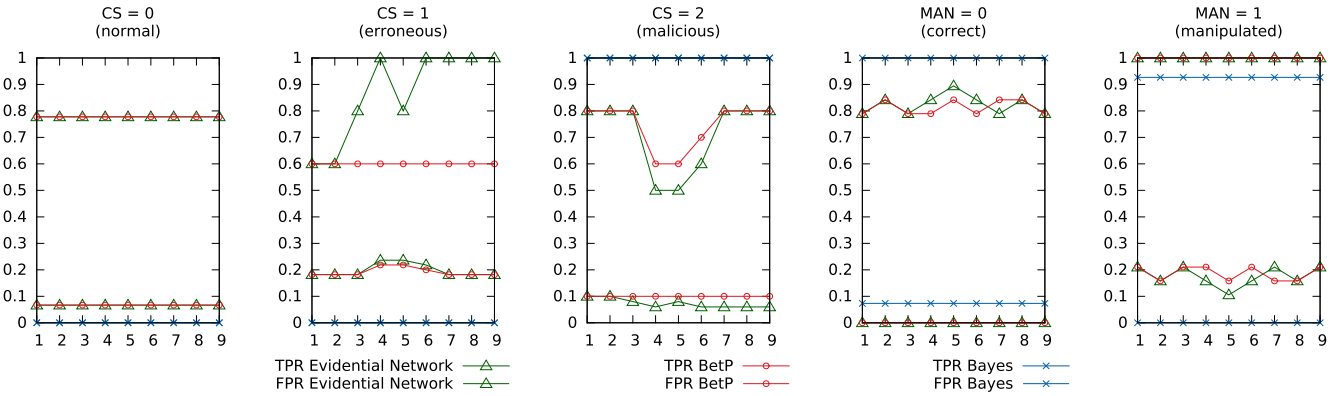

FIGURE 4. **Comparative study of EN performance.**(***Each figure presents the detection accuracy with respect to one specific variable state. They show the TPR and FPR of the compared approaches on the y-axis. The different sensor reliability configurations (see Tab. 4) that are evaluated are shown on the x-axis.***)

To evaluate the performance of the evidential network in different configurations, each scenario is evaluated and the result is compared to the expected result from the respective core scenario. Subsequently, TPR and FPR are computed for each variable state of the two variables in D^0 separately. This approach is necessary to evaluate the detection accuracy with respect to each system state. To be correctly detected, the pignistic transform (see Eq. 8 in Sect. IV-G) of the expected variable state has to be 10% above the pignistic probability for any other variable state of the same variable. Otherwise, the evaluation is counted as a false negative for the expected state and a false positive for any state with a pignistic probability above that threshold. Our research showed that this threshold is appropriate. System states are not always completely clear and the same set of sensor states can have different causes. Therefore, it is desirable that the results are somewhat ambiguous in some states.

B. COMPARISON TO BAYESIEN NETWORKS

Two approaches based on classic Bayesian reasoning have been implemented to provide a comparison for the performance of the EN presented in Sect. III-B. To this extent, relation implication rules that form the evidential network are transformed in two ways to remove uncertainty from the reasoning process. Given a general relation implication rule (see Sect. IV-E), the degree of uncertainty is given by $\beta - \alpha$. This probability needs to be reassigned with the introduction of additional relation implication rules so that $\beta = \alpha$ for every relation implication rule in the EN. Given a general rule (see Eq. IV-F in Sect. IV it is replaced by $|\Theta_B|$ rules; one rule for each focal element in the implied domain. The question then is about the confidence placed in each rule.

In the first approach $\alpha_1 = \alpha/|B|$ and $\alpha_2 = (\beta-\alpha)/|\Theta_B \backslash B|$, where α_1 is the confidence in all rules where $B' \in B$ and α_2 is the confidence in all rules where $B' \notin B$. We will call this approach *Bayes* for the rest of this work.

The second approach introduces the same number of variables, but assigns different confidences in each rule. It is based on the computation of the pignistic transformation (see Eq. 8 in Sect. IV-G) and divides uncertainty equally between all variable states. In this approach $\alpha_1 = \alpha/|B| + (\beta - \alpha)/|\Theta_B|$ and $\alpha_2 = (\beta - \alpha)/|\Theta_B|$.

Consider the following example for relation implication rule m^a_{10} given in Tab. 3 in Sect. III-B.

$$(HI = 1) \Rightarrow (PE = 1) \quad \texttt{with confidence between } 0.72 \texttt{ and } 1.$$

In *Bayesian* reasoning it would be replaced by two rules to remove uncertainty.

$$(HI = 1) \Rightarrow (PE = 1) \texttt{ with confidence } 0.72.$$
$$(HI = 1) \Rightarrow (PE = 0) \texttt{ with confidence } 0.28.$$

Note that $\Theta_B = \Theta_{PE} = \{0, 1\}$ and $B = \{1\}$. Similarly, the *BetP* conversion would replace m^a_{10} with two rules as well. However, their confidence would be different.

$$(HI = 1) \Rightarrow (PE = 1) \texttt{ with confidence } 0.86.$$
$$(HI = 1) \Rightarrow (PE = 0) \texttt{ with confidence } 0.14.$$

Figure 4 shows the TPR and FPR for each variable state in D^0 and for each sensor reliability case in Tab. 4. The results show that the EN can detect all system states of interest with good accuracy. The TPR is (with some exceptions, based on sensor reliability) around 80% and the FPR below 20% (for many states even below 10%). These are very good results considering that for some results only limited sensor evidence is available. Further, as argued previously in Sect. III, the same sensor evidence can indicate various higher level states; some degree of uncertainty is therefore wanted. While a completely accurate detection would be desirable, the results show that the presented approach is able to indicate if sensor evidence is not sufficient to make clear statements about the causality. This is indicated by a moderate FPR. Given the way TPR and FPR are computed, this also explains why the TPR is not higher.

The steep rise in the TPR of erroneous control state detection (i.e., $CS = 1$), as well as the dip in malicious control

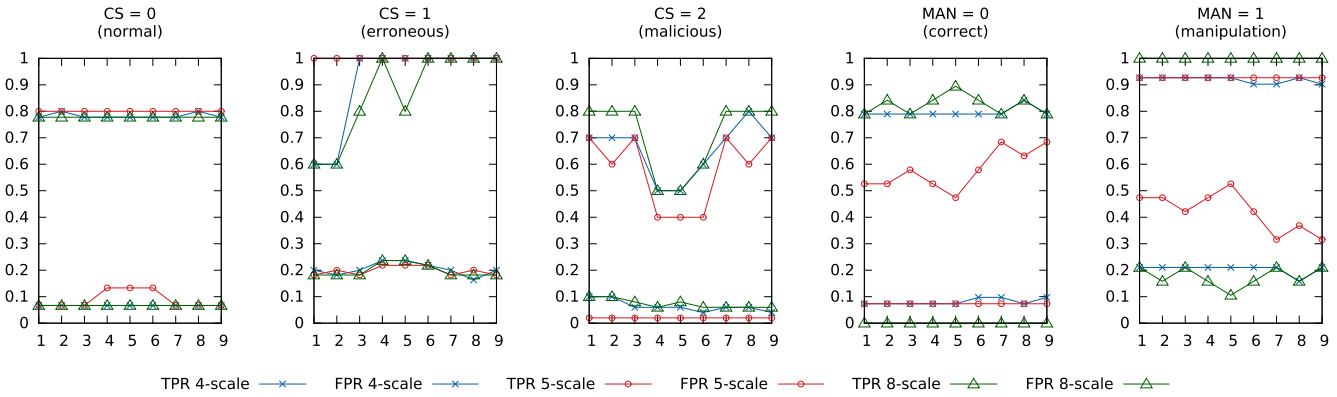

FIGURE 5. **Evaluation of scales of different size for the relation implication rule design process (see Tab. 1).(*Each figure presents the detection accuracy with respect to one specific variable state. They show the TPR and FPR of the compared approaches on the y-axis. The different sensor reliability configurations (see Tab. 4) that are evaluated are shown on the x-axis.*)**

state detection, is a result of the changes in sensor reliability. In Tab. 4 it is shown that the reliability of both *AV* and *HI* in normal states is lowest in configuration 1 and 2; for malicious detection it is lowest in configuration 4 and 5. Therefore, it can be concluded that the accuracy of the EN depends on the reliability of the sensors. However, the unreliability of one sensor can be compensated for by other sensors. This is shown by the fact that only the configurations where both sensors perform worst show a significant change in detection rates.

What is more, in comparison to the two Bayes approaches, the proposed evidential network can compete through all states. The *Bayes* transformation is unable to detect any case of $CS = 0$ or $CS = 1$. It highly overfits to malicious control behavior.

The *BetP* transformation is much more accurate than the Bayesian approach over all system states. The similarity of the results to those of the EN is expected; the use of the pignistic transformation on all relation implication rules already makes use of DS theory, and considers the uncertainty of the relationships. There is no state where this approach shows notable improvements over the evidential network. Furthermore, with increasing sensor reliability, the accuracy in erroneous state detection improves for the EN, while no such effect can be seen for the BetP transformation. This is a key finding. Evidential networks enable a much more intuitive design of the knowledge base. But even if this intuitive design is used to derive Bayesian relation rules, the inference process of the EN results in higher accuracy. Finally, the results from the EN are initially given in DS belief structures that provide additional information to traditional Bayesian probabilities. Only for this evaluation the belief structures are put through the pignistic transformation to make the results comparable. This additional information is lost when Bayesian reasoning is used.

C. EVALUATION OF DIFFERENT CONFIDENCE SCALES

In Sect. V-A, the concept of scales was introduced to simplify the design of relation implication rules. This evaluation considers the impact that the number of elements in these scales has on the EN's accuracy. In work by Ou *et al.* [15], the authors claim that the reliability placed on rules in intrusion detection systems can be accurately classified by a scale of four elements. This approach was later adopted by Zomlot *et al.* [11] in the context of DS theory for alert correlation. Figure 5 shows TPR and FPR for each variable state in D^0, but with relation implication rules of different confidences (based on the scale used). The results show no conclusive evidence, that a higher number of elements in the scale results in consistently better results. The detection rates for $CS = 0$ (normal operation) are stable through all nine sensor reliability combinations. While the five element scale appears more accurate with lower sensor reliability for $CS = 1$ (shown by the steeper rise) it performs consistently worse in detection of $MAN = 0$ (absence of manipulation) than the other two scales, and produces a higher false positive rate for $MAN = 1$. However, the four element scale does not show the same decrease in accuracy that we would expect if the cause would be the number of elements in the scale.

However, the results do show that the design of the relation implication rules is very critical to system performance. For fine grained relationships, a scale with more elements can make it easier during the design process to accurately model causality.

D. SENSOR MAPPING EVALUATION

The final evaluation considers the integration of sensor evidence in the EN. Section V-B presented a set of rules that can be applied to transfer Bayesian a-priori probabilities about sensor reliability into relation implication rules. Furthermore, Sect. V-C applied these rules to IT security sensors, which are traditionally evaluated through TPR and FPR. Figure 6 compares the performance of the EN when sensors are integrated the way it is suggested in Sections V-B and V-C to two other approaches. In the BetP approach, the relation implication rules used to map the sensor evidence are strictly Bayesian. The mapping is performed according to the transformation introduced as *BetP* in Sect. VII-B (see also Sect. VII-C for details about the transformation of relationship rules).

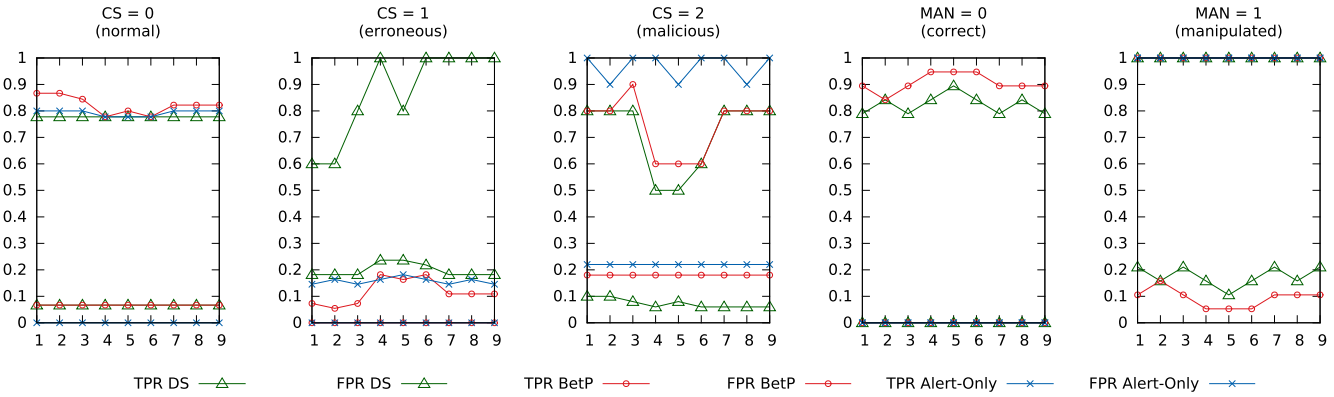

FIGURE 6. Evaluation of different approaches to integrate sensor evidence based on sensor reliability. (*Each figure presents the detection accuracy with respect to one specific variable state. They show the TPR and FPR of the compared approaches on the y-axis. The different sensor reliability configurations (see Tab. 4) that are evaluated are shown on the x-axis.*)

The third approach (called *Alert-Only*) was presented by Zomlot *et al.* [11]. The authors claim that IDS rules are designed only to detect a specific condition. An alert from that rule can be an indicator for a specific (often malicious) behavior. However, the absence of an alert has to be considered completely irrelevant. The argument is that the rules are not designed to detect the absence of attacks and should therefore not be used to provide evidence for anything else than attacks. As an example, consider again an original relation implication rule: m^b_{10}.

$$(HI = 0) \Rightarrow (PE = 0)$$
$$\texttt{with confidence between } 0.16 \texttt{ and } 1.$$

This rule would be eliminated in this approach because the absence of an alert $HI = 0$ provides evidence about a system state.

Figure 6 show the evaluation results in comparison. While the detection rates of $CS = 0$ are consistent among the different sensor mapping methods, significant performance differences can be seen for $CS = 1$ and $CS = 2$. The *Alert-Only* mapping highly overfits towards the detection of malicious states. Only malicous control behavior and communication manipulations are detected; all other states have a TPR of 0, which is not acceptable. Additionally, the FPR for malicious states is consistently higher than those of other approaches. This result can be expected because the approach ignores information about normal behavior.

The performance of the *BetP* mapping approach is similar to the *DS* approach for the detection of manipulations. However, the *BetP* approach for sensor integration is unable to detect erroneous behavior in the controller. To detect all possible states in a system (normal, erroneous and malicious), the sensor integration approach that is presented in this work shows the best results.

E. SUMMARY

The presented evaluation shows that evidential networks provide a very suitable solution to state inference in cyber-physical systems. The accuracy of ENs is at least equal to that of reasoning approaches with classical probabilities, but with several proven practical benefits. First, the design of the EN through relation implication rules is much more intuitive, because uncertainty about the encoded relationships can be considered. This design process can be further simplified with the use of confidence scales. Where previous work [11], [15] lacked clear information on the relative performance of confidence scales, this work has experimentally compared a range of previously proposed scales in practice. We show that the number of elements in the scale has no significant impact on the accuracy of the EN (see Sect. V-A and VII-C). Further, the results of the state inference is given in belief structures, rather than classical probabilities. These belief structures provide more information, because they represent the level of uncertainty in the results but can, at the same time, estimate classical probabilities. The degree of uncertainty provides information about the level of trust that can be placed in the inferred results; something that can increase the acceptance of the solution for operators. Finally, we were able to show that the way in which sensor evidence is considered in ENs is critical to their performance. Following a detailed discussion, we were able to provide a generalized solution how knowledge about sensor reliability should be leveraged in an EN in Sect. V-B. In Sect. VII-D, we were able to show that this approach led to a high increase in detection accuracy when compared to approaches from related work (see [11], [14], [21]).

VIII. CONCLUSION

This work proposes evidential networks for state inference in cyber-physical systems. State inference aims to identify the causality between low-level evidence and higher-level system states. This goes beyond the goals of most widely adopted correlation techniques (see Sect. II and [7]), and is critical to support control decisions of human operators or automated algorithms in systems of ever increasing complexity. The presented results show that evidential networks are an improvement on approaches that use classical

Bayesian probabilities to describe confidence in hypotheses (see [7], [20], [21]). They allow a more intuitive design of the system model with the use of uncertain information. In addition to a representation of the results in the classical probability domain, belief structures also provide information about the degree of certainty in these probabilities. This is useful for decision makers, because it provides a measure for the trust that can be placed in the results. At the same time, experimental results show that the detection accuracy is better or comparable with approaches that perform reasoning with Bayesian probabilities.

Although this work evaluates evidential networks on a single use-case, we argue in Sect. III that the use-case investigated experimentally is representative of a wide range of CPS. It covers general problems for state inference in cyber-physical systems, such as different sensor types or system-specific thresholds that lead to specialized system states. At the same time, the use-case is not artificially designed – it was shown by Kang *et al.* [2] that it is based on a number of real world scenarios.

Solutions for state inference are of vital importance for an informed, timely and accurate response by operators in systems of increasing complexity. We argue that dependable operation of cyber-physical systems requires a holistic view on sensor evidence and system states that is able to accurately differentiate between normal operation, specific error states and specific attack states. This work shows that EN provides better visibility about complex system states, which is a requirement for more accurate control decisions.

APPENDIX

The following table completes the relation implication rules that form the knowledge base in the evidential network (see Sect. III-B for details).

Mass	Relationships in rules
m_1	Relationships of CS with EPR, MAC@CTL, NS@HMI and NS@CTL: $(NS@CTL = 0) \Rightarrow (CS = 0)$ with confidence between *verylikely* and 1; $(NS@CTL = 1) \Rightarrow (CS = 1)$ with confidence between *likely* and 1; $(NS@CTL = 2) \Rightarrow (CS = 2)$ with confidence between *likely* and 1; $(MAC@CTL = 1) \Rightarrow (CS \in \{0, 1\})$ with confidence between *likely* and 1; $(NS@HMI = 2) \Rightarrow (CS = 0)$ with confidence between *improbable* and 1; $(EPR = 1, MAC@CTL = 0) \Rightarrow (CS \in \{1, 2\})$ with confidence between *feasible* and 1; $(EPR = 1, MAC@CTL = 1) \Rightarrow (CS = 0)$ with confidence between *feasible* and 1; $(EPR = 1, MAC@CTL = 1, NS@HMI = 2) \Rightarrow (CS = 0)$ with confidence between *probable* and 1; $(EPR = 1, MAC@CTL = 0, NS@HMI \in \{0, 1\}) \Rightarrow (CS \in \{1, 2\})$ with confidence between *verylikely* and 1; $(EPR = 0) \Rightarrow (CS \in \{0\})$ with confidence between *potentially* and 1; $(NS@HMI \in \{0, 1\}) \Rightarrow (MAN = 0)$ with confidence between *improbable* and 1; $(MITM = 1) \Rightarrow (MAN = 1)$ with confidence between *probable* and 1; $(EPR = 1) \Rightarrow (MAN = 1)$ with confidence between *potentially* and 1; $(EPR = 0) \Rightarrow (MAN = 0)$ with confidence between *verylikely* and 1; $(MITM = 0, NS@HMI \in \{0, 1\}) \Rightarrow (MAN = 0)$ with confidence between *likely* and 1;
m_3	Relationships of MITM with MAC@CTL and MAC@INV: $(MAC@CTL = 1, MAC@INV = 0) \Rightarrow (MITM = 1)$ with confidence between *potentially* and 1; $(MAC@CTL = 0, MAC@INV = 1) \Rightarrow (MITM = 1)$ with confidence between *potentially* and 1; $(MAC@CTL = 0, MAC@INV = 0) \Rightarrow (MITM = 0)$ with confidence between *probable* and 1; $(MAC@CTL = 1, MAC@INV = 1) \Rightarrow (MITM = 1)$ with confidence between *probable* and 1;
m_7, m_{21}	Relationships of NS with CM, EP, and SP : $(CM = 1) \Rightarrow (NS = 2)$ with confidence between *verylikely* and 1; $(CM = 0) \Rightarrow (NS \in \{0, 1\})$ with confidence between *likely* and 1; $(SP = 1) \Rightarrow (NS = 2)$ with confidence between *verylikely* and 1; $(SP = 0) \Rightarrow (NS \in \{0, 1\})$ with confidence between *potentially* and 1; $(EP = 1) \Rightarrow (NS \in \{1, 2\})$ with confidence between *probable* and 1; $(EP = 0) \Rightarrow (NS = 0)$ with confidence between *potentially* and 1; $(CM = 1,\ SP = 1,\ EP = 1) \Rightarrow (NS = 2)$ with confidence between *probable* and 1; $(CM = 0,\ SP = 0,\ EP = 1) \Rightarrow (NS = 1)$ with confidence between *probable* and 1; $(CM = 0,\ SP = 0,\ EP = 0) \Rightarrow (NS = 0)$ with confidence between *probable* and 1.
m_8, m_{22}	Relationships of CM with MW: $(MW = 1) \Rightarrow (CM = 1)$ with confidence between *likely* and 1; $(MW = 0) \Rightarrow (CM = 0)$ with confidence between *improbable* and 1.
m_9, m_{23}	Relationships of CM with PE: $(PE = 1) \Rightarrow (CM = 1)$ with confidence between *possible* and 1; $(PE = 0) \Rightarrow (CM = 0)$ with confidence between *improbable* and 1.

REFERENCES

[1] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1244–1253, Sep. 2013.

[2] B. Kang *et al.*, "Investigating cyber-physical attacks against IEC 61850 photovoltaic inverter installations," in *Proc. IEEE 20th Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2015, pp. 1–8.

[3] G. Dán, H. Sandberg, M. Ekstedt, and G. Björkman, "Challenges in power system information security," *IEEE Security Privacy*, vol. 10, no. 4, pp. 62–70, Jul./Aug. 2012.

[4] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir, "The future of human-in-the-loop cyber-physical systems," *Computer*, vol. 46, no. 1, pp. 36–45, Jan. 2013.

[5] P. P. Shenoy, "A valuation-based language for expert systems," *Int. J. Approx. Reason.*, vol. 3, no. 5, pp. 383–411, 1989.

[6] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[7] S. Salah, G. Maciá-Fernández, and J. E. Díaz-Verdejo, "A model-based survey of alert correlation techniques," *Comput. Netw.*, vol. 57, no. 5, pp. 1289–1317, 2013.

[8] A. Benavoli, B. Ristic, A. Farina, M. Oxenham, and L. Chisci, "An application of evidential networks to threat assessment," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 2, pp. 620–639, Apr. 2009.

[9] C. Simon and P. Weber, "Evidential networks for reliability analysis and performance evaluation of systems with imprecise knowledge," *IEEE Trans. Rel.*, vol. 58, no. 1, pp. 69–87, Mar. 2009.

[10] X. Hong, C. Nugent, M. Mulvenna, S. McClean, B. Scotney, and S. Devlin, "Evidential fusion of sensor data for activity recognition in smart homes," *Pervasive Mobile Comput.*, vol. 5, no. 3, pp. 236–252, 2009.

[11] L. Zomlot, S. C. Sundaramurthy, K. Luo, X. Ou, and S. R. Rajagopalan, "Prioritizing intrusion analysis using Dempster–Shafer theory," in *Proc. 4th ACM Workshop Secur. Artif. Intell. (AISec)*, New York, NY, USA, Oct. 2011, pp. 59–70.

[12] T. M. Chen and V. Venkataramanan, "Dempster–Shafer theory for intrusion detection in ad hoc networks," *IEEE Internet Comput.*, vol. 9, no. 6, pp. 35–41, Nov. 2005.

[13] W. Hu, J. Li, and Q. Gao, "Intrusion detection engine based on Dempster–Shafer's theory of evidence," in *Proc. IEEE Int. Conf. Commun., Circuits Syst.*, Jun. 2006, pp. 1627–1631.

[14] A. C. Squicciarini, G. Petracca, W. G. Horne, and A. Nath, "Situational awareness through reasoning on network incidents," in *Proc. 4th ACM Conf. Data Appl. Secur. Privacy (CODASPY)*, New York, NY, USA, 2014, pp. 111–122.

[15] X. Ou, S. R. Rajagopalan, and S. Sakthivelmurugan, "An empirical approach to modeling uncertainty in intrusion analysis," in *Proc. Annu. Comput. Secur. Appl. Conf.*, 2009, pp. 494–503.

[16] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *J. Netw. Comput. Appl.*, vol. 36, no. 1, pp. 16–24, Jan. 2013.

[17] R. Mitchell and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Comput. Surv.*, vol. 46, no. 4, pp. 55:1–55:29, Mar. 2014.

[18] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 303–336, 1st Quart., 2014.

[19] H. T. Elshoush and I. M. Osman, "Alert correlation in collaborative intelligent intrusion detection systems—A survey," *Appl. Soft Comput.*, vol. 11, no. 7, pp. 4349–4365, Oct. 2011.

[20] S. A. Mirheidari, S. Arshad, and R. Jalili, "Alert correlation algorithms: A survey and taxonomy," in *Cyberspace Safety and Security*. Cham, Switzerland: Springer, 2013, pp. 183–197.

[21] Y. Zhai, P. Ning, P. Iyer, and D. S. Reeves, "Reasoning about complementary intrusion evidence," in *Proc. 20th Annu. Comput. Secur. Appl. Conf.*, Dec. 2004, pp. 39–48.

[22] X. Hong *et al.*, "Evidential event inference in transport video surveillance," *Comput. Vis. Image Understand.*, vol. 144, pp. 276–297, Mar. 2016.

[23] F. Aguirre, M. Sallak, W. Schön, and F. Belmonte, "Application of evidential networks in quantitative analysis of railway accidents," *Proc. Inst. Mech. Eng. O, J. Risk Rel.*, vol. 227, no. 4, pp. 368–384, 2013.

[24] S. Qiu, R. Sacile, M. Sallak, and W. Schön, "On the application of valuation-based systems in the assessment of the probability bounds of hazardous material transportation accidents occurrence," *Safety Sci.*, vol. 72, pp. 83–96, Feb. 2015.

[25] B. Kang, K. McLaughlin, and S. Sezer, "Towards a stateful analysis framework for smart grid network intrusion detection," in *Proc. 4th Int. Symp. ICS SCADA Cyber Secur. Res. (ICS-CSR)*, Aug. 2016, pp. 124–131.

[26] D. Wei and K. Ji, "Resilient industrial control system (RICS): Concepts, formulation, metrics, and insights," in *Proc. 3rd Int. Symp. Resilient Control Syst. (ISRCS)*, Aug. 2010, pp. 15–22.

[27] R. Haenni and N. Lehmann, "Probabilistic argumentation systems: A new perspective on the Dempster–Shafer theory," *Int. J. Intell. Syst.*, vol. 18, pp. 93–106, Jan. 2003.

[28] P. Smets, "Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem," *Int. J. Approx. Reason.*, vol. 9, no. 1, pp. 1–35, 1993.

[29] B. Ristic and P. Smets, "Target identification using belief functions and implication rules," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 1097–1103, Jul. 2005.

[30] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," in *Proc. UAI*, 1990, pp. 29–40.

[31] B. Schneier, "Attack trees," *Dr. Dobb's J.*, vol. 24, no. 12, pp. 21–29, 1999.

[32] D. H. Stamatis, *Failure Mode and Effect Analysis: FMEA From Theory to Execution*. Milwaukee, WI, USA: American Society for Quality, Quality Press, 2003.

[33] N. G. Leveson, *Engineering a Safer World—Systems Thinking Applied to Safety*. Cambridge, MA, USA: MIT Press, 2011.

[34] *Systems and Software Engineering—Vocabulary*, document ISO/IEC/IEEE 24765:2010(E), 2010, pp. 1–418.

[35] J. Molina, "Evaluating host intrusion detection systems," Ph.D. dissertation, Dept. Mech. Eng., Univ. Maryland, College Park, MD, USA, 2007.

[36] P. P. Shenoy, "Valuation-based systems: A framework for managing uncertainty in expert systems," in *Fuzzy Logic for the Management of Uncertainty*, L. A. Zadeh and J. Kacprzyk, Eds. New York, NY, USA: Wiley, 1992, pp. 83–104.

IVO FRIEDBERG received the master's degree in software engineering and Internet computing from the Vienna University of Technology in 2014. He is currently pursuing the Ph.D. degree with the AIT Austrian Institute of Technology and Queen's University Belfast, with a focus on resilience of smart grids under cyber-attacks. His research interests lie in intrusion response, machine learning, resilience and cyber-physical control systems.

XIN HONG received the B.Sc. degree (Hons.) from Fudan University, China, and the Ph.D. degree in artificial intelligence from the University of Ulster, U.K. She is currently a Research Fellow with the Center for Secure Information Technologies, Queen's University Belfast. Her research interests include reasoning under uncertainty, intelligent diagnosis, pattern recognition, information fusion, the application of evidential networks for event inference in transport video surveillance, and cyber-physical security.

KIERAN MCLAUGHLIN is currently the Research Leader in industrial control system (ICS) network security, for SCADA, smart grid, and critical infrastructure. His interests include technologies for intrusion detection/prevention using ICS/SCADA protocol, behavioral and stateful analysis techniques, and ICS vulnerability and attack behavior analysis. He is also a Co-I for the current EU FP7 projects PRECYSE and SPARKS. These involve leading European industrial and academic partners and are addressing protection of critical infrastructures and smart grids against cyber-attacks. He is also a Co-I for the EPSRC funded CAPRICA, NIMBUS, and ARIES research programs and a Co-Director of the Queen's M.Sc. in cyber security.

PAUL SMITH received the Ph.D. degree in 2003. He was a Senior Research Associate with Lancaster University, U.K. He is currently a Senior Scientist with the Center for Digital Safety and Security of AIT Austrian Institute of Technology. He has been involved in the area of network resilience for several years, authoring numerous conference and journal articles. He is currently coordinating the EU-funded SPARKS project, which is investigating the security and resilience of the smart grid, with a research focus in the project on cyber-security risk assessment.

PAUL C. MILLER was born in Belfast, U.K., in 1964. He received the B.Sc. (Hons.) and Ph.D. degrees in pure and applied physics from Queen's University Belfast, Belfast, in 1985 and 1989, respectively. From 1989 to 1991, he was a Research Fellow with the School of Electrical and Electronic Engineering, Queen's University Belfast. From 1991 to 1999, he was initially a Research Scientist and then a Senior Research Scientist with the Defense Science and Technology Organization, Adelaide, Australia. In 1999, he rejoined the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, as a Lecturer, where he has been a Senior Lecturer since 2008. Since returning to academia, he has been involved in video analytics for both defense and civilian CCTV applications, and biomedical image analysis. He has authored over 100 articles. His current research interests include image restoration, segmentation, multitarget tracking, person re-identification, the gender/age profiling of subjects in video, and the convergence of cyber and physical security. He was a recipient of the IPRCS International Machine Vision and the Image Processing Conference Best Paper Award in 2008.