Data Fusion-Based Anomaly Detection in Networked Critical Infrastructures

Béla Genge, Christos Siaterlis, and Georgios Karopoulos
Institute for the Protection and Security of the Citizen, Joint Research Centre
Via E. Fermi 2749, Ispra (VA), Italy
{bela.genge,christos.siaterlis,georgios.karopoulos}@jrc.ec.europa.eu

***Abstract*—The dramatic increase in the use of Information and Communication Technologies (ICT) within Networked Critical Infrastructures (NCIs), e.g., the power grid, has lead to more efficient and flexible installations as well as new services and features, e.g., remote monitoring and control. Nevertheless, this has not only exposed NCIs to typical ICT systems attacks, but also to a new breed of cyber-physical attacks. To alleviate these issues, in this paper we propose a novel approach for detecting cyber-physical anomalies in NCIs using the concept of cyber-physical data fusion. By employing Dempster-Shafer's "Theory of Evidence" we combine knowledge from the cyber and physical dimension of NCIs in order to achieve an Anomaly Detection System (ADS) capable to detect even small disturbances that are not detected by traditional approaches. The proposed ADS is validated in a scenario assessing the consequences of Distributed Denial of Service (DDoS) attacks on Multi Protocol Label Switching (MPLS) Virtual Private Networks (VPNs) and the propagation of such disturbances to the operation of a simulated power grid.**

***Index Terms*—Anomaly Detection System; Networked Critical Infrastructures; SCADA; Data Fusion; MPLS; DDoS.**

I. Introduction

In the last years we have witnessed a dramatic increase in the use of Information and Communication Technologies (ICT) within Networked Critical Infrastructures (NCIs), e.g., power plants, water plants and energy smart grids. As a result, it has been possible to implement more efficient and flexible installations as well as new services and features such as remote monitoring and maintenance, energy markets, and the newly emerging smart grid. Although the advantages of this trend are indisputable, the dramatical shift from a completely isolated environment, to a "system of systems" integration with existing infrastructures, e.g., the Internet, has lead to the exposure of NCIs to significant cyber threats. This has been highlighted by many studies on Supervisory Control And Data Acquisition (SCADA) systems [1], [2], [3], i.e., the core infrastructure providing monitoring and control of physical processes, which proved that NCIs are not only subject to typical IT systems attacks, but also to a new breed of *cyber-physical* attacks.

Cyber-physical attacks exploit the cyber and physical dimensions of NCIs and can have serious consequences on their normal operation. Stuxnet [4], the first malware specifically designed to attack the control hardware of NCIs, was a clear demonstration in this sense and showed a new level of sophistication in malware development. Stuxnet raised many open questions, but most importantly it highlighted the lack of an efficient approach to detect complex cyber-physical attacks on NCIs. Unfortunately, as stated in [3] the peculiarities of NCIs can render traditional ICT security techniques ineffective when faced with cyber-physical attacks. Therefore, this paper alleviates the aforementioned issues by proposing a novel approach for detecting cyber-physical anomalies in NCIs using the concept of *Cyber-Physical Data Fusion* and Dempster-Shafer's "Theory of Evidence". Anomaly-based intrusion detection is well-suited for scenarios in which the behavior of malicious attacks is not known beforehand. In traditional ICT systems anomaly-based detections might have a high false-positive rate due to dynamic traffic and unpredictable behavior. However, since NCIs expose a more predictable behavior, the use of anomaly detection systems is well suited in this case.

This work is motivated by the following arguments: (i) the complexity of cyber-physical attacks targeting NCIs requires techniques addressing all their dimensions; (ii) interdependencies between the cyber and physical require approaches that fuse the two dimensions in order to aggregate evidence and to provide a holistic view of NCIs; and (iii) existing techniques do not address the complexity of the entire system. The novelty of the proposed approach is that it combines reports of various cyber and physical sensors in order to provide a unified view of an entire NCI installation; an important aspect that is missing from related techniques. The approach is experimentally validated in a scenario assessing the consequences of Distributed Denial of Service (DDoS) attacks on Multi Protocol Label Switching (MPLS) Virtual Private Networks (VPNs) and the propagation of such disturbances to the normal functioning of a simulated power grid.

The paper is organized as follows: Section II provides an overview of related work, while the proposed approach is detailed in Section III. Then, the performance of the proposed ADS is evaluated in Section IV, and the paper concludes in Section V.

II. Related Work

Anomaly detection is a well established field of research. For NCIs and especially for SCADA systems we can find several approaches that can be categorized based on the addressed NCI dimension: (i) approaches addressing the cyber dimension; (ii) approaches addressing the physical dimension; and (iii) approaches addressing both dimensions.

Most of the purely cyber-oriented approaches assume highly deterministic communications patterns, which is a distinctive characteristic of NCIs. The recent work of Pleijsier [5] shows that by inspecting connection parameter patterns between client and server stations we can detect abnormal communications behaviors in NCIs. In a similar way Barbosa, *et al.* [6] and Garitano, *et al.* [7], proved that the periodicity of network traffic in NCIs can be the basis for detecting anomalies in the cyber dimension of NCIs. As shown in [3], the main disadvantage of the previously mentioned approaches is that an attacker might employ legitimate traffic flows with a devastating effect on physical processes. Therefore, as proposed in this paper, such techniques need to be coupled with the physical dimension of NCIs in order to ensure an accurate detection of both cyber and physical anomalies.

The presence of the physical dimension in the architecture of NCIs opened the way for model predictive techniques to be applied in detecting anomalies in the operation of physical processes. In this direction, we find the work of Svendsen and Wolthusen [8], [9] which employed physical process models together with approaches from feedback control theory to predict future states and ultimately to detect physical anomalies. A similar approach has been developed by Cárdenas, *et al.* [10], where a model of a chemical plant and feedback control loops were used to predict the state of the physical process and to detect attacks against NCIs. Although such approaches might accurately detect the presence of anomalies, they require a complete and highly detailed model of the physical process, which is not always available. To address this issue, the work of Nai Fovino, *et al.* [11], [12], [13] builds on the assumption that every attack on NCIs will ultimately lead to a transition of the system from a *secure state* to a *critical state*. Although this approach eliminates the need of a highly detailed physical process model it does not take into account the cyber dimension of NCIs. Furthermore, since the detection engine is running in the cyber space, it builds on the assumption that communications with the physical process sensors are always intact. However, disruptive cyber attacks can cause the complete interruption of communications, which will render this approach ineffective. In contrast, the approach proposed in this paper takes into account not only evidence from both dimensions of NCIs, but also the lack of it.

Finally, we mention approaches addressing both the cyber and physical dimensions of NCIs. In [14], Raciti and Nadjm-Tehrani proposed an alert aggregation technique that collects data from cyber and physical sensors. Unfortunately the approach does not fuse the evidence from the two dimensions and therefore a combined view with this technique is not possible. Conversely, in [15], Levorato and Mitra proposed a unified view of several smart grid dimensions, e.g., energy market and weather conditions, by using sparse approximation and wavelet theories. The approach is demonstrated to be well-suited for high-level analysis of the smart grid, but it might loose its effectiveness when dealing with low-level aspects, e.g., detecting SYN attacks. Furthermore, since it requires a priori training with real data, the technique could be highly error-prone and might not detect anomalies generated by malware that is already in place. In contrast, the technique proposed in this paper fuses together the cyber and physical dimensions of NCIs and most importantly it does not require an initial training data set.

III. Proposed Approach: Cyber-Physical Data Fusion

After reviewing available data fusion mechanisms we selected the Dempster-Shafer (D-S) "Theory of Evidence" based on the following considerations. First, as we do not have a good model for the normal network and physical system we excluded physical methods, like the Kalman filter that requires the knowledge of the state transition matrix. Second, we avoided methods that make strong assumptions about the measured data, like a naive Bayesian classifier that assumes knowledge of the "a priori" probability distribution of the observed random variables. Another factor that influenced our choice was that the D-S approach allows the use of information from multiple heterogeneous sources with different sensitivity, reliability and false alarm rates. By using it, we were also able to incorporate expert knowledge from network administrators, without building an extremely complicated expert system.

D-S enables the combination of evidence generated from multiple *sensors*, e.g., basic detection elements. Within the architecture of the proposed Anomaly Detection System (ADS) each sensor monitors, detects and reports its own perspective (belief) of the observed cyber and/or physical attributes. The beliefs of several sensors are then combined (fused) in order to provide a unified view of the system state.

From our perspective, NCIs are seen as having a stochastic behavior without assuming any underlying functional model. The attempt to infer the unknown state of the system is based on knowledge reported by sensors, that may have been acquired based on totally different criteria. Possible sources of information are signature-based IDS, custom DDoS detection programs, control hardware, or physical sensors. Therefore, the proposed technique could also be seen as complementary to existing ADSs.

A. Overview of the Theory of Evidence

Dempster-Shafer's "Theory of Evidence" can be considered an extension of Bayesian inference. There are many different ways to interpret the basic mathematical formulations of the theory that was introduced by Shafer [16]. It can be viewed either from a probabilistic or an axiomatic point of view and all these approaches are concisely surveyed in [17]. Besides the different theoretical approaches and interpretations, all of them boil down to the same mathematical formulas. The "Theory of Evidence" has been analyzed in the fields of statistical inference, diagnostics, risk analysis and decision analysis. Our methods and notations are mostly inspired from the field of "Diagnostics" [18].

Let us have a set of possible states of a system $\theta_1, \theta_2, ..., \theta_N \in \Theta$, which are mutually exclusive and complete (exhaustive). The set Θ is often called *the frame of discernment*. We will call hypotheses H_i subsets of Θ, in other words elements of the powerset 2^{Θ}. Our goal is to infer the true system state without having an explicit model of the system, just based on some evidence (measurements) $E_1, ..., E_M$. Such evidence can be considered as hint (with some uncertainty) toward some system state. Based on one evidence E_j we assign a probability that it infers a certain hypothesis H_j. A *basic probability assignment (bpa)* is a mass function m which assigns beliefs in a hypothesis or as Shafer stated "the measure of belief that is committed exactly to *H*" [16]:

$$m : 2^{\Theta} \to [0, 1]. \tag{1}$$

This membership function *m* has to satisfy the following conditions:

$$m(\emptyset) = 0 \; and \; m(H) \geq 0, \forall H \subseteq \Theta \; and \qquad \sum_{H \subseteq \Theta} m(H) = 1 \quad . \tag{2}$$

At this point we have to underline the flexibility and advantages of this theory in contrast to the Bayesian approach, where we can only assign probabilities on single elements of Θ and not on elements of the powerset of the possible states. This theory gives us the opportunity to model uncertainty and the fact that some observations can distinguish between some system states, while they might not be able to provide any hints about others. For example, we might know that an evidence points to hypothesis $H = \theta_1, \theta_2$ with a high probability but on the same time it might provide no information (complete ignorance) whether the system is in θ_1 or θ_2. Furthermore it is crucial that the "Theory of Evidence" calculates the probability that the evidence supports a hypothesis rather than calculating the probability of the hypothesis itself (like the traditional probabilistic approach).

We define *Bel* as a *belief function* related to a hypothesis H:

$$Bel(H) = \sum_{B \subseteq H} m(B). \tag{3}$$

This definition says intuitively that a portion of belief committed to a hypothesis B must also be committed to any other hypothesis that it implies, i.e., to any $H \supseteq B$. A Belief function has the following properties:

$$Bel(\emptyset) = 0 \; and \; Bel(\Theta) = 1. \tag{4}$$

The *Plausibility* of H is defined as:

$$Pl(H) = \sum_{B \cap H \neq \emptyset} m(B) \tag{5}$$

and can be correlated to the doubt in the hypothesis H:

$$Doubt(H) = Bel(H^c) = 1 - Pl(H), \tag{6}$$

where H^c is the complement of H. Intuitively, this relation means that the less doubt we have in a hypothesis H the more plausible it is. Generally we can characterize $Bel(H)$ as a quantitative measure of all our supportive evidence and $Pl(H)$ as a measure of how incompatible our evidence is with H in terms of doubt (refuting evidence). The true belief in H lies in the interval $[Bel(H), Pl(H)]$. Our degree of ignorance is represented by the difference $Bel(H) - Pl(H)$.

The second important element of Dempster-Shafer theory is that it provides a rule to combine independent evidence E_1, E_2 into a single more informative hint:

$$m_{12}(H) = \frac{\sum_{B \cap C = H} m_1(B) m_2(C)}{\sum_{B \cap C \neq \emptyset} m_1(B) m_2(C)}. \tag{7}$$

Based on this formula we can combine our observations to infer the system state based on the values of belief and plausibility functions. In the same way we can incorporate new evidence and update our beliefs as we acquire new knowledge through observations.

"Theory of Evidence" makes the distinction between uncertainty and ignorance, so it is a very useful way to reason with uncertainty based on incomplete and possibly contradictory information extracted from a stochastic environment. It does not need "a priori" knowledge or probability distributions on the possible system states like the Bayesian approach and as such it is mostly useful when we do not have a model for our system. In comparison with other inference processes, like first order logic which assumes complete and consistent knowledge and exhibits monotonicity or probability theory which requires knowledge in terms of probability distributions, the "Theory of Evidence" has a definite advantage in a vague and unknown environment.

The "Theory of Evidence" from a computational point of view is in worst case exponential, because Dempster's rule of combination (Eq. 7) requires to find all pairs of sets B, C such that $B \cap C = H$ which is $o(2^{|\Theta| - |H|} \times 2^{|\Theta| - |H|})$. Thus it may be hard to compute in the general case, although some efficient algorithms for fast computation exist. Nevertheless for many practical applications with few focal elements, an exhaustive approach is still feasible.

B. Proposed Anomaly Detection System

Based on the Dempster-Shafer "Theory of Evidence" we propose a novel architecture to enable anomaly detection in cyber-physical systems. The proposed architecture, depicted in Figure 1, illustrates the collection of data from both the physical and cyber realms. The system fuses the knowledge that is collected from the reports of various sensors in order to infer the state of the system. One important aspect that should be emphasized is that our sensors do not only collect data, but they also provide a first level of detection. Their outputs are translated to basic probability assignments which are fused by the Dempster-Shafer inference engine.

As in any data fusion system, the performance of the implementation depends on the selected sensors. Possible cyber sensor types could include: TCP-SYN packet monitoring; UDP and ICMP packet monitoring; and router traffic monitoring. For the physical dimension we can expand the list with sensors monitoring physical parameters such as: pressure,

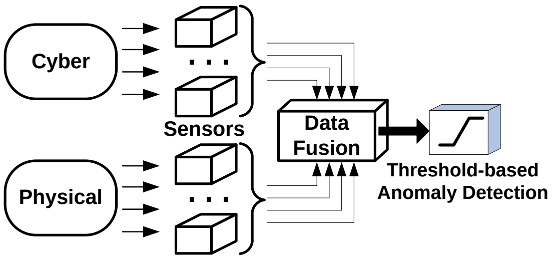


Fig. 1: Architecture of the proposed Anomaly Detection System

temperature, liquid level, and valve position. Most importantly, the flexibility of the "Theory of Evidence" allows engineers to expand this list with other application-specific sensors.

In the proposed system we use Θ to denote the set of all possible states of the system, also known as the *Frame of Discernment* in the terminology of the "Theory of Evidence". Each sensor has the ability to detect a specific set of attacks which can be expressed by defining a mass function *m* for 2 possible sets:

- the set H of states that the sensor can recognize or is sensitive to, for which $m(H)$ denotes the sensor's belief in the states from H;
- the set Θ as previously defined, for which $m(\Theta)$ denotes the degree of uncertainty associated to this sensor.

It follows from equation 2 that $m(H) + m(\Theta) = 1$. Based on these assumptions engineers can use the modeling power of "Theory of Evidence" to include expert knowledge about each sensors' detection ability. A simple guideline to help engineers define individual m-values is shown in Figure 2. The intuition behind this guideline is that although going over and under certain thresholds leads us towards a quite certain decision, in the interval between these low and high thresholds our beliefs should be treated with an increased uncertainty. Figure 2 shows two basic probability assignment possibilities. Figure 2 (a) defines one threshold interval ($[T_{low}, T_{high}]$) and can be applied in scenarios such as TCP-SYN-flooding attacks, where an increasing number of SYN-requests can lead to a DoS attack. In this case the level of uncertainty given by $m(\Theta)$ increases in between the two thresholds, denoting the sensor's uncertainty related to the value of $m(H)$. Figure 2 (b) defines two threshold intervals ($[T_{low}, T_{high}]$ and $[T'_{low}, T'_{high}]$) and can be applied in the physical dimension of NCIs, where parameters are usually bound to an interval, e.g., steam pressure. In this case the uncertainty appears in two different settings, as there are two threshold intervals. Assuming a typical sensor, moving the curve up and to the left yields a more sensitive sensor increasing possibly the false positive alarm rate. On the other hand a down or right movement makes the sensor less sensitive, increasing the false-negatives. By condensing the curve along the x-axis we move towards binary detection, whereas expanding it yields a sensor with greater uncertainty but more sensitivity.

Besides the aforementioned issues, it is of utmost importance that the detection engine takes into account missing evidence as well. In real scenarios communications packets relevant to the decision process might be delayed or most of the times dropped by networking hardware due to congestions caused by disruptive cyber attacks. Furthermore, a fully compromised sensor might not send any updates at all to the detection engine, which needs to be properly modeled and incorporated in the decision process. Within the proposed ADS we model missing evidence by dynamically increasing the degree of uncertainty $m(\Theta)$ for a given sensor. The detection engine takes into account the last value reported by a specific sensor, but with a dynamic uncertainty degree that changes according to its freshness. Consequently, we define a saturated linear increase of the uncertainty degree $m(\Theta)_\tau \in [\lambda_{min}, \lambda_{max}]$ for time freshness τ such that:

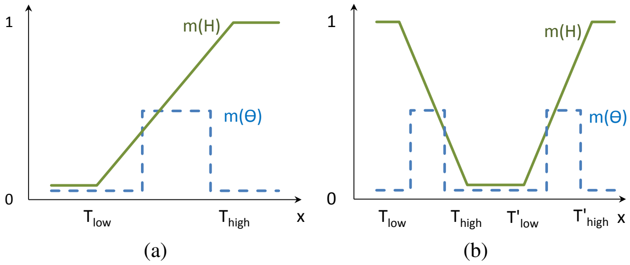


Fig. 2: Generic guidelines to define basic probability assignments: (a) Single -, and (b) dual - threshold interval

$$m(\Theta)_\tau = \frac{\lambda_{max} - \lambda_{min}}{t_{max} - t_{min}}(\tau - t_{min}) + \lambda_{min}, \tag{8}$$

where λ_{min} and λ_{max} are the minimum and maximum possible values for $m(\Theta)_\tau$, $[t_{min}, t_{max}]$ is the freshness interval on which the function is defined and τ is calculated as the difference between the system time and the time stamp of the last received value.

As a final note we mention that unlike in a purely cyber system, within the context of NCIs, the properties of the physical dimension open a different way for defining hypothesis. More specifically, since the interactions within industrial processes are governed by well-established laws of physics it is possible to infer process states that are not directly monitored. For instance, by simply measuring the pressure in a gas tank we could also infer the temperature of the gas knowing that there is a direct proportional dependency between the two. Consequently, this provides a powerful way to define a larger spectrum of hypothesis from a reduced set of physical sensors, which in turn will lead to more accurate detections.

C. System Implementation Details

The authors' previous experience [19] proves that real-time simulation can be a powerful technique in the context of experimental cyber security assessment of NCIs. Therefore, a prototype of the data fusion engine was developed in Matlab Simulink, since this is a general simulation environment

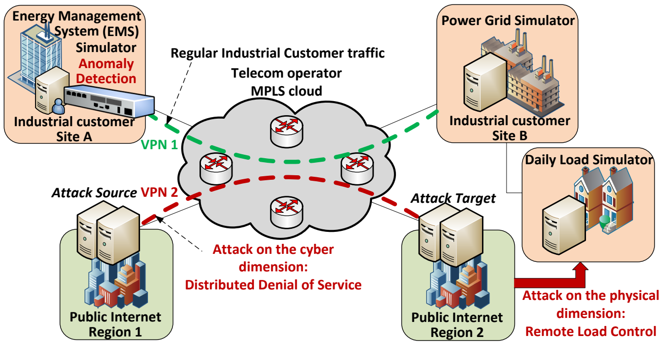

Energy Management System (EMS) Simulator
Anomaly Detection
Industrial customer Site A
Regular Industrial Customer traffic
Telecom operator MPLS cloud
Power Grid Simulator
Industrial customer Site B
VPN 1
Daily Load Simulator
Attack Source
VPN 2
Attack Target
Attack on the cyber dimension: Distributed Denial of Service
Public Internet Region 1
Public Internet Region 2
Attack on the physical dimension: Remote Load Control

Fig. 3: Experiment setup including a remotely controlled power grid and public traffic routed through an MPLS Telco communications infrastructure

for dynamic and embedded processes. From Simulink, we generate the corresponding 'C' code using Simulink Coder, which is then integrated into the framework developed in our previous work [19], [20]. Our previously developed framework combines emulation testbeds based on Emulab [21], [20] with real-time simulation in order to enable disruptive experiments on physical processes while ensuring a high fidelity of the cyber dimension. With this approach the model can interact in real-time with the rest of the system, it can receive data from real sensors and it can produce alerts according to the implemented data fusion decision algorithm.

Prototypes of cyber and physical sensors have been developed in the Python scripting language. Typical cyber sensors monitor network traffic throughput, a method also employed by related work [6], [7], while physical sensors process packets originating from control hardware in order to monitor the state of the physical dimension.

IV. Performance Evaluation

In order to verify the validity of the proposed Anomaly Detection System and to prove the superior performance of data fusion over state of the art, we have conducted a series of tests configured in a protected environment within our laboratory. We recreated the typical architecture of an NCI installation in which a simulated power grid is controlled remotely [22] (Figure 3) and we launched several attacks that test: (i) the detection of cyber anomalies; (ii) the detection of physical anomalies; and (iii) the detection of cyber or physical anomalies based on the combined evidence from the two dimensions.

A. Description of the Experimental Scenario

The experimental scenario consists of a remotely controlled power grid and several attacks causing severe telecommunications service degradation which propagates across Critical Infrastructures. As illustrated in Figure 3 we define two hypothetical sites. *Site A* runs a simplified model of an Energy Management System (EMS) [23] to ensure voltage stability. The EMS continuously monitors and adjusts the operational parameters of the power grid model (IEEE 39-bus New England system) running at *Site B*. The daily load imposed to our system derives from real data [24] and the intervention of the EMS is required to keep the grid stable.

TABLE I: Sensors with hypothesis and monitored parameters

Type	Count	Hypothesis	Monitored
Physical	21	H_i^1 ={PHYSICAL-Anomaly} H_i^2 ={CYBER-Anomaly, NORMAL} $H_i^3 = \Theta$	Voltages
Cyber	1	H_{22}^1 ={CYBER-Anomaly} H_{22}^2 ={PHYSICAL-Anomaly, NORMAL} $H_{22}^3 = \Theta$	Throughput

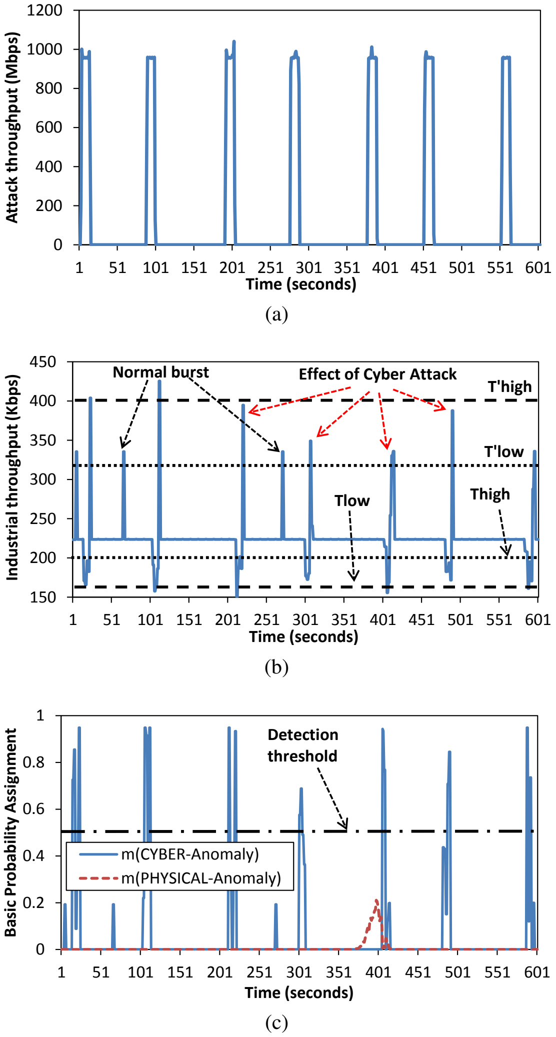
To provide a realistic communications infrastructure between the EMS and power grid simulator we assumed that the service provider uses an MPLS (Multi Protocol Label Switching) network. MPLS is a protocol that telco operators already use to replace older implementations based on Frame Relay and Asynchronous Transfer Mode (ATM) [25]. We created a minimal MPLS network with four Cisco 6503 routers, on which we defined two MPLS Virtual Private Networks (VPNs). VPN 1 acted as a protected virtual circuit between *Site A* and *Site B*, an approach that is usually followed by telco operators to isolate customer traffic. Since telco operators route diverse traffic, e.g., public Internet traffic, through the same MPLS cloud, we used VPN 2 to create a virtual circuit between two different "public" regions.

The Anomaly Detection System monitors the network traffic and the state of the grid by processing packets passing through the edge router of *Site A*. The detection engine detects both cyber and physical anomalies by monitoring network traffic throughput and power grid voltages.

B. System States and Threshold Intervals

In order to illustrate the applicability of the approach we defined the following system states: Θ ={CYBER-Anomaly, PHYSICAL-Anomaly, NORMAL}. Based on the proposed guidelines for basic probability assignments and several trial and error procedures for tuning the system parameters we have designed two types of sensors, as depicted in Table I. 21 *Physical* sensors monitor voltages on 21 different substations. Physical sensors can clearly detect a physical anomaly by inspecting voltage levels, but cannot distinguish between a cyber anomaly and a normal state. On the other hand, we defined one *Cyber* sensor that monitors the network throughput and can clearly detect a cyber anomaly, but it is unable to say anything about a physical anomaly or to clearly identify a normal system.

For both sensor types we defined two threshold intervals since the monitored parameters from both the cyber and the physical dimensions must remain within well-defined boundaries. For physical sensors we used $[T_{low}, T_{high}] = [0.86, 0.92]$ and $[T'_{low}, T'_{high}] = [1.08, 1.12]$ (in p.u.), since operators usually run the electrical grid with voltage levels that range from 0.9 p.u. to 1.1 p.u.. For the cyber sensor we used $[T_{low}, T_{high}] = [160, 200]$ and $[T'_{low}, T'_{high}] = [320, 400]$

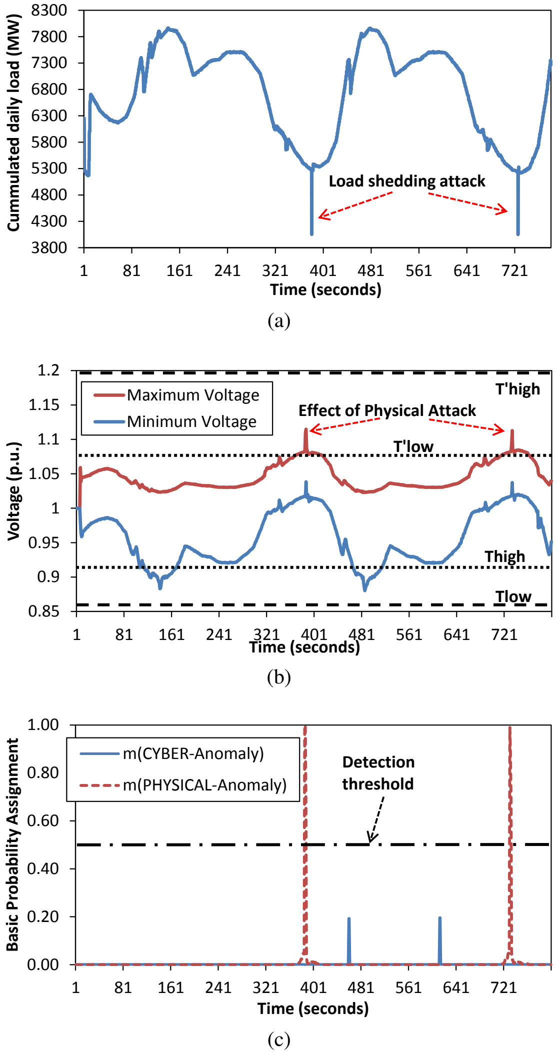


(a)

(b)

(c)

Fig. 4: Detection of cyber attacks: (a) Cyber attack throughput, (b) effect on the industrial customer throughput, and (c) the detection of the attack



(a)

(b)

(c)

Fig. 5: Detection of attacks on physical processes: (a) Physical attack through dynamic load shedding, (b) effect on power grid voltages, and (c) the detection of the attack

(in Kbps), since the average measured throughput was of 224Kbps, but at the same time we measured random bursts reaching up to 336Kbps that were caused by delayed replies originating from the power grid simulator.

C. Detection of Cyber Anomalies

We launched several random bandwidth consuming DDoS attacks using typical tools such as TCPReplay and Scapy in VPN 2 and we measured their effect on the industrial network traffic in VPN 1. Each attack ran for 10 seconds and consumed close to 1Gbit/s of communications bandwidth (see Figure 4 (a)).

The effect of the attack on the network traffic in VPN 1 can be clearly seen in Figure 4 (b), where we observe two types of changes in the network throughput: (i) normal bursts, caused by delayed replies from the power grid simulator, and (ii) network bursts caused by the attack. In the later case we see that bursts are preceded by a reduction of the network throughput, which is the expected behavior for congested links.

The changes in the network throughput are successfully detected by the proposed ADS. As shown in Figure 4 (c) for a threshold of 0.5 the detection engine can accurately identify cyber anomalies and can filter out normal network traffic bursts. For the first case the CYBER-Anomaly Basic Probability Assignment (BPA) increases well-above the 0.5 threshold and reaches a value of 0.92, which is clearly a sign of a cyber anomaly. Normal traffic bursts are also visible in this figure. However, since in these cases the BPA increases only up to 0.2, they are not detected as cyber anomalies.

As a final note we should underline the fact that the implemented attack also illustrates that MPLS VPNs alone cannot ensure a proper isolation between virtual circuits. For this purpose Telco operators usually implement well-established mechanisms such as Quality of Service (QoS) and network

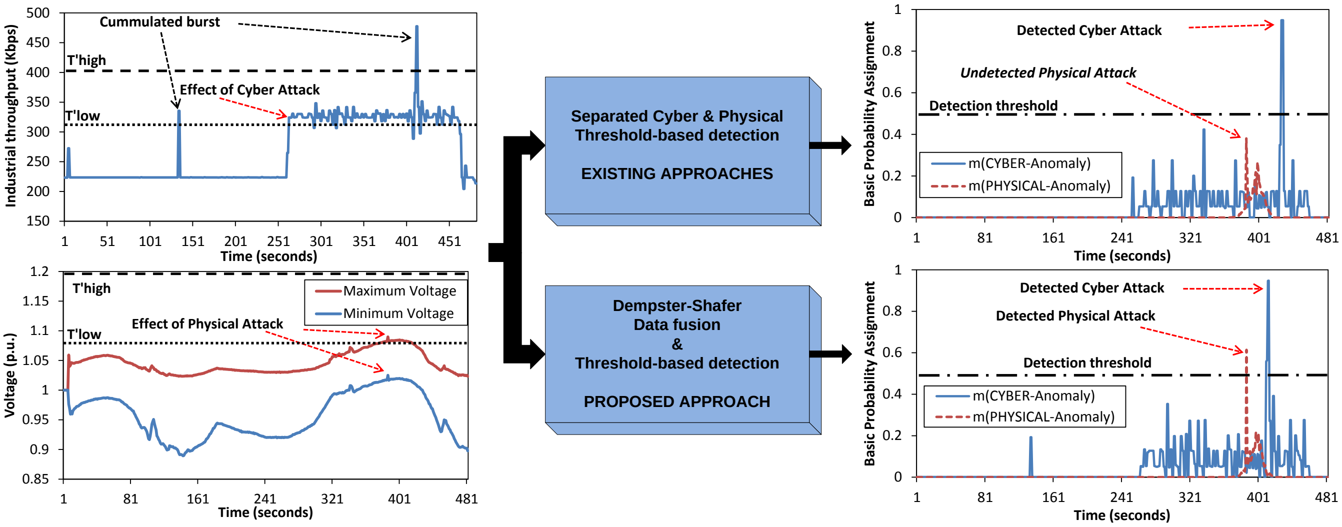


Fig. 6: Separate cyber & physical anomaly detection vs data fusion-based anomaly detection

traffic engineering. Nevertheless, such protective measures are not compulsory, e.g., through policies and regulation. Therefore, the severe risks that are involved if such protective measures are not implemented were clearly demonstrated by this particular phase in our experiment which highlighted the potential impact of ICT disruptions. More information on this topic and similar studies can be found here [22], [26], [27].

D. Detection of Physical Anomalies

The attack on the physical process, i.e., the simulated power grid, exploits the remote control capabilities of control hardware. Within our laboratory we recreated a hypothetical scenario in which the attacker forges several control packets and injects them into the control network of several substations in order to dynamically shed loads at a specific time of day. In this procedure we used legitimate packets (similar to those used by the EMS) to trigger load shedding for short time periods, i.e., 0.2s, at five different substations.

As shown in Figure 5 (a) the daily load exhibits significant loss of load, which triggers over-voltages in the electricity grid, shown in Figure 5 (b). The effects of such an attack are obvious. Excessively high voltage levels can trigger protective mechanisms, e.g., circuit breakers, to disconnect devices from the grid. However, they can also cause short-circuits and can damage vital equipment, which could lead to cascading failures propagating to the entire grid.

The implemented ADS monitors the state of the physical process and successfully detects the attacks on it. As shown in Figure 5 (c), the dynamic load shedding attack causes the PHYSICAL-Anomaly BPA to increase up to 1, which is a clear indication of a physical anomaly. Although the cyber attack is not detected, since it is launched from a separate network that is not monitored (see the experimental setting in Figure 3), the ADS is able to trigger alarms caused by a physical anomaly.

E. Fusion vs Separated Cyber-Physical Anomaly Detection

Based on the results presented so far we can clearly state that the performance of the proposed ADS depends on the quality of sensors and that of additional ADSs that are supplying evidence to the data fusion engine. However, as already stated, the added-value of data fusion over existing ADSs is that it can combine evidence from various sensors and provide an aggregated view of the system. In the context of cyber-physical systems the ADS fuses together evidence from the cyber and physical dimensions and, as shown in the remaining of this section, it provides a more effective detection of anomalies.

For illustration purposes we ran a slightly modified version of the dynamic load shedding attack mentioned in the previous section. The attack was launched in VPN 1 and was implemented to cause load shedding in only one substation, thus providing a more "subtle" version of the previous attack that would not trigger alarms with regular ADSs.

As shown on the left side of Figure 6, the industrial traffic throughput increases to an average of 330Kbps, while voltages exhibit a barely visible increase. On the right side of Figure 6 we can see that existing approaches, i.e., that separate the cyber realm from the physical, might effectively detect cyber attacks. However, attacks on the physical process will not be detected. Conversely, the proposed ADS fuses the evidence from the cyber and physical dimensions, which leads to the detection of the physical attack as well. As shown in Figure 6, in this case the PHYSICAL-Anomaly BPN increases above the detection threshold up to 0.6, which is clearly a sign of a physical anomaly. It should be noted that in this particular case the cyber sensor provides a significant amount of evidence for detecting the physical anomaly. This is mainly due to hypothesis H_{22}^{2}, which states that in case there is no cyber anomaly the sensor cannot distinguish between a physical anomaly and a normal state. Consequently, this brings an additional proof that there *might* be a physical anomaly, which is then fused together with the evidence provided by physical sensors and leads to the aforementioned detection.

The two attack scenarios presented in this section proved that data fusion can be a good candidate for implementing anomaly detection in complex multi-dimensional systems such as NCIs. Furthermore, it should be noted that although the study was limited to these two types of attacks, i.e., Denial of Service and packet forging, the proposed ADS can also detect other, possibly unknown attacks, i.e., zero-day attacks, on the cyber or physical dimension of NCIs. This is mainly due to the specific characteristics of ADSs that, opposed to signature-based intrusion detection systems, perform system monitoring in search for anomalous states that could trigger alarms.

V. Conclusion

We proposed a novel approach to detect anomalies in cyber-physical systems by combining reports of various cyber and physical sensors using Dempster-Shafers' "Theory of Evidence". In the proposed Anomaly Detection System (ADS) sensors act as autonomous agents that send periodic reports to a central unit that fuses together evidence from the cyber and physical dimensions and provides a unified view of the entire system. Consequently, the combined knowledge from the majority of sensors can effectively filter malicious/missing reports from compromised/defective sensors and can indicate an increased belief of an anomaly state. The validity and effectiveness of the proposed ADS were demonstrated by experimental results conducted on a simulated power grid and communications infrastructure based on real Multi Protocol Label Switching (MPLS) Virtual Private Networks (VPNs). The results also confirmed the superior performances of data fusion-based ADS over separated cyber and physical ADS. As future work we intend to continue to enhance the proposed ADS with automated parameter tuning techniques and to integrate existing ADSs with the proposed detection engine.

References

[1] D. Fidler, "Tinker, Tailor, Soldier, Duqu: Why cyberespionage is more dangerous than you think," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 1, pp. 28–29, 2012.

[2] B. Reaves and T. Morris, "Analysis and mitigation of vulnerabilities in short-range wireless communications for industrial control systems," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 34, pp. 154 – 174, 2012.

[3] I. Nai Fovino, A. Carcano, M. Masera, and A. Trombetta, "An experimental investigation of malware attacks on SCADA systems," *International Journal of Critical Infrastructure Protection*, vol. 2, no. 4, pp. 139–145, 2009.

[4] T. Chen and S. Abu-Nimeh, "Lessons from Stuxnet," *Computer*, vol. 44, no. 4, pp. 91–93, april 2011.

[5] E. Pleijsier, "Towards anomaly detection in scada networks using connection patterns," in *18th Twente Student Conference on IT*, 2013.

[6] R. Barbosa, R. Sadre, and A. Pras, "Towards periodicity based anomaly detection in SCADA networks," in *17th IEEE Conference on Emerging Technologies & Factory Automation (ETFA 2012)*, 2012, pp. 1–4.

[7] I. Garitano, C. Siaterlis, B. Genge, R. Uribeetxeberria, and U. Zurutuza, "A method to construct network traffic models for process control systems," in *17th IEEE Conference on Emerging Technologies & Factory Automation (ETFA 2012)*, 2012, pp. 1–8.

[8] N. Svendsen and S. Wolthusen, "Using physical models for anomaly detection in control systems," in *Critical Infrastructure Protection III*, ser. IFIP Advances in Information and Communication Technology, C. Palmer and S. Shenoi, Eds. Springer Berlin Heidelberg, 2009, vol. 311, pp. 139–149.

[9] ——, "Modeling and detecting anomalies in scada systems," in *Critical Infrastructure Protection II*, ser. The International Federation for Information Processing, M. Papa and S. Shenoi, Eds. Springer US, 2009, vol. 290, pp. 101–113.

[10] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '11. New York, NY, USA: ACM, 2011, pp. 355–366.

[11] I. Nai Fovino, A. Coletta, A. Carcano, and M. Masera, "Critical state-based filtering system for securing scada network protocols," *Industrial Electronics, IEEE Transactions on*, vol. 59, no. 10, pp. 3943–3950, 2012.

[12] A. Carcano, A. Coletta, M. Guglielmi, M. Masera, I. Fovino, and A. Trombetta, "A multidimensional critical state analysis for detecting intrusions in scada systems," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 2, pp. 179–186, 2011.

[13] I. Nai Fovino, A. Carcano, T. De Lacheze Murel, A. Trombetta, and M. Masera, "Modbus/dnp3 state-based intrusion detection system," in *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on*, 2010, pp. 729–736.

[14] M. Raciti and S. Nadjm-Tehrani, "Embedded cyber-physical anomaly detection in smart meters," in *7th International Conference on Critical Information Infrastructures Security*, 2012.

[15] M. Levorato and U. Mitra, "Fast anomaly detection in SmartGrids via sparse approximation theory," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, 2012, pp. 5–8.

[16] G. Shafer, "A mathematical theory of evidence," *Princeton University Press*, 1976.

[17] J. Kohlas and P. Monney, "Theory of evidence - a survey of its mathematical foundations, applications and computational analysis," *ZOR-Mathematical Methods of Operations Research*, vol. 39, pp. 35–68, 1994.

[18] K. Tomsovic and B. Baer, "Fuzzy information approaches to equipment condition monitoring and diagnosis," in *IEEE Press*, 1998, pp. 59–84.

[19] B. Genge, C. Siaterlis, I. Nai Fovino, and M. Masera, "A cyber-physical experimentation environment for the security analysis of networked industrial control systems," *Computers & Electrical Engineering*, vol. 38, no. 5, pp. 1146 – 1161, 2012.

[20] C. Siaterlis, A. Garcia, and B. Genge, "On the use of Emulab testbeds for scientifically rigorous experiments," *IEEE Communications Surveys and Tutorials*, vol. PP, no. 99, pp. 1–14, 2012.

[21] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar, "An integrated experimental environment for distributed systems and networks," in *Proceedings of the 5th Symposium on Operating Systems Design and Implementation*, 2002, pp. 255–270.

[22] B. Genge and C. Siaterlis, "Analysis of the effects of distributed denial-of-service attacks on MPLS networks," *International Journal of Critical Infrastructure Protection*, no. 0, pp. –, 2013.

[23] T. Tuan, J. Fandino, N. Hadjsaid, J. Sabonnadiere, and H. Vu, "Emergency load shedding to avoid risks of voltage instability using indicators," *Power Systems, IEEE Transactions on*, vol. 9, no. 1, pp. 341–351, feb 1994.

[24] M. Manera and A. Marzullo, "Modelling the load curve of aggregate electricity consumption using principal components," *Environ. Model. Softw.*, vol. 20, no. 11, pp. 1389–1400, Nov. 2005.

[25] IBM and Terna, "Cisco and IBM provide high-voltage grid operator with increased reliability and manageability of its telecommunication infrastructure," *IBM Case Studies*, 2007.

[26] C. Deccio, "Maintenance, mishaps and mending in deployments of the domain name system security extensions (DNSSEC)," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 2, pp. 98–103, 2012.

[27] S. Rahimi and M. Zargham, "Analysis of the security of VPN configurations in industrial control environments," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 1, pp. 3–13, 2012.