Theory of Evidence-Based Automated Decision Making in Cyber-Physical Systems

Christos Siaterlis and Béla Genge
Joint Research Centre, European Commission
Institute for the Protection and Security of the Citizen
Via E. Fermi, 2749, Ispra (VA), 21027, Italy,
Email: {christos.siaterlis, bela.genge}@jrc.ec.europa.eu

***Abstract*—The Smart Grid is a complex cyber-physical system that is evolving rapidly from a relatively isolated to an open and diverse environment. Within this context, enhancing the security of the future Smart Grid becomes a major priority. In this paper we introduce the use of data fusion for automated decision making in cyber-physical systems such as the Smart Grid. One of the most important applications of decision making is in the field of anomaly detection. This can enable the detection of attacks in cyber-physical systems without requiring a complete description of the physical process. The novelty of our approach is that it combines reports of various cyber and physical sensors, rather than focusing on either one single metric, or one singe realm, as was the case of similar techniques. Based on the proposed architecture we implement a new cyber-physical anomaly detection system. We show that data fusion is much more effective if it combines both cyber and physical realms, rather than focusing on the two realms separately.**

***Index Terms*—Cyber-Physical systems, Smart Grid, data fusion, anomaly detection systems.**

I. Introduction

The Smart Grid is a complex cyber-physical system that is evolving rapidly from a relatively isolated to an open and diverse environment. The adoption of Information and Communication Technologies (ICT) has led to cost optimization as well as greater efficiency, flexibility and interoperability between components. It is expected that Supervisory Control And Data Acquisition (SCADA) systems will provide the communications architecture for substation and distribution automation, advanced metering and home area networking applications for the future Smart Grid [1]. This will expose many challenging problems in the security of Smart Grid as current SCADA systems are exposed to significant cyber-threats; a fact that has been highlighted by many studies [2], [3]. For example, the recently discovered Stuxnet worm [4] is the first malware that is specifically designed to attack SCADA systems. Its ability to reprogram the logic of control hardware in order to alter physical processes demonstrated how powerful such threats can be; it has served as a wakeup call for the international security community.

Within this context, enhancing the security of the future Smart Grid becomes a major priority. Consequently, we find a growing need for tools and methods that identify possible security issues in the architecture of the future Smart Grid. In this paper we support this direction by introducing the use of data fusion for automated decision making in cyber-physical systems such as the Smart Grid. One of the most important applications of decision making is in the field of anomaly detection. This can enable the detection of attacks in cyber-physical systems without requiring a complete description of the physical process. Other applications range from modeling the behavior of plant operators to cyber-physical Intrusion Detection Systems (IDS). In our previous work [7] we have shown that data fusion can be successfully applied to the implementation of Distributed Denial of Service (DDoS) detection systems. This paper extends our previous work with elements from the physical realm and proposes an architecture to support decision making in complex cyber-physical systems such as the Smart Grid.

The mathematical foundation for the data fusion process is provided by the Dempster-Shafer "Theory of Evidence" (D-S). D-S enables the combination of evidence generated from multiple *sensors*, e.g. basic detection elements. 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. Sensors act as thin autonomous agents which collaborate by sharing their beliefs about the observed attributes. From our perspective, the cyber-physical system is 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 (e.g. Programmable Logical Controllers - PLCs) or physical sensors.

The novelty of our approach is that it combines reports of various cyber and physical sensors, rather than focusing on either one single metric, or one singe realm, as was the case of similar techniques. Moreover, based on the proposed architecture we implement a new cyber-physical anomaly detection system. We show that data fusion is much more effective if it combines both cyber and physical realms, rather than focusing on the two realms separately.

The paper is structured as follows. After a short overview of related works in Section II and of the Dempster-Shafer "Theory of Evidence" in Section III, the proposed architecture

is presented in Section IV. Based on this architecture we propose a new Anomaly Detection System in Section V and we conclude in Section VI.

II. Related Work

We begin by mentioning the work of Svendsen and Wolthusen [11] that use an explicit model of a SCADA system for anomaly detection. The detection process is enhanced by using feedback control theory to predict future values and ultimately detect physical anomalies in the system. A similar approach has been developed by Cárdenas, *et al.* [12], 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 the system. The main disadvantage of these approaches is the requirement of a complete model for the physical process. In contrast, our approach does not require such a model and it can be used not only on physical systems but also on cyber-physical systems.

In our previous work [7] we have also experimented with the Dempster-Shafer "Theory of Evidence" to develop an Intrusion Detection System for cyber systems. The proposed architecture monitored parameters such as SYN, UDP and ICMP packets but also received regular reports on network traffic from switches and routers. This paper is an extension of our previous work in [7] and it proposes data fusion for cyber-physical systems. Another approach that monitors cyber parameters has been proposed by Nai Fovino, *et al.* [14], where the authors defined a set of rules to describe the Modbus/DNP3 commands that can cause the system to go into a critical state. The proposed approach can effectively detect anomalies from the cyber realm (e.g. invalid Modbus/DNP3 packets), however, it does not monitor the physical realm. Consequently, physical anomalies that have different causes than Modbus/DNP3 packets can not be detected.

Finally, we mention an approach that focuses on cyber and physical realms, proposed by Zimmer, *et al.* [13]. The main assumption of the authors is that the system consists of hardware running real-time Operating Systems (OSs) where the execution of applications can be estimated at design time. Zimmer, *et al.* propose to enhance the OS with instrumentation mechanisms to detect anomalies in the system based on application execution timings. Although this procedure can be effective for real-time OSs and malware that is not designed to overwrite the OS kernel, it can not detect much more sophisticated attacks such as the recently reported Stuxnet that was able to completely rewrite the control logic code. In contrast, our approach is not linked to a specific hardware or software and is able to detect anomalies caused by much more sophisticated attacks.

III. Dempster-Shafer 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 in 1976 [8]. It can be viewed either from a probabilistic or an axiomatic point of view and all these approaches are concisely surveyed in [9]. 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" [10].

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*" [8]:

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

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

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

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.

IV. Proposed System Architecture

Based on the Dempster-Shafer "Theory of Evidence" we propose a novel architecture to enable decision making in cyber-physical systems. The proposed architecture, depicted in Fig. 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. In our previous work [7] we have focused on a DDoS system where we have identified several sensor types that could be used for the cyber realm, such as: TCP-SYN packet monitoring sensors; UDP and ICMP packet monitoring sensors; and router traffic monitoring sensors. For the physical realm we can expand the list with sensors monitoring physical parameters, such as: pressure sensors; temperature sensors; liquid level sensors; and valve positions. Most importantly, the flexibility of the "Theory of Evidence" allows engineers to expand this list with other application-specific sensors.

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. Fine-tuning each sensor and translating their measurements to 'basic probability assignments' (bpa's) is not trivial and it might require several trials to be able to express beliefs about the state of the system. Nevertheless, as stated in our previous work [7], administrators can bypass this problem by using a supervised learning approach and inserting a minimal neural network at the sensor level.

A simple guideline to help engineers define individual m-values is shown in Fig. 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. Fig. 2 shows two basic probability assignment possibilities. Fig. 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 $m(H)$. Fig. 2 (b) defines two threshold intervals ($[T_{low}, T_{high}]$ and $[T'_{low}, T'_{high}]$) and can be applied in the physical realm, where parameters are usually bound to an interval (e.g. steam pressure, liquid level). In this case the uncertainty appears in two different settings, as there are two threshold intervals.

The main novelty of the proposed architecture is that it fuses together the cyber and the physical realms to provide an overall view of the system state. As shown later in the next section,

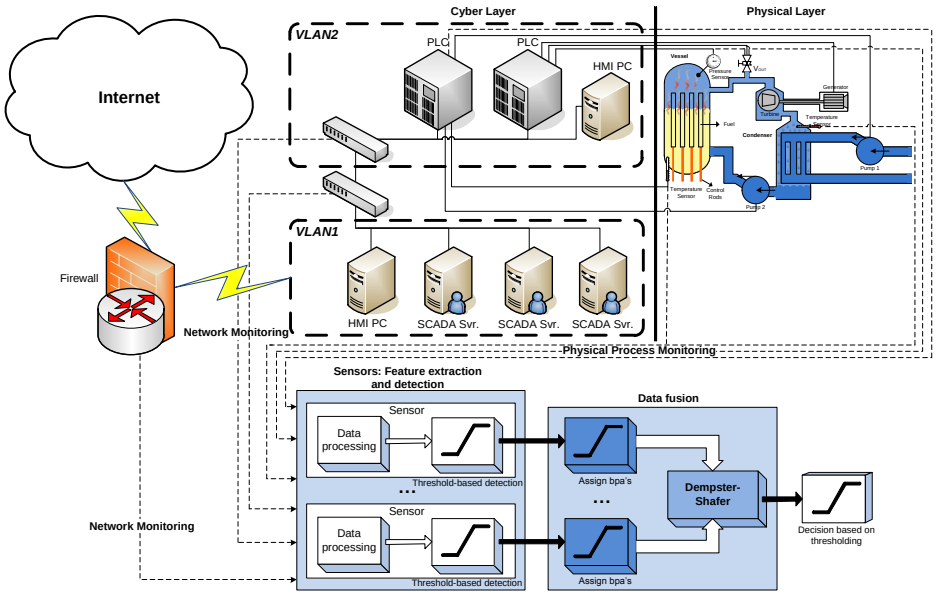
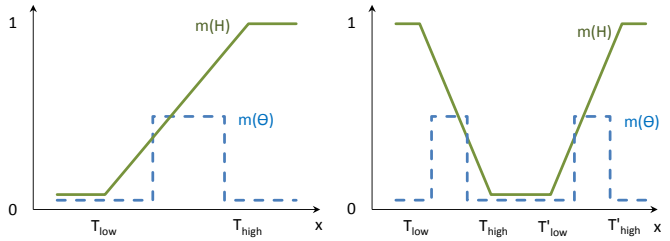


Fig. 1: Proposed system architecture



(a) (b)

Fig. 2: Generic guidelines to define basic probability assignments: (a) Single threshold interval (b) Dual threshold interval

the added value of this procedure is that even if one sensor fails to report the state of the system, combined knowledge gathered from other sensors can guide the operator to take a decision.

V. Towards a New Cyber-Physical Anomaly Detection System

In this section we present a new anomaly detection system (ADS) for cyber-physical systems, based on the proposed system architecture. The ADS monitors parameters from both the cyber and the physical realms. From the cyber realm it monitors parameters such as TCP-SYN requests, UDP traffic, ICMP traffic and Modbus packet traffic, while from the physical realm it monitors three parameters of a Boiling Water Power Plant: boiler steam pressure, water level and generated electricity. The monitored physical parameters are the ones found in the model of a 160MW oil-fired electric power plant based on the Sydsvenska Kraft AB plant in Malmö, Sweden,

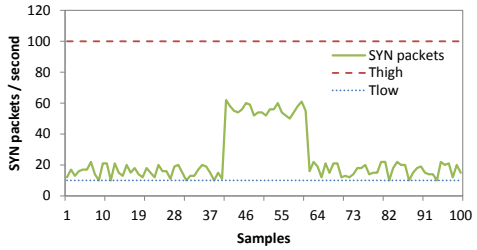
TABLE I: Sensors with hypothesis and monitored parameters

Sensor	Hypothesis	Monitored parameters
Sensor 1	H_1^1 ={CYBER-Anomaly} H_1^2 ={PHYSICAL-Anomaly, NORMAL} $H_1^3 = \Theta$	TCP-SYN requests, UDP traffic and ICMP traffic
Sensor 2	H_2^1 ={PHYSICAL-Anomaly} $H_2^2 = \Theta$	Modbus packet traffic
Sensor 3	H_3^1 ={PHYSICAL-Anomaly} H_3^2 ={CYBER-Anomaly, NORMAL} $H_3^3 = \Theta$	Steam pressure, water level and generated electricity

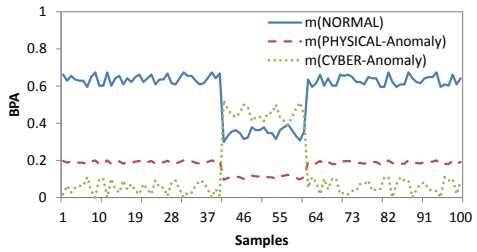
developed by Bell and Åström [6].

In our simplified implementation 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 3 sensors, depicted in Table I. The first sensor monitors three parameters from the cyber realm, based on which it can detect a cyber anomaly, but it can not differentiate between a physical anomaly and a normal state of the system. The second sensor is able to detect a physical anomaly by inspecting Modbus packets, however, it is not able to say anything about the state of the system if there is no anomaly related to Modbus packets. Finally, the third sensor can clearly detect a physical anomaly, but it can not detect a cyber anomaly, as it only receives information from physical process sensors.

In the remaining of this section we show that one of the main added values of our proposals is that even if one sensor fails to detect a cyber or physical anomaly, combined knowledge gathered from other sensors clearly indicates an increased belief of an anomaly state. For this purpose we have developed

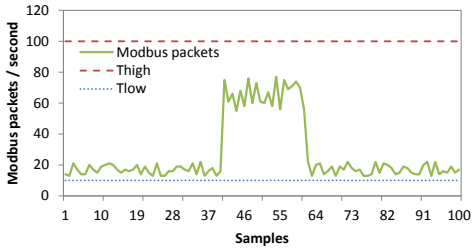


(a)

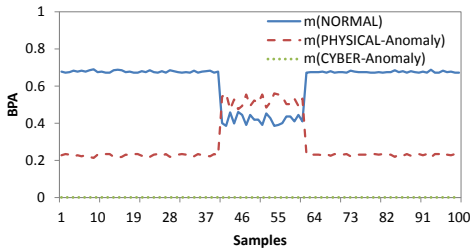


(b)

Fig. 3: SYN-flood attack: (a) Number of SYN packets (b) BPA function values



(a)



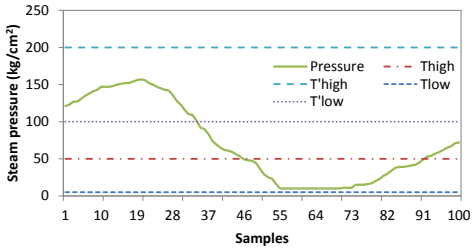
(b)

Fig. 4: Modbus packet attack: (a) Number of Modbus packets (b) BPA function values

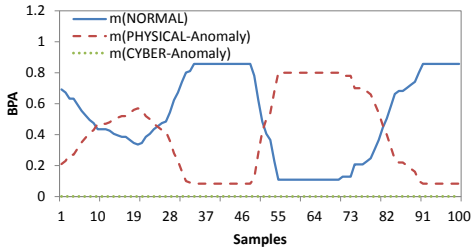
a prototype of the proposed sensors and inference engine in Matlab Simulink. One of the main reasons for choosing Matlab Simulink was that implementations can benefit from all the advantages of a well-established modeling tool. Moreover, further developments can enable Simulink models to interact in real time with other software components through Matlab Real Time Workshop, as shown in our previous work [5].

The results from Fig. 3 (a) and (b) illustrate the detection of a cyber anomaly, where we considered a detection interval for SYN packets of $[T_{low}, T_{high}] = [10, 100]$. As shown in Fig. 3 (a), we increased the number of SYN packets / second from 20 to 60 in order to simulate a SYN flood attack. The implemented engine is able to detect this anomaly and classifies it as a CYBER-Anomaly, as shown in Fig. 3 (b). Going further, we also simulated an attack with Modbus packets, by increasing the number of Modbus packets from 20 to 70, as shown in Fig. 4 (a). In this case we used the same detection interval of $[T_{low}, T_{high}] = [10, 100]$ and the system was able to detect a PHYSICAL-Anomaly. Finally, we simulated an attack on the physical process by increasing the steam pressure from 120 kg/cm^2 to 160 kg/cm^2 and later on decreasing it to 5 kg/cm^2, as shown in Fig. 5 (a). We have used two steam pressure anomaly detection intervals, one of $[T_{low}, T_{high}] = [5, 50]$ and the other one of $[T'_{low}, T'_{high}] = [100, 200]$. By excessively increasing or decreasing the pressure we simulate an attack on the physical process. This anomaly is detected by the proposed system, as shown in Fig. 5 (b). Here, the value of m(PHYSICAL-Anomaly) exceeds the value of m(NORMAL) for the samples where a physical anomaly is present.

These scenarios illustrated the functionality of the implemented data fusion engine and the applicability of the proposed approach for cyber-physical anomaly detection. However, in our discussion we have not mentioned other factors such as false positive or false negative detection rates because our intention was only to illustrate one possible application for the proposed architecture. The full implementation of an entire ADS requires deeper analysis of sensor design and a much more sophisticated sensor tuning, aspects that were not the focus of this paper, but are considered to be further possible developments.



(a)



(b)

Fig. 5: Boiler pressure attack: (a) Steam pressure (b) BPA function values

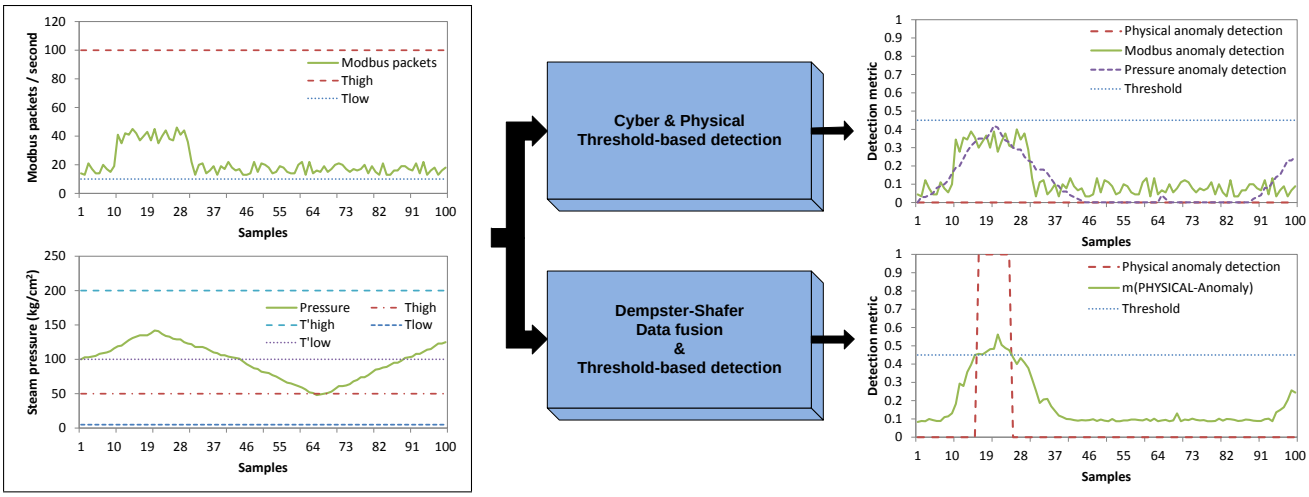


Fig. 6: Separate cyber & physical anomaly detection vs data fusion-based anomaly detection

The previous scenarios have demonstrated the applicability of our proposal, but they do not show the main advantage of cyber-physical data fusion-based detection systems over separate cyber and physical anomaly detection systems. For this purpose we have considered two parameters: Modbus packet count and steam pressure, with a detection threshold of 0.45. The Modbus packet count is increased from 20 to 50, at the same time increasing the steam pressure from 100 kg/cm^2 to 140 kg/cm^2. As shown in Fig. 6, by using separate threshold-based detections for the physical and cyber realms, the detection engine concludes that there is no attack on the system, although there are both cyber and physical anomalies present in the system. In contrast, by using the proposed data fusion engine, the evidence from the two realms accumulates, leading to an increase in the value of m(PHYSICAL-Anomaly) above the threshold. This example demonstrates the power of cyber-physical data fusion and provides a clear motivation of our proposal.

VI. Concluding Remarks

In this paper we have presented a new method to support automated decision making in cyber-physical systems such as the Smart Grid. The main novelty of the proposal is that it combines reports of various cyber and physical sensors using Dempster-Shafers' "Theory of Evidence". In our proposal sensors act as autonomous agents that send periodic reports to a central unit that fuses together evidence from the cyber and physical realms to provide a unified view of the system. The applicability of the approach has been demonstrated by the development of a new Anomaly Detection System (ADS) in Matlab Simulink. Simulation results show that one of the main added values of our proposals is that even if one sensor fails to detect a cyber of physical anomaly, combined knowledge gathered from other sensors can indicate an increased belief of an anomaly state. Nevertheless, further research must be conducted in the direction of sensor design and parameter tuning. For this purpose well-established methods from the field of Neural Networks could also be considered.

References

[1] Pike Research's report, "Smart Grid Communications Architecture," April, 2011.

[2] I. Nai Fovino, A. Carcano, M. Masera, 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.

[3] S. East, J. Butts, M. Papa, and S. Shenoi, "A Taxonomy of Attacks on the DNP3 Protocol," in *Proc. IFIP Advances in Information and Communicaiton Technology*, Vol. 311, pp. 67–81, 2009.

[4] The Symantec Stuxnet Dossier, 2010, http://www.wired.com /images_blogs/threatlevel/2010/11/w32_stuxnet_dossier.pdf

[5] B. Genge, I. Nai Fovino, C. Siaterlis, and M. Masera, "A Framework for Analyzing Cyber-Physical Attacks on Networked Industrial Control Systems," IFIP International Conference on Critical Infrastructure Protection, To Appear, 2011.

[6] R.D. Bell, K.J. Åström, "Dynamic models for boiler-turbine alternator units: data logs and parameter estimation for a 160MW unit," Lundt Institute of Technology, Report TFRT–3192, Sweden, 1987.

[7] C. Siaterlis, and V. Maglaris, "One step ahead to multisensor data fusion for DDoS detection," *Journal of Computer Security - Special issue on security track at ACM symposium on applied computing 2004*, Vol. 13, Issue 5, pp. 779–806, 2005.

[8] G. Shafer, "A Mathematical Theory of Evidence," Princeton University Press, Princeton, 1976.

[9] 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.

[10] K. Tomsovic and B. Baer, "Fuzzy information approaches to equipment condition monitoring and diagnosis," *Electric Power Applications of Fuzzy Systems*, IEEE Press, pp. 59-84, 1998.

[11] N. Svendsen and S. Wolthusen, "Using Physical Models for Anomaly Detection in Control Systems," *IFIP Advances in Information and Communication Technology*, Vol. 311/2009, pp. 139–149, 2009.

[12] A. Cárdenas, S. Amin, Z.S. Lin, Y.L. Huang, Chi-Y. Huang, and S. Sastry, "Attacks Against Process Control Systems: Risk Assessment, Detection, and Response," *in Proc. 6th ACM Symposium on Information, Computer and Communications Security*, pp. 355-366, 2011.

[13] C. Zimmer, B. Bhat, F. Mueller, and S. Mohan, "Time-Based Intrusion Detection in Cyber-Physical Systems," *in Proc. 1st ACM/IEEE International Conference on Cyber-Physical Systems*, 2010.

[14] I. Nai Fovino, A. Carcano, T. De Lacheze Murel, A. Trombetta, M. Masera, "Modbus/DNP3 State-Based Intrusion Detection System," *24th IEEE International Conference on Advanced Information Networking and Applications*, pp.729–736, 2010.