Last Line of Defense: A Novel IDS Approach Against Advanced Threats in Industrial Control Systems

Mark Luchs[1] and Christian Doerr[2]

[1] Delft University of Technology
Department of Offshore and Dredging Engineering
Mekelweg 4, 2628CD Delft, The Netherlands
Email: m.luchs@tudelft.nl
[2] Delft University of Technology
Cybersecurity Group
Mekelweg 4, 2628CD Delft, The Netherlands
Email: c.doerr@tudelft.nl

Abstract. Industrial control systems are becoming increasingly interconnected, and with it their vulnerability to malicious actors. While intrusion detection systems are suited to detect network-based attacks, they remain unable to detect more sophisticated attacks against control systems, for example a compromise of the PLCs. This paper makes the case that the evolving landscape of threats such as the Stuxnet malware requires an alternative approach to intrusion detection in industrial control systems. We argue that effective control of such advanced threats needs to happen in the last link of the control network, hence building a last line of defense. A proof of concept of this new paradigm was implemented for the control system of a dredging vessel, and we describe main lessons learned and pose open research questions we find based on these experiences for ICS intrusion detection.

Keywords: Cyber Physical Security, Intrusion Detection, Industrial Control Systems

1 Introduction

Industrial control systems (ICS) monitor and control physical systems, forming a cooperative bond between the digital and the physical world. They are to be found, amongst others, in critical infrastructures, building monitoring or production systems [1]. In recent years these systems are becoming increasingly connected to IP-based networks or even the Internet, either indirect through corporate networks or by direct connection. As such these systems are exposed to much of the same weaknesses as traditional IT systems. The effects of their failure, though, are potentially much more severe. Causing irreparable harm to the physical system being controlled, its environment, and to the people who depend on said system [2].

Unfortunately this potential for abuse is no longer just speculation. There exists numerous incidents resulting in significant damage, ranging from the flooding of a water treatment facility in Maroochy Australia caused by a disgruntled ex-employee with knowledge of the system and old access credentials [3], the often referenced attack on the Natanz nuclear facility in Iran by the Stuxnet malware [4–6], to the massive damage to a blast furnace in a steel mill in Germany after an attacker gained access to the control systems [7, 8].

Industrial control systems security is still in its infancy [9]. Although the topic is now attracting significant attention and there are many technical solutions to protect IT environments, controls such as intrusion detection systems or firewalls are not easily bridged to ICS systems. As we argue in this paper, many of the classic IT controls are not directly applicable to ICS or would actually not provide effective mitigation against the kind of real-world attacks listed above. While firewalls, intrusion detection systems and packet inspection tools are capable of filtering out unusual traffic in regular IT systems in terms of origin, access destinations and content, this defense is less applicable for ICS systems as packets will only flow between the programmable logic controllers (PLCs) and the ICS control node. As past attacks have compromised the control node and used this host to inject malicious commands or upload malicious binaries to the field devices, these types of attacks would not stand out from a network-packet analysis: packets are still flowing between the authorized control host and the ICS devices, and also in terms of packet sizes or from a protocol-interaction standpoint no anomalies would stand out.

Given these new types of threats and potential attack vectors, we argue that a new type of intrusion detection system is needed. As intrusion detection cannot successfully detect and mitigate such advanced vectors on the process network by looking at traffic between the controller and the PLCs, ICS defense needs to move closer towards the field devices that have actually been compromised in past incidents. It is necessary to (also) apply detection and mitigation on the last link of the field network, hence drawing a last line of defense that is difficult to subvert.

This paper contains three main contributions: first, we propose the idea of rethinking IDS approaches to meet advanced threats in industrial control systems, and argue that advanced vectors can only effectively be met deep inside the control network. Together with placing an IDS deeper into the ICS, it is also necessary to extend the current approach of anomaly detection in terms of packets and network flows, by interpreting the content and context of the packets and adding knowledge about the actual ICS process into the anomaly detection. Second, we have implemented this new paradigm for the control system of a dredging vessel and evaluated in extensive simulations the utility of this new IDS paradigm. Since detection rates are highly system- and model-specific and difficult to abstract, we do not go into the performance results in this paper, but rather present a number of observations and lessons learned about building intrusion detection for industrial control systems and discuss in our view open research challenges to solve. This is our third contribution.

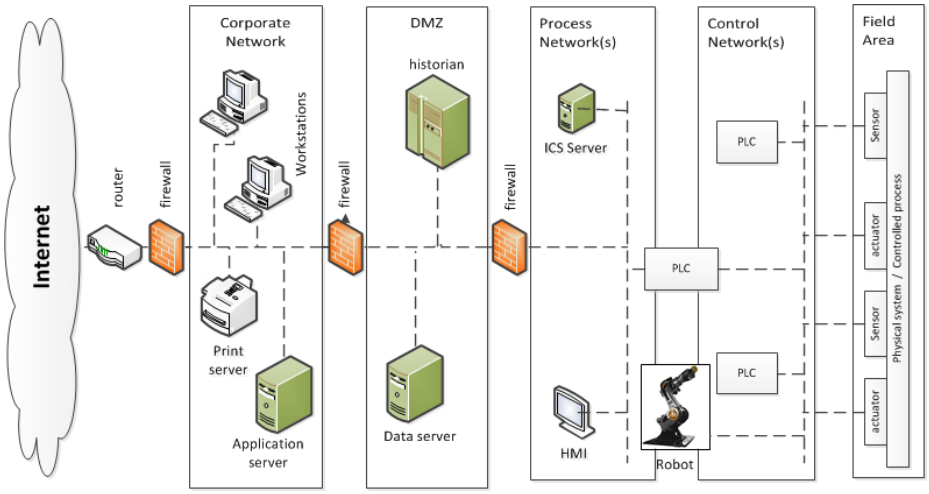

Fig. 1. Security zones and components of an industrial control system and corporate network.

The remainder of this paper is structured as follows: Section 2 describes the new threat model industrial control systems face, and reviews current approaches to ICS intrusion detection. Section 3 discusses related work. Section 4 presents an alternative approach to IDS design for ICS. Section 5 briefly introduces the evaluation of a prototype built to test the new design, while section 6 discusses the major lessons learned. Section 7 concludes and summarizes our work.

2 Threat Model

Although industrial control systems have always posed high value targets, three major developments over the last decade have greatly increased the attack surface and risk of these systems: First, previously entirely separated systems are now routinely furnished with remote access possibilities, allowing the continuous retrieval of measurements and statistics from the industrial processes inside the business operations. Second, increased value opportunities from cyber crime have resulted in a steadily increasing influx of actors, as well as a continuously growing specialization and sophistication of these actors [10]. Third, originally state-sponsored activities such as the Stuxnet incursion have demonstrated to a wider audience the general vulnerability of these systems. We have already seen in the recent past that the ideas and source code developed by nation-state actors have proliferated towards cyber criminals, and interactions of cyber criminals [11] with ICS – accidental or intentional – such as [8] thus need to be considered in the risk analysis of such systems.

Figure 1 shows the typical system architecture of an organization operating an industrial control network inside its perimeter. The different security

zones are typically enforced by firewalls or network diodes, and meant to ensure that no unauthorized traffic flows take place across a trust zone boundary or within a particular compartment. Techniques such as firewalls, diodes, or authenticated proxies have proven their merit for securing regular enterprise networks, and it is common practice to identify malicious activities or compromised hosts through automatic anomaly detection of network traffic based on traffic destination, packet types, payload sizes, or otherwise coinciding request patterns.

An industrial process network however provides almost no opportunities for such traffic anomaly-based threat detection. Although in the special case of a zone breach between the DMZ and process network some unusual traffic patterns could be observed, process network traffic will always flow between the exact same hosts. The human-to-machine-interface (HMI) will query the status of the PLCs in regular intervals, which are reported back by these units to the HMI and archived at the historian. In case of process changes, new logic is uploaded by the HMI directly to the PLCs. As industrial control systems are secured by an intrusion detection system monitoring the network traffic inside the process network, the following threats will remain undetected:

- **Incorrect operational process instructions.** Intentionally malicious or a benign operational mistake, the person controlling the system from the HMI could send incorrect commands to the devices in the field area, triggering a situation with severe consequences. From the perspective of the network there would be no anomalous flows, as commands – although unsafe ones – would be issued from the operator's station and sent to the field devices as usual. A problem could however be detected if the IDS would go beyond network-flow analysis, and parse and interpret the content of the network packets. While the packets are compliant to the protocol and also unlikely to be matched by a signature database of known threats (as commonly used by IDS), an IDS *with* knowledge about the physical process model as proposed in this paper would be able to spot a deviation from the expected state.
- **Malicious control software on PLCs.** In the example of the Natanz facility, the Stuxnet malware uploaded modified program code to the PLCs which executed malicious process logic in addition to the normal program [6]. When the devices were queried by the control host, the PLCs reported back incorrect values and did not reveal the presence of the malicious additions to the control program. Viewed from a network perspective, none of these activities would typically be flagged as abnormal: the control system would routinely upload new program code to the PLCs, query the devices and receive packets back. Even for a protocol-aware IDS, any of the packets are valid, and completely compliant with the access policy of such control network.

The sober conclusion, especially from the latter example, is that against advanced threats with the ability to modify the behavior of the PLCs itself, a mitigation approach centered around the process network is unlikely to provide

merit, unless the scope of detection is greatly extended to include in-depth verification of control logic before application, as well as stringent access control and supply chain security of the field devices. Given the changing threat profile and documented instances of such incidents, we believe it is necessary to extend the risk analysis and mitigation plans for industrial control systems towards these potential threats, and given the difficulty in achieving threat localization in the process network build up an additional layer of defenses one layer deeper inside the field network. Although an advanced adversary might in theory launch an attack in which all process control variables and sensor values would remain identical to the benign scenario, a pervasive tracking of the system at the process level will reduce the degree of freedom drastically, thereby eliminating most adversaries and slowing attack progressing down significantly. The proposed IDS should only be one of many techniques in a portfolio of controls. It may however augment and complement existing controls and address other operational questions, for example the detection of wear and tear, with the additional use cases and added benefit making an adoption more likely. The following section will describe current practices in industrial control system security, section 4 describes our alternative proposal for introducing security in the last link of the ICS network.

3 Related Work

Two areas of existing work are relevant for the design proposed in this paper, (a) security challenges for control systems, and (b) intrusion detection systems.

3.1 Challenges faced by ICS

Despite the similarities between control and IT systems, such as basic components used, the challenges they face in securing them are very different. As are their responses to security breaches [2, 12–14]. Three challenges which must be faced to strengthen control networks are identified; improving access controls, security inside the network, and the security management of control networks. Despite the similarities between control and IT systems, the challenges they face in securing them are quite different. Cheminod [14] additionally raises the challenge that while most ICS security studies focus on prevention and/or detection, there is relatively little research available into the response to threats. Historically these threats originate from the inside, these days this is shifting to the majority of security threats emanating from the outside.

3.2 Intrusion Detection Systems

Historically, the origin of intrusion detection systems evolved out of a set of tools mostly intended to help administrators review audit trails such as user access logs. In 1987 Denning published a paper titled “An intrusion-detection model” [15], describing what to this day remains the basis for many monitoring

systems. Today, there is a large body of IDS related research available. Almost all of these focus exclusively on IT-based environments, where the offered solutions are not directly suitable for ICS environments. This applies even when underlying protocols and infrastructure used are the same [2, 12–14, 16].

Research investigating monitoring solutions for ICS environments are not completely absent, however almost all of these make use of IT based approaches such as network traffic analysis and packet inspection [17–20]. These proposed solutions are thus focussing on the protocols utilized and ignore the physical domain entirely. As such a large and presumable the most important resource is missed. Researchers have however suggested to utilize this resource and incorporate knowledge of the physical system into the workings of the IDS itself [19,20]. By understanding the network traffic it is possible to simulate the physical, the result of which can then be used to take the physical state into account [21]. Research investigating a direct tap into the physical state by going directly to the field devices (sensors and actuators) seems to be missing however. [19] presents a taxonomy of ICS security related work. Besides presenting a new validation metric, they look into the advantages and disadvantages of different evaluation setups. Simulations here have the benefit that they are easily adaptable, and possibly provide the best method for initial proof-of-concept work. This is also reflected by the majority of the taxonomy, which rely on simulation for their validation step. Especially noteworthy is the conclusion by Urbina et al. [19] about the untreated risk if an attacker can falsify readings from the field devices itself, an issue that is mitigated by the work in this paper.

Change detection. As part of a series of studies into the security of ICS systems, Cardenas concluded that only limited research into ICS security is available and that what is published are generally tweaks of solutions aimed at an IT environment [2,13,22]. As such, incorporating knowledge of the physical system might very well trigger a paradigm shift in the sector. This realization leads to the proposal of a linear mathematical model which is used to analyse the actual system and determine if an attack is ongoing. Their main aim being to "*protect the operational goals from a malicious party attacking our cyber infrastructure*", which they separate into a two stage process: first, the detection of attacks on cyber-physical infrastructures, and second their survival [16, 23].

For the detection problem Cardenas suggest that when having knowledge on how the output sequences should react to the control input sequence it is probably possible to detect an attack by comparing this expected output with the actual received output signal. The effectiveness of this idea will depend on the quality of the estimated output signal. Further investigating this idea they created a model of a physical system and formulated an anomaly detection algorithm based on change detection. Change detection works under the assumption that a set of measurements starts out under the normal hypothesis, H_0. The thought is that this hypothesis will then be rejected in favor of the attack hypothesis H_1 at a certain measurement. To avoid making any assumptions on the probability of an actual attacker their work does not assume a parametric distribution but only puts mild constraints on the measured sequence.

Virtual image. Having researched the effects that malware has on industrial control systems, both for the case of IT malware as samples specifically targeting ICS, [21] suggests to leverage available knowledge on the physical system to enhance its protection. Following this suggestion [24, 25] does exactly that by creating an new intrusion detection system which maintains an internal representations of the physical state of the controlled systems in a virtual image. To define and build this virtual image a new formalized language has been specifically defined, which has been named Industrial State Modeling Language (ISML) [24]. The virtual image is meant to operate parallel to the monitored system, providing real time insights and analysis. At start-up the IDS will load the systems model from an XML file which comes with predefined settings and values. During operation, the IDS received a copy of the network traffic which it scans for ICS protocols. The ICS packets are then processed and the commands they contain send on towards the virtual image where they are used to update the internal representation of the control system. Each update also triggers a monitoring module that compares the (virtual) system state to a list of predefined critical states. If a match is found the IDS will raise an alarm. Implementation of a prototype and conducting of experiments have demonstrated that the proposed solution is successful, proving the approach has merit.

By adding a multidimensional metric that provides a parametric measure of the distance between a given state and the set of critical states further extends the IDS, giving it the ability to estimate future instability in the system [26]. Conducting experiments using a prototype implementation of the extended IDS demonstrated the improved functionality and that the approach indeed has merit.

Network based. The solutions given by the research above operate on the control network, which also applies to the research discussed by [19]. These solutions obtain their information on which their system works from the network traffic between controllers and the larger ICS systems. This means that their effectiveness directly builds on the security and trust placed upon the controllers within the system. If an malicious entity is able to compromise such devices the previous solutions can be evaded.

4 A Cyber-Physical IDS Architecture

In this section we present an alternative approach for intrusion detection in control systems. Previous work such as [19, 20, 24, 25] is centered around an anomaly detection based on information flows over the network connecting PLCs and the larger control network, which we have seen in the discussion of the threat model would not detect malicious or accidental instructions sent to the field devices or an upload of malicious control programs to the PLC. By moving intrusion detection one level deeper into the industrial control network, the field network, intrusion detection can also exploit the physical state of equipment - such as temperature and pressure readings -, as opposed to looking at network

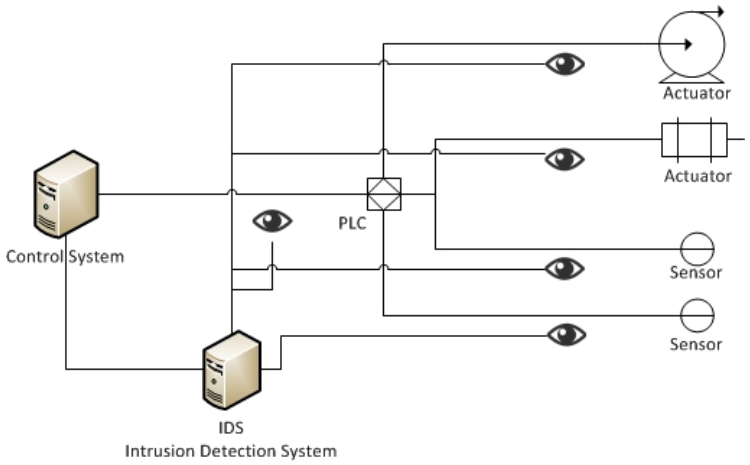

Fig. 2. The concept architecture

traffic solely. As "communication" with the field network is analog only, a sensor would map temperature into a voltage and an actuator would switch depending on the voltage level applied to the control wire, the absence of more sophisticated data exchange provides a number of advantages for defense: first, as there is no digital communication link this makes it much harder for malware to compromise the system. Second, by monitoring these lines it is possible to obtain the raw values measured in the field area, and in combination with an analysis of packets on the control network thereby also find faulty or maliciously acting PLCs.

Figure 2 shows a conceptual view of the shift to the last link. In addition to passively observing the communication between the control system and the PLC, values are taken from all or the most important sensor and actuator connections on the other side of the PLC. This design approach provides a number of distinct advantages:

1. **Extensibility and Multi-Vector Detection.** The system may be extended to include more sensors to accommodate evolving attacks and new vectors, including sensors not connected to the ICS system such as a microphone listing to the acoustics of machinery, sensors reading power usage and output of equipment, or even radio frequencies readings. One of Stuxnet's attack vectors for example cause damages by changing the spinning speed of centrifuges, causing mechanical damage. While a microphone could have easily recognized such changes, it is exemplary for additional types of sensors that are (normally) not providing input to an ICS system, but would help within the context of our IDS to detect abnormal events, making it significantly harder to launch attacks that would go unnoticed by other off-the-ICS sensors.
2. **Unmodified signal path.** Existing network-based IDS are located in between devices (here the controller and the PLCs) to scan and block malicious

traffic. As scans however change the latency of communication messages, may in some cases change packet order and the blocking of select packets within a larger control packet train may cause significant side effects, control engineers in our pre-study voiced concerns about placing such devices inside a production environment. As the proposed system does not interfere with IDS communication and control messages are interpreted off the bus, timings, packet order and the integrity of packet trains remain unmodified.

3. **Ability to detect compromised ICS infrastructure.** As the Stuxnet malware compromised both the control system as well as the PLCs reading and responding to sensor values, it was able to send back falsified sensor readings and remain unnoticed while bringing the ICS outside of the safe operating context. An IDS system reading both the state of actuators and sensor readings at the last line (which are assumed to be analog voltages) and comparing them with the readings reported by the regular control infrastructure has the ability to detect malfunctioning or purposely compromised equipment. While this would seem like a duplication of the control infrastructure, the additional expenditures for such an approach are actually minimal: they simply require a single microcontroller per group of sensors and actuators digitizing analog voltages and reporting them cryptographically-signed via Ethernet to the IDS.
4. **Upgrade without compliancy issues.** As no changes to the existing ICS infrastructure are necessary, the approach would allow for an effective upgradability of existing legacy systems. Note that since nothing is placed into the signal path that may intercept or alter its behavior, no compliancy issues or the necessity of re-certify the system would arise which would make a roll-out within certain critical infrastructures very expensive or time consuming.

From Network Anomalies to Physical Process Knowledge

The other main modification that is necessary to accommodate today's threat landscape is to move away from a detection solely based on network anomalies and include information about the physical processes and their behavior into an IDS. As discussed during section 2, several relevant threats in IDS would not deviate in terms of communication endpoints or packet sizes from normal traffic, nor trigger any exception in terms of protocol or access policy compliance. While this complicates the design of IDS, and means that instead of short training of off-the-shelf an extensive customization period of the IDS to the system at hand is necessary, an IDS with information about the physical processes can evaluate how a command sent from the control system to the PLCs would play out and thus be able to stop these threats previously uncovered.

In our work we implemented the physical model behind a general trailing suction hopper dredger design, which will be briefly introduced in section 5. Ultimately though, the last line of defense approach can be adapted to any kind of ICS given some knowledge of the underlying system. In the following we will briefly discuss four detection strategies we apply in our prototype system.

These strategies are basic and generic and cover different information sources available in system operation, as well as the design, engineering and control process. When the physical system is designed, we know from the engineering documentation the boundary conditions under which the system is designed to operate. A monitoring system can directly apply these values to maintain process safety and security, as deviations from the normal are cause for concern. In the next step, it is meaningful to compare the consistency between the action and values of the devices in the field area, and those being reported on the control network and inside the control system. This uncovers faulty devices, as well as those operating with malicious hard- and software. Finally, in case of a physical process (e.g., a chemical reaction), detailed knowledge of how the process will behave based on changes in input. Based on this information, an internal representation of the ICS can be kept by the IDS. This allows commands be evaluated for their potential effect and an evaluation whether the obtained sensor values would be feasible and expected from the current state of the system. Incoming sensor values and events, either input from a monitoring point such as an I/O measurement or communication and control messages, are subjected to four analysis methods in the proposed architecture: consistency comparison, value and signature analysis and envelope escalation. Each method is explained in further detail below:

Consistency comparison. In a first step, the value emitted by an instrument is compared to the value in the control system, as assuming the PLC and/or control system is not tampered with these values should match up. Any deviations are thus a basis for alarm and further investigation, indicating a malfunction or deliberate action.

Value analysis. In addition to a basic consistency check, instrument readings are compared against the device, component and system level specifications, describing the minimum/maximum operating context for specific components or the rating alarm and trip setting (RATS) list for the entire design. Think for example of the flow rate in a pipeline, which the control system only monitors for a lower bound value - no flow for example. There is also an physical upper limit though, which could be the maximum capacity the pumps can sustain. Any deviations from this are flagged for immediate investigation, especially if not reported as such by the main ICS control system.

Signature analysis. Industrial control systems run very structured processes, onto which a form of signature analysis is applied. The value under consideration and the context in which it appeared is matched against a list of logic rules and reference traces, raising an alarm when deviations are encountered. These rules can take any form, as long as they can be specified in a machine-readable and -interpretable format. Within this initial proof-of-concept we are considering (a) timing analysis of ICS control packets (since PLCs will show a different answer time and deviations from their otherwise exact response patterns in case they

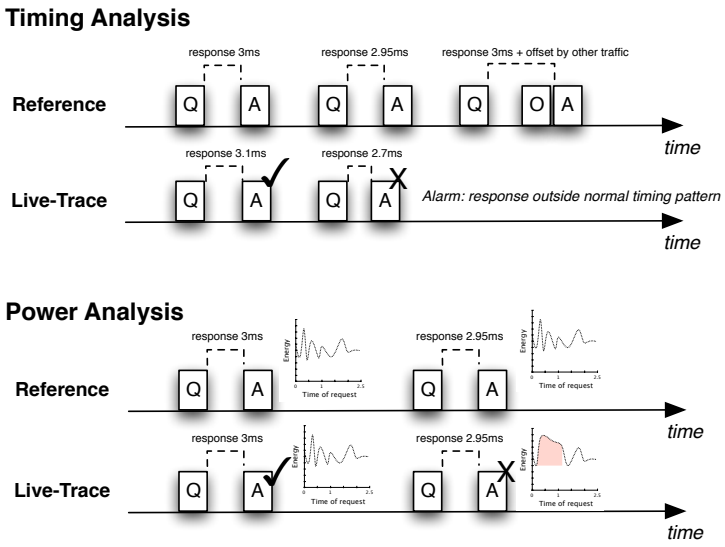

Fig. 3. Signature detection for timing analysis of request-response packets, as well as power consumption during the servicing of requests

are executing different software branches than usual), (b) request-responsive-sequence analysis of packets (PLCs communicate in set intervals status messages to which other devices hen react), and (c) power-trace analysis of PLCs (as PLCs execute other code branches than usual their relative power consumption over time will change). Figure 3 shows a schematic representation of this method.

Envelope escalation. As in control engineering the various states of the system, the failure domains and safe operating conditions are well defined, we can utilise knowledge of the system's secure and insecure states to follow the actions and reactions in a multi-level multi-dimensional model, thereby creating a safe-state envelope. Envelopes have been used for dependability engineering of communication networks to provide hard performance guarantees during challenge events [27, 28], however the method may directly be applied to control engineering as well. Each independent subcomponent of an ICS is described by one or more envelopes, which are defined by a set of metrics assessing a particular component from various angles. Each envelope hence captures an N-dimensional state space, which is annotated based on the system specifications indicating which operating context is safe or not.

Figure 4 shows this concept in a two-dimension plot for 2 independent metrics, with green indicating a safe operating context, red an unsafe system condition and yellow an operation outside norm values which may be temporarily acceptable or after a legitimate operator override. As can be seen in the figure, the IDS system monitors the development of the system's status based on the envelope specification and tracks whether it can still be considered safe. As com-

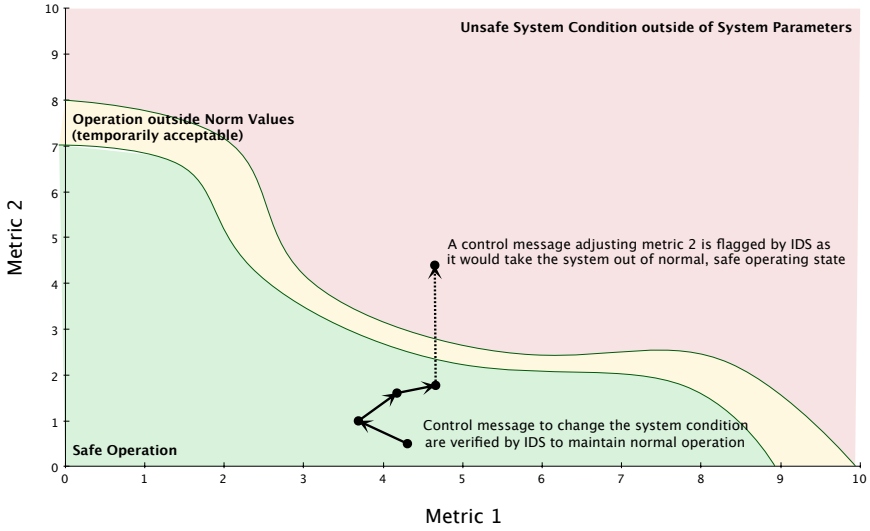

Fig. 4. Envelope escalation

mands are issued and controls are actuated that would take it into an unsafe condition, an alarm is generated. This tracking and detection can be done in two ways: First, many control processes are well defined, i.e., it is possible to determine before-hand how say a chemical process will change given a particular change in input variables. Second, in case such information is not directly modelable, it is usually generated and available after the testing period during a system's commissioning, during which normal and various boundary cases of a system are being tested.

Both cases will let the IDS directly flag a command as abnormal to the operator, and in case an actor conducts a previously unknown attack a comparison with historical commands and system responses will help the system maintainer to at least identify, where and how things went out of the ordinary with concrete pointers on how to roll back.

5 Experiments

To verify that our proposed new IDS approach has validity and can detect malicious tampering within an ICS, we implemented and evaluated a proof-of-concept prototype. This prototype is based on a generic trailing suction hopper dredger design. As the prototype is subjected to intentional tampering of the control systems, it is for security reasons not running on any production systems, but in an simulated environment. This has the additional advantage of enabling us to control system state and repeat scenarios under identical circumstances, and evaluate the performance of different detection strategies.

The remainder of this section will further elaborate on the source model, followed by a discussion of the experiments and the obtained results.

Source model

The source model can be created using three approaches: 1) first principles[3], 2) empirical input and output obtained from the field, and 3) a combination of both. This quickly narrows down to first principles as obtaining such empirical data is infeasible within this works scope. The approach used is to start out with the most simplistic first principles model with the option to increase complexity after initial evaluation. Reasoning for this is that the goal is to evaluate the IDS system and not to create the most complex and realistic TSHD model.

A TSHD has either a fixed or dynamic overflow system. Within the generic design use is made of a dynamic overflow system, which will also be used in the model as this offers a possible vector for malicious behaviour. The main risk within the loading cycle is that this overflow does not work as intended. This has the potential to cause the system to overload and sink the vessel. There are other parts that could malfunction, such as a suction pump not turning off. In those cases however lowering the overflow would win time and safety by simply ensuring excess cargo is discharged overboard.

The source model is built in such a way that there is a physical model representing the TSHD, and a separate controller that influences the state of the modeled THSD. This mimics the functionality of a real controller, which also operates on a process. The physical model includes the following main components: the hopper, the dynamic overflow and the inflow pipeline. In reality there would be many other parts involved but for the purposes of the evaluation these are not required and will be presented abstractly within the model.

The control network receives the sensor information from the physical model and processes this. After processing the controller computes the required change and sends a control message that influences the state of the source system. This has been represented with figure 5, displaying the field devices in play and their connection to the controller.

Experiments

There is a total of three experiments which aim to evaluate the prototype and the proposed detection strategies. Each is based on vulnerabilities identified by the threat model and inherent weaknesses in the source model. These experiments are cyber incident, manipulation attack and envelope escalation, each targeting a specific detection strategy.

Cyber Incident. A cyber incident occurs when a unwanted situation occurs but there is no malicious intent trying to cause the situation. An example is the overflowing of a tank because a control engineer has entered the wrong maximum volume for said tank. While these events might resemble a cyber attack,

[3] In physics the first principles approach relates to something which is based directly on established science.

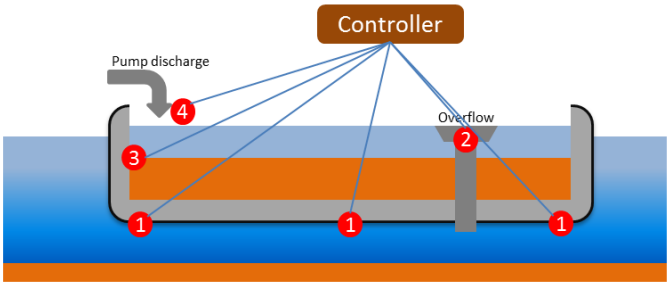

Fig. 5. The THSD modeled network.

the differentiator is intent. The evaluation is accomplished by introducing two incidents that each target a different part of the system. During each iteration one of these is randomly selected. The two incidents used are then: The first case mimics the changing of a system set-point. The second case alters process logic used within the controller. Both of these incidents mimic an operator, engineering or operational error which can cause the physical process to move outside of its (safe) operating specification.

Manipulation Attack. An manipulation attack is said to have occurred when a malicious entity manipulates a controller to act in a malicious way. For example by ignoring specific sensor reading and always reporting all is safe. This experiments implements three attacks: the first is a controller that thinks the cargo is not changing, the second is where the controller will effectively keep the overflow static and the final where the controller has access to a misleading draught. This produces three possible incidents, one of which is chosen at random during each iteration.

Envelope Escalation. An envelope incident has occurred when either a cyber incident or attack causes a monitored part of the physical system to move outside of the safety envelope. This experiment makes use of two incidents (set-point and static overflow) selected from the previous experiments, and additionally implements a new situation where the flow speed of the slurry is reduced to below a critical value.

Results

The main aim behind the proof-of-concept and the experiments was to demonstrate that the idea to reframe intrusion detection from a network-based view to a process-aware approach has merit. The obtained results are based on the initial experiments, without improvements, and simply marking a sample as malicious when somewhere in the model an unwanted situation was occurring. For each proposed analysis method, the results are as follows:

Method	True positive rate	False alarm rate
Value Analysis	88.7%	1.31%
Consistency Analysis	100%	$6.3 \cdot 10^{-6}$%
Envelope Analysis	92.3%	10.7%

It is important to note here that the signature method, as mentioned earlier, is not included in the experiments. The reason for this is the implementation of a model instead of making use of an actual control system and hardware. As this approach does not directly allow for signature detection, it was decided to keep it for future work.

It is evident that a system protected by the last-line-of-defense approach will be somewhat more complex than a design where a single IDS module monitors the flow of packets on the system's network, after all additional wires are needed from sensors and actuators to devices that cryptographically sign and report sample values to the main IDS component. This requirement in deployment complexity needs however be weighted against the added operational complexity when intrusion detection were to be deployed right into the system's signal path, potentially necessitating re-certification and measures to deal with dropped and modified control messages as discussed in section 4. Whether such an approach should be pervasively deployed or only a few sensors covered on the last line of defense is subject to a risk and cost-benefit analysis, a solution could be to cover the most essential sensors and parts of the process whose deviation would result in the highest impact, or instead redefine the measurement strategy by measuring multiple properties through orthogonal sensors not otherwise included in the control system. This point will be further elaborated in the lessons learned in the next section.

Regardless of the placement and coverage, the resulting defense in depth will increase the resilience of the system. In case where no formal description of an underlying process is available, a detection of process deviation can to some extent be accomplished by empirical learning of sensor values at the expense of loosing the predictive power of the IDS.

6 Lessons Learned and Open Research Questions

Based on the experiences building the proof of concept for the dredging hopper and the evaluation of the prototype with practitioners, we did find the alternative approach to have merit, but also raise a number of interesting aspects that have not found consideration in the academic literature. In the following we discuss five of the most important lessons learned and open research questions we identified.

Measuring ICS intrusion detection success
Coming from a traditional network-security background, measuring the performance of an intrusion detection system is straight forward. Given a number of packets, some of which are tagged as malicious, success is easily quantified

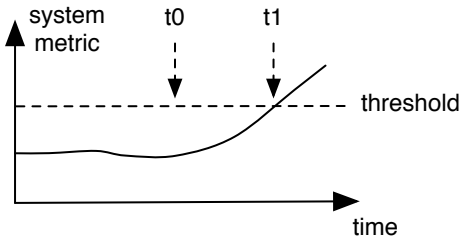

Fig. 6. In many contexts, ICS intrusion detection can be more efficiently measured by time-to-detect, instead of benchmarks such as TPR or FPR.

by means of the true/false positive and true/false negative ratios. While we may attach benign and malicious labels to packets on the control network, this packet-driven view on IDS success is however less expressive when viewed from the operational side of industrial control system owner.

To demonstrate this issue imagine some metric measuring one particular aspect of an industrial process which is monitored over time as shown in figure 6. At some point in time t0, a command is sent from the HMI to the PLCs that will guide the system to an undesired operating condition. At the scale of most industrial processes, the effects are however not immediately visible but will only manifest in time, for example a closed value will lead to a build up in pressure that will ultimately raise an alarm once a predefined threshold is exceeded. We see from this example that a packetized view to measure the success of intrusion detection may be rethought in case of ICS. While a single packet may be the root cause for an developing issue, categorizing packets as malicious and non-malicious is not straight forward. Although the system state could be defined as "infected" at any point after t0, the problem would only be detected as soon as the metric exceeds the tolerance and detection threshold at t1. By definition samples between t0 and t1 could thus be seen as false negatives as the IDS was unsuccessful in detecting the compromised system state. Even when counting packets, changing the sampling and reporting rates of the field units will change the TPR and FPR of the system without changing the IDS performance in itself. We believe that a better approach for ICS IDS evaluation would be operational characteristics such as the time to detect an issue, or the time between false positives as suggested by [19].

Rethinking detection with orthogonal sensor inputs

A recurring principle in IT security is the principle of "defense in depth", in which multiple layers of control require an adversary to circumvent multiple defenses thus slowing down the attack progression and increasing the likelihood of detection. The ideal defense in depth scenario would contain a set of complementary detection and defense mechanisms, so that any attempt to bypass one layer would be detected by another.

Industrial control system intrusion detection could embrace this principle of orthogonal detection at comparatively low complexity and overhead. While the sensors connected to a PLC and control network are all directly process related, there exist a plethora of other industrial sensor types that can be leveraged for asserting the process' stability. With sensors measuring the ship engine's rotations per minute and fuel consumption, an attacker could still change the parameters of the fuel injection process and increase the wear and tear. Although invisible in the existing measurements, an additional microphone and a fingerprint of the engine's sound would add an additional dimension to the attack, drastically increasing the complexity to successfully complete the compromise without being detected and remain invisible to the adversary as these additional sensors are not connected to the regular process network and thus do not show up in the HMI.

While not directly stopping attack progression, such additional sensory values significantly reduce the available attack surface, a malicious actor would need to account for the monitoring of multiple process variables while pursuing the attack. This will address unintentional misconfiguration incidents, eliminate most adversaries, and slow down the attack progression of advanced ones. As augmenting today's systems with additional sensory capabilities will increase both cost and complexity of such deployments, deployment of these sensors then depends on a risk evaluation of a control system and limited to the most critical failure points or processes with the highest impact upon failure.

Slow response time reduces urgency of comprehensive detection
When a command is executed to change the state of a physical system this change does not happen instantly, some amount of time will transition between the initial state and the resting state. This passing of time means that in the event of a malicious command the system will not instantly transition into an unwanted state. This "extra" time reduces the urgency for instant, and extremely, accurate detection as multiple samples can be taken and even combined for analysis, prior the system actually transitioning into the unwanted state.

Digital sensor security
While the bulk of sensors and actuators we found in practice were analog, newer types of sensors are increasingly making the transition to a digital integrated system. Basic building blocks such as pressure gauges that used to translate the measured quantity into an output voltage, now frequently include a network interface to stream out data independently, as well as extensive embedded software stacks such as a web server to control and configure the device. This additional software does not only introduce new features, but likely also new vulnerabilities and the possibility for an attacker to compromise the sensor itself. Attacks can thus be embedded even one layer deeper into an industrial control system, if a compromise of the firmware of digital sensors and actuators cannot be effectively mitigated by an IDS operating anywhere in the field network.

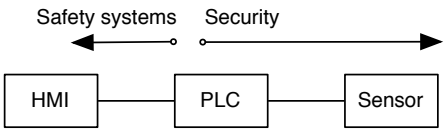

Fig. 7.

Additional value provided by ICS IDS
A significant portion of the use cases sought after by current IDSes is also accomplished by safety systems in industrial control systems. Based on the values provided by the PLC, redundant backups monitor the correct behavior of the primary control and raise an alert if the behavior of the primary deviates from the actions considered appropriate by the safety system.

Current safety systems however do not extend beyond the control network into the field network, and typically do not establish the correct functioning of the field devices and their embedded software as shown in figure 7. Such deviations of sensor values from the actual process state may be the result of two causes, either a malicious attack or the result of malfunctioning devices and instrumentation.

As it can be expected that in most operating contexts, the likelihood of sensor and actuator failure will significantly outweigh the likelihood of a device failure caused by an intentional attack, an intrusion detection system that includes the last line of defense has the potential to extend the scope of safety systems and augment them, for example by detecting a malfunctioning sensor device or a component damaged by wear and tear, ultimately leading to a merging of safety and security in ICS. The fusion of these two domains will both cut costs and result in more comprehensive coverage of the system, and the simultaneous view of the network on both sides of the PLCs also enables new functionality, such as the detection of faulty sensors or the measurement of wear and tear. These aspects might create a sufficiently attracting selling proposition as it can lower the operational costs, allow for better maintenance scheduling, and entice the development of new detectors and controls in ICS, and ultimately be less expensive even when considering the additional costs for extra instrumentation.

Still, in situations with malicious intent, the components in the process network will be the first ones to be targeted, and research results and practical experiences – such as the steel mill incident – highlight that existing controls are unsuited to stop advanced adversaries, requiring in addition to orthogonal detection extra rings of security beyond the current ones implemented by network-based IDS and safety systems.

7 Conclusion

This paper presents a novel intrusion detection system for industrial control systems which exploits their well defined nature and physicality. Our approach

differentiates from related research in that it also exploits the physical state of the system, as opposed to the network traffic between PLC and control system. This state information is then analyzed by three detection strategies: (a) value comparison that compares actual instrument values to what the control system reports, deviation of which could indicate a faulty or compromised controller; (b) signature analysis which checks the instrument value(s) to pre-set and state independent rules; (c) an envelope escalation strategy where a multi-level multi-dimensional envelope is created depending on the actual system state, checking if each sensor reading is still within this envelope from a system operation context. We built a working prototype of our IDS, for which initial validation experiments demonstrated the general feasibility of this approach. Although this research is a work in progress the initial results are promising and indicate that approaching intrusion detection from the physical instead of the networking side is indeed feasible and provides additional detection capabilities not existing in current solutions. As with all security research it will not provide a catch-all solution, but used along side other strategies offers a firm last line of defense.

References

1. D. Hadziosmanovic, *The Process Matters: Cyber Security in Industrial Control Systems*. PhD thesis, Universiteit Twente, 2014.
2. A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Conference on Hot Topics in Security*, 2008.
3. M. Abrams and J. Weiss, "Malicious control system cyber security attack case study-maroochy water services, australia," July 2008.
4. N. Falliere, L. O. Murchu, and E. Chien, "W32.stuxnet dossier," tech. rep., Symantec, February 2011.
5. G. McDonald, L. O. Murchu, S. Doherty, and E. Chien, "Stuxnet 0.5: The missing link," tech. rep., Symantec, 2013.
6. R. Langner tech. rep., The Langner Group, November 2013.
7. R. M. Lee, M. J. Assante, and T. Conway tech. rep., SANS ICS, 2014.
8. BSI, "Die Lage der IT-Sicherheit in Deutschland 2014," 2014.
9. R. Radvanovsky and J. Brodsky, *Handbook of SCADA / Control Systems Security*. CRC Press, 2013.
10. R. J. Anderson, *Security Engineering*. Wiley, 2008.
11. D. Goodin, "Stepson of stuxnet stalked kaspersky for months, tapped iran nuke talks." arstechnica.com, Februari 2017.
12. V. M. Igure, S. A. Laughter, and R. D. Williams, "Security issues in SCADA networks," *Computers & Security*, vol. 25, no. 7, pp. 498 – 506, 2006.
13. A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on Future Directions in Cyber-physical Systems Security*, DHS, July 2009.
14. M. Cheminod, L. Durante, and A. Valenzano, "Review of security issues in industrial networks," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, 2013.
15. D. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. SE-13, no. 2, 1987.

16. 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 *Symposium on Information, Computer and Communications Security*, 2011.
17. S. Etalle, C. Gregory, D. Bolzoni, E. Zambon, and D. Trivellato, "Monitoring industrial control systems to improve operations and security," tech. rep., Security Matters, 2013.
18. S. Etalle, C. Gregory, D. Bolzoni, and E. Zambon, "Self configuring deep protocol network whitelisting," tech. rep., Security Matters, 2013.
19. D. I. Urbina, J. A. Giraldo, A. A. Cardenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, and H. Sandberg, "Limiting the impact of stealthy attacks on industrial control systems," in *SIGSAC Conference on Computer and Communications Security*, 2016.
20. D. Hadžiosmanović, R. Sommer, E. Zambon, and P. H. Hartel, "Through the eye of the plc: Semantic security monitoring for industrial processes," in *Annual Computer Security Applications Conference*, 2014.
21. I. N. Fovino, A. Carcano, M. Masera, and A. Trombetta, "An experimental investigation of malware attacks on scada systems," *Critical Infrastructure Protection*, November 2009.
22. A. Cardenas, J. Baras, and K. Seamon, "A framework for the evaluation of intrusion detection systems," in *Security and Privacy, 2006 IEEE Symposium on*, pp. 15 pp.–77, May 2006.
23. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Distributed Computing Systems Workshops*, pp. 495–500, June 2008.
24. I. N. Fovino, A. Carcano, T. D. L. Murel, A. Trombetta, and M. Masera, "Modbus/dnp3 state-based intrusion detection system," in *International Conference on Advanced Information Networking and Applications*, 2010.
25. A. Carcano, I. N. Fovino, M. Masera, and A. Trombetta, "State-based network intrusion detection systems for scada protocols: A proof of concept," in *Critical Information Infrastructures Security*, vol. 6027, Springer-Verlag, 2010.
26. A. Carcano, A. Coletta, M. Guglielmi, M. Masera, I. N. Fovino, and A. Trombetta, "A multidimensional critical state analysis for detecting intrusions in scada systems," in *Transactions on Industrial Informatics*, vol. 7, 2011.
27. C. Doerr and J. M. Hernandez, "A computational approach to multi-level analysis of network resilience," in *Third International Conference on Dependability*, DEPEND, 2010.
28. C. Doerr, "Challenge tracing and mitigation under partial information and uncertainty," in *Communications and Network Security (CNS)*, 2013.