Challenges and Opportunities in Detection of Safety-Critical Cyber-Physical Attacks

Hui Lin, hlin2@unr.edu, CSE Department, University of Nevada at Reno;

Homa Alemzadeh, ha4d@virginia.edu, ECE Department, University of Virginia (co-first author);

Zbigniew Kalbarczyk, kalbarcz@illinois.edu, ECE Department, University of Illinois Urbana-Champaign;

Ravishankar Iyer, rkiyer@illinois.edu, ECE Department, University of Illinois Urbana-Champaign

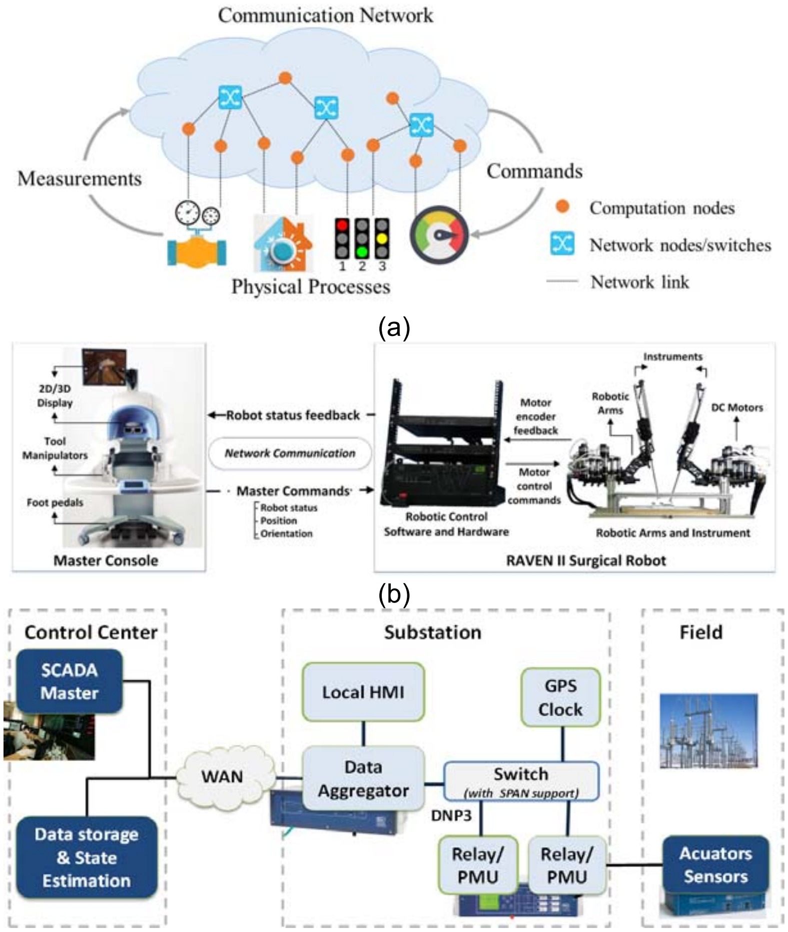
Abstract: Cyber-physical systems (CPS) control and monitor physical processes through off-the-shelf computing components and network infrastructure. CPS are increasingly used in various application domains, e.g., smart power grids, vehicular networks, interconnected medical devices, and smart manufacturing systems, with very different characteristics regarding control and computing algorithms, underlying physical infrastructures, communication protocols, timing constraints, and level of autonomy. Despite these variations, CPS face the threat of cyber-physical attacks, which might exploit the common vulnerabilities in cyber layer to introduce safety violations in physical domain. In this paper, we discuss the common challenges in detecting CPS attacks by presenting representative related work and analyze how diverse characteristics of CPS impact the efficacy of detection mechanisms. To clarify our analysis, we use two different example CPS, i.e., power grids and surgical robots. Finally, we use this analysis to identify the ongoing challenges and future research directions in ensuring resilience of CPS.

Keywords: security, attacks, cyber-physical, safety-critical, resilience, detection.

Introduction

Cyber-physical systems (CPS) are systems controlling and monitoring physical processes through the tight interconnection of off-the-shelf computing components and network infrastructure. Despite the differences in communication networks and physical processes, today's CPS are control systems with two common types of interactions between cyber and physical layers, as shown in Figure 1(a). One type of interactions involves collecting measurements from the physical processes and using them as an input to the control algorithms to update the models of the physical processes in the cyber layer. Another type of interactions involves the commands generated by the control algorithms based on the most current model and estimated the state of the physical process to ensure system's operation and long-term stability.

One of the major threats to the resilience of CPS is the *safety-critical cyber-physical attacks*, a class of malicious attacks that exploit the vulnerabilities in the cyber domain as footholds to introduce safety violations in the physical layer. By compromising measurements or control commands in a legitimate manner, adversaries can leave few detectable traces in the cyber and physical domains and evade detection by the commonly used intrusion and malware detection techniques. To present these attacks in a unified way, we use Figure 1(a) to depict the most likely entry points for attackers to penetrate into the system. In the first type of attacks, which are often referred to as *control-related attacks*, adversaries maliciously modify the control fields of commands delivered through communication networks to cause damage or disruption in the operation of the physical processes[1]. These attacks are no longer only the subject of study in research as their occurrence has been reported in real incidents, e.g., the Stuxnet attack on Iranian nuclear facilities and the attacks to the Ukrainian power plants. In the second type of attacks,



(a)

(b)

(c)

Figure 1. (a) Cyber-physical system control and example communication structures for (b) robotic surgical systems and (c) power grid infrastructures.

which are referred to as *false data injection* attacks, adversaries compromise the measurements related to the state of physical processes. Research studies have shown that the false data injection attacks can either (i) mislead control algorithms into issuing unsafe control commands to the physical layer[2] or (ii) hide the real physical state of the CPS (potentially malicious state caused by the control-related attacks) to delay the response and recovery from malicious states.

Different attack models on CPS, including the attack entry points and attack profiles, have been subject to study in previous work[3]. In this paper, we are specifically focusing on the CPS attacks with the common objective of causing disruptions in the physical layer. These disruptions can be classified into the following categories:

- **Physical malfunction:** Adversaries causing the CPS fail to deliver the contracted service, e.g., power system outage. These malfunctions might not introduce severe damage to the system but can jeopardize the reputation of service providers in the long-term.
- **Personal safety:** Adversaries injecting control-loop "Trojan” to perform unexpected operations[4], causing threats to personal safety (e.g., attacks to industrial or surgical robotics).
- **Economic loss:** Adversaries targeting the optimization procedures in CPS to introduce economic loss and/or directly obtain economic benefits (e.g., attacks to the optimal power flow analysis in electric power grids that try to satisfy the load demands with the smallest generation costs).
- **Altered observability:** Adversaries (or false data injection attacks) hindering the observability of CPS to either (i) mislead system operator into issuing false control operations and introduce other consequences or (ii) hide the malicious consequences to delay remedy procedures.

Unlike previous work that took ad-hoc approaches to proposing attack models and defense mechanisms for specific CPS, we strive to establish scientific foundations for modeling of attacks and design of defense mechanisms that can further guide us in designing future attack-resilient CPS. Specifically, we aim at developing a set of generalized design principles for resilience against cyber-physical attacks. To achieve this goal, we have reviewed representative literature on a wide range of CPS applications and system characteristics and a variety of possible attack scenarios.

Scope of Review: We specifically focus on the related work presented recently at different academic venues, including, security, dependability, control, power systems, and the Internet of things (IoT) conferences. Due to space limitations, the works which only focused on abstract mathematical models of CPS attacks without precise threat models, those with no experimental evaluation of the impact of attacks, or attack models with no concrete detection methods were excluded from our analysis. This survey helps us to identify the unique characteristics of the CPS across different application domains and characterize different categories of attack detection methods. We then analyze the correlations between the CPS characteristics and the detection techniques to characterize trends, alternative approaches, limitations, and ongoing challenges in ensuring resilience in CPS.

CPS Diversity

In this section, we characterize CPS systems that have been the target of cyber-physical attacks in previous research, in terms of control and computing algorithms, underlying physical processes, communication infrastructures, timing and resource constraints, and level of autonomy versus involvement of human operators in their supervision and control.

Table 1 presents the breakdown of diverse characteristics in CPS. The first row of the table lists different target systems studied in the reviewed literature, including industrial control systems (e.g., power systems, chemical and water plants), robotic systems (e.g., surgical robots or industrial robots used in smart manufacturing), autonomous and platoon-based vehicles, automated building systems, Internet of things, and augmented/virtual reality systems. The first column of the table includes different dimensions that we consider for characterization of these systems and are further described below:

- **Cyber-domain.** A major component of the cyber-domain in CPS is the communication network. We use “Central” control to indicate that there exists a central control unit across a wide area network collecting states of underneath physical processes and making decisions on control operations. A typical example of central control is SCADA (Supervisory Control And Data Acquisition) systems, which are often used in electric power grids and water plants. “Distributed” control refers to regional communication networks involving components that are physically near to each other, such as platoon-based communication networks inter-connecting the autonomous vehicles. In distributed CPS, each physical component makes a control decision based on the information collected from its neighbor components. The last type of communication network considered here is the “Local” communication, where a physical component communicates only with the control decision unit in cyber layer without sharing information with other physical components.
- **Physical-domain Models.** The physical domain characteristic refers to the state of the underneath physical processes and their trajectory over time. We classify the physical domain based on the mathematical models used to specify the physical state of a CPS. Researchers can develop a closed form analytical model to specify the “Static” state for some CPS and to specify the “Dynamic” state for others. For some CPS, such as devices for monitoring human behavior or physiology, establishing a closed form model that accurately describes the underneath physical processes is very challenging. Those cases are specified as “*No Analytical Model*” in the table.
- **Level of Autonomy.** Some CPS require the active involvement of human decision and supervision in their control loops. The human involvement is closely related to the level of autonomy of the control system. “Autonomous” control (listed in Table 1) indicates that control decisions are automated with little involvement of human operators unless an extreme emergency case occurs. In the “Semi-autonomous” control, the human operators and autonomous control algorithms are collaboratively involved in supervision and decision making and human decisions can prioritize the automated decisions. In the “Manual” control, the decisions made by human operators directly impact the critical path of the control loop. These human involvement levels can be mapped to the levels of autonomy defined by the international standards for self-driving cars and surgical robots (e.g., "manual" mapped to levels 0-1, "semi-autonomous" to levels 2-3, and "autonomous" to levels 4-5).

Table 1. Diverse Characteristics in CPS.

Related Work	5, 1, 6, 7	8	9, 10, 11	4, 12	13	14, 15	16, 17	18	19	20
Application Domains **Characteristic Dimensions**	*Power Systems*	*Chemical Plants*	*Water Plants*	*Medical Devices*	*Smart Manufacturing*	*Autonomous Vehicles*	*General Process Control*	*Building Automation*	*Internet of Things*	*Augmented Virtual Reality*
Cyber-domain										
• Central	■	□	□		□		■			
• Distributed									■	
• Local				■		■	■	□		■
Physical-domain Model										
• Static Model	■				□					
• Dynamic Model	■	■	□	■			■			
• No Analytical Model				□		■		□	■	□
Level of Autonomy										
• Autonomous				■	□		■	■	■	
• Semi-Autonomous	■	□	□			■	□			
• Manual				■						□
Time constraints										
• Tight (< 10ms)				■		■				□
• Medium (10~100ms)	■				□		□		■	
• Loose (> 1 sec)	■	■	■				□	■		

Legend: ■ indicates that the characteristics are explicitly mentioned in the literature; □ indicates that the characteristics are implicitly inherited from the literature or its related work.

- **Time Constraints.** We use this characteristic to specify the requirements of computation and communication latency in CPS. These requirements not only impact the interactions with the human operators but also the trade-offs that should be made in the design of detection and response mechanisms. Based on the reviewed literature, we classify these requirements into three levels of tight (< 10ms), medium (10-100ms), and loose (> 1sec).

There could be close correlations between certain characteristics of CPS. For example, in case of the tight involvement of human operators in control, the communication latency is demanding to ensure that operators can observe the run-time state of the physical process and take timely appropriate actions. Example of CPS with such characteristics include surgical robots, autonomous vehicles, and augmented reality systems.

Example CPS

In Table 2, we present a detailed description of two example CPS, including robotic surgical systems (Figure 1(b)) and power grid infrastructures (Figure 1(c)) and demonstrate their similarities (inherited from the common communication structure shown in Figure 1(a)) and their very different characteristics. We then discuss the detection of control related attacks in these two CPS. Both these systems rely on a feedback control loop, in which a control decision software and/or human operators rely on measurements from the physical systems to decide the appropriate control operations.

Table 2. Characteristics of Two Example CPS.

Characteristics	*Robotic Surgical Systems*	*Power Grid Infrastructure*
System Description	The surgical robots typically consist of a master teleoperation console, the robot control system, and the robotic arms and surgical instruments. • The master console is controlled by the human operator to send the desired position and orientation of robotic arms to the robot control system which translates them into control commands for moving robotics joints. • The control system consists of software modules running on top of commodity operating systems and robotics middleware, communicating with hardware and electronic components (e.g., motor controllers, DACs, PLC) via custom interface devices (e.g., USB).	The communication structure includes control center, substations, and field sites. • The control center uses a SCADA master to collect data from substations, estimate system state, and issue control operations. • A substation can contain various intelligent devices, which can run off-the-shelf operating systems and communicate with each other over IP-based network.
Physical Domain	In each control loop, the current state of the end effector on each robotic arm is estimated based on the encoder readings from the joints using the forward cable coupling and kinematics functions. The end-effector positions and orientations are translated to the joint and motor positions using inverse kinematics and cable coupling calculations and are sent to the motor controllers in the form of torque commands obtained using a Proportional-Integral-Derivative (PID) controller.	To describe system state, we can formulate at each bus two power-flow equations, which specify the mathematic relations among the system state, the generated power, the consumed power, and the power delivered to other buses at each timestamp.
Cyber Domain	The user commands are transferred from the master console to the robot control software over the network using TCP/UDP based protocols. The communication network is usually implemented as point-to-point connections between devices, to achieve a complete isolation from the rest of the network infrastructure.	The control center is connected to substations through a *wide area network* (WAN). Traditionally, this control-network is not open to the public Internet. However, to boost control efficiency, the control network is now connected through corporate networks of a power system or through personal devices.
Level of Autonomy	The state-of-the-art surgical robots are semi-autonomous systems requiring real-time interactions with human operators. The transition between the control states and robotic movements can only occur upon the commands issued from the surgeon at the master console.	Today's power systems can operate autonomously, as formally specified in multiple industrial standards, e.g. IEC 61850. The example of the automation includes fault isolation and generation controls, automatic reclosing breakers, etc. The grid operators monitor the real-time system state and will only interfere upon certain anomaly, e.g., load shedding decisions.
Time Constraints	The robot control software must complete each iteration of control and computing the new position of the robotic arms within a time less than or equal to 1ms.	In power grids, the requirements to deliver measurements or control commands can range between hundreds of milliseconds to several seconds.

Classification of Detection Approaches

We classify the detection approaches in the reviewed literature into two main categories of *Data-centric* and *Specification-centric*. The *data-centric* approaches refer to the detection methods that rely on the statistical characteristics extracted from the measurements collected from CPS. The typical examples include anomaly-based intrusion detection methods and network/device fingerprinting methods[5,9,14]. The major advantage of the data-centric approach is that it can apply machine-learning and analytic techniques to data from either cyber or physical domains or both, despite different CPS implementations. However, without considering the domain-specific characteristics of a CPS, when cyber-attacks happen, there exists a semantic gap between statistical deviations and the physical impact of attacks on CPS.

The *specification-centric* approaches refer to the detection methods that rely on the established standards, rules, or known specifications and models of target CPS to detect any inconsistent or anomalous behavior. These approaches can directly reveal adversary's intentions and apply remedy mechanisms based on the detected attack scenario. However, developing accurate models and specification-centric detection techniques are often very challenging tasks, as many CPS and industrial control systems rely on proprietary communication protocols and the manufacturers are often reluctant to reveal details of their design documents to third-party entities that perform security monitoring.

In Table 3, we present the breakdown of different detection approaches in data-centric and specification-centric categories as well as a hybrid of the two, as described below:

- ***Data-centric Detection***. We further classify the data-centric detection approaches into two categories, i.e., "cyber-domain" and "physical domain," to refer to the source of data used to extract statistics from. The traditional CPS often use proprietary network protocols for cyber and physical communications. Consequently, approximately 40% of the reviewed literature use the cyber-domain data-centric approaches, i.e., the data extracted from the transport or IP layer of network packets. This group of work uses the cyber layer data to indirectly reflect domain-specific behavior and state of the physical layer. For example, based on the observation that physical operations are usually periodic, Markman et al. divide time-stamped traffic flows into a sequence of bursts and then build a deterministic finite automaton for each burst of the traffic. Compared to treating all traffic equally, the automata built for each burst can reflect the characteristics of the underlying physical process[9]. Another approach uses time stamp differences as fingerprints as demonstrated by Formby et al. where time stamps recorded at consecutive TCP layer packets are used for fingerprinting and to infer the execution time of certain physical operations[5].

  The physical-domain data-centric approaches use the measurements collected from the physical processes to infer the state of CPS. If the application protocol used in a CPS is open, system operators can also extract the physical-domain data from the application layer of the run-time network packets[14]. For example, Urbina et al. used the statistical models based on the physical-domain data to not only detect attacks but also limit their impacts on physical processes[11].
- ***Specification-centric Detection.*** Similar to the data-centric approaches, we classify the specification-centric detection approaches into two categories of "cyber-domain" and "physical-domain," based on the scope of the specifications used to detect attacks. The cyber-domain specifications use the standards of communication channel protocols as references to model normal application activities. The typical examples include IP based networks such as DNP3 or Modbus, which are widely used in industrial control systems. These protocols define not only the syntax of network packets but also state transitions in end-devices. One major challenge of using cyber-domain specifications is how to translate protocol definitions into rule sets that are deployable in communication channel monitors. For example, Caselli et al. has presented automated methods based on machine learning to facilitate this procedure[18].

  The physical-domain specification approaches, on the other hand, rely on mathematical models that can describe the dynamic state of the physical processes. These mathematical models can provide valuable references to estimate the potential consequence of attacks in the physical domain. However, the major challenge is that complexity of these mathematical models may prevent system administrators from using them for real-time monitoring. For example, in our recent work, we have demonstrated that it is applicable to relieve the complexity of these mathematical models by reducing the number of considered parameters and the complexity of computations through approximation[4].
- ***Hybrid Detection.*** This category of detection methods use the advantages of both data-centric and specification-centric approaches across different domains of a single CPS. For example, Fauri et al. use physical models as guidelines to build anomaly-based detection mechanisms[13]. The physical

Table 3. CPS Characteristics vs. Detection Methods.

Characteristic Dimensions	*Cyber-domain*			*Physical-domain*			*Human Involvement*			*Time Constraints*		
	Central	*Distributed*	*Local*	*Static*	*Dynamic*	*No Analytical Model*	*Autonomous*	*Semi-autonomous*	*Manual*	*Tight (< 10ms)*	*Medium (10~100ms)*	*Loose (> 1 sec)*
Data-centric: Detection based on statistical models of activities observed in either cyber or physical domains.												
• Cyber-domain	5, 9	19	15	5	9	19, 15	19	9, 5, 15		15	19, 5	9
• Physical-domain	13†, 10†, 11, 6		14, 15	13†, 6	10†, 11	14, 15	13†	14, 10†, 11, 15, 6		14, 15	13†, 6	10†, 11
Specification-centric: Detection using established standards, rules, or known specifications and mathematical models of a target CPS												
• Cyber-domain	10†		18, 20, 12		10†	18, 20, 12	18, 12	10†, 20		20, 12		10†, 18
• Physical-domain	13†, 1, 6, 16, 7		4, 8, 17	13†, 1, 6	4, 8, 16, 17, 7		13†, 8, 17	1, 6, 16, 7	4	4	13†, 1, 6, 17	8, 16, 7

Legend: † indicates that the method is hybrid and uses both data-centric and specification- centric approaches.

models can help identifying the critical parameters to consider for anomaly detection in the cyber layer. On the other hand, the data-centric approaches can be effective for estimating the unknown model parameters or the relations between the parameters. By combining statistical characteristics observed in the cyber domain with the mathematical models of physical domain, we believe that hybrid methods can capture a more thorough picture of CPS operational logic and thus be more accurate and efficient in detection of attacks.

Detection Methods and CPS Characteristics

In this section, we analyze the detection methods proposed for categories of CPS with different characteristics and application domains shown in Table 1. The results of this characterization are presented in Table 3. It appears that there exist correlations among the types of detection methods used in the literature and the characteristics of the target CPS. The following are some of the major observations made based on this analysis:

- Data-centric detection methods based on cyber-domain measurements dominate the literature, while specification-centric methods based on physical-domain models are emerging in recent years. This trend shows that researchers have realized the importance of applying domain-specific knowledge of CPS into design of detection methods.
- For the CPS that modeling of the underlying physical system is challenging (*"No Analytical Model"* category), the data-centric methods or cyber-domain specification-centric methods are the only viable options. Similarly, in CPS such as power grid where constructing accurate dynamic models of physical system is challenging and not very efficient, the static models of physical domain were considered[1,5,13].
- For some CPS, although the static[5] and dynamic[9,10,11] models of physical processes exist in the literature, some proposed detection techniques still focused on data-centric methods based on cyber or physical domain measurements.

- Little attention has been paid to design of resilient CPS with tight real-time constraints and most of the existing work rely on data-centric approaches. This is because *timely* detection of the adverse consequences of attacks and damage prevention requires the runtime execution of complex mathematical models, which cannot be easily achieved within the real-time constraints of system.
- Many detection methods rely heavily on “central” cyberinfrastructure and only a few focus on “distributed” CPS such as IoT systems. In "distributed" CPS, the communication overhead of information reaching the decision-making points can become a hurdle for real-time detection and response.

Detecting Control-Related Attacks in Example CPS

In Table 3, we demonstrated the correlations of CPS characteristics and detection methods. In this section, we use the example CPS from Table 2 to discuss those correlations in details. Here we mainly focus on control-related attacks that are initiated by malicious modification of control commands. In these example CPS, we use the following common set of detection principles: (i) keeping track of the updated physical states in the cyber layer; (ii) continuous monitoring of communication network/links (in cyber-domain) to assess the control commands; and (iii) using the dynamic behavioral model of physical system to estimate the potential consequence of executing control commands before they are actually executed in the physical layer. However, the same principles lead to very different designs and implementations of the detection methods based on the characteristics of the target CPS.

Attack Detection in Robotic Surgical Systems

In our previous work[4], we demonstrated that malicious modification of control commands in a surgical robot could cause abrupt jumps of a few millimeters in the robotic arms in only a couple of milliseconds. If the attacker mounts such attacks at a critical time during surgery, it could cause catastrophic damage to the robot and potential harms to patients. Among four discussed characteristics of CPS, the cyber and physical domain characteristics of surgical robots play a dominant role in design of detection methods. For timely detection of unsafe abrupt jumps before they occur in the physical layer, we designed a hybrid detection approach consisting of: i) a dynamic behavioral model of the robotic actuators; and ii) an anomaly detection module for continuous monitoring and fusion of real-time measurements from cyber-layer.

Cyber-domain. To minimize the gap between the time of safety checks to the time of execution of control commands and reducing the attack surface, we retrofitted the hardware interface board in the control system of the surgical robot as the last computational stage for deploying dynamic models and anomaly detection mechanisms. All control commands and sensor measurement sent by the control software are received and monitored before the commands are executed on the physical robot.

Physical-domain and time constraints. To estimate the impact of control commands, we developed a software module that estimates the next position of the robotic actuators based on the control commands. Two sets of second-order ordinary differential equations were used to describe the dynamics of the robotic joints, and DC motors and the cable tension for the joints. The fourth-order Runge-Kutta and explicit Euler methods were used to calculate the solutions to these equations using numerical integration solvers at runtime. The main challenge in developing the dynamic model was to perform estimations within the tight time constraints of the robot control loop. To reduce the computational cost while maintaining the model accuracy and real-time guarantees, we modeled the robot manipulator dynamics using only the first three (out of seven) degrees of freedom (two rotational joints plus one translational joint). This approximation is reasonable because the first three joints are positioning joints that contribute the most to the instruments’ end effectors’ positions[4].

Attack Detection in Power Grid Infrastructure

As shown in [1], the malicious modification of control commands can impact power system's steady state and dynamic behavior, similar to what happened in the Ukrainian power grids incident, where malicious commands injected by attackers resulted in safety violation of the grid and causing the grid to be down for several hours.

Cyber-domain. Using off-the-shelf communication infrastructure makes it easy to tap into the power system's network to monitor measurements and commands in SCADA systems. Overcoming proprietary network protocol used in power systems can be achieved by designing new toolsets, such as extensions of Bro, a runtime network traffic analyzer, to support DNP3 and Modbus, the network protocols widely used in U.S. power grids. However, the wide area communication network with a large number of sensors can introduce large overheads on collecting measurements and make the real-time detection challenging.

Physical-domain and time constraints. To estimate the consequences of control commands in physical layer, we used power-flow analysis to estimate the state of power grids upon execution of the commands. To shorten detection latency while preserving detection accuracy, we proposed a new adaptive power-flow analysis and integrated it with real-time network analyzers[1]. Specifically, we adapted the number of iterations that classical AC power-flow analysis used to estimate the power system state. Instead of statically fixing this parameter, we dynamically adapted the number of iterations based on scale of the communication networks, the topology of power system transmission network, and the parameters of control commands observed at runtime.

Discussion

Summary: In this paper, we presented a summary of literature on detection of cyber-physical attacks by reviewing the papers presented in recent years across different academic communities. We found that:

- A large body of work in the control theory and CPS communities focused on analysis of attacks and detection methods based on abstract mathematical models of CPS. These studies presented robust mathematical understanding of the potential attacks and their success probabilities or robust state estimation techniques for handling noise and uncertainty. However, they often lacked precise threat models (describing the required steps for the actual implementation of attacks) and experimental evaluation of impact of attacks on the physical system. We did not include those papers in our analysis due to space limitations.
- Related work presented in the security and dependability communities mainly focused on attack implementations with realistic threat models and assessing the impact of attacks. But the proposed models and detection methods only targeted specific CPS without generalization to other domains.
- Further, there exist many research efforts focusing on proposing new attack models with no concrete detection methods or only detecting attacks that target exclusively the cyber layer or the physical layer. We did not include these works in our analysis due to space limitations.

Research Challenges and Opportunities: As CPS evolve with advanced sensing, computing and network technologies, some of the CPS characteristics and attack detection techniques can be unified within and generalized across different application domains to provide new opportunities for design of resilient CPS. For example, in recent work we applied the distributed communication infrastructure (a characteristic of CPS such as platoon-based vehicles and IoTs) to the power grids to detect false data injection attacks that often evade detection in central network infrastructures[2]. Another example includes adapting the existing physical-domain models from an application domain to design of generalized specification-centric methods for all CPS in that domain. Further, similar to the examples shown in the previous section, detection methods and principles can be generalized and applied to different applications.

The study of different attack detection methods provides us with valuable insights into the ongoing challenges and opportunities in ensuring resilience of CPS:

- **Hybrid Detection Methods:** There are very few research efforts focusing on the hybrid detection methods, e.g., data-centric approaches considering both cyber and physical data[15], or both data-centric and specification-centric approaches considering both the measurements and models from the physical system[6]. It has been shown that hybrid approaches provide improved detection accuracy by capturing a more comprehensive picture of system state[8].
- **Resilience in Human-in-the-loop CPS:** There is little work on the resilience of CPS which require tight human interactions and involvement in control and decision making[4,20]. This research trend indicates the on-going challenges in accurate modeling of human behavior, psychological status, and decision-making procedures as well as design of safe procedures for timely transfer of supervisory control.
- **Attack Response and Recovery:** Response and recovery are important parts of CPS resilience to maintain continuous operation, but it is rarely addressed in the reviewed literature. The challenges in designing appropriate response mechanisms are tightly correlated with the time/resource constraints and the level of autonomy or human involvement in CPS. Although, this topic is beyond the scope of designing detection methods, we argue that the necessary conditions for effective response should be taken into consideration when designing the detection mechanisms, e.g., the maximum response time can be used as a requirement for design of intrusion detection systems[1], or the operator response time

can be used as a constraint for optimization of detection latency in real-time semi-autonomous systems[4].

References

1. H. Lin, et al., "Runtime Semantic Security Analysis to Detect and Mitigate Control-Related Attacks in Power Grids," in *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 163-178, Jan. 2018.
2. E. Amullen, et al., "Multi-agent System for Detecting False Data Injection Attacks Against the Power Grid," in *Proceedings of the Second Annual Industrial Control System Security Workshop* (*ICSS '16*), pp. 38-44, 2016.
3. M. Rocchetto and N. Tippenhauer, "On Attacker Models and Profiles for Cyber-Physical Systems" In *Computer Security – ESORICS 2016. ESORICS 2016*. Lecture Notes in Computer Science, vol 9879. Springer, Cham, 2016.
4. H. Alemzadeh, et al., "Targeted Attacks on Teleoperated Surgical Robots: Dynamic Model-Based Detection and Mitigation," *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (*DSN*), Toulouse, pp. 395-406, 2016.
5. D. Formby, et al., "Who's in Control of Your Control System? Device Fingerprinting for Cyber-Physical Systems." *NDSS*, 2016.
6. Z. Zhang, et al., "Combating time synchronization attack: a cross layer defense mechanism," In *Cyber-Physical Systems (ICCPS), 2013 ACM/IEEE International Conference on*, pp. 141-149. IEEE, 2013.
7. F. Dörfler, J.W. Simpson-Porco, F. Bullo, "Breaking the hierarchy: Distributed control and economic optimality in microgrids," *IEEE Transactions on Control of Network Systems*, *3*(3), pp.241-253, 2016.
8. A. Cárdenas et al., "Attacks against process control systems: Risk assessment, detection, and response," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, pp. 355-366, 2011.
9. C. Markman, A. Wool, and A. A. Cardenas, "A New Burst-DFA model for SCADA Anomaly Detection," In *Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and Privacy* (*CPS* '17), 2017.
10. M. A. Umer, et al., "Integrating Design and Data Centric Approaches to Generate Invariants for Distributed Attack Detection," In *Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and Privacy* (*CPS* '17), 2017.
11. D. I. Urbina, et al., "Limiting the Impact of Stealthy Attacks on Industrial Control Systems," In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (*CCS* '16), ACM, pp. 1092-1105, 2016.
12. D. Halperin, et al, "Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses," in *IEEE Symposium on Security and Privacy (S&P)*, pp. 129-142, 2008.
13. D. Fauri, et al., "From System Specification to Anomaly Detection (and back)," In *Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and Privacy* (*CPS* '17), 2017.
14. K. Cho and K. G. Shin, "Viden: Attacker Identification on In-Vehicle Networks," In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (*CCS* '17), 2017.
15. T. P. Vuong, G. Loukas, and D. Gan, "Performance evaluation of cyber-physical intrusion detection on a robotic vehicle," *IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing* (*CIT/IUCC/DASC/PICOM*), 2015.
16. A. Teixeira, I. Shames, H. Sandberg, K.H. Johansson, K.H., "A secure control framework for resource-limited adversaries," Automatica, *51*, pp.135-148, 2015.
17. M. Pajic, et al., "Robustness of attack-resilient state estimators," *ACM/IEEE 5th International Conference on Cyber-Physical Systems (with CPS Week 2014),* pp. 163-174, 2014.

18. M. Caselli, et al., “Specification Mining for Intrusion Detection in Networked Control Systems,” 25th USENIX Security Symposium, 2016.
19. E. Ronen, et al., “IoT Goes Nuclear: Creating a ZigBee Chain Reaction,” 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, 2017, pp. 195-212.
20. K. Lebeck, et al., “Securing Augmented Reality Output,” 2017 IEEE Symposium on Security and Privacy (SP), pp. 320-337, 2017.

Hui Lin is an Assistant Professor at the Computer Science and Engineering Department in the University of Nevada at Reno. He earned his Ph.D. degree from the University of Illinois at Urbana-Champaign in 2017 in electrical and computer engineering. His research interests include cyber security, intrusion detection systems, and software-defined networking (SDN) in the areas of cyber-physical systems, such as power systems. He has successfully adapted Bro, a runtime network traffic analyzer, to support network protocols (e.g., DNP3) commonly used in power grid infrastructure. The DNP3 analyzer that he developed has been included in Bro and can be downloaded freely by utility companies. His current work focuses on applying SDN in cyber-physical systems; he intends to use SDN’s network programmability to design flexible cyber-physical systems which can quickly respond to cyber-attacks and accidents. Contact him at hlin2@unr.edu.

Homa Alemzadeh is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Virginia. Before joining UVA, she was a research staff member at the IBM T. J. Watson Research Center. Her research interests are at the intersection of computer systems dependability and data science, in particular data-driven resilience assessment of cyber-physical systems and safety and security validation and monitoring in medical CPS. Alemzadeh received her Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign and her B.Sc. and M.Sc. degrees in Computer Engineering from the University of Tehran. She is a member of IEEE, the IEEE Computer Society, the IEEE Engineering in Medicine and Biology Society, the IEEE Women in Engineering, and ASEE. Contact her at alemzadeh@virginia.edu.

Zbigniew Kalbarczyk is a Research Professor at the Coordinated Science Laboratory of the University of Illinois at Urbana-Champaign. Dr. Kalbarczyk’s research interests are in the area of design and validation of reliable and secure computing systems. His current work explores emerging technologies, such as resource virtualization to provide redundancy and assure system resiliency to accidental errors and malicious attacks. His research also involves analysis of data on failures and security attacks in large computing systems, and development of techniques for automated validation and benchmarking of dependable and secure computing systems using formal (e.g., model checking) and experimental methods (e.g., fault/attack injection). He served as the Program Chair for the International Conference on Dependable Systems and Networks (DSN) in 2002 and 2007. He is an Associate Editor of IEEE Transactions on Dependable and Secure Computing. Dr. Kalbarczyk has published over 130 technical papers and is regularly invited to give tutorials and lectures on issues related to design and assessment of complex computing systems. He is a member of the IEEE, the IEEE Computer Society, and the IFIP Working Group 10.4 on Dependable Computing and Fault Tolerance. He can be reached at kalbarcz@illinois.edu.

Ravishankar Iyer is the George and Ann Fisher Distinguished Professor of Engineering at the University of Illinois at Urbana-Champaign. He holds appointments in the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory (CSL), and the Department of Computer Science, serves as Chief Scientist of the Information Trust Institute, and is affiliate faculty of the National Center for Supercomputing Applications (NCSA). He currently co-leads the CompGen Center at Illinois. Professor Iyer is a Fellow of the American Association for the Advancement of Science, the IEEE, and the ACM. He has received several awards, including the AIAA (American Institute for Aeronautics and Astronautics) Information Systems Award, the IEEE Emanuel R. Piore Award, and the 2011 Outstanding Contributions award by the Association of Computing Machinery’s Special Interest Group on Security. Professor Iyer is also the recipient of the degree of Doctor Honaris Causa from Toulouse Sabatier University in France. Contact him at rkiyer@illinois.edu.