Safety-Critical Cyber-Physical Attacks: Analysis, Detection, and Mitigation

Hui Lin, Homa Alemzadeh, Daniel Chen, Zbigniew Kalbarczyk, Ravishankar K. Iyer
Coordinated Science Laboratory, University of Illinois at Urbana-Champaign,
1308 W. Main Street, Urbana, IL, 61801
{hlin33, alemzad1, dchen8, kalbarcz, rkiyer}@illinois.edu

ABSTRACT

Today's cyber-physical systems (CPSs) can have very different characteristics in terms of control algorithms, configurations, underlying infrastructure, communication protocols, and real-time requirements. Despite these variations, they all face the threat of malicious attacks that exploit the vulnerabilities in the cyber domain as footholds to introduce safety violations in the physical processes. In this paper, we focus on a class of attacks that impact the physical processes without introducing anomalies in the cyber domain. We present the common challenges in detecting this type of attacks in the contexts of two very different CPSs (i.e., power grids and surgical robots). In addition, we present a general principle for detecting such cyber-physical attacks, which combine the knowledge of both cyber and physical domains to estimate the adverse consequences of malicious activities in a timely manner.

1. INTRODUCTION

In today's cyber-physical systems (CPSs), control operations involve complex interactions between cyber domain controls and physical domain processes. As shown in Figure 1, measurements collected from the physical processes are used as an input to the control algorithms to update the models of the physical processes in the cyber domain. Based on the current model and estimation of the state of physical processes, the control algorithms generate commands to adjust the state of the physical processes.

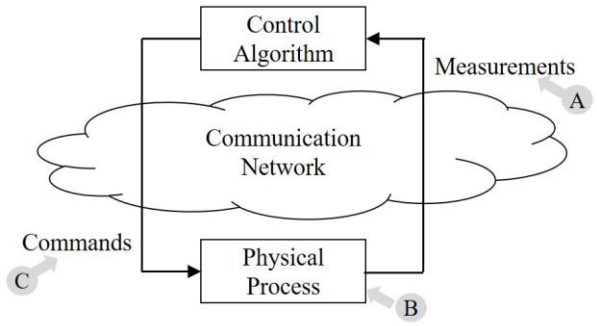


Figure 1. Cyber-physical system control.

HotSoS '16, April 19-21, 2016, Pittsburgh, PA, USA

DOI: http://dx.doi.org/10.1145/2898375.2898392

Figure 1 depicts a generic CPS's control loop and the most likely entry points (marked as *A*, *B*, and *C* in Figure 1) for attackers to penetrate the system. In attacks that compromise measurements (often referred to as false or bad data-injection attacks, marked as type *A* in Figure 1), the attackers try to mislead the control algorithm by corrupting the cyber system states [13][19] and, thus cause a wrong command to be issued to the physical process. Examples of the impact of false data injection attacks, in terms of disrupting control operations and potential economic losses, are studied in [24][26].

Type *A* attacks frequently aim at indirect changes of the commands issued to the physical process. However, in today's CPSs, commands are often transmitted over IP-based control network on unprotected communication channels. If an attacker can gain access to the control network or the communication link between the cyber and physical components, the attacker can disrupt the system by directly compromising the control commands (type *C* attack). This is not to say that the attacks on sensor measurements are not important. Quite the opposite, compromised measurements can be used to hide the real (potentially anomalous) state of the power grid in order to delay the detection of the attacks before the actual damage to the system (as seen in the example of Stuxnet [9] and in the recent study [16]).

To identify and rank the attacks that exploit the vulnerabilities in physical components (marked as type *B* in Figure 1), many researchers proposed metrics, which can be used to uncover different types of vulnerabilities [27][28]. For example, power system's electrical characteristics, such as the load of substation or transmission lines, can be used to understand how an overloading event, caused by cyber-attacks, could cause a safety violation. Additionally, previous research studied the characteristics of the transmission network (e.g., connectivity or the length of the shortest path between substations) to specify how malicious attacks can propagate through CPSs [11][15].

Instead of perturbing physical components simultaneously, previous research analyzes in type *B* attacks that an adversary perturbs physical components in sequence. A brief discussion on the risk of the cascaded outage caused by accidents or attacks is presented in [25]. Zhang et al. experimentally demonstrate that the cascaded attack can introduce more significant damage than the attacks that perturb multiple physical components simultaneously [28]. Note that, type *B* attacks often require physical access to the actual CPS device, which is not easy, less practical, and has a higher risk of being detected.

Our research focuses on studying type *C* attacks, in which the control fields of commands delivered over communication channels are maliciously modified, and assessing the impact of the attacks on the resiliency of CPSs. Unlike type *B* attacks that consider attacks on all combinations of physical components, we narrow down the search space to only include the components that attackers can compromise through cyber domain, to reduce the

analysis time and computation power. Unlike type *A* attacks that affect the process indirectly, modifying control fields can directly affect the physical process and thus, introduce safety violation. To make things worse, it is difficult to detect this class of attacks by solely monitoring in the cyber domain, because their modifications do not introduce any anomalies in the control flow and communication protocols.

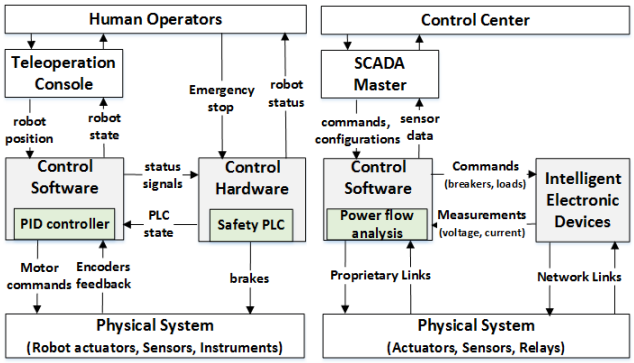
As shown in [18], the malicious modification of control commands can impact power system's steady state and dynamic behavior. In [1] we demonstrated that malicious modification of control commands in a surgical robot could cause abrupt jumps of a few millimeters in the robotic arms. If the attacker mounts the attack during a surgical procedure, it could cause catastrophic damage to the robot and harm the patient in the middle of a surgery. Another example of this type of attack is the recent incident in Ukrainian power grids, where attackers used the cyber domain to inject malicious commands, which resulted in safety violation of the grid and caused the grid to be down for several hours [2][4].

To detect such attacks in a timely manner, our approach is to combine the information from both cyber-domain simulations with physical domain process state in a smart way. Contrary to previous work, which mainly focuses on analysis and monitoring of malicious activities in the cyber domain, we believe that combining the modeling and simulation of both cyber and physical infrastructures is the key to predict the potential safety violation and can be beneficial to comprehensive study of attacks and their impacts.

In this paper, we focus on a class of attacks that impact the physical processes without introducing anomalies in the cyber domain (type *C* attacks). We discuss the common challenges in detecting this type of attacks in the contexts of two very different CPSs, namely, *power grids* and *surgical robots*. We discuss general principles for detecting such cyber-physical attacks, which combine the knowledge of both cyber and physical domains to estimate the adverse consequences of malicious activities in a timely manner.

2. OVERVIEW OF TARGET CPSs

In Figure 2, we show the control structures of two example CPSs (i.e., robotic surgical systems and power grid infrastructures) side by side to demonstrate their similarities. Both CPSs rely on a feedback control loop, in which human system operators rely on measurements from the physical systems to decide the appropriate operations.



(a) (b)

Figure 2. Example control structures for (a) robotic surgical systems and (b) power grid infrastructures.

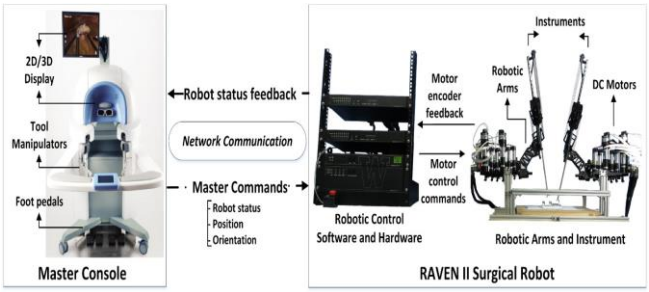
In Figure 2(a), which shows the typical control structure of a robotic system used in minimally invasive surgery, the control software receives the user commands (e.g. the desired position and orientation of the robot) through a teleoperation console and translates them into surgical movements by issuing motor commands. In Figure 2(b), which shows the common control structure used in a power grid, the control software receives the measurements of current, voltage, and power usage, estimates the electronic state, and issues commands which can adjust power system's operational conditions.

Both robotic surgical systems and power grid infrastructure share the similar feedback loops, which allow us to propose a general detection principle on common CPSs (details in Section 4). However, the implementation of the control structure and algorithms can vary dramatically in different CPSs, which implements the detection principle into ad-hoc methods for cyber-physical attacks in different systems.

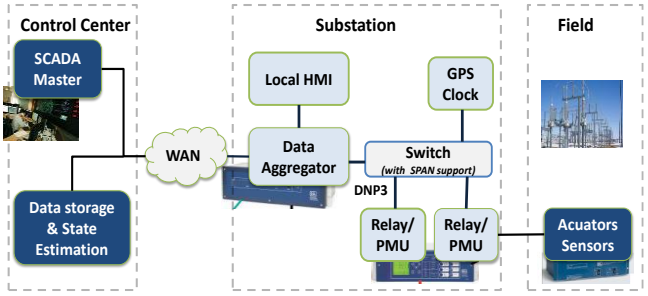
2.1 Surgical Robotic System

Surgical robots are designed as human-operated robotically controlled systems, consisting of a teleoperation console, a control system, and a patient-side cart (which hosts the robotic arms, holding the surgical endoscope and instruments).

Figure 3(a) shows the common configuration of the RAVEN II system, an open-source surgical robot [20][22]. The desired position and orientation of robotic arms, foot pedal status, and robot control mode are sent from the master console to the robotic control software over the network using the Interoperable Teleoperation Protocol (ITP), a protocol based on the UDP [12]. The control software receives the user packets, translates them into motor commands, and sends them to the control hardware, which enables the movement of robotic arms and surgical instruments.



(a)



(b)

Figure 3. Example communication structures for (a) robotic surgical systems and (b) power grid infrastructures.

The control software runs as a node (process) on the Robotic Operating System (ROS) middleware [21] on top of a real-time (RT-Preempt) Linux kernel. It communicates with the physical motor controllers and a Programmable Logic Controller (PLC) through a hardware interface (two custom USB interface boards).

The interface boards include commodity programmable devices, digital to analog converters, and encoder readers. The PLC controls the fail-safe brakes on the robotic joints and monitors the system state by communicating with the control software.

The RAVEN control system starts with an initialization phase before getting ready for the operation. During the initialization phase, the mechanical and electronic components of the system are tested to detect any faults or problems. After successful initialization, the robot enters the "Pedal Up" state, in which the robot is ready for teleoperation but the brakes are engaged. When the human operator presses the foot pedal on the master console, the robot moves to the "Pedal Down" state. In this state the brakes are released from the motors, allowing the teleoperation console to control the robot.

Control algorithm. In each control loop, the current state (position and orientation) of the end effector on each robotic arm is estimated based on the encoder readings from its joints using the forward kinematics function. The user-desired end-effector positions and orientations (received from the surgeon console) are translated to the joint and motor positions using inverse kinematics calculations. Then the amount of torque needed for each motor to reach its new position is obtained from a Proportional-Integral-Derivative (PID) controller that minimizes the error between the desired and measured torque values. Finally, the motor torque commands are transferred in the form of DAC commands to the motor controllers on the USB interface boards, to be executed on the physical motors on each arm.

Time constraints. The robot control software must complete each iteration of computing the new position of the robotic arms within time less or equal to 1ms.

2.2 Power Grid Infrastructure

A power system is composed of buses (representing substations) that are connected through transmission lines. We use the voltage magnitude and angle for each bus to represent the operational conditions of a power grid.

Figure 3(b) shows a common communication structure used in a power grid, which has three major parts: control center, substations, and field sites. The Control Center uses SCADA (Supervisor Control And Data Acquisition) Master, which collects data from Substations, analyzes the data (using the state estimation software), and issues commands (opening/closing breakers or adjusting generations) to devices in substations to maintain and control operation of the grid.

A substation can contain various intelligent electronic devices (IEDs; e.g., relays, phasor measurement units (PMU), GPS clock, etc.). These IEDs can run off-the-shelf operating systems and communicate with each other over IP-based network. On the other hand, IEDs are also connected to actuators and sensors through proprietary links to monitor the electric state at field sites.

The control center is connected to substations through a wide area network (WAN) as substations can be distributed in a large geographic area. Traditionally, this control-network is not open to the public Internet. However, to boost control efficiency, the control network is often connected through corporate networks of a power system or through personal devices (e.g., field engineering laptop operated by engineers working at field sites).

Control algorithm. To describe the physial process of a power grid, 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. The power-flow equations are nonlinear; solving them can obtain the steady state of a power system. There are two groups of approaches to solve power-flow equations. AC power-flow analysis uses iterative algorithms (e.g., Newton-Raphson algorithm) to calculate solutions that are within a predefined error threshold. DC power-flow analysis solves the linear approximation of the power-flow equations in order to get the solution more quickly.

Time constrains. In power grids, the requirements to deliver measurements or control commands can range between hundreds of milliseconds to several seconds [10]. For example, commands to protect devices against short-circuit faults are required to deliver with 166 milliseconds while commands issued by control centers to operate devices in substations can take several seconds to finish.

Discussion. The intrusiveness of the control algorithms vary in different CPSs. Some cyber domain commands may only tune the inputs to the physical process while others may significantly modify the configuration of the physical process [6][7]. For example, in surgical robotic systems, control commands are input values of differential equations, which specify the movement of rotors and joints. In power grids, however, a system administrator can directly control circuit breakers responsible for connecting/disconnecting transmission lines and thus, change the topology of transmission networks. The consequence is that the parameters, instead of inputs, of power-flow equations are changed.

3. CHALLENGES

The control operations in CPSs rely on continuous interaction between cyber and physical components, which present new challenges in detecting potential attacks launched against the system.

3.1 Attack Detectability

Cyber-physical attacks in CPS are difficult to detect by monitoring the cyber or physical domains separately from each other. Table 1 uses power grids and robotic surgery systems as examples to describe the challenges in the attack detection based on monitoring cyber or physical domains alone.

It is difficult to detect and mitigate attacks based solely on the activities from the cyber domain, due to two reasons. *First*, in many CPSs, the communication protocol in the cyber domain usually lacks security characteristics, such as encryption/authentication, due to use of legacy devices and demanding requirements of delivery time in network communication. Consequently, attackers can easily perform reconnaissance by passively monitoring the communication without generating anomaly in the cyber domain. For example, the DNP3 protocol, which is widely used in the U.S. power grids, still do not have any encryption features. *Second*, the compromises of the physical process can be crafted by changing one valid control command to another valid command, without violating any protocol syntax, control flow, or the performance of communication. For example, modification of a single bit in the DNP3 packets that deliver commands to control the circuit breakers, can change the on/off state of the breaker. Consequently, the existing intrusion detection systems that usually rely on the anomaly of the syntax (such as the length of the commands or range of a field in network packets) or signatures of abnormal events can become ineffective against such compromises [8]. Similarly, surgical robots rely on unprotected serial links (e.g., USB, RS232, or FireWire) for transferring commands and feedback between the cyber and physical components. A maliciously crafted change in new coordinates delivered to the motors through a USB channel might not raise any anomalies in the communication protocols, but

Table 1. Challenge in Detection of Attacks in Cyber-Physical Systems

Challenges		Example Cyber-Physical Systems	
		Power Grids	**Surgical Robots**
Cyber domain	Lack of encryption and authentication mechanisms for legacy devices	Communication is in a plain text.	Leaking of user commands and state information from the unencrypted data transferred through network and serial links.
	Malicious and unsafe commands can be encoded in legitimate formats	Modification of a few bits in network traffic can maintain the correct communication syntax.	TOCTTOU (time of check to time of use) vulnerability allowing malicious modification of the control commands after they are checked by the software and before are communicated to the hardware.
	Inconsistency between the state estimation in the cyber domain and the actual state in physical process.	False data injection attacks on measurements	Lack of complex models for accurate estimation of the system dynamics and behavior of robotic joints in real-time.
	Real-time constraints on control systems	Control operations should be delivered in a few hundred milliseconds.	Real-time constraint of 1 millisecond per control iteration.
Physical domain	Attacks are hard to distinguish from incidental failures and human induced safety hazards.	Contingency analysis evaluates the consequence of incidents, in which one or two physical components are out of service.	Similar safety-critical impact might occur due to unexpected physical failures or unintentional human errors.
	Inadequate knowledge of the global system state.	Periodically performing state estimation can detect the consequence of attacks based on the collected measurements. However, it is difficult for each substation to decide the impact of a command on the whole power grid.	There are limited hardware resources on the embedded computational units in the interface and the physical layer of the robot to perform sophisticated computations for estimating system state.

could cause a sudden jump in the robotic arms and damage to the physical system [1].

It is also difficult to detect and mitigate the attacks based solely on the activities from the physical domain. Today's CPSs rely on traditional safety procedures that are originally designed to remedy accidents caused by unexpected physical failures, which occur locally. However, the safety procedures can become ineffective against malicious attacks. In power grids, traditional contingency analysis considers only low-order incidents (i.e., the "*N*-1" or "*N*-2" contingency in which one or two devices are out of service). Consequently, it is impractical to construct a black list of the possible attacks for a large-scale system, which could cause coordinated failure across the grid. On the other hand, surgical robots have a hard limit on the maximum allowable torque threshold for the physical motor; however, this cannot detect malicious modification of the motor command values that are within the range specified by the threshold but still cause deviations that result in safety violation.

3.2 Diagnosis

Attacks are hard to distinguish from incidental failures and human-induced safety hazards. For example, a malicious attack on a surgical robot by carefully changing the motor torque commands could result in a sudden jump of the robotic arm. Similar sudden jump behavior due to unexpected physical failures or unintentional human errors are also observed in actual practice [1]. Furthermore, although many cyber-physical attacks cause safety violations, the violations themselves do not reveal the entry point of the attacks and the malicious activities in the cyber domain. Without such information, it is a challenge to identify the vulnerability exploited by attackers and thus, to perform the appropriate response or remedy actions (e.g., software patching or updating operational procedures).

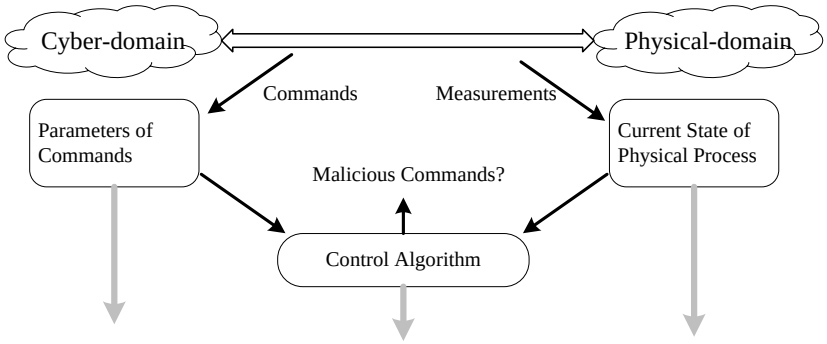
3.3 Real-Time Constraints

Cyber-physical systems usually have strict requirements on timely delivery of control operations. However, those requirements can span across different ranges. For example, power grids need to deliver the commands in the range from several hundred milliseconds to several seconds [10], while the surgical robots are required to perform control computations within only a few milliseconds [1]. As a result, it is difficult to propose a runtime detection mechanism that is appropriate for all range of CPSs. With stringent real-time constraints on the control system operation, any real-time detection and mitigation actions must complete within those constraints to avoid deviation in system dynamics, leading to potential damage [1].

4. DETECTION PRINCIPLE

In this section, we describe the detection principle (see Figure 4) and its realization in the context of power grid infrastructure and surgical robotic systems. Because attacks are initiated in the cyber domain and manifest in the physical domain, the detection mechanisms should combine the knowledge (and runtime data) from the two domains to capture a complete system view and enable attack detection. Specifically, we integrate security monitoring in the cyber domain with the control algorithms used by the physical domain to estimate the consequences of suspicious activities.

As shown in the top flow chart in Figure 4, we obtain two pieces of information from the communication between cyber domain and physical domain (i.e., *commands* and *measurements*). From the



	Commands	Control Algorithm	Measurements
Common Principle	Increase visibility in the cyber domain	Estimate (ahead of time) the consequence of command execution	Increase the integrity of measurements
Surgical Robotic Systems	Intercept commands sent by control software	Model the robot manipulator dynamics with selected degrees of freedoms (e.g., three degrees of freedoms in our case)	Retrofit hardware interface board (custom USB board) in order to deliver measurements to the detection module.
Power Grids	Integrate network monitors (e.g., Bro) with SCADA protocol (e.g., DNP3 or Modbus) analyzers Classify critical and noncritical commands	Use "*N*-1" contingency analysis to decide relative severity of an attack Dynamically adjust the number of iterations in AC power flow analysis to balance detection accuracy and latency	Compare measurements observed by network monitors placed at different locations on the network in order to validate the integrity of measurements

Figure 4. Detection principle and its application to target CPSs.

measurements, we estimate the current state of physical processes; from the commands, we extract the parameters related to control operations. Based on the measurements and the commands' parameters, the control algorithm estimates (ahead of time) the system impact of the control command execution and hence, allows us to determine whether the command is malicious.

In the table in Figure 4, we explain the detection principle and its application in the two target CPSs. The first row of the table ("Common Principle") gives common principles that can be applied to accurately observe commands, collect trusted measurements, and build control algorithms. The second ("Robotic Surgery") and the third ("Power Grids") rows summarize the implementation of the identified principles in the two target CPSs.

Observability of commands. In order to accurately obtain the parameters of commands, we need to increase the visibility in the cyber domain, which includes the control software, communication network, and computing platforms. Many current CPSs use proprietary protocols, which network monitors cannot fully understand. The goal of increasing the visibility is to improve our awareness and understanding of *what is really happening* rather than *what we believe should have happened* in the cyber domain. Also, we can obtain a better understanding of the interactions between the cyber and physical components, which can help in designing efficient and effective detection mechanisms against the targeted attacks.

Collection of measurements. Trusted measurements are essential to make an accurate estimate of the impact of the control commands on the system state. However, collecting trusted measurements is not easy, as many attacks (marked by "*A*" in Figure 1) focus on compromising measurements of CPSs to reduce observability of physical domain. Consequently, on detecting cyber-physical attacks, we can take advantage of the detection methods proposed to protect the integrity of measurements [5].

Control algorithm. We need to employ the control algorithms and estimation techniques to look ahead to the changes in states and the dynamics of the physical system upon execution of control commands. The operation of physical systems (e.g., the power flow in power grids or the movements of robotic arms in surgical robots) can be accurately estimated using nonlinear dynamic models of the system. Most control algorithms rely on the computation of differential equations to run such models, which can take a long time to finish and thus, make real-time monitoring difficult. Even though existing optimization techniques and linearized models can reduce the computation cost of the state estimation, fusing the information on the activities observed in the cyber domain (e.g., the network activities) with multiple estimated measurements from the physical domain can further optimize the computation and reduce the detection latency.

Discussion. Note that this detection principle complements the ongoing efforts to secure the computing environment in CPSs, such as using virtual private networks and adding encryption and authentication features to communication channels. In the cases in which an insider attacker can bypass such security mechanisms (e.g., the Stuxnet attackers obtained a valid security certificate [9]), the detection technique proposed here can help to reveal the malicious intentions behind activities that appear normal to system operators but are unsafe when propagated to the physical system.

4.1 Detection in Robotic Surgical Systems

Observability of commands. We retrofitted the hardware interface board (custom USB board) in the control system of the RAVEN surgical robot such that the detection mechanism based on the

dynamic model (details in the next paragraph) receives all control commands sent by the control software and monitors them before they are executed on the physical robot.

Collection of measurements. As shown in [1], software running in the programmable microcontroller (e.g., firmware) of the hardware interface board can become an attack target. Once attackers penetrate the interface board, they can compromise measurements, to indicate the wrong physical state. However, this is less likely compared with attacks targeting the control software running in the cyber domain, since gaining remote access to the interface board and changing the firmware requires passing through several more barriers. One solution to ensure the integrity of the firmware is to apply remote attestation periodically [14] or to compare measurements observed at different locations.

Control algorithm. To estimate the impact of the control commands, we enhanced the control algorithm and safety mechanisms of the surgical robot by developing a software module that models the dynamical behavior of the robotic actuators. To describe the physical process of the surgical robotic system, two sets of second-order ordinary differential equations were used to describe the dynamics of the robot joints, and DC motors and the corresponding cable tension for the joints, respectively. The fourth-order Runge-Kutta and explicit Euler methods were used to calculate the solutions to these equations using the numerical integration solvers. The challenge in developing the model was to be able to perform estimations within the time constrains of the robot's single iteration through the control loop (1 *ms* for the RAVEN II robot). To reduce the computational cost while maintaining the model accuracy as well as the system real-time guarantees, we modeled the robot manipulator dynamics using the first three (out of seven) degrees of freedom only (two rotational joints plus one translational joint). This is reasonable because the first three joints are positioning joints that contribute most to the instruments' end effectors' positions, whereas the other four degrees of freedom are instrument joints, mainly affecting the orientation of the end effectors.

Our experiments showed that we can more accurately and preemptively detect the adverse consequences of control commands in the physical system (e.g., abrupt jumps of robotic arms) compared with the existing software safety checks and emergency stop in the RAVEN II robot. Furthermore, with the help of the simplified model, we can also complete the state estimation and detection within the real-time constraints of each control loop.

4.2 Detections in Power Grid Infrastructure

Observability of commands. To accurately obtain the parameters of control commands, we extended Bro, a runtime network traffic analyzer, to support DNP3 and Modbus, network protocols widely used in U.S. power grids. The analyzers allowed us to extract semantics related to control operations from network traffic [17]. Consequently, we distinguished critical control commands that can operate devices in substations and thus, change operational conditions of the power grid.

Collection of measurements. To obtain trusted measurements from substations, we depolyed network analyzers in both control center and substations in the power grid. By comparing the measurements observbed at different locations, we ensure that the measurements are free from corruptions. Furthermore, we can apply methods proposed to detect false data injection attacks to further protect the inegrity of measurements [5].

Control algorithm. To estimate the consequence of commands, we used power-flow analysis to estimate the state of power grids upon executions of the commands. One critical challenge was that existing algorithms proposed for power-flow analysis have fixed parameters; using these algorithms, the detection latency could not always meet the real-time requirements of delivering control 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 [18]. Specifically, we adapted the number of iterations that the iterative algorithm in AC power-flow analysis used to estimate the power system state. Instead of statically fixing this parameter (e.g., being fixed by one loop of iteration in [3]), we dynamically adapted the number of iterations based on the parameters of control commands observed at runtime. Specifically, when a disturbance of multiple devices is observed, the number of iterations to analyze it is assigned as the *average number of iterations* that the classical AC power-flow analysis takes to analyze the disturbance of each involved device (i.e., the *N*-1 contingency analysis). By dynamically adjusting the number of iterations, we can save computation time to perform accurate detection on more severe perturbations. Our experiments demonstrate that the adaptive algorithm can reduce computation time by fifty percent compared with the classical AC power-flow analysis and increase the accuracy by two orders of magnitudes compared with DC power-flow analysis.

4.3 Response to Attacks

Unlike the general computing environment, it is difficult to remedy the impact of safety-critical attacks in CPSs. Consequently, responses to detections need careful design.

In this section, we study the timeline of steps that occur when a control command is executed in a CPS (shown in Figure 5); we use this timeline to analyze two categories of response mechanisms: *stop command execution* and *reverse command execution*.

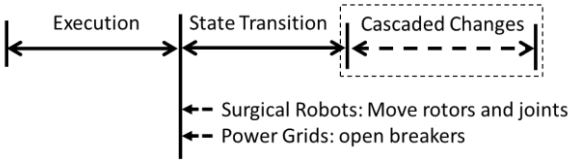


Figure 5. Timeline of steps of executing a command.

The *Execution* stage encompasses the delivery and execution of a control command. In surgical robotic systems, commands can result in the movement of rotors or joints. In power grids, commands make changes to substation devices (e.g., opening circuit breakers). The length of the execution stage can vary for the two CPSs. Surgical robots and their control platforms are usually located in the same network environment (e.g., in a hospital network); its execution stage can last not more than a few milliseconds. In a power grid, substations can be located in a large geographical area; the wide area network communication can make the execution of commands last for hundreds of milliseconds.

After command execution, CPSs can experience transient changes and ultimately reach a new steady state. This period is represented by the *State Transition* stage. The state transition can be described by differential equations, which can estimate the new steady state if the corresponding CPSs become stable. Even if CPSs are stable, their new steady state can still introduce safety violations. Examples include the rotors in surgical robots moving out of safety range or transmission lines in a power grid being overloaded. When safety violations happen, CPSs can use existing safety procedures to remove the violations. For example, surgical robots can perform emergency stops if rotors move out of the safety range. In power

grids, safety procedures can avoid safety violations within a substation (e.g., by disconnecting the overloaded transmission lines). However, these safety procedures in power grids can introduce cascaded changes (represented by the *Cascade Changes* stage) as more overloaded transmission lines are disconnected. These cascaded changes can put power grids in an unexpected state or even cause a physical damage.

Although the *stop-command* mechanism prevents malicious physical changes from being initiated, it requires the attack detection to complete before a command is executed. Such a response mechanism puts a strict time constraint on detection. On the other hand, the *reverse-command* mechanism allows commands to execute first, and then remedy its impact after the commands are determined to be malicious. This response mechanism gives the detection algorithm slightly more time to evaluate the impact of the command. However, it increases the risk of being unable for timely recovery.

The implementation of control structures in surgical robotic systems and power grid infrastructures requires them to use different response mechanisms. In surgical robotic systems, the consequence of the command can have an instant negative impact on patients (e.g., an abrupt jump of the robotic arm may cause serious injury to the patient). Consequently, in the surgical robot, we use the *stop-command* mechanism to handle malicious commands. The proposed dynamic model-based detection was directly integrated with surgical robots and can stop the malicious movement of rotors and joints [1].

In power grids, the intrinsic inertia of devices in substations and the use of wide area network for communications can take the grids a long time (e.g., on the order of minutes for the automatic generation control) to reach steady state. In this CPS, we use the *reverse-command* mechanism. We integrated the adaptive power-flow analysis with network traffic analyzers to enable timely detections of malicious commands. Also, the adaptive power-flow analysis reduced the detection latency, which allowed us to take advantage of the existing mechanism in power grids to cope with accidental commands (e.g., reclosing logics deployed in intelligent relays [18][23]).

5. CONCLUSIONS

Even though CPSs can have very different characteristics in terms of control algorithms, configurations, underlying infrastructure and communication protocols, and real-time requirements, they share similar challenges in protection against malicious attacks. In this paper, we discuss two CPSs, namely surgical robotic systems and power grids infrastructure.

To overcome the challenge of detecting cyber-physical attacks, we introduce a general principle for the detection, which combines the knowledge of both cyber and physical domains to estimate the adverse consequences of malicious activities on the physical processes and prevent system damage. We discuss how to apply the identified principles to implement detection methods specifically designed for the two target CPSs.

In future work, we plan to further explore how this detection principle can be applied to other CPSs (e.g., nuclear plant or water plant) to increase their resilience against cyber-physical attacks.

6. ACKNOWLEDGMENTS

This material is based in part upon work supported by the National Security Agency under Grant Number H98230-14-C-0141, the National Science Foundation under Award Numbers CNS 13-14891 and CNS 15-45069, and the Department of Energy under Grant Number DE-OE0000780 (NETL).

REFERENCES

[1] Alemzadeh, H., Chen, D., Li, X., Kesavadas, T., Kalbarczyk, Z. T., and Iyer, R. K. Targeted Attacks on Teleoperated Surgical Robots: Dynamic Model-based Detection and Mitigation. To appear in *Proceedings of the 46th IEEE/IFIP International Conference on Dependable Systems and Networks* (Toulouse, the France's, June-July, 2016). DSN '16.http://web.engr.illinois.edu/~alemzad1/papers/Surgical_Robots_Attacks_2015.pdf

[2] Albert, R., Albert, I., and Nakarado, G. L. Structural vulnerability of the North American power grid. *Physical review E*. 69, 2 (Feb. 2004).

[3] Albuyeh, F., Bose, A., and Heath, B. Reactive power considerations in automatic contingency selection. *IEEE Trans. Power Apparatus and Systems*. PAS-101, 1 (Jan. 1982), 107–112.

[4] Assante, M. J. Confirmation of a Coordinated Attack on the Ukrainian Power Grid. January, 2016, [Online] available: https://ics.sans.org/blog/2016/01/09/confirmation-of-a-coordinated-attack-on-the-ukrainian-power-grid.

[5] Bobba, R. B., Rogers, K. M., Wang, Q., Khurana, H., Nahrstedt, K., and Overbye, T. J. Detecting false data injection attacks on DC state estimation. In *Preprints of the First Workshop on Secure Control Systems* (Stockholm, Sweden, April 12, 2010). SCS '10.

[6] Cardenas, A. A., Amin, S., and Sastry, S. Secure Control: Towards Survivable Cyber-Physical Systems. In *Proceedings of 28th International Conference on Distributed Computing Systems Workshops* (Beijing, China, June 17-20, 2008). ICDCS '08. IEEE, pp. 495-500.

[7] Cardenas, A. A., Amin, S., and Sastry, S. Research Challenges for the Security of Control Systems. In *Proceedings of 3rd Usenix Workshop on Hot Topics in Security*, (San Jose, CA, July 28-August 1st, 2008). Usenix HotSec '08.

[8] Cheung, S., Dutertre, B., Fong, M., Lindqvist, U., Skinner, K., and Valdes, A. Using model-based intrusion detection for SCADA networks. In *Proceedings the SCADA Security Scientific Symposium* (Miami Beach, FL, January, 2007). 127–134.

[9] Falliere, N., Murchu, L., and Chien, E. 2011. *W32.Stuxnet dossier*. Symantec Security Response.

[10] IEEE standard communication delivery time performance requirements for electric power sub-station automation, IEEE Std. 1646-2004, 2005.

[11] Hines, P., Cotilla-Sanchez, E., and Blumsack, S. Do topological models provide good information about electricity infrastructure vulnerability? *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 20, 3 (2010).

[12] King, H., Hannaford, B., Kwok, K. W., Yang, G. Z., Griffiths, P., Okamura, A., Farkhatdinov, I., Ryu, J. H., Sankaranarayanan, G., Arikatla, V., Tadano, K., Kawashima, K., Peer, A., Schauss, T., Buss, M., Miller, L., Glozman, D., and Rosen, J. Plugfest 2009: Global interoperability in telerobotics and telemedicine. In *Proceedings of IEEE International Conference on Robotics and Automation* (Anchorage, AK, May 3-7, 2010). ICRA '10, 1733-1738.

[13] Kosut, O., Jia, L., Thomas, R. J., and Tong, L. Malicious data attacks on the smart grid. *IEEE Trans. Smart Grid*. 2, 4 (Oct. 2011), 645-658.

[14] LeMay, M., and Gunter, C. A. Cumulative Attestation Kernels for Embedded Systems. *IEEE Trans. Smart Grid*. 3, 2 (June 2012), 744-760.

[15] Lesieutre, B. C., Pinar, A., and Roy, S. Power system extreme event detection: The vulnerability frontier. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (Waikoloa, HI, January 7-10, 2008), HICSS '08, 184-184.

[16] Li, Z., Shahidehpour, M., Alabdulwahab, A., Abusorrah, A. Bilevel model for analyzing coordinated cyber-physical attacks on power systems. In *IEEE Trans. Smart Grid*. 99 (Aug. 2015).

[17] Lin, H., Slagell, A., Di Martino, C., Kalbarczyk, Z. T., and Iyer, R. K. Adapting bro into scada: building a specification-based intrusion detection system for the DNP3 protocol. In *Proceedings of Cyber Security and Information Intelligence Research Workshop* (Oak Ridge, TN, January, 2013). CSIIRW '13.

[18] Lin, H., Slagell, A., Kalbarczyk, Z. T., Sauer, P. W., and Iyer, R. K. Runtime Semantic Security Analysis to Detect and Mitigate Control-related Attacks in Power Grids. To appear in *IEEE Trans. Smart Grid.*

[19] Liu, Y., Ning, P., and Reiter, M. False data injection attacks against state estimation in electric power grids. In *Proceedings of 2009 ACM Conference on Computer and Communications Security* (Chicago, IL, November 9-13, 2009). CCS '09, ACM, New York, NY, pp. 21 – 32.

[20] Lum, M. J., Friedman, D. C., Sankaranarayanan, G., King, H., Fodero, K., Leuschke, R., Hannaford, B., Rosen, J., and Sinanan, M. N. The RAVEN: Design and validation of a telesurgery system. *The International Journal of Robotics Research*. 28, 9 (May 2009), 1183-1197.

[21] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, F., Leibs, J., Wheeler, R., and Ng, A. Y. ROS: an open-source Robot Operating System. *ICRA workshop on open source software*. 3,3.2 (June 2009), 5-5.

[22] Rosen, D. Friedman, H. King, P. Roan, L. Cheng, D. Glozman, J. Ma, S. Kosari and L. White. Raven-II: An Open Platform for Surgical Robotics Research. *IEEE Trans. Biomedical Engineering*. 60, 4 (April 2013), 954-959.

[23] Schweitzer Engineering Laboratories, Inc. June 27th 2013. *SEL-421-4-5 Relay Protection and Automation System, Instruction Manual*.

[24] Tan, R., Nguyen, H. H., Foo, E. Y., Dong, X., Yau, D. K., Kalbarczyk, Z. T., Iyer, R. K., and Gooi, H. B. Optimal False Data Injection Attack against Automatic Generation Control in Power Grids. In *Proceedings of the 7th ACM/IEEE International Conference on Cyber-Physical Systems* (Vienna, Austria, April 11- 14, 2016). ICCPS '16.

[25] Vaiman, M., Bell, K., Chen, Y., Chowdhury, B., Dobson, I., Hines, P., Papic, M., Miller, S., and Zhang, P. Risk assessment of cascading outages: Methodologies and challenges. *IEEE Trans. Power Systems*. 27, 2 (Dec., 2012), 631–641.

[26] Xie, L., Mo, Y., and Sinopoli, B. Integrity data attacks in power market operations. *IEEE Trans. Smart Grid.* 2, 4 (Dec. 2011), 659-666.

[27] Zhu, Y., Yan, J., Sun, Y., and He, H. Revealing Cascading Failure Vulnerability in Power Grids Using Risk-Graph. *IEEE Trans. Parallel and Distributed Systems.* 25, 12 (Jan., 2014), 3274-3284.

[28] Zhu, Y., Yan, J., Tang, Y., Sun, Y., and He, H. Resilience Analysis of Power Grids Under the Sequential Attack. *IEEE Trans. Information Forensics and Security.* 9, 12 (Oct. 2014), 2340-2354.