Safety-critical Cyber-physical Attacks: Analysis, Detection, and Mitigation

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

Abstract (Poster)

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 use the vulnerabilities in the cyber domain as footholds to introduce safety violations in the physical processes. In this poster, we specifically 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 combines 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 process 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.

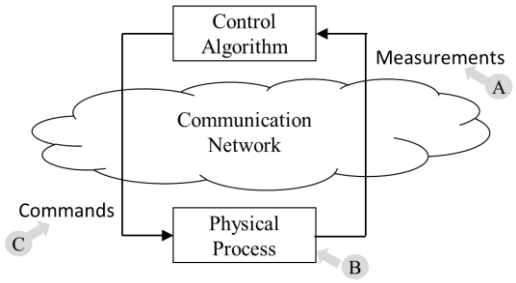

Figure 1. Cyber-physical System Control

Figure 2(a) 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. The motor commands are then sent to the hardware controllers that enable the movement of robotic arms and surgical instruments. The PLC safety processor controls the fail-safe brakes on the robotic joints and monitors the system state by communicating with the robotic software.

Figure 2(b) shows a common control structure used in a power grid. In the control center, the state estimation software collects from sensors the measurements of voltages, currents, and power usage to estimate power system's state. Based on the result of the state estimation, SCADA master can issue commands to adapt the physical configuration of power grids, for safe operation, maintenance purposes, or economic benefits.

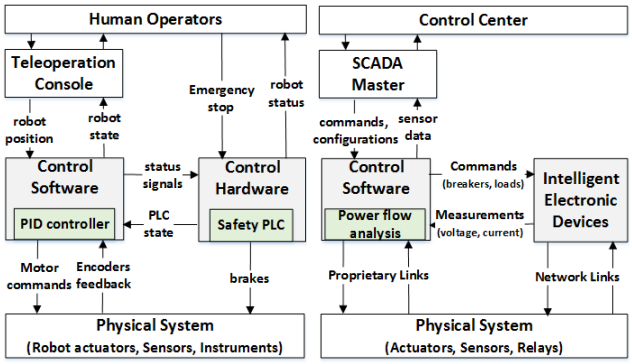

Figure 2. Example control structures for (a) robotic surgical systems and (b) power grid infrastructures

The implementation of control algorithms depends on the characteristics of the target system. For example, in the control system of a surgical robot, the amount of motor torque needed for each robotic arm to reach its new position is obtained from a controller that minimizes the error between the measured state variables (the current motor positions and velocities) and the desired states (next motor positions and velocities) [1]. In a power grid, an integer-programming model can be constructed to decide the most economical power generations.

Consequently, the intrusiveness of the control algorithms on the physical processes varies. Some cyber domain commands may only tune the inputs to the physical process while others may significantly modify the configuration of the physical process [2][3]. A typical example is in the power grid, where the system administrator can directly control circuit breakers responsible for connecting/disconnecting transmission lines and thus, change the topology of the transmission network.

Figure 1 depicts a generic cyber-physical system's control loop and shows most likely entry points (marked as A, B, and C in Figure 1). In attacks that compromise measurements (often referred to as false or bad data injection attacks, marked as type "*A*" in the Figure 1), the attackers try to mislead the control algorithm by corrupting the

HotSOS '16,

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.

cyber system states [4][5] 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 [6][7].

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 [8][9]. 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 [10]. 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.

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).

Our research focuses on studying Type C attacks, in which the control fields of commands delivered over the communication channels are maliciously modified, and assessing the impact of the attacks on the resiliency of CPSs. In particular, we focus on a class of attacks that cannot be detected 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 [11], 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 [12][13].

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.

2. 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.

2.1 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. 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 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 [14]. Similarly, surgical robots rely on unprotected serial links for transferring commands and feedback between the cyber and physical domains. A maliciously crafted change in the in new coordinates delivered to the motors through a USB channel could cause a sudden jump in the robotic arms and damage 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. 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 value that are within the threshold but still cause deviation that results in safety violation.

2.2 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-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 the attacker and thus, to perform the appropriate response or remedy actions, e.g., software patching or updating operation procedure.

2.3 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 [15], while the surgical robots are required to perform control computations within 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].

3. Detection Principle

In this section, we describe the detection principle (see Figure 3) and its realization in the context of the power grid and surgical robot CPSs. As 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 the attack detection.

For the cyber-domain, which includes the control software, communication network, and computing platforms, we need to improve our awareness and understanding of *what is really happening* rather than *what we believe should have happened* in the cyber domain through better monitoring of the network communications. Many CPSs use proprietary protocols, which network monitors cannot fully understand. By increasing the visibility in the cyber-domain, we can obtain a better understanding of the interactions between the cyber and physical components [16], which can help in designing efficient and effective detection mechanisms against the targeted attacks.

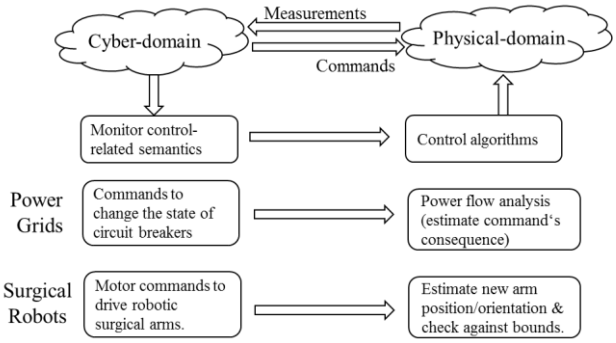

Figure 3. Detection Principle of Attacks against CPSs

On the other hand, the knowledge of physical domain is needed to estimate the real impact of attacks on the CPSs. Specifically, we need to use the control algorithms and estimation techniques to look-ahead the changes in states and dynamics of physical system upon execution of control commands. The operation of physical systems (e.g. the power flow in the smart grid or the movements of robotic arms in a surgical robot) can be accurately estimated using non-linear dynamic models of the system. Most control algorithms rely on the computation of differential equations to run such models, which can take long latency to finish and thus, make the real-time monitoring impossible. Even though existing optimization techniques and linearized models can reduce the computation cost of 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. For example, in [11], we propose adapting the power flow analysis based on the control semantics extracted from the communication network. In [1] we use a simplified dynamic model to estimate the next state of the first three joints in the surgical

robot, but with a fusion of multiple measurements, such as motor velocities and positions and joint positions, we can accurately detect the abnormal changes to the surgical arm.

4. Conclusion

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. We propose a general principle for detection of cyber-attacks, which combines the knowledge of both cyber and physical infrastructures to estimate the adverse consequences of malicious activities in the physical domain and prevent system damage.

5. REFERENCES

[1] H. Alemzadeh, D. Chen, X. Li, T. Kesavadas, Z. T. Kalbarczyk, R. K. Iyer, "Targeted Attacks on Teleoperated Surgical Robots: Dynamic Model-based Detection and Mitigation," Technical Report, Dec 2015; http://web.engr.illinois.edu/~alemzad1/papers/Surgical_Robots_Attacks_2015.pdf

[2] A. A. Cardenas, S. Amin, S. Sastry, "Secure Control: Towards Survivable Cyber-Physical Systems," in *Distributed Computing Systems Workshops, 2008. ICDCS '08. 28th International Conference on*, June 17-20, 2008, pp.495-500.

[3] A. A. Cárdenas, S. Amin, and S. Sastry, "Research Challenges for the Security of Control Systems," In *Proc. Usenix HotSec*, 2008.

[4] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 2009 ACM Conference on Computer and Communications Security (CCS)*, pp. 21 – 32.

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

[6] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Trans. Smart Grid*, vol. 2, issue. 4, pp. 659-666, Dec. 2011.

[7] R. Tan, H. H. Nguyen, E. Y. Foo, X. Dong, D. K. Yau, Z. T. Kalbarczyk, R. K. Iyer, H. B. Gooi, " Optimal False Data Injection Attack against Automatic Generation Control in Power Grids," in *Proc. The 7th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, April 11-14, 2016, Vienna, Austria.

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

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

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

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

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

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

[14] S. Cheung, B. Dutertre, M. Fong, U. Lindqvist, K. Skinner, and A. Valdes, "Using model-based intrusion detection for SCADA networks," In *Proc. the SCADA Security Scientific Symposium*, pp. 127–134, Jan 2007.

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

[16] H. Lin, A. Slagell, C. Di Martino, Z. Kalbarczyk, and R.K. Iyer, "Adapting bro into scada: building a specification-based intrusion detection system for the DNP3 protocol," in *Proc. Cyber Security and Information Intelligence Research Workshop (CSIIRW)*, 2013.