Are You Threatening my Hazards?

Marina Krotofil[1] and Jason Larsen[2]

[1] Hamburg University of Technology, Hamburg, Germany
[2] IOActive, Inc., Seattle, WA 98104, USA,

Abstract. This paper presents a framework for discussing security in cyber-physical systems, built on a simple mental model of the relationship between security and safety that has protection flows at its core. We explain their separation of concerns and outline security issues which can yield a violation of the protection flow, supporting the discussion with real world examples. We conclude the paper with a discussion on matters which are beyond our control, subjected to contradictory requirements, or do not have easy solutions. We also identify novel research challenges in the emerging field of cyber-physical security.

1 Introduction

Advances in computing and networking have rendered possible the addition of new capabilities to physical systems that could not be feasibly added before. This led to the emergence of engineering systems called cyber-physical systems (CPS): collaborative environments consisting of computational and communication elements controlling physical entities with the help of sensors and actuators. Aircrafts, robots, utilities, chemical and food plants are examples of such systems. Some cyber-physical systems are termed critical infrastructures because their functionality is critical to modern society.

While "cyberfication" contributes to improving the efficiency of physical processes, it is also a source of concerns about vulnerabilities to both random cyber failures and security attacks. On one hand embedded computers have enabled the governing of physical applications to achieve desired outcomes. On the other hand physical systems can be instructed in the same way to perform actions that are not intended. As a result software code which does not inherently possess tangible force can potentially acquire destructive capacity through the ability to instruct physical systems to malfunction. Cyber attacks on physical systems are correspondingly called *cyber-physical attacks*. The implications of this class of cyber attacks (the ability to inflict physical damage) is the main difference between cyber-physical and conventional cyber attacks. What is not always understood is that breaking into the cyber-physical system and taking over its component(s) is not enough to carry out an attack. Actually abusing the system requires additional knowledge such as a good understanding of mechanics, physics, signal processing, control principles, etc. Moreover, different types of CPS are subjected to fundamentally dissimilar failure modes which first need to be discovered.

In the context of CPS, safety systems have the critical function of detecting dangerous or hazardous conditions and taking actions to prevent catastrophic consequences on the users and the environment. The industrial control community has substantial experience in identifying and addressing potential hazards and operational problems in terms of plant design and human error, used to minimize the effects of atypical situations and to achieve a safe outcome from a situation that could have resulted in a major accident. However, the evolution of safety systems is largely built on the ability to interconnect systems and to automate notifications and alarms in the event of safety breaches. As a result, safety systems became vulnerable to cyber attacks. In the past the relationship between safety and security was studied in the context of dependable computing (Fig. 1). Compared to previous work on determining common approaches to safety and security, which had its focus on IT or system-design, see e.g. [24], [18], we also include the underlying physical processes in our considerations.

Fig. 1: Dependability and security attributes, based on [5]

Both cyber security and safety have distinct histories and have developed their own bodies of work. In both disciplines basic concepts have developed into a language that can be used to describe best practices. However, the current efforts to secure critical infrastructure have used the language of cyber security drawing little from the language of safety. Architectures are most often described in terms of security boundaries and not in terms of hazards. This cyber-oriented view of the world has been codified into standards and regulations governing process control.

One regulation illustrating this point are the NERC CIP standards [22]. Under this regulation a control system is broken down into a set of "control centers". The communications between control centers and to outside entities defines an electronic security perimeter (ESP). Not all control centers are required to be defended. Simple tests are used to determine whether a particular control center is required to be defended in compliance with the standard. However, most of the tests are cyber-oriented. The only safety-oriented test is that the control system should have the ability to shed 300 MW of load. All other hazards such as bringing a generator out of phase [29] or energizing a line during maintenance work are ignored by the standard.

The NIST 800-53a standard has a similar flavor [23]. Its general hardening recommendations such as password lengths are applied broadly to the devices

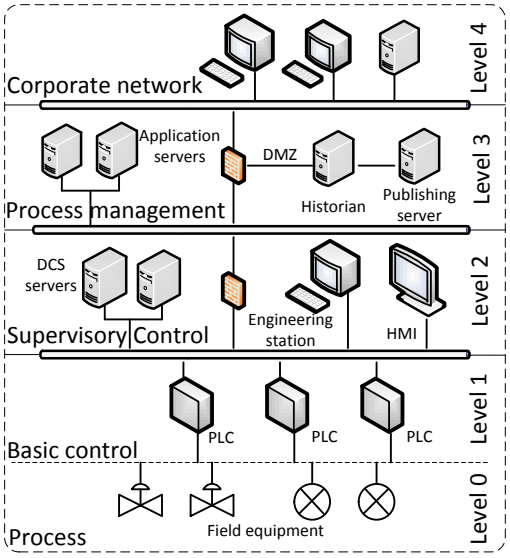

Fig. 2: ICS reference architecture

used in process control. The standard is meant to be applied to all industrial processes without any modification for the specific product being manufactured.

In both cases, there is no need for the implementer to understand the inherent hazards of the system. Hazards are simply part of the nameless devices represented by the lowest level of the Industrial Control Systems (ICS) architecture (Fig. 2) whereas cyber security exists as a barrier on top of those devices (predominantly in the form of the firewalls). One of the key dangers of this style of thinking is that all parts of the process can be grouped together into a single security compartment without any regard to how the parts of the process interact with each other and, specifically, how those interactions are key parts of the safety of the process.

The purpose of this paper is threefold. First, we propose an easily comprehensible mental model of the relationship between safety and security in cyber-physical systems. We explain the most important differences between the protective roles of safety and security and justify concerns about the violations of the direction of the protection flow. Second, we show that cyber-physical systems cannot be secured only by the means of canonical IT security approaches; the physical world of the CPS needs to be taken into consideration. In this respect the cyber-part of a cyber-physical system should not be seen as an infrastructure that needs protection from attacks in cyberspace as this leads to losing view of what is happening in the physical world and whether the system remains safe. Third, we identify novel research challenges which require solutions for improving the security posture of cyber-physical systems.

2 Disassembling the Safety and Security Nexus

'Take care of the sense and the sounds will take care of themselves.'
-Lewis Carroll, Alice's Adventures in Wonderland (1865)

In the physical world safety is a primary concern. Even before somebody is allowed to visit a plant, they usually watch a safety training video. Many industrial companies have a large screen displaying the number of days elapsed since the last safety accident. Very few, if any, companies address cyber security concerns in the same way. Security is still traditionally seen as an IT-issue, predominately concerned with protecting emails and the data on the enterprise servers. Security is often accompanied with the term privacy reflecting its information-centric approach.

Connection of safety systems has allowed processes to become more efficient and has become de facto a component of a facility's infrastructure. The growing body of regulations and standards is a direct result of the importance being placed on safety. Safety systems are not just reactive alert systems responding to a crisis, but also a proactive and predictive way of avoiding disastrous situations. It is unthinkable that these systems may fail to alert when the need arises. However software intense industrial systems and communication technologies have opened up pathways for external security threats to impact the safety of the system. Physical systems can now be attacked through cyberspace and cyberspace can be attacked though physical devices. In order to meet the challenge of securing cyber-physical systems, the security community needs to develop an understanding of safety and security that is not wholly derived from either computer science or safety engineering. It should, in fact, evolve into a new discipline which merges the fundamental concepts from each and inject new ideas of its own.

Safety and security are sometimes described as two sides of the same coin [9]. If the attacker can compromise safety systems through cyberspace and prevent them from performing their intended protection function, a security incident may lead directly to a catastrophic event. Security and safety are interconnected but both have different missions and employ different vocabularies. In order to understand their "separation of concerns" we examine the purpose and the properties of both security and safety.

2.1 Safety

Safety measures are intended to protect against *hazards*, while security measures protect against *threats*. In the safety field, hazards present a risk to a tangible entities such as human health, environment and machinery. Hazards are closely related to the concept of energy release or its change. The energy might be mechanical, chemical, electrical, thermal, kinetic, etc. An incident develops when an uncontrolled energy hits a human body, environment or material assets. Hazardous situations are generally assumed to be random events caused by natural conditions such as mechanical or human failures or as a result of disturbances to

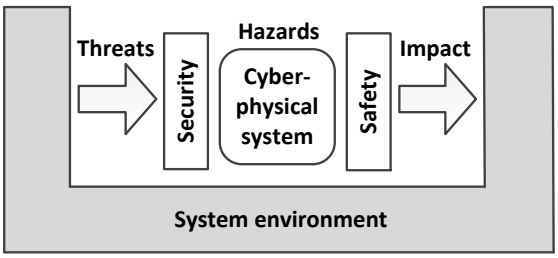

Fig. 3: Relationship between security and safety

the environment. Due to the assumption of independent failures, the safety field employs statistical methods used in reliability engineering. Once the system is in a safe state, it stays so if untouched.

2.2 Security

In contrast to hazards, IT security threats are directed at data and the supporting communication infrastructure, and do not present a direct risk of physical damage. Threats can always be traced back to humans and their will to perform a certain action. Threats may be further divided into external (e.g., hackers) and insider threats (e.g., employees of an organization). Most security incidents are caused by deliberate acts[3]. The purpose of a deliberate malicious act is forcing and incident to happen with the desire of a beneficial outcome for the attacker. It is impossible to control security threats (where, when and how an attack happens) but an organization can apply its best effort in protecting itself. Security depends on continuous updates to counter the current threats. Based on the above discussion the relationship between security and safety can be summarized in the simple model depicted in Fig. 3.

2.3 Protection Flow

The direction of the arrows in Fig. 3 represents the ideal protection flow. If security measures fail and a security incident occurs, safety precautions kick in to prevent major losses (Fig. 4). A real-life example which demonstrates accordance with the intended flow of protection is an accident at the Hatch nuclear power plant [12]. The plant was forced into an emergency shutdown for 48 hours after an engineer applied a software update to a single computer on the plant's business network. The computer was used to collect diagnostic data from the process control network and the update was designed to synchronize data between both networks. When the engineer rebooted the computer, the synchronization program reset the data on the control network. The control systems interpreted the

[3] Security also deals with unintended incidents, but the methods of preventing accidental security violations are the same as dealing with deliberate violations.

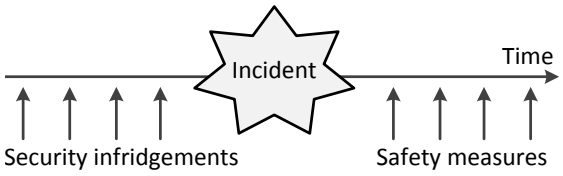

Fig. 4: Temporal relationship between security and safety

reset as a sudden drop in the reactor's water reservoirs and triggered an automatic shutdown. The nuclear plant's emergency systems performed as designed, and the cyber incident did not endanger the safety of the nuclear facility.

In practice, cyber-physical systems are complex and it is often not easy to draw a distinct line between the protective boundaries of safety and security. The physical layer can expose digital systems to cyber attacks. For example a frame relay link to a remote substation exposes the control layer to third party manipulation. The cyber layer also directly influences the safety of the physical layer. For example an interlock that stops a generator from starting when its oil pump is off is necessary for the physical protection of a generator. The field of cyber-physical security is concerned with these interdisciplinary cases. In the following sections we will discuss issues which can violate the flow of protection as well as identify interesting research problems.

3 Violations of the Model

'If everybody minded their own business, the world would go around a great deal faster than it does.'
-Lewis Carroll, Alice's Adventures in Wonderland (1865)

As safety systems were integrated into common infrastructures, it became essential to examine the way in which safety-critical data flows as it is collected, transferred, and shared. Manipulations of data may have a domino effect on the rest of the of the system from that point onward. Control data flow from the cyber systems towards the physical systems. Status and measurement data flow in the reverse direction, giving another opportunity for an attacker to impact the safety of a process. Once a cyber-physical system is created, an attacker may use it to impact the veracity, integrity, or availability of the data (Fig. 5).

3.1 Veracity

The standard armory in network security such as firewalls, secure tunnels, digital signatures, and access control provide no defense if sensor readings were manipulated before reading. When false data is submitted to the cyber infrastructure, false data will be delivered securely to the intended application. In the context

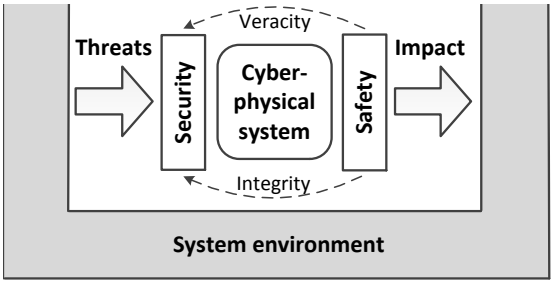

Fig. 5: Violation of the protection flow

of cyber-physical systems process data originates in the physical space and no canonical IT security solution can guarantee that the inputs from a sensor faithfully capture reality. The identification of a hazardous situation depends on the measurements of physical phenomena. If the attacker manages to manipulate sensor signals, e.g. through the manipulation of the surrounding environment or through sensor miscalibration, safety systems will not take over in a critical situation. Such an attack would violate one crucial but predominantly overlooked security property of information called trustworthiness or *veracity* [11]. To give an example, an explosion at BP Texas City Refinery killed 15 people and injured another 180 [27]. The root cause of the tragedy was critical alarms and control instrumentation providing false indications. Due to calibration errors the splitter tower level indicator showed that the tower level was declining when it was actually overfilling with flammable liquid hydrocarbons. As a result the operator kept filling in the tower. This chain of events eventually led to an explosion.

3.2 Integrity and Availability

At any point in the processing or transmission of a critical data, the integrity of that data can be compromised. In a cyber-physical system, this starts from the first point that a measurement is converted into digital form. Certain pieces of the process data must be accurate at all times in order to maintain the safety of the process. For example if a value is manipulated an interlock designed to protect the equipment may not engage. A common design in sensors is to vary the voltage or amperage in response to a physical phenomenon. The analog signal is then converted into a digital number by the controller. While a sensor can be purely analog, it is now more common for a microcontroller to be embedded into the sensor itself [21]. That microcontroller then produces a separate signal which is consumed by the controller. In all cases, the analog signal must be calibrated and scaled to transform it into a useful unit of measurement. This represents the first step where malicious actors may manipulate the data. The scaled data is then transmitted to a controller that uses it to make process control decisions. Data may be processed and combined with other data and transferred as inputs to yet additional controllers. Understanding the data source and its pathway is

essential to understanding how an attacker might cause a negative outcome. The Stuxnet worm masked its destructive operation by invading a Siemens controller and hijacking a read function. When an outside computer asked for the state of the system, the malware returned pre-recorded well-behaved process data. Such violation of the integrity property of the process data stopped higher level safety mechanisms from shutting down the centrifuges resulting in their physical damage [13].

Data can be simply missing instead of being incorrect. Especially where safety is concerned, the availability of the data is important. Safety system are designed to *detect* an unsafe condition and then *respond* to the condition appropriately. If there is a loss of data, the process can be programmed to shut down safely, but in many cases the correct action may be determined by the data. Availability is particularly threatened by the adoption of wireless technologies by process control vendors. Even if a third party cannot understand or manipulate the data, interfering with a radio transmission is simple and often done by accident.

3.3 Hidden Impact Data

A hazard is usually contained within a particular cyber-physical system. In order to determine the scope of the infrastructure that must be protected to mitigate a particular hazard, the extent of the cyber-physical systems that contains the hazard must be defined. A granular definition of a cyber-physical system can be defined by starting with the physical equipment that embodies the hazard and working back through the data flows and the logic that are required for the cyber-physical system to operate. If this analysis is done with sufficient fidelity, the cyber-physical system will be bounded to a specific set of devices, the logic that is contained within those devices, the specific pieces of data the logic operates on, and finally the communications infrastructure that transmits the data from one device to another.

Any nontrivial process will be composed of more than one cyber-physical system operating on a connected set of equipments. A cyber-physical system as defined above does not operate independently. The physics of the process connect a cyber-physical system to other cyber-physical systems. For example one cyber-physical system may generate steam and another will use that steam to produce electricity. That physical connection between two cyber-physical systems is a dependency relationship and can be seen as a data flow. This hidden data flow can lead to an improperly described cyber-physical system.

When the logic of a cyber-physical system operates on a particular datum, that datum may actually be an *aggregate* of other data even if that value was directly measured from the process. In such case the unseen data may have a negative impact on the cyber-physical system. Consider as an example a process unit that produces ammonia in which a pressure in a vessel is maintained with a pump. Maintaining the right ammonia pressure is critical to the financial health of the plant. Another pump is responsible for the outflow of the ammonia from the vessel but is not considered critical to the economy of the plant. It may be possible to set up a standing wave between the two pumps that has a direct

impact on the ability of the first pump to perform its function. In that case the state of the non-critical pump can be considered *hidden impact data* for the CPS that contains the critical pump. The unseen state of the second pump is critical to the functionality of the first pump even though there are no data flows between the two control loops. Controlling for these types of resonances is part of the standard training for process control engineers and may be obvious to them, but if the cyber-physical system was defined by tracing its electronic data flows, the state of the second pump would not be identified.

Narrowly defining a cyber-physical system may lead to cases where each individual system is protected but the system-of-systems is improperly protected. The best way to identify hidden impact data has not been investigated and may be a topic of future research.

3.4 Wireless Communication

Major vendors are adopting short range wireless networks in their product lines. This represents a major shift in both the architecture and the security of cyber-physical systems. Davis [10] showed that a practical worm could be created to attack electric power meters. The worm used the wireless interface on the meters to spread from device to device eventually giving control of the remote disconnect switches on the meters over a wide geographic area. In this case the field equipment was the source and the target of the attack. The wireless interfaces allowed the attacker to completely bypass the outer layers of security and approach the hazards directly. Wireless sensors hope to use encryption to protect the communication between nodes. History suggest that flaws will be found in these new wireless protocols [2].

While the wireless protocol certainly differ, many of them are based on the IEEE 802.15.4 standard. Mobile phones and other gadgets are already collecting data on nearby wireless networks as they travel including those based on 802.15 antennas. The databases are designed to map wireless devices that are only in close proximity periodically. This is sometimes referred to as "device social networks". The wireless sensors used in cyber-physical systems may be first mapped by a mobile device and then later attacked by it. This would be a logical extension to the work done by Leverett [16] to identify SCADA targets using the popular device search engine Shodan [20]. Encryption has long promised to make wireless hops as secure as physically wired devices, but have always fallen short of perfect security.

While this may still seem far in the future, cell phone makers are already adding software to collect proximity data of wireless devices [1]. Google will soon know that a doctor rides a bus with an industrial control technician every day even if they never interact with each other. The attackers cannot be far behind. In a recent case, the pumping stations in Oslo were converted to bluetooth for control adding a wireless component to this particular CPS [8]. In the very near future, it will be possible to plan attacks that use the antenna on the doctor's phone to compromise the technician's phone and then compromise

the pumping station. These new pathways completely bypass the known and understood security boundaries in place to stop cyber attacks.

4 Preserving flow of protection

'My dear, here we must run as fast as we can, just to stay in place. And if you wish to go anywhere you must run twice as fast as that.'
-Lewis Carroll, Through the Looking-Glass (1871)

The model suggests that control is meant to flow from the cyber systems to the physical system. In cyber-physical systems backwards data flows give the opportunity for attackers to impact the process. Minimizing or eliminating those backward flows hardens the process in respect to cyber attacks.

Logically commands come from the operator and then instruct the field equipment to perform some action. Data then flows from the field equipment back to the operator. When data is consumed by a computing device that then issues a command based on that data, a cyber-physical system has been created. These loops are the basis of modern process control. The scope of a cyber-physical system is bounded by those flows. The system contains every device, wire, and physical actuator that is involved in these flows. That would include the networking gear that carries the data as well as the valve that controls flow.

It is important to note that CPSs are fractal in structure. Consider the feedback mechanisms in a smart valve. A feedback loop is performed between the pressure sensor and the air controller to ensure the valve closes at the correct speed. The valve on the other hand is seen as a single element in a larger subcomponent of a factory. The scope of a CPS depends greatly on the definition of the "system". The choices made in designing a communications architecture can quickly add a large number of devices to a CPS, as a common infrastructure is used in the control of individual systems. A properly defined CPS should include all of the devices that can be used to manipulate it. Under this definition an attacker must interact with some part of the CPS to achieve her goals.

4.1 Ensuring Veracity

At the lowest level a physical measurement is turned into data. Veracity can be achieved in two ways. First, if the environment warrants a sufficient degree of physical protection tamper-resistant sensors can be deployed which protect the physical sensor. Second, if the environment cannot be easily controlled, countermeasures can take the form of consistency or plausibility checks on received sensor inputs [11]. Sensor readings can be false on purpose (due to attack) or by accident (e.g. wrong calibration of sensor). Although this distinction does not matter for the application, it matters for the design of the countermeasures. Redundancy and consistency checks such as majority voting have been used for detecting accidental sensor failures. With intentional attacks defenses cannot be

built on the basis of statistical independencies and may take the form of plausibility checks. In this case the models of the physical space under observation are used to judge to which extent individual sensor readings [19] are consistent with the overall state of the system derived from all the readings. It is possible to further model the relationship between different aspects of the physical process (e.g. temperature and pressure) in order to detect impossible sensor readings and flag them as suspicious. Changes in the plant configuration are not required to implement such countermeasures which make make them more practical.

4.2 Security Zones

As noted above, a granular architecture can be created by tracing specific hazards back through a cyber-physical system matching specific devices and specific pieces of data with the hazard. When the components involved have been identified, the maximum impact of an attack can be determined by examining the hazards assigned to the compromised components. If the hazard may potentially lead to a loss of life, the components should be protected more vigorously than those relating to a hazard that only results in financial loss. Components related to similar negative outcomes can be placed within a common security boundary. A new network diagram might detail a "loss-of-life security boundary" and a "financial-damage security boundary". Such boundaries would be much more useful to all parties than the traditional boundary between the "Process Control LAN" and the "Business Network".

What constitutes a hazard depends on where the boundaries of a system are drawn. The boundaries will determine which conditions and components are considered as part of the hazard. As suggested in [18] the most useful way to define the boundaries is to draw them to include the conditions related to the accident over which a system engineer has some control. This will allow to avoid the accident through eliminating or controlling the hazard.

As an example consider a robotic arm installed in an automotive factory. The robotic arm is used to paint cars on an assembly line. It poses both a chemical hazard to plant personnel (painting them) and a physical hazard (knocking them on the head). The painting chamber has a door sensor that powers off the robot if a human enters the room maintaining the safety of the system. The whole control system for the robot is complex containing multiple computers and an array of optical equipment used to automate the painting process. In contrast, the actual equipment needed to protect human life is a much smaller subset. The door sensor must accurately detect a human, the communications infrastructure must transmit the door status to the controller, and the controller logic must instruct the power relay to power off the robot. This control circuit is simple, concise, and easy to understand. Only these parts of equipment need to be secured from cyber attacks to mitigate these particular hazards.

This model has the additional utility that it immediately alerts the maintainers to the risks associated with modifications of the system. As in the above example, if the devices necessary for the protection of human life are within a separate boundary, it is inherently obvious that adding a new function to those

devices may impact human safety. If the entire robotic system is contained within the same boundary, a small change vital to safety may go overlooked. Such granular security zoning facilitates better harmonization of safety and security life cycles. Whereas security relies on frequent updates such as installing patches, upgrading firmware or adding new firewall rules, any such change in software or operational practices must be followed by a cumbersome safety revision. Failing to do so can result in casualties. Thus, after update of the SAP-based maintenance software at DuPont (without review), an alarm notifying on a hose change due date "disappeared". As a result, a hose used to transfer phosgene from a cylinder to a process wore out and catastrophically failed spraying a worker in the face and resulting in his death [26].

Once a process is analyzed there will be a number of security zones ranging from "public safety hazards" to "public relations problems". A process could be designed from the start to separate those hazards into different equipment, but in most situations individual devices will have multiple hazards associated with them. A single controller could be involved in a hazard to human life and a catastrophic financial event. In the robot painting example a single controller is likely responsible for painting the automobiles and for powering off the robot. The logical choice would be to place that controller into the most risky category namely hazards to human life. Small changes to the design of the process could be used to "downgrade" the controller from human life to financial loss. If the function of shutting down the robot was moved to a separate controller, the rest of the robot network is only a threat to the financial resources of the company. The new controller and its communications could then be more conservatively protected.

Greater protection often leads to greater costs and less flexibility. Identifying the hazards tied to each individual piece of equipment can be used to reduce costs and increase flexibility. In the original case, the entire painting network needs to be aggressively protected. Installing a vendor VPN to streamline support would open the possibility for an outside entity to endanger human life. Understanding the hazards and separating them into zones allows greater flexibility for the rest of the network. If the greatest harm from the VPN is merely financial, a simple cost/benefit argument could be made when sizing the defenses.

Even with the additional costs caused by adding and maintaining an additional device, the solution could end up being more cost effective for the implementer. Cyber security resources are scarce and expensive. More complex systems take greater resources. Being able to allocate more resources to protecting higher risk but simpler devices could result in an overall cost savings to the implementer. Also, better protected security zones, assuming that they are not invaded by an attacker, can be used for the detection of cyber-physical attacks as proposed in [3].

4.3 Eliminating Cyber-Physical Systems

Hazards with particularly severe negative outcomes can be mitigated by removing them from digital control entirely. As an example the petroleum refining

industry has a requirement that remotely operated valves be present to interrupt fuel supplies during a jet fire. The valves must be remotely operated since it is too hazardous for a human to approach a burning column of petroleum [28]. This example illustrates a key conflict. The valves need to be remotely controlled for safety, but the critical nature of the valves makes them a target for an attacker. In these cases, the remote valve could be replaced with a non-digital circuit to perform the same function. If there is no digital circuitry there is no chance for disruption via cyber means.

The U.S. Nuclear Regulator Commission is investigating the use of field-programmable gate arrays (FPGA) for critical controls [6]. Traditionally a Programmable Logic Controller (PLC) or other logic controller is used in the safety systems that protect critical hazards within a plant. The use of digital systems is flexible allowing the mitigation to be updated as the plant changes and hazards are more completely understood. It also opens up those critical controls to potential cyber attack. At some time in the near future, those controller may be replaced with FPGAs running a very discrete set of logic embodying the safety requirement. If the safety logic needs to be updated, the FPGA can be updated without the need to build and test new physics-based mitigations. This approach may provide a middle ground between purely analog systems and digital systems.

5 Discussion

'It would be so nice if something made sense for a change.'
-Lewis Carroll, Alice's Adventures in Wonderland (1865)

Even in the formal world of process control, there can be competing goals. Some conflicts are obvious. During a cyber attack, a compromised device cannot simply be unplugged. Disconnecting a part of a cyber-physical system does not guarantee that the system will eventually enter a safe state. For example a chemical reaction does not stop simply because its controller is no longer regulating the temperature. Larsen [14] has shown that a full attack payload can be miniaturized to fit into a small microcontroller located directly on the field equipment. Shutting down a controller does not guarantee that code execution stops on all the attached field devices. In addition, shutting down a cyber-physical system in response to a cyber event results in a loss-of-control for the operator. The desire to stop the attack and the desire to control the process are in conflict. Leverett [17] has shown that an attacker can gain code execution on an PLC which may be used to modify the state of an operator's display. It follows that the operator cannot rely on her displays during a cyber event. There is no consensus answer about the best course of action during such an event.

These competing goals exist at the macro level as well. The safe state for a nuclear reactor is to shut down. Nearly all the logic in a reactor control system is used to detect an unsafe condition and perform a controlled shutdown. If that same reactor is considered as a part of the larger electric grid the correct action become less clear. First consider the reactor's role during a severe winter

snowstorm. If the state of the grid is already unstable and the reactor goes offline, a blackout will occur resulting in a loss of life [4] and other externalities [25]. Second consider the same reactor during the East coast blackout. During that event reactor operators were slow to shut down even though excessive generation had resulted in the frequency of the grid had rising to 63.4Hz eventually resulting in additional outages.

Without knowledge of the state of the larger system, the safe course of action may not be known. If the state of the larger system is used as part of the safety logic, it must be imported from an external and therefore untrusted connection. If the connection is untrusted, how can it be used as part of a critical safety decision? This is an area for further study.

Currently there is no way for attackers to remotely analyze a cyber-physical system using the physics of the process. This is an area wide open for research. Although preliminary research has been already done [15], the full potential of a cyber-physical system to affect the physical world has not been explored yet. Bratus [7] has defined security violations as "unexpected computations" that can be described by so called *weird machines*. This has lead to exploitation techniques that use everything from ELF (Executable and Linkable Format) loaders to cryptographic validators to invade cyber systems. No such abstraction exists for cyber-physical systems.

Just as in the early days of cyber exploitation, it is clear that unexpected results can come from manipulating the controls of a process. If these unexpected results can be predicted and chained together, an attacker may be able to achieve heretofore unexplored results. What is lacking is an equivalent set of primitives for cyber-physical systems. In the future those primitives may be chained together to produce "unexpected physics".

6 Conclusions

Securing cyber-physical systems is challenging in the sense that they are all very dissimilar and most security problems (and solutions) exist only in a particular context. Nevertheless, we should start studying them trying to identify common patterns so that we can investigate unified solutions to address those patterns.

References

1. Configure access points with Google Location Service. `https://support.google.com/maps/answer/1725632?hl=en`
2. Project KillerBee. `https://code.google.com/p/killerbee/`
3. Safety securing approach against cyber-attacks for process control system. Computers & Chemical Engineering 57, 181 – 186 (2013)
4. Anderson, G., Bell, M.L.: Lights Out: Impact of the August 2003 Power Outage on Mortality in New York. Epidemiology 23(2), 189 –193 (2012)
5. Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. IEEE Transactions on Dependable and Secure Computing 1, 11–33 (2004)

6. Bobrek, M., Bouldin, D., Holcomb, D., Killough, S., Smith, S., Ward, C.and Wood, R.: Review Guidelines for Field-Programmable Gate Arrays in Nuclear Power Plant Safety Systems. U.S.RNC (2010)
7. Bratus, S., Locasto, M., Patterson, M.L., Sassaman, L., Shubina, A.: Exploit Programming: From Buffer Overflows to 'Weird Machines' and Theory of Computation. USENIX ;login 36(6), 13 –21 (2011)
8. connectBlue: Bluetooth Technology in Oslo Pump Stations. http://www.connectblue.com/fileadmin/Connectblue/Web2006/Documents/References/ABB_Norway.pdf (2011)
9. Cusimano, J., Byres, E.: Safety and Security: Two Sides of the Same Coin. ControlGlobal (2010)
10. Davis, M.: SmartGrid Device Security: Adventures in a new medium. Black Hat USA (2011)
11. Gollmann, D.: Veracity, plausibility, and reputation. In: Information Security Theory and Practice. Security, Privacy and Trust in Computing Systems and Ambient Intelligent Ecosystems, LNCS, vol. 7322, pp. 20–28 (2012)
12. Kesler, B.: The vulnerability of nuclear facilities to cyber attack. Strategic Insights 10(1), 15–25 (2011)
13. Langner, R.: To kill a centrifuge. Tech. rep., Langner Communications (2013)
14. Larsen, J.: Going Small When Attacking a Process. http://vimeopro.com/s42012/s4x14/video/84632472
15. Larsen, J.: Breakage. Black Hat USA (2008)
16. Leverett, É.P.: Quantitatively Assessing and Visualising Industrial System Attack Surfaces. Master's thesis, University of Cambridge, UK (2011)
17. Leverett, É.P., Wightman, R.: Vulnerability Inheritance Programmable Logic Controllers. The 2nd International Symposium on Research in Grey-Hat Hacking (GreHack) (2013)
18. Leveson, N.G.: Engineering a Safer World: Systems Thinking Applied to Safety. The MIT Press (2012)
19. Linda, O., Manic, M., McQueen, M.: Improving control system cyber-state awareness using known secure sensor measurements. In: 7th International Conference on Critical Information Infrastructure Security (2012)
20. Matherly, J.C.: SHODAN. http://www.shodanhq.com/ (2009)
21. McIntyre, C.: Using Smart Instrumentation. Plant Engineering: online magazine (2011), http://www.controleng.com/single-article/using-smart-instrumentation/a0ec350155bb86c8f65377ba66e59df8.html, retrieved: December, 2013
22. NERC: Critical Infrastructure Protection Standards. http://www.nerc.com/pa/Stand/Pages/CIPStandards.aspx
23. NIST: Guide for Assessing the Security Controls in Federal Information Systems and Organizations (2010)
24. Novak, T., Gerstinger, A.: Safety- and Security-Critical Services in Building Automation and Control Systems. IEEE Transactions on Industrial Electronics 57(11), 3614–3621 (2010)
25. Rinaldi, S., Peerenboom, J., Kelly, T.: Identifying, understanding, and analyzing critical infrastructure interdependencies. Control Systems, IEEE 21(6), 11 –25 (2001)
26. U.S. Chemical Safety and Hazard Investigation Board: DuPont Corporation Toxic Chemical Releases: Investigation Report. Tech. rep., U.S. Chemical Safety Board (CSB) (20011)

27. U.S. Chemical Safety and Hazard Investigation Board: Bp America Refinery Explosion: Final Investigation Report. Tech. rep., U.S. Chemical Safety Board (CSB) (2007)
28. U.S. Chemical Safety and Hazard Investigation Board: LPG Fire ar Valero– McKee Refinery: Final Investigation Report. Tech. rep., U.S. Chemical Safety Board (CSB) (2007)
29. Zeller, M.: Myth or reality - does the Aurora vulnerability pose a risk to my generator? In: Protective Relay Engineers, 2011 64th Annual Conference for. pp. 130 –136 (2011)