

**Position Paper for
Workshop on Future Directions in Cyber-physical Systems Security**

**Aloysius K. Mok
Department of Computer Science
The University of Texas at Austin
July 10, 2009**

Cyber-physical system (CPS) security presents new problems inasmuch as cyber-physical systems are different from either information systems (IS) or physical control systems (PCS) in that there are interactions between the information processing elements and the physical control elements that neither IS nor PCS alone may incur. In computational terms, CPS systems presents new failure modes that have not been anticipated by the designer of control systems or information processing systems. An attacker who understands these interactions may exploit the new failure modes to cause major system upsets. In particular, the likely use of wireless technology as the infrastructure for communication and control opens up a huge gap for attackers to mount unconventional attacks. New models and techniques must be developed to counter these attacks.

We can think of these attacks as strategies to disrupt the coordination between the physical state information flow and the command and control information flow of the CPS system. For example, in the Smart Grid of the future, reliable phase measurements of the power generators are critical for the safe and efficient control of the power grid. An attacker of the Smart Grid may insert itself in the information flow infrastructure of the control network and thereby able to create havoc. This can be done by, for example, jamming the wireless control signals at crucial moments and at different points of the control network, or worse still, an attacker that has succeeded in impersonating the identity of a transmitter can mount Byzantine attacks on the Smart Grid. The resulting erroneous phase measurements could cause serious damage if appropriate counter measures are not taken. While there has been significant work in applying defenses against Byzantine attacks in the literature, all the solutions must depend on the assumption that not all the nodes in the network are compromised simultaneously and that at least some fraction of the nodes (above a lower bound) must function

correctly in order to defend against the attack. As such, there is no defense that we are aware of that is effective against common-mode attacks inasmuch an attacker that can exploit a common vulnerability to compromise every node can potentially mount an attack to simultaneously compromise all the nodes and disable all the control network elements. This can happen if for example, a bad version of the control software has been downloaded into every node of the system waiting to become activated on command of the attacker. This type of attack can in theory overcome all known defenses against Byzantine attacks.

On the other hand, there are characteristics of CPS systems that are advantageous to the defender. Some of these characteristics are:

(I) The dynamics of the physical system under control is determined by the laws of physics that no attacker can alter.

(II) The rates of change of the physical state variables of the process under control may be slow in comparison to the rate of information flow in the control system, and this is usually the case with mechanical and chemical systems.

(III) The value of information in a CPS system is usually a function of time and may degrade substantially with the passage of time.

An implication of (I) is that the defender has a source of information that of itself is not subject to subversion by a human adversary through cyber attacks alone, although the information may be compromised during transmission and transformation. If we can devise security checks that use only information directly from the physical process measurements, then such checks cannot be subverted by an adversary. These checks can be the foundation of the secure computing base of the system.

An implication of (II) is the possibility for the defender to maintain through measurements and real-time simulation techniques an approximate but sufficiently accurate global view of the physical state of the system or at least be able to know within bounded time if sufficient accuracy has been lost, so that the defender can know that the state information cannot be trusted. This is analogous to the late Flaviu Cristian's approach in system design vis-a-vis his timed asynchronous model of computation [4]. The issue is how we can effectively leverage the relatively slower rate of change in the physical state to increase the probability of attack detection. We think that this is possible unless the attacker can corrupt physical state measurements on a global scale and in a synchronized fashion. In any case, it should be possible to design detection systems that build on local measurements in a hierarchical fashion and on different time scales that would make it extremely difficult for an attacker to synchronize an attack.

An implication of (III) is that the attacker has limited time to exploit the information s/he has obtained about the physical system state. Therefore the protection of state variable information does not have to be permanent. Specifically, it is only necessary for the defender to deny the attacker the information s/he seeks for as long as the information has time value. This should open up new ways to look at the design of information protection methods and architecture.

We believe that the crucial point in mounting an effective defense is in being able to correlate between global control information and local physical state measurements in time, and in switching the defensive strategy faster than the attacker can incur unacceptable physical damage to the CPS. This assumes that certain real-time constraints can be monitored either by the information flow infrastructure itself or by correlation to physical state measurements. Contrary to common wisdom, we do not regard hard timing constraints as a vulnerability in the design of secure systems. We take the view that hard timing constraints define the integrity of a system so that failure to meet a hard timing constraint means that system integrity has been breached, but any single timing failure should not result in total system failure. This approach to design secure systems is akin to the late Flaviu Cristian's "timed asynchronous system" design philosophy. The bedrock of this approach is that time synchronization failure is detectable. In the

case of CPS systems, the physical measurements performed by locally controlled physical devices can be hardened, say by hardware means, to detect timing failures as exhibited by any unexpected behavior of the physical processes under control as a function of time. A successful attacker must now coordinate the timing of the attack on the physical measurement system in order to mask the attack. This race against time turns the timing constraint requirements into an advantage for the defender inasmuch as the physics of the processes under control help define what is normal behavior against which anomalous system behavior can be compared. This gives us hope that we can design low-overhead intrusion detection and prevention systems that can defend against new classes of denial-of-service attacks such as allergy attacks [1] which would otherwise have been very difficult to defend against.

The other half of an effective defense against attacks is the design of recovery mechanisms, given that some attacks will succeed. Central to the notion of recovery is the idea of failure semantics. Here the critical insight is that a system should be designed to function according to specification not only under normal conditions but that the system should also function with high probability in some specific way when a failure occurs. For CPS systems, the traditional failure mode classification is not satisfactory inasmuch as the classification was devised with computer (hardware) systems in mind. For CPS systems, we need a classification system that is more specific with respect to the interaction between the physical components of the system under control and the computer and communications components that control them. For example, we should be able to take advantage of the fact that mechanical components usually fail on a time scale in seconds) that is large compared with the reaction time of computer systems (much less than a second) to define a CPS-specific failure classification by taking into account the amount of time that is available for defensive actions. We may for example add a time dimension to the failure mode classification by including explicitly a time-to-failure parameter. There are also other dimensions that a CPS-specific failure mode classification scheme can exploit by focusing on the application domain specifics. For example, when a traffic light fails, we do not stop all traffic but instead we put the lights in a blinking-red mode. This signals to all the cars entering the intersection to follow a pre-agreed protocol, namely, the intersection

should be regarded as being regulated by all-way stop signs. The blinking red lights thus impose a specific type of behavior on all traffic. This suggests a way to generalize failure mode semantics: we can define failure semantics in terms of protocols that the failed system must follow. These protocols can be formalized by the plethora of techniques from computer science, hybrid systems and other branches of engineering. In other words, we can define failure modes by the protocols that a failed system must follow and these protocols can be application-specific as well as have a time dimension. No defense is perfect in that the effectiveness of a defense is necessarily relative to the assumptions made about the damage that can be caused by an attacker. In CPS systems, the damage to the system can be quantified by the behavioral deviation from a healthy system. Since the speed at which an effective attack on a CPS system can be mounted is also limited by the physics of the system under control, this allows the defense to define and enforce application-specific failure semantics that will aid in the design of secure systems.

We believe that secure CPS design can benefit from the assumption that system failures are detectable by local physical measurements and that detection of security attacks can be predicated on the detection of such failures. Given this assumption, the defense against security attacks becomes a discipline of designing systems with well chosen physical failure semantics and the timely detection of such failures. What is required are design principles for specifying physically induced timing requirements and the tools for monitoring these constraints [2]. The monitoring facility should be incorporated into the secure computing base of any CPS system. This is the philosophy we have adopted in our ongoing research on wireless process control systems.

In the domain of industrial process automation, the most challenging security problems that we must face are related to the introduction of wireless technology in process control. We have been working in the software architecture for wireless process control in the last four years. In collaboration with Emerson Process Management, we have been actively involved in the engineering of the WirelessHART protocol which is one of the two major industry standards in the process management area. With the expected proliferation of wireless technology in industrial process management, it is critical that we have a solid theoretical

and engineering foundation for securing our wireless systems against attacks. We have taken the first step in this direction in our ongoing work with the WirelessHART process control stack [3]. There is a lot of work that needs to be done including:

- (1) Collaboration with the utility industry to understand and exploit the domain knowledge that is required to formulate security checks directly from the sensors to provide a first line of defense against cyber attacks.
- (2) Creating an information system architecture to exploit the correlation of process and timing information of both the IS and PCS sides of the utility system, including real-time simulation facilities.
- (3) Engineering domain-specific real-time database services that are resilient to cyber attacks, especially denial-of-service attacks.
- (4) Understanding quantitatively the tradeoff between the cost of timely detection and the damage that a utility can sustain as a function of time-to-detection and the time to execute counter-measures.
- (5) Investigating what is appropriate failure semantics that can minimize damage as well as the time to recovery.
- (6) Exploring information protection methods and architectures that can exploit the fact that the attacker can have only a bounded time window in which to synchronize and mask an attack before the attacker's data becomes stale and loses its value, e.g., by inventing light-weight cryptographic techniques that can guarantee short-term protection for any new data.
- (7) Understanding and exploiting the technical characteristics of wireless protocols and technologies for process control to facilitate implementing points 1-6 above.

The above issues call for inter-disciplinary research by researchers from the control, process automation, computer and communications security, real-time system, mathematical optimization and other areas. A sustained effort will be needed to make headway in securing current and future cyber-physical systems.

References

- [1] Simon P. Chung, Aloysius K. Mok, "Advanced Allergy Attacks: Does a Corpus Really Help?", Proceedings of the 10th International Symposium on Recent Advances in Intrusion Detection (RAID'07), LNCS 4637, pp. 236-255, Sep 5, 2007.
- [2] Honguk Woo and Aloysius K. Mok, "Real-Time Monitoring of Uncertain Data Streams using Probabilistic Similarity", Proceedings of 28th IEEE Real-Time Systems Symposium (RTSS'07), pp. 288-297, Dec 2007.
- [3] Jianping Song, Song Han, A.K. Mok, Deji Chen, M. Lucas, and M. Nixon, "Wirelesshart: Applying Wireless Technology in Real-time Industrial Process Control", Proceedings of 14th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'08), pp. 377-386, April 2008.
- [4] F. Cristian and C. Fetzer, "The Timed Asynchronous System Model" IEEE Transactions on Parallel and Distributed Systems, vol. 10 no. 6, pp. 642-657, June 1999.