Security Analysis for Process Control Systems

Zong-Syun Lin[†], Alvaro A. Cárdenas[‡], Saurabh Amin[‡], Hsin-Yi Tsai[†],
Yu-Lun Huang[†] and Shankar Sastry[‡]

[†] National Chiao Tung University, Taiwan
[‡] University of California, Berkeley

ABSTRACT

We present security analysis of process control systems (PCS) when an attacker can compromise sensor measurements that are critical for maintaining the operational goals. We present the general sensor attack model that can represent a wide variety of DoS and deception attacks. By taking example of a well studied process control system, we discuss the consequences of sensor attacks on the performance of the system and important implications for designing defense actions. We develop model-based detection methods that can be tuned to limit the false-alarm rates while detecting a large class of sensor attacks. From the attacker's viewpoint, we show that when the detection mechanisms and control system operations are understood by the attacker, it can carry stealth attacks that maximize the chance of missed detection. From the defender's viewpoint, we show that when an attack is detected, the use of model-based outputs maintains safety under compromised sensor measurements.

1. INTRODUCTION

Control systems are computer-based systems that *monitor* and *control* physical processes. These systems represent a wide variety of networked information technology (IT) systems connected to the physical world. The overall objectives of control systems are: (1) to maintain safe operational goals by limiting the probability of undesirable behavior, (2) to meet the production demands by keeping certain process values within prescribed limits, (3) to maximize production profit.

Control systems are more vulnerable today than in the past due to the increased standardization of technologies, the increased connectivity of control systems to other computer networks and the Internet, insecure connections, etc. Because of the increasing risk to computer attacks, there has been a significant effort in recent years to discuss and identify the security issues of control systems

In this proposal we focus on attacks on the *regulatory layer*. The regulatory control layer has direct access to the sensors that measure the process variables and is responsible for nominal safety and operation of the processes in the system. Since the regulatory layer controllers are required to demonstrate faster response, they are traditionally based on the classic proportional-integral-derivative (PID) algorithms.

2. OUR APPROACH

We believe that most of the previous work in the security of control systems has three goals: (1) create awareness of security issues with control systems, (2) help control systems operators and IT security officers design a security policy, and (3) recommend basic security mechanisms for prevention (authentication, access controls, etc), detection, and response to security breaches.

While these recommendations and standards have placed significant importance in the *survivability* of control systems; we argue that they have not considered new research problems that arise when control systems are under attack. In particular, researchers have not considered how attacks affect the estimation and control algorithms -and ultimately, how attacks affect the physical world.

In this work we argue that the major distinction of control systems with respect to other IT systems is the interaction of the control system with the physical world. We propose to incorporate the physical process dynamics in the security analysis of the control system and focus on an attacker that compromises sensor readings. We have two major goals (1) to develop a threat assessment methodology, and (2) to design attack detection and response mechanisms.

3. ATTACK MODELS

In this proposal we focus on attacks on sensor networks and the effects they can have on the process control system. We consider the case when the state of the system is measured by a sensor network of p sensors that observes the measurement vector $y(k) = \{y_1(k), \dots, y_p(k)\}$, where $y_i(k)$ denotes the measurement by sensor i at time k. All sensors have a dynamic range that defines the domain of y_i for all k. That is, all sensors have defined minimum and maximum values $\forall k, y_i(k) \in [y_i^{\min}, y_i^{\max}]$. Let $\mathcal{Y}_i = [y_i^{\min}, y_i^{\max}]$. We assume each sensor has a unique identity protected by a cryptographic key.

Let $\tilde{y}(k) \in \mathbb{R}^p$ denote the *received measurements by the controller* at time k. Based on these measurements the control system defines control actions to maintain certain operational goals. If some of the sensors are under attack, $\tilde{y}(k)$ may be different from the real measurement $y(k)$; however,

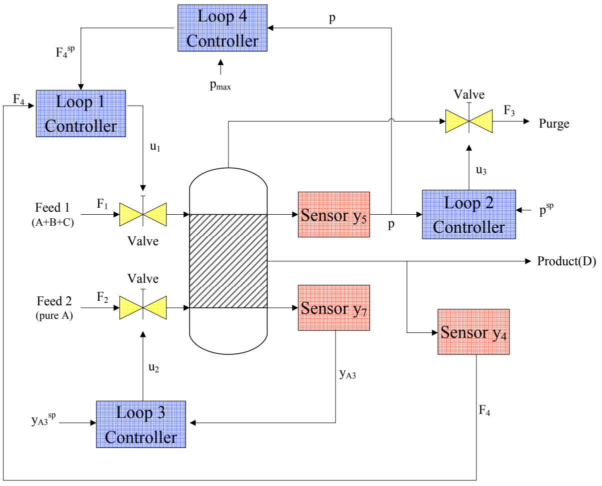


Figure 1: Architecture of the Simplified TE Plant

we assumed that the attacked signals $\tilde{y}_i(k)$ also lie within $\mathcal{Y}_i$ (signals outside this range can be easily detected by fault-tolerant algorithms).

Let $\mathcal{K}_a = \{k_s, \ldots, k_e\}$ represent the attack duration; between the start time k_s and stop time k_e of an attack. A general model for the observed signal is the following:

$$\tilde{y}_i(k) = \begin{cases} y_i(k) & \text{for } k \notin \mathcal{K}_a \\ y_i(k) + \lambda_i(k) & \text{for } k \in \mathcal{K}_a \\ y_i^{\min} & \text{for } k \in \mathcal{K}_a,\ y_i(k) + \lambda_i(k) < y_i^{\min} \\ y_i^{\max} & \text{for } t \in \mathcal{K}_a,\ y_i(k) + \lambda_i(k) > y_i^{\max} \end{cases}$$

This general sensor attack model can be used to represent a variety of attacks such as additive injection, multiplicative scaling, replay attacks and DoS attacks.

4. PROCESS DESCRIPTION

To test our attacks, we use the Tennessee-Eastman process control system (TE-PCS) model and the associated multi-loop PI control law as proposed by Ricker [1]. The process architecture and the control loops are described in Figure 1. The *control objective* is to *regulate* F_4, the rate of production of the product D, at a set-point F_4^{sp}, while maintaining P, the operating pressure of the reactor, below the shut-down limit of 3000 kPa as dictated *safety* considerations, such that C, the *operating cost* is minimized.

There are four *input variables*, denoted as u_1, u_2, u_3 and u_4, available to achieve the above control objective. Ricker [1] suggests the input-output pairings (or *control loops*) as seen in Figure 1. The PI control law for the loop$-i$ controller for the k^{th} sampling period is given by

$$u_i(k) = u_i(k-1) + K_i \left(e_i(k) - e_i(k-1) + \frac{\Delta t}{\kappa_i} e_i(k) \right) \quad (1)$$

where $e_i(k) =$ setpoint $-$ measured value of controlled variable for loop$-i$ controller at k^{th} sampling period. The controller settings K_i and κ_i are pre-tuned and given [1]. The control input vector for k^{th} sampling period is denoted as $u(k) = (u_1(k), \ldots, u_4(k))^{\top}$. We also add a Gaussian disturbance to the control inputs $u(k)$ so that the system is never in a complete steady state.

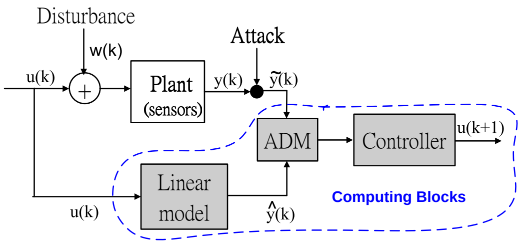


Figure 2: The proposed detection module.

5. THREAT ASSESSMENT

We study the security issues of control systems by experimenting and simulating cyber attacks on sensor signals in the TE-PCS model. Because operating the chemical reactor with a pressure larger than 3000 kPa is unsafe, it may lead to an explosion or damage of the equipment. Assume that the goal of the attacker is to raise the pressure level of the tank to a value larger than 3000 kPa, we attack a single sensor or a single controller at a given time. From the experimental results, we found that the most effective of these attacks were the max / min attacks (make the forged signals the extreme values, i.e, $y^{\max}$ or $y^{\min}$); however, not all of them were able to drive the pressure to unsafe levels. We found out that, in general, the DoS attacks do not affect the plant. We conclude that if the plant operator wants to prevent an attack from making the system operate in an unsafe state, it should prioritize the integrity of the sensors rather than their availability.

6. MODEL-BASED ATTACK DETECTION

Detecting attacks to control systems can be formulated as anomaly-based intrusion detection systems. Our proposed attack detection system is presented in Figure 2. The control input sequence $u(k)$ is fed to the physical system after being perturbed by an additive Gaussian process noise sequence $w(k)$. The process noise sequence can be thought as unmodeled factors that affect the evolution of system state. The input sequence $u(k)$ is also fed to a system model that is representative of the physical system and is internal to the detection system. The internal model will produce an output sequence $\hat{y}(k)$. The anomaly detection module (ADM) will compare the two measurement sequences: the sequence $\tilde{y}(k)$ that is received from the sensor measurements and may have been influenced by the attacker with the sequence $\hat{y}(k)$ that is obtained from the internal model. The ADM raises an alert if the deviation between the two sequences is significant.

To formalize this problem, we need (1) a linear model that is representative of the physical system, and (2) an anomaly detection algorithm. We use the linear model, characterized by the matrices A, B, and C, obtained by linearizing the non-linear TE-PCS model about the steady-state operating conditions. The model dynamics that are linear in state $x(k) \in \mathbb{R}^n$ and control input $u(k) \in \mathbb{R}^m$ are

$$x(k+1) = Ax(k) + Bu(k) \quad (2)$$

Assume that the system (2) is monitored by a *sensor network* with p sensors. We can obtain the representative measurement sequence, $\hat{y}(k) \in \mathbb{R}^p$, from the observation equations

$$\hat{y}(k) = Cx(k), \tag{3}$$

For our anomaly detection algorithm we use a change detection formulation. The problem formulation is: given a time series sequence $z(1), z(2), \ldots, z(N)$, determine the minimum number of samples, N, the anomaly detection scheme should observe before making a decision d_N between two hypotheses: H_0 (normal behavior) and H_1 (attack). Let

$$z_i(k) := |\tilde{y}_i(k) - \hat{y}_i(k)| - b_i \tag{4}$$

where b_i is a small positive constant chosen such that

$$\mathbb{E}_{H_0}[|\tilde{y}_i(k) - \hat{y}_i(k)| - b_i] < 0 \tag{5}$$

The nonparametric CUSUM statistic for sensor i is

$$S_i(k) = (S_i(k-1) + z_i(k))^+,\ S_i(0) = 0 \tag{6}$$

and the corresponding decision rule is

$$d_{N,i} \equiv d_\tau(S_i(k)) = \begin{cases} H_1 & \text{if } S_i(k) > \tau_i \\ H_0 & \text{otherwise} \end{cases} \tag{7}$$

where τ_i is the threshold selected based on the false alarm rate for sensor i.

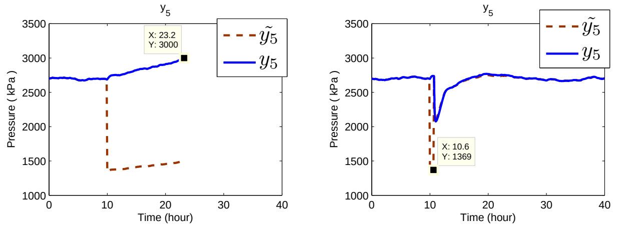
Our response strategy (shown in Fig 2) can be summarized as follows: For sensor i, if $S_i(k) > \tau_i$, the ADM replaces the sensor measurements $\tilde{y}_i(k)$ with measurements generated by the linear model $\hat{y}_i(k)$ (that is the controller will receive as input $\hat{y}_i(k)$ instead of $\tilde{y}_i(k)$). Otherwise, it treats $\tilde{y}_i(k)$ as the correct sensor signal.

7. EXPERIMENTS

In this section, we briefly discuss how our defense system works under attacks. We omit the details for determining the two parameters (b and τ) of the nonparametric CUSUM statistic. We also have to make sure that if there is a false alarm, controlling the system by using the estimated values from the linear system will not cause any safety concerns. We found that while a false response mechanism increases the pressure of the tank, it never reaches dangerous levels.

We now test the detection and response performance of the ADM for certain attacks. Because operating the chemical reactor with a pressure larger than 3000 kPa is unsafe, all our attacks attempt to raise the pressure in the tank. In order to quantify the magnitude of the attack we use multiplicative scaling attacks with parameter λ^m and attack each sensor. Our attacks start at time $T = 10$ hours. We only report attacks on y_5 here. The results for y_4 and y_7 are similar. Sensor y_5 monitors pressure of the reactor. Attacking sensor y_5 by lowering the value makes controller turn down the purge valve to increase pressure. In an unprotected system the safety of the system is compromised at time $T = 23.5$ (hr) if we set parameter of scaling attack λ^m_{y5} to 0.5. With ADM enabled, the attack can be detected at time $T = 10.7$ (hr) and the plant remains stable.

If an attacker compromises two more sensors then he can mount multiple attacks; however these attacks can also be detected independently by each statistic $S_I(k)$. As an anecdote we attack y_5 with a replay attack and y_7 with a scaling attack. In the original plant system (without ADM), Fig 4 shows that plant goes to an unsafe state at time $T = 16.2$ (hr). Compared with just launching an scaling attack on y_7, the combined attack takes much less time to drive the pressure past safety levels. The reason is that the replay attack on y_5, gives an erroneous information to the controller that tries to prevent an increase in pressure.

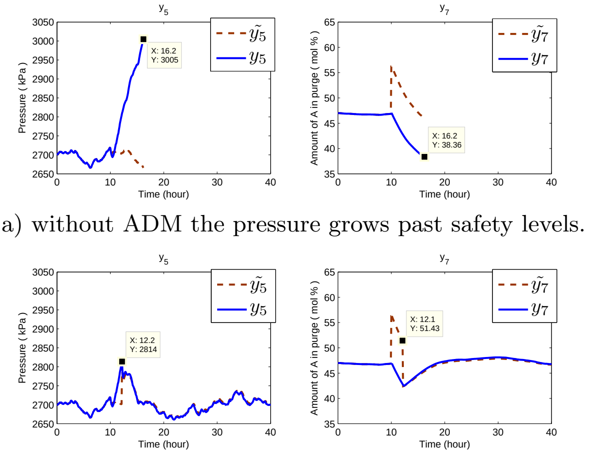


(a) Without ADM

(b) ADM detects and responds to the attack at $T =$ 10.7 (hr)

Figure 3: $\tilde{y_5} = y_5 * 0.5$

If we have an ADM the attack is detected by $S_{y5}(k)$ at time $T = 12.2$ (hr) and independently by $S_{y7}(k)$ at time $T = 12.1$ (hr).



(a) without ADM the pressure grows past safety levels.

(b) The statistics for y_5 and y_7 independently detect the attack.

Figure 4: $\tilde{y_5}(t) = y_5(t-10)$ & $\tilde{y_7} = y_7 * 1.2$

8. STEALTH ATTACKS

Although the proposed ADM can detect a wide range of attacks, we consider a more powerful adversary that knows about the detection scheme. We take a conservative approach in our models by assuming a very powerful attacker with knowledge of: (1) the exact linear model that we use, the parameters of the ADM, and (3) the control command signals. Such a powerful attacker may be unrealistic in some scenarios, but one may want to test the resiliency of our system to such an attacker to guarantee safety for a wide range of attack scenarios. The goal of a stealth attacker is to raise the pressure in the tank without being detected. We define and analyze three such attacks in our work.

9. REFERENCES

[1] N. Ricker. Model Predictive Control of a Continuous, Nonlinear, Two-Phase Reactor. *Journal of Process Control*, 3:109–109, 1993.