available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ijcip

Understanding the physical and economic consequences of attacks on control systems

Yu-Lun Huang[c,*], *Alvaro A. Cárdenas*[a], *Saurabh Amin*[b], *Zong-Syun Lin*[c], *Hsin-Yi Tsai*[c], *Shankar Sastry*[a]

[a] *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, USA*
[b] *Department of Civil and Environmental Engineering, University of California, Berkeley, California 94720, USA*
[c] *Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, 30010, Taiwan*

ARTICLE INFO

Article history:
Received 11 June 2009
Accepted 11 June 2009

Keywords:
Control systems
Integrity attacks
Denial-of-service attacks
Consequences

ABSTRACT

This paper describes an approach for developing threat models for attacks on control systems. These models are useful for analyzing the actions taken by an attacker who gains access to control system assets and for evaluating the effects of the attacker's actions on the physical process being controlled. The paper proposes models for integrity attacks and denial-of-service (DoS) attacks, and evaluates the physical and economic consequences of the attacks on a chemical reactor system. The analysis reveals two important points. First, a DoS attack does not have a significant effect when the reactor is in the steady state; however, combining the DoS attack with a relatively innocuous integrity attack rapidly causes the reactor to move to an unsafe state. Second, an attack that seeks to increase the operational cost of the chemical reactor involves a radically different strategy than an attack on plant safety (i.e., one that seeks to shut down the reactor or cause an explosion).

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Control systems are computer-based systems used to monitor and control physical processes. They are usually composed of a set of networked devices such as sensors, actuators, controllers, and communication devices.

Control systems and networks are essential to monitoring and controlling many critical infrastructure assets (e.g., electric power distribution, water treatment, and transportation management) and industrial plants (e.g., those used for manufacturing chemicals, pharmaceuticals, and food products). Most of these infrastructures are safety-critical – an attack can impact public health, the environment, the economy, and even lead to the loss of human life.

Control systems are becoming more complex and interdependent and, therefore, more vulnerable. The increased risk of computer attacks has led to numerous investigations of control system security (see, e.g., [1–11]). Most of the technical solutions involve extensions and improvements to traditional information technology (IT) mechanisms. However, very few solutions consider the interactions between security and the physical processes being controlled. In particular, researchers have not considered how attacks affect the estimation and control algorithms that regulate physical systems, and, ultimately, how the attacks affect the physical environment.

The goal of this paper is to initiate the development of new threat models for control systems. We argue that a threat assessment must include an analysis of how attacks on control systems can affect the physical environment in order to: (i) understand the consequences of attacks, (ii) estimate the possible losses, (iii) estimate the response time required by defenders, and (iv) identify the most cost-effective defenses.

* *Corresponding author.* Tel.: +886 3 5131476.
E-mail address: ylhuang@cn.nctu.edu.tw (Y.-L. Huang).

1874-5482/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijcip.2009.06.001

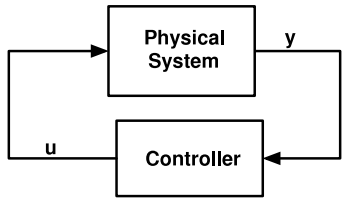

Fig. 1 – Control system abstraction.

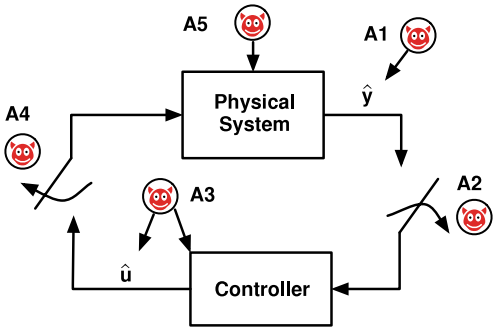

Fig. 2 – Attacks on control systems.

The paper is organized as follows. The next section, Section 2, focuses on formal models of cyber attacks in control systems. Section 3 describes the experimental setup and analyzes the experimental results. The final section, Section 4, summarizes our conclusions and highlights areas for future research.

2. Modeling Attacks

This section defines the control system abstraction and formally models integrity and denial of service (DoS) attacks.

2.1. *Notation*

A control system is composed of sensors, controllers, actuators, and the physical system (plant). Sensors monitor the physical system and send measurements to a controller. The controller sends control signals to actuators. Upon receiving a control signal, an actuator performs a physical action (e.g., opening a valve). Fig. 1 clarifies the relationships between the physical system, sensor signals (y), the controller, and control signals (u).

The following notation is used to formally model attacks on control systems.

- Time (t): The term t denotes an instant of time. A process runs from $t = 0$ to $t = T$.
- Sensor Measurement ($y_i(t)$): The term $y_i(t)$ denotes the value measured by sensor i at time t. Note that, $\forall i, t$, $y_i(t) \in \mathcal{Y}$, where $\mathcal{Y} = [y_i^{min}, y_i^{max}]$ (y_i^{min} and y_i^{max}) are the reasonable minimum and maximum values representing the plant state, respectively. Also, $Y = [y_1, y_2, \ldots, y_n]^T$, where n is the number of sensors.
- Manipulated Variable ($u_i(t)$): The term $u_i(t)$ denotes the output of controller i at time t. Note that, $\forall i, t$, $u_i(t) \in \mathcal{U}_i$, where $\mathcal{U}_i = [u_i^{min}, u_i^{max}]$ is the allowable range of controller output values.
- Attack Duration ($\mathcal{T}_a$): The term $\mathcal{T}_a$ denotes the duration of an attack. An attack starts at $t = t_s$ and ends at $t = t_e$. Note that $\mathcal{T}_a = [t_s, t_e]$.

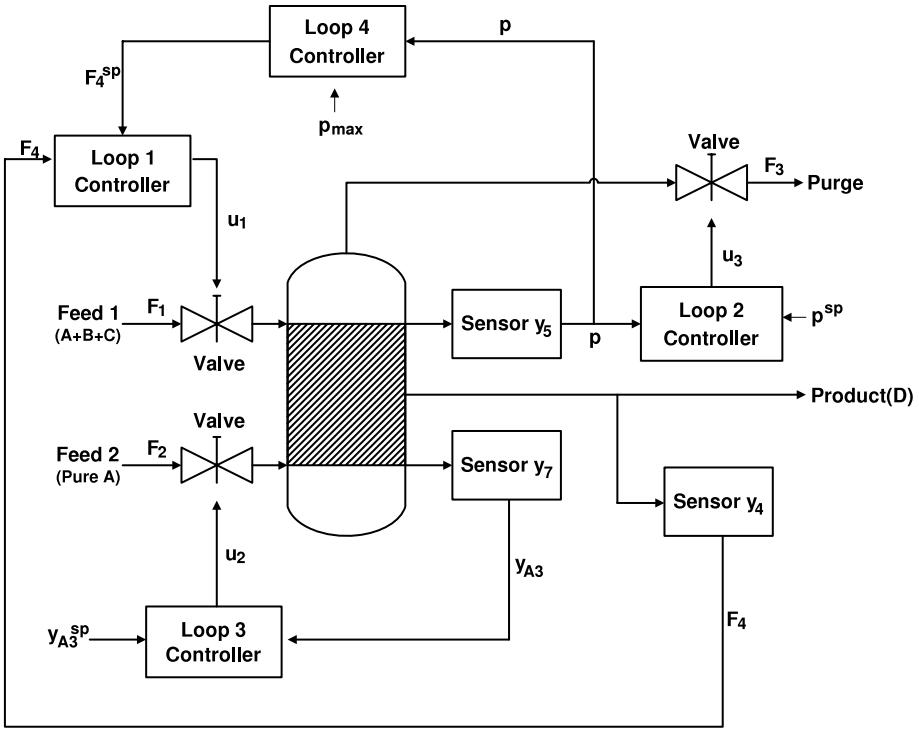

Fig. 3 – Chemical plant.

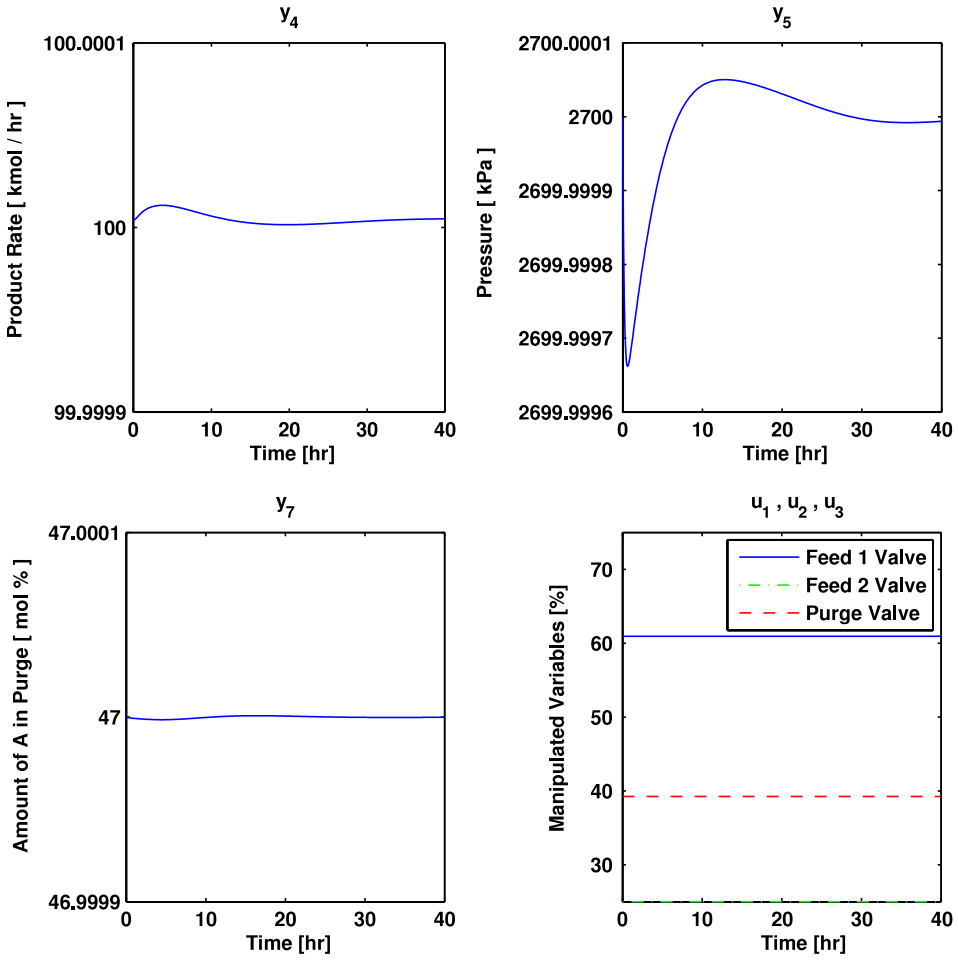

Fig. 4 – Plant outputs without noise.

Fig. 2 identifies several attacks on control systems. A1 and A3 correspond to *integrity attacks*, where the adversary sends false information $\hat{y} \neq y$ or $\hat{u} \neq u$ from (one or more) sensors or controllers. The false information may be an incorrect measurement, an incorrect time when the measurement was observed, or an incorrect sender identifier. The adversary can launch these attacks by obtaining the secret keys used by the devices or by compromising sensors (A1) or controllers (A3). We assume that each device is uniquely authenticated. Therefore, an attacker who compromises the secret key of a device is able to impersonate only that device.

A2 and A4 correspond to *DoS attacks*, where the adversary prevents the controller from receiving sensor measurements or prevents actuators from receiving control commands. The adversary can launch a DoS attack by jamming communication channels, compromising devices and preventing them from sending data, attacking routing protocols, or flooding the network.

A5 corresponds to a *direct attack* against actuators or an *external physical attack* on the plant. From an algorithmic perspective, it is not possible to defend against such attacks (aside from detecting them). Therefore, significant efforts must be implemented to deter and/or prevent attacks against physical systems (e.g., by implementing physical security controls).

2.2. *Modeling integrity attacks*

A successful integrity attack on sensor i modifies the real sensor signal, causing the input to the control function u to be changed from y to $\hat{y}$. In an integrity attack, the adversary sends a value $\hat{y}$ or $\hat{u}$ to a sensor or actuator based on the information available to the adversary.

In an effort to develop a systematic – and trackable – treatment of attack strategies, we propose the investigation of *max attacks*, *min attacks*, *scaling attacks*, and *additive attacks*. We assume that all these attacks lie within $\mathcal{U}_i$ and $\mathcal{Y}_i$. Note that signals outside this range are easily detected by fault-tolerant algorithms.

The following attacks can be launched against sensors:

- Min and Max Attacks:

$$\hat{y}_i^{\min}(t) = \begin{cases} y_i(t) & \text{for } t \notin \mathcal{T}_a \\ y_i^{\min} & \text{for } t \in \mathcal{T}_a, \end{cases}$$

and

$$\hat{y}_i^{\max}(t) = \begin{cases} y_i(t) & \text{for } t \notin \mathcal{T}_a \\ y_i^{\max} & \text{for } t \in \mathcal{T}_a. \end{cases}$$

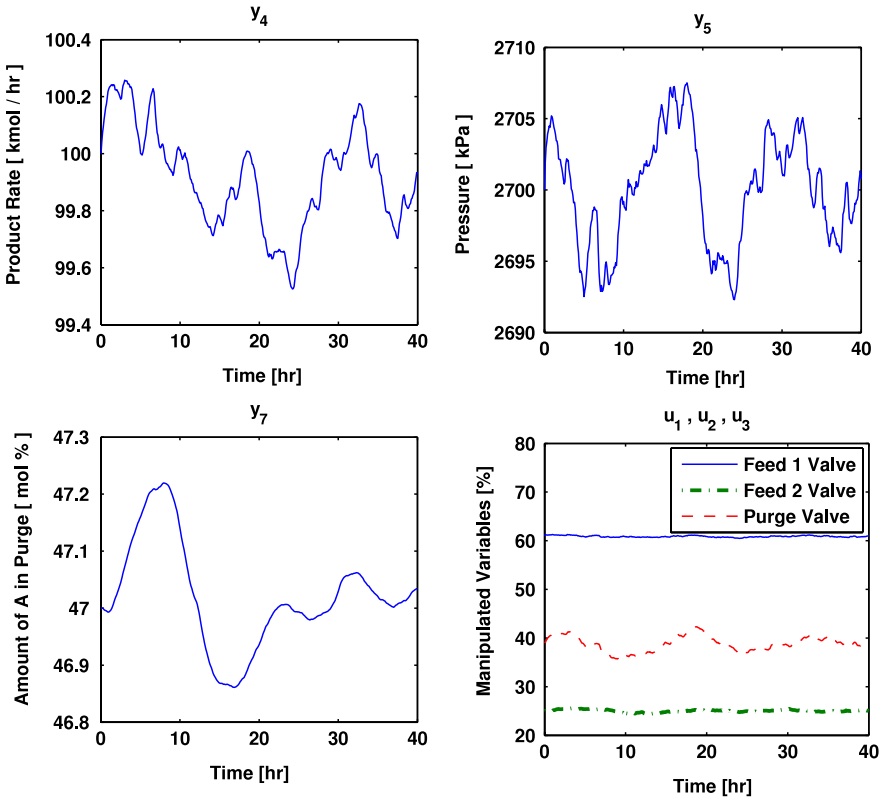

Fig. 5 – Plant outputs with Gaussian noise.

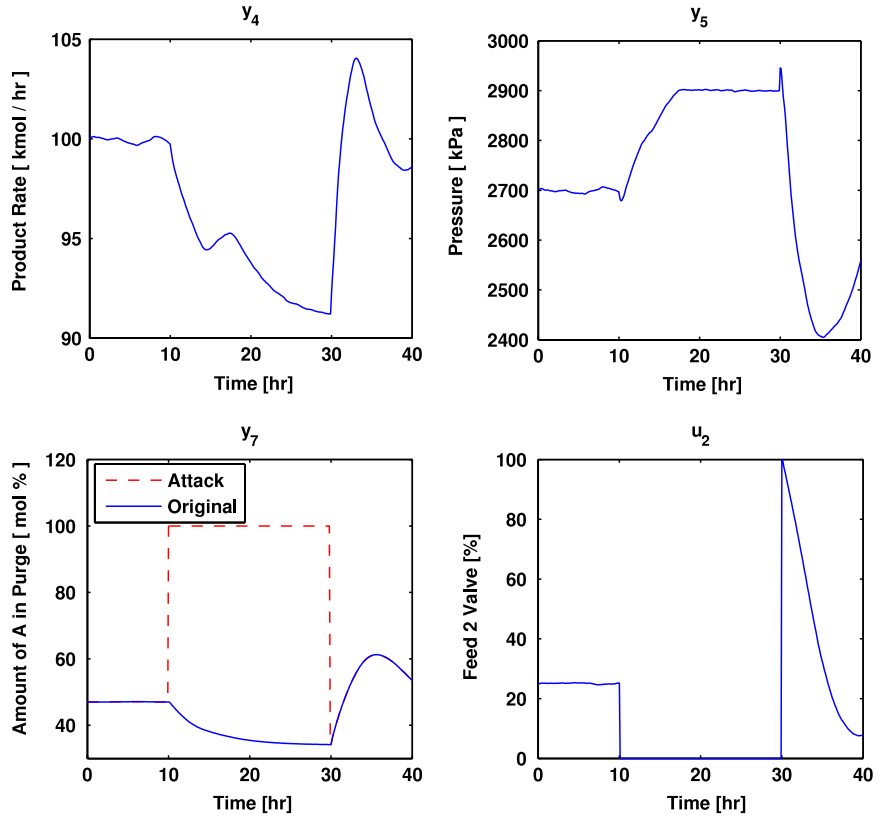

Fig. 6 – Integrity attack y_7^{max} from $t = 0$ to $t = 30$.

Fig. 7 – Integrity attack $y_5^{\min}$ from $t = 0$ to $t = 30$.

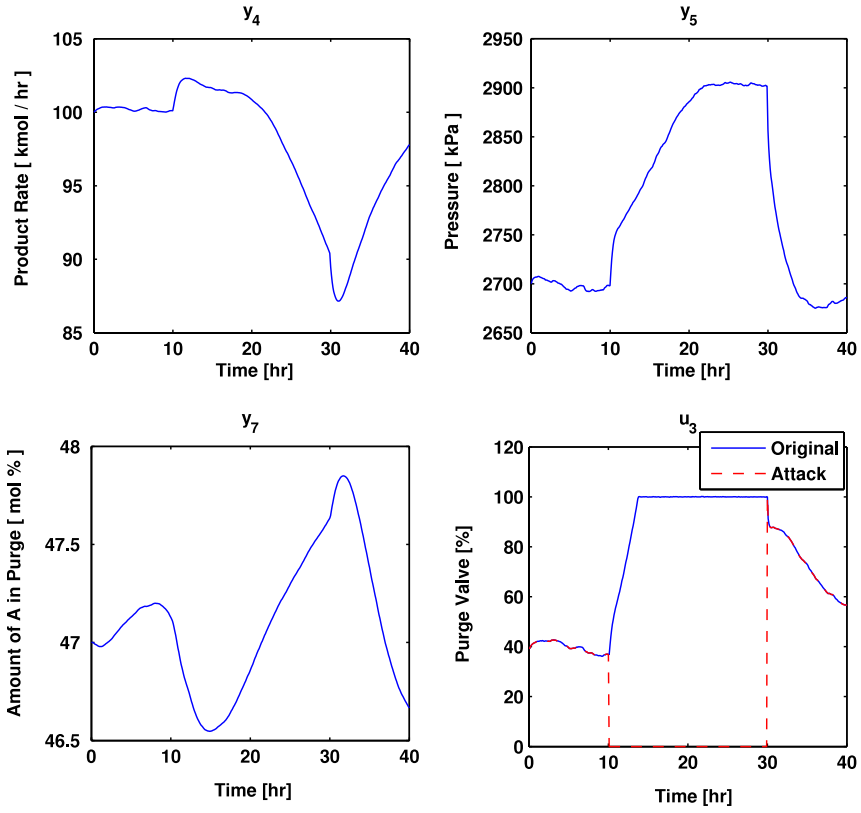

Fig. 8 – Integrity attack $u_3^{\min}$ from $t = 0$ to $t = 30$.