6-14-2019 11:00 AM

Analysis, Design and Demonstration of Control Systems Against Insider Attacks in Cyber-Physical Systems

Xirong Ning, *The University of Western Ontario*

Supervisor: Jiang, Jin, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Electrical and Computer Engineering

Abstract

This dissertation aims to address the security issues of insider cyber-physical attacks and provide a defense-in-depth attack-resilient control system approach for cyber-physical systems.

Firstly, security analysis for cyber-physical systems is investigated to identify potential risks and potential security enhancements. Vulnerabilities of the system and existing security solutions, including attack prevention, attack detection and attack mitigation strategies are analyzed.

Subsequently, a methodology to analyze and mathematically characterize insider attacks is developed. An attack pattern is introduced to represent key features in an insider cyber-physical attack, which includes attack goals, resources, constraints, modes, as well as probable attack paths. Patterns for such attacks are analyzed for different attack stages. Impacts and consequences of these attacks are analyzed by using an attack tree. Stealthy conditions of insider attacks are identified through temporal and spatial analysis, respectively.

On the defense side, a cross-layered detection scheme is developed to reveal stealthy insider attacks, and an attack-resilient control scheme is proposed to mitigate impacts of these attacks. The detection scheme includes a hierarchical approach by incorporating different detection methods in multiple layers to provide a defense-in-depth detection against the attacks. A model-based anomaly detection method is used to uncover the anomalies caused by temporal stealthy attacks, while a data-driven clustering detection method is used to recognized anomalies induced by spatial stealthy attacks. The attack-resilient control scheme consists of a decision logic and multiple attack-resilient controllers. The decision logic responds to the anomalies identified by the detection scheme and subsequently switches to suitable controllers. These controllers are designed to respond to these attacks and mitigate or minimize their impacts.

To validate the above methodologies, a general guideline for designing an experimental security assessment platform has been developed in this dissertation. Furthermore, a modular approach is proposed to design and implement a platform to simulate various insider attacks

and to evaluate corresponding defense mechanisms on a cyber-physical system. The designed platform has been implemented on a physical component based dynamic system simulator, known as Nuclear Process Control Test Facility (NPCTF). The proposed vulnerability assessment and security enhancement techniques have been validated under different insider attacker scenarios.

Keywords

Cyber-physical system security; vulnerability analysis; insider threat; cross-layer detection; attack-resilient control; and security assessment platform

Acknowledgments

There are lots of people who have helped, supported, and inspired me for the past six years, to make this dissertation possible.

First and foremost, I would like to express the sincere gratitude to my supervisor Dr. Jing Jiang for offering the opportunity to work in this group. I am grateful to his continuous guidance, support and patience throughout the whole study and research.

Second, I am very grateful to Dr. Xinhong Huang for providing lots of help, advise and friendship in many ways. Her management skills and kindness have helped me steer the journey in the right direction.

Third, I'd like to thank all CIE group members and all my friends for their help and encouragement. I have never felt alone because of their friendship. They have made this journey colorful. A special thanks for my close friend Dr. Xiaoyan Wang who has given me lots of care and help in the past years.

Last and most importantly, I would like to express the greatest gratitude to my family for their everlasting love, share, respect and understanding throughout my life. My beloved family have been giving me countless courage, hope and happiness to motivate me to explore and adventure the unknown wonderland. Special thanks for my parents Qifan Ning and Fenggai Wang who are always there for me with forever love and support.

Thanks for all the love, support and help, which have motivated me to complete this dissertation.

Table of Contents

List of Tables

List of Figures

Abbreviations

AA	Actuator Attack
AI	Analog Input
AO	Analog Output
AP	Attack Pattern
AT	Attack Tree
CIP	Common Industrial Protocol
CPS	Cyber-Physical System
CUSUM	CUmulative SUM
DI	Digital Input
DO	Digital Output
DoS	Denial of Service
ECCS	Emergency Core Cooling System
FDI	False-Data Injection
HMI	Human-Machine Interface
IDS	Intrusion Detection System
I/O	Input/ Output
IP	Internet Protocol
KF	Kalman Filter
NPCTF	Nuclear Process Control Test Facility
OPC	OLE for Process Control
P	Proportional
PD	Proportional-Derivative
PID	Proportional-Integral-Derivative
PLC	Programmable Logic Controller
SA	Sensor Attack

SE	State Estimate
TCP/IP	Transmission control protocol/ Internet Protocol
UDP	User Datagram Protocol

Nomenclature

A	System state transition matrix
B	System input matrix
C	System output matrix
C_i	Centroid of a cluster i
L	Steady-state Kalman filter gain
N	Total number of observations
Q	Estimate error covariance
R	Measurement noise covariance
R_i	Radius of a cluster i
$x(k)$	System state
$\hat{x}(k)$	State estimate
$y(k)$	System measured output
$\tilde{y}(k)$	System observed output
$\hat{y}(k)$	Output prediction
$u(k)$	System input
$\tilde{u}(k)$	Received input in actuators
$r(k)$	Measurement residual
$a(k)$	Attack input
$I(k)$	Measurement data set
$P(k)$	The covariance of state vector estimate
$P^{-}(k)$	The error covariance ahead in Kalman filter prediction
$S(k)$	Safety set
Γ^{u}	The binary incidence matrix to control channels
Γ^{y}	The binary incidence matrix to measurement channels
$\|x\|_p$	p-norm of a vector x
b	Residual bias value
d	Cluster parameter

ε	Predefined threshold for a safety set
τ	Detection threshold

Chapter 1

1. Introduction

1.1 Background

Cyber-physical systems (CPSs) can be essentially viewed as a physical process and its corresponding control systems connected through some form of common communication networks [1], as is shown in Figure 1.1. Data between the physical process and the control system are transmitted through communication networks for monitoring and control purposes [2]. Because the networks can also be used by potential adversaries, it opens up potential entry points for them to tamper with the transmitted data. Adversaries might even gain access to safety-critical system information by exploiting weaknesses of networks or communication protocols. Due to cyber-physical interactions, malicious adversaries might manipulate the transmitted data to disrupt the physical process through cyber means, which is referred herein as cyber-physical attacks [3]. If these attacks are targeted to safety-critical processes, they can cause immense damage in the physical parts of the system and even endanger human lives.

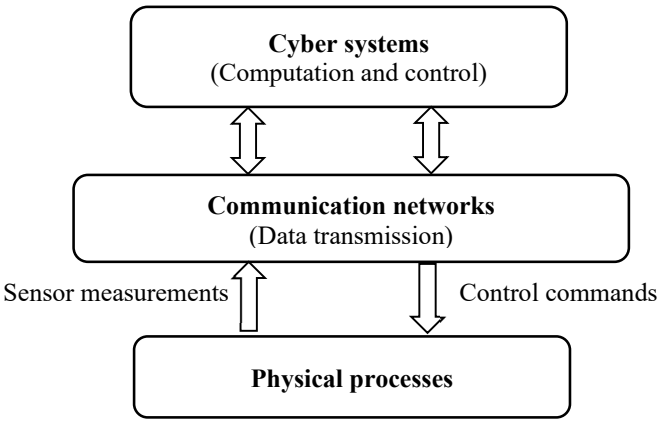


Figure 1.1 Diagram of a cyber-physical system

Cyber-physical attacks can come from either an insider threat or an external threat. The insider threat is the most daunting challenge to handle [4], it is because that insider attackers usually have legitimated access to the targeted resources and may even know how to carry out

destructive actions while avoiding being detected [5]. A well-publicized insider attack is on the Maroochy Shire Council's sewage control system in Austria [3]. There are also many cyber-attacks targeted on safety-critical systems that take advantage of the insider knowledge and cause adverse effects on physical processes. Stuxnet on Siemens PLC systems by introducing a malware capable of modifying internal commands [6]. In 2013, Havex attack was meticulously prepared to remotely compromise the industrial control systems and caused massive damages in safety-critical infrastructures [7]. Black energy attacked Ukrainian power grids in 2015 by seizing control of SCADA systems to deliberately switch off substations to cause wide-area blackouts [8].

There are two unique features commonly in these attacks:

1) **All these attacks are stealthy.** In all the cases, attackers are able to gain access into the system and leverage their inside knowledge about the system to bypass the anomaly detection schemes and achieve their attack goals without triggering any alarms [9]. Even though there are many security measures implemented to fence off external attackers in the cyber-physical systems, such as intrusion detection, data encryption and access control mechanisms, they may be ineffective to insider attacks.

   Therefore, it is of great significance to find security solutions that can extract the features of insider attacks, identify system vulnerabilities related to insider attacks, and manage the security risks respect to insider attacks.

2) **All these attacks are enabled due to cyber-physical interaction.** Attackers have taken advantage of the cyber-physical interactions inside the system. They have compromised the network and tampered the transmitted data in the cyber layer, then used the cyber-physical interdependencies to manipulate the process operation and caused severe physical damage without being detected.

Hence, the security of cyber-physical systems requires analysis of both cyber layer and physical process, and their interactions [10]. Methods that integrate cyber-physical security

and control theory are needed to provide attack detection and resilient control against insider cyber-physical attacks, which is the focus of this dissertation.

1.2 Motivations

Based on the analysis of the reported attack accidents, it is necessary to address the security challenge of insider cyber-physical attacks against cyber-physical systems, where the attacker can (1) tamper sensor readings or (2) manipulate control commands [11]. This kind of attacks are referred as insider attacks in this dissertation.

Even though there are many researches on modeling and analysis methods of insider attacks, it is still challenging to describe features of attacks mathematically because attacks usually happen in unpredicted ways. Therefore, instead of identifying a specific model of attacks, it is necessary to analyze the resulting impacts of insider attacks on the CPS and develop corresponding countermeasures.

A major distinction of cyber-physical security with respect to cyber security is the cyber-physical interaction of the control system with the physical processes. Cyber-physical attacks originate from cyber space but have impacts on the physical processes.

Traditionally, security issues of cyber-physical systems are mainly investigated from the perspective of information security with a focus on confidentiality, availability, integrity of the information in the cyber space [1]. While information security studies are key elements in the cyber space, they have less consideration on the interdependencies between the physical process and the cyber space. Moreover, such information security methods are not effective against insider attacks and they also fail against attacks targeting directly to the physical system dynamics. Thus, information security methods are not enough to secure cyber-physical systems. It is required to handle the cyber-physical coupling relationships and interactions. Therefore, one needs to consider cyber-physical interdependencies from a control system perspective, to enhance the security of cyber-physical systems.

For safety-critical CPS, the ultimate objective is to secure process and control mechanisms [12]. It is motivated to develop an attack-resilient control approach that can provide attack detection, protection and control for both cyber and physical aspects of the system. In order to analyze the impacts of cyber-physical attacks, and to validate the implemented security enhancement strategies, there is also a need of a security assessment platform to conduct experimental evaluation.

1.3 Research scope

This dissertation considers the security problem of cyber-physical systems against insider attacks from a control perspective. The focus of this work is on insider cyber-physical attacks, whereby the attacker is assumed to be able to interrupt the communications during the data transmission and tamper with the data of sensor measurements or control commands.

The scope and assumptions of the research are listed as follows.

- The inside attacker considered has access to the system or already inside the system, can compromise communication networks, tamper with the exchanged sensor measurements or control commands being sent and received, and cause damages to the physical process.
- Attacks are considered being stealthy for the anomaly detection scheme that only detects if the transmitted data meets the physical laws or relationships which will not trigger an alarm.
- The attack goal is to initialize an attack in the cyber space to cause impacts and damage on the physical process.
- Attacks are assumed to happen in a single channel at a time, coordinated attacks are not considered in this dissertation.
- The supervisory station is isolated from the rest of the system and is assumed to be secure. It contains a control system and an anomaly detection scheme.

- It is assumed that the supervisory layer is secure and could not be penetrated by the attacker.

The focus of this dissertation is on providing security analysis on system vulnerabilities and threats with respect to insider attacks, and designing security enhancement methods to prevent, detect and mitigate the impacts of such attacks. This research consists of the following three core tasks, as is shown in Figure 1.2.

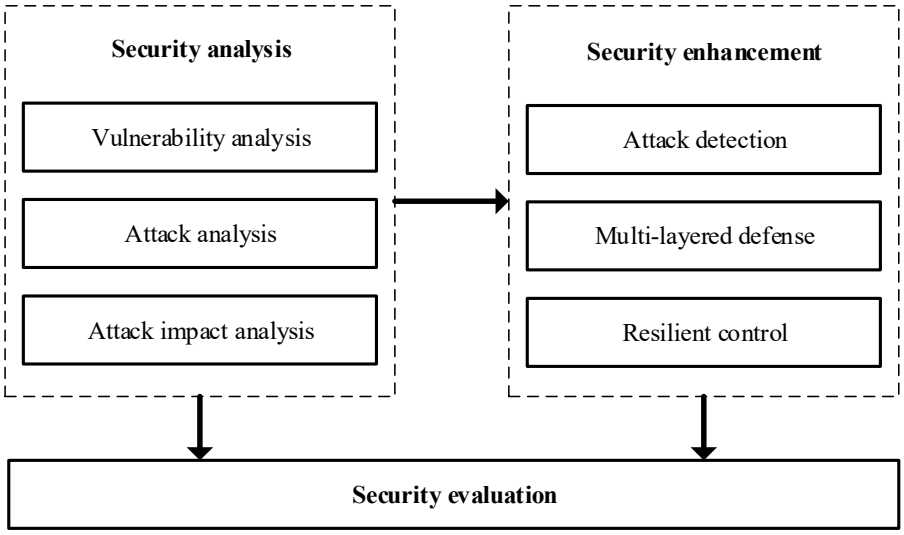


Figure 1.2 Research focus of the dissertation

Please note, terminologies used in the dissertation have been defined based on the industry standards ISA/IEC-62443: *Security for Industrial Automation and Control Systems: Technical Security Requirements for IACS Components* [13], and related technical references.

1.4 Contributions of the dissertation

Based on the research tasks in Figure 1.2, the contributions of the dissertation can be summarized into three main groups: (1) security analysis, (2) security enhancement, and (3) security evaluation.

(1) Security analysis

The contributions of this aspect are threefold.

First, potential ways of breach of security in cyber-physical systems have been investigated and analyzed. Second, a unified formulation against insider attacks has been proposed. Features of insider attacks are extracted using an attack pattern. Lastly, stealthy conditions of insider attacks are identified based on a temporal and spatial analysis. Different attack scenarios and their impacts are represented through an attack tree.

This analysis links attack threats with system vulnerabilities. The outcome of the analysis can then be used to improve the security of CPSs against potential insider attacks. Moreover, the dissertation has improved the existing work on attack pattern and stealthiness analysis against insider attacks.

(2) Security enhancement

The contributions on this topic can be highlighted in the following three aspects.

First, an online cross-layered detection scheme has been designed to reveal potential anomalies in multiple layers. The detection scheme takes a hierarchical approach by combining different detection methods in respective layers to provide a defense-in-depth detection against attacks of different forms. A state estimation with CUSUM based detection method and a data-driven detection method are proposed to work together to detect stealthy attacks. The cross-layered detection scheme is proved to be effective, as shown by examples how attack-inflicted anomalies can be detected before the attack can cause significant impacts on the wellbeing of the physical process.

The above cross-layered design has made notable improvements to the existing detection techniques that merely focus on network intrusion detection or anomaly detection in physical processes. The current design fuses data from both the cyber layer and the physical layer, integrates them with model-based and data-driven methods to provide a stronger and more robust defense-in-depth detection.

Second, an attack defensive framework has been developed in this work. This framework, combining attack prevention, anomaly detection and mitigation strategies, offers a defense-in-depth protection against insider attacks to maintain the CPS in a safe state. By using the proposed framework, system security has been enhanced as attack anomalies are detected quickly, and the system operator can be alerted promptly to take actions and to mitigate impacts of the attacks.

Third, this dissertation introduces an attack-resilient control scheme to mitigate effects of attacks, which includes an attack response scheme, a decision-making scheme and a set of switchable controllers. The attack response scheme can isolate and replace the corrupted data, the decision-making scheme can switch in appropriate controller into the system, and the controller can mitigate the attack and bring the system to a safe state. This work provides a temporary solution to protect the system before more permanent solutions can be taken by human operators to secure the system.

(3) Security evaluation

The contributions in this topic have two parts.

First, a general design methodology for developing a security assessment platform has been developed, which provides an overview on how to develop a security platform on a cyber-physical system. Modular design makes the development and implementation flexible.

To the best of our knowledge, there are no such platforms reported in the open literature.

Second, security experimentation and associated performance evaluation techniques on a specific cyber-physical system have been carried out. Experimental case studies have demonstrated that the platform is capable of identifying system vulnerabilities, validating various detection and mitigation strategies, and evaluating system security conditions and providing insights for security enhancement.

1.5 Structure and organization of the dissertation

The overall framework of the dissertation is outlined in Figure 1.3.

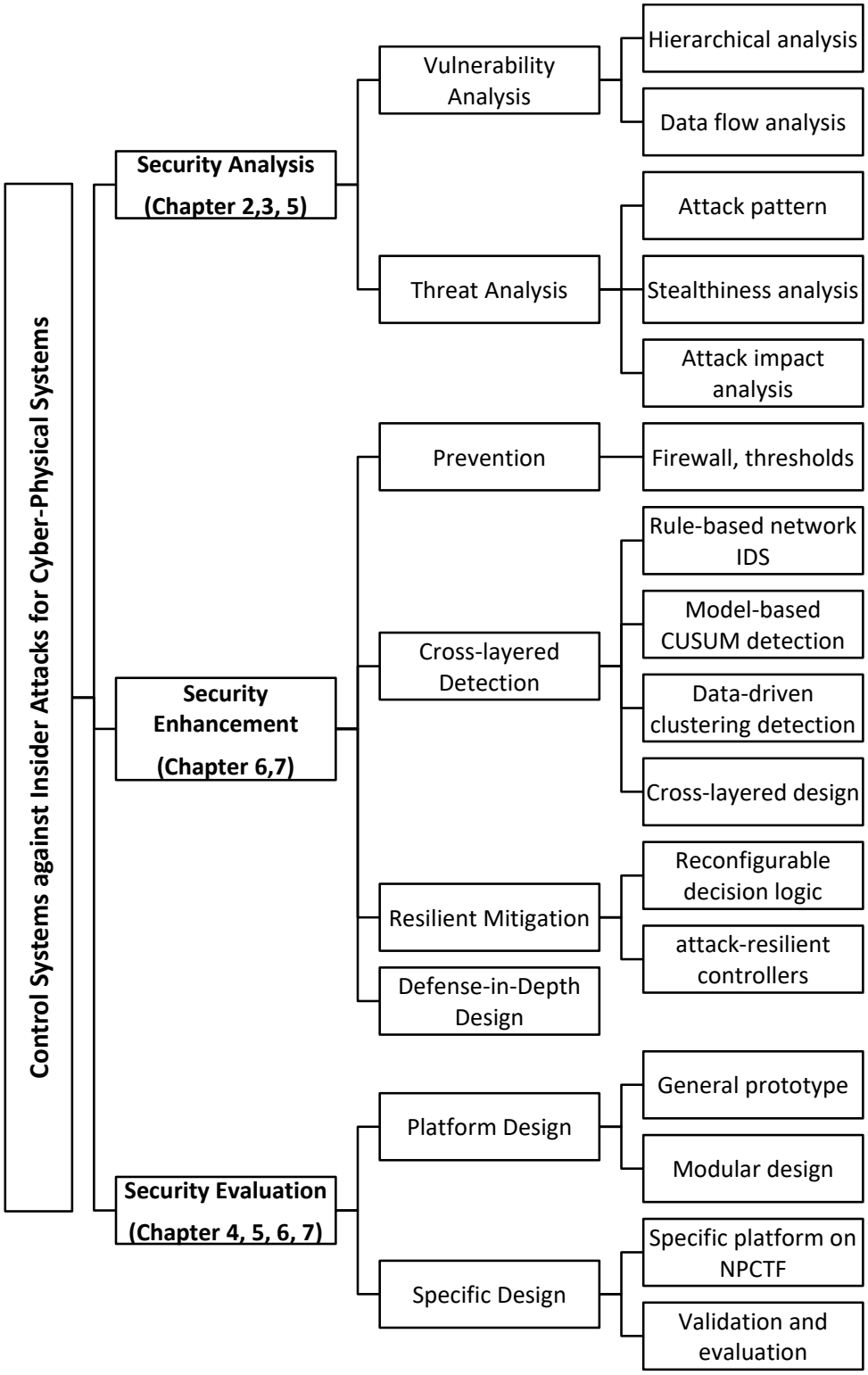


Figure 1.3 Overview of the topics covered in the dissertation

The dissertation is structured as follows:

Chapter 2 investigates the cyber-physical security issues related to insider attacks based on existing research work and literature. Risk assessment methods including vulnerability analysis, threat assessment and impact analysis are investigated and analyzed. Security enhancement strategies including topics on attack prevention, attack detection, and attack mitigation are surveyed and discussed. Meanwhile, security issues and challenges are also analyzed.

Chapter 3 analyzes system vulnerabilities and potential insider attacks on the system. Cyber-physical interactions and attack impacts are examined in the form of an attack tree. A specific analysis is demonstrated on a nuclear process control test facility (NPCTF).

Chapter 4 presents a modular design of an experimental security platform. It develops a general design guideline for a cross-layered experimental security platform, and proposes a modular approach to design and implement a platform for security tests on cyber-physical systems. This chapter also describes the process of constructing a security platform prototype for a specific cyber-physical environment and the way to use it for various security assessments.

Chapter 5 introduces a methodology on analysis and formulation of insider attacks through data tampering. Attack features are characterized by an attack pattern, stealthy conditions are analyzed, and impacts are also discussed.

Chapter 6 provides a cross-layered detection approach to detect anomalies from different layers. It integrates network intrusion detection with physical process detection, and combines model-based and data-driven detection algorithms to reveal various stealthy attacks.

Chapter 7 presents an attack-resilient control system design, which includes a decision-making scheme to respond to the attacks resiliently, and an attack-resilient controller to mitigate the impact of attacks. This chapter also presents an attack defensive framework to provide defense-in-depth protection for cyber-physical systems.

Chapter 8 concludes the dissertation and provides some discussions on future work in this area.

Chapter 2

2 Investigation on Security of Cyber-Physical Systems under Insider Attacks

2.1 Introduction

Security is critically important to ensure a reliable operation of cyber-physical systems. The purpose of this chapter is to investigate techniques for security analysis and enhancement solutions, which can provide some references and guidance as how to design defensive strategies.

Topics covered in this chapter are summarized in Figure 2.1. These topics can be classified into three categories: vulnerability analysis, threat assessment, and security enhancement strategies. System vulnerabilities and features of insider threats are analyzed and surveyed. Existing security solutions, including attack prevention, attack detection and attack mitigation strategies are also investigated.

The remainder of this chapter is organized as follows. In Section 2.2, system vulnerabilities are surveyed. In Section 2.3, features of insider threats are discussed, attack models are investigated and methods to evaluate impacts of attacks are also summarized. In Section 2.4, a variety of security solutions including attack prevention, detection and mitigation techniques are surveyed and compared, secure architectures are discussed to develop a defense-in-depth control system. Finally, Section 2.5 presents some discussion and potential solutions on security of CPSs.

2.2 Survey on vulnerability analysis related to insider attacks

Cyber-physical systems are featured as tight coupling of cyber-physical components. This cyber-physical interaction has induced security vulnerabilities that might be exploited by attackers. Different approaches to identify potential vulnerabilities related to insider attacks have been studied recently.

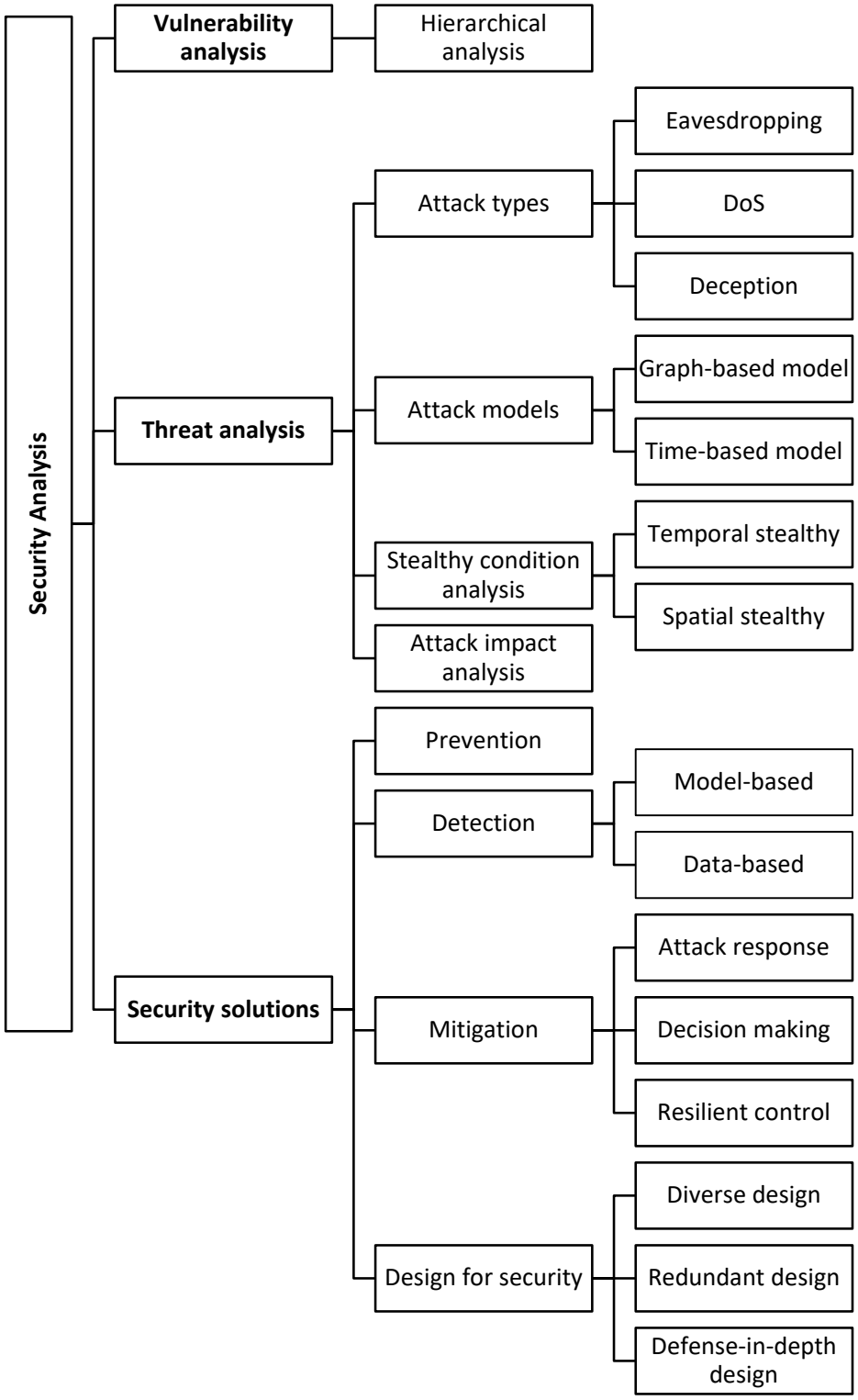


Figure 2.1 Techniques investigated on security analysis and enhancement

A method based on fault tree analysis is used to identify process vulnerabilities to insider attacks in [4]. A graph-based model is proposed to determine inherent network vulnerabilities that could be exploited by a malicious insider in [14]. Several behavior-based models are proposed to establish the relation between the vulnerabilities and insider attacks in [4] and [15]. An insider threat model in [16] is established to acquire cyber situational awareness. Since a cyber-physical attack is initiated from the cyber domain, and then manifested to the physical domain, a successful insider attacker will have to combine knowledge from both domains to explore the vulnerabilities to inflict physical damage to the process. It is necessary to analyze the vulnerabilities among the cyber-physical couplings and interactions. Unfortunately, the interactions and dependencies between the cyber and physical components have not been considered in these techniques.

2.3 Survey on threat assessment

2.3.1 Cyber-physical attacks and their characteristics

Insider attacks on cyber-physical systems can be classified by attack types, entry points and stealthy conditions, as summarized in Figure 2.2.

Adversaries may interrupt the communication networks, tamper with the data packets being sent to the controller or eavesdrop to gain information on the system state [17]. Thus, the type of attacks can correspondingly be categorized into three groups: eavesdropping, denial-of-Service (DoS) attacks, and deception attacks [1].

Eavesdropping attacks aim to intercept the network traffic and capture relevant information from the network traffic for later analysis, however, this kind of attacks will not have an impact on the physical process. DoS attacks aim to disrupt the data transmission by interrupting the communication networks. The deception attacks can compromise the integrity of data packets by tampering with the transmitted data between the physical layer and the cyber layer. Deception attacks can further be classified as false-data injection attacks, replay attacks and covert attacks. The characteristics of these attacks are described in Table 2.1.

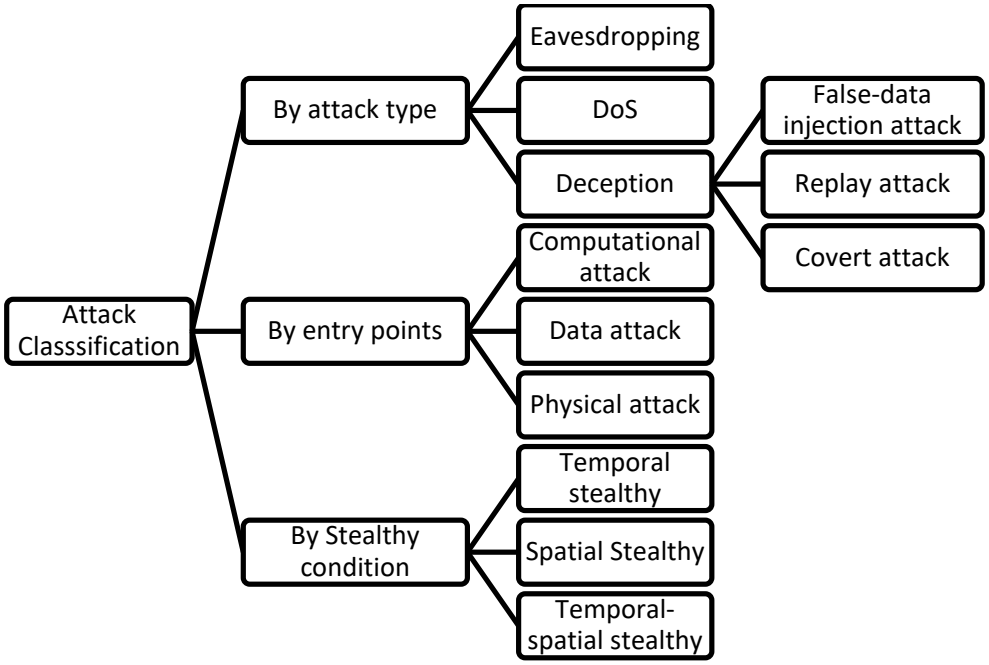


Figure 2.2 Classification of insider attacks on cyber-physical systems

A malicious attack could be carefully designed to compromise transmitted sensor data and tamper with the data that is sent to the detection scheme without being detected. A general methodology for synthesizing stealthy attacks is presented in [18]. Impacts of false-data injection attacks are characterized as a reachable set in [19], this reachable set are compared with the safety set of the system to evaluate the attack impacts. Detectability and identifiability of attacks are characterized analytically using an extended observer in [20]. To get the maximum attack impact, an DoS attack schedule is proposed to bypass intrusion detection mechanisms in [23]. In [24], a maximum DoS attack schedule is designed to deteriorate system performance indexes and maximize the trace of expected average estimation error. In [25], a maximum false-data injection attack is studied to determine the minimum number of the manipulated variables so as to cause the maximal damage.

To study the stealthiness of an attack, a graph-theoretic approach is constructed to characterize the smallest set of the attacked variables to make the network states unobservable to control

center in [21]. Two data-driven attack strategies based on the subspace of the estimated system states are presented to construct the unobservable attack in [22].

Table 2.1 Insider attacks on cyber-physical systems

Attack type		**Description**	**Reference**
Eavesdropping		Compromise the system and eavesdrop the transmitted data	[26]
Denial of Service (DoS)		Jam the networks traffic to make the communication channels unavailable	[23] [24]
Deception attack	General deception attack	Interrupt the data transmission and inject a malicious action	[27, 28] [26]
	False-data injection attack	Modify the transmitted data in a stealthy way	[19] [18] [29] [30] [20] [31] [25]
	Replay attack	Use historical data to hide the current malicious action	[32] [33-35]
	Covert attack	Coordinate control signals and sensor measurements to hide the attack action	[36]

2.3.2 Modeling methods for insider attacks

Attack models are used to map the insider threats to cyber-physical system vulnerabilities. In Table 2.2, attacks are analyzed based on control-theoretic, cyber security, and hybrid approaches separately.

Several works in this field have focused on identifying models to characterize an insider attacker based on his/her psychological and behavioral characteristics. For example, an attack model is defined by attacker's knowledge, disclosure resources and disruption resources in [43]. An insider deception model based on a grounded theory method is used to identify the technical and behavioral features of insider attacks [44]. Attack vectors are identified based on the policy violations in [45]. A framework based on insider attacker-related behaviors and

symptoms is proposed to describe insider attackers based on socio-economic aspects rather than the system architecture have been discussed in [46]. A game-theoretic model is proposed to model and analyze the insider threats in [47].

Table 2.2 Impact analysis of insider attacks from different perspectives

Approaches	Proposed techniques	Reference
Control-theoretic	Attack models, stealthy condition of attacks	[32]
	Physical watermarking detection to replay attacks	[37]
	Moving target detection	[38]
	Characterization of robust estimation and control	[39]
Cyber security	Sequence-aware intrusion detection system	[40]
	Big data analytics for attacks on PLC	[41]
Hybrid approaches	Attack graph generation	[42]

Most of these modeling work have been focused on modeling attacker's behavior, there is fewer considerations which assess the insider threat in a control-theoretical manner. System vulnerabilities exploited by the insider attacks, physical impacts of the attacks, and system resources used by an attacker need to be studied in order to provide indications for a secure control system design.

Furthermore, it is of the utmost importance to study the dynamics of the physical process under attacks, and to capture features caused by such attacks. An insider pattern is defined by its type, capabilities, objective, and strategy in [48]. A model based on a semi-Markov chain is presented to predict possible decisions by attackers and to evaluate the system security in [49]. An attack space has been defined according to the system knowledge, disclosure information, and disrupted resources in [32]. Some illustrative examples have been presented to show how an attack signal is injected into a state estimator in a stealthy way in [50, 51]. These research are focused on analysis of how insiders might attack from the perspective of an attacker's behavior [4]. However, to secure the physical process, it is necessary to analyze the impact of the attacks on the physical process from a system point of view, to identify anomalies that an attack might manifest on the system. Since attacks are initiated from the cyber domain, and then manifested to the physical domain, a successful insider attacker will have to combine

knowledge from both domains to explore the vulnerabilities to inflict physical damage to the process. It is necessary to analyze the vulnerabilities among the cyber-physical couplings and interactions. Unfortunately, the interactions and dependencies between the cyber and physical components have not been considered in these techniques.

In order to capture the impacts of an insider attacker, a tuple has been used based on organization structure [52], to trace a sequence of attack actions leading up to safety violations. Attack models are represented using dataflow-based directed graphs in [53]. Similarly, attack trees [54], attack graphs [55], integrated fault-attack trees [56], and attack pattern trees [57] are all used to characterize insider attacks and their attack paths and steps. These researches help to identify system vulnerabilities under insider attacks.

2.3.3 Techniques for stealthy condition analysis

The ultimate goal of an insider attacker is to drive some critical system variables into unsafe states without triggering any alarms by keeping the attack stealthy or delaying any detection or responses. A well-planned attacker might bypass the anomaly detection system or hide his/her actions for a long period. It is possible for an attacker to create false sensor signals that will not raise an alarm, some examples are presented in [28, 58] [59] and [60]. Works on stealthy condition analysis are summarized in Table 2.3.

Table 2.3 Review of techniques for stealthy condition analysis

Type of attacks	Description of stealthy condition	Reference
Replay	Bypass the anomaly detector	[32] [33] [62]
False data injection attack	Tamper the anomaly detector	[51] [58] [65]
Zero-dynamics attack	Modify control commands to hide attacker's actions	[66]
Covert attack	Bypass traditional anomaly detectors	[96]
Surge attack, bias attack, and geometric attack	Bypass traditional anomaly detectors	[27]

So far as keeping a cyber-physical attack stealthy, there are two main approaches: the first is temporal stealth attack to tamper the anomaly detection mechanism by injecting deceptive data,

such as a deception attack [61], or a false-data injection attack [58]. The second approach is spatial stealth attack to conceal malicious attacks using healthy historical data, such as replay attacks in [62]. A methodology is presented to study stealthy attacks in [63]. Detectability and identifiability of a stealthy attack are defined in [64]. However, none of these studies have taken into account essential features of an insider attacker. As a result, many assumptions made in these works may not be directly applicable for attacks committed by an insider. Therefore, common characteristic of attacks and their impacts on cyber-physical systems need to be analyzed.

2.3.4 Attack impact analysis

Analyzing impact of attacks will provide an overall view of CPSs security status and a guideline to design mitigating methods. Research work on impact analysis is summarized in Table 2.4. A game theory method is used to analyze the cyber threats within a cyber-physical system in [67]. Impacts of attacks on critical networks are evaluated in [68] and [69] to increase the resilience of cyber-physical systems. In order to analyze how an cyber attack can affect the physical process, a threat model is proposed in [70], possible consequences of DoS attacks and deception attacks are assessed. Impacts of combination attacks are considered in [71], and an aspect-oriented method is proposed to model these impacts in [72]. To better understand the attack impacts timely, an algorithm is presented in [73] to predict the possible consequences by attacks. In order to develop the characteristics of attacks, an attack description language is proposed in [74], however, this method can only be applied to known attacks.

The above-mentioned methods mainly focus on analysis of insider attacks in the cyber domain. By combining information from both cyber and physical domains, it is more likely that a pattern from a cyber-physical attack can be revealed, and subsequent impacts can be alleviated. Therefore, impacts on physical process, as well as the interactions of the cyber system with the physical world should be considered to develop a general and systematic framework for securing cyber-physical systems.

Table 2.4 A literature review on methods for impact analysis

Methods	References
Game theory method to analyze cyber-physical attacks	[67]
Analysis of attack impacts on networks	[68]
Cause-consequence relationship analysis	[69]
Impact analysis on DoS and integrity attacks	[70]
Statistical analysis for various attack scenarios	[71]
Aspect-oriented risk assessment	[72]
Predictive risk assessment method	[73]
Qualitative and quantitative analysis for cyber-physical attacks	[74]

2.4 Investigation of security enhancement solutions

Reviews on security solutions include prevention, detection and mitigation. Attack prevention is defined as the first barrier against insider attacks starting from the entry point. Attack detection techniques need to be built for all layers of a cyber-physical system, and mitigation mechanisms are initiated to respond and mitigate the impacts of the attacks.

2.4.1 Attack prevention and detection

Attack detection is to identify anomalies of the system. Attack detection techniques can be classified into two groups: (1) passive detection techniques to prevent attacks, and (2) active detection techniques to identify the anomalies of attacks. Active detection techniques can be designed as data-based methods or model-based methods. Related techniques are investigated in Table 2.5.

Passive detection techniques mainly focus on protecting the information security include firewalls, demilitarized zones and network intrusion detections to prevent intrusions and misuse of access privileges. Guidelines are proposed in [75] to design specific firewalls and demilitarized zones to prevent the intrusions from the external network to the physical process. Intrusion detection methods are proposed in [76] and [77] to monitor the network traffics. These passive techniques can help to prevent intrusions form external or local networks.

However, they might be ineffective for sophisticated attacks and insider attacks. it is necessary to employ defense-in-depth detection strategies to provide a layered detection.

Active anomaly detection techniques can be classified into data-based and model-based techniques. Data-based approaches does not require system and attack models, they detect the anomalies through machine-learning [80] and pattern recognition techniques [78, 79] for analyzing hidden patterns in the observed training data set. Model-based approaches are based on the parametric models under normal operations and under different attack scenarios. The detection decision rules are made on the residuals between system observations and model-based system outputs, such as game theory [85], physical watermarking [90] and state estimation techniques [91-100]. However, the residuals are often not obvious due to the model uncertainties and noises, and the model might be utilized to bypass the detection schemes by sophisticated attackers. It is required to consider the cumulative effects of insider attacks and the constraints of system models when designing a detection framework.

Table 2.5 A literature review of prevention and detection methods

Attack detection	Techniques	References
Attack prevention	Firewalls and demilitarized zones	[75]
	Network intrusion detection	[76] [77]
Data-based detection	Clustering	[78, 79]
	Machine learning	[80]
	Data fusion	[81] [82] [83]
Model-based detection	Graph theoretic methods	[84]
	Game theory	[85]
	Gaussian authentication	[86] [87]
	Fast greedy algorithm	[88]
	Physical watermarking	[37, 89] [90]
	State estimation	[91-94] [95] [96] [97] [98] [99] [100]
	Rule-based detection	[51]
	Hybrid detection	[40, 42]

2.4.2 Mitigation methods

Once safety violations or anomalies are detected, corresponding mitigation actions will be triggered. The objective of attack mitigation is to minizine impacts of the attack and recover the system operation as much as possible.

There are two types of mitigation strategies: (1) proactive methods that mitigate the system prior to the detection of an attack and (2) reactive mitigation that takes actions only when an attack has been detected. This chapter investigates the related work of proactive methods, which is summarized in Table 2.6.

Table 2.6 Review on mitigation methods

Techniques	Approaches	References
Game-theory control	Dynamic zero-sum game theory and a jamming strategy	[101]
	A receding horizon Stackelberg control law	[102]
Resilient control	Attack-resilient control with a distributed control methodology	[103,104,105]
	Attack-resilient control through a time-trigger strategy	[106]
	Attack-resilient control with a Kalman state estimator	[107]
	A multiple-task robust controller	[108]
	Attack-resilient control using a hybrid model	[109]
	A two-stage attack-resilient control system	[110]
	Reconfiguration control for safety violations	[111]
Optimal control	An optimal decoder to minimize the attack effects	[112]
	Design of an optimal estimator to minimize the worst-case impact	[113]
	Horizon linear–quadratic control	[114]
Predictive control	A predictive control system to compensate for adverse effects	[115]
Networked control	Contingency analysis to detect malicious control commands	[116]

These methods include game-theory methods, resilient control method, optimal control method, predictive method, and network control method. A dynamic zero-sum equilibrium control strategy is proposed to defend DoS attacks in [101] and a receding horizon control law against replay attacks is presented in [102]. Attack-resilient control designs are studied based on various strategies, such as distributed controllers in [103, 104, 105], multi-agent time-trigger strategies in [106], and state estimation through Kalman filter in [107]. Hybrid controllers are designed in [108] and [109] to defend stealthy attacks, a two-stage resilient control system is designed to respond and mitigate attack impacts. In order to minimize the attack impacts, optimal control is considered in [112], [113] and [114]. In order to compensate the adverse effects of attacks, a predictive control system is demonstrated in [115], and a contingency analysis is given in [116].

However, most of these mitigation methods are designed based on known attacks, since the attacks are unknown and hard to predict, some of the impacts of attacks may not be acquired and mitigated effectively. A resilient defensive framework should be performed in multiple layers to secure the cyber-physical system.

2.4.3 Security architecture development

A secure architecture is also necessary to ensure the security of a cyber-physical system. Security enhancement solutions should be considered from the cyber layer to physical layer. Table 2.7 summarizes the design of security architectures from different perspectives. A data fusion-based framework is proposed in [117] in order to enhance the robustness of networks. The cyber-physical interactions of a resilient cyber-physical system architecture are discussed in [118]. A authentication architecture for a IoT system is studied to presented to enhance the end-to-end security. A cyber-physical security architecture is proposed in [119] from an information security perspective. A layered architecture is analyzed in [120] to improve the security of communication protocols.

Security in one layer may not satisfy the required security requirements, hence there should be multi-layer security solutions to secure the cyber-physical systems. A defense-in-depth

security architecture is in need to accommodate various security solutions in multi-layer systems. Different defense-in-depth designs for CPS are presented in Table 2.8.

Table 2.7 A literature overview of security architecture design

Proposed approach	References
A fusion-based defense mechanism	[117]
Qualitative and quantitative analysis for cyber-physical interactions	[118]
IoT-based security architecture	[119]
An information security framework	[120]
Security architectures to study security of heterogeneous protocols	[121]

Table 2.8 Defense-in-depth solutions for security enhancement

Design methods	Techniques	References
Single-layer solutions	A comprehensive review on IDS techniques	[122]
	IDS in cyber layer	[35]
Multi-layer solutions	Distributed management and control of security	[123]
	A framework for attack-resilient industrial control systems: Attack detection and controller reconfiguration	[124]
	A comprehensive analysis of security objectives	[125]
	A CPS security framework including multiple security mechanism	[126]
	A cross-layer context-aware security framework	[127]

2.5 Discussions

This chapter has summarized the related research works to secure control of cyber-physical systems against insider attacks. Based on the review of existing work, there are mainly two aspects need to be studied for further research and improvements: (1) there is a need for risk assessment methods to address the attack impacts on physical processes from a control point

of view, provide indicators for security enhancement strategies; and (2) attack mitigation methods may be improved when a defense-in-depth structure and multi-layer redundant design are considered.

The remaining chapters on this dissertation will focus on these two directions to enhance the security of the system.

Chapter 3

3 Vulnerability Analysis under Insider Attacks

3.1 Introduction

One of the core features of cyber-physical systems is the tight cyber-physical connectivity and interactions. Malicious adversaries might use the cyber-physical couplings to launch cyber-physical attacks on safety-critical processes and cause disruptions in operations of physical processes. To understand insider attacks on cyber-physical systems and develop a corresponding defensive framework, it is necessary to map out the relations from the insider attacks to the vulnerabilities within a cyber-physical system.

There are two questions need to be answered when assessing the system security.

(1) What assets in the cyber-physical system are vulnerable to insider attacks?

This question is related to system vulnerabilities that might be taken advantage of by inside attackers.

(2) What are the threats from insider attacks?

This question can be answered by the analysis of possible attacks, including analysis of attack models and their impacts.

To answer these two questions, this chapter analyzes system vulnerabilities under insider attacks.

3.2 Definition of an insider threat

An insider threat is defined by a unique set of attributes, which includes [128]:

- **Access:** Insiders are those who have legitimate access to the targeted system or already gain control of the system. Malicious insiders might abuse such access to the targeted

resources and avoid being detected by access control strategies that are designed mainly to prevent against external intrusions [5].

- **Authorized resources:** Insiders already have authorized resources to conduct operations for their assigned duties, which also give them accessible to the targeted resources and carry out destructive attacks.

- **Knowledge:** Insiders already have certain degree of the knowledge of the targeted system and its security countermeasures. They may even know how to exploit the system vulnerabilities and carry out their malicious actions without being detected, which makes detecting, mitigating, or recovering from insider attacks extremely challenging [13].

Security issues associated with insider attacks normally have two unique traits:

- **Cyber-physical coupling:** attacks launched from cyber space can cause physical damage in the processes.

- **Stealthy attacks**: insiders could design their attacks in such a way to avoid being detected.

3.3 Vulnerability analysis of cyber-physical systems

Vulnerabilities are the weaknesses that an adversary could exploit and use to cause damages to the systems [126]. Analysis of vulnerabilities can identify the potential entry points and understand how an attacker might take advantage of the vulnerabilities to launch malicious attacks. The vulnerabilities of cyber-physical systems can be classified into five categories [129]:

- architectural vulnerabilities;

- security policy vulnerabilities;

- software and hardware vulnerabilities;

- communication network vulnerabilities; and

· detection and control related vulnerabilities.

In this chapter, vulnerabilities of cyber-physical systems are analyzed in two aspects: hierarchical analysis and data flow analysis.

3.3.1 Hierarchical analysis

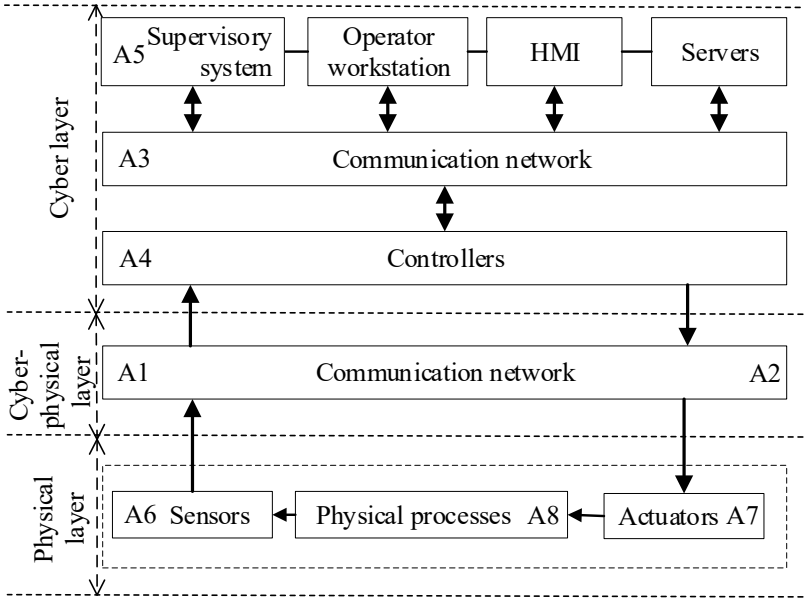


Figure 3.1 Architecture and composition of a cyber-physical system

The architecture of a typical cyber-physical system can be conceptually illustrated in Figure 3.1 in three layers: cyber layer, cyber-physical layer and physical layer [130]. The cyber layer contains high-level human machine interface, control algorithms, information and data processing devices. Its functions include data processing, control command generation, and high-level process management and optimization [131]. The physical layer typically consists of sensors, actuators, and physical processes. These elements are generally in hardware forms. Cyber-physical layer consists of network infrastructure that facilitates data exchanges between the cyber layer and the physical layer. Communication protocols are used to ensure smooth cyber-physical interactions [132]. To analyze security of cyber-physical systems, Cyber-physical security, all three layers have to be involved: (1) data processing and control in the

cyber layer, (2) data transmission in the cyber-physical layer, and (3) sensor measurements and control commands in the physical layer.

3.3.2 Data flow analysis in a control loop

In cyber-physical systems, there are two types of data flow in a control loop: sensor data flow and actuator data flow. In a cyber-physical system, the cyber system interacts with the physical system by reading the sensor data and sending the control commands through cyber-physical interactions. Sensor data flow and actuator data flow are interdependent, a change in one side will lead to changes in the other side [10].

1) **Sensor data flow.** Sensor data are sensor measurements from the physical system, compromised sensor data may mislead the controllers to make false control commands and result in security violations at the physical system.

2) **Actuator data flow.** Actuator data flow are transmitted from the cyber layer to the physical system. Such information in the cyber space can be used by attackers to cause undesired deviation in the operation of the physical system.

3.4 Attack analysis

3.4.1 Insider attacks on cyber-physical systems

3.4.1.1 Potential entry points for insider attacks

Potential entry points of cyber-attacks are labelled as A1- A8 in Figure 3.1. The nature of these attacks is explained in Table 3.1.

Attacks initiated from points A1, A2 and A3 target the sensor measurement data and control commands. Under such attacks, adversaries may interrupt the communication connection, eavesdrop to gain information on the system state, or tamper with the transmitted data packets [17]. These intrigue activities can lead to: denial of service attacks(DoS), deception attacks, false-data injection attacks (FDI), and replay attacks. The adversaries might conceal other

illegitimate activities from human operators or event detection algorithms implemented in the supervisory system. Such attacks are called stealthy attacks.

Attacks initialized at A4 and A5 can compromise the controller or the supervisory system or alter some system configurations [17]. Attacks launched from A6, A7 and A8 can be viewed as physical attacks.

This dissertation mainly focuses on security issues as a result of cyber-physical attacks that tamper with the data streams to cause damages in the physical process. The attack surface is originated from A1, A2 and A3, possible attack scenarios can be summarized in Table 3.1.

Table 3.1 Potential entry points for insider attacks on CPS

CPS layers	Label	Entry points of attacks	Security issues	Attack scenarios
Cyber-physical layer	A1	Communication network between the sensors and the controllers	Interrupt the communication between the sensors and the controllers; Manipulate or/and eavesdrop measurement data sent to the controllers	Denial of Service (DoS) attack Deception attack False-data injection attack Replay attack
	A2	Communication network between the controllers and the actuators	Interrupt the communication between the controllers and the actuators; Manipulate or/and eavesdrop the data package sent to the actuators	Denial of Service (DoS) attack Deception attack False-data injection attack Replay attack

Cyber layer	A3	Communication network between controllers and supervisory systems	Interrupt the communication between the controllers and the supervisory systems; Manipulate or/and eavesdrop the data package between the controllers and the supervisory systems	Denial of Service (DoS) attack Deception attack False-data injection attack Replay attack
	A4	Controllers	Interrupt normal operations of the controlled process, manipulate the control logics in the controllers, or send tampered data to the supervisory system/detection system	Denial of Service (DoS) attack Deception attack False-data injection attack Ladder logic bombs[133]
	A5	Supervisory system	Compromise the supervisory systems, change system configurations, or disrupt detection systems	Malware, code or program injection
Physical layer	A6	Physical process	Physical attack on physical processes	Direct physical attacks
	A7	Sensors	Physical attack on sensors	Direct physical attacks
	A8	Actuators	Physical attack on actuators	Direct physical attacks

3.4.1.2 Definition of a successful attack

In this dissertation, a successful attack is defined as: (1) the attack goal has been achieved; and (2) the attack is stealthy before its goal is achieved.

Given the attack goal, the attacker utilizes available resources to carry out a sequence of malicious actions. A successful insider attack can be marked by an action or actions that drive

a physical process beyond its safety limits while remaining undetected. This definition can be used to evaluate whether the attack is successful or not.

A failed attack: The insider attack fails if it is detected before the attack causes safety issues in process variables.

3.4.2 Attack trees

A comprehensive attack tree that integrates anti-models are constructed to show the logical sequence of an attack. Attack tree relate the system vulnerabilities from the attacker's entry points in cyber layer to physical processes. Anti-goals are used to model an attacker's malicious intentions related to system vulnerabilities [164]. By constructing a comprehensive attack tree, attack anti-goals and steps can be mapped to system vulnerabilities.

An attack tree is shown in Figure 3.2.

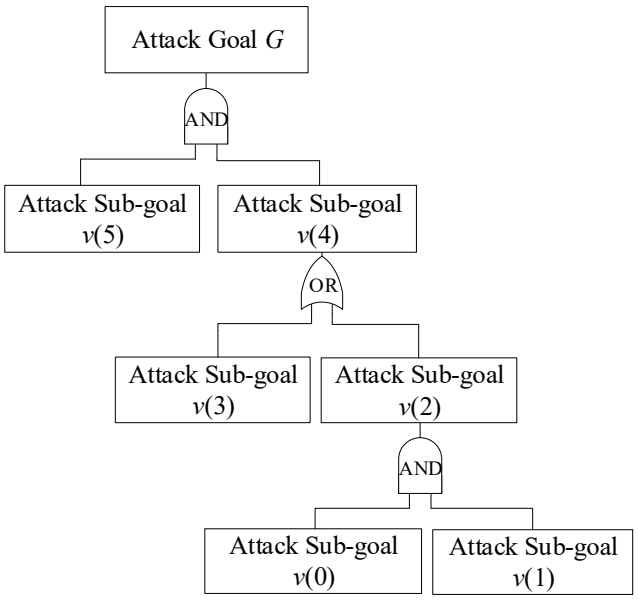


Figure 3.2 An attack tree

The construction of an attack tree starts from the identification of the attack anti-goal and sub anti-goals. The attack anti-goal, sub anti-goals, and attack steps are linked by logical

connective functions and subsequently a tree structure is formed. The synthesis of the tree is described graphically using connective symbols (AND, OR). The node that represents the attack goal is referred as a root node. When an attack sub goal is broken down further, the corresponding node is called a non-leaf node. Once an attack sub anti-goal is exhaustive, or when it is decided not to expand the analysis further, the corresponding branch is terminated with a leaf node.

In this chapter, an attack tree with system state and attack scenarios has been constructed to identify the attack pattern from the cyber domain to the physical process.

Procedures to construct a comprehensive attack tree is as follows:

(1) Set attack goals and sub-goals;

(2) Design attack mode scenarios;

(3) Define the attack steps;

(4) Link the attack steps as a chain to form a complete attack path;

(5) Integrate common attack steps for different attack mode scenarios; and

(6) Construct a complete attack tree.

The attack tree is constructed from the top to the bottom, but the execution sequence of an attack is from the bottom to the top. The attack tree results from a graph theoretic analysis of the network, security of a network and its interaction with the physical processes can be analyzed based on the attack tree.

The AND-OR refinement structure can be used to link sets of sub goals in an attack scenario. Each sub-goal in the attack tree is considered to be a vulnerability point in the system, $v(0)$ is the entry point of an insider attack on the CPS, and G is the final attack goal.

Thus, the vulnerability vector of each path $P(i)$ is:

$$V(P(i)) = \{v(0), v(1), \ldots, G\} \tag{3.1}$$

In order to identify the vulnerabilities, all the nodes and variables in each attack path are needed to form a valid attack path.

3.4.2.1 Attack path identification using attack tree

Attack paths and the corresponding steps can be identified based on the attack tree from the bottom to the top. An attack path not only reflects the cyber-physical interactions of the system, but also reveals the attack sequence hidden within the system. Attack steps based on the corresponding attack path can also be identified.

For rest of this chapter, a practical cyber-physical system is used to illustrate the conception. This system is known as Nuclear Process Control Test Facility (NPCTF), as is shown in Figure 3.3.

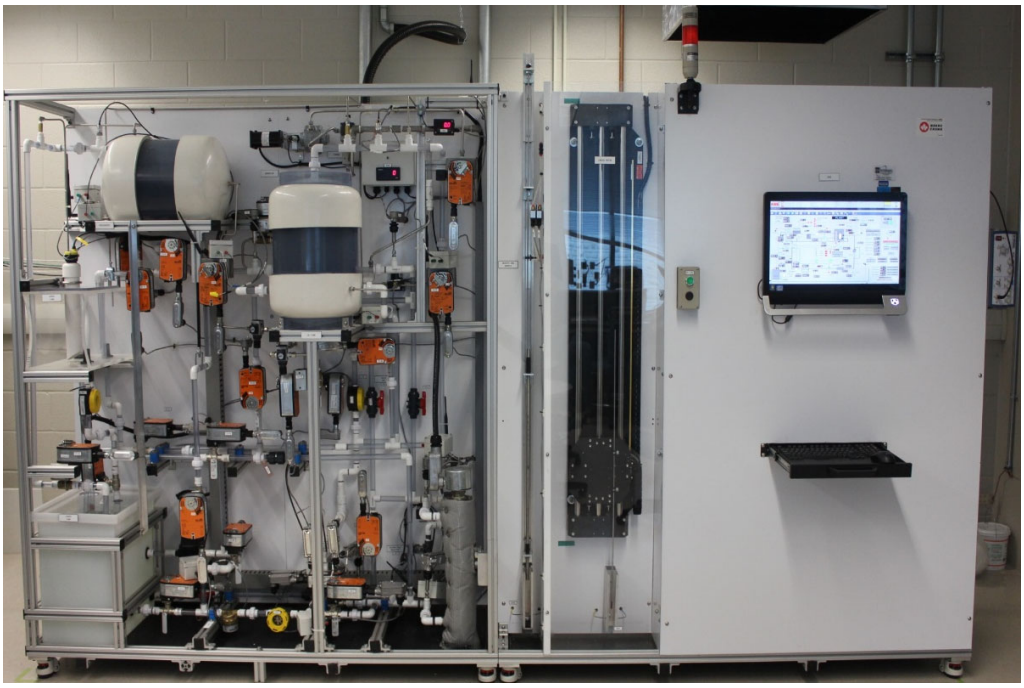


Figure 3.3 Overview of the NPCTF

3.5 NPCTF environment

In this dissertation, all case studies are implemented on NPCTF. As a foundation of the case study, this chapter analyzes the vulnerabilities and potential insider attacks. For generality and simplification, the following chapters will give brief introductions on NPCTF.

3.5.1 Cyber-physical aspects of NPCTF

NPCTF is designed as a general-purpose process control test facility, supporting research in instrumentation and control (I&C) at the Control, Instrumentation, and Electrical Systems (CIES) Research Laboratory at The University of Western Ontario (UWO) [134].

NPCTF is a fully operational scaled version of a physical plant, which represents the relevant portions of a cyber-physical system. In order to provide accurate information and real-life experimentation capabilities, this facility consists of a physical simulator to mimic the dynamics of a nuclear power plant, real field devices placed in the physical environment, and real programmable logic controllers (PLCs), which generate and exchange data packets via a communication switch in the cyber layer.

To identify the vulnerabilities related to insider attacks, the environment for a cyber-physical security platform used in the NPCTF is described as follows.

3.5.1.1 Physical Process

In NPCTF, sensor measurements and actuator signals are associated with the analog or digital data in thermal-hydraulic processes, controllers generate actual data packets and interact with field devices to carry out detection and control tasks. There are totally 19 AIs, 30 AOs, 8 DIs, 14 DOs and 12 control loops in the NPCTF. Detailed description of the control loops can be found in [134].

Sensor readings and control signals are transmitted between field devices and PLC. Analog I/O messages are in the form of 4-20 mA signals that must be converted back and forth to their corresponding physical values.

Insider attacks on sensors and actuators will choose I/O messages between devices and PLC in NPCTF as input to attacks, which are exchanged through TCP/IP packets.

3.5.1.2 Control System and Anomaly Detection System

An ABB Freelance AC700 PLC is chosen to implement the protection and control for the NPCTF. The PLC receives measurements from the sensors and computes the corresponding control actions. Anomaly detection system is set based on the safety limits of the physical processes. Anomalies can be detected if a process variable exceeds the designated limits. The Anomaly detection system is also used to detect anomalies caused by attacks on NPCTF.

The control algorithm and anomaly detection system are programmed in a Ladder logic diagram and sequential event logic diagram, using the ABB Control Builder F. Since there were no security checks for performing logic updates, an attacker can tamper with the sensor readings or control signals to the actuators through this vulnerability.

3.5.1.3 Communication Network

Measurement data from the sensors and actuator data sent to the actuators are collected as (AI, DI) and (AO, DO) and transmitted over the control network in the NPCTF. Information between NPCTF process and the ABB PLC is communicated via a field bus, and communication between PLC and HMI is based on TCP and UDP protocols.

3.5.2 Vulnerability analysis of NPCTF

For clarity of presentation, the heater control loop on NPCTF is selected as an example to analyze the system vulnerabilities. The heater control loop is shown in Figure 3.4.

There are one actuator (C2) for regulation of heater power, and two sensors (T_1, T_2) for inlet and outlet temperature measurement in the heater control system, respectively, as shown in Figure 3.4. The outlet temperature T_2 is regulated by the heater current C2 through a proportional (P) controller, when the sensor readings of T_2 to the PLC decreases, the current signal C2 in the heater will increase accordingly. The anomaly detection system is designed

according to the minimum (LL) and maximum (HH) bounds defined by the system safety limit. The safety limit of water temperature T_1 and T_2 are set at 37℃, which should not be surpassed, otherwise potential damages to the system can occur and force a system shutdown.

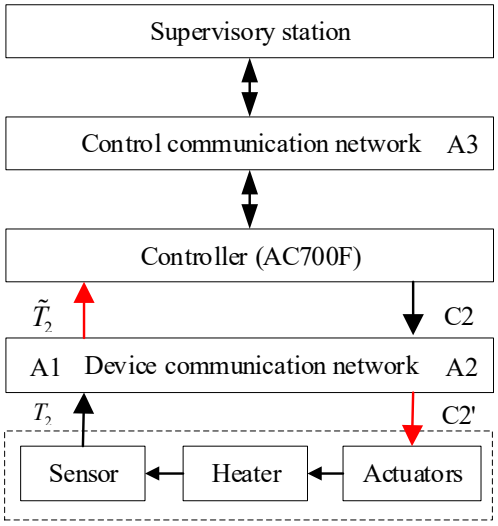


Figure 3.4 Cyber-physical attacks on the heater control loop

ECCS (emergency core cooling system) is used as an emergency control when the system is in an unsafe state.

The safety setting on the heater control system is given as follows: the safety boundary of T1 and T2 is set as <HH=35℃, LL=15℃>, and the set point is 37℃. The current C2 ranges from 0 to 100%.

The vulnerabilities of the heater control loop are analyzed by considering the following aspects.

1) Architectural vulnerabilities

There are three potential entry points on NPCTF for insider attacks, as listed in Figure 3.4. A1 is the entry point to the sensor communication channel from the sensor measurements to PLC. A2 is the entry point to the control communication channel from PLC to the actuator of the

heater. A3 is the entry point to the communication network between PLC and the supervisory station.

2) Security policy vulnerabilities

Currently, there is only passive security policy, which is the firewalls and safety thresholds. Attackers may take advantage to intrude the communication network and attack the heaters of the system.

3) Software and hardware vulnerabilities

All the hardware on NPCTF have no safety and security protection. Most of the software are open-sourced and have no access control or encryption. This vulnerability may open some backdoors due to the lack of security policies.

4) Communication network vulnerabilities

TCP and UDP are used in the communication network between PLC and the supervisory station, UDP protocol is vulnerable to most of the sniff tools. The attackers can compromise the communication network and deliver the attacks into the channel.

5) Detection and control related vulnerabilities

There are only passive detection measures, such as safety limit, firewalls are used in NPCTF, which is ineffective for insider attacks. The existing controller in the heater control loop is a PD controller for normal operation, which cannot maintain the system performance under various situations.

Based on the knowledge of these vulnerabilities, an attacker can take advantage of cyber-physical interactions and identified these vulnerabilities to design stealthy attacks and drive the system into unsafe state.

3.5.3 Insider attacks on NPCTF

Based on the potential entry points, the insider can attack in the heater control system by three means: (1) attack by tampering with sensor data T_2; (2) manipulation of the control commands C2; and (3) modification to the setpoints of T_2. Types of attacks may include false-data injection attack, replay attacks, and other more sophisticated deception attacks.

Given that the attacker aims to attack the heater outlet temperature of T_2, an attack tree can be constructed to analyze these possible attacks in Figure 3.5.

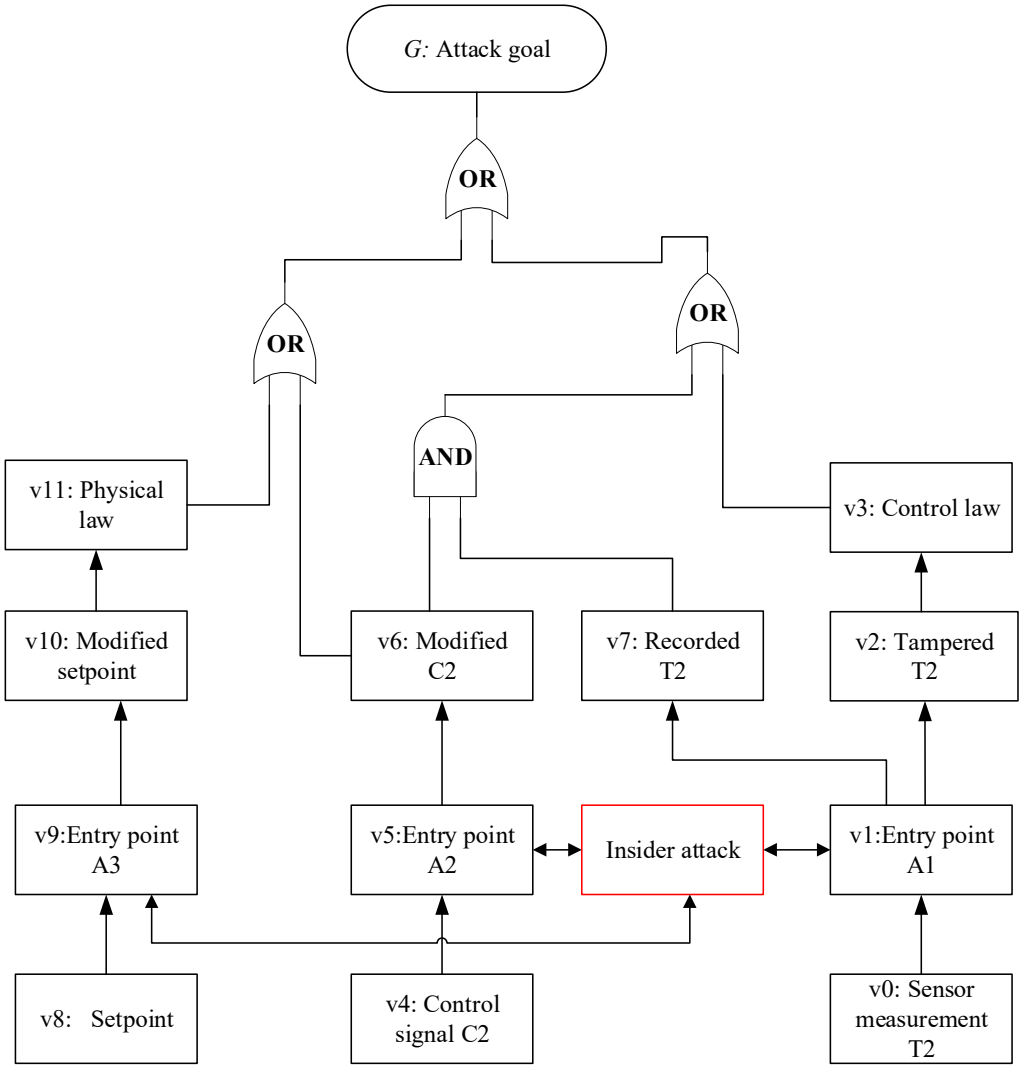


Figure 3.5 Attack tree analysis of the heater control system

Based on the attack tree, it can be observed how an attacker can advantage of system vulnerabilities from the cyber space to physical process. Attack steps and attack path can also be derived from the attack tree.

3.6 Conclusions

In this chapter, a general analysis of the vulnerabilities subject to insider attacks are presented. Possible entry points and the corresponding insider attacks are listed based on the system architecture. An attack tree is used to analyze impacts of attacks and to demonstrate the relationship between system vulnerabilities and insider attacks.

To demonstrate some basic concepts, a specific cyber-physical system, NPCTF, is used as an example in this chapter. A heater control loop in NPCTF is selected to demonstrate how an attack tree can be constructed. Based on the analysis and discussion in this chapter, various case studies will be carried out for specific aspects in attack generation, detection and defense in the following chapters.

Chapter 4

4 Design of a Modular Platform for Security Assessment of Cyber-Physical Systems

4.1 Introduction

Many research works have been done recently to investigate security aspects of CPSs by developing techniques to identify vulnerabilities in existing systems that could potentially be exploited by attackers [135, 136] and assessing impacts in an event of a security compromise [137]. Subsequently, various detection and mitigation strategies are proposed to boost the security and to minimize the consequences of these attacks [32, 76]. However, before these techniques can be deployed in practice, it is necessary to evaluate their effectiveness first in an environment that resembles the realistic situation as much as possible. Due to destructive nature of some of the cyber-physical attacks, it may not be safe, nor practical to carry out some attacks on the real process being protected, even under strict control, just to validate the security measures. As a result, many results in the previous research works remain in an idealized and a theoretical level until they are fully battle-tested by experiments in a physical environment [138]. Hence, it is safe to say that the nature and effectiveness of many existing security protection, detection and mitigation techniques are still not yet truly dependable.

To ensure the effectiveness of these techniques, it is essential to have a security assessment platform to analyze vulnerabilities in a cyber-physical system, and to experimentally validate and evaluate these techniques in a safe and controlled manner. Through this platform, one should be able to generate various attack scenarios after exploiting system vulnerabilities, and to implement different defense strategies, and finally to evaluate the strength of the security under various operating scenarios [139]. To meet the needs of simulating variety of attacks for different cyber-physical systems, it is highly desirable that the platform be modular and flexible.

In this chapter, a generalized guideline for testing security of cyber-physical systems is developed. The platform is composed of four main modules. Various types of attacks can be

modeled in the Attack Scenario Generation Module. Detection of potential security threats and corresponding defense strategies are implemented in the Security Enhancement Module. The level of security for a cyber-physical system can be analyzed and assessed in the Security Evaluation Module. The Platform Management Module ensures smooth operation of these three functional modules in real-time.

To validate the effectiveness of the proposed platform, an experimental demonstration has been carried out using a cyber-physical system in a laboratory environment. The case studies have shown that security test experiments can be tailored to evaluate various scenarios on such a platform. The proposed platform can be used to explore system vulnerabilities, to evaluate security enhancement strategies, and to assess the system security.

The chapter is organized as follows. Section 4.2 describes some of the existing work. Section 4.3 presents technical requirements and desirable features of a security assessment platform. Section 4.4 describes the platform design in detail. Construction of prototype platform is covered in Section 4.5. Section 4.6 presents results of case studies to demonstrate the features and effectiveness of the platform. Finally, Section 4.7 concludes this chapter.

4.2 Existing work

Development of experimental cyber-physical security test platforms has been an active topic of research over the last few years. Several institutions have developed such platforms for validating and evaluating various cyber security tools and technologies. These platforms also create realistic environments for testing attack/defense scenarios. Some of the existing platforms are compiled in Table 4.1, categorized by their intended use, implementation details, and application domains.

Most existing security platforms are focused on cyber-attacks originated from communication protocols or network configurations in cyber layers [147]. The main protections against such attacks are intrusion prevention and detection in the network, which strongly lean towards cyber security aspects. However, in a cyber-physical system, information in the cyber layer is closely coupled with the behaviors of physical process [153]. An evil goal of a perpetrator is

no longer merely to cause a network disruption, rather to inflict maximum damage to the physical process. Hence, further to the information in the cyber layer and the physical process, cyber-physical interactions and their interdependencies need to be considered when securing a cyber-physical system. For this purpose, a cross-layer platform is needed to support in-depth study of various aspects of cyber-physical security issues.

Table 4.1 Existing security testing platforms

Classification	Categories	References
Use of the platform	Cyber security	[139], [140], [141], [139, 142-145], [146, 147], [148], [149, 150]
	Control theoretic-based security	[151],[152]
Physical vs simulation in implementation	Real cyber, real physical	[140], [141], [142]
	Real cyber, simulated physical	[139, 143, 144]
	Simulated cyber, real physical	[146, 147]
	Simulated cyber, simulated physical	[148],[152]
	Hybrid (hardware-in-the loop)	[151], [149, 150]
Targeted domains	Smart grids	[139, 143, 144] [146, 147], [148]
	Power systems	[142], [149, 150] , [152]
	SCADA	[146, 147], [148]
	Water treatment plants	[151],[141]

From an operational safety point of view, a control system for the CPS should be designed such that, when a malicious attack is detected, the safety of the system should be maintained. For this reason, the platform must be implemented in such a way to automate test workflows and accommodate cyber-security evaluation through various test cases, while maintaining safety for entire system. In other words, the platform needs to integrate both cyber-physical security functionalities and control system actions within the same framework. To the best of our knowledge, there are no such platforms reported in the open literature.

There are various implemented methods to present a cyber-physical environment with different research concerns. With heavily inclined focus on cyber security over cyber-physical security, many existing security platforms are implemented using high-fidelity models or even real cyber components for the cyber parts, but with much simplified or even software simulated physical processes, such as in [139, 143, 144]. Unfortunately, an overly simplified physical process may not be able to provide in-depth information on the behavior of actual physical process, its control systems, and more importantly cyber-physical interactions during an attack.

It is shown that the platforms built on physical installations in [146, 147] do provide more insightful responses from the physical processes. One can also capture interactions among cyber and physical parts for realistic cyber-physical system interactions in the security experiments. However, the use of physical components does not always guarantee repeatability as there are so many uncontrollable factors involved. It is also difficult to maintain original system functionalities, especially when attack tests are underway.

On the other hand, platforms based entirely on simulation in [148] and [152] provide strong repeatability, but they only represent a limited number of practical scenarios. The results of tests may not be representative, and test credibility could be in question for general cases, especially when cyber-physical interactions are strong and interdependent. However, although these implemented methods provide various solutions on how to reproduce a specific cyber-physical environment, discussions on how to develop and conduct security tests based on the implemented cyber-physical environment are very limited.

Thus, it is necessary to extract key features of a security platform and to design a generalized prototype security platform that is applicable to different cyber-physical systems. In addition, it would be helpful if the common aspects could be extracted into a modular design, as it provides flexibility to add or change features to support various test scenarios for different security concerns.

Furthermore, many existing researches on cyber-physical security platforms are domain-specific, such as power systems [142], [149, 150] , [152] and water treatment plants [151],

[141]. One of the reasons might be that domain knowledge and specific implementation details may be relevant to understand the mind of a potential attacker. However, there are many attack scenarios are common across different application domains. It would be useful to develop some general design guidelines for test platforms to investigate various security issues independent of domain of applications.

After a literature review, the following shortcomings in the existing platforms have been identified:

- Most of works in the context of security platforms are focused on cyber security, lack of works to study cyber-physical security, and to combine cyber-physical security and control systems for securing a cyber-physical system.
- There are currently no generalized design methods on how to design a security assessment platform.
- There are no modular design and implementation of security testing platforms in the published work.

This chapter provides a design method for an experimental security assessment platform to address the above-mentioned shortcomings and meet the comprehensive requirements of a cyber-physical security test platform. Key features of the proposed platform are:

- Provide general design methodologies for different cyber-physical systems found in different domain of applications;
- Adopt a modular design philosophy so that different modules can be selected and assembled to meet unique needs in different security evaluation scenarios; and
- Support cross-layer tests for cyber-physical security and combine control system design with the consideration of cyber-physical security.

4.3 Platform requirements

The functionalities of the platform are as follows.

- Identify vulnerabilities in both cyber and physical layers that might be exploited by attackers;
- Generate various attack scenarios to expose and identify vulnerabilities of the cyber-physical system and to understand the cascading effects of an attack;
- Develop and validate different cyber-physical security enhancement solutions to increase system resilience; and
- Evaluate the results of security tests and provide insights and procedures for mitigating the effects of the attacks and minimizing their impacts.

According to the expected functions, the proposed platform is decomposed into three main functional modules and one Platform Management Module. The three functional modules are: (1) Attack Scenario Generation Module, (2) Security Enhancement Module, and (3) Security Evaluation Module. The modules and their respective functionalities are summarized in Table 4.2. Requirements of each module is analyzed according to the expected functionalities.

4.3.1 Requirements of functional modules

4.3.1.1 Requirements for Attack Scenario Generation Module

Vulnerabilities can be identified by analyzing potential avenues that an attacker could take to mount an attack. For this purpose, attack scenarios need to be generated and their profiles need to be extracted. This module is known as Attack Scenario Generation Module. The module should be able to generate both preprogrammed attacks and customized attacks based on the specific research interests and practical concerns. The preprogrammed attack scenarios can be generated automatically or manually, other attacks can be generated by the users based on their specific knowledge and acquired resources.

Table 4.2 Summary of functional modules

Functional Modules	Functionalities	Description
Attack Scenario Generation Module	Vulnerability analysis	Explore existing vulnerabilities
	Attack tests	Generate different attack scenarios to examine how an attack could inflict physical damage from the cyber domain
	Impact assessment	Assess the impact of various attacks
Security Enhancement Module	Tests of detection methods	Implement different detection rules
	Tests of defense strategies	Reconfigure and perform various decision-making logics and defense strategies
Security Evaluation Module	Security assessment metrics	Develop various metrics to evaluate the security related performance
	Security evaluation	Assess effectiveness for different defense strategies
Platform Management Module	Monitoring of tests	Ensure a safe experimental environment for security tests
	Adding/removing functional modules	Modularized design, flexible to add/remove scenarios for specific security issues
	Data collection and analysis	Log data of each test scenario

A process to generate a cyber-physical attack scenario is shown in Figure 4.1. This process is developed based on an attack kill-chain in [165]. An attack can be launched in two stages. At the preparation stage, the attacker needs to get access to the communication channel, gathers required information, develops attack strategies and builds the attack path to deliver the attacks. Therefore, to generate an attack scenario, the Attack Scenario Generation Module should contain communication interface to capture network traffic and gain network information, design attack scripts to generate attack scenarios, and triggering schemes to trigger the attack

scenarios. During the attack execution stage, to capture and extract the attack profiles, the Attack Scenario Generation Module should be able to monitor and record the attack activities.

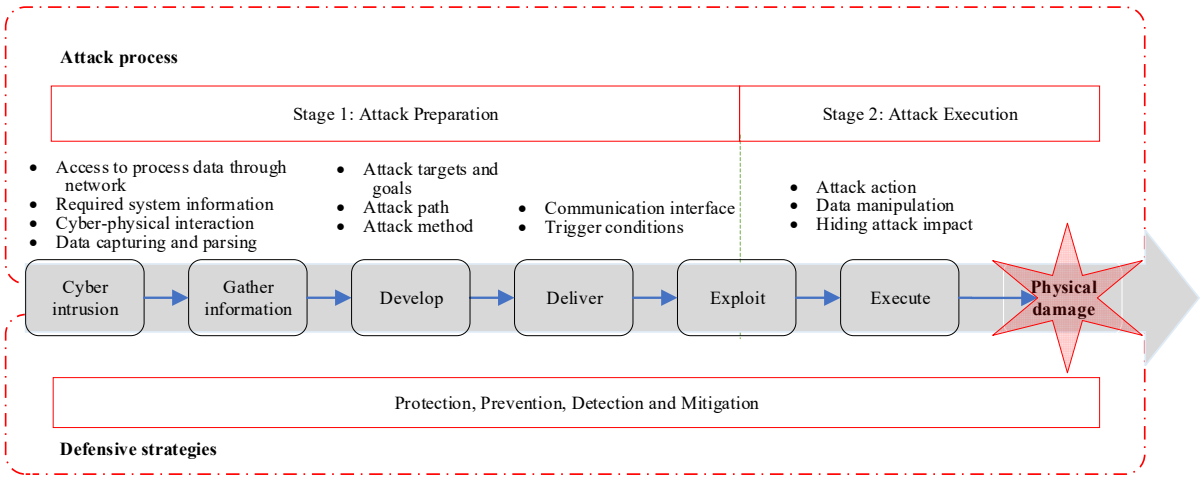


Figure 4.1 Process to generate an attack

4.3.1.2 Requirements for Security Enhancement Module

To foil an imminent cyber-physical attack, a cross-layered detection scheme and defense-in-depth mitigation system is needed. This is carried out by a Security Enhancement Module. The module should be able to accommodate various strategies for security enhancement and flexible enough to change detection or defense strategies. Integration of cyber-physical security and control should be taken into account in the meanwhile. An example of a defense framework is shown in Figure 4.2 [67]. Once the attack detection scheme reveals an imminent attack, and attack mitigation scheme can be activated by the detection mechanism to respond to the detected threat and reduce its adverse effects.

4.3.1.3 Requirements for Security Evaluation Module

A comprehensive evaluation framework, together with a set of user-friendly tools is also needed for examining and evaluating the security and defense-readiness levels. This is performed by a Security Evaluation Module.

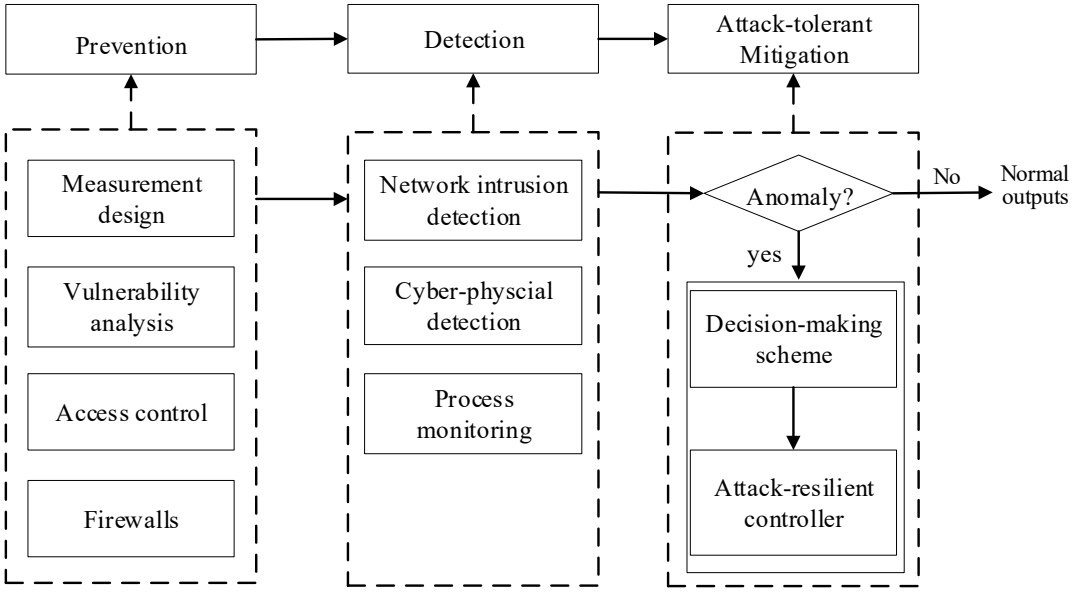


Figure 4.2 Defense framework for a cyber-physical system

Development of these techniques and tools may require data interaction and aggregation from different sources. Some in-depth security analysis and evaluation of the related systems is also needed within the platform. Furthermore, measurable security metrics also needs to be defined to assess the effectiveness of detection and mitigation schemes.

4.3.2 Overall design of the proposed platform

To meet the technical requirements, an overall framework of the proposed platform is proposed as in Figure 4.3. The locations of modules and data interactions are also illustrated in Figure 4.3.

The Attack Scenario Generation Module is connected with the cyber-physical system via an activation switch. The Security Enhancement Module is connected to the device communication channel and controllers to detect anomalies and execute attack mitigation strategies. The Security Evaluation Module gathers information from the process and other modules, and defines suitable security metrics for system security evaluation. The Platform

Management Module consists of an off-line part and an on-line part. Based on specific test requirements of a given scenarios, the off-line part determines suitable module compositions to form an effective test environment, while the on-line part oversees the entire operation of the platform during the test process to ensure safety and operational effectiveness.

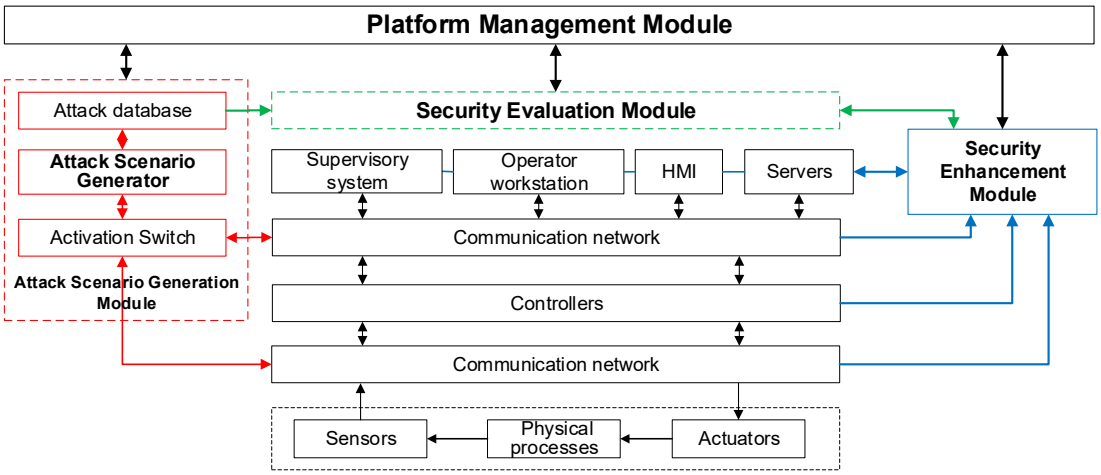


Figure 4.3 Proposed architecture of a cyber-physical security platform

4.4 Design of functional modules

4.4.1 Attack Scenario Generation Module

To identify potential vulnerabilities and to trace consequences of an attack, an Attack Scenario Generation Module is designed to mimic realistic cyber-physical attacks. The details of this module are shown in Figure 4.4. It contains four sub-blocks stored in the form of an Attack Library, i.e. attack scripts, targeted channel selection, attack duration setting, and attack trigger logics. Descriptions of these sub-blocks are further listed in Table 4.3.

In addition, there is an activation switch within this module, which captures and transmits data between this module and the cyber-physical part of the system. A user can gain access to the targeted network channel via this activation switch either remotely through a network or by tapping into the network physically through pre-defined open ports.

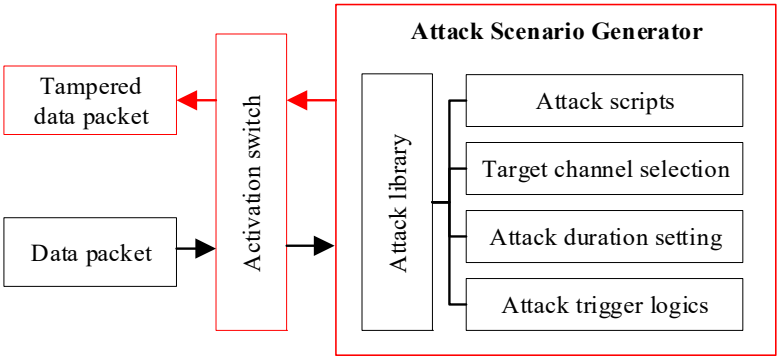


Figure 4.4 Organization of an Attack Scenario Generation Module

Table 4.3 Attack library in an Attack Scenario Generation Module

Sub-blocks	Description
Attack scripts	Design details of the attack scenarios
Target channel selection	Selection of access points for targeted communication channels
Attack duration setting	Record of the process for a staged attack scenario
Attack trigger mechanism [133]	Attack triggered when a pre-determined input is detected
	Attack triggered when a particular trigger sequence is detected
	Attack triggered when the timer has ended its count sequence
	Attack triggered when a particular internal state is achieved

This module is connected to the communication network between the field devices and the controller, and between the controller and the supervisory workstation. It is capable of interrupting or manipulating sensor readings and control flow through this switch. The implemented work includes gathering and parsing the communication data packets, designing attack scripts, building up attack paths, and setting up target communication channels, attack duration, and trigger conditions.

This module is designed as an open-source attack library. Pre-programmed attack scripts and trigger logics are constructed within the attack library. This module is generic to generate various attack scenarios in different communication channels. All the sub-blocks within the module can be edited, added or removed based on the requirements in the tests. Attack scripts

in the module can be written and stored in the library, it can also be generated by the user of this platform.

This module is implemented on a separate computer. It provides a wide attack surface including data attacks on sensors, actuators and controllers. It is an open source platform not only for the pre-programed attack scenarios, but also for other customized tests.

Outcomes of the Attack Scenario Generation Module are as follows:

- Vulnerability analysis of an existing architecture, cyber access, communication protocols, data flows between the control system and the physical process;
- Vulnerability indicators for security enhancement; and
- Index for evaluating attack impacts.

4.4.2 Security Enhancement Module

To mitigate adverse effects of an attack, detection and mitigation schemes are implemented in this module. The composition of the module is shown in Figure 4.5. This module contains several sub-functional blocks to support testing and validation of various detection and mitigation strategies. The detailed functional units are described in Table 4.4.

There are three main functional parts in the Security Enhancement Module. The first one is for data collection and processing. The second one is for cross-layer attack detection. It supports process anomaly detection in physical layer, anomaly detection in cyber-physical layer, and network intrusion detection in the cyber layer. These detection results are forwarded to decision-making unit in the supervisory station. The third part is for attack mitigation. It consists of a decision-making unit and control algorithms. Defense-in-depth strategies that combine control and cyber-physical security can be implemented in these sub-blocks.

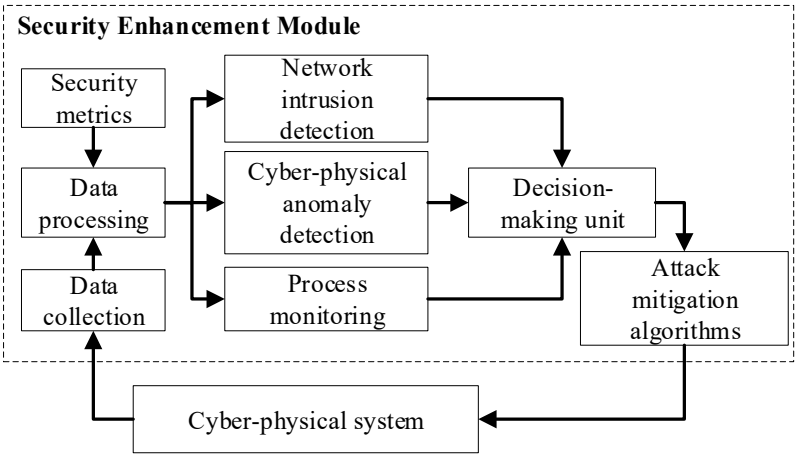


Figure 4.5 Function blocks of a Security Enhancement Module

Table 4.4 Cross-layer design of a Security Enhancement Module

Sub-blocks		Description
Data preparation	Data collection	Collecting data from different layers and channels
	Data processing	Process and forward data to detection blocks
	Security boundary	Defining security boundaries and detection rules
Detection	Network intrusion detection	Detecting intrusions of security boundary in cyber layers
	Anomaly detection in data transmission	Detecting anomalies in cyber-physical layer
	Process monitoring	Detecting anomalies in physical layer
Defense	Decision-making unit	Situation awareness, reconfiguration of operating conditions, and execution of control actions
	Attack mitigation algorithms	Executing attack-resilient control algorithms

Each functional unit in the Security Enhancement Module can vary in locations and with different implementation details. Detection schemes are deployed in multiple locations for cross-layer detection. Defense schemes are deployed in the supervisory system, it includes a decision-making unit and a control scheme. The decision-making unit is implemented within

the controllers using specific programming languages, while control algorithms are often implemented in a Distributed Control System (DCS) or a Programmable Logic Controller (PLC) and exchange data with the CPS through OPC or other communication channels.

To provide flexibilities for security tests, all the sub-blocks are designed as an independent modular, which can be operated as a combined unit or individually.

Outcomes of the Security Enhancement Module are potential anomalies and mitigation results provided by detection methods and mitigation strategies.

4.4.3 Security Evaluation Module

After attack scenarios are generated, and security enhancement strategies are conducted on the platform, one needs to (1) analyze system vulnerabilities associated with these attack scenarios, (2) assess their impacts, (3) analyze experimental results, and finally (4) evaluate the system security under specific mitigation strategies.

These functions are realized in the Security Evaluation Module as outlined in Figure 4.6. There are three sub-blocks in this module, details are described in Table 4.5. Data sources of this module are from the Attack Scenario Generation Module, the Security Enhancement Module and the supervisory system. All data are collected to evaluate the system security, and security metrics are calculated for security tests.

In Security Evaluation Module, the data from different layers are extracted and sorted out into two data streams. One is the actual measurements and signal values that are transmitted through the communication channels; and the other is the attacked values that are observed by the supervisory system or those forwarded to the physical process. In this module, security metrics are used as evaluation rules, and system evaluation methods are implemented according to different objectives.

Outcomes of the Security Evaluation Module include identified vulnerabilities, assessment of attack impacts, effectiveness of defense strategies, system security awareness and insights of security enhancement.

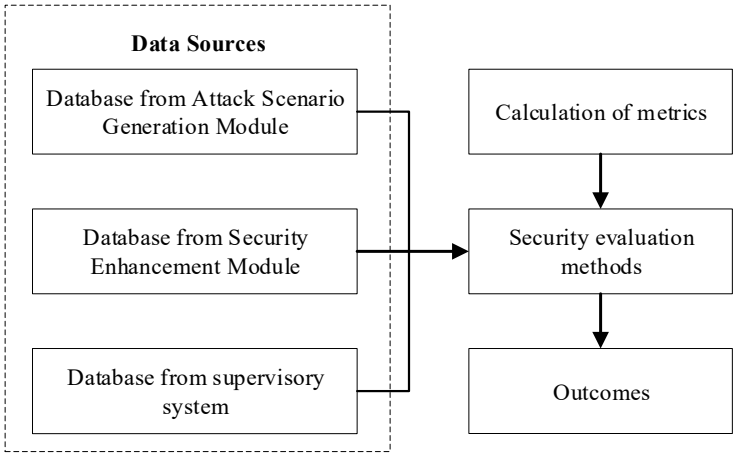


Figure 4.6 Function blocks of a Security Evaluation Module

Table 4.5 Design consideration in a Security Evaluation Module

Sub-blocks		Description
Data sources	Database from Attack Scenario Generation Module	Attack information and compromised data
	Database from Security Enhancement Module	Defense information and mitigated data
	Database from supervisory system	Historian of process status from the communication network
Calculation of metrics		Metrics to measure experimental results and performance validation
System evaluation methods		Evaluation methods used
Evaluation outcome		System security awareness

4.4.4 Platform Management Module

The Platform Management Module is used to manage other modules as well as the real-time monitoring of security and vulnerabilities within the platform during tests. Its interfaces with other modules are shown in Figure 4.7. Since the platform is designed in a modular fashion,

modules can be easily reconfigured, deployed and initialized according to the needs of specific security tests.

This module can be implemented in a separate computer to manage all the functional modules and monitor events and scenarios if necessary.

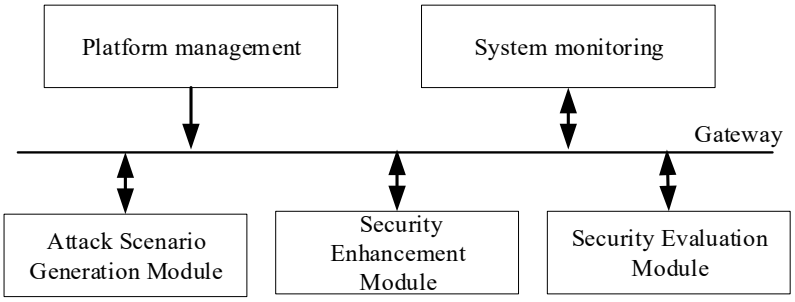


Figure 4.7 Function blocks of a Platform Management Module

4.5 Construction of a prototype platform

A well designed cyber-physical test platform should cover two aspects: (1) an experimental environment of the cyber-physical system, and (2) required functionalities for security tests.

To demonstrate the inner workings of the proposed platform, a prototype platform is constructed by integrating all the modules into an experimental environment in Figure 4.3. The environment is designed to represent key features of a cyber-physical system. The implementation details for each functional module are presented next.

4.5.1 Composition of a cyber-physical environment

Construction of the experimental environment can be divided into three main layers: (1) industrial control facilities and software in the cyber layer, (2) communication networks in the cyber-physical layer, and (3) a physical system including sensors and actuators in the physical layer.

Multiple sources for data collection and multiple access points for security tests are needed in this environment. As such, the functional modules can be connected to the cyber-physical environment to perform security tests.

4.5.2 Construction of a specific platform

Once the prototype platform is constructed, it is then connected to a cyber-physical environment known as the nuclear power control test facility (NPCTF). The constructed security platform on NPCTF is shown in Figure 4.8.

Details for each functional module are as follows. Tools used for each model is listed in the tables below, implementation details are described in the following sections. Considering the security of the designed platform and security of NPCTF, the developed code is not publicized, all the codes are stored in and managed by UWO CIE Lab. The codes package written for this dissertation are listed in Appendix A, demo videos for the platform and tests are listed in Appendix B.

4.5.2.1 Implementation of the Attack Scenario Generation Module

The Attack Scenario Generation Module is constructed on a separate attack computer under Kali Linux environment, procedures to generate an attack scenario is illustrated in Figure 4.9. Each sub-block in Figure 4.4 is implemented with details given in Table 4.6. In order to generate an attack scenario based on Figure 4.1, a user needs to get access to the communication network in NPCTF first. An activation switch is implemented as the communication interface between the Attack Scenario Generation Module and the communication network in NPCTF. The user can capture and gather the transmitted data packets through this activation switch. In order to read and inject attack scenarios, communication protocols are parsed, and attack scripts are compiled in Python language. Currently, three types of cyber-physical attack scripts are developed in this module, i.e. deception attack, false-data injection attack and replay attack. Since this module is an open-source platform, users can customize it and generate other cyber-physical attack scenarios as situations require.

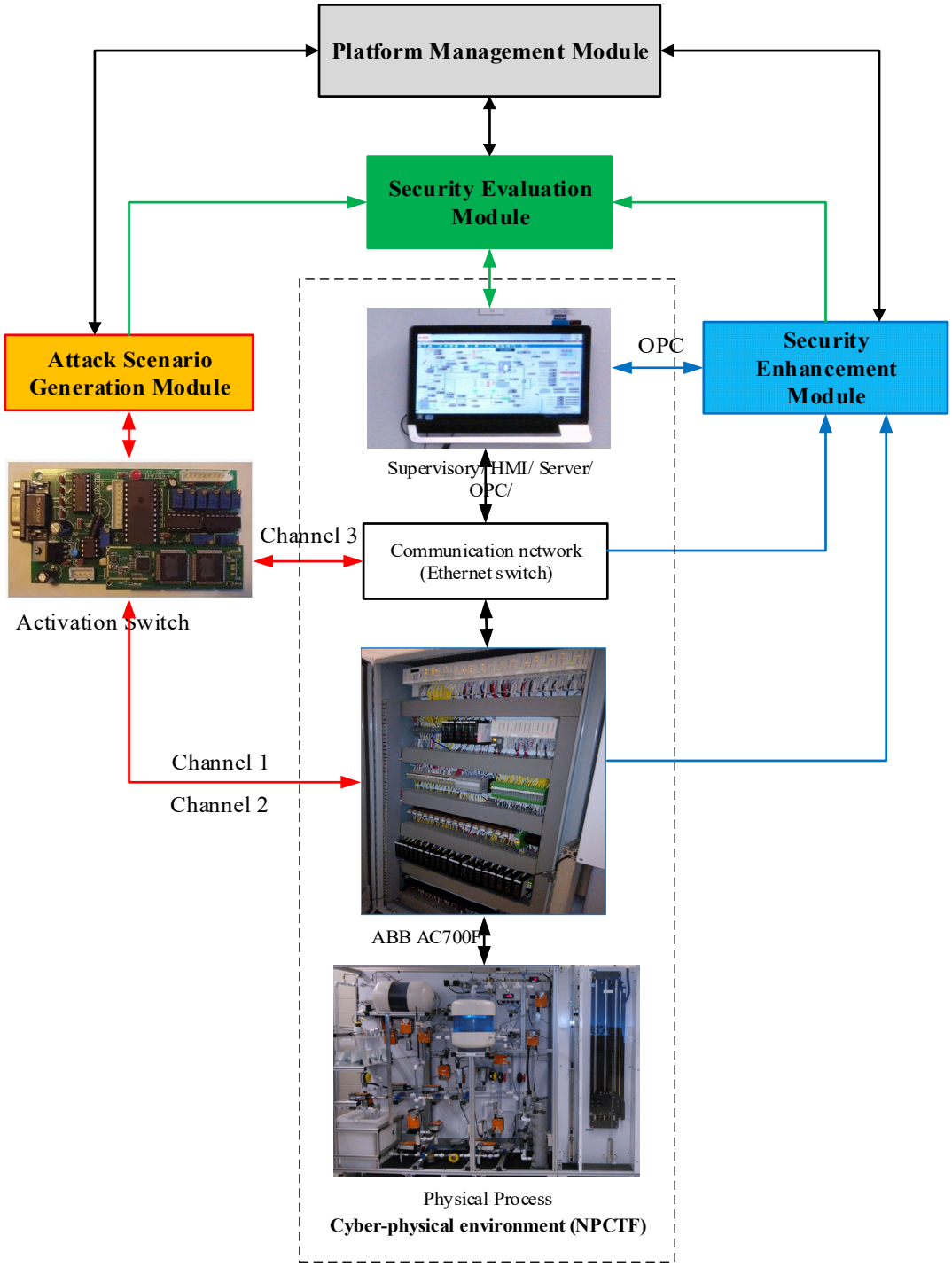


Figure 4.8 Composition of the prototype security platform

There are three communication channels that can be selected in the Attack Scenario Generation Module: (1) Channel 1 connects the network between the sensors to the controllers, (2) Channel 2 connects the network between the controller and the actuators, and (3) Channel 3 connects the network between the controllers and the supervisory station. These channels represent the attack entry points A1, A2 and A3 in Figure 3.1, respectively.

Attacks can be generated and launched from a separate attack computer. HMIs and PLCs are on two different subnets and connected through an Ethernet using TCP/IP. Bi-direction communication channels have been constructed by reversing the control protocols, transmitted data packet can be extracted through the activation switch. In this implementation, three types of cyber-physical attack scenarios are developed: deception attack, false-data injection attack and replay attack. Two trigger logics are also designed in the attack library.

The attack duration setting block and the attack trigger block are implemented using Visual basic language, which can generate the attack duration setting and trigger the attack scenarios.

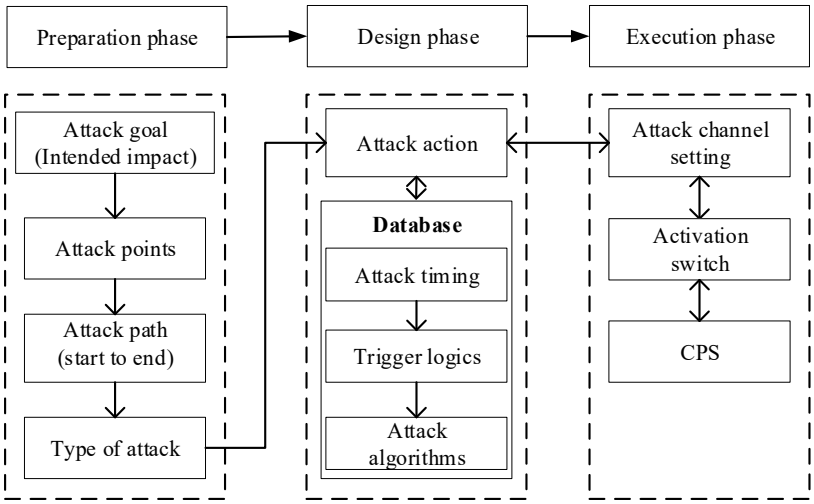


Figure 4.9 Procedures to generate an attack scenario

Table 4.6 Implementation of an Attack Scenario Generation Module

Sub-blocks	Functions	Tools
Attack scripts	Capture transmission data packet	Wireshark
	Parse communication protocols and construct attack scenarios	Python language
Target channel selection	Attack data transmission	Industrial activation switch
	Build graphical user interface (GUI) for channel selection	Visual Basic
Attack duration setting	Build GUI for attack duration setting	Visual Basic
Attack trigger schemes	Build GUI for selection of trigger logics	Visual Basic

4.5.2.2 Implementation of the Security Enhancement Module

To validate the detection and defense schemes, the detection and defense function block is implemented on the NPCTF through an OPC server, as shown in Figure 4.10. Data transmission and algorithms for detection and defense function unit are detailed in Table 4.7.

Different detection methods in different layers are implemented in the detection sub-blocks. Network intrusion detection unit and anomaly detection for cyber-physical interactions are deployed in the supervisory station using Snort. The anomaly detection unit for process data is implemented in a separate workstation. It collects the process data online through an OPC server in the supervisory station and performs the detection algorithms real-time in MATLAB. When there is an alarm, the detection unit will send the detected anomaly to the supervisory station.

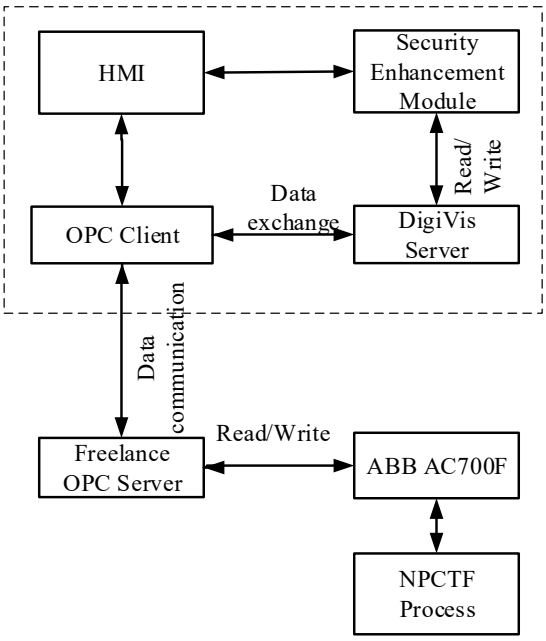


Figure 4.10 Implementation of a Security Enhancement Module on NPCTF

Table 4.7 Construction of a Security Enhancement Module

Sub-blocks		**Functions**	**Tools**
Data preparation	Physical process data collection	Collect process data	Freelance OPC 2000
	Network data collection	Capture transmitted data packets	Wireshark
	Data processing	Analyze and process datasets	BASE
	Security boundary	Define security metrics and safety thresholds as detection rules	Snort
Detection	Network intrusion detection	Scan port Detect network intrusions	Nmap Snort
	Anomaly detection	Monitor cyber-physical interactions Construct detection methods	Snort MATLAB
	Process data anomaly detection	Monitor process data	DigiVis
Defense	Decision-making unit	Configuration in Supervisory station	Control Builder F
	Attack mitigation algorithms	Compute the control parameters	MATLAB Control Builder F

Defense schemes are implemented in the supervisory station. Decision-making logics and controller structures are reconfigured through ABB Control Builder F and are loaded to PLC. Control algorithms are implemented in MATLAB to calculate the parameters of the reconfigured controller. The calculated parameters are sent back to PLC through OPC server.

4.5.2.3 Implementation of the Security Evaluation Module

During the operation, the status of the platform needs to be logged for further analysis and evaluation. The logged dataset contains the physical properties related to the process, as well as the network traffic including those in the midst of attacks. Security metrics are based on a specific area and concerns, and evaluation methods are implemented in the supervisory station and connected to NPCTF through an OPC server.

Implementation procedures for the Security Evaluation Module is shown in Figure 4.11. It resides on the supervisory station.

In the Security Evaluation Module, data from different layers are extracted and sorted into two streams. One is for control commands and parameters that compromise the network protocols; and the other is for the current state of the observed process variables [154].

In this chapter, NPCTF is served as a target physical process to demonstrate how the proposed design methods can be used to construct a specific security assessment platform. The proposed modular design is not restricted to only NPCTF, it can also be applied to other cyber-physical systems. It provides guidelines and methods to build up a cyber-physical security assessment platform.

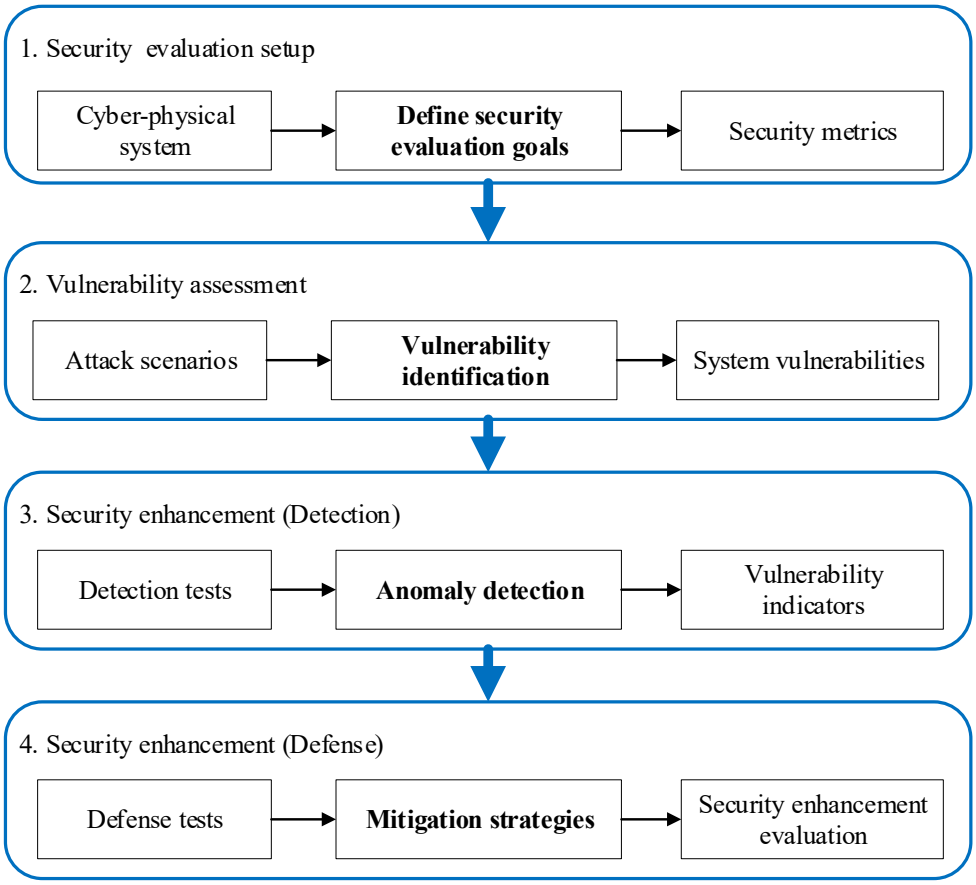


Figure 4.11 Implementation procedures for a Security Evaluation Module

4.6 Case Studies

To study the features and effectiveness of the proposed platform, case studies including various cyber-physical attack scenarios, detection methods and mitigation strategies are performed on NPCTF using the platform, as shown in Figure 4.8. The selected system to mount attacks is the heater control loop of NPCTF, as shown in Figure 4.12.

In the heater control loop, the outlet temperature T_2 is regulated by the heater current C2 through a proportional (P) controller. When the temperature T_2 is below the setpoint value, the controller (AC700F) will send out a command to increase the heater current. An anomaly detection scheme is designed according to the minimum (LL=15℃) and maximum (HH=37℃)

thresholds defined by the system safety specifications. If the measured temperature T_2 exceeds 37°C, the alarm will be triggered to cut off the current supply and subsequently trip the system shut down.

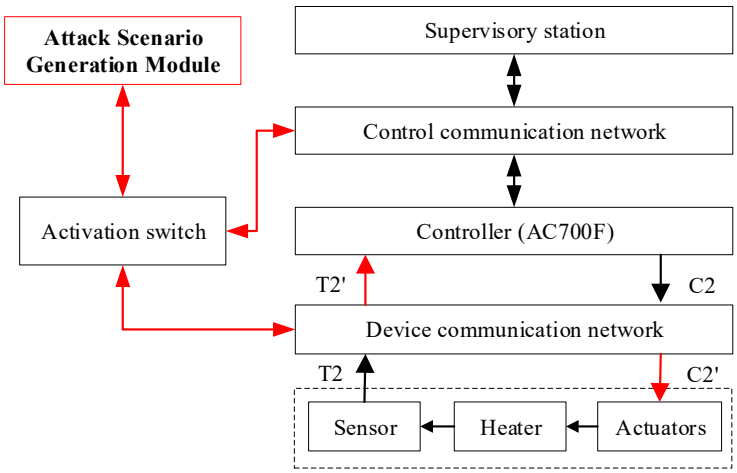


Figure 4.12 Cyber-physical attacks on the heater control loop

The attack goal is to drive T2 beyond its safety limit without being detected.

4.6.1 Experiment design

The experiment consists of three cases.

Case #1 is to validate the functionality of the Attack Scenario Generation Module. System vulnerabilities are explored through six attack scenarios. In order to reflect different attack surfaces and scenarios supported by the Attack Scenario Generation Module, three different attack scenarios are launched on sensor measurement T_2, and three different attack scenarios are launched on the control signals to actuator C2. The implemented attack scenarios are described in Table 4.8.

Case #2 is to validate the functionality of the Security Enhancement Module and test various cross-layered detection methods. There are four detection methods, D1-D4, that are deployed to detect the anomalies induced by various attack scenarios on NPCTF, as listed in Table 4.9.

Case #3 is to evaluate the performance of attack mitigation strategies. Different mitigation methods M1-M3, are development and tested. Design methods are listed in Table 4.9.

Table 4.8 Attack scenarios launched on the platform

Attack entry point	Attack scenario	Attack type	Attack description
Sensor measurement T_2	SA1	False-data injection attack	Modify $\tilde{T}_2(t) = T_2(t) - 0.05t$
	SA2	Replay attack	Record and replay historical data when injecting attacks
	SA3	DoS Attack	Blocking sensor measurements for 20s
Control commands to actuator C2	AA1	False-data injection attack	Inject a deviation of 10% to C2
	AA2	Replay attack	Record and replay historical data when injecting attacks
	AA3	DoS Attack	Blocking control commands to actuator for 10s

Table 4.9 Security enhancement methods used in Case #2

Security enhancements	Methods	Techniques
Detection methods	D1	Safety threshold (HH)
	D2	Rule-based network intrusion detection
	D3	CUSUM method: τ=0.5, b=1
	D4	Physical watermarking method [37]
Mitigation methods	M1	P controller
	M2	Decision-making unit
	M3	A PI controller with a state estimator

Mitigation methods consist of a decision-making unit M2 and a resilient control system M3. When M2 receives anomaly alarms generated from detection methods, it will decide how to respond to the attacks for the given situation.

Evaluation work are constructed based on these three case studies. Security evaluation metrics together with evaluation methods EV1-EV3 are defined in Table 4.10.

Outline of the experimental designs are presented in Table 4.11.

Table 4.10 Definition of security metrics

Evaluation	Concerns	Security metrics	Description
Case #1: EV1	Effectiveness of the Attack Scenario Generation Module	Test effectiveness	Are correct attack scenarios generated?
	Exploring vulnerabilities Impact of attacks	Attack successful criteria Attack duration	Achieve attack goal before being detected Start time and end time
		Attack impacts	System dynamics and consequences
Case #2: EV2	Evaluating performance of detection methods	Detection effectiveness	If attack is detected timely?
		Detection speed	Time to detect an attack measured from the moment the attack starts
Case #3: EV3	Evaluating performance of mitigation methods	Mitigation effectiveness	If attacks are mitigated?
		Response to attacks	System dynamics

Table 4.11 Case studies on the platform

Security tests	Case #1	Case #2	Case #3
Implemented attacks	SA1, SA2, SA3	SA1, SA2, SA3	SA1, SA2
	AA1, AA2, AA3	AA1, AA2, AA3	
Detection methods	D1	D2	D2+D3+D4
		D3	
		D4	
Mitigation methods	M1: P Controller	M1: P Controller	M2+M3
Evaluation Methods	EV1	EV2	EV3

4.6.2 Experimental results

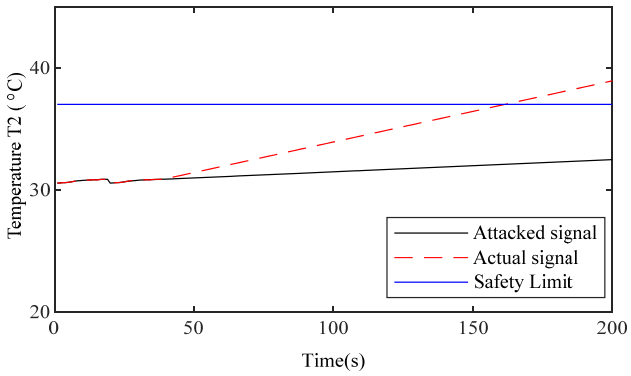
4.6.2.1 Case #1: Exploring system vulnerabilities

In this case study, six attack scenarios have been implemented to investigate system vulnerabilities. The experimental results of three attacks on sensor measurements (SA1, SA2, SA3) are shown in Figure 4.13.

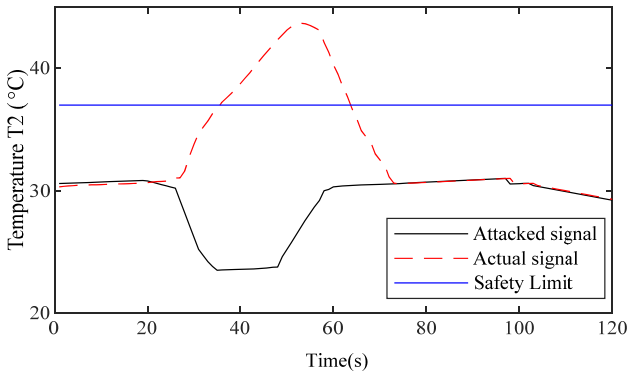
In SA1, the method of the attack is to manipulate the sensor data from T_2 by adding a slowly varying negative offset ΔT_2 = -0.05t and then sending it to PLC. The attack is set to be triggered at a predefined time t_{start}=25s. As shown in Figure 4.13(a), when the actual signal T_2 exceeds its safety limit of 37℃ at 164s, the tampered $\tilde{T}_2$ (Attacked signal) sent to the PLC is only appeared to be 24℃, the anomaly detection scheme is fooled.

In SA2, a replay attack is implemented during the transmission of the sensor measurement, the actual T_2 is replaced by a recorded historical data. When SA2 begins at time t=25s, the heater is running at its setpoint T_2 =30℃. The recorded data is sent to the controller, the deviation between the fake $\tilde{T}_2$ and the setpoint of T_2 leads to an increase of actual T_2. As shown in Figure 4.13(b), the replay attack continues being undetected until the actual T_2 is out of its safety limit of 37℃.

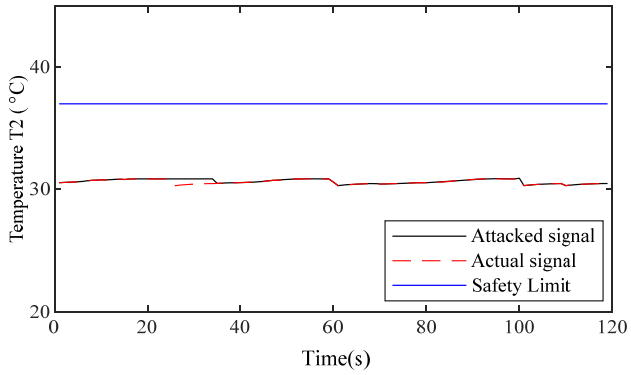
In SA3, the data packet is blocked for 20s when transmitted from the sensor to the controller. The last received data of T_2 is used during the communication interruption. It can be observed in Figure 4.13(c)that the DoS attack does not have a major impact if the system is at a steady state when the attack is launched.



(a) False-data injection attack on sensor (SA1)



(b) Replay attack on sensor (SA2)



(c) DoS attack on sensor (SA3)

Figure 4.13 Attack scenarios on the temperature sensor data

Attacks on the actuator data (AA1, AA2, AA3) are implemented and the results are demonstrated in Figure 4.14. It can be observed that attacks have compromised the control signals and sent tampered control commands to the actuators (Left part in Figure 4.14). The

tampered control commands then drive the physical process out of the safety limit (Right part in Figure 4.14).

This case study has validated that the Attack Scenario Generation Module can generate various cyber-physical attacks on different attack surfaces. Attack impact and system vulnerabilities can be extracted and analyzed from the generated attack scenarios.

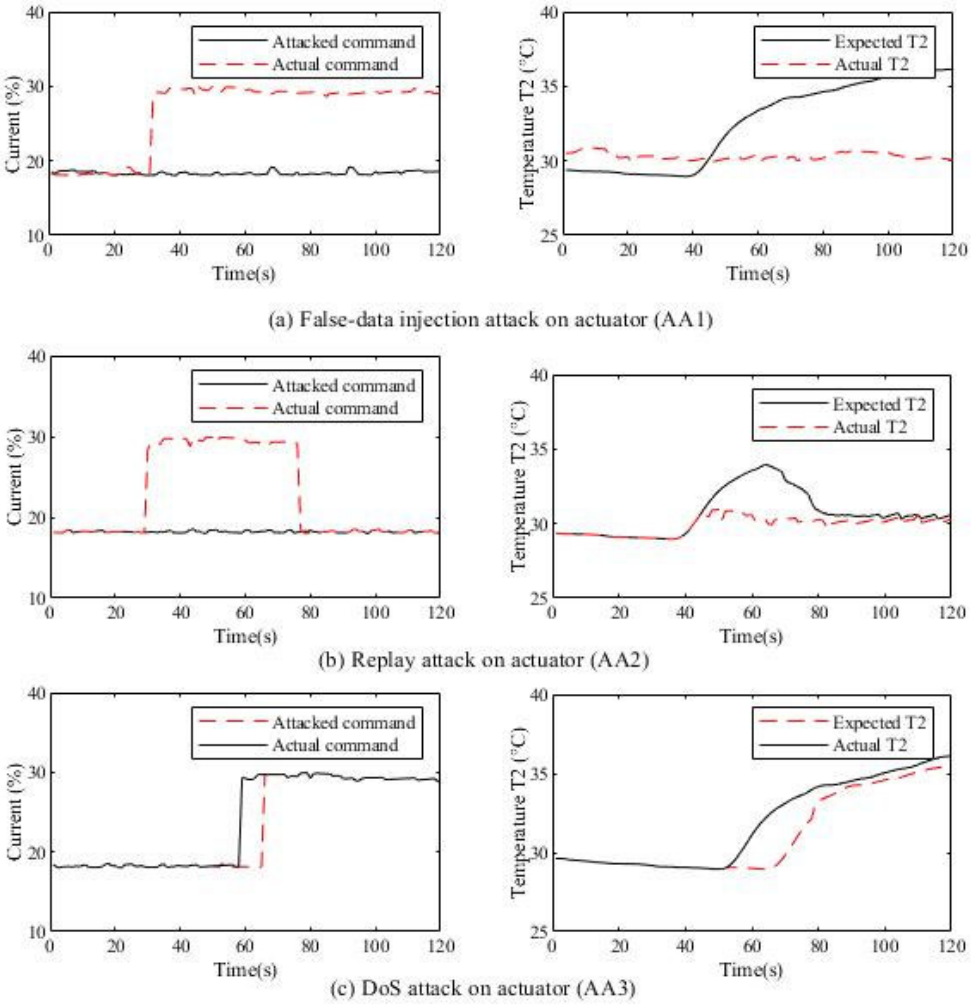


Figure 4.14 Attack scenarios on the heater actuator

4.6.2.2 Case #2: Developing and testing multi-layer detection methods

In this case, different detection methods have been configured to reveal the anomalies from cyber-physical attacks. Based on the defined security metrics for Case #2, the results of the implemented detection methods are summarized in Table 4.12.

The detection results have shown that various detection methods can be tested and evaluated on this platform. Different detection methods can be compared or integrated with respect to specific vulnerabilities. For example, the safety threshold method (D1) can only detect attacks that break system safety limit. Rule-based methods are effective for attacks that do not conform to the rules, but it does not work for stealthy attacks. CUSUM method (D3) provides fast detection speed, and physical watermarking method (D4) is more effective in detecting replay attacks. Furthermore, this platform can also be used to test the comprehensive performance of an integrated detection scheme that includes various different detection methods.

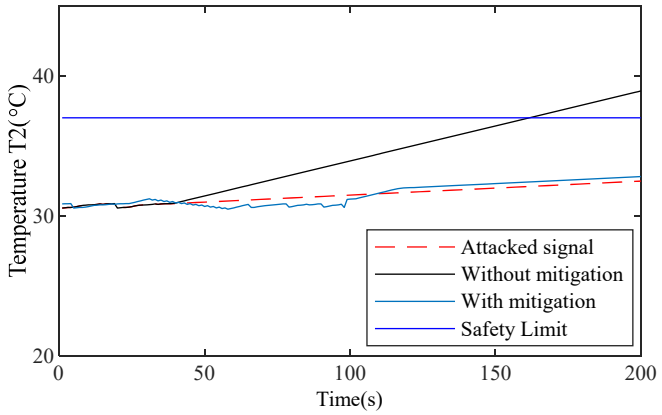
Table 4.12 Results of detection methods on the platform

Attack scenario	**Attack start time**	**Detection effectiveness**				**Detection speed**			
		D1	D2	D3	D4	D1	D2	D3	D4
SA1	t=25s	Undetected	Undetected	Detected	Detected	--	--	6s	13s
SA2	t=30s	Undetected	Undetected	Detected	Detected	--	--	4s	3s
SA3	t=30s	Undetected	Detected	Undetected	Detected	--	2s	--	8s
AA1	t=35s	Detected	Detected	Detected	Detected	97s	12s	3s	8s
AA2	t=30s	Undetected	Undetected	Detected	Detected	--	--	6s	6s
AA3	t=60s	Undetected	Detected	Detected	Detected	--	3s	10s	10s

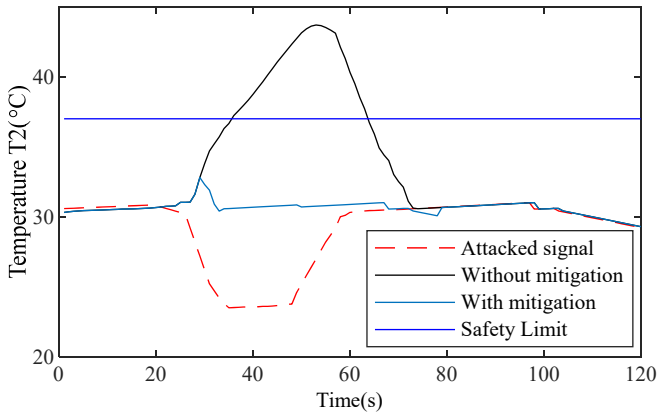
4.6.2.3 Case #3: Evaluating the performance of mitigation strategies

In order to mitigate the effects of attacks, a multi-layer attack-resilient control system is implemented within the Security Enhancement Module. In this case, a false-data injection attack (SA1) and a replay attack (SA2) are implemented on the system, respectively. Detection methods D2, D3 and D4 work together to detect anomalies, decision-making logic M2 is used

to determine which controller should be switched. It is designed that the original P controller is operating when there is no attack detected. When an anomaly alarm is triggered, the controller M3 will be switched to reduce the impacts of attacks. The experimental results are shown in Figure 4.15.



(a) Attack mitigation under SA1



(b) Attack mitigation under SA2

Figure 4.15 Attack mitigation against attacks

The results have shown that the control system design incorporates the cyber-physical security. When the anomaly is detected, the decision-making logic will respond to these anomalies and select the corresponding control algorithm. When the false-data attack deviates the measurement of T_2, CUSUM method (D3) detects the deviation and trigger an attack alarm.

The corresponding controller will respond and compensate the deviation to maintain the system variables within the normal operating range. Hence, the effect of the deviation caused by the attack is neutralized, and the measurement of T_2 will return to its desired value. When the replay attack hides the actual sensor measurement T_2 by a recorded data, physical watermarking detection (D4) reveals the anomaly. The decision-making unit will use the setpoint of T_2 to substitute $\tilde{T}_2$, and the controller will bring the heater temperature back to its setpoint.

Implementation of the presented security platform demonstrates that the proposed design method for a security platform can be easily applied to a specific cyber-physical environment. The case studies have shown that the proposed platform is effective to perform various security tests. The modularized design makes the security tests flexible. The platform provides a cross-layer Security Enhancement Module, which could take security in cyber and cyber-physical layer into consideration during controller design.

4.7 Conclusions

This chapter provides a design guideline for an experimental security platform, and proposes a modular approach to design and implement such a platform for security study of cyber-physical systems. The developed platform consists of three functional modules: (1) Attack Scenario Generation Module, (2) Security Enhancement Module, and (3) Security Evaluation Module. The first module can be used to mimic attack scenarios to expose potential system vulnerabilities. The second module supports various strategies to detect, prevent, and mitigate potential attacks. Finally, the third module creates a multi-layer systematic environment to analyze and evaluate the identified security issues. The platform also consists of a Platform Management Module to manage the three functional modules and monitor the test in process.

To demonstrate the effectiveness of the proposed systems and techniques, a specific prototype platform has been designed and implemented by using a physical component based dynamic system simulator, known as nuclear process control test facility (NPCTF). Case studies have been carried out on this platform to demonstrate the features and feasibilities of the proposed

platform. Different security scenarios have been implemented and their effects have been evaluated to study the effectiveness of the three functional modules. Experimental results have validated this modular design approach and demonstrated that the platform can be an effective tool to analyze vulnerabilities, and to evaluate the effectiveness of different security enhanced strategies for cyber-physical systems. Test results can also provide insights to security strengthening strategies.

Chapter 5

5 Analysis and Formulation of Insider Attacks through Data Tampering

5.1 Introduction

To ensure that insider attacks do not cause major disruptions to cyber-physical systems, it is critical to understand how the system is impacted by an attacker and how to detect these attacks. This chapter focuses on analysis and modeling of insider attacks through data tampering, to be more precise, attacks that may try to disable or tamper with sensor measurements or control signals during transmission process.

In this chapter, a method to analyze and characterize the features of insider attacks is proposed. Firstly, the model of a cyber-physical system subject to insider attacks is analyzed in the framework of a cyber-physical system. Then, an attack pattern is captured in terms of attack goals, resources, constraints, modes, as well as potential attack paths. Next, conditions to achieve stealthy attacks are analyzed. Attack process is analyzed based on these attributes in the attack pattern. potential impacts of such attacks on the system behavior are analyzed using an attack tree. Finally, case studies are carried out to demonstrate effectiveness of the proposed work.

5.2 System analysis

5.2.1 Cyber-physical systems

A cyber-physical control system can be conceptually divided into four main parts as shown in Figure 5.1: a physical process, a communication network, a controller, sensors and actuators. An anomaly detection scheme can also be introduced to such a system.

In the control loop, the sensor measurements and control commands are transmitted through a cyber-enabled communication network. The sensor and actuator signals on the physical side

can be represented by $y(k) \in R^p$ and $\tilde{u}(k) \in R^m$, respectively. The sensor data and the control commands at the cyber side are denoted as $\tilde{y}(k) \in R^p$ and $u(k) \in R^m$, respectively.

Anomaly detection scheme uses the observed sensor data $\tilde{y}(k)$ and the control commands $u(k)$ in the cyber side to detect any anomalies based on the normal operations [156].

The nominal system behavior under normal operations can be defined as $\tilde{u}(k) = u(k)$ and $y(k) = \tilde{y}(k)$.

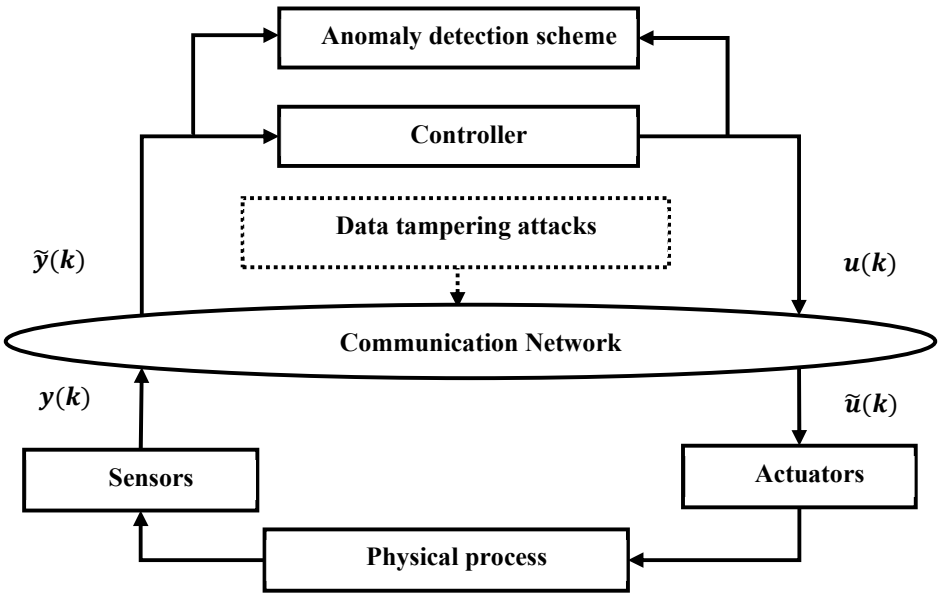


Figure 5.1 A cyber-physical system with an anomaly detection scheme

5.2.1.1 Physical process model

Assume that the model of the physical process can be represented as:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) \end{aligned} \tag{5.1}$$

where $x(k) \in R^n$ are the system state variables, $u(k) \in R^m$ is the control command applied to the process, $y(k) \in R^p$ is the output of the system, A, B, and C are the system matrices of

appropriate dimensions, and $k \in \{0,1,\cdots,N\}$ denotes the discrete-time index, taking values from the time horizon [0, N].

From Equation (5.1), the output of the system can also be derived as:

$$\begin{aligned} y(k+1) &= Cx(k+1) \\ &= CAx(k) + CBu(k) \end{aligned} \tag{5.2}$$

From a security point of view, Equation (5.2) links potential anomalies in the control commands $\tilde{u}(k)$ to the observation of $y(k+1)$. Therefore, it is possible to detect anomalies from the sensor measurements through proper data processing.

5.2.1.2 Anomaly detection scheme

An anomaly detection scheme is used to monitor the system behavior and detect possible anomalies. Sensor measurements $\tilde{y}(k)$ and control commands $u(k)$ are collected in an anomaly detection scheme.

Given sensor data $\tilde{y}(k)$ and control commands $u(k)$ at time k, the system state at time (k+1) can be estimated as

$$\hat{x}(k+1) = L_1(\hat{x}(k), u(k), \tilde{y}(k+1)) \tag{5.3}$$

where $L_1(.)$ is a state estimator of the system.

The sensor output of the system at time (k+1) can be predicted based on model of the physical process.

$$\hat{y}(k+1) = CA\hat{x}(k) + CBu(k) \tag{5.4}$$

If the system has been attacked, the attack detection scheme will compare the compromised data $\tilde{y}(k)$ with the estimated output $\hat{y}(k)$. The difference is known as a residual and then can be used to detect existing anomalies.

The residual is defined as:

$$r(k) := \tilde{y}(k) - \hat{y}(k) \tag{5.5}$$

The detection scheme is defined as:

$$Z = f(r(k)) \tag{5.6}$$

where $f(.)$ is a detection algorithm.

Attack detection decision rule can be defined by testing the following hypothesis:

$$Detection = \begin{Bmatrix} H_0 \ \text{(No Attack)} & Z \le \tau \\ H_1 \ \text{(Under Attack)} & Z > \tau \end{Bmatrix} \tag{5.7}$$

where $\tau > 0$ is a pre-selected detection threshold. If the deviation exceeds the detection threshold, H_1 is accepted and the detection scheme will arise an anomaly alarm, otherwise under the hypothesis H_0, it means no anomaly has been detected.

5.2.2 Cyber-physical systems under insider attacks

Let an insider attack $a(k)$ represent the attack at time *k*, the system input and output under this attack can now be characterized as $\tilde{y}(k)$ or $\tilde{u}(k)$. The system model now becomes:

$$\begin{aligned} x(k+1) &= Ax(k) + B\tilde{u}(k) \\ y(k) &= Cx(k) \end{aligned} \tag{5.8}$$

If the attacks are only on control commands sent to the actuators, then $\tilde{u}(k) \neq u(k)$, $\tilde{y}(k) = y(k)$. The attack offset $a_u(k)$ becomes:

$$a_u(k) = \tilde{u}(k) - u(k) \tag{5.9}$$

If the attacks are only on sensor measurements sent to the controllers, $\tilde{u}(k) = u(k)$, $\tilde{y}(k) \neq y(k)$. The attack offset $a_y(k)$ becomes

$$a_y(k) = \tilde{y}(k) - y(k) \tag{5.10}$$

If the attacks are both on control commands sent to actuators and sensor measurements, $\tilde{u}(k) \neq u(k)$, $\tilde{y}(k) \neq y(k)$. The attack offset becomes a vector $\boldsymbol{a}(k)$ as follows:

$$\boldsymbol{a}(k) = \begin{bmatrix} a_u(k) \\ a_y(k) \end{bmatrix} = \begin{bmatrix} \tilde{u}(k) - u(k) \\ \tilde{y}(k) - y(k) \end{bmatrix} \tag{5.11}$$

5.3 Formulation of insider attacks

In this section, a general analytical formulation of insider attacks is presented and described firstly, and then three specific attack scenarios are formulated case by case.

5.3.1 Formulation of an attack pattern

An attack pattern describes attack features associated with performing a particular type of attack [157]. Attack patterns represent a set of undesirable and unexpected operational behaviors. In this section, an attack pattern is defined as a representation used to model different insider attack scenarios. Based on the identified vulnerabilities in Chapter 3, each attack pattern contains six attributes, based on system vulnerabilities, which is defined as a tuple in **Definition 5.1**.

Definition 5.1 (Insider attack pattern): An insider attack pattern AP is defined as a tuple with six attributes.

$$AP = \{G_s, R_s, C_s, M, P, I\} \tag{5.12}$$

where G_s is the attack goal.

R_s represents the accessible disruption resources related to the attack, which may affect the integrity of the system components.

C_s represents the conditions used to keep the insider attack stealthy.

M is the attack scenario, which the attacker may take to achieve the goals.

P represents the entry points, attack steps and attack paths for a successful attack.

I represents attack impacts on the system.

The following sub-sections contain detailed descriptions for the above six attributes in an insider attack pattern.

5.3.1.1 Attack goals

The attack goals might be the penetration process, or a set of exploitation of system vulnerabilities, or impacts on the behaviors of the systems. In this dissertation, the attack goal is to drive safety-critical variables out of their safety boundary and cause dangerous impacts in physical process while keeping stealthy.

5.3.1.2 Attack resources

Attack resources include knowledge of system model, interactions among various sub-systems, and those resources that the attacker may possess relating to targeted components in the system. Attackers can compromise CPS information with specific objectives. It is assumed herein that the adversary: (1) has knowledge of the system dynamics, (2) the information of the control and/ or anomaly detection strategies; and (3) aims to conduct a malicious action that will compromise system if not being detected in time.

5.3.1.3 Stealthy conditions

Stealth conditions can be viewed as constraints from an attacker point of view. For the linear system described in Equation (5.1), the attack is stealthy if there is no anomaly alarm detected by the anomaly detection scheme during an attack.

In this dissertation, it is assumed that the attack is launched without violating the constraints of stealthy conditions. Stealthy conditions will further be analyzed in detail in Section 5.4.1.

5.3.1.4 Attack methods

Attack methods refer to the ways that an adversary may take to carry out an attack. Formulation Three insider attack methods have been considered in this chapter: deception attack, false-data injection attack, and replay attack. These attack methods have been formulated in Section 5.3.2.

5.3.1.5 Attack paths and attack steps

Attack paths and steps describe how an attack is to be carried out to achieve its intended goals. In this chapter, a comprehensive attack tree that integrates attack scenarios and vulnerabilities of the system is developed in Section 5.4.2 to plan out the possible attack paths and identify the corresponding attack steps from the cyber domain to the physical processes.

5.3.1.6 Attack impacts

Attack impacts are analyzed in an attack tree in Section 5.4.2 as well.

5.3.2 Formulation of insider attacks

In order to characterize the features of insider attacks from a system point of view, this section discusses two types of insider attack strategies and their related mathematical models. The first is a deception attack and a false-data injection attack. The objective is to mislead the anomaly detection mechanism and inject false data stealthily. The second is a replay attack, which hides its malicious attack action by replaying a healthy historical data sequence in the system.

5.3.2.1 Deception attack

During a deception attack, sensor measurements $y(k)$ and control commands $u(k)$ are tampered to $\tilde{u}(k)$ and $\tilde{y}(k)$, respectively.

Considering the attacker's access to the communication channels in cyber layers, a deception attacks can be modeled as:

$$\begin{aligned} \tilde{u}(k) &= u(k) + \Gamma^u a_u(k) \\ \tilde{y}(k) &= y(k) + \Gamma^y a_y(k) \\ a(k) &= \begin{bmatrix} \Gamma^u a_u(k) & \Gamma^y a_y(k) \end{bmatrix} \end{aligned} \tag{5.13}$$

where $a_u(k)$ and $a_y(k)$ represent the attack signals to the corresponding sensor and control channels, $\Gamma^u \in \{0,1\}$ and $\Gamma^y \in \{0,1\}$ are the binary index matrix that indicate the connectivity status between the attack signals and the corresponding communication channels.

5.3.2.2 False-data injection attack

A false-data injection attack can manipulate the state estimator and insert certain signals into an unknown subset of sensors and actuators without being detected. A false-data injection attack on control signals can be modeled as:

$$\begin{aligned} \tilde{u}(k) &= u(k) + \Gamma^u a_u(k) \\ a(k) &= \Gamma^u a_u(k) \end{aligned} \tag{5.14}$$

where $\Gamma^u a_u(k)$ is the attack signal injected by the insider attacker to the control channel; and $\Gamma^u \in \{0,1\}$ represents the binary incidence matrix mapping the data corruption to the respective data channels.

A false-data injection attack on a subset of sensor nodes can be modeled as:

$$\begin{aligned} \tilde{y}(k) &= y(k) + \Gamma^y a_y(k) \\ a(k) &= \Gamma^y a_y(k) \end{aligned} \tag{5.15}$$

where $\Gamma^y a_y(k)$ is the tampered signal sent by the inside attacker to the sensor measurement

channel; and $\Gamma^y \in \{0,1\}$ is a binary index matrix that indicate the connectivity status between the attack signals and the corresponding communication channels.

5.3.2.3 Replay attack

In replay attacks, the adversary first records a sequence of historical data, then replays the recorded data to hide his or her malicious actions. A diagram of this replay attack is described in Figure 5.2. Replay attack mode can be described in two steps:

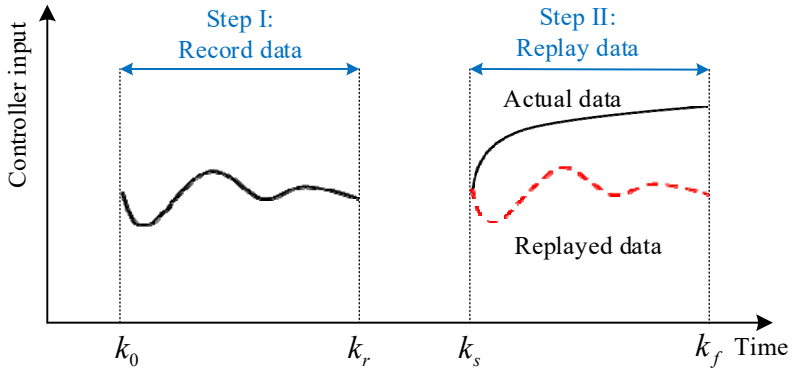


Figure 5.2 Diagram of a replay attack

Let $I_y(k)$ be the set of sensor data to the controller at time k, $I(k_0,k_r)$ represents a set of sensor data from k_0 to k_r.

Step I: The attacker records a sequence of sensor measurements $I(k_0,k_r)$ from k_0 to k_r, there is no action at this stage.

$$\text{Step I: } \begin{cases} a(k)=0 \\ I(k)=\begin{bmatrix} \Gamma^u u(k) \\ \Gamma^y y(k) \end{bmatrix} \end{cases} \tag{5.16}$$

where $k_0 < k < k_r$ and $I(0,k_0)$ is an empty set before time k_0.

Step II: Starting at time $k_s > k_r + 1$, the attacker modifies the sensor signals to the controllers with the recorded historical data, and inject attacks to control commands.

$$\text{Step II: } \begin{cases} a(k) = \begin{bmatrix} \Gamma^u a_u(k) \\ \Gamma^y(y(k_{k-T})) - y(k) \end{bmatrix} \\ I(k) = I(k-T) \end{cases} \tag{5.17}$$

where $T = k_s - k_0,\ k_s < k < k_s + k_r - k_0$.

Meanwhile, starting at time k_s, the attacker might inject a control input $\Gamma^u b^u(k)$ to achieve their malicious objective, all measurement data during the attack interval may not be available to the anomaly detection scheme, which helps the attacker to keep stealthy by designing the attack $\Gamma^u a_u(k)$ to achieve the malicious goal.

Given the attack pattern and the corresponding attack methods from different insider attacks, it is necessary to analyze the stealthy conditions and impacts on the CPS, and to capture their features from a control and system point of view.

5.4 Analysis of insider attacks

There are two aspects that need to be considered from a control system point of view when analyzing the impact of an insider attack. One is its stealthy conditions, which can show limit of the attack, and provide clues for improving the resilience of the control system. Another is its impact on the system, which includes system performance degradation under insider attacks as well as the corresponding vulnerabilities related to the attack.

5.4.1 Analysis of stealthy conditions

5.4.1.1 Stealthy condition with attack process

To analyze the stealthy conditions of the attack, attack processes need to be analyzed firstly. An attack can be divided into several stages from having access to the entry points to achieving the attack goals. The process can be summarized in Figure 5.3.

Along with the attack process, stealthy attacks can be composed of three preceding phases: (1) stealthiness at communication, (2) stealthiness at execution, (3) stealthiness at propagation [158].

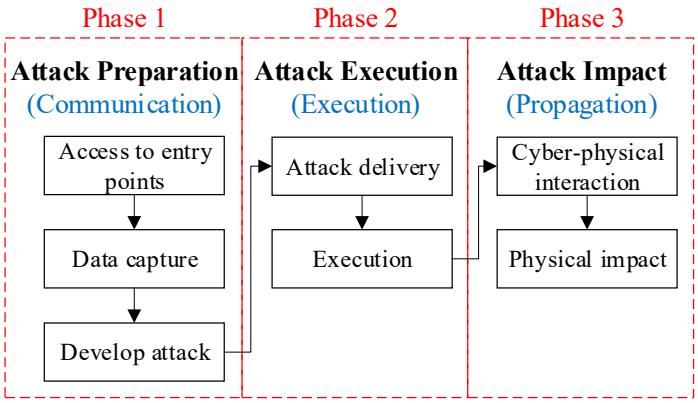


Figure 5.3 Stealthiness in an attack process

For the insider attacks considered in this dissertation, the stealthy condition in the Phase 1 is to get access to the targeted communication channel and compromise the data while not being detected by the network intrusion detection schemes and firewalls. In the Phase 2, the attacker delivers the attack (tampering data) from the cyber space to the system. The stealthy condition for this phase is that the attack cannot trigger any alarms before the attack goal is achieved. The last phase of stealthiness means that the attack has achieved the attack goal successfully before being detected.

5.4.1.2 Stealthy conditions

Given an adversary has the access to the system and knowledge of the network and can inject false sensor readings and manipulate the state variables. Under this assumption, the adversary has already achieved stealthiness in the phases of communication and attack execution. It is necessary to analyze the stealthiness of the attack propagation phase.

There are two types of stealthy conditions at the propagation stage: (1) spatial stealthy conditions and (2) temporal stealthy conditions.

One simple cyber-physical attack is to add a nonzero attack vector $a(k)$ to the original sensor measurements vector $y(k)$. The observed sensor measurement $\tilde{y}(k)$ can be manipulated as $\tilde{y}(k) = y(k) + a(k)$. If the anomaly detection scheme in the system is based on the norm of the

residual, the tampered measurements $\tilde{y}(k)$ would be indistinguishable from the nominal values $y(k)$. As a result, the attack would be undetectable.

For both the deception attack and false-data injection attack, an attack will only be detected when the residual of the anomaly alarm system exceeds the detection threshold. Therefore, stealthiness of the attacks is dependent of the detection schemes employed in the system.

5.4.1.2.1 Spatial stealthy attack

A spatial stealthy attack usually takes advantage the coupling relationship among the physical process variables to bypass the anomaly detection schemes.

Let $\hat{y}_a(k) = C(\hat{x}(k) + \delta)$ be the output estimation when an attack is under way, δ is the error of the state estimation caused by the attack.

The residual under attacks can be calculated as

$$\begin{aligned} r_a(k) &= \tilde{y}(k) - \hat{y}_a(k) \\ &= y(k) + a(k) - C(\hat{x}(k) + \delta) \\ &= y(k) - C\hat{x}(k) + (a(k) - C\delta) \\ &= y(k) - \hat{y}(k) + (a(k) - C\delta) \\ &= r(k) + (a(k) - C\delta) \end{aligned} \tag{5.18}$$

When the attack is designed as $a(k) = C\delta$, $r_a(k) = r(k)$, the residual under the attack $r_a(k)$ is the same as the residual $r(k)$ under a normal operation. This means that the attack is camouflaged in the system measurements. The attack will not cause any deviations in the system estimation and therefore will not be detected by the anomaly detection scheme.

5.4.1.2.2 Temporal stealthy attack

A temporal stealthy attack usually hides its deviation to make the observed output adopts to the dynamics of the system [159]. Stuxnet replay attack is an example of such an attack.

Given that the attacker has the knowledge of the model, the attacker can simulate the system output based on this model, and then use the simulated output to deceive the detection scheme. Since the simulated output and sending input satisfy the control law, the residual will always be zero. Consequently, arbitrary data can be injected into the system without affecting residuals.

A temporal stealthy attack can be described as

$$\begin{aligned} x(k+1) &= Ax(k) + B(u(k) + a_u(k)) \\ y(k) &= Cx(k) - a_y(k) \end{aligned} \tag{5.19}$$

where $a_y(k)$ is given by

$$\begin{aligned} x_a(k+1) &= Ax(k) + Bu_a(k) \\ a_y(k) &= Cx_a(k) \end{aligned} \tag{5.20}$$

where $u_a(k) = u(k) + a_u(k)$.

5.4.2 Impact analysis of insider attacks

5.4.2.1 Analysis by attack trees

The impact of an insider attack on the system can been analyzed using attack trees. Details about attack trees has been discussed in Chapter 3.

Recall that vulnerability vector of each path can be expressed as:

$$V(P(i)) = \{v(0), v(1), \cdots, G\} \tag{5.21}$$

In order to identify the vulnerabilities, all the nodes and variables need to be analyzed along each attack path.

5.4.2.2 Similarities and differences among attack patterns

The purpose to analyze the features of attacks is to analytically identify system vulnerabilities relating to the attacks from a system and control perspective. Based on the identified vulnerabilities, an effective defensive strategy can be developed.

In this section, similarities and differences of the above attack methods are summarized as an input for the design of subsequent attack defensive schemes. The above two attack methods have the same goals, namely to disrupt the integrity of the system without being detected. All of them are constrained by the safe operating boundaries. They all require knowledge of the system in terms of models, interactions, and inside attackers' resources for the targeted system. Although their attack methods are different, their final goals are the same.

In an attack tree, some vulnerabilities are used in different attack paths. These vulnerabilities can be identified and used to design anomaly detection schemes. Furthermore, common attack paths and attack steps can be established by using the collected similarities of different attack scenarios.

5.5 Case studies

5.5.1 Experimental setup

To demonstrate the proposed methodology, case studies are conducted on the security platform developed in Chapter 4. In this section, NPCTF architecture is analyzed first to identify system vulnerabilities. Then, two stealthy attack scenarios are analyzed. One is to illustrate the impact of a temporal stealthy false-data injection attack on sensors. The other is to illustrate the impact of a spatial replay attack on sensors and actuators.

5.5.1.1 System model

The selected system is the heater control loop on NPCTF, as shown in Figure 4.13. In the heater control loop, the safety limit of T_2 is set at 37℃, the residual magnitude is set to be 0.5℃, the change rate is less than 0.05℃/s.

A model of the physical system is identified as a first order transfer function:

$$T_2(k+1) = -0.8T_2(k) + 3.347C_2(k) + 2.0556T_1(k) + 1.59F_1(k) \tag{5.22}$$

The prediction of $\hat{T}_2$ is computed online based on a Kalman filter estimation method.

$$\hat{T}_2(k+1) = -0.8\hat{T}_2(k) + 3.347C_2(k) + 0.67(T_2(k) - \hat{T}_2(k)) + 2.0556T_1(k) + 1.59F_1(k) \tag{5.23}$$

The detection system is

$$Detection = \begin{cases} Accept \quad H_0 \quad if \quad \left|\hat{T}_2 - T_2\right| \le 0.5 \\ Accept \quad H_1 \quad if \quad \left|\hat{T}_2 - T_2\right| > 0.5 \end{cases} \tag{5.24}$$

5.5.1.2 Analysis of attack scenarios and stealthy condition analysis

Assume that an attacker has managed to gain access to the communication network between the sensors and the PLC via the activation switch, and subsequently modified the data packets being sent to PLC. In other words, the PLC is spoofed with the tampered temperature reading.

The attacker's goal is to tamper the sensor measurement of T_2 to drive critical system variables out of the safety limit before an alarm is triggered. Two attack scenarios are implemented to test the effectiveness of the proposed detection methods.

The first scenario is a false-data injection attack. The attacker tampers the sensor reading from T_2 to the controller by injecting a negative deviation to the actual value. The attack is described can be modeled as:

$$\begin{aligned} \tilde{T}_2(k) &= T_2(k) + a_y(k) \\ a_y(k) &= -0.03(k - 80) \end{aligned} \tag{5.25}$$

The system is at the steady operating state when this attack is injected. This negative term will deceive the controller to increase the power of the heater to regulate the system back to the steady state, herein causing the actual temperature T_2 rise to go beyond the safety limit. This attack is temporal stealthy by hiding its deviation within the safety threshold.

The second scenario is a replay attack. Before the attack is launched, the transient history $T_2^{'}(k)$ of the heater control loop is recorded when the setpoint of T_2 changes from 30℃ to 25℃. The attack is formulated as:

$$\text{Step I: } \begin{cases} \tilde{T}_2(k) = T_2(k) \\ a(k) = 0 \end{cases} \quad k \le 70 \tag{5.26}$$

$$\text{Step II: } \begin{cases} \tilde{T}_2(k) = T_2^{'}(k) \\ a(k) = \tilde{T}_2(k) - T_2^{'}(k) \end{cases} \quad 70 < k \le 200 \tag{5.27}$$

When the control loop is at the steady-state for its setpoint T_2=30℃, the actual T_2 is replaced with a recorded historical data. The replayed historical data deceives the controller to make wrong control actions and cause damage to the system. Since the historical data satisfies the safety threshold and the physical laws governing the heater, the attack will not be detected. It achieves spatial stealthy.

5.5.1.3 Analysis using an attack tree

The above two attack scenarios have been examined to generate an attack tree for the heater control system. Two attack scenarios were obtained by corrupting the measurement signal T_2 from the attack tree in Figure 5.4. Based on the experiment and attack tree, the insider attack pattern, attack path, and attack impact on the control loop are analyzed effectively.

For the false-data injection attack, the attack path is: $P(1) = \{v0, v1, v2, v3, G\}$.

For the replay attack, the attack path is: $P(2) = \{v0, v1, v7\} \cup \{v4, v5, v6, G\}$.

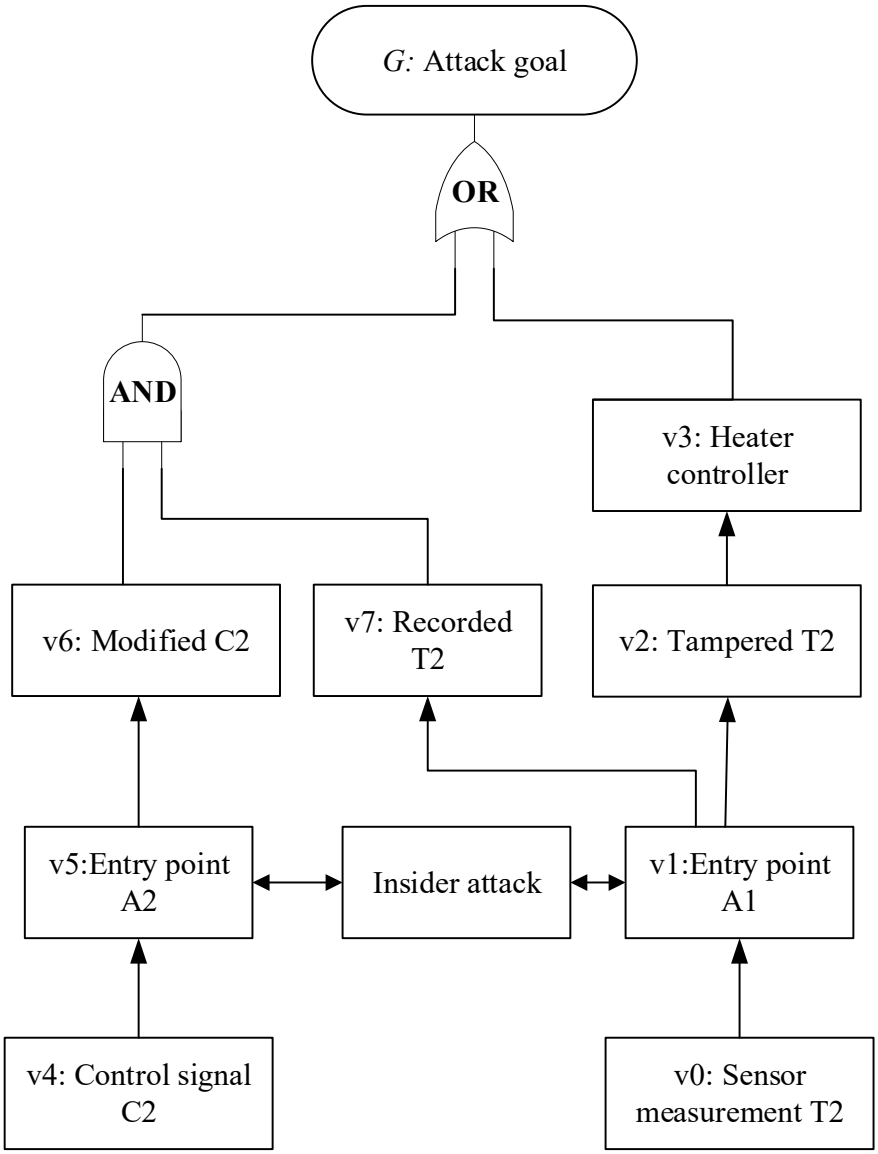


Figure 5.4 Attack tree analysis of the two attack scenarios

Based on the attack tree, attack patterns of these two attacks are analyzed to extract their similarities and differences in Section 5.5.3.

5.5.2 Experimental results

The steps of a false-data injection attack and a replay attack are further described in Table 5.1. The propagation diagram of these attacks and their impacts on the heater behavior are shown in Figure 5.5.

In the false data injection attack, the attacker tampers with the actual T_2 slightly to deceive the controller, while keeping the magnitude of the residual within the threshold 0.5℃. Results of the false data injection attack is demonstrated in Figure 5.6. It can be observed that when the actual T_2 exceeds its safety limit of 37℃ at 251s, the deceived $\tilde{T}_2$ sent to the PLC is only 27℃. The anomaly detection scheme is not triggered, and the attack goal has been achieved. As such, a sophisticated attacker can spoof the measurement data to fit the physics of the system, while still driving the critical system variable out of the safety limit unsuspiciously.

Table 5.1 Steps for mounting insider attacks on the heater control loop

Attack steps	FDI attack action	Replay attack action	Impact on the system
ATKS_1	Access to the network		Network compromised
ATKS_2	Port scanning		Network compromised
ATKS_3	Capturing communication data packets		Information eavesdropped
ATKS_4	Parsing data packets		Information compromised
ATKS_5	Creating authenticated packets from Attack Scenario Generation Module to PLC	Recording a period of data and prepare to send to PLC	Communication protocol compromised
ATKS_6	Triggering the attack		Vulnerabilities exploited
ATKS_7	Launching the attack scenario		Vulnerabilities exploited
ATKS_8	False-data injection to PLC	Replay the recorded data to PLC	Data manipulated
ATKS_9	Deceive PLC to increase T_2		PLC deceived
ATKS_10	Hiding impact of attacks		PLC deceived
ATKS_11	Safety limit exceeded		Triggering shutdown system
ATKS_12	Attack end		System disrupted

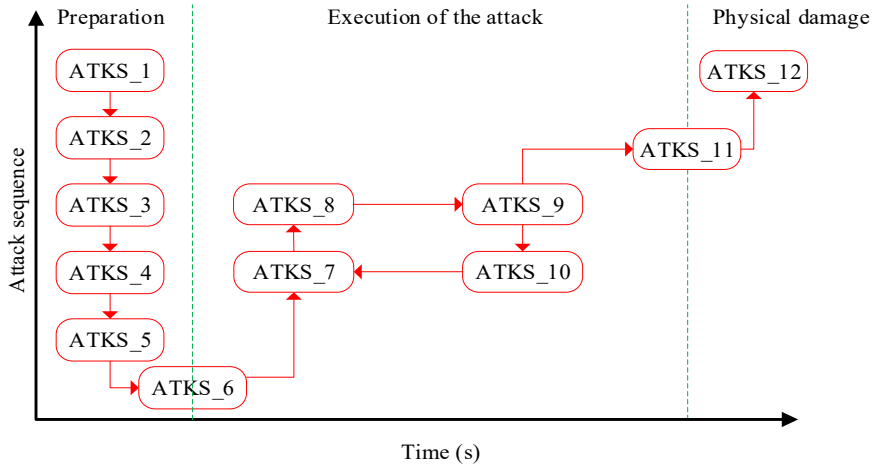


Figure 5.5 Attack process analysis

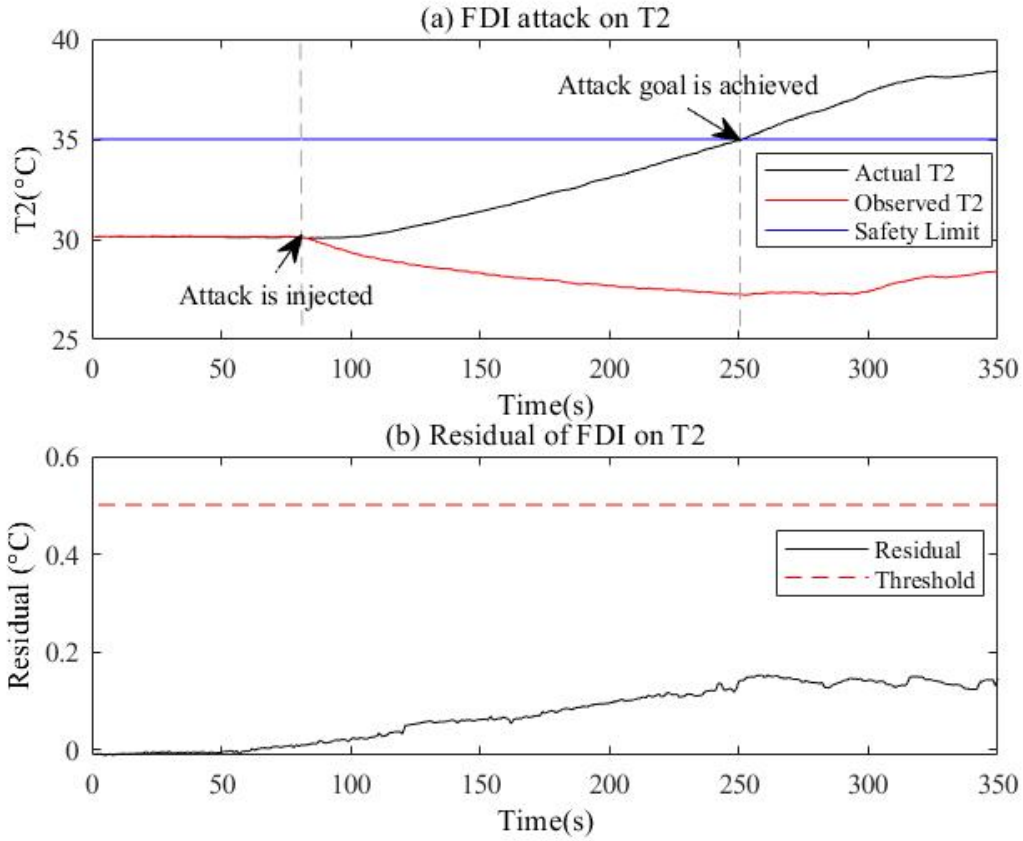


Figure 5.6 A false-data injection attack on the sensor

Results of the replay attack is shown in Figure 5.7. In order to keep the replay attack stealthy, in the early stage of the attack, the replayed data is chosen to be the actual data. When the changes start at k=78s, the controller is deceived by the attacks to make a wrong command to increase T_2. Results have shown that when the actual T_2 is beyond its limit at k=163s, the system haven't identified any anomalies. The attack has achieved its desired attack goal.

Results demonstrate that the proposed method is effective to analyze system vulnerabilities related to insider attacks and extract steps of these attacks on physical processes. These results can be used for assessing the extent of cyber-physical attacks and for designing potential attack detection and defense mechanisms.

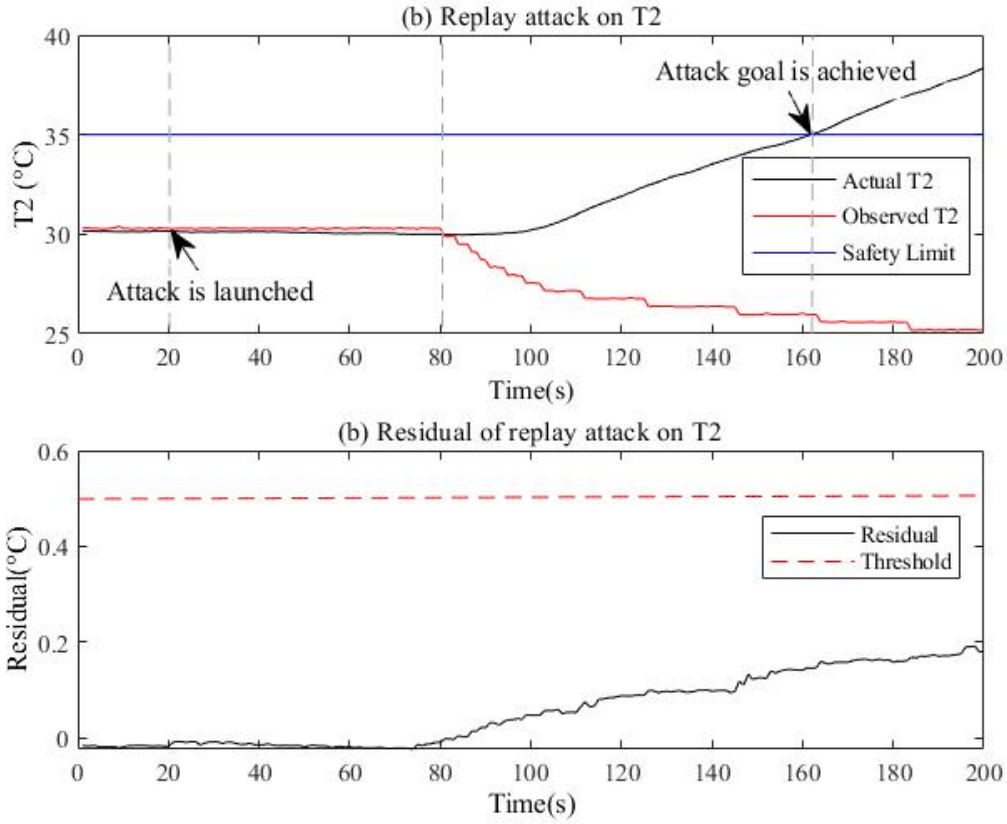


Figure 5.7 A replay attack on the sensor

5.5.3 Analysis of similarities and differences based on attack pattern

The similarities and differences for the two stealthy attack scenarios are analyzed according to the tuple of the attack pattern $AP = \{G_s, R_s, C_s, M, P\}$.

(1) Attack goals G_s

Two different scenarios share the same attack goals, i.e., drive the outlet temperature of the heater beyond the safety limit.

(2) Resources R_s

False-data injection attacks need disruption resources to obtain sensor measurements, the replay attack still needs the knowledge of the control loop playing back the recordings.

(3) Attack stealth conditions C_s

Both attacks have the same constraints for keeping their acts undetected.

(4) Attack mode M

While their attack methods are different, it should be mentioned that both the false-data injection attack path and the replay attack path use the same entry point A1 and goes to the same target *G*.

(5) Attack paths and steps P

The attack paths and steps are different for each scenario, but they have some common attack steps along different paths. These common steps would be considered as the critical steps. From the simulation results and analysis, the characteristics of insider attacks can be analyzed in both the physical and cyber domains. The attack patterns, including insider attack strategies and attack paths, can be extracted effectively using the proposed framework. The goal of the insider attack can be achieved while keeping the attack process stealthy. It can be seen from

the attack tree that one attack scenario may be executed using different attack paths, and that different attack scenarios may use the same attack path.

5.6 Conclusions

A detailed analysis and formulation of insider attacks are described in this chapter to characterize the insider attacks and to identify vulnerabilities that the insider attack could use to disrupt physical process from a cyber space. There are two main conclusions drawn from this research:

First, a formulation methodology for insider attacks can be analyzed by attack patterns, which are defined by the adversary's goal, attack mode, attack path, attack resources, and attack constraints. To understand impacts of an insider attack, features have been captured and analyzed. A generic attack pattern for insider attacks, applicable to different attack scenarios, has been modeled and analyzed to characterize the essential features of insider attacks.

Second, stealthy conditions of insider attacks are analyzed from temporal and spatial perspectives. To capture the mapping relationship between cyber-physical attacks and the resulting impacts on physical process, impact analysis can be performed using attack tree methods. Data-tampering attack scenarios including deception attack, replay attack, and false data injection attack are formulated and analyzed using the proposed framework.

With this proposed formulation methodology for insider attacks, it is possible to understand and to model insider cyber-physical attacks against CPSs, and to analyze impacts of these attacks, hence helping to strengthen the security of cyber-physical systems.

Chapter 6

6 Cross-layered Anomaly Detection of Insider Attacks

6.1 Introduction

In Chapter 5, security issues with respect to insider attacks have been discussed. Due to cyber-physical interactions and unique nature of insider attacks, an adversary is assumed to have the resources and authorized access to tamper the sensor and actuator data in a stealthy way. Traditional attack detection methods, such as firewalls, network intrusion detection methods may be ineffective to insiders who have legitimate access to the networks. To secure such systems, it is imperative is to detect and determine any anomalies caused by insider attacks before they lead to unacceptable consequences in the physical process. This chapter will investigate the situation and develop corresponding methods to detect any anomalies caused by stealthy insider attacks.

Anomaly detection schemes can be categorized into three facets according to three CPS layers: (1) the cyber layer; (2) the physical process; and (3) the cyber-physical interactions. Using the information of the cyber layer, one can create preventive measures to potential attacks. With the knowledge of the physical process, one can analyze and mitigate potential effects of attacks. By observing the cyber-physical interactions, it is possible to detect potential cyber-physical anomalies.

In this chapter, a cross-layered anomaly detection framework is developed and implemented with the focus on sensor and actuator data tampering by an insider attack. The detection scheme can identify the anomalies by analyzing the observed data in the cyber layer, transmitted data in the cyber-physical layer and process data in the physical layer.

This chapter considers anomaly detection of data tampering attacks on sensor data and control signals. Please note, since the insider attacker has authorized access to the system and has knowledge of the system, access control strategies, such as encryption, are not within the scope in this work.

The organization of this chapter is as follows:

1) Development of a general framework for cross-layered detection scheme;

2) Design two detection algorithms to recognize the anomalies both in temporal and spatial dimension.

3) Evaluation of the proposed framework and methods by performing case studies on the experimental security assessment platform.

6.2 Problem formulation

6.2.1 System model

The mathematical models of a cyber-physical system under insider attacks can be described as

$$\begin{aligned} x(k+1) &= Ax(k) + B(u(k) + a_u(k)) \\ y(k) &= Cx(k) \\ \tilde{y}(k) &= Cx(k) + a_y(k) \end{aligned} \tag{6.1}$$

The attack $a_u(k) \in \mathbb{R}^m$, $a_y(k) \in \mathbb{R}^p$ are added to actuators and sensors, $y(k)$ is the sensor signal on the physical side, $\tilde{y}(k)$ is the sensor data received on the cyber side. Note that the system matrix pairs (A, B) and (C, A) satisfy the controllability and observability conditions.

Assume that a set of attack sequences can be written as

$$\begin{aligned} a_u &:= \begin{bmatrix} a_u^T(0) & a_u^T(1) & \dots & a_u^T(k) \end{bmatrix}^T \\ a_y &:= \begin{bmatrix} a_y^T(0) & a_y^T(1) & \dots & a_y^T(k) \end{bmatrix}^T \end{aligned} \tag{6.2}$$

These attack sequences can cause some critical system variables to go out of its normal range. The objective of an anomaly detection scheme is to recognize anomalies caused by such attack sequences.

6.2.1.1 Safety set

When a system is attacked, the first task is to secure the physical process in a safe state. The safety of the physical process can be defined by a set of boundaries on the process variables. If all the variables are inside of the boundaries, it is said that the system is safe, otherwise, the system is in unsafe state. A concept of safety boundary is described in Figure 6.1.

Definition 6.1: Safety boundary

A safety boundary Φ can be defined by a set of constraints given by

$$\Phi = \{\phi_i(x(k)) \le 0 \,|\, i = 0,1,\ldots,h\} \tag{6.3}$$

where h is the number of boundaries.

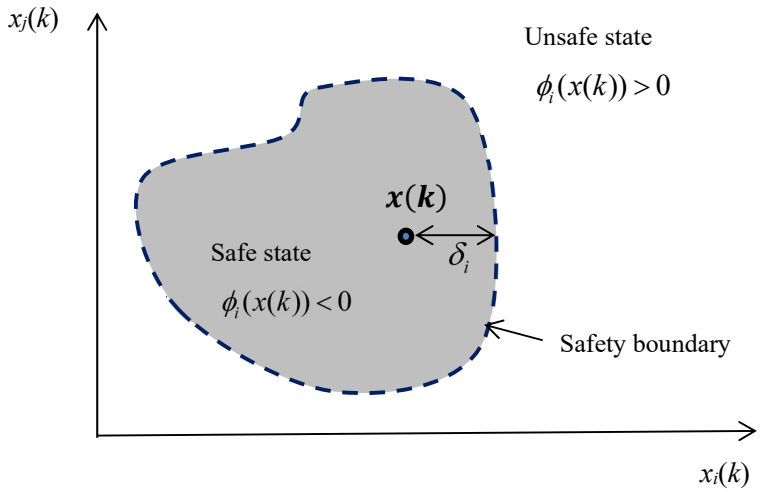


Figure 6.1 Definition of a safety boundary

Define δ_i as the distance between the system specific variables and a predefined safety boundary at the sampled instant k

$$\delta_i = sig(k)\left|x\,|_{\phi_i(x(k))=0} - x\right| \tag{6.4}$$

where $sig(k) = 1$ if $x(k)$ is inside the safety boundary, $sig(k) = -1$ if $x(k)$ is outside the safety boundary.

Definition 6.2: Residual

Let $\hat{x}_a(k)$ be the state estimate of the system when the system is under attack. Then, the dynamics of the estimator can be represented as:

$$\begin{aligned} \hat{x}_a(k+1) &= A\hat{x}_a(k) + Bu(k) + L(\tilde{y}(k+1) - CA\hat{x}_a(k) - CBu(k)) \\ \hat{y}(k+1) &= CA\hat{x}_a(k) + CBu(k) \end{aligned} \tag{6.5}$$

where $\tilde{y}(k+1)$ is the sensor data received on the cyber side, L is an observer gain.

While the system is being attacked, the residual between the compromised data and the estimated data can be defined as

$$r(k+1) := \tilde{y}(k+1) - \hat{y}(k+1) \tag{6.6}$$

Definition 6.3: Safety set

At a time instant k, a subset $S(k)$ is assigned as a safety set given by

$$S(k) = \left\{ r : \left\| r(k) \right\|_p \leq \varepsilon \right\} \tag{6.7}$$

where $\|.\|_p$ with $1 \leq p \leq \infty$ is the specified safety metrics, ε is a predefined threshold for the safety set.

For a stealthy attack, before the attacker reaches the final target, the tampered data are kept within the safety boundary to avoid being detected.

6.2.1.2 Faults and attacks

The detection problems of insider attacks in cyber-physical systems have some similarities with that of faults from sensors and actuators. Faults and attacks can both seen as threats that

will have impacts on physical processes. Therefore, many fault detection and isolation techniques can be used to detect the adverse effects of insider attacks.

However, there are some conceptual differences between faults and attacks, traditional faults detection and isolation methods cannot be directly applied to detect insider attacks. The most distinct difference is that a fault is considered as a non-colluding physical event that occurs in components of the system randomly, while an attack often happens in the cyber layer and causes effects in the physical process with an malicious intent. An fault happens in a specific component and cannot disappear before the component is repaired or replaced, while an attack can be performed in many potential points and in a coordinated and stealthy way, and it can happen and disappear according to the attack scenarios. The impact of a fault is merely on physical processes, while an attack can affect the transmitted both in cyber and physical layers. These differences motivate the need to address cross-layer detection problems in a cyber-physical system.

6.2.2 Anomaly detection problem

From the equation (5.2), it can be seen that the attack impacts will be reflected on the output sequences $y(0), y(1), \cdots, y(k)$, regardless the attack is on the sensor data or the actuator data. Therefore, it is possible to detect anomalies from the sensor measurements.

Given that the set of received output sequences at time *k* on the cyber side can be described as:

$$\tilde{Y}(k) = (\tilde{y}_1(k), \tilde{y}_2(k), \ldots, \tilde{y}_n(k)) \tag{6.8}$$

The detection problem can be stated as: to detect a possible change in $\tilde{Y}(0), \tilde{Y}(1), \ldots, \tilde{Y}(k)$ with the minimum time and to determine if there is an anomaly.

The decision of the anomaly detection scheme can be formed in a hypothesis: normal hypothesis (H_0) and attack hypothesis (H_1).

There are two aspects need to be considered for this detection problem:

1) Development of a cross-layered framework to deal with anomalies in multiple layers; and

2) Design of diverse detection algorithms to identify these anomalies and to issue an alarm.

6.3 Design of a cross-layered anomaly detection scheme

To recognize anomalies in a multi-layers CPS, it is necessary to consider data in different layers to determine the detection rules, that is, a cross-layered detection scheme. This chapter will develop a cross-layered detection framework and design corresponding detection algorithms.

6.3.1 Cross-layered detection framework

The proposed detection system utilizes a defense-in-depth concept to detect anomalies, as is shown in Figure 6.2. It integrates the cyber data, network data and process data to provide a cross-layer detection. It employs three different detection algorithms across the three layers. The first layer is a traditional rule-based intrusion detection method, to prevent external intruders and limit the resources available to inside attackers. Insiders may bypass this detection layer in some situations. The second layer is in the cyber-physical layer, a data-driven algorithm is proposed to detect the anomalies among the transmitted data. The data flow transmitted in the cyber-physical network and the inherent physical laws of the process system are integrated for detection in this layer. The third layer is in the physical layer, a model-based state estimation algorithm and a temporal-based detection algorithm are proposed to detect anomalies in the physical processes.

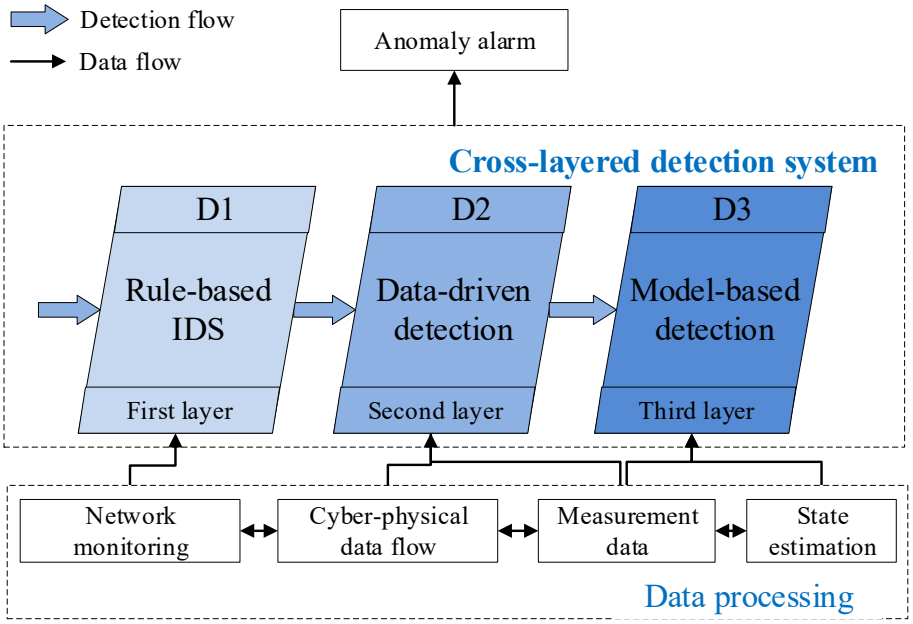


Figure 6.2 A cross-layered anomaly detection framework

A detailed flowchart of the proposed anomaly detection scheme is shown in Figure 6.3. In order to find the anomalies caused by attacks as early as possible, the scheme reads both network and process data online, performs different detection methods in multiple layer, and identifies if there are any abnormal changes to the data.

Due to the distinct differences between attacks and faults, this detection framework considers a cross-layered detection scheme, which is also different from the traditional fault detection and isolation techniques that only consider the physical layers. For a fault detection scheme, only D3 is considered, which is not sufficient nor effective to detect insider attacks.

The proposed detection scheme provides a cross-layered detection framework: D1 is focused on the network intrusion detection to provide the initial detection of attacks. D2 integrates the data both in physical layer and physical layer to identify the anomalies among the cyber-physical interactions in case that D1 fails, which is different from the traditional fault detection

methods. D3 is originated from a traditional method, which provides the anomaly detection in the physical process. In this chapter, the focus is on the development of the second and third layers in the proposed detection framework.

Please note that, although the proposed detection framework is for detection of attacks, the model-based detection system in the third layer in this chapter can also be used to detect anomalies from both attacks and faults.

6.3.2 Cross-layered detection methods

Based on the discussion in Chapter 5, insider attacks attempt to keep the attack stealthy using known information about the spatial and temporal-based knowledge of the system. To reveal the anomalies by stealthy attacks, two diverse methods are designed. One is a model-based method for temporal anomalies, in which the accumulated residuals of the compromised measurements are calculated and evaluated. The other is data-driven method to detect spatial anomalies, in which patterns of variables are learned and analyzed.

6.3.2.1 Model-based anomaly detection

A model-based anomaly detection method can leverage measured process data to detect malicious deviations from the expected process behavior. Most of the model-based detection methods are based on residuals, and only the current state of the system is considered. In this chapter, a state estimation-based CUSUM method that takes into account both the current state and the history state is used for anomaly detection.

The basic idea of this detection scheme is demonstrated in Figure 6.4. The scheme consists of three stages: state prediction, residual generation, and anomaly detection. The state of the process is predicted based on the process model, residuals are generated based on the estimated data and the measured data, and the cumulative sum (CUSUM) algorithm is used to process the residuals and detect the anomalies.

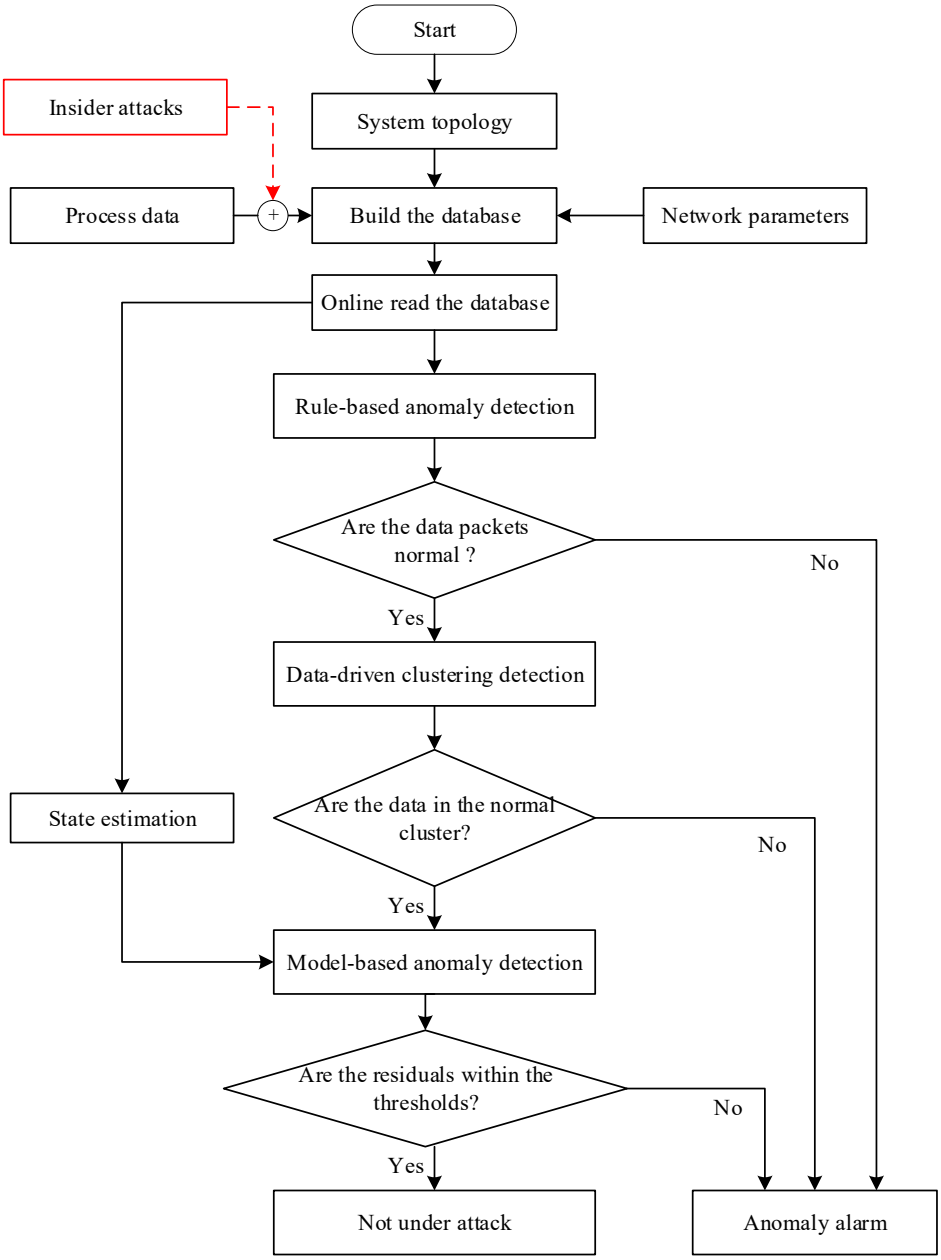


Figure 6.3 Flow chart of the proposed methodology

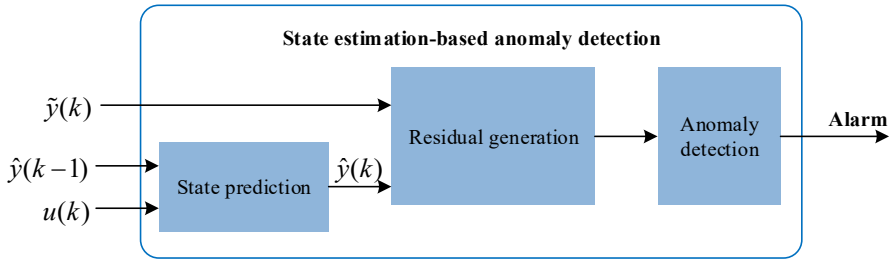


Figure 6.4 Functional elements in a model-based anomaly detection

To formalize the anomaly detection problem, the detection algorithm–a nonparametric cumulative sum (CUSUM) algorithm [162] is as follows: if the way how the output sequence of the physical system $y(k)$ reacts to the control input sequence $u(k)$ is known, any attacks by tampering the sensor data can then be potentially detected by comparing the estimated output $\hat{y}(k)$ with the measured sensor output $\tilde{y}(k)$.

1) State prediction

This step is to provide an prediction of the system. In this dissertation, a Kalman filter is used to estimate the system state and predict the system output. The Kalman filter algorithm is listed in the following.

$$\begin{aligned} &\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + LD \\ &L = P^{-}(k)C^{T}(CP^{-}(k)C^{T} + R)^{-1} \\ &D = y(k+1) - CA\hat{x}(k) - CBu(k) \\ &P^{-}(k) = AP(k-1)A^{T} + Q \\ &P(k) = (I - LC)P^{-}(k) \end{aligned} \tag{6.9}$$

where L is Kalman gain, $P(k)$ is the covariance of state vector estimate, $P^{-}(k)$ is the error covariance ahead, Q is the process noise covariance, and R is measurement noise covariance.

The output can be predicted based on the current estimation can be calculated as:

$$\hat{y}(k+1) = CA\hat{x}(k) + CBu(k) \tag{6.10}$$

This estimation method takes into account of the historical output when predicting the future output.

2) Residual computation

Define norm of the residual $\left\|\tilde{y}(k)-\hat{y}(k)\right\|$ as the detected sequence. When there are no attacks, $\left\|\tilde{y}(k)-\hat{y}(k)\right\|$ is identically to be zero.

Considering the measurement noise and error of the process model, under normal operation $\left\|\tilde{y}(k)-\hat{y}(k)\right\|$ is normally less than a predefined value.

In order to reduce false alarm rate, a positive constant b is used as an offset to compensate the measurement noises and errors of process models.

The offset residual can be represented as:

$$z(k)=\left\|\tilde{y}(k)-\hat{y}(k)\right\|-b \tag{6.11}$$

3) Computation of b

Given that $b_{\min}$ is the averaged residual computed using historical data of the physical process under the normal operation, b is selected to be larger than $b_{\min}$.

$$b_{\min}=E\left[\left\|\tilde{y}(k)-\hat{y}(k)\right\|\right] \tag{6.12}$$

4) CUSUM parameters

Based on the offset residual in Equation (6.11), a nonparametric CUSUM is calculated as

$$S(k)=(S(k-1)+z(k))^{+}, S(0)=0 \tag{6.13}$$

where $S(k)$ is cumulative sum of the offset residuals, $(S(k-1)+z(k))^+$ is the max of $(0, S(k-1)+z(k))$.

The corresponding decision rule becomes

$$\text{Detection Logic} = \begin{cases} H_0 & S(k) \le \tau \\ H_1 & S(k) > \tau \end{cases} \tag{6.14}$$

where τ is a threshold selected based on the false alarm rate.

The observation $\tilde{y}(k)$ starts under normal operation hypothesis H_0. When the CUSUM surpasses the threshold, the detection scheme changes to hypothesis H_1 to raise an anomaly alarm.

A false alarm is when the detection scheme identifies the observed data as an anomaly, but the activity is a normal behavior.

From Equation (6.11) - (6.14), the time to detect an attack (detection time) increases as *b* increases, but false alarm rate decreases .

From Equation (6.14), the threshold τ of CUSUM method presents a trade-off between the detection time and the false alarm rate. The false alarm rate decreases as τ increases, but the detection time will increase as well.

6.3.2.2 Data-driven clustering-based detection

Although the model-based detection method can predict system outputs and detect anomalies, they require an accurate model of the physical process, which is often not available. To address this issue, a data-driven clustering-based detection method is proposed. The clustering-based detection method is based on historical dynamic behaviors. They can capture the correlations or relationships among variables and check the data consistency to identify anomalies of the system.

The clustering approach classifies measurements into several groups, each group is called a cluster. Within a cluster, the data shares similar patterns. Outliers are defined as the data that does not belong to the predefined clusters. The inputs to the clustering algorithm are the measurement data. The output of the clustering algorithm is a subset containing a specific operating state of the physical system.

A flow chart of the clustering detection scheme can be summarized in Figure 6.5. The detection algorithm [78] is described as follows.

Define N as the total number of observations of y, l as the number of clusters, m as the number of data in a cluster, $C(i)$ as cluster i, and $y(j)$ as the vector of observations at moment j.

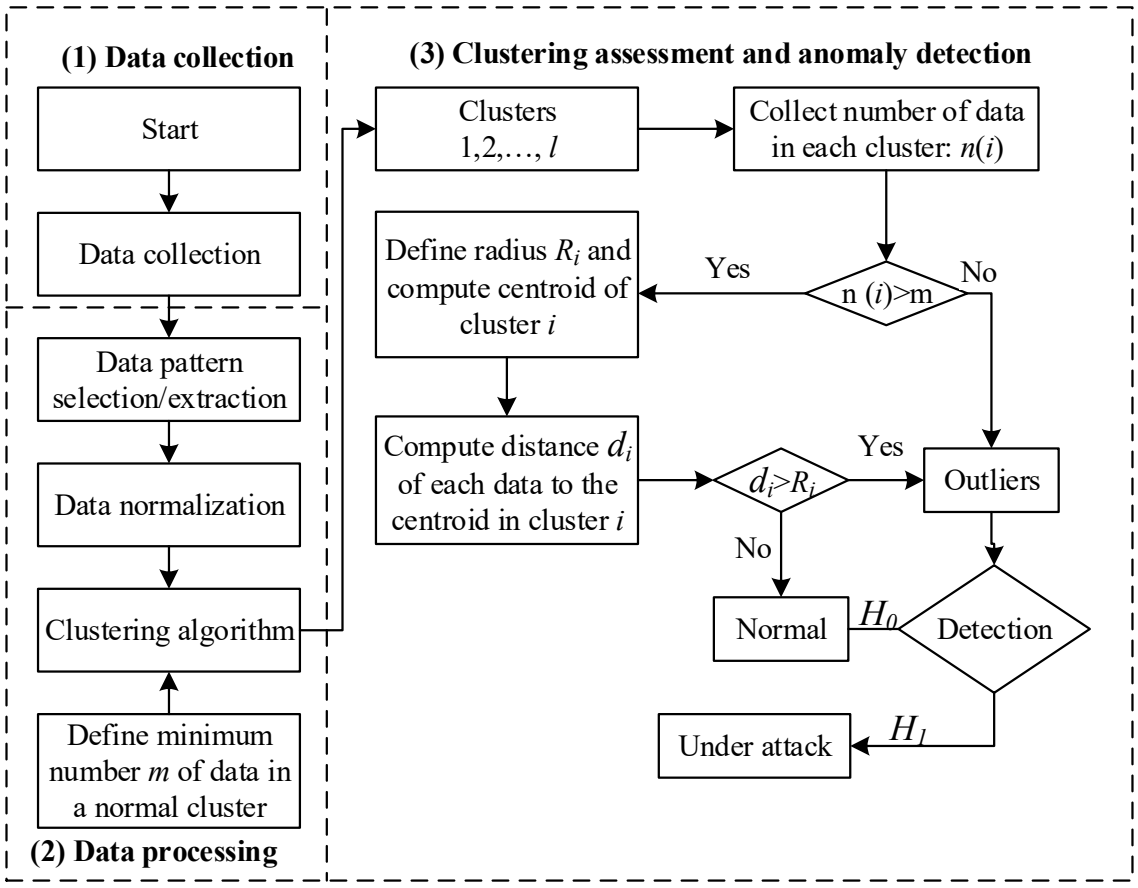


Figure 6.5 Data-driven clustering-based detection method

1) Cluster extraction

The clustering detection method collects the raw data under normal operation and trains the data for clustering. The purpose of this step is to get the parameters for the clustering.

The centroid C_i and radius R_i of a cluster is defined and computed as:

$$\begin{aligned} C_i &= \frac{1}{N}\sum_{j=1}^{N} y(j) \\ R_i &= \sqrt{\frac{1}{N}\sum_{j=1}^{N}(y(j)-C_i)^2} \end{aligned} \tag{6.15}$$

where R_i denotes the average distance of all measurements to the centroid C_i.

For a set of measurement data, *l* clusters will be generated according to the computed centroids and radius.

2) Cluster classification

In this step, the detection scheme collects the measurement data online and partitions them into *l* clusters. Each data belongs to a cluster with the nearest centroids.

Given that $n(i)$ measurements are selected as the initial cluster $C(i)$. If $n(i) > m$, the selected cluster meet the required minimum number of data in a cluster. Otherwise, it is treated as an outlier.

For each initialized cluster, the Euclidean distance between the measurement and the centroid is calculated as:

$$d_i = \left\| y(j) - C_i \right\| \tag{6.16}$$

Outliers can be classified by comparing the distance d_i with the radius R_i of the cluster.

3) Anomaly detection

For the anomaly detection, various operational states are classified into different clusters, the clustering detection algorithm evaluates the clusters and provides a binary classification as H_0 normal or H_1 abnormal. In this chapter, clusters are classified as two categories: (1) Cluster 1 for normal operating state, and (2) Cluster 2 for abnormal state.

6.4 Case studies

This section presents case studies and performance analysis of the proposed anomaly detection framework with two detection algorithms.

A security platform on NPCTF environment as implemented in Chapter 4 is used. The same heater control loop as shown in Figure 4.13 is used. The safety limit of T_2 is set to be lower than 37℃, and the change rate is set to be less than 0.05℃/second.

6.4.1 Experimental setup

The experiment scenarios are outlined in Table 6.1.

Table 6.1 Case studies for the cross-layer detection scheme

Experimental scenarios	
Attack scenarios	Stealthy false-data injection (FDI) attack on T_2
	Stealthy replay attack on T_2
Detection methods	Implementation of the cross-layered anomaly detection
	Model-based CUSUM method
	Data-driven clustering method

6.4.1.1 Attack scenarios

Two stealthy attack scenarios have been implemented to test the effectiveness of the proposed detection methods. The attacker's goal is to tamper the sensor measurement of T_2 to drive the heater system out of the safe region before an alarm is triggered.

The first scenario is to tamper the sensor data T_2 going to the controller by injecting a negative deviation ΔT_2 = -0.02(*k*-25) to the actual value at *k*=25s. The system has been operating at the steady state, when the attack is initiated. The deviation-0.02(*k*-25) will deceive the controller to increase the power of the heater to regulate the system back to the desired steady state, herein causing the rise of T_2 beyond its safety limit. This attack is temporal stealthy because it hides its deviation under the safety threshold.

The second scenario is a replay attack. Before the attack is launched, the transient response of the heater control loop subjected to a change in the setpoint of T_2 from 30℃ to 25℃ has been recorded. When the control loop operates steadily at its setpoint T_2 =30℃, the actual sensor measurement of T_2 is replaced with the recorded historical data. The controller will regulate the system according to the replayed data, which then cause the heater to increase its power and drives the actual T_2 beyond its safety limit. Since the historical data satisfies the safety threshold and the physical laws of the heater, this attack is essentially spatial stealthy.

6.4.1.2 Cross-layered detection scheme

The proposed cross-layer anomaly detection scheme is implemented aside the supervisory station of NPCTF. It collects and processes the data online through an OPC client, it monitors the networks through Snort, the sampling period is chosen to be 1 second. Rule-based method in the first layer is implemented in the Snort environment, white lists are set to prevent network intrusion from external attackers. A data-driven clustering-based detection method in the second layer and a CUSUM detection method in the third layer are implemented in MATLAB environment. The data-driven detection scheme triggers an alarm when outliers are classified as an abnormal cluster. The model-based detection scheme arises an anomaly alarm after the detection variable goes beyond a specific threshold calculated in term of CUSUM.

1) Parameters in the model-based detection algorithm

To implement the proposed detection framework in Figure 6.1, selection of the CUSUM parameters are described as follows.

The following model is used to estimate the system state by using a Kalman filter.

$$T_2(k+1) = -0.8T_2(k) + 3.347C_2(k) + 2.0556T_1(k) + 1.59F_1(k) \tag{6.17}$$

where $T_2(k)$ is the outlet temperature of the heater, $T_1(k)$ is the inlet temperature of the heater, $C_2(k)$ is the current of the heater, $F_1(k)$ is the water flow rate of the heater.

The prediction of $\hat{T}_2(k)$ is computed online based on a Kalman filter method.

$$\begin{aligned} &\hat{T}_2(k+1) = -0.8\hat{T}_2(k) + 3.347C_2(k) + L(T_2(k) - \hat{T}_2(k)) + 2.0556T_1(k) + 1.59F_1(k) \\ &L = P^-(k)(P^-(k) + R)^{-1} \\ &P^-(k) = 0.64P(k-1) + Q \\ &P(k) = (1+L)P^-(k) \end{aligned} \tag{6.18}$$

where $P(0)=0$, $Q = 4\times10^{-4}$, and $R=0.25$.

For the parameters in CUSUM method, in order to select a suitable value for b, the residual $\left\|\hat{T}_2(k) - T_2(k)\right\|$ between the estimation $\hat{T}_2(k)$ and sensor measurement $T_2(k)$ is computed based on 24 hours of historical data under the normal operation, the empirical value of b_{min} is computed to be 0.316℃, b is chosen to be 0.5 in the case studies.

The threshold τ of CUSUM method is selected as 0.5℃.

2) Data-driven clustering-based detection algorithm

The proposed clustering-based detection method is implemented with N=200, m=30, l = 2. The steady state and transient state are trained in advance based on the historical data,

the centroid for the normal state cluster is set to be 29.85℃, and the radius is calculated to be 0.5℃. The decision threshold is selected as 1℃.

Implementation of the proposed clustering detection scheme undertakes the detection of two distinct clusters. The first cluster reflects a normal steady-state operation, while the second cluster reflects observations which may be attributed to attacks.

In order to reflect the nature of the clusters, a cluster index is defined as:

$$\text{Cluster index} = \begin{cases} 1, & \text{normal state} \\ 2, & \text{abnormal state} \end{cases}.$$

3) Selection of evaluation metrics

To evaluate the proposed methods, detection effectiveness and detection time are considered as the evaluation metrics. Detection effectiveness is evaluated if the attack is detected before it has driven the critical system variables out of the safety set. Detection time is the time that it takes to detect an anomaly caused by the attack.

6.4.2 Performance results

6.4.2.1 Detection results under a FDI attack

Results of the model-based detection method and data-driven detection method under a stealthy FDI attack are shown in Figure 6.6 and Figure 6.7, respectively.

The results in Figure 6.6 have demonstrated that the model-based detection scheme can detect the anomaly effectively. The attack is injected at k=25s, the observed T_2 is close to its estimation to keep the attack stealthy. Although the residual is still within the threshold, the cumulative sum of residuals has indicated an anomaly and arises the alarm at k =49s.

The results in Figure 6.7 have shown that the clustering-based method can also identify the abnormal cluster at k =90s. When an attack is launched at k=25s, the observed measurements

matches with the cluster of normal operation. As the deviation between the observed $\tilde{T}_2$ and the expected T_2 increases, an abnormal cluster is declared.

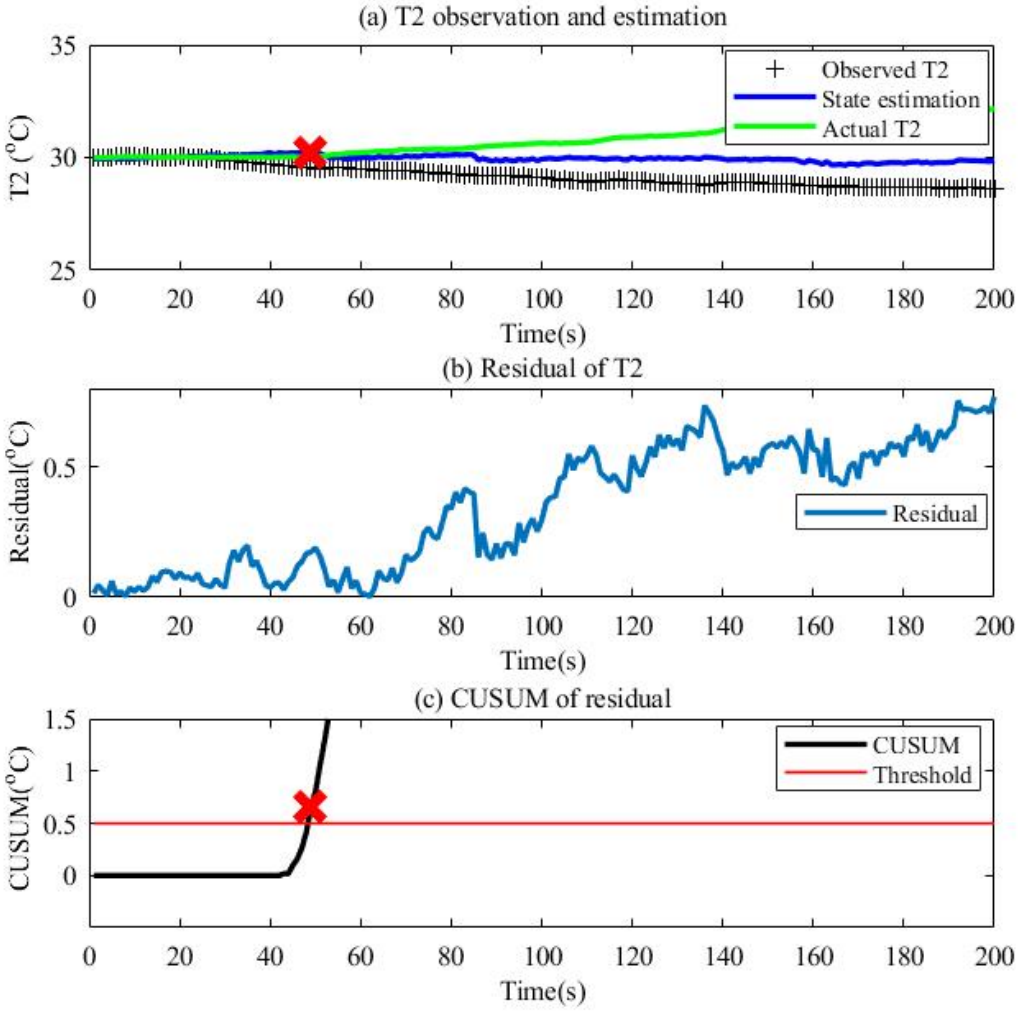


Figure 6.6 Model-based anomaly detection under a FDI attack

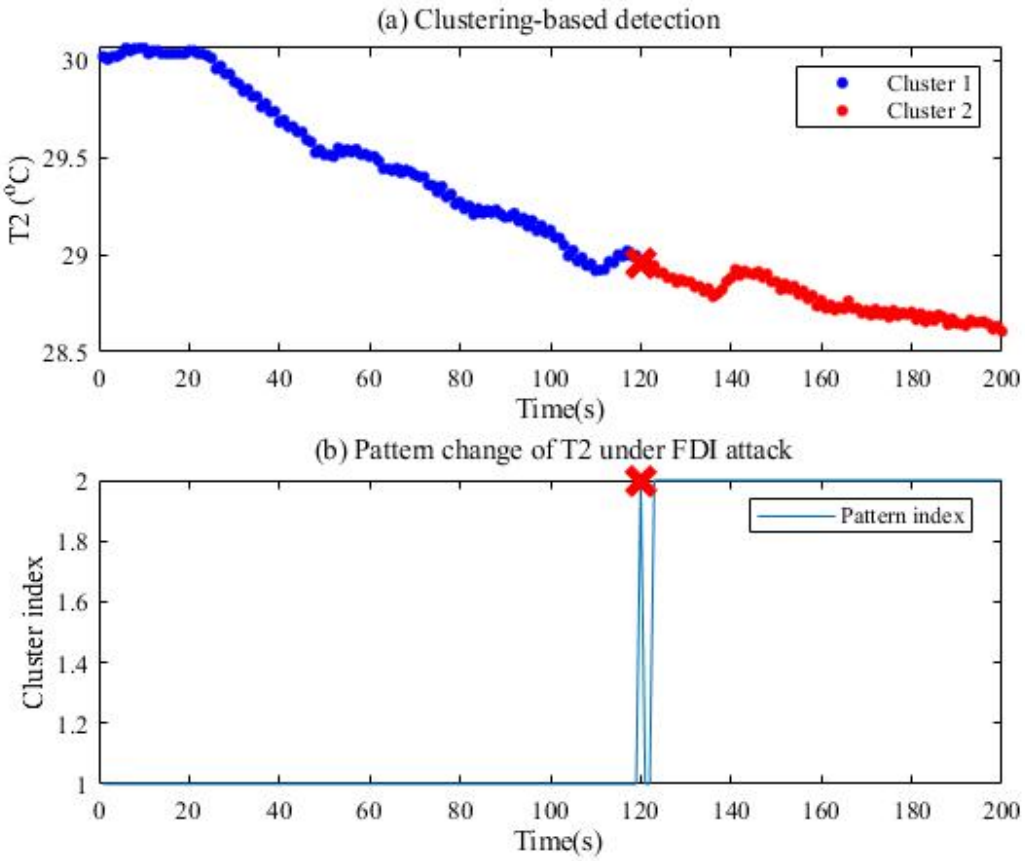


Figure 6.7 Clustering-based anomaly detection under a FDI attack

6.4.2.2 Detection results under a replay attack

Results of the model-based detection scheme under a replay attack is demonstrated in Figure 6.8. In order to be stealthy, the replayed data to the controller meets the steady operating condition at the beginning of the attack. At k =85s, the recorded data starts to change from 30℃ to 25℃, while the controller still maintains the system at its setpoint T_2=30℃. Although the replayed data $\tilde{T}_2$ is still normal, the cumulative sum of the residuals recognizes that there is an anomaly and raises the alarm at k =94s.

Results of the data-driven detection scheme under the replay attack are shown in Figure 6.9. Experimental results have demonstrated that the data-driven detection scheme identifies two

clusters. The first cluster includes both steady-state and transient dynamics, the second cluster is identified as abnormal states because the replayed data does not match the pre-defined steady state. Since the replayed data comes from the historical data in normal state, the steady state at 30℃ is classified as a normal cluster. When replayed $\tilde{T}_2$ starts to change from 30℃ to 25℃, it is still within the radius at the beginning, hence the generated cluster is still considered as a normal cluster. However, the replayed $\tilde{T}_2$ already deviates significantly from the normal state afterwards, the cluster index changes from normal to abnormal at k =120s, and the generated cluster is classified to be abnormal and the alarm is triggered.

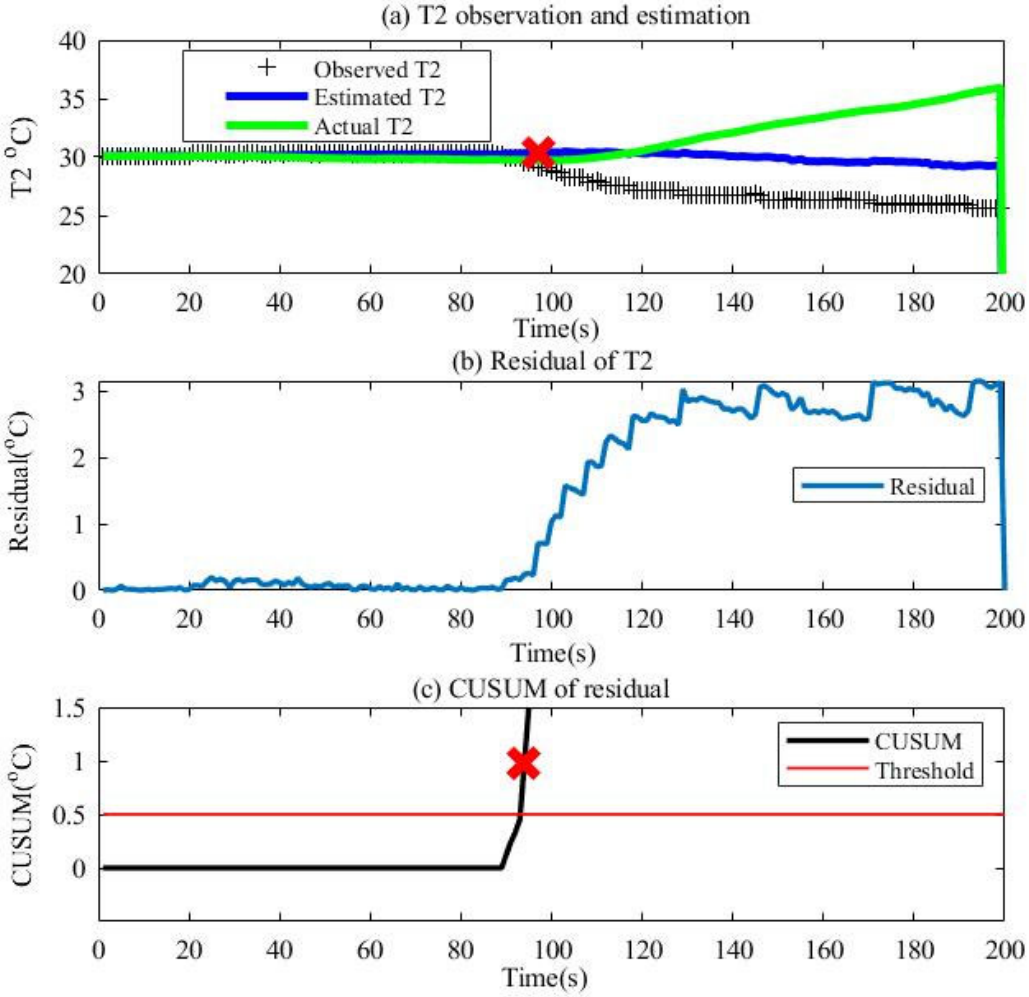


Figure 6.8 Model-based anomaly detection under a replay attack

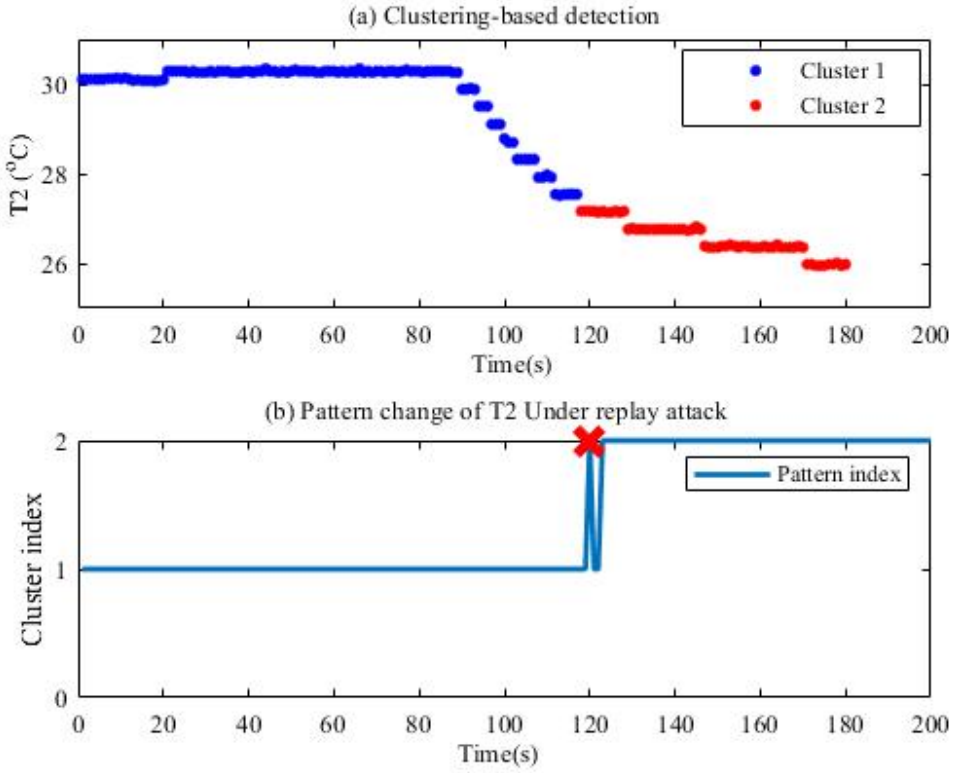


Figure 6.9 Clustering-based anomaly detection under a replay attack

6.4.2.3 Detection Effectiveness

Based on the defined security metrics, the results of the implemented detection methods are summarized in Table 6.2. The results have shown that both detection methods are effective in detecting attacks.

Table 6.2 Results of detection methods

Attack scenario	**Attack start time**	**Detection time**		**Detection effectiveness**	
		Model-based	Data-driven	Model-based	Data-driven
FDI attack	k=25s	k=49s	k=90s	Detected	Detected
Replay attack	k=75s	k=94s	k=120s	Detected	Detected

Model-based detection method arises the alarm soon after the data is tempered during the attacks. The detection time of the data-driven clustering method is longer than that of the model-based method, especially for a FDI attack. This is because that the injected FDI attack generates a very small deviation at each time step, it is difficult for the clustering method to classify the attacked data from the normal data in a short time. But for CUSUM methods, the cumulative sum of the residuals includes the historical residuals, which is more effective than clustering methods which use the current state of the system only.

In the case studies, the threshold τ of CUSUM method is conservatively selected to be 0.5℃, no false alarm is encountered during the experiments.

6.5 Conclusions

In this chapter, a cross-layered detection scheme is developed to detect anomalies caused by different stealthy insider cyber-physical attacks. The detection scheme considers a hierarchical approach by combining different detection methods in different layers to provide a defense-in-depth detection against the attacks. In the proposed detection scheme, the first layer is a rule-based detection, only the authorized users can gain access to the system. The second layer includes a data-driven clustering-based method, which is to identify anomalies from cyber-physical interactions. The third layer is a state estimation-based CUSUM method to detect the anomalies based on physical process data. These methods work can together to provide a defense-in-depth detection scheme.

Results have shown that the proposed detection scheme is effective in detecting insider attacks. The model-based CUSUM detection method can detect anomalies quickly and effectively. For situations where physical model of the system is difficult to be identified, data-driven approach can provide with adaptivity and flexibility. These different detection methods can work independently or can also be integrated to detect anomalies in multi-layers.

Chapter 7

7 An Attack Defensive Scheme against Insider Attacks

7.1 Introduction

If an insider attack is launched on a cyber-physical system, a detection scheme should identify the anomaly and alert the operators and activate attack mitigation strategies at the same time. Attack mitigation system should then respond to the attack and secure the physical system in a safe state during and after this attack. This chapter is focused on designing an attack defensive framework to provide a defense-in-depth detection, response and mitigation strategies to insider attacks.

In order to mitigate impacts of these attacks, an attack-resilience control is one that can react, tolerate and reconfigure the system [103]. It is important for a cyber-physical system to incorporate with some attack-resilient capabilities into its control systems so that the system can be maintained to be within the safe operation range.

In this chapter, an attack-resilient control system is designed to mitigate the impacts of attacks. The resilient control system includes an attack response scheme, a decision-making scheme, and a bank of controllers. The attack response scheme responds to the detected anomalies. The decision scheme enables the control system switch to a suitable controller in response to the identified attack anomalies.

It is assumed that the supervisory layer is secure and could not be penetrated by the attacker. The supervisory station is isolated from the rest of the system and is assumed to be secure. It contains a control system and an anomaly detection scheme.

This chapter is organized as follows. First, an attack defensive framework is presented in Section 7.2. Second, design of a resilient control system in response to an attack, a decision-making scheme and corresponding controllers are discussed in Section 7.3. Subsequently, case studies are included to demonstrate the effectiveness in Section 7.4 and finally conclusions are drawn in Section 7.5.

7.2 An attack defensive framework

An effective defense-in-depth strategy requires a holistic approach to leverage all of the resources in order to provide effective layers of protection against attacks [163]. In this section, an attack defensive framework is proposed to address multi-layered defense strategies as shown in Figure 7.1.

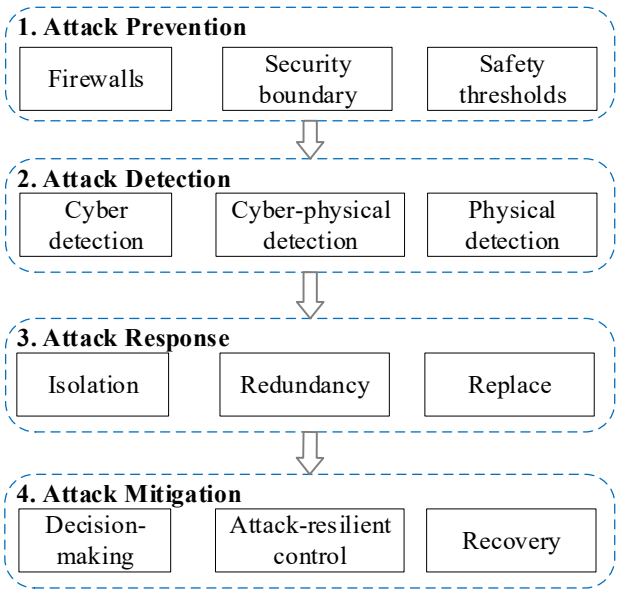


Figure 7.1 A defense-in-depth framework

There are four defense layers in the proposed framework. The first layer includes passive prevention strategies against insider attacks, the second layer consists of active detection techniques to recognize anomalies. The third layer contains responses to attacks, and the fourth layer is composed of different attack-resilient control systems to mitigate attack impacts and maintain system performance.

This attack-resilient control framework has a similar structure with a fault tolerant control framework. However, due to the different features of faults and attacks, different defensive approaches need to be considered. In general, a fault tolerant control may be achieved through safety boundary control, physical process monitoring and repair and replacement of the

physical components, respectively. On the other hand, an attack-resilient control system needs to consider both the cyber and physical enhancements and addressing multi-layer security issues. Within the proposed defensive framework, the focus includes both cyber security enhancements and physical process control.

1) **Attack prevention**

Attack prevention aims to decrease the likelihood of attacks through a combination of multiple approaches, such as security boundaries, firewalls and safety limits. These are performed offline before the system is attacked.

2) Attack detection

For attacks that are not preventable, online detection methods can be applied to identify anomalies caused by attacks. A cross-layered detection scheme has been considered in Chapter 6. A model-based detection method and a data-driven method can work together to provide a defense-in-depth detection scheme.

3) Attack response

Once an attack is detected, the corrupted measurement data needs to be isolated and replaced. In the current framework, the tampered data can be corrected through estimated measurements or through redundant measurements.

4) Attack mitigation

Once an anomaly is detected and its nature is diagnosed, mitigation strategies can be taken to reduce the attack impacts and maintain the system in a safe state. In the proposed framework, attack mitigation can be achieved through a decision logic scheme and a set of attack-resilient controllers.

It is worthwhile to mention that, in this framework, attack detection and prevention schemes do not affect normal operations of the system, only attack response and attack mitigation

strategies can reconfigure the system. For a safety-critical system, even if the system is kept in a safe state by an attack mitigation scheme, the employed attack response scheme and resilient control strategy are still considered as a temporary solution before a human operator reacts to the situation.

This chapter is focused on design of attack responses and mitigation strategies to secure the cyber-physical system in an event of an attack.

7.3 Design of an attack-resilient control system

Given the existence of an attack, one of the important requirements for safety-critical systems is to be attack resilient. Thus, design of an attack-resilient control system has two objectives. One is to isolate the corrupted data from the attacks, the other one is to reduce its attack impacts and maintain the system safety and performance at an acceptable degree.

7.3.1 Attack response scheme

Attack response is a follow-up action soon after the detection scheme arises an alarm, it means that the measurement data might have been corrupted. An attack response scheme has to evaluate the consequence of the attack and to isolate the tampered data to prevent further damages to the system.

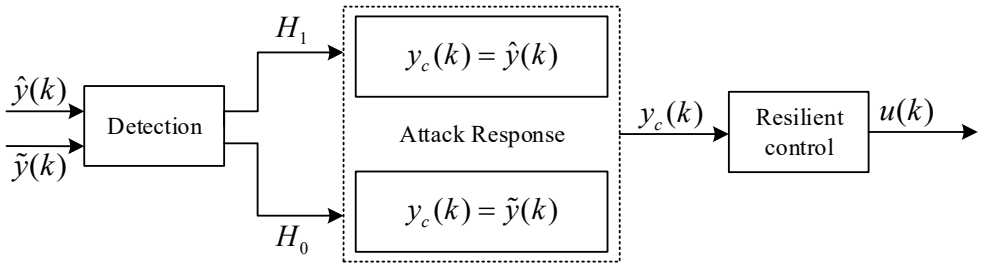


Figure 7.2 A conceptual diagram of the attack response scheme

Given that the state estimation $\hat{y}(k)$ is available, an attack response scheme is shown in Figure 7.2. It consists of two sequences of measurements, one is the estimated output $\hat{y}(k)$ and the

other is the measured signal $\tilde{y}(k)$. Once an attack is detected (H_1), the scheme will replace the measurement $\tilde{y}(k)$ by the estimated data $\hat{y}(k)$ to the controller instead.

Since there is a risk of false alarms in the detection scheme, it is important to make sure that the estimated output $\hat{y}(k)$ will not cause safety concerns to rest of the system. Therefore, this proposed attack response method can only be considered as a temporary solution to isolate the potential attacks before a human operator confirms and responds to the situation.

Another potential solution is to add redundant communication and measurement channels for safety-critical variables, which are independent of the current networks and measurements. When an attack happens, these redundancies can be used to correct the corrupted data which is sent to the controller.

One difference between a fault isolation and an attack response scheme is redundancy consideration. In a fault isolation scheme, redundant components in physical process are considered, while redundant communication channels are considered in an attack response.

7.3.2 Resiliency in mitigation

Once the anomalies are detected, an attack-resilient control scheme should be triggered. Since insider attacks also have impacts on physical processes, fault-tolerant control can be applied to attack-resilient control scheme.

Since attacks are difficult to predict and may drive system to various dangerous conditions, the focus of this section is on design of an attack-resilient control scheme to mitigate anomalies resulted from an attack. There are two aspects to be considered in the design an attack-resilient control scheme: one is a decision-making scheme to determine which control mode should be activated to defend the attack; the other is the designing of corresponding controller to realize the control objectives. A design diagram of a resilient control system is shown in Figure 7.3.

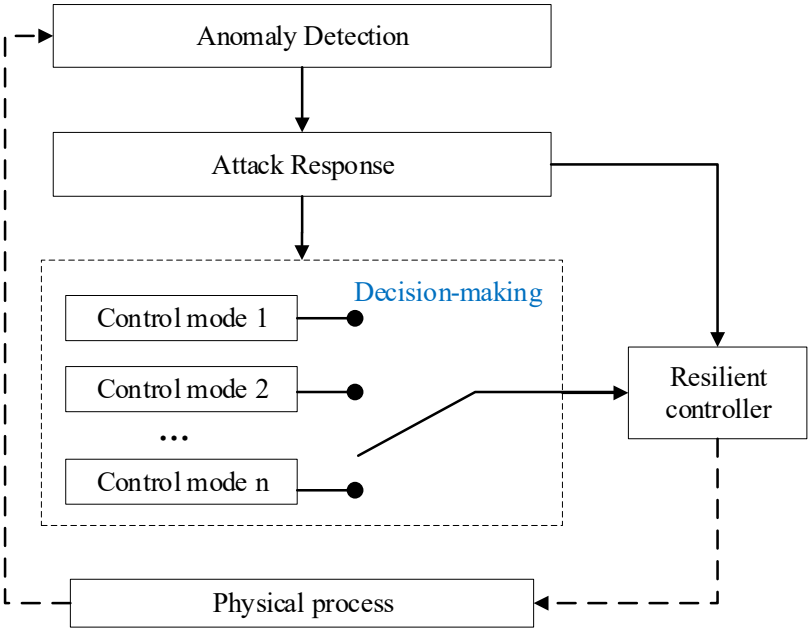


Figure 7.3 Structure of an attack-resilient control system

The decision-making scheme can be designed based on multiple criteria: safety, security and system performance [156]. Based on the replaced or corrected data from the attack response scheme, the decision-making scheme will analyze the security and safety conditions and determine a control mode to mitigate the attack. With a switchable control mode, the controller can mitigate the attack and maintain the system performance in an acceptable degree.

In this chapter, a decision-making scheme is designed based on current safety region of the system once an attack is detected. There are three regions for a system state, as is shown in Figure 7.4.

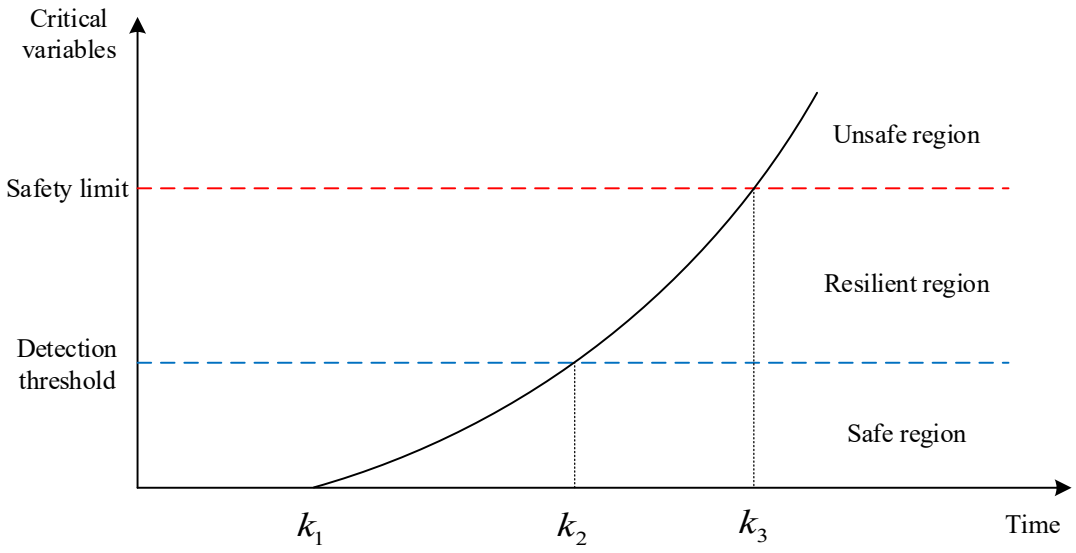


Figure 7.4 Safety region for the decision-making scheme [155]

1) Safe region: in this region, the system is safe and there are no anomalies detected.

2) Resilient region: an anomaly has been detected while the system is in still safe state, the detected anomaly can be mitigated through a resilient control method.

3) Unsafe region: an anomaly has been detected, while system is already in an unsafe state, the detected anomaly is unacceptable. This is a forbidden region.

Based on the system state, there are three control modes that can be triggered by the decision-making scheme.

1) When the system is in the safe region, control mode #1 is used as a normal control. The goal of the controller is to maintain the current performance without any reconfiguration.

2) When the system is in the resilient region, control mode #2 is triggered as a resilient control. The goal of the controller is to mitigate system state back to the safe region or degrade the system performance to reduce impacts of the attack.

3) When the system is in the unsafe region, control mode #3 is triggered as an emergency control. The goal of the controller is to take immediate actions to minimize damages to the system.

7.3.3 Automated mitigation and supervised mitigation

As is mentioned in Section 7.2, the mitigation is considered as a temporary solution to protect the system in case there are anomalies. There are two ways to perform the mitigation strategies. One is automated mitigation, which means that the resilient control scheme will be triggered automatically once an anomaly is detected. The other is supervised mitigation, in which the security situation needs to be confirmed by a human operator when the detection scheme raises an alarm, and then the proposed resilient control scheme will be activated if the operator confirms the security situation.

Since there might be false rate of detection schemes, automated mitigation should ensure the system variables are maintained within their safety limits. Otherwise the automated mitigation may make a wrong decision on the controller selection.

7.4 Case studies

To validate the effectiveness of the proposed scheme, case studies are presented in this section. System performance is analyzed based on the proposed defensive framework and a resilient control on NPCTF. The heater control loop as described in Figure 4.13 is used again.

7.4.1 Experiment design

The purpose of the case studies is to experimentally validate the developed attack defensive framework and resilient control techniques. The model and the state estimation of the heater control system is given in Equation (6.17) and Equation (6.18) , details can be found in Section 6.4.1.2.

7.4.1.1 Attack scenarios

Two attack scenarios have been considered in this chapter.

The first scenario (SA1) is a false-data injection attack on the sensor data of $T_2(k)$. The tampered data sent to the controller is

$$\tilde{T}_2(k) = \begin{cases} T_2(k) - 0.03(k-30) & k > 30 \\ T_2(k) & k \le 30 \end{cases} \tag{7.1}$$

The second scenario (SA2) is a replay attack. The tampered data sent to the controller is

$$\tilde{T}_2(k) = \begin{cases} T_2^{'}(k) & 70 < k \le 200 \\ T_2(k) & k \le 70 \end{cases} \tag{7.2}$$

where $T_2^{'}(k)$ is a set of recorded historical data on the transient response of $T_2(k)$ when the setpoint is changed from 30℃ to 25℃.

7.4.1.2 Attack defensive framework

The attack defensive strategies are presented in Table 7.1.

Table 7.1 Implemented attack defensive strategies

Framework	Methods	Techniques
Attack prevention	D0	Safety limit and security boundary
Attack detection	D1	Rule-based network intrusion detection
	D2	CUSUM method: $\tau = 1, b = 0.5$
	D3	Data-driven clustering-based method
Attack response	R1	Attack response
Attack mitigation	M1	Decision-making unit
	M2	Resilient controllers

1) Attack prevention D0

Two boundaries are set in this layer. One is the safety limit of T_2 as 37℃, which is set within the detection system; the other is the security boundary to prevent external or unauthorized users, which is set in the firewalls inside the supervisory station.

2) Attack detection

The cross-layered detection scheme is implemented in this framework, which includes three detection methods referred as D1, D2 and D3.

D1 is a rule-based detection method, which is implemented in a Snort environment to define the scope of the attack defined in this dissertation. The detection rules include a whitelist for the authorized users and consistency monitoring of network traffics. This method can detect network intrusions and interruptions to the system.

D2 and D3 are implemented with the same settings as in Section 6.4.1.2.

3) Attack response

R1 is an attack response scheme to react to the detection anomalies. Since there are no redundant sensors and communication channels in the heater control loop, in order to isolate the measurement data $\tilde{T}_2(k)$, the estimated $\hat{T}_2(k)$ is used to replace $\tilde{T}_2(k)$ instead when there is an anomaly detected.

Measurement data to the controller $T_2^c(k)$ is set as

$$T_2^c(k) = \begin{cases} \tilde{T}_2(k) & H_0 \\ \hat{T}_2(k) & H_1 \end{cases} \tag{7.3}$$

where the estimated temperature $\hat{T}_2(k)$ is computed using Equation (6.18), $\tilde{T}_2(k)$ is the measured sensor data in the control side.

4) Attack mitigation

To mitigate impacts of an attack, the decision-making scheme needs to determine a resilient control mode based on the detected anomalies, and trigger the corresponding controller. In this

case, M1 is a decision-making unit, to decide which controller should be activated. M2 is a set of PID controllers to mitigate the system back to a safe state.

Safety region of this case study is set to be:

a) Safe region: H_0 is true.

b) Resilient region: H_1 is true, but $\tilde{T}_2(k)$ and $\hat{T}_2(k)$ are lower than 37 ℃

c) Unsafe region: H_1 is true, and $\tilde{T}_2(k)$ or $\hat{T}_2(k)$ is higher than 37 ℃.

The decision-making logic in this case study has three control modes.

a) Control mode #1: normal control.

When there is no attack detected, normal control mode is selected. The observed data $\tilde{T}_2(k)$ is sent to the controller, a PD controller is employed with parameters $K_P = 6.5, K_D = 1$.

b) Control mode #2: attack-resilient control.

When the system is in resilient region, an attack-resilient controller is selected to mitigate impacts of the attack. The observed data $\tilde{T}_2(k)$ is isolated first and the estimated $\hat{T}_2(k)$ is sent to the controller, a PID controller is employed to regulate the system back to its setpoint at 30 ℃. Parameters of the PID controller are $K_P = 7.6, K_I = 0.1, K_D = 1$

c) Control mode #3: emergency control.

When the system is in the unsafe state, an emergency control mode is triggered. Under this situation, the heater will be shut down to avoid further damage, and the emergency control system ECCS in NPCTF will be triggered to ensure that the physical process is safe.

In this case study, the false-data injection attack and the replay attack have been implemented on the heater control system, respectively. The detection scheme including D0- D3 is used to

detect anomalies. Attack response R1 is to respond to the attacks, and mitigation methods M1 and M2 are to mitigate the system. The proposed attack-resilient control system is reconfigured in the supervisory station and then loaded into PLC in advance.

7.4.2 Experimental results

To study the effectiveness of the framework, the experiments are carried out under two cases. The first case is an automated attack response and mitigation, in which all the defensive strategies are executed automatically. The second case is a supervised attack response and mitigation, which requires a human operator to confirm and release the response after an attack is detected.

7.4.2.1 Automated attack response and mitigation

Performance of the proposed automated attack defensive scheme under two attack scenarios have been demonstrated in Figure 7.5.

1) Results of automated mitigation to a FDI attack

The performance of the system when the sensor data of T_2 is tampered by the stealthy false-data injection attack is shown in Figure 7.5(a).

The FDI attack starts at t k=30s, the detection system detects the anomaly at k=54s. Although the detection threshold has been exceeded at this time, the actual $T_2(k)$ is still around its setpoint, and is close to its estimation $\hat{T}_2(k)$. The automated attack response scheme replaces $\tilde{T}_2(k)$ by the estimated $\hat{T}_2(k)$, and the resilient control system keeps $T_2(k)$ in its steady state despite of the attack.

2) Results of automated mitigation to a replay attack

The results of automated mitigation against a replay attack are shown in Figure 7.5(b). Before the attack is launched, the replayed data is recorded in advance by the attacker. System is operating at the steady state. At k=70s, the replay attack starts to send the historical data of T_2

to the controller. At *k*=80s, the historical data starts to decrease the temperature from 30 ℃ to 25 ℃.

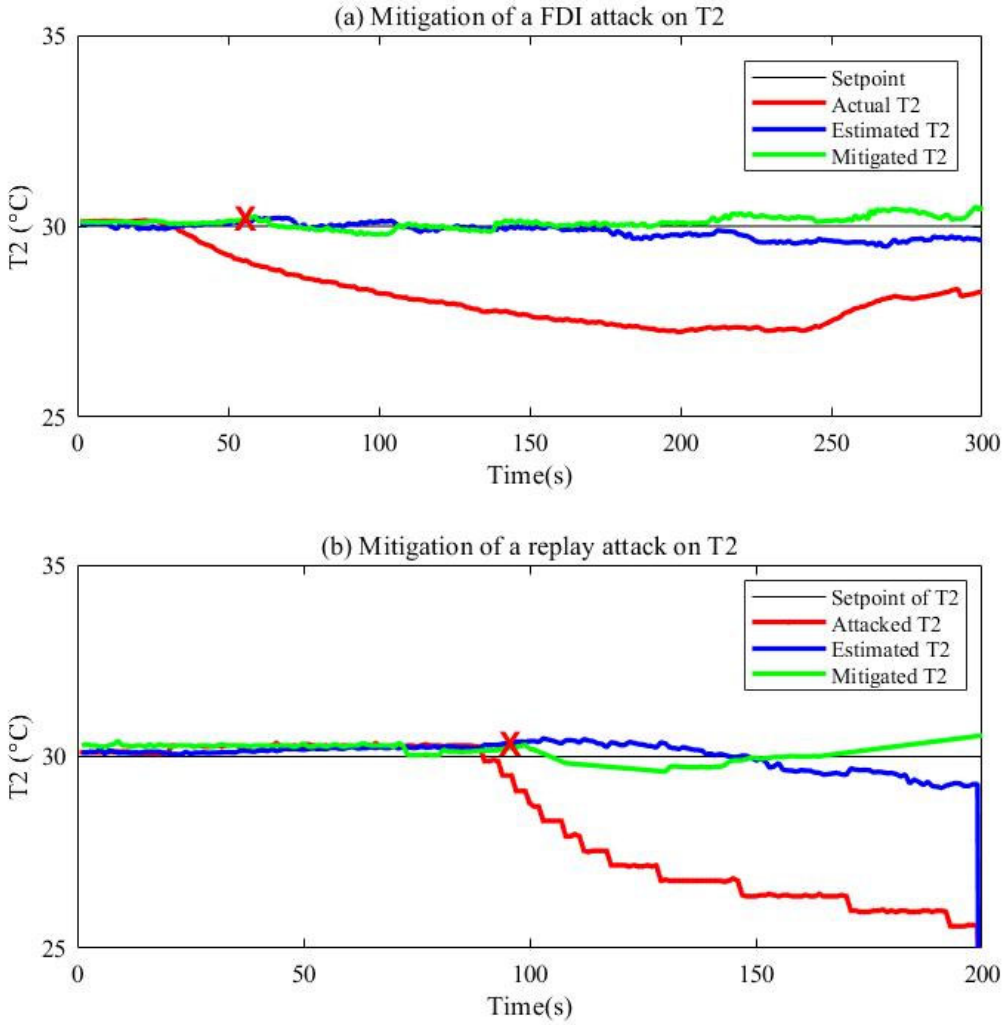


Figure 7.5 Performance of the automated attack-resilient control scheme

The detection system identifies this attack at *k*=93s, and the attack response scheme is triggered to isolate the attacked $\tilde{T}_2(k)$ and the estimated $\hat{T}_2(k)$ is used by the controller instead. Since

the estimated $\hat{T}_2(k)$ is close to the actual $T_2(k)$, it is relatively easy for the controller to maintain the system in its safe region.

It can be observed that the automated attack response and mitigation strategies are effective to eliminate the effects of these attacks and maintain the system performance.

7.4.2.2 Supervised attack response and mitigation

This case study involves three stages: (1)before an attack (normal operation), (2)during an attack without mitigation (validation of detection effectiveness), and (3) during an attack with mitigation (validation of mitigation effectiveness). Experimental results are shown in Figure 7.6.

1) Results of supervised mitigation to a FDI attack

Results of a supervised mitigation against the stealthy false-data injection attack are provided in Figure 7.6(a).

As can be seen, prior to the launch of the attack, the measurement data of T_2 is in its steady state 30℃. At *k*=30s, a FDI attack is initialized to the process the tampered $\tilde{T}_2$ decreases by a deviation of -0.03 (*k*-30) ℃. No mitigation is used in the period from *k*=30s to *k*=100s, the attacker deceives the controller to increase the power of the heater and drives the actual temperature T_2 higher than its setpoint. The attack is detected at *k*=56s and an operator confirms this situation and responds to this attack shortly after.

To validate the effectiveness of the proposed defensive strategies, the resilient control system is activated by the operator at *k*=100s. The measured $\tilde{T}_2(k)$ is replaced by the estimated $\hat{T}_2(k)$, the resilient control mode is triggered. It can be seen that the actual $T_2(k)$ is moving back to its steady state at around 30 ℃ starting from *k*=126s. The resilient control scheme recovers the system to its normal operating condition after the attack is detected and isolated.

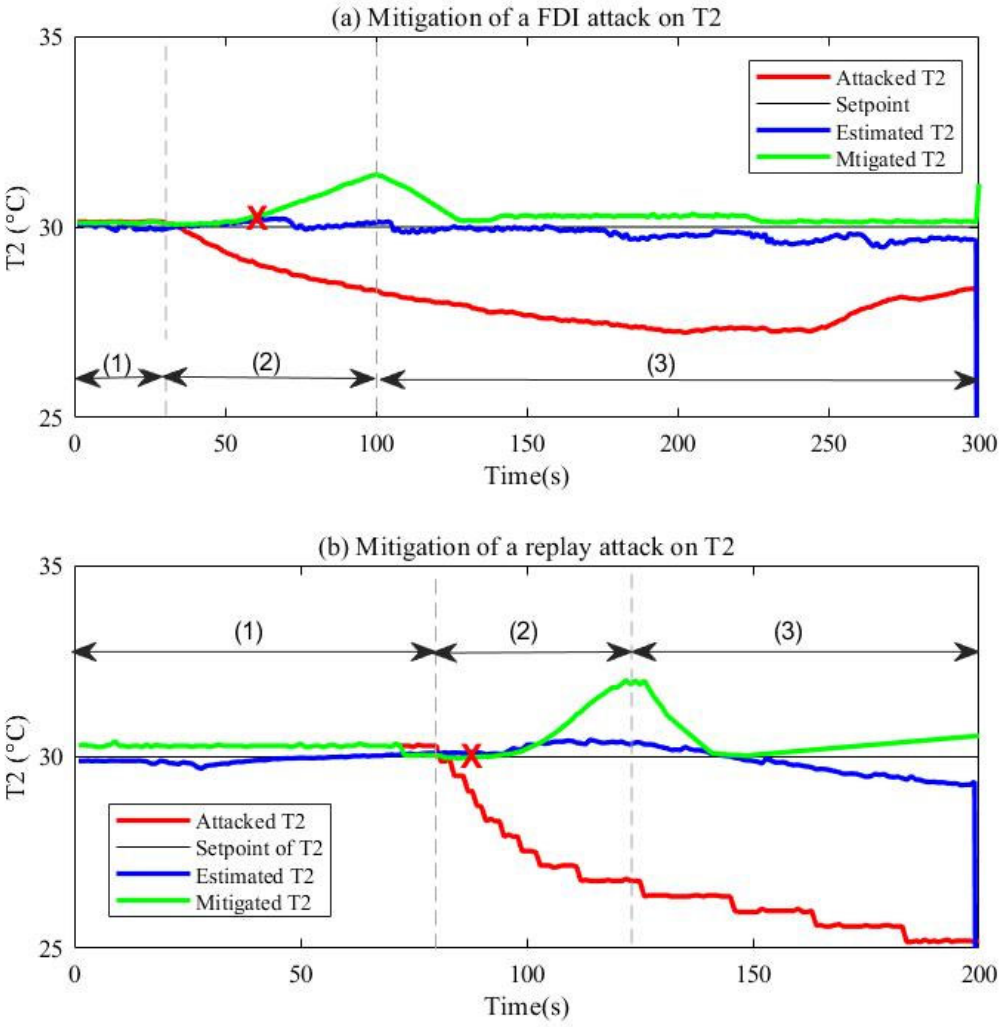


Figure 7.6 Performance of the supervised attack-resilient control scheme

2) Results of supervised mitigation to a replay attack

Results of mitigation against a replay attack are shown in Figure 7.6(b).

Before an attack happens, the system is operating at the steady state. At k=70s, the replay attack starts to send the historical data of T_2 to the controller. To keep the attack stealthy, the replayed data is in the same steady state at the beginning of this attack. At k=80s, the historical data

starts to change the setpoint of T_2 from 30 ℃ to 25 ℃. Since the controller regulates the system based on the replayed data, which causes the continued increase of the actual temperature T_2. The detection system raises an alarm at *k*=88s.

The attack-resilient control system starts operation at *k*=121s, the estimated $\hat{T}_2(k)$ is used by the controller. Results have shown that the proposed attack defensive strategies can bring the system back to its steady state from *k*=147s onward.

Under the case of the replay attack, when the measured T_2 is replayed by the attacker, the detection scheme arises an alarm and the attacked data is replaced. Since the estimated value of $\hat{T}_2$ has deviated from the setpoint 30℃, the decision scheme switches the control objective to maintain the system in a safe state.

Results in Figure 7.5 and Figure 7.6 have demonstrated that the proposed framework is effective to defend against stealthy attacks on sensor data. When an anomaly is detected, the corrupted data is replaced by its estimated values, the decision-making logic will respond to these anomalies and select the most appropriate control algorithm, and the resilient control scheme can recover the system characteristics and return it to the normal operating conditions.

Because the resilient control requires system reconfiguration when responding to attacks, it is often required an operator's confirmation for safety-critical processes, instead of automatically react to the attack. Hence, although the proposed mitigation methods can maintain the system in a safe state, it is considered to be a temporary solution at the best. Human operator will make the final decision.

7.5 Conclusions

In this chapter, an attack defensive framework and an attack-resilient control scheme are proposed to mitigate impacts of insider attacks. The overall framework provides a defense-in-depth defense approach against the studied attacks. The resilient control scheme consists of an

attack response scheme, a decision-making logic and a set of controllers. The attack response isolates the tampered measurements and replaces them by estimated or reconstructed values. The decision-making logic responds to the anomalies identified by the detection scheme and subsequently triggers the desirable control modes. Corresponding controllers are then switched to mitigate the attacks and maintain the safety of the system.

Results have shown that the proposed framework is effective to defend against insider attacks on sensors. This multi-layered defensive framework addresses the security enhancement strategies including attack prevention, detection, response and mitigation, which provide a defense-in-depth protection against insider attacks. The attack-resilient control system integrates the security and safety solutions together, to mitigate attack impacts and maintain the system in a safe state effectively.

Chapter 8

8 Conclusions and Future Work

8.1 Conclusions

The research reported in this dissertation is comprised of theoretical study and experimental evaluation of an attack-resilient control system design, analysis and demonstration. The contributions of this work can be summarized as follows.

8.1.1 Theoretical analysis and design

(1) Security of cyber-physical systems has been investigated and related issues are analyzed.

This work presents the existing research work related to insider attacks. Vulnerabilities of systems are analyzed to determine potential ways for security enhancements. Existing security solutions, including attack prevention, attack detection and attack mitigation strategies are investigated.

System vulnerabilities are important factors for the security enhancement solutions, and from the existing security solutions, some insights can be extracted to strength the system security situation.

(2) A methodology to analyze features of potential insider attacks and their impacts has been proposed.

The methodology is based on system-theoretic and graph-theoretic approaches. Firstly, vulnerability analysis related to insider attacks are analyzed for a general cyber-physical system. Then, an attack pattern is described for such attacks, which includes attack goals, resources, constraints, modes, as well as possible attack paths. Stealthy conditions are analyzed in temporal and spatial dimensions, potential impacts of such attacks on the system are analyzed using an attack tree. Similarities among different cyber-physical attack scenarios and system vulnerabilities have also been examined.

This methodology links the attack impacts with system vulnerabilities, which provides insights into design of security enhancement strategies. Analysis results of the stealthy conditions demonstrate that the limits of insider attacks, and the stealthy condition can also be used as an indicator for attack detection schemes.

(3) An online cross-layer detection scheme has been developed with respect to stealthy insider attacks.

The detection scheme takes on a hierarchical approach by using different detection methods in different layers to provide a defense-in-depth detection against the attacks. A model-based detection method and a data-driven detection method are employed to detect various anomalies.

A cross-layer design provides detections from a cyber layer to a physical process. In this detection scheme, data-driven and model-based detection methods cooperate to reveal the stealthiness of attacks. This methodology has been proven to be effective in detecting both spatial stealthy attacks and temporal attacks.

(4) An attack defensive framework and an attack-resilient control scheme have been proposed.

To make the system resilient to various insider attacks, a multi-layered defensive framework is presented. The framework includes attack prevention, detection, response and mitigation. To mitigate the impacts of attacks, an attack-resilient control scheme is provided, in which a decision-making scheme is designed to make decisions under various threats, and select a suitable attack-resilient controller to mitigate the impacts of attacks.

The defense-in-depth deployment of the attack-resilient control structure provides layered protection for the system. The attack-resilient control system can ensure that the safety-critical physical process remains in the safe state in case of attacks.

8.1.2 Experimental validation and evaluation

(1) A design guideline on how to develop a security platform on a cyber-physical system has been developed, and a modular approach to design such a platform has been proposed for security assessment of cyber-physical systems.

The developed platform consists of three functional modules: (1) Attack Scenario Generation Module, (2) Security Enhancement Module, and (3) Security Evaluation Module. The first module can be used to synthesize attack scenarios to identify system vulnerabilities. The second module provides various strategies to prevent, detect and mitigate attacks. The third module creates a multi-layer systematic environment to analyze and evaluate cyber-physical security issues.

The generalized methodology provides a guideline to develop a security assessment platform. Modular design makes the development and implementation flexible. In addition, this platform proposes a cross-layer framework, it supports not only cyber-physical security assessment but also security enhancement, which makes a diverse and defense-in-depth security study possible.

(2) A prototype platform has been designed and implemented.

A prototype platform has been implemented by using a physical component based dynamic system simulator, known as nuclear process control test facility (NPCTF). To demonstrate the effectiveness of the proposed platform, case studies have been carried out on the proposed platform to demonstrate how to perform different security tests for vulnerability assessment and security enhancement. Different security scenarios have been designed and evaluated on this platform, which bridges a gap between academic research and engineering applications.

The prototype platform can be extended to other cyber-physical systems. Due to the modular design, the proposed generalized modular design is not restricted only to NPCTF, it can be used with other cyber-physical systems with appropriate configurations.

8.2 Limitations of this work

Considering the scope and assumptions of this work, there are a few limitations in the current work.

(1) The studied attacks are insider attacks on tampering on sensor data or control signals. Detection of these kinds of attacks are based on the safety limits of the system, and the data in cyber domain are extracted from the data base in the supervisory station. If the safety limits of the system have been manipulated, the detection scheme and mitigation scheme would have been misled and lead up to wrong decisions. If the database is attacked, the detection and mitigation schemes might be deceived as well. Therefore, it is of importance to secure the safety limits and data base in the supervisory station.

(2) Since the work is focused on attacks a in a single communication channel at a time, when multiple sensors are being tampered, the closed loop control might be interrupted into an open loop control, the attacker might tamper with the data arbitrarily and/ or launch coordinated attacks to bypass or hide their malicious actions. In such situations, the detection and mitigation schemes might not be effective anymore. Therefore, study the number of sensor attacks that can be detected by the proposed system is necessary in the future work.

(3) The purpose of this work is to design an attack-resilient control system against the insider cyber-physical attacks on sensor data and control commands. Detection and mitigation schemes are deployed in supervisory station. It is assumed that the supervisory layer is isolated from rest of the system and is secure with respect to the studied attacks. If the supervisory station is attacked, the attacker might get control of the whole system and arbitrarily change the system configuration. All the data that are received and sent over the communication channel can be obtained and modified by the attacker, the proposed work might not work anymore.

8.3 Future work

Security of cyber-physical systems is an emerging area of research. While the current work presents multiple contributions, continued efforts are still needed. Based on the work so far, future research can be directed to the following topics:

(1) Security analysis

An attack-defense tree could be used to analyze the security situation of system based on game-theoretic tools.

(2) Anomaly detection scheme

- For detection methods that are based on state estimation, credibility of the observed measurement data and the estimated data could be evaluated, to improve the detection rate and false alarms.
- Machine learning techniques can be considered for anomaly detection to online predict the system output or classify the system patterns using measurement data.

(3) Resilient control of cyber-physical systems

- Security control of communication protocols and attack-resilient control of physical process should be integrated to provide a more effective solution against insider attacks.
- To further enhance the ability against stealthy attacks, online cross-layered detection and supervised resilient control techniques should be considered together. A flexible reconfiguration structure is needed to accommodate this research.
- After the detection scheme triggers an alarm, it is important to have techniques to isolate and reconstruct the tampered data to ensure the safety operation of the system. More research is needed in this area.

- Redundant communication and measurement channels can be used for safety-critical variables. More research is needed in this area.

(4) Security platform

The designed prototype platform can be extended to other applications, such as study on attack penetration tests, online reconfiguration of defense strategies, and synthesis of control strategies.

References

[1] G. Wu, J. Sun, and J. Chen, "A survey on the security of cyber-physical systems," *Control Theory and Technology,* vol. 14, pp. 2−10, 2016.

[2] F. Zhang, H. A. D. E. Kodituwakku, W. Hines, and J. B. Coble, "Multi-layer data-driven cyber-attack detection system for industrial control systems based on network, System and Process Data," *IEEE Transactions on Industrial Informatics,* 2019.

[3] S. A. A. A. Cárdenas, and S. S. Sastry, "Research challenges for the security of control systems," in *Proceedings of the 3rd Conference on Hot Topics Security*, Berkeley, pp. 1−6, 2008.

[4] M. Bishop, H. M. Conboy, H. Phan, B. I. Simidchieva, G. S. Avrunin, L. A. Clarke*, et al.*, "Insider threat identification by process analysis," in *2014 IEEE Security and Privacy Workshops (SPW)*, pp. 251−264, 2014.

[5] S. Furnell, "Enemies within: the problem of insider attacks," *Computer fraud & security,* vol. 2004, pp. 6−11, 2004.

[6] R. R. Farwell J P., "Stuxnet and the future of cyber war," *Survival,* vol. 53, pp. 23−40, 2011.

[7] Y. Cherdantseva, P. Burnap, A. Blyth, P. Eden, K. Jones, H. Soulsby*, et al.*, "A review of cyber security risk assessment methods for SCADA systems," *Computers & security,* vol. 56, pp. 1−27, 2016.

[8] R. M. Lee, M. J. Assante, and T. Conway, "Analysis of the cyber attack on the Ukrainian power grid," Electricity Information Sharing and Analysis Center (E-ISAC), 2016.

[9] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Systems Magzine,* vol. 35, pp. 24−45, 2015.

[10] R. Akella, H. Tang, and B. M. McMillin, "Analysis of information flow security in cyber–physical systems," *International Journal of Critical Infrastructure Protection,* vol. 3, pp. 157−173, 2010.

[11] M. Pajic, J. Weimer, N. Bezzo, O. Sokolsky, G. J. Pappas, and I. Lee, "Design and implementation of attack-resilient cyber-physical systems: with a focus on attack-resilient state estimators," *IEEE Control Systems Magzine,* vol. 37, pp. 66−81, 2017.

[12] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, 2009, vol. 5, pp. 1−7.

[13] "ISA/IEC-62443: Security for Industrial Automation and Control Systems", https://www.isa.org/isa99/, accessed June 16, 2019.

[14] R. Chinchani, A. Iyer, H. Q. Ngo, and S. Upadhyaya, "Towards a theory of insider threat assessment," in *International Conference on Dependable Systems and Networks*, pp. 108−117, 2005.

[15] P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese, "Automated insider threat detection system using user and role-based profile assessment," *IEEE Systems Journal*, 11(2), pp. 503−512*,* 2017.

[16] K. Tang, M.T. Zhou, and W.Y. Wang, "Insider cyber threat situational awareness framwork using dynamic Bayesian networks," in *4th International Conference on Computer Science & Education*, pp. 1146−1150, 2009.

[17] R. Taormina, S. Galelli, N. O. Tippenhauer, E. Salomons, and A. Ostfeld, "Characterizing cyber-physical attacks on water distribution systems," *Journal of Water Resources Planning and Management*, 143(5): 04017009, 2017.

[18] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *Proceedings of the IEEE Conference on Decision and Control*, 2010, pp. 5991−5998.

[19] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proceedings of the IEEE Conference on Decision and Control*, 2010, pp. 5967−5972.

[20] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical security via geometric control: Distributed monitoring and malicious attacks," in *Proceedings of the IEEE Conference on Decision and Control*, 2012, pp. 3418−3425.

[21] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid,* vol. 2, pp. 645−658, 2011.

[22] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach," *IEEE Transactions on Signal Processing,* vol. 63, pp. 1102−1114, 2015.

[23] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal DoS attack policy against remote state estimation," in *Proceedings of the IEEE Conference on Decision and Control*, 2013, pp. 5444−5449.

[24] H. Zhang, P. Cheng, L. Shi, and J. Chen, "Optimal Denial-of-service attack scheduling against linear quadratic Gaussian control," in *Proceedings of the American Control Conference*, 2014, pp. 3996−4001.

[25] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On false data-injection attacks against power system state estimation: Modeling and countermeasures," *IEEE Transactions on Parallel and Distributed Systems,* vol. 25, pp. 717−729, 2014.

[26] Q. Shafi, "Cyber physical systems security: A brief survey," in *Proceedings of 12th International Conference on Computational Science and Its Applications, ICCSA 2012*, pp. 146−150, 2012.

[27] A. A. Cárdenas, S. Amin, Z. S. Lin, Y.L. Huang, C.Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and communications security*, 2011, pp. 355−366.

[28] D. Kundur, X. Feng, S. Mashayekh, S. Liu, T. Zourntos, and K. L. Butler-Purry, "Towards modelling the impact of cyber attacks on a smart grid," *International Journal of Security and Networks,* vol. 6, pp. 2−13, 2011.

[29] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm),* pp. 226−231, 2010.

[30] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *2012 IEEE Global Communications Conference (GLOBECOM),* pp. 3153−3158, 2012.

[31] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Sparse malicious false data injection attacks and defense mechanisms in smart grids," *IEEE Transactions on Industrial Informatics,* vol. 11, pp. 1198−1209, 2015.

[32] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica,* vol. 51, pp. 135−148, 2015.

[33] C. R. Mo Y, Sinopoli B., "Detecting integrity attacks on SCADA systems," *IEEE Transactions on Control Systems Technology,* vol. 22, pp. 1396−1407, 2014.

[34] R. Mahmoud, T. Yousuf, F. Aloul, and I. Zualkernan, "Internet of things (IoT) security: Current status, challenges and prospective measures," in *2015 10th International Conference for Internet Technology and Secured Transactions, ICITST 2015*, pp. 336−341, 2015.

[35] K. Zhao and L. Ge, "A survey on the internet of things security," in *Proceedings of 9th International Conference on Computational Intelligence and Security, CIS 2013*, pp. 663−667.

[36] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks," *IEEE Transactions on Smart Grid,* vol. 4, pp. 1244−1253, 2013.

[37] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magzine,* vol. 35, pp. 93−109, 2015.

[38] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *Proceedings of the IEEE Conference on Decision and Control*, 2015, pp. 5820−5826.

[39] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks," *IEEE Transactions on Automatic Control,* vol. 59, pp. 1454−1467, 2014.

[40] E. Mousavinejad, F. Yang, Q. L. Han, and L. Vlacic, "A novel cyber attack detection method in networked control systems," *IEEE transactions on cybernetics,* pp. 1−11, 2018.

[41] S. Zonouz, J. Rrushi, and S. McLaughlin, "Detecting industrial control malware using automated PLC code analytics," *IEEE Security and Privacy,* vol. 12, pp. 40−47, 2014.

[42] S. Zonouz, K. M. Rogers, R. Berthier, R. B. Bobba, W. H. Sanders, and T. J. Overbye, "SCPSE: Security-oriented cyber-physical state estimation for power grid critical infrastructures," *IEEE Transactions on Smart Grid,* vol. 3, pp. 1790−1799, 2012.

[43] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems a quantitative risk management approach," *IEEE Control Systems Magazine,* vol. 35, pp. 24−45, 2015.

[44] B. O. Nurse J R C, Legg P A, et al., "Understanding insider threat: a framework for characterising attacks," presented at *2014 IEEE Security and Privacy Workshops (SPW)*, 2014.

[45] F. Kammüller and C. W. Probst, "Modeling and verification of insider threats using logical analysis," *IEEE Systems journal,* 11(2), pp. 534−545, 2015.

[46] M. Rocchetto and N. O. Tippenhauer, "On attacker models and profiles for cyber-physical systems," in *European Symposium on Research in Computer Security*, pp. 427−449, 2016.

[47] D. Liu, X. Wang, and J. Camp, "Game-theoretic modeling and analysis of insider threats," *International Journal of Critical Infrastructure Protection,* vol. 1, pp. 75−80, 2008.

[48] K. N. Kim, M. S. Yim, and E. Schneider, "A study of insider threat in nuclear security analysis using game theoretic modeling," *Annals of Nuclear Energy,* vol. 108, pp. 301−309, 2017.

[49] H. Orojloo and M. A. Azgomi, "A method for evaluating the consequence propagation of security attacks in cyber–physical systems," *Future Generation Computer Systems,* vol. 67, pp. 57−71, 2017.

[50] P. Y. Chen, S. Yang, J. A. McCann, J. Lin, and X. Yang, "Detection of false data injection attacks in smart-grid systems," *IEEE Communications Magazine,* vol. 53, pp. 206−213, 2015.

[51] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, pp. 5967−5972, 2010.

[52] S. Duttagupta, K. Ramamritham, and P. Kulkarni, "Tracking dynamic boundaries using sensor network," *IEEE Transactions on Parallel and Distributed Systems,* vol. 22, pp. 1766−1774, 2011.

[53] A. Sarkar, S. Köhler, S. Riddle, B. Ludaescher, and M. Bishop, "Insider attack identification and prevention using a declarative approach," in *2014 IEEE Security and Privacy Workshops (SPW),* pp. 265−276, 2014.

[54] T. Chee-Wooi, G. Manimaran, and L. Chen-Ching, "Cybersecurity for critical infrastructures: attack and defense modeling," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,* vol. 40, pp. 853−865, 2010.

[55] S. Abraham and S. Nair, "A predictive framework for cyber security analytics using attack graphs," *arXiv preprint arXiv:1502.01240,* 2015.

[56] I. N. Fovino, M. Masera, and A. De Cian, "Integrating cyber attacks within fault trees," *Reliability Engineering & System Safety,* vol. 94, pp. 1394−1402, 2009.

[57] I. Agrafiotis, J. R. Nurse, O. Buckley, P. Legg, S. Creese, and M. Goldsmith, "Identifying attack patterns for insider threat detection," *Computer Fraud & Security,* vol. 2015, pp. 9−17, 2015.

[58] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC),* vol. 14, pp. 13, 2011.

[59] A. Teixeira, G. Dán, H. Sandberg, and K. H. Johansson, "A cyber security study of a SCADA energy management system: Stealthy deception attacks on the state estimator," in *IFAC Proceedings Volumes*, pp. 11271−11277, 2011.

[60] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid,* vol. 5, pp. 580−591, 2014.

[61] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Stealthy deception attacks on water SCADA systems," in *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control*, 2010, pp. 161−170.

[62] M. Zhu and S. Martínez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Transactions on Automatic Control,* vol. 59, pp. 804−808, 2014.

[63] A. Teixeira, "Toward cyber-secure and resilient networked control systems," KTH Royal Institute of Technology, 2014.

[64] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control,* vol. 58, pp. 2715−2729, 2013.

[65] A. Teixeira, H. Sandberg, and K. H. Johansson, "Networked control systems under cyber attacks with applications to power networks," in *American Control Conference (ACC)*, pp. 3690−3696, 2010.

[66] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1806−1813, 2012.

[67] A. Ashok, A. Hahn, and M. Govindarasu, "Cyber-physical security of wide-area monitoring, protection and control in a smart grid environment," *Journal of advanced research,* vol. 5, pp. 481−489, 2014.

[68] B. Genge, C. Siaterlis, and M. Hohenadel, "Impact of network infrastructure parameters to the effectiveness of cyber attacks against Industrial Control Systems," *International Journal of Computers, Communications and Control,* vol. 7, pp. 674−687, 2012.

[69] B. Genge, I. Kiss, and P. Haller, "A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures," *International Journal of Critical Infrastructure Protection,* vol. 10, pp. 3−17, 2015.

[70] Y. L. Huang, A. A. Cárdenas, S. Amin, Z. S. Lin, H. Y. Tsai, and S. Sastry, "Understanding the physical and economic consequences of attacks on control

systems," *International Journal of Critical Infrastructure Protection,* vol. 2, pp. 73−83, 2009.

[71] M. Krotofil, A. Cárdenas, J. Larsen, and D. Gollmann, "Vulnerabilities of cyber-physical systems to stale data-determining the optimal time to launch attacks," *International Journal of Critical Infrastructure Protection,* vol. 7, pp. 213−232, 2014.

[72] A. Wasicek, P. Derler, and E. A. Lee, "Aspect-oriented modeling of attacks in automotive cyber-physical systems," in *Design Automation Conference*, pp. 1−6, 2014.

[73] W. Wu, R. Kang, and Z. Li, "Risk assessment method for cyber security of cyber physical systems," in *Proceedings of 2015 the 1st International Conference on Reliability Systems Engineering, ICRSE 2015*, 2015.

[74] M. Yampolskiy, P. Horváth, X. D. Koutsoukos, Y. Xue, and J. Sztipanovits, "A language for describing attacks on cyber-physical systems," *International Journal of Critical Infrastructure Protection,* vol. 8, pp. 40−52, 2015.

[75] R. Piggin, "Emerging good practice for cyber security of industrial control systems and SCADA," in *7th IET International Conference on System Safety, incorporating the Cyber Security Conference 2012*, pp. 1−6, 2012.

[76] D. F. Pasqualetti F, Bullo F. , "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control,* vol. 58, pp. 2715−2729, 2013.

[77] S. Sundaram, M. Pajic, C. N. Hadjicostis, R. Mangharam, and G. J. Pappas, "The wireless control network: Monitoring for malicious behavior," in *Proceedings of the IEEE Conference on Decision and Control*, 2010, pp. 5979−5984.

[78] I. Kiss, B. Genge, and P. Haller, "A clustering-based approach to detect cyber attacks in process control systems," in *Proceeding of 2015 IEEE International Conference on Industrial Informatics, (INDIN 2015)*, 2015, pp. 142−148.

[79] M. Krotofil, J. Larsen, and D. Gollmann, "The process matters: Ensuring data veracity in cyber-physical systems," in *ASIACCS 2015 Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, pp. 133−144, 2015.

[80] T. Liu, Y. Sun, Y. Liu, Y. Gui, Y. Zhao, D. Wang*, et al.*, "Abnormal traffic-indexed state estimation: a cyber-physical fusion approach for smart grid attack detection," *Future Generation Computer Systems,* vol. 49, pp. 94−103, 2015.

[81] S. Ntalampiras, "Automatic identification of integrity attacks in cyber-physical systems," *Expert Systems with Applications,* vol. 58, pp. 164−173, 2016.

[82] F. Sakiz and S. Sen, "A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV," *Ad Hoc Networks,* vol. 61, pp. 33−50, 2017.

[83] H. Vincent, L. Wells, P. Tarazaga, and J. Camelio, "Trojan detection and side-channel analyses for cyber-security in cyber-physical manufacturing systems," *Procedia Manufacturing,* vol. 1, pp. 77−85, 2015.

[84] A. Teixeira, H. Sandberg, and K. H. Johansson, "Networked control systems under cyber attacks with applications to power networks," in *Proceedings of the 2010 American Control Conference,* pp. 3690−3696, 2010.

[85] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 1854−1859, 2013.

[86] R. Chabukswar, Y. Mo, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, pp. 11239−11244, 2011.

[87] B. Tang, L. D. Alvergue, and G. Gu, "Secure networked control systems against replay attacks without injecting authentication noise," *Proceedings of the American Control Conference, Montreal: IEEE,* pp. 60280−66036, 2012.

[88] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid,* vol. 2, pp. 326−333, 2011.

[89] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Transactions on Control Systems Technology,* vol. 22, pp. 1396−1407, 2014.

[90] Y. Mo, J. P. Hespanha, and B. Sinopoli, "Resilient detection in the presence of integrity attacks," *IEEE Transactions on Signal Processing,* vol. 62, pp. 31−43, 2014.

[91] M. Garcia, A. Giani, and R. Baldick, "Smart grid data integrity attacks: Observable islands," in *2015 IEEE Power & Energy Society General Meeting,* pp. 1−5, 2015.

[92] B. Li, R. Lu, W. Wang, and K. K. R. Choo, "Distributed host-based collaborative detection for false data injection attacks in smart grid cyber-physical system," *Journal of Parallel and Distributed Computing,* vol. 103, pp. 32−41, 2017.

[93] Q. Yang, L. Chang, and W. Yu, "On false data injection attacks against Kalman filtering in power system dynamic state estimation," *Security and Communication Networks,* vol. 9, pp. 833−849, 2016.

[94] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control,* vol. 58, pp. 2715−2729, 2013.

[95] A. N. Bishop and A. V. Savkin, "Set-valued state estimation and attack detection for uncertain descriptor systems," *IEEE Signal Processing Letters,* vol. 20, pp. 1102−1105, 2013.

[96] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Transactions on Smart Grid,* vol. 4, pp. 160−169, 2013.

[97] L. Liu, M. Esmalifalak, and Z. Han, "Detection of false data injection in power grid exploiting low rank and sparsity," in *IEEE International Conference on Communications*, pp. 4461−4465, 2013.

[98] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid,* vol. 5, pp. 612−621, 2014.

[99] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *2010 1st IEEE International Conference on Smart Grid Communications, Smart Grid Communication*, pp. 1−6, 2010.

[100] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *Proceedings of IEEE SmartGridComm*, 2010, pp. 220−225.

[101] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical security via geometric control: Distributed monitoring and malicious attacks," *IEEE Conference on Decision and Control, Atlanta,* pp. 1096−1101, 2010.

[102] M. Zhu and S. Martínez, "Stackelberg-game analysis of correlated attacks in cyber-physical systems," in *Proceedings of the American Control Conference*, 2011, pp. 4063−4068.

[103] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 5469, pp. 31−45, 2009.

[104] M. Zhu and S. Martinez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Transactions on Automatic Control,* vol. 59, pp. 804−808, 2014.

[105] M. Zhu and S. Martínez, "On distributed constrained formation control in operator-vehicle adversarial networks," *Automatica,* vol. 49, pp. 3571−3582, 2013.

[106] H. Shisheh Foroush and S. Martinez, "On event-triggered control of linear systems under periodic denial-of-service jamming attacks," in *Proceedings of the IEEE Conference on Decision and Control*, 2012, pp. 2551−2556.

[107] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *Proceedings of the American Control Conference*, 2013, pp. 3344−3349.

[108] C. Kwon and I. Hwang, "Hybrid robust controller design: cyber attack attenuation for cyber-physical systems," in *Proceedings of the IEEE Conference on Decision and Control*, 2013, pp. 188−193.

[109] Q. Zhu and T. Başar, "Robust and resilient control design for cyber-physical systems with an application to power systems," in *Proceedings of the IEEE Conference on Decision and Control*, 2011, pp. 4066−4071, 2011.

[110] S. Amin, G. A. Schwartz, and S. Shankar Sastry, "Security of interdependent and identical networked control systems," *Automatica,* vol. 49, pp. 186−192, 2013.

[111] S. McLaughlin, "CPS: Stateful policy enforcement for control system device usage," in *29th Annual Computer Security Applications Conference, ACSAC 2013,* New Orleans, LA, United states, pp. 109−118, 2013.

[112] H. Fawzi, P. Tabuada, and S. Diggavi, "Security for control systems under sensor and actuator attacks," in *Proceedings of the IEEE Conference on Decision and Control*, 2012, pp. 3412−3417.

[113] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Transactions on Automatic Control,* vol. 60, pp. 1145−1151, 2015.

[114] S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska, and J. Dong, "Finite energy and bounded attacks on control system sensor signals," in *Proceedings of the American Control Conference*, 2014, pp. 1716−1722.

[115] Z. H. Pang and G. P. Liu, "Design and implementation of secure networked predictive control systems under deception attacks," *IEEE Transactions on Control Systems Technology,* vol. 20, pp. 1334−1342, 2012.

[116] H. Lin, A. Slagell, Z. Kalbarczyk, P. W. Sauer, and R. K. Iyer, "Semantic security analysis of SCADA networks to detect malicious control commands in power grids," in *Proceedings of the ACM Conference on Computer and Communications Security*, 2013, pp. 29−34.

[117] P. Y. Chen, S. M. Cheng, and K. C. Chen, "Information fusion to defend intentional attack in Internet of things," *IEEE Internet of Things Journal,* vol. 1, pp. 337−348, 2014.

[118] F. Hu, Y. Lu, A. V. Vasilakos, Q. Hao, R. Ma, Y. Patil, "Robust cyber-physical systems: concept, models, and implementation," *Future Generation Computer Systems,* vol. 56, pp. 449−475, 2016.

[119] S. R. Moosavi, T. N. Gia, A. M. Rahmani, E. Nigussie, S. Virtanen, J. Isoaho, "SEA: A secure and efficient authentication and authorization architecture for IoT-based healthcare using smart gateways," *Procedia Computer Science*, pp. 452−459, 2015.

[120] P. Venkitasubramaniam, J. Yao, and P. Pradhan, "Information-theoretic security in stochastic control systems," in *Proceedings of the IEEE*, 2015, vol. 103(10), pp. 1914−1931.

[121] H. Yoo and T. Shon, "Challenges and research directions for heterogeneous cyber-physical system based on IEC 61850: Vulnerabilities, security requirements, and security architecture," *Future Generation Computer Systems,* vol. 61, pp. 128−136, 2016.

[122] R. Mitchell and I. R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Computing Surveys (CSUR),* vol. 46, p. 55, 2014.

[123] C. Neuman, "Challenges in security for cyber-physical systems," *DHS Workshop on Future Directions in Cyber-Physical Systems Security,* pp. 22−24, 2009.

[124] K. Paridari, N. O'Mahony, A. E. D. Mady, R. Chabukswar, M. Boubekeur, and H. Sandberg, "A framework for attack-resilient industrial control systems: Attack detection and controller reconfiguration," in *Proceedings of the IEEE*, 2018, vol. 106, pp. 113−128.

[125] T. Lu, J. Lin, Zhao, Y. Li, and Y. Peng, "A security architecture in cyber-physical systems: Security theories, Analysis, Simulation and application fields," *International Journal of Security and its Applications,* vol. 9, pp. 1−16, 2015.

[126] T. Lu, B. Xu, X. Guo, L. Zhao, and F. Xie, "A new multilevel framework for cyber-physical system Security," *First international workshop on the swarm at the edge of the cloud,* 2013.

[127] E. K. Wang, Y. Ye, X. Xu, S. M. Yiu, L. C. K. Hui, and K. P. Chow, "Security issues and challenges for cyber physical system," in *Proceedings of the 2010 IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social Computing*, pp. 733−738, 2010.

[128] J. Hunker and C. W. Probst, "Insiders and insider threats-an overview of definitions and mitigation techniques," *JoWUA,* vol. 2, pp. 4−27, 2011.

[129] I. N. Fovino, L. Guidi, M. Masera, and A. Stefanini, "Cyber security assessment of a power plant," *Electric Power Systems Research,* vol. 81, pp. 518−526, 2011.

[130] Y. Ashibani and Q. H. Mahmoud, "Cyber physical systems security: Analysis, challenges and solutions," *Computers & Security,* vol. 68, pp. 81−97, 2017.

[131] S. C. Genge B L, Fovino I N, et al., "A cyber-physical experimentation environment for the security analysis of networked industrial control systems," *Computers & Electrical Engineering,* vol. 38, pp. 1146−1161, 2012.

[132] A. Ashok, "Attack-resilient state estimation and testbed-based evaluation of cyber security for wide-area protection and control," Iowa State University, 2017.

[133] N. Govil, A. Agrawal, and N. O. Tippenhauer, "On ladder logic bombs in industrial control systems," *Computer Security*, Springer, Cham, 2017: 110−126.

[134] Jin Jiang, Ataul Bari, and Drew J. Rankin, "A physical simulator in supporting of research and development for instrumentation and control systems in nuclear power plants," presented at *the Proceedings of the 9th International Conference on Nuclear Plant Instrumentation, Control & Human–Machine Interface Technologies*, Charlotte, USA, 2015.

[135] T. Lu, J. Lin, L. Zhao, Y. Li, and Y. Peng, "A security architecture in cyber-physical systems: security theories, analysis, simulation and application fields," *International Journal of Security and Its Applications,* vol. 9, pp. 1−16, 2015.

[136] X. Jin, W. M. Haddad, and T. Yucelen, "An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems," *IEEE Transactions on Automatic Control,* 62(11), pp. 6058−6064, 2017.

[137] R. Liu, C. Vellaithurai, S. S. Biswas, T. T. Gamage, and A. K. Srivastava, "Analyzing the cyber-physical impact of cyber events on the power grid," *IEEE Transactions on Smart Grid,* vol. 6, pp. 2444−2453, 2015.

[138] P. W. Tsai and C. S. Yang, "Testbed@ TWISC: A network security experiment platform," *International Journal of Communication Systems,* vol. 31, 2018.

[139] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu, "Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid," *IEEE Transactions on Smart Grid,* vol. 4, pp. 847−855, 2013.

[140] "National SCADA Test Bed: Fact Sheet," Idaho National Laboratory, 2009.

[141] A. P. Mathur and N. O. Tippenhauer, "SWaT: A water treatment testbed for research and training on ICS security," in *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater),* 2016, pp. 31−36.

[142] M. M. I. Nai Fovino, L. Guidi, and G. Carpi., "An experimental platform for assessing SCADA vulnerabilities and countermeasures in power plants," in *In Human System Interactions (HSI), 2010 3rd Conference on*, pp. 679−686, 2010.

[143] T. Edgar, D. Manz, and T. Carroll, "Towards an experimental testbed facility for cyber-physical security research," Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2012.

[144] C. Siaterlis, B. Genge, and M. Hohenadel, "EPIC: a testbed for scientifically rigorous cyber-physical security experimentation," *IEEE Transactions on Emerging Topics in Computing,* vol. 1, pp. 319−330, 2013.

[145] C. Vellaithurai, A. Srivastava, S. Zonouz, and R. Berthier, "CPIndex: cyber-physical vulnerability assessment for power-grid infrastructures," *IEEE Transactions on Smart Grid,* vol. 6, pp. 566−575, 2015.

[146] M. McDonald, J. Mulder, B. Richardson, R. Cassidy, A. Chavez, N. Pattengale*, et al.*, "Modeling and simulation for cyber-physical system security research, development and applications," *Sandia National Laboratories, Tech. Rep. Sandia Report SAND2010−0568,* 2010.

[147] M. A. Queiroz C, Tari Z., "SCADASim—A framework for building SCADA simulations," *IEEE Transactions on Smart Grid,* vol. 2, pp. 589−597, 2014.

[148] M. Mallouhi, Y. Al-Nashif, D. Cox, T. Chadaga, and S. Hariri, "A testbed for analyzing security of SCADA control systems (TASSCS)," in *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, pp. 1−7, 2011.

[149] J. Mirkovic, T. V. Benzel, T. Faber, R. Braden, J. T. Wroclawski, and S. Schwab, "The DETER project: Advancing the science of cyber security experimentation and test," in *Technologies for Homeland Security (HST), 2010 IEEE International Conference on*, pp. 1−7, 2010.

[150] S. Poudel, Z. Ni, and N. Malla, "Real-time cyber physical system testbed for power system security and control," *International Journal of Electrical Power & Energy Systems,* vol. 90, pp. 124−133, 2017.

[151] G. Bernieri, E. E. Miciolino, F. Pascucci, and R. Setola, "Monitoring system reaction in cyber-physical testbed under cyber-attacks," *Computers & Electrical Engineering,* 59, pp.86−98, 2017

[152] J. Hong, S. S. Wu, A. Stefanov, A. Fshosha, C. C. Liu, P. Gladyshev*, et al.*, "An intrusion and defense testbed in a cyber-power system environment," in *IEEE power and energy society general meeting*, pp. 1−5, 2011.

[153] I. Friedberg, K. McLaughlin, P. Smith, D. Laverty, and S. Sezer, "STPA-SafeSec: safety and security analysis for cyber-physical systems," *Journal of Information Security and Applications,* vol. 34, pp. 183−196, 2017.

[154] D. Hadžiosmanović, R. Sommer, E. Zambon, and P. H. Hartel, "Through the eye of the PLC: semantic security monitoring for industrial processes," in *Proceedings of the 30th Annual Computer Security Applications Conference*, 2014, pp. 126−135.

[155] Sun H, Peng C, Zhang W, "Security-based resilient event-triggered control of networked control systems under denial of service attacks," J*ournal of the Franklin Institute*, 2018.

[156] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths*, et al.*, "A survey of physics-based attack detection in cyber-physical systems," *ACM Computing Surveys (CSUR),* vol. 51(4), No. 76, 2018.

[157] N. J. R. C. Agrafiotis I, Buckley O, et al., "Identifying attack patterns for insider threat detection," *Computer Fraud & Security, 2015,* vol. 2015, pp. 9−17, 2015.

[158] L. Cazorla, C. Alcaraz, and J. Lopez, "Cyber stealth attacks in critical information infrastructures," *IEEE Systems Journal,* 12(2), pp. 1778−1792, 2018.

[159] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *2015 IEEE 54th Annual Conference on Decision and Control (CDC)*, pp. 5820−5826, 2015.

[160] C. Kwon, "Cyber attack analysis on cyber-physical systems: Detectability, severity, and attenuation strategy," Purdue University, 2013.

[161] C. Kwon and I. Hwang, "Cyber attack mitigation for cyber–physical systems: hybrid system approach to controller design," *IET Control Theory & Applications,* vol. 10, pp. 731−741, 2016.

[162] H. R. Ghaeini, D. Antonioli, F. Brasser, A. R. Sadeghi, and N. O. Tippenhauer, "State-aware anomaly detection for industrial control systems," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 2018, pp. 1620−1628.

[163] D. M. Cappelli, A. P. Moore, and R. F. Trzeciak, “The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud)," Addison-Wesley, 2012.

[164] Kordy B, Piètre-Cambacédès L, Schweitzer P. "DAG-based attack and defense modeling: Don’t miss the forest for the attack trees," *Computer science review*, 2014, 13: 1-38.

[165] Hahn A, Thomas R K, Lozano I, et al. “A multi-layered and kill-chain based security analysis framework for cyber-physical systems,” *International Journal of Critical Infrastructure Protection*, 2015, 11: 39-50.

Appendices

Appendix A: Code packages*.

No.	File Name	Files
A1.1	Attack HMI	Attack HMI design codes
		Attack data processing codes
		UWO Security test interface
A1.2	Attack scripts	Communication protocol parsing codes
		Attack data processing code package
		Trigger logic design
		Attack scripts for attack scenarios in the dissertation
A1.3	Activation switch	Activation switch PCB design and configuration
		Activation switch configuration codes
A1.4	ABB AC700F DCS program configuration	OPC interface design
		Anomaly detection HMI design
		Mitigation reconfiguration program
A1.5	Defense programs	OPC communication codes
		Data collection and processing codes
		Detection codes for different detection methods
		Mitigation codes
		Results demonstration codes
A1.6	Snort environment settings	Snort rules

* Considering the security of the designed platform and security of NPCTF, the developed code is not publicized. Please contact UWO CIE Lab for more information if necessary.

Appendix B: Demo videos.

No.	Video Name	Links
B1.1	Demo for security platform design	http://cies-western-eng.ca/xirong/file1.mp4
B1.2	Demo for security tests	http://cies-western-eng.ca/xirong/file2.mp4

Curriculum Vitae

Name: Xirong Ning

Post-secondary Education and Degrees:

Tsinghua University
Beijing, China
2003-2006 M.A.

The University of Western Ontario
London, Ontario, Canada
2013-2019 Ph.D.

Related Work Experience

Research Assistant and Teaching Assistant
The University of Western Ontario
2013-2019

Control System Engineer
Chinergy Co. LTD
Beijing, China
2006-2013.

Publications:

[1] **Xirong Ning** and Jing Jiang, *"*Defense-in-depth against insider attacks in cyber-physical systems*," Internet of Things and Cyber-Physical Systems,* 2022, 2: 203-211.

[2] **Xirong Ning** and Jing Jiang, "Design, analysis and implementation of a security assessment platform for cyber-physical systems," *IEEE Transaction on Industrial Informatics*, 2022, 18(2):1154-1164.

[3] **Xirong Ning** and Jing Jiang, "In the mind of insider attackers on cyber-physical systems and how not being fooled," *IET Cyber-Physical Systems: Theory & Applications*, 2020, 5(2):153-161.

[4] **Xirong Ning** and Jing Jiang, "Methods for Identification and Analysis of Safe Operating Boundary for Dynamic Systems", *NPIC & HMIT 2015 Conference*, 2015, Charlotte, USA.