Resilient Monitoring and Control Systems: Design, Analysis, and Performance Evaluation

by

Maruthi T Ravichandran

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2015

Doctoral Committee:

Professor Semyon M. Meerkov, Chair
Assistant Professor Laura K. Balzano
Professor Ilya V. Kolmanovsky
Assistant Professor Necmiye Ozay
Associate Professor S. Sandeep Pradhan

To my parents and the University of Michigan

ACKNOWLEDGEMENTS

The first person whom I would like to thank is my advisor, Professor Semyon Meerkov. His tireless guidance, thoughtful mentorship, and unwavering support were crucial to my success as a PhD student. I consider myself fortunate to have been a part of his research group at the University of Michigan (UM).

I would like to express my gratitude to the members of my dissertation committee, Professors Laura Balzano, Ilya Kolmanovsky, Necmiye Ozay, and Sandeep Pradhan, for their interest in my research and helpful feedback about this dissertation.

I am grateful to the Idaho National Lab (INL) and the UM Department of Electrical Engineering and Computer Science for their support of our research. I also wish to thank Dr. Humberto Garcia (INL) and Dr. Wen-Chiao Lin (formerly with INL) for their partnership in our research.

I would like to thank Professor Achilleas Anastasopoulos and Becky Turanski for their help and encouragement, especially during a critical phase of the initial part of my PhD program. Thanks are also due to Ann Pace and Michelle Feldkamp for their assistance in dealing with logistical issues during the program.

I am grateful to my friends, Raj Suryaprakash, G. K. Chaitanya, Deepika Mutukuri, Kirithika Rajendran, Satya Chandrasekar, Karthik Kameshwaran, and several others, for their company, which made life in Ann Arbor fun and interesting.

Finally, I would like to thank my parents and Megha for their unceasing love and encouragement, without which this research would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Resilient Monitoring and Control Systems: Design, Analysis, and Performance Evaluation

by

Maruthi T Ravichandran

Chair: Professor Semyon M. Meerkov

Critical infrastructure systems (i.e., power plants, power grids, transportation networks, chemical plants, etc.) and their sensor networks are vulnerable to cyber-physical attacks. Cyber attacks refer to the malicious manipulation of the sensor data, while physical attacks refer to the intentional damage of the plant components, by an adversary. The goal of this dissertation is to develop monitoring and control systems that are resilient to these attacks.

The monitoring system is termed resilient if it provides the least uncertain (in terms of the minimum entropy) process variable estimates and plant condition assessment. Similarly, the feedback control system is termed resilient if it identifies the actuators under attack and generates the best possible control signals (in terms of the largest probability of maintaining the process variables in the desired range).

The resilient monitoring system (RMS) developed in this research consists of five layers: Data quality acquisition, process variable assessment, plant condition assessment, sensor network adaptation, and decentralized knowledge fusion. The techniques involved in each of these layers are rigorously analyzed and are shown to identify the plant condition - normal or anomalous - in a reliable and timely manner. The developed RMS is applied to a model of a power plant, and its performance is evaluated under several cyber-physical

attack scenarios. The measure of resiliency is quantified using Kullback-Leibler divergence and is shown to be high in all scenarios considered.

The resilient control system (RCS) is developed based on two approaches: Model predictive control (MPC)-based approach and synchronous detection (SD)-based approach. In the MPC-based approach, a control input is calculated using the information provided by the RMS. The goal here is to steer the process variable to the desired value, while ensuring that it always remains within a safe domain. In the SD-based approach, the condition of the sensor and actuator is assessed using the method of synchronous detection. Then, using this assessment, the controller is modified (if possible) so that the effects of the attacks on the closed loop system response are eliminated. Using simulations, it is shown that both these approaches are viable for the design of RCS.

Thus, the main contribution of this research is in providing the theoretical foundation for the design of resilient monitoring and control systems applicable to a class of critical infrastructure systems, characterized by complex interactions of continuous process variables.

CHAPTER I

Introduction

Resilient monitoring and control systems is a relatively new area of research. In this section, we briefly characterize these systems and describe the goals of our investigations. In addition, the organization of this dissertation is outlined at the end of this section.

1.1 Resilient Monitoring Systems

Plant monitoring systems are wired or wireless sensor networks intended to measure process variables (e.g., temperature, pressure, flow rates, etc.), analyze them, and inform the plant operator about the plant conditions – normal or anomalous. Based on this information, the operator or the automatic control system takes corrective actions, if needed. When some of the sensors are captured by an attacker, forcing them to project misleading information (possibly, statistically unrelated to the actual values of process variables), the identified plant conditions could be erroneous. This may lead to wrong actions on the part of the operator/control system and, possibly, a disaster. To prevent this situation, the monitoring system must possess a capability of autonomously identifying the attacked sensors and mitigating their effect (by discounting or disregarding completely the data they project). Although the loss of sensors may lead to *degradation* of plant condition assessment, in a well-designed system this degradation should be "proportional" to the severity of the attack, i.e., *graceful*. Plant monitoring systems that possess such a property are referred

to as *resilient*.

This research is intended to develop techniques that can be used to ensure resiliency, analyze their properties and, on this basis, design and evaluate the performance of a resilient monitoring system. A specific application, in terms of which the development is carried out, is a simplified model of a power plant, although a similar approach can be used for other applications as well.

While the designed resilient monitoring system exhibits a high level of resiliency, it exposed a shortcoming of the approach developed – the time required to compute the plant condition assessment increases exponentially with the number of sensors in the sensor network. (This problem was termed by Richard Bellman as the *curse of dimensionality*.) Clearly, the above shortcoming may result in an unacceptably long assessment time in many applications, and, thus, its reduction is a central problem of improving the resilient monitoring system design. This problem is addressed in the dissertation.

1.2 Resilient Control Systems

Resilient control systems are feedback systems that maintain an acceptable level of performance in the presence of attacks on the plant, sensors, and actuators. This research addresses the design of resilient control systems based on two approaches, described below.

The first approach involves the calculation of the resilient control input, using the information provided by the resilient monitoring system. As described in details in a subsequent chapter of the dissertation, this approach is similar to that of *model predictive control* [1].

In the second approach, the condition of the sensors and actuators are first assessed using the method of *synchronous detection* [2]. Then, based on this assessment, the controller is modified (if possible) so that the effects of the attacks on the closed loop system response are eliminated.

1.3 Organization of Dissertation

The remainder of this dissertation is organized as follows: The design, analysis, and performance evaluation of the resilient monitoring system is described in Chapter II. The issue of combating the curse of dimensionality is addressed in Chapter III. The model predictive control approach to resilient feedback systems is presented in Chapter IV. The synchronous detection approach to resilient control systems is described in Chapter V. Finally, the conclusions and directions for future research are given in Chapter VI. All proofs and the parameters involved in simulating the power plant are included in the Appendices.

CHAPTER II

Resilient Monitoring Systems: Architecture, Design, Analysis, and Performance Evaluation

2.1 Introduction

This section describes the specific scenario addressed in resilient monitoring systems, and outlines the techniques developed in this work.

2.1.1 Scenario and problem addressed

Briefly, the scenario considered in this research is as follows:

- The monitored plant process variables, $\mathbf{V}_i$, $i = 1, ..., M$, are characterized by probability density functions (pdf's) $f_{\tilde{V}_i}(\tilde{v}_i)$, $i = 1, ..., M$. In practice, the *status* of the process variables is often characterized as being Normal (N) or Anomalous (A). The latter could be, for instance, Low (L) or High (H). In this case, $f_{\tilde{V}_i}(\tilde{v}_i)$ induces a random event with the outcomes in $\{\mathrm{L}_{V_i}, \mathrm{N}_{V_i}, \mathrm{H}_{V_i}\}$, $i = 1, ..., M$. With a slight abuse of terminology, we refer to this event (and similar events throughout this dissertation) as a discrete random variable, V_i, $i = 1, ...M$, with the probability mass function (pmf), $p[V_i]$, defined on the universal set $\Sigma_{V_i} = \{\mathrm{L}_{V_i}, \mathrm{N}_{V_i}, \mathrm{H}_{V_i}\}$, $i = 1, ..., M$.

- The plant, $\mathbf{G}$, is also characterized by its status, which is a discrete random variable, G, with the pmf $p[G]$ defined by the pmf's of process variables and taking values on

$\Sigma_G = \{\mathrm{N}_G, \mathrm{A}_G\}$, where N_G and A_G denote the normal and anomalous plant statuses, respectively. Depending on the plant, the anomalous status can be further characterized by specific anomalies, e.g., boiler insulation damaged, turbine malfunctioning, etc. In each status, plant dynamics may be different, e.g., described by different transfer functions.

- Each process variable, $\mathbf{V}_i$, is monitored by a sensor, $\mathbf{S}_i$ (multiple sensors of a process variable are also considered in the sequel). If a sensor is under attack, its projected data may have a pdf, $f_{\tilde{S}_i}(\tilde{s}_i)$, statistically unrelated to $f_{\tilde{V}_i}(\tilde{v}_i)$. In this situation, utilizing the sensor data in order to assess the process variable may lead to a pmf, $\hat{p}[V_i]$, qualitatively different from $p[V_i]$. For instance, $\hat{p}[V_i]$ may indicate that the process variable is Normal, while in reality it is Low or High.

- The plant status assessment is based on the process variable assessments, $\hat{p}[V_i]$, $i = 1, ..., M$, and is quantified by a pmf denoted as $\hat{p}[G]$, $G \in \{\mathrm{N}_G, \mathrm{A}_G\}$. Since, as indicated above, the process variable assessments may be erroneous, $\hat{p}[G]$ may be quite different from the actual $p[G]$ and, thus, lead to erroneous actions by the plant operator.

In this scenario, the *optimal* resilient monitoring system must be able to identify the status of the plant, $\mathbf{G}$, in such a manner that the "distance" between the estimated and the actual pmf's, $\hat{p}[G]$ and $p[G]$, is minimized, as quantified by an appropriate measure of distance between the two pmf's. While this research is not intended to solve this problem, here we design a plant monitoring system that degrades gracefully under an attack (i.e., is resilient), and demonstrate that it *performs favorably in comparison with a non-resilient one* (as quantified by a measure of resiliency based on the *Kullback-Leibler divergence* [3]).

2.1.2 Contributions of this work: Techniques developed and resilient monitoring system designed

The techniques developed in this work are as follows:

- The "trustworthiness" of a sensor is quantified by a parameter referred to as *data quality* (DQ), which takes values on $[0, 1]$, with 1 indicating that the sensor is totally trustworthy and 0 not trustworthy at all. To identify DQ, we develop an *active data quality acquisition procedure,* whereby probing signals are applied to process variables, and the level of disagreement between the anticipated and the actual response of the sensors is used to quantify their DQ's.

- The estimates of process variables pmf's, $\hat{p}[V_i]$, $i = 1, ..., M$, are calculated based on the data projected by the sensors and their DQ's. Since DQ is not a statistical quantity, classical statistics cannot be used for this purpose. Therefore, we introduce a model of the DQ's effect on the coupling between sensors data and process variables and, using this model, develop the so-called *h-procedure* (which is a modified stochastic approximation algorithm [4]). Analyzing this procedure, we show that it converges to a steady state defined by the DQ's. Specifically, if $DQ = 1$, it converges to the actual process variable pmf; as DQ tends to 0, the steady state of the h-procedure converges to a uniform pmf, implying that in this limit the sensor measurements carry no information at all. For all other DQ's, the conditional pmf of V_i given the sensor data is an affine function of DQ. When multiple sensors monitor a process variable, the *Dempster-Shafer rule* [5] is used to combine the steady states of the h-procedures associated with each sensor.

- The estimate of the plant status pmf, $\hat{p}[G]$, is calculated based on the statistical plant model (typically given as a set of conditional pmf's $P[V_i|G]$, $i = 1, ..., M$, or a joint conditional pmf $P[V_1, V_2, ..., V_M|G]$), the estimates of the process variables pmf's, $\hat{p}[V_i]$, $i = 1, ..., M$, and the *Jeffrey rule* [6].

- The above assessments are carried out at each state of the sensor network, where the state is a vector of 1's and 0's, with 1 indicating that the corresponding sensor is taken into account for process variable assessment and 0 that it is not. The quality of each state is quantified by the entropy (i.e., the level of uncertainty) of either $\hat{p}[G]$ or $\hat{p}[V_i]$. The adaptation of the sensor network to the optimal state, i.e., the state with the smallest entropy, is carried out using the so-called *rational controllers* [7], which are decision making devices that reside mostly in states, where the penalty function (i.e., entropy) is minimized.

- As mentioned above, the adaptation can be carried out using the entropy of either $\hat{p}[G]$ or $\hat{p}[V_i]$. The former, which we refer to as *centralized*, suffers from the curse of dimensionality: the adaptation time grows exponentially with the number of sensors in the network. To combat this problem, a *decentralized* system (see, e.g., [8–12]), with adaptation based on $\hat{p}[V_i]$, could be used. In the case of a power plant, this decentralized system is comprised of *sub-plants*, e.g., boiler, turbine, reheat pipe, etc. Such a decomposition, however, impedes the derivation of inferences among the sub-plants, which, as it turns out, are important to ensure resiliency. Therefore, we develop a decentralized system based on plant *decomposition with knowledge fusion* and show that it leads to both mitigation of the curse of dimensionality and derivation of the previously mentioned inferences.

The above techniques were introduced in our previous work, [13–19]. Using these techniques, we design a resilient plant monitoring system consisting of the following five layers: data quality acquisition, process variable assessment, adaptation, knowledge fusion, and sub-plant assessment. The subsequent sections describe in details each of the developed techniques, along with its application to the power plant monitoring system.

2.1.3 Related literature

The literature related to the topic of this chapter can be classified into six groups. The first one is devoted to foundational issues, where the problems of resilient monitoring and control are motivated and formulated, [20–25]. The second group includes publications on control-theoretic methods for attack identification and alleviation, [26–31]. In these publications, the authors consider LTI systems with a given state space realization (A, B, C, D) and disturbances interpreted as attack vectors. The problem addressed is to identify the attack and, if possible, mitigate its effect, for instance, by designing a controller that makes the closed-loop system invariant with respect to the disturbance attack. The main difference of the current work is that the plant may be either normal or anomalous (i.e., described by several state space realizations), and the problem is to identify the true plant status, in spite of the misleading information projected by the sensors.

The third group consists of publications on fault tolerant control, [32–34]. In these works, it is assumed that a closed-loop system has multiple sensors and actuators, some of which could be faulty due to natural or malicious causes. The typical problem here is to determine the conditions (e.g., the number of sensors and actuators) under which the closed-loop system performance is maintained without degradation. The difference of the current work is that, although multiple sensors may be present, the goal is to determine the status of the plant and, if otherwise impossible, tolerate degradation.

The fourth group consists of research on monitoring the communication channels or the sensor measurements in order to capture anomalous data and correlate it with a possible attack, [35–40]. In terms of the current work, this implies the identification of DQ. While the results of these publications may be useful for resilient plant monitoring, they do not provide methods for process variable and plant condition assessment pursued in the current work.

The fifth group consists of papers on identification of and protection against data injection attacks intended to mislead state estimation algorithms, [41–47]. The emphasis of

the research here is on determining optimal positions of "known-secure" sensors, which prevent the damage of the attack, or on utilizing game-theoretic approaches as quantitative techniques for risk management.

The sixth group consists of publications on the analysis of vulnerability of the cyber-physical system to attacks, [48–50]. In these papers, tools such as graph theory and discrete event systems theory are utilized to determine "vulnerability points" in the system. However, these works do not provide methods to identify the plant condition under the misleading information projected by the sensors.

Although the areas of robust estimation and robust statistics (see, e.g., [51]) may seem related to the topic of this dissertation, they are, in reality, not, since the data provided by the attacked sensors could be statistically unrelated to the process variable.

To summarize, the current literature does not offer any methods of identifying the plant status under misleading information provided by the sensors. The methods to accomplish that are developed in this dissertation.

2.1.4 Chapter outline

The remainder of this chapter is structured as follows: Section 2.2 addresses the issue of active data quality acquisition. In Section 2.3, the h-procedure and associated techniques for process variable assessment are described. Section 2.4 is devoted to plant pmf assessment. The sensor network adaptation is discussed in Section 2.5, where a practical consequence of the curse of dimensionality is quantified. An approach to combatting the curse of dimensionality based on a decentralized system with knowledge fusion is developed in Section 2.6. The resulting five-layer monitoring system architecture is presented in Section 2.7. An application to a power plant is discussed and investigated by simulations in Section 2.8. All proofs and the parameters of the power plant model are included in the Appendix.

2.2 Active Data Quality Acquisition

In this section, we describe an approach to DQ evaluation briefly mentioned in Subsection 2.1.2.

Consider sensor $\mathbf{S}$ intended to monitor process variable $\mathbf{V}$ and assume that the following holds:

Assumption II.1. (i) Process variable $\mathbf{V}$ is quantified by a continuous random variable $\tilde{V}$, taking values in the domain $\tilde{V} \in [V_{\min}, V_{\max}]$; its pdf, $f_{\tilde{V}}(\tilde{v})$, is unknown.

(ii) The random variable $\tilde{V}$ induces a discrete random variable V, which describes the status of $\mathbf{V}$ and takes values on

$$\Sigma_V = \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\} \tag{2.1}$$

with the pmf given by

$$\begin{aligned} &p[V = \mathrm{L}_V] = \textstyle\int_{V_{\min}}^{R_1} f_{\tilde{V}}(\tilde{v})\, d\tilde{v},\ p[V = \mathrm{N}_V] = \textstyle\int_{R_1}^{R_2} f_{\tilde{V}}(\tilde{v})\, d\tilde{v}, \\ &p[V = \mathrm{H}_V] = \textstyle\int_{R_2}^{V_{\max}} f_{\tilde{V}}(\tilde{v})\, d\tilde{v}, \end{aligned} \tag{2.2}$$

where R_1 and R_2 are known and $V_{\min} < R_1 < R_2 < V_{\max}$ (V's with outcomes other than Low, Normal, and High can be introduced similarly). Since $f_{\tilde{V}}(\tilde{v})$ is unknown, the pmf of V is also unknown.

(iii) The d.c. gain, $\alpha_{\mathbf{V}}$, of $\mathbf{V}$ with respect to its control input, $\mathbf{U}_{\mathbf{V}}$ (e.g., fuel valve of the boiler), depends on the status of $\mathbf{V}$, i.e., whether it is Low, Normal, or High. This is formalized by assuming that $\alpha_{\mathbf{V}}$ is a priori known piecewise constant function of the

expected value of $\tilde{V}$ (denoted as $\mu_{\tilde{V}}$):

$$\alpha_{\mathbf{V}} = \begin{cases} \alpha_{\mathbf{V}}^{\mathrm{L}}, & \text{if } \mu_{\tilde{V}} \in [V_{\min}, R_1) \\ \alpha_{\mathbf{V}}^{\mathrm{N}}, & \text{if } \mu_{\tilde{V}} \in [R_1, R_2) \\ \alpha_{\mathbf{V}}^{\mathrm{H}}, & \text{if } \mu_{\tilde{V}} \in [R_2, V_{\max}]. \end{cases} \tag{2.3}$$

In the case of other than L, N, and H anomalies, $\alpha_{\mathbf{V}}$ is introduced similarly. (Note that we use here the d.c. gain, rather than the full transfer function, in order to require as little information about the plant as possible. Also, various other dependencies of $\alpha_{\mathbf{V}}$ on $\mu_{\tilde{V}}$ can be considered; for instance, $\alpha_{\mathbf{V}}$ could be assumed to be a piecewise linear function of $\mu_{\tilde{V}}$; expression (2.3) is used here for simplicity.)

(iv) The data projected by sensor $\mathbf{S}$ is quantified by a continuous random variable $\tilde{S}$, taking values on $\tilde{S} \in [V_{\min}, V_{\max}]$; its pdf, $f_{\tilde{S}}(\tilde{s})$, can be evaluated using the classical statistical methods (based on the sensor measurements).

(v) The random variable $\tilde{S}$ induces a discrete random variable S taking values on

$$\Sigma_S = \Sigma_V = \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\} \tag{2.4}$$

with the pmf given by

$$\begin{aligned} &p[S = \mathrm{L}_V] = \textstyle\int_{V_{\min}}^{R_1} f_{\tilde{S}}(\tilde{s})\, d\tilde{s},\ p[S = \mathrm{N}_V] = \int_{R_1}^{R_2} f_{\tilde{S}}(\tilde{s})\, d\tilde{s}, \\ &p[S = \mathrm{H}_V] = \textstyle\int_{R_2}^{V_{\max}} f_{\tilde{S}}(\tilde{s})\, d\tilde{s}, \end{aligned} \tag{2.5}$$

where R_1 and R_2 are the same as in (2.2). Since $f_{\tilde{S}}(\tilde{s})$ may be viewed as known, the pmf of S is known as well.

(vi) If $\mathbf{S}$ is not attacked, $\mu_{\tilde{S}} = \mu_{\tilde{V}}$, where $\mu_{\tilde{S}}$ is the expected value of $\tilde{S}$. If $\mathbf{S}$ is under attack, $\mu_{\tilde{S}} \neq \mu_{\tilde{V}}$ and the pmf's of S and V may be qualitatively different; for instance,

$\max_{\sigma\in\Sigma_S} p[S=\sigma]$ may be achieved at $\sigma = \mathrm{L}_V$, while $\max_{\sigma\in\Sigma_V} p[V=\sigma]$ at $\sigma = \mathrm{N}_V$. (The expression $\mu_{\tilde{S}} \neq \mu_{\tilde{V}}$ can be viewed as a definition of the attacker; other types of attackers can be considered as well.)

■

Under Assumption II.1, the active data quality acquisition is carried out as follows: Introduce a probing signal using the control input $\mathbf{U_V}$. Any type of deterministic or random probing signals could be used. Here, we use the simplest probe – a rectangular pulse with amplitude $A_{\mathbf{V}}$ and duration T, applied at the time instant t_0, i.e.,

$$u_{\mathbf{V}}(t) = A_{\mathbf{V}}\mathrm{rect}_T(t-t_0). \tag{2.6}$$

The value of $A_{\mathbf{V}}$ is selected sufficiently small so that $A_{\mathbf{V}} << \min\{[R_1 - V_{\min}], [R_2 - R_1], [V_{\max} - R_2]\}$. The value of T is selected so that $\tilde{V}$ reaches a small vicinity of its steady state defined by the probe.

If the sensor is not under attack, i.e., $\mu_{\tilde{S}} = \mu_{\tilde{V}}$, the following takes place:

$$\mu'_{\tilde{S}} - \mu_{\tilde{S}} = A_{\mathbf{V}}\alpha_{\mathbf{V}}(\mu_{\tilde{S}}), \tag{2.7}$$

where $\mu'_{\tilde{S}}$ is the expected value of $\tilde{S}$ after the probe and $\alpha_{\mathbf{V}}$ is the d.c. gain defined in (2.3). If the sensor is attacked, (2.7) does not hold. In order to quantify the severity of the attack, introduce the notion of *probing inconsistency* $(PIC_{\mathbf{S}})$ defined by:

$$PIC_{\mathbf{S}} := \left|(\mu'_{\tilde{S}} - \mu_{\tilde{S}}) - A_{\mathbf{V}}\alpha_{\mathbf{V}}\left(\mu_{\tilde{S}}\right)\right|. \tag{2.8}$$

Clearly $PIC_{\mathbf{S}} = 0$ implies that the sensor is not attacked; $PIC_{\mathbf{S}} > 0$ indicates an attack and its severity. Given this $PIC_{\mathbf{S}}$, the DQ of sensor $\mathbf{S}$ is defined as:

$$DQ_{\mathbf{S}} = e^{-F(PIC_{\mathbf{S}})}, \tag{2.9}$$

where $F(\cdot)$ is a strictly increasing function of $PIC_{\mathbf{S}}$ with $F(0) = 0$. Note that if $F(PIC_{\mathbf{S}})$ grows too fast, then DQ will be small even for relatively small $PIC_{\mathbf{S}}$'s; if it grows too slow, DQ is relatively large even for large $PIC_{\mathbf{S}}$'s. Our numerical study, reported in [14], indicates that a quadratic $F(\cdot)$ provides better results for subsequent utilization than a linear one. Therefore, we introduce this function as

$$F(PIC_{\mathbf{S}}) := -\frac{\ln \epsilon}{PIC^2_{\max,\mathbf{S}}} PIC^2_{\mathbf{S}}, \tag{2.10}$$

where ϵ is a sufficiently small positive number and $PIC_{\max,\mathbf{S}}$ is the largest value attainable by $PIC_{\mathbf{S}}$. Clearly, due to (2.9) and (2.10), $\min DQ_{\mathbf{S}} = \epsilon$, which can be viewed as a design parameter.

Expressions (2.1)-(2.10) characterize the active DQ acquisition procedure utilized in this work. As mentioned above, numerous modifications of this procedure are possible by considering different properties of V, different types of probing signals and their effect on process variables, various definitions of probing inconsistency, etc. Specific selections may depend on intended applications. The ones used here are motivated by the application to a power plant.

2.3 Process Variable pmf Assessment

In this section, we describe an approach to the evaluation of process variable pmf, $\hat{p}[V]$. As mentioned in Subsection 2.1.2, this pmf is evaluated based on the sensors data and their DQ's. If the DQ were 1, this could be accomplished using classical statistics. However, these methods would lead to erroneous results if $0 \leq DQ < 1$. Therefore, to carry out this evaluation, a model of the effect of DQ on the coupling between V and S must be postulated and then, in the framework of this model, a novel statistical method for pmf's evaluation should be developed. Below, this development is carried out, and methods for pmf evaluation using a single and multiple sensors are introduced.

2.3.1 Model of V and S coupling

We introduce the notion of sensor believability as follows:

$$\beta_{\mathbf{S}} = \frac{|\Sigma_V| - 1}{|\Sigma_V|} DQ_{\mathbf{S}} + \frac{1}{|\Sigma_V|}, \tag{2.11}$$

where $|\Sigma_V|$ is the cardinality of the universal set of V. If, as indicated in (2.1), $|\Sigma_V| = 3$, then

$$\beta_{\mathbf{S}} = \frac{2}{3} DQ_{\mathbf{S}} + \frac{1}{3}.$$

The last two equations imply that when $DQ = 1$, believability is also 1; when $DQ = 0$, believability is $\frac{1}{|\Sigma_V|}$, implying that every status of V is equally likely. Using the believability, introduce

Assumption II.2. The coupling between V and S is as follows:

$$\begin{aligned} P[V = \sigma | S &= \sigma] = \beta_{\mathbf{S}}, \\ P[V = \bar{\sigma} | S &= \sigma] = \tfrac{1-\beta_{\mathbf{S}}}{|\Sigma_V|-1}, \end{aligned} \tag{2.12}$$

where $\bar{\sigma}$ implies 'not σ' and $\sigma, \bar{\sigma} \in \Sigma_V$. ■

Clearly, this implies that if $DQ = 1$, then V has the same status as S with probability 1; if $DQ = 0$, every status of V is equally probable, irrespective of the status of S. The coupling (2.12) is used throughout this work.

2.3.2 Process variable pmf assessment using a single sensor

Consider a sensor $\mathbf{S}$ intended to monitor process variable $\mathbf{V}$. As indicated above, our goal is to evaluate the pmf of V, based on the sensor data, $s_1, s_2, ..., s_n, ...$ (where the subscript

is the time index) and its data quality $DQ_{\mathbf{S}}$. In other words, we are interested in

$$\hat{p}[V=\sigma] = \lim_{n\to\infty} P[V=\sigma|s_1, s_2, ..., s_n; DQ_{\mathbf{S}}],\ \forall\sigma\in\Sigma_V. \tag{2.13}$$

To accomplish this, consider

$$\hat{p}_n[V=\sigma] = P[V=\sigma|s_1, s_2, ..., s_n; DQ_{\mathbf{S}}],\ \forall\sigma\in\Sigma_V, \tag{2.14}$$

and introduce, for convenience, the notation

$$h_\sigma(n) := \hat{p}_n[V=\sigma],\ \forall\sigma\in\Sigma_V.$$

Obviously, the limit of $h_\sigma(n)$, $\forall\sigma\in\Sigma_V$, as $n\to\infty$ (if it exists) is the sought pmf, $\hat{p}[V]$. Define the evolution of $h_\sigma(n)$ as follows:

$$h_\sigma(n+1) = h_\sigma(n) + \epsilon_h\left[h^*_\sigma(s_{n+1}) - h_\sigma(n)\right],\ h_\sigma(0) = \frac{1}{|\Sigma_V|},\ \forall\sigma\in\Sigma_V, \tag{2.15}$$

where the set point, $h^*_\sigma(s_{n+1})$, is given by

$$h^*_\sigma(s_{n+1}) = \begin{cases} \beta_{\mathbf{S}}, & \text{if } s_{n+1}=\sigma \\ \frac{1-\beta_{\mathbf{S}}}{|\Sigma_V|-1}, & \text{if } s_{n+1}\neq\sigma, \end{cases} \tag{2.16}$$

and the step, ϵ_h, is either a small number,

$$0 < \epsilon_h << 1, \tag{2.17}$$

or a function of n, monotonically converging to 0 such that

$$0 < \epsilon_h(n) \leq 1,\ \sum_{n=0}^{\infty}\epsilon_h(n) = \infty,\ \sum_{n=0}^{\infty}\epsilon_h^2(n) < \infty. \tag{2.18}$$

As it follows from (2.16), the evolution of $h_\sigma(n)$ depends on both the sensor data and $DQ_{\mathbf{S}}$ (through $\beta_{\mathbf{S}}$). The system of equations (2.15), (2.16) is referred to as the h-procedure. It can be viewed as a stochastic approximation algorithm [4] with a random set point.

Theorem II.1. *Under Assumptions II.1 and II.2, the recursive procedure* (2.15), (2.16) *converges to*

$$\lim_{n\to\infty} h_\sigma(n) = p[S=\sigma]DQ_{\mathbf{S}} + \frac{1-DQ_{\mathbf{S}}}{|\Sigma_V|},\ \forall\sigma\in\Sigma_V, \tag{2.19}$$

where

1. *The convergence is in probability under* (2.17)*;*

2. *The convergence is almost sure under* (2.18).

Proof. See the Appendix. ■

Equation (2.19) implies that

$$\hat{p}[V=\sigma] = p[S=\sigma]DQ_{\mathbf{S}} + \frac{1-DQ_{\mathbf{S}}}{|\Sigma_V|},\ \forall\sigma\in\Sigma_V. \tag{2.20}$$

Thus, according to the above theorem, if DQ is close to 1, the pmf of process variable, $\hat{p}[V]$, is close to the pmf of the sensor, $p[S]$. However, if DQ is close to 0, the same sensor data result in $\hat{p}[V]$ being practically uniform and independent of the sensor measurements. For all intermediate values of DQ, the pmf $\hat{p}[V]$ is an affine function of DQ.

Recursive procedure (2.15), (2.16) is the basis of process variable assessments used throughout this work.

2.3.3 Process variable pmf assessment using multiple sensors

Assume that process variable $\mathbf{V}$ is monitored by two sensors, $\mathbf{S}_1$ and $\mathbf{S}_2$, having data quality, $DQ_{\mathbf{S}_1}$ and $DQ_{\mathbf{S}_2}$, respectively. The goal is to evaluate $\hat{p}[V]$ based on the data projected

by both sensors, i.e.,

$$\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V=\sigma] = \lim_{n\to\infty} P[V=\sigma|s_1^1,...,s_n^1;DQ_{\mathbf{S}_1};s_1^2,...,s_n^2;DQ_{\mathbf{S}_2}],\ \forall\sigma\in\Sigma_V. \quad (2.21)$$

This can be accomplished by combining the two pmf's, evaluated based on the h-procedure, i.e., $\hat{p}^{\mathbf{S}_1}[V]$ and $\hat{p}^{\mathbf{S}_2}[V]$, into a single pmf, $\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V]$, using the Dempster-Shafer rule [5]:

$$\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V=\sigma] = \frac{\hat{p}^{\mathbf{S}_1}[V=\sigma]\hat{p}^{\mathbf{S}_2}[V=\sigma]}{\sum_{\sigma\in\Sigma_V}\hat{p}^{\mathbf{S}_1}[V=\sigma]\hat{p}^{\mathbf{S}_2}[V=\sigma]},\ \forall\sigma\in\Sigma_V. \quad (2.22)$$

A question arises: Is $\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V]$ "better" than the constituent $\hat{p}^{\mathbf{S}_1}[V]$ and $\hat{p}^{\mathbf{S}_2}[V]$ from the point of view of the uncertainty in the process variable assessment, i.e., entropy? Calculations show that this may or may not be the case (depending on $DQ_{\mathbf{S}_1}$ and $DQ_{\mathbf{S}_2}$). Therefore, having the three estimates $\hat{p}^{\mathbf{S}_1}[V]$, $\hat{p}^{\mathbf{S}_2}[V]$, and $\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V]$, we select as the final estimate, $\hat{p}^*[V]$, the one with the smallest entropy, i.e.,

$$\hat{p}^*[V] = \arg\min\left[I\{\hat{p}^{\mathbf{S}_1}[V]\}, I\{\hat{p}^{\mathbf{S}_2}[V]\}, I\{\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V]\}\right], \quad (2.23)$$

where the entropy is the *Shannon entropy* [52], defined as

$$I\{p[V]\} = -\sum_{\sigma\in\Sigma_V} p[V=\sigma]\log_{|\Sigma_V|} p[V=\sigma]. \quad (2.24)$$

2.4 Plant pmf Assessment

As mentioned in Subsection 2.1.2, the plant status assessment is quantified by $\hat{p}[G]$, $G\in\Sigma_G$. To describe a method for its evaluation, let the plant model be given by $P[V_i|G]$, $i=1,...,M$, and let $\hat{p}[V_i]$, $i=1,...,M$, denote the process variable pmf's evaluated as described in Section 2.3. Then, $\hat{p}[G]$ can be computed using the following:

Algorithm II.1. (a) Assign the initial plant pmf:

$$p_0[G] = \left[\frac{1}{3},\ \frac{1}{3},\ \frac{1}{3}\right]. \tag{2.25}$$

(b) Calculate the initial joint pmf of V_i and G:

$$p_0[V_i, G] = P[V_i|G]p_0[G],\ i = 1, 2, ..., M. \tag{2.26}$$

(c) Calculate the marginal probability:

$$p_0[V_i] = \sum_{G \in \Sigma_G} p_0[V_i, G],\ i = 1, 2, ..., M. \tag{2.27}$$

(d) Apply the Jeffrey rule [6]:

$$\hat{p}[V_i, G] = p_0[V_i, G]\frac{\hat{p}[V_i]}{p_0[V_i]},\ i = 1, 2, ..., M. \tag{2.28}$$

(e) Marginalize to obtain the plant pmf estimate:

$$\hat{p}^{V_i}[G] = \sum_{V_i \in \Sigma_{V_i}} \hat{p}[V_i, G],\ i = 1, 2, .., M. \tag{2.29}$$

(f) If $M > 1$, combine the pmf's obtained in (2.29) using the Dempster-Shafer rule:

$$\hat{p}[G = \sigma_{\mathrm{G}}] = \frac{\prod_{i=1}^{M} \hat{p}^{V_i}[G = \sigma_{\mathrm{G}}]}{\sum_{\sigma_{\mathrm{G}} \in \Sigma_G} \prod_{i=1}^{M} \hat{p}^{V_i}[G = \sigma_{\mathrm{G}}]},\ \sigma_{\mathrm{G}} \in \Sigma_G. \tag{2.30}$$

■

If the plant model is given as $P[V_1, V_2, ..., V_M|G]$, marginalize it to obtain $P[V_i|G]$,

$i = 1, 2, ..., M$, and then follow steps (a)-(f) above.

Algorithm II.1 is carried out after the h-procedure has converged and $\hat{p}[V_i]$, $i = 1, ..., M$, is evaluated. To speed up the process of $\hat{p}[G]$ evaluation, it is tempting to apply this algorithm recursively, i.e., using $\hat{p}_n[V_i]$, instead of $\hat{p}[V_i]$, at step (d). As it turns out, however, this may lead to a paradox: the entropy of $\hat{p}_n[G]$ may tend to 0 as $n \to \infty$, irrespective of the sensors data and their DQ's. This paradox can be explained by the fact that when $\hat{p}_n[V_i]$ approaches its limit (i.e., is practically constant), the dynamics of $\hat{p}_n[G]$ are defined not by the sensor measurements and their DQ's, but by the eigenvalues of the recursive version of Algorithm II.1, defined as follows:

Algorithm II.2. (a) Assign the plant pmf at time n as:

$$\hat{p}_n[G], \text{ where } \hat{p}_0[G] = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]. \tag{2.31}$$

(b) Calculate the joint pmf of V_i and G:

$$\hat{p}_n[V_i, G] = P[V_i|G]\hat{p}_n[G], \ n = 0, 1, 2, ...; \ i = 1, 2, ..., M. \tag{2.32}$$

(c) Calculate the marginal probability:

$$\hat{p}_n^G[V_i] = \sum_{G \in \Sigma_G} \hat{p}_n[V_i, G], \ n = 0, 1, 2, ...; \ i = 1, 2, ..., M. \tag{2.33}$$

(d) Apply the Jeffrey rule:

$$\hat{p}_{n+1}[V_i, G] = \hat{p}_n[V_i, G]\frac{\hat{p}_{n+1}[V_i]}{\hat{p}_n^G[V_i]}, \ n = 0, 1, 2, ...; \ i = 1, 2, ..., M. \tag{2.34}$$

(e) Marginalize to obtain the plant pmf estimate:

$$\hat{p}_{n+1}^{V_i}[G] = \sum_{V_i \in \Sigma_{V_i}} \hat{p}_{n+1}[V_i, G],\ n = 0, 1, 2, ...;\ i = 1, 2, ..., M. \tag{2.35}$$

(f) If $M > 1$, combine the pmf's obtained in (2.35) using the Dempster-Shafer rule:

$$\hat{p}_{n+1}[G = \sigma_{\mathrm{G}}] = \frac{\prod_{i=1}^{M} \hat{p}_{n+1}^{V_i}[G = \sigma_{\mathrm{G}}]}{\sum_{\sigma_{\mathrm{G}}} \prod_{i=1}^{M} \hat{p}_{n+1}^{V_i}[G = \sigma_{\mathrm{G}}]},\ n = 0, 1, 2, ...;\ \sigma_{\mathrm{G}} \in \Sigma_G. \tag{2.36}$$

(g) Update n to $n + 1$. Return to (a).

■

To investigate the performance of this algorithm, consider a plant $\mathbf{G}$ with process variable $\mathbf{V}$, monitored by sensor $\mathbf{S}$. Assume that the universal sets of G, V, and S are given by:

$$\Sigma_G = \{\mathrm{N}_G, \mathrm{A}_G\},\ \Sigma_V = \Sigma_S = \{\mathrm{N}_V, \mathrm{A}_V\}. \tag{2.37}$$

Further, assume that the plant model is characterized by the conditional pmf

$$P[V|G] = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}, \tag{2.38}$$

where $a < 0.5$. Denote the pmf's of the process variable and the plant at time n as

$$\hat{p}_n[V] = [h_{\mathrm{N}_V}(n), h_{\mathrm{A}_V}(n)],\ \hat{p}_n[G] = [k_{\mathrm{N}_G}(n), k_{\mathrm{A}_G}(n)], \tag{2.39}$$

where $h_{\mathrm{N}_V}(n)$ and $h_{\mathrm{A}_V}(n)$ are calculated using the h-procedure (2.15), (2.16) and $k_{\mathrm{N}_G}(n)$ and $k_{\mathrm{A}_G}(n)$ are evaluated using Algorithm II.2. To specify the evolution of $k_{\mathrm{N}_G}(n)$ and

$k_{\mathrm{A}_G}(n)$, substitute (2.38) and (2.39) in steps (a)-(e) of this algorithm to obtain

$$k_{\mathrm{N}_G}(n+1) = \left[\frac{1-a}{C(n)}\right] k_{\mathrm{N}_G}(n) + \left[\frac{a k_{\mathrm{N}_G}(n)}{D(n)} - \frac{[1-a]k_{\mathrm{N}_G}(n)}{C(n)}\right] h_{\mathrm{N}_V}(n+1), \quad (2.40)$$

with $k_{\mathrm{N}_G}(0) = 0.5$ and $C(n)$ and $D(n)$ given by

$$\begin{aligned} C(n) &:= [1-a]k_{\mathrm{N}_G}(n) + a[1-k_{\mathrm{N}_G}(n)], \\ D(n) &:= a k_{\mathrm{N}_G}(n) + [1-a][1-k_{\mathrm{N}_G}(n)]. \end{aligned} \quad (2.41)$$

Denote the steady state values of $h_{\mathrm{N}_V}(n)$ and $h_{\mathrm{A}_V}(n)$, evolving according to the h-procedure (2.15),(2.16), as $h^{\mathrm{ss}}_{\mathrm{N}_V}$ and $h^{\mathrm{ss}}_{\mathrm{A}_V}$, respectively. Then, the steady state values of $k_{\mathrm{N}_G}(n)$ and $k_{\mathrm{A}_G}(n)$ are quantified as follows:

Theorem II.2. *The steady state, $k^{\mathrm{ss}}_{\mathrm{N}_G}$, of the recursion* (2.40) *is characterized by:*

1. $k^{\mathrm{ss}}_{\mathrm{N}_G} = 1$, *if* $h^{\mathrm{ss}}_{\mathrm{N}_V} > 1-a$;
2. $k^{\mathrm{ss}}_{\mathrm{N}_G} = 0$, *if* $h^{\mathrm{ss}}_{\mathrm{N}_V} < a$;
3. $k^{\mathrm{ss}}_{\mathrm{N}_G} = \frac{h^{\mathrm{ss}}_{\mathrm{N}_V} - a}{1-2a}$, *if* $h^{\mathrm{ss}}_{\mathrm{N}_V} > a$ *and* $h^{\mathrm{ss}}_{\mathrm{N}_V} < 1-a$.

Proof. See the Appendix. ■

This theorem exhibits the paradoxical nature of the recursive Jeffrey rule. Namely, if, for instance, $h^{\mathrm{ss}}_{\mathrm{N}_V} = 0.7$, i.e., $\hat{p}[V] = [0.7, 0.3]$, and $a = 0.4$, then, according to Part 1 of Theorem II.2, $\hat{p}[G] = [1, 0]$, implying that the plant status is normal with certainty, while the process variable status is uncertain. Similarly, for the same a, if $h^{\mathrm{ss}}_{\mathrm{N}_V} = 0.3$, i.e., $\hat{p}[V] = [0.3, 0.7]$, then, according to Part 2, $\hat{p}[G] = [0, 1]$, implying that the plant status is anomalous, again with certainty, while the process variable status is uncertain. In other words, this theorem implies that a recursive version of Jeffrey rule may "create erroneous information" rather than transfer it from one quantity, V_i, into another, G.

2.5 Sensor Network Adaptation and Measure of Resiliency

As mentioned in Subsection 2.1.2, the adaptation of sensor network to the state with minimal entropy can be carried out using either the plant or the process variable pmf's. In this section, we describe the former and in Section 2.7 the latter.

2.5.1 Sensor network

Consider the plant $\mathbf{G}$ with M process variables, $\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_M$, monitored by N_S sensors, $\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_{N_S}$, under Assumption II.1. Each sensor may or may not be utilized for the process variable pmf's assessment. This induces the sensor network state space, X, where each state, x, is an N_S-tuple of 1's and 0's, with 1 in the i-th place indicating that $\mathbf{S}_i$ is used for process variable pmf's assessment and 0 that it is not. Thus, the cardinality of the state space, $|X|$, is 2^{N_S}. (A practical consequence of this exponential growth of $|X|$ as a function of N_S is discussed in Subsection 2.5.4.) The process variable pmf's and the plant pmf assessed in state x of the sensor network are denoted as $\hat{p}_x[V_i]$, $i = 1, ..., M$, and $\hat{p}_x[G]$, $x \in X$, respectively. The goal of the sensor network adaptation is to converge to the state, where the entropy of $\hat{p}_x[G]$ is minimal.

2.5.2 Adaptation using a rational controller

As mentioned in Subsection 2.1.2, the adaptation technique used in this work is based on rational controllers introduced in [7] and further developed in [53, 54]. Rational controllers are decision making devices that possess two properties: *ergodicity* and *rationality*. The ergodicity property implies that each state, x, of the decision space, X, is visited with a non-zero probability. The rationality property implies that the residence time in states with a smaller value of the *penalty function* is larger than in those with a larger one. The degree to which this distinction takes place is referred to as the *level of rationality* and quantified by a positive integer, N.

If the sensor network adaptation is based on the plant assessment pmf, $\hat{p}_x[G]$, the

penalty function is selected as its entropy, $I\{\hat{p}_x[G]\} := \hat{I}_x(G)$. Various types of rational controller dynamics can be defined to ensure rationality and ergodicity. In this work, to ensure the former, the following residence time in each state $x \in X$ is introduced:

$$T_x = \begin{cases} T_{\max}, & \text{if } \hat{I}_x(G) \leq \beta \\ \left(\frac{\beta}{\hat{I}_x(G)}\right)^N T_{\max}, & \text{if } \hat{I}_x(G) > \beta, \end{cases} \tag{2.42}$$

where $\beta > 0$ is a small number (design parameter) and $T_{\max}$ is the largest residence time (also a design parameter). To ensure ergodicity, when T_x expires, the controller moves to the next state in a deterministic, round-robin manner.

Let τ_x be the relative residence time in state $x \in X$, i.e.,

$$\tau_x = \frac{T_x}{\sum_{x \in X} T_x}. \tag{2.43}$$

Then, the *average* plant assessment pmf, to be reported to the plant operator after each complete round-robin cycle, is evaluated as

$$\bar{p}[G] = \sum_{x \in X} \tau_x \hat{p}_x[G]. \tag{2.44}$$

It can be shown that if N is sufficiently large, $\bar{p}[G]$ is arbitrarily close to $p^*[G]$ at which $\min_{x \in X} \hat{I}_x(G)$ is attained (see [7]). Note that although under the deterministic, round-robin transition rule, the state with the minimal entropy could be selected by various other methods, we use (2.42)-(2.44) since it is equally applicable to random transitions, which may be necessary in other applications.

2.5.3 Measure of resiliency

The measure of resiliency employed in this work is based on the Kullback-Leibler divergence, [3], of two pmf's, $p_1[G]$ and $p_2[G]$, given by:

$$D\left(p_1[G]||p_2[G]\right) = \sum_{\sigma_G \in \Sigma_G} p_1\left[G = \sigma_G\right] \log_{|\Sigma_G|} \frac{p_1[G = \sigma_G]}{p_2[G = \sigma_G]}. \tag{2.45}$$

Let $p_1[G]$ be the true pmf of the plant, $p[G]$. As for $p_2[G]$, we consider two cases. In the first one, $p_2[G]$ is $\bar{p}[G]$ calculated according to (2.44) and based on the DQ's of the sensors. In the second, $p_2[G]$ is the pmf of the plant assessed under the assumption that the DQ of all sensors is 1; we refer to such a system as *non-resilient* and denote the resulting pmf as $p_{\rm nr}[G]$. Then, the measure of resiliency (MR) considered in this paper is given by

$$MR = \frac{D\left(p[G]||p_{\rm nr}[G]\right) - D\left(p[G]||\bar{p}[G]\right)}{D\left(p[G]||p_{\rm nr}[G]\right)}. \tag{2.46}$$

Clearly, $MR \leq 1$, and the equality is attained when $\bar{p}[G] = p[G]$. Thus, to test the resiliency of a monitoring system, one has to assume that $p[G]$ is known, evaluate $\bar{p}[G]$ and $p_{\rm nr}[G]$, and then use (2.46). This is carried out in Section 2.8 for the case of the power plant.

2.5.4 Temporal properties of adaptation and curse of dimensionality

From the temporal point of view, the adaptation process consists of *epochs*; $|X|$ epochs (where, as before, X is the sensor network state space) comprise a *cycle*; at the end of each cycle, $\bar{p}[G]$ is reported to the plant operator.

For each $x \in X$, the epoch consists of three periods: DQ acquisition (T_{DQ}), process variable(s) and plant pmf evaluation ($T_{\rm eval}$), and residence in state x (T_x). Assuming that the sensor data are provided every 0.01sec and using the procedure described in Section 2.2, T_{DQ} can be evaluated as 5sec (if the time constant of the process variable is 1sec and 100

measurements are utilized to calculate the sensor mean). Using the procedures described in Sections 2.3 and 2.4, the duration of process variable and plant assessment, $T_{\rm eval}$, can be calculated as 6sec (if the stopping rule of the h-procedure is $|h_\sigma(n+1) - h_\sigma(n)| < 10^{-4}$). The maximum residence period, $T_{\max}$, can be selected as desired. If it is selected to be 1sec, the duration of each epoch is less than or equal to 12sec.

As mentioned above, $|X|$ epochs constitute a cycle, so that the cycle duration is, at most, $12|X|$sec. Thus, the resilient monitoring system provides the plant assessment pmf, $\bar{p}[G]$, within a reporting period $T_{\rm report} = 12|X|$sec. If a network consists of 5 sensors, $T_{\rm report} = (2^5)12{\rm sec} \approx 6$min, whereas in a network of 10 sensors, $T_{\rm report} \approx 3$hr, which is clearly unacceptable. This curse of dimensionality is the main drawback of the centralized system based on $\hat{p}_x[G]$ adaptation.

2.6 Decentralized System with Knowledge Fusion

This section provides a method for combatting the curse of dimensionality based on the plant decomposition with knowledge fusion. Note that while the current development is carried out in terms of a power plant, a more general characterization of this method is provided in Chapter III.

2.6.1 Power plant

A simplified model of a power plant is shown in Figure 2.1, where B is the boiler, HT and LT are the high and low pressure turbines, respectively, RP is the reheat pipe, C is the condenser, FP is the feedwater pump, and $\mathbf{S}_{ij}$'s are the sensors. For simplicity, it is assumed that only B, HT, RP, LT may be under a physical attack or malfunction, while C and FP are assumed to operate normally; hence, their sensors are not included in Figure 2.1.

Having 8 sensors, the number of network states is 256. Thus, based on the temporal properties discussed in Subsection 2.5.4, a report to the plant operator could be produced in about every 51min. To combat this drawback, a decentralized system could be considered,

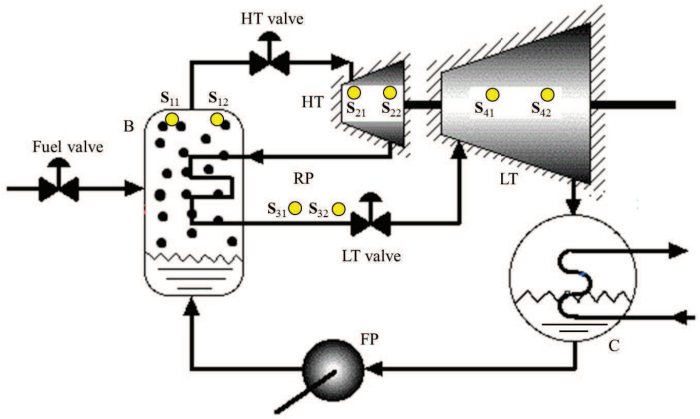
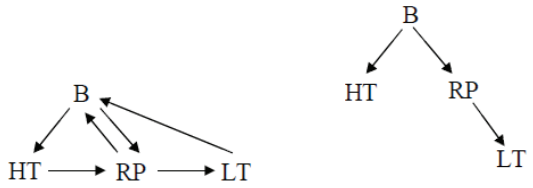


Figure 2.1: Schematics of the power plant

where B, HT, RP, and LT are viewed as separate *sub-plants* monitored by their respective sensor *sub-networks* (i.e., B by sensors $\mathbf{S}_{11}$ and $\mathbf{S}_{12}$, etc.). The problem with such a decentralized system is that inferences arising from coupling of process variables that belong to various sub-plants are neglected. In other words if, for example, all boiler sensors are captured by an attacker, no information about the boiler could be derived, even if all other sensors operate normally. To alleviate this problem, we develop another approach – based, as it is mentioned in Subsection 2.1.2, on a decentralized system with knowledge fusion and show that it leads to reliable and timely plant condition assessments (see Section 2.8).

2.6.2 Developing the decentralized system with knowledge fusion

Assume, for simplicity, that B, HT, RP, and LT are characterized by a single process variable, e.g., its temperature, denoted as $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$, and $\mathbf{V}_4$, respectively, each monitored by two sensors. Mutual influences of the temperature among sub-plants can be represented by a directed *cyclic graph* shown in Figure 2.2(a). Assuming that the heat-generating capacity of B is large enough to maintain RP temperature independent of HT conditions (normal or anomalous), the influence HT $\rightarrow$ RP can be omitted. Similarly, under the above assumption, one may ignore the influence RP $\rightarrow$ B, since B is capable of maintaining its own temperature independent of HT and RP conditions. Further, if the heat-absorbing capacity of C is large enough to maintain a constant water temperature at its outlet independent of



(a) Cyclic graph representation

(b) Tree graph representation

Figure 2.2: Influence diagrams

LT condition, the influence LT $\rightarrow$ B can also be ignored. Under these assumptions, the cyclic graph of Figure 2.2(a) is reduced to the *tree graph* of Figure 2.2(b). This implies that the power plant can be represented as four sub-plants, denoted as $\mathbf{G}_{\mathrm{B}}$, $\mathbf{G}_{\mathrm{HT}}$, $\mathbf{G}_{\mathrm{RP}}$, and $\mathbf{G}_{\mathrm{LT}}$, interrelated as shown in Figure 2.2(b). This partitioning induces a corresponding partitioning of the sensor network $\mathbf{SN}$ into four sub-networks, $\mathbf{SN}_{\mathrm{B}}$, $\mathbf{SN}_{\mathrm{HT}}$, $\mathbf{SN}_{\mathrm{RP}}$, and $\mathbf{SN}_{\mathrm{LT}}$, consisting of $\{\mathbf{S}_{11}, \mathbf{S}_{12}\}$, $\{\mathbf{S}_{21}, \mathbf{S}_{22}\}$, $\{\mathbf{S}_{31}, \mathbf{S}_{32}\}$, and $\{\mathbf{S}_{41}, \mathbf{S}_{42}\}$, respectively. If X_k, $k \in \{\mathrm{B}, \mathrm{HT}, \mathrm{RP}, \mathrm{LT}\}$, denotes the state space of each sub-network, then the number of states in each of them is 4, and, if the evaluation of each state takes 12sec, a report to the operator is produced in approximately 48sec (rather than 51min, as in the centralized case). Clearly, under this decomposition, the aforementioned report would consist of the pmf's of the sub-plants, i.e., $\bar{p}[\mathrm{B}]$, $\bar{p}[\mathrm{HT}]$, $\bar{p}[\mathrm{RP}]$, and $\bar{p}[\mathrm{LT}]$, rather than of a single pmf $\bar{p}[G]$.

Note that in this decentralized architecture, the sensor sub-networks adaptation is carried out based on $\hat{p}[V_i]$ (rather than $\hat{p}[G]$). This is because $\hat{p}[G_i]$, $i \in \{\mathrm{B}, \mathrm{HT}, \mathrm{RP}, \mathrm{LT}\}$, becomes available only after the knowledge fusion of $\hat{p}[V_i]$'s is carried out.

To implement knowledge fusion calculations, couplings among process variables must be introduced. This is accomplished based on the conditional probabilities $P[V_i|V_j]$. While specific matrices representing these conditional pmf's are given in Subsection 2.8.1, below we describe the knowledge fusion calculations used in this work.

2.6.3 Knowledge fusion calculations

Let $\bar{p}_{\mathbf{G}_{\text{B}}}[V_1]$, $\bar{p}_{\mathbf{G}_{\text{HT}}}[V_2]$, $\bar{p}_{\mathbf{G}_{\text{RP}}}[V_3]$, and $\bar{p}_{\mathbf{G}_{\text{LT}}}[V_4]$ be the process variable pmf's of the sub-plants, evaluated using the techniques described in Sections 2.2, 2.3, and 2.5. Then, fusion of this information, leading to the sought inferences, is carried out as follows:

Algorithm II.3. Inferences for $\mathbf{V}_1$:

(a) Calculate the pmf of V_1 based on the sensors of LT (denoted as $\bar{p}_{\mathbf{G}_{\text{LT}}}[V_1]$):

$$\bar{p}_{\mathbf{G}_{\text{LT}}}[V_1] \;=\; \sum_{\sigma_3 \in \Sigma_{V_3}} P[V_1|V_3=\sigma_3]\bar{p}_{\mathbf{G}_{\text{LT}}}[V_3=\sigma_3], \tag{2.47}$$

where $\bar{p}_{\mathbf{G}_{\text{LT}}}[V_3]$ is calculated as

$$\bar{p}_{\mathbf{G}_{\text{LT}}}[V_3] = \sum_{\sigma_4 \in \Sigma_{V_4}} P[V_3|V_4=\sigma_4]\bar{p}_{\mathbf{G}_{\text{LT}}}[V_4=\sigma_4]. \tag{2.48}$$

(b) Calculate the pmf of V_1 based on the sensors of RP:

$$\bar{p}_{\mathbf{G}_{\text{RP}}}[V_1] = \sum_{\sigma_3 \in \Sigma_{V_3}} P[V_1|V_3=\sigma_3]\bar{p}_{\mathbf{G}_{\text{RP}}}[V_3=\sigma_3]. \tag{2.49}$$

(c) Calculate the pmf of V_1 based on the sensors of HT:

$$\bar{p}_{\mathbf{G}_{\text{HT}}}[V_1] = \sum_{\sigma_2 \in \Sigma_{V_2}} P[V_1|V_2=\sigma_2]\bar{p}_{\mathbf{G}_{\text{HT}}}[V_2=\sigma_2]. \tag{2.50}$$

(d) Calculate the pmf of V_1 based on all sensors of the sensor network (using the Dempster-Shafer rule):

$$\bar{p}_{\mathbf{G}_{\text{B,HT,RP,LT}}}[V_1=\sigma_1] = \frac{\prod\limits_{k=\text{B,HT,RP,LT}} \bar{p}_{\mathbf{G}_k}[V_1=\sigma_1]}{\sum\limits_{\sigma_1 \in \Sigma_{V_1}} \prod\limits_{k=\text{B,HT,RP,LT}} \bar{p}_{\mathbf{G}_k}[V_1=\sigma_1]}. \tag{2.51}$$

(e) Finally, select $\bar{p}^*[V_1]$ as the one of the five pmf's obtained above, which has the smallest entropy, i.e.,

$$\bar{p}^*[V_1] = \arg\min \left\{ I\left\{\bar{p}_{\mathbf{G}_{\mathrm{B}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{HT}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{RP}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\mathrm{B,HT,RP,LT}}}[V_1]\right\}\right\}. \tag{2.52}$$

■

Fusion of other process variable pmf's is carried out similarly, leading to $\bar{p}^*[V_2]$, $\bar{p}^*[V_3]$, and $\bar{p}^*[V_4]$.

2.7 Decentralized Resilient Monitoring System for Power Plant

Turning now to the issue of computing the pmf's of B, HT, RP, and LT, we introduce a five-layer architecture shown in Figure 2.3. It consists of four parallel sub-architectures, each corresponding to a sub-plant, $\mathbf{G}_{\mathrm{B}}$, $\mathbf{G}_{\mathrm{HT}}$, $\mathbf{G}_{\mathrm{RP}}$, and $\mathbf{G}_{\mathrm{LT}}$, which could be under a physical attack (or malfunction). The inputs to each sub- architecture are the sensor data provided by the sub-networks $\mathbf{SN}_{\mathrm{B}}$, $\mathbf{SN}_{\mathrm{HT}}$, $\mathbf{SN}_{\mathrm{RP}}$, and $\mathbf{SN}_{\mathrm{LT}}$, which could be under a cyber attack. The physical and cyber attacks might be either coordinated or not. The outputs of the overall architecture are the assessed sub-plant pmf's, i.e., $\bar{p}[\mathrm{B}]$, $\bar{p}[\mathrm{HT}]$, $\bar{p}[\mathrm{RP}]$, $\bar{p}[\mathrm{LT}]$.

The five layers of this architecture can be characterized as follows (using the sub-plant B, as an example):

- The DQ acquisition layer remains the same as in Section 2.2.
- The process variable assessment layer consists of two parts. The first one represents the evaluation of $\hat{p}_{x_{\mathrm{B}}}[V_1]$ using the methods of Section 2.3. The second part evaluates $\bar{p}_{\mathbf{G}_{\mathrm{B}}}[V_1]$ using the expression (2.44) applied to the sub-plant (i.e., $\bar{p}_{\mathbf{G}_{\mathrm{B}}}[V_1] = \sum_{x_{\mathrm{B}} \in X_{\mathrm{B}}} \tau_{x_{\mathrm{B}}} \hat{p}_{x_{\mathrm{B}}}[V_1]$, where $\tau_{x_{\mathrm{B}}}$ is the output of the adaptation layer).

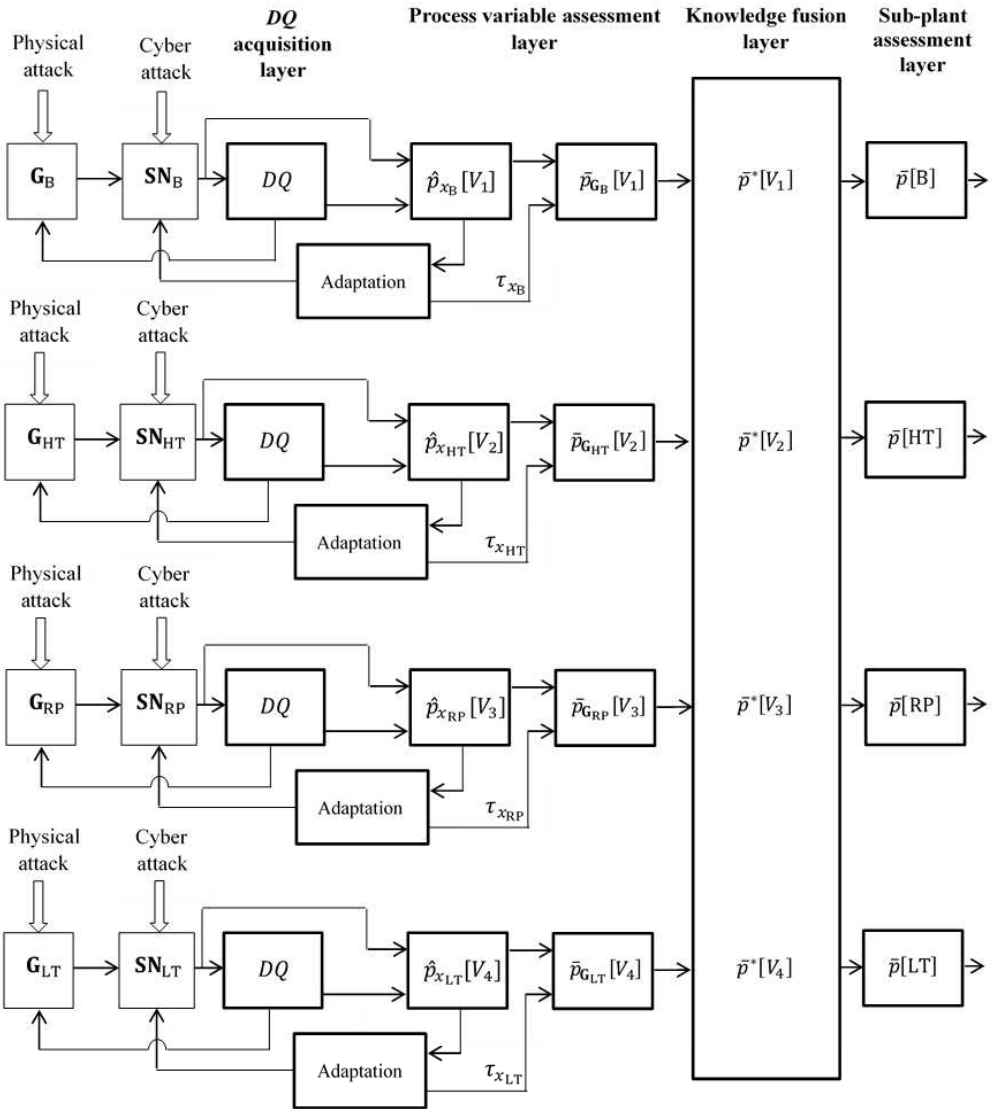


Figure 2.3: Five-layer resilient monitoring system architecture based on decentralization with knowledge fusion

- The sub-network adaptation layer operates as described in Section 2.5, but using the entropy of $\hat{p}_{x_{\mathrm{B}}}[V_1]$ as the penalty function.

- The knowledge fusion layer implements the calculations described in Subsection 2.6.3.

- The sub-plant assessment layer evaluates $\bar{p}[\mathrm{B}]$, $\bar{p}[\mathrm{HT}]$, $\bar{p}[\mathrm{RP}]$, and $\bar{p}[\mathrm{LT}]$ using the technique of Section 2.4.

The measure of resiliency is evaluated using (2.46) applied separately to each sub-plant, e.g.,

$$MR_{\text{B}} = \frac{D\left(p[B]||p_{\text{nr}}[B]\right) - D\left(p[B]||\bar{p}[B]\right)}{D\left(p[B]||p_{\text{nr}}[B]\right)}. \tag{2.53}$$

The MR's for HT, RP, and LT are computed similarly, resulting in the following vector:

$$\overrightarrow{MR} = \left[MR_{\text{B}},\ MR_{\text{HT}},\ MR_{\text{RP}},\ MR_{\text{LT}}\right]. \tag{2.54}$$

Based on the calculations of Subsection 2.5.4, the assessment time in each of the sub-architectures of Figure 2.3 is $(12\text{sec})2^2 = 48\text{sec}$. Note that the centralized assessment of this plant, having 8 sensors, would be $(12\text{sec})2^8 = 3072\text{sec} = 51.2\text{min}$.

2.8 Performance Evaluation of Decentralized Resilient Monitoring System with Knowledge Fusion

In this section, we apply the resilient monitoring system of Figure 2.3 to the power plant of Figure 2.1. While the statistics of process variables and the parameters of the monitoring system are specified in the Appendix, below we introduce the sub-plant anomalies (Subsection 2.8.1), describe the attack scenarios and the resulting system performance (Subsection 2.8.2), and discuss qualitative features of the results obtained (Subsection 2.8.3).

2.8.1 Sub-plant anomalies and process variable coupling

2.8.1.1 Boiler

The anomaly of B is insulation fracture. Since the fracture results in a lower than normal temperature, the universal set of V_1 is $\Sigma_{V_1} := \{\text{L}_{\text{V}_1}, \text{N}_{\text{V}_1}\}$.

2.8.1.2 High pressure turbine

The anomaly of HT is also the insulation fracture. Taking into account the influence B $\rightarrow$ HT, we assume that V_2 takes progressively increasing values under the following conditions: Both B and HT are damaged; only B is damaged; only HT is damaged; and both B and HT operate normally. As it follows from the above, the universal set of V_2 is $\Sigma_{V_2} := \{\mathrm{VL_{V_2}}, \mathrm{L_{(1)V_2}}, \mathrm{L_{(2)V_2}}, \mathrm{N_{V_2}}\}$, where VL stands for Very Low, and $\mathrm{L_{(1)V_2}}$ and $\mathrm{L_{(2)V_2}}$ indicate Low HT temperature due to B and HT damage, respectively.

2.8.1.3 Reheat pipe

The anomaly of RP is similar to that of B and HT, i.e., the insulation fracture. Regarding V_3, we assume that it takes increasing values under the following conditions: Both B and RP are damaged; only B is damaged; only RP is damaged; and both B and RP operate normally. From the above, $V_3 \in \Sigma_{V_3} := \{\mathrm{VL_{V_3}}, \mathrm{L_{(1)V_3}}, \mathrm{L_{(2)V_3}}, \mathrm{N_{V_3}}\}$.

2.8.1.4 Low pressure turbine

Since LT operates at a low pressure, we assume that the anomaly is not due to the fracture of its insulation, but due to the inefficient transfer of energy to the output shaft, leading to the temperature being higher than normal. Taking into account the chain of influences B $\rightarrow$ RP $\rightarrow$ LT and the above assumption, V_4 takes progressively increasing values under the following conditions: LT operates normally, while RP and B are damaged; LT malfunctions, while RP and B are damaged; LT and RP operate normally, while B is damaged; LT malfunctions and B is damaged, while RP operates normally; LT and B operate normally, while RP is damaged; LT malfunctions and RP is damaged, while B operates normally; LT, RP, and B operate normally; and LT malfunctions, while RP and B operate normally. As it follows from the above, $V_4 \in \Sigma_{V_4} := \{\mathrm{VL_{(1)V_4}}, \mathrm{VL_{(2)V_4}}, \mathrm{L_{(1)V_4}}, \mathrm{L_{(2)V_4}}, \mathrm{M_{(1)V_4}}, \mathrm{M_{(2)V_4}}, \mathrm{N_{V_4}}, \mathrm{H_{V_4}}\}$, where M stands for Medium and H for High.

2.8.1.5 Coupling of process variables

As described in Subsection 2.6.2, the couplings of the process variables are characterized by the conditional pmf's $P[V_i|V_j]$. Taking into account the universal sets introduced above, these pmf's are as follows:

$$P[V_1|V_2] = P[V_1|V_3] = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}}_{A}, \; P[V_2|V_1] = P[V_3|V_1] = \underbrace{\begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}}_{B}, \tag{2.55}$$

$$P[V_3|V_4] = \begin{bmatrix} A & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{2\times 4} & A \end{bmatrix}, \; P[V_4|V_3] = \begin{bmatrix} B & \mathbf{0}_{4\times 2} \\ \mathbf{0}_{4\times 2} & B \end{bmatrix}, \tag{2.56}$$

where the matrices A and B are given in (2.55).

2.8.1.6 Universal sets of the sub-plants

Since each sub-plant is characterized by a single anomaly, the random variable G_i, $i \in \{\mathrm{B}, \mathrm{HT}, \mathrm{RP}, \mathrm{LT}\}$, which represents its status, has the universal set comprised of two outcomes, $\{\mathrm{N}_{G_i}, \mathrm{A}_{G_i}\}$, $i \in \{\mathrm{B}, \mathrm{HT}, \mathrm{RP}, \mathrm{LT}\}$, where, as before, N_{G_i} and A_{G_i} stand for normal and anomalous status of the sub-plant $\mathbf{G}_i$, respectively.

2.8.2 Attack scenarios and the resulting monitoring system performance

In this section, we introduce seven cyber and cyber-physical attack scenarios selected so as to exhibit the main features of the resilient monitoring system designed herein. As it may be expected, physical attacks on the sub-plants are less damaging for resilient monitoring than cyber attacks on the sensors. Nevertheless, to illustrate that every sub-plant

status (normal or anomalous) can be identified with or without a physical attack, we include cyber-physical attacks into consideration as well.

Scenario 1: *Cyber attack on the boiler:* All sub-plants operate normally. All sensors monitoring B are captured and project misleading information that the boiler is damaged. All other sensors operate normally.

Performance: The resilient monitoring system computes the following pmf's:

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\text{LT}}] = [0.92, 0.08],\end{aligned} \tag{2.57}$$

correctly indicating that all sub-plants operate normally with large probability. The non-resilient monitoring system (i.e., the system with DQ's of all sensors equal to 1 – see Subsection 2.5.3) evaluates the pmf of B as $p_{\text{nr}}[G_{\text{B}}] = [0.05, 0.95]$, erroneously indicating that the boiler is damaged. Using (2.53) and (2.54), the measure of resiliency under this scenario is calculated as $\overrightarrow{MR} = [0.98, -, -, -]$, where "$-$" indicates that none of the sensors of the corresponding sub-plant are attacked.

Scenario 2: *Cyber attack on the low pressure turbine:* All sub-plants operate normally. All sensors of LT are under attack, reporting that it is malfunctioning. All other sensors operate normally.

Performance: The resilient monitoring system computes the following pmf's:

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\text{LT}}] = [0.49, 0.51],\end{aligned} \tag{2.58}$$

implying that, while the status of B, HT, and RP is ascertained correctly, the status of LT is undetermined (i.e., either normal or anomalous with almost equal probabilities). The non-resilient monitoring system evaluates the pmf of LT as $p_{\text{nr}}[G_{\text{LT}}] = [0.09, 0.91]$, erroneously

indicating that LT is malfunctioning. The measure of resiliency in this case is $\overrightarrow{MR} = [-, -, -, 0.7]$. Note, however, that if only one sensor of LT was captured, the status of all sub-plants would be assessed correctly with the pmf's

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],$$
$$\bar{p}[G_{\text{LT}}] = [0.91, 0.09].$$

Scenario 3: *Coordinated cyber-physical attack on the reheat pipe:* RP is under attack, resulting in insulation fracture. All other sub-plants operate normally. Since RP is attacked, the temperature of LT is $\text{M}_{(1)\text{V}_4}$. All sensors of RP are captured, forcing them to indicate that RP is normal. All other sensors are not attacked.

Performance: The pmf's of B, HT, RP, and LT are computed as follows:

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.12, 0.88],\\ &\bar{p}[G_{\text{LT}}] = [0.92, 0.08],\end{aligned} \tag{2.59}$$

correctly identifying the status of all sub-plants. The non-resilient monitoring system evaluates the pmf of RP as $p_{\text{nr}}[G_{\text{RP}}] = [0.91, 0.09]$, i.e., erroneously. The measure of resiliency is $\overrightarrow{MR} = [-, -, 0.95, -]$. Note that if the attack was not coordinated, e.g., physical attack on RP and cyber attack, say, on LT, the status of LT would be undetermined, i.e.,

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.12, 0.88],$$
$$\bar{p}[G_{\text{LT}}] = [0.49, 0.51].$$

Scenario 4: *Coordinated cyber-physical attack on the high pressure turbine:* HT is under attack, resulting in fracture of its insulation, with V_2 being $\text{L}_{(2)\text{V}_2}$. All other sub-plants operate normally. All sensors of HT are captured, forcing them to indicate that its status is normal. All other sensors are not attacked.

Performance: The pmf's of the sub-plants are computed as follows:

$$\begin{aligned}&\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.51, 0.49],\ \bar{p}[G_{\mathrm{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\mathrm{LT}}] = [0.92, 0.08],\end{aligned} \tag{2.60}$$

correctly identifying the status of B, RP, and LT, while that of HT is undetermined. The non-resilient monitoring system evaluates the pmf of HT as $p_{\mathrm{nr}}[G_{\mathrm{HT}}] = [0.9, 0.1]$, i.e., erroneously indicating that HT is normal. The measure of resiliency is $\overrightarrow{MR} = [-, 0.69, -, -]$. If only one sensor of HT was captured, the status of all sub-plants would be ascertained correctly with the pmf's

$$\begin{aligned}&\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.11, 0.89],\ \bar{p}[G_{\mathrm{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\mathrm{LT}}] = [0.92, 0.08].\end{aligned}$$

If the attack was not coordinated, e.g., a physical attack on HT and a cyber attack on all sensors of B, the resulting performance would be

$$\begin{aligned}&\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.1, 0.9],\ \bar{p}[G_{\mathrm{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\mathrm{LT}}] = [0.92, 0.08],\end{aligned}$$

indicating that all sub-plants are assessed correctly.

Scenario 5: *Coordinated cyber-physical attack on the boiler and low pressure turbine:* B and LT are under attack, resulting in insulation damage of the former and malfunctioning of the latter, with V_1 being $\mathrm{L}_{\mathrm{V}_1}$ and V_4 being $\mathrm{L}_{(2)\mathrm{V}_4}$. All other sub-plants operate normally, with V_2 being $\mathrm{L}_{(1)\mathrm{V}_2}$ and V_3 being $\mathrm{L}_{(1)\mathrm{V}_3}$. All sensors of B and LT are captured, forcing them to indicate that their status is normal. All other sensors are not attacked.

Performance: The pmf's of the sub-plants are computed as follows:

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.05, 0.95],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\text{LT}}] = [0.51, 0.49],\end{aligned} \tag{2.61}$$

correctly identifying the status of B, HT, and RP, while the status of LT is undetermined. The non-resilient monitoring system evaluates the pmf's of B and LT as $p_{\text{nr}}[G_{\text{B}}] = [0.95, 0.05]$ and $p_{\text{nr}}[G_{\text{LT}}] = [0.92, 0.08]$, erroneously assessing them as normal. The measure of resiliency is $\overrightarrow{MR} = [0.98, -, -, 0.72]$. If only one sensor of LT was captured, the status of all sub-plants would be ascertained correctly with the pmf's

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.05, 0.95],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\text{LT}}] = [0.1, 0.9].\end{aligned}$$

Note also that if the attack was not coordinated, e.g., physical attack on LT and cyber attack on all sensors of B, the resulting performance would be

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\\ &\bar{p}[G_{\text{LT}}] = [0.09, 0.91],\end{aligned}$$

indicating that all sub-plants are assessed correctly.

Scenario 6: *Coordinated cyber-physical attack on the boiler, reheat pipe, and low pressure turbine:* B, RP, and LT are under attack, with V_1, V_3, and V_4 being L_{V_1}, VL_{V_3}, and $\text{VL}_{(2)\text{V}_4}$, respectively. The remaining sub-plant, HT, operates normally. All sensors that monitor B, RP, and LT are captured, forcing them to indicate that their status is normal. The sensors of HT are not attacked.

Performance: The pmf's of the sub-plants are computed as follows:

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.05, 0.95], \bar{p}[G_{\text{HT}}] = [0.9, 0.1], \bar{p}[G_{\text{RP}}] = [0.51, 0.49],\\&\bar{p}[G_{\text{LT}}] = [0.5, 0.5],\end{aligned} \tag{2.62}$$

correctly identifying the status of B and HT, while the status of RP and LT is undetermined. The non-resilient monitoring system evaluates the pmf's of B, RP, and LT as $p_{\text{nr}}[G_{\text{B}}] = [0.95, 0.05]$, $p_{\text{nr}}[G_{\text{RP}}] = [0.9, 0.1]$, and $p_{\text{nr}}[G_{\text{LT}}] = [0.92, 0.08]$, erroneously assessing them as normal. The measure of resiliency is $\overrightarrow{MR} = [0.98, -, 0.7, 0.72]$. If only one sensor of LT was captured, the status of all sub-plants would be ascertained correctly with the pmf's

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.05, 0.95], \bar{p}[G_{\text{HT}}] = [0.9, 0.1], \bar{p}[G_{\text{RP}}] = [0.12, 0.88],\\&\bar{p}[G_{\text{LT}}] = [0.09, 0.91].\end{aligned}$$

If the attack was not coordinated, e.g., physical attack on LT and all sensors of B and RP being captured, the status of all sub-plants would be assessed correctly with the pmf's

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.95, 0.05], \bar{p}[G_{\text{HT}}] = [0.9, 0.1], \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\\&\bar{p}[G_{\text{LT}}] = [0.09, 0.91].\end{aligned}$$

Scenario 7: *Coordinated cyber-physical attack on all sub-plants:* All sub-plants are attacked, resulting in their anomalous operation. All sensors are captured, forcing them to indicate that their status is normal.

Performance: The status of all sub-plants is undetermined with the pmf's being close to $[0.5, 0.5]$. The non-resilient monitoring system evaluates erroneously that all sub-plants are normal. The measure of resiliency is $\overrightarrow{MR} = [0.76, 0.7, 0.7, 0.72]$. If one sensor of HT was not captured, the pmf's of the sub-plants would be

$$\begin{aligned}&\bar{p}[G_{\text{B}}] = [0.05, 0.95], \bar{p}[G_{\text{HT}}] = [0.1, 0.9], \bar{p}[G_{\text{RP}}] = [0.5, 0.5],\\&\bar{p}[G_{\text{LT}}] = [0.5, 0.5],\end{aligned}$$

i.e., B and HT are assessed correctly, while RP and LT are undetermined. If one sensor of HT and one sensor of LT were not captured, the pmf's of the sub-plants would be

$$\bar{p}[G_{\text{B}}] = [0.05, 0.95], \bar{p}[G_{\text{HT}}] = [0.1, 0.9], \bar{p}[G_{\text{RP}}] = [0.12, 0.88],$$
$$\bar{p}[G_{\text{LT}}] = [0.09, 0.91],$$

i.e., all are assessed correctly.

2.8.3 Discussion

The above results lead to the following conclusions:

- Under all attack scenarios considered, *the resilient monitoring system provides no erroneous assessments* (as insinuated by the attacker).

- As evidenced by Scenarios 1-4, *cyber attacks on HT and LT are more dangerous than those on B and RP*. This is due to the structure of the conditional probability matrices (2.55), which permit inferences from HT and LT to B and RP, but not vice-versa. In other words, cyber-attacking the terminal nodes of the graph of Figure 2.2(b) is more dangerous than attacking the initial and/or intermediate ones.

- As evidenced from Scenarios 3 and 4, *coordinated cyber-physical attacks may not be more dangerous than non-coordinated ones*. More important is not the coordination, but the nature of a cyber attack $-$ involving or not the terminal nodes of the graph.

- As follows from Scenario 7, the *minimum number of non-attacked sensors necessary and sufficient to correctly assess all sub-plants is* 2*: one for HT and one for LT*. If these sensors were made "known-secure", the plant assessment would never be compromised.

- In all cases considered, *the measure of system resiliency is quite high:* from 0.69

(when some sub-plants status remains undetermined) to close to 1 (when all sub-plants status is assessed with certainty).

CHAPTER III

Combating Curse of Dimensionality in Resilient Monitoring Systems: Conditions for Lossless Decomposition

3.1 Introduction

As described in Chapter II, the adaptation of the sensor network can be carried out either in a centralized or decentralized manner. The former, which is applied in our previous work [17], suffers from the curse of dimensionality, namely, the assessment time of the plant condition, T_a, behaves as

$$T_a = \kappa 2^{N_{\mathbf{SN}}}, \tag{3.1}$$

where the pre-exponential factor, κ, depends on the assessment algorithms involved, and $N_{\mathbf{SN}}$ is the number of sensors in the network. This implies that even if κ is relatively small, say, $\kappa = 1$sec, T_a is 17min if $N_{\mathbf{SN}} = 10$ and 12 days if $N_{\mathbf{SN}} = 20$. Clearly, such a long assessment time may be unacceptable in most applications. To address this shortcoming, the decentralized system was introduced in Section 2.7, wherein the development was car-

ried out in terms of a power plant application. While the resulting decentralized resilient monitoring system was shown to provide both timely and reliable assessments, a rigorous analysis of the developed approach was not provided. The current chapter is devoted to this issue.

The decentralized system is based on a decomposition of the sensor network into subnetworks, each monitoring a small subset of plant process variables. In the sequel, we assume that each of these subnetworks monitors a single process variable, although the case of subnetworks monitoring a group of process variables can be considered similarly. Thus, if a plant has M process variables (e.g., temperatures, pressures, flow rates, etc.), the sensor network, $\mathbf{SN}$, is decoupled into M subnetworks, $\mathbf{SN}_1, ..., \mathbf{SN}_M$, leading to the assessment time in each subnetwork given by

$$T_{a,i} = \kappa_i 2^{N_{\mathbf{SN}_i}}, \tag{3.2}$$

where $N_{\mathbf{SN}_i}$ is the number of sensors monitoring the process variable i. Therefore, even if the pre-exponential factor is somewhat increased (i.e., $\kappa_i > \kappa$, $\forall i$), the assessment time would decrease substantially if $N_{\mathbf{SN}_i} << N_{\mathbf{SN}}$, $\forall i$. For instance, if $N_{\mathbf{SN}_i} \leq 2$, $\forall i$, and $\kappa_i = 1.5$sec, $\forall i$, the process variable assessment time is less than 6sec, irrespective of $N_{\mathbf{SN}}$. In other words, the assessment of each process variable could be carried out sufficiently rapidly, and the resulting information could be used for plant condition assessment practically instantaneously (based on the algorithm provided in Chapter II).

Clearly, this decomposition may reduce the quality of process variable and plant condition assessment. For example, if all sensors monitoring a process variable are attacked, no assessment of its state would be made. To avoid this deficiency, we employ the so-called *decentralized inference calculations* (or, as it is termed in Chapter II, knowledge fusion calculations), whereby mutual influences of process variables are taken into account. In terms of the power plant, this implies that even if all boiler sensors are attacked, the sensors of the

other components may be used to infer information about the boiler status. The question arises: Under which conditions this decomposition leads to no information losses, as compared with the *centralized inference calculations* utilized in [17]? The main contribution of this chapter is in providing an answer to this question.

The outline of this chapter is as follows: Section 3.2 introduces the model considered and formalizes the problem addressed. In Section 3.3, the algorithms used in the centralized and the decentralized process variable assessments are described. Section 3.4 provides a sufficient condition under which the decentralized inference calculations lead to no information losses as compared with the centralized ones. Finally, in Section 3.5, a sufficient condition for lossless decentralization is derived.

3.2 Model and Problem Formulation

Consider the plant $\mathbf{G}$ with process variables $\mathbf{V}_i$, $i = 1, ..., M$, each viewed as a random variable, V_i, with the universal set $\Sigma_{V_i} = \{\mathrm{N}_{V_i}, \mathrm{A}_{V_i,1}, ..., \mathrm{A}_{V_i,n_i-1}\}$, where N_{V_i} stands for Normal and $\mathrm{A}_{V_i,l}$ for an anomaly of type l (induced either by a physical attack or malfunction), and n_i is the cardinality of Σ_{V_i}. The coupling among the process variables is characterized by a set of conditional probabilities $P[V_i|V_j]$, $i \neq j$, $i, j = 1, ..., M$.

The plant $\mathbf{G}$ is monitored by the sensor network $\mathbf{SN}$ comprised of $N_{\mathbf{SN}}$ sensors, which could be either under a cyber-physical attack or malfunction. The $\mathbf{SN}$ can be viewed as a set of subnetworks, $\mathbf{SN}_i$, $i = 1, ..., M$, each monitoring the process variable $\mathbf{V}_i$ and consisting of $N_{\mathbf{SN}_i}$ sensors, so that $\sum_{i=1}^{M} N_{\mathbf{SN}_i} = N_{\mathbf{SN}}$. Since the state space, X, of $\mathbf{SN}$ consists of vectors comprised of 1's and 0's, the cardinality of X is $2^{N_{\mathbf{SN}}}$. Similarly, the state space of $\mathbf{SN}_i$ is Y_i with the cardinality $2^{N_{\mathbf{SN}_i}}$, $i = 1, ..., M$. Clearly, X can be viewed as the Cartesian product of Y_i's,

$$X = Y_1 \times Y_2 \times \cdots Y_M, \tag{3.3}$$

and each state $x \in X$ can be viewed as the ordered concatenation of the states $y_i \in Y_i$, i.e.,

$$x = (y_1, y_2, ..., y_M). \tag{3.4}$$

Given this model, the centralized and the decentralized assessments of process variables V_i, $i = 1, ..., M$, can be symbolically represented as follows:

Centralized:

$$\left\{\hat{p}_{y_i}[V_i] \otimes \hat{p}_{y_j}[V_i],\ \forall j \neq i\right\} \Rightarrow \hat{p}_x[V_i] \xRightarrow{\text{optimization over } X} \hat{p}_{x_i^*}[V_i]. \tag{3.5}$$

Decentralized:

$$\hat{p}_{y_i}[V_i] \xRightarrow{\text{optimization over } Y_i} \left\{\hat{p}_{y_i^*}[V_i] \otimes \hat{p}_{y_j^*}[V_i],\ \forall j \neq i\right\} \Rightarrow \hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i]. \tag{3.6}$$

The notations involved in these expressions are:

Centralized:

- $\hat{p}_{y_i}[V_i]$ is the probability mass function (pmf) of V_i, estimated when $\mathbf{SN}_i$ is in the state $y_i \in Y_i$.
- $\hat{p}_{y_j}[V_i]$ is the centralized inferred pmf of V_i, calculated when $\mathbf{SN}_j$ is in the state $y_j \in Y_j$.
- Symbol $\otimes$ denotes the Dempster-Shafer combination of the pmf's involved.
- $\hat{p}_x[V_i]$ is the pmf of V_i, estimated when $\mathbf{SN}$ is in the state $x = (y_1, y_2, ..., y_M) \in X$.
- $\hat{p}_{x_i^*}[V_i]$ is the centralized optimal pmf of V_i, estimated when $\mathbf{SN}$ is in the state $x_i^* \in X$, resulting in the smallest entropy of $\hat{p}_x[V_i]$.

Decentralized:

- $\hat{p}_{y_i}[V_i]$ is the same as in the centralized case.

- $\hat{p}_{y_i^*}[V_i]$ is the decentralized optimal pmf of V_i, estimated when $\mathbf{SN}_i$ is in the state $y_i^* \in Y_i$, resulting in the smallest entropy of $\hat{p}_{y_i}[V_i]$.

- $\hat{p}_{y_j^*}[V_i]$ is the decentralized inferred pmf of V_i, estimated when $\mathbf{SN}_j$, $j = 1, ..., M$, $j \neq i$, is in its state $y_j^* \in Y_j$, resulting in the smallest entropy of $\hat{p}_{y_j}[V_j]$.

- Finally, $\hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i]$ is the decentralized optimal pmf of V_i, estimated when $\mathbf{SN}$ is in the state $(y_1^*, y_2^*, ..., y_M^*) \in X$.

While the calculations involved in evaluating these pmf's are described in Section 3.3, below we comment on the main differences between the centralized and the decentralized process variable assessments:

(α) The centralized system uses all the pmf's, $\hat{p}_{y_i}[V_i]$ and $\hat{p}_{y_j}[V_i]$, $\forall j \neq i$, in order to evaluate x_i^*, whereas the decentralized one uses only the locally optimal pmf's, $\hat{p}_{y_i^*}[V_i]$ and $\hat{p}_{y_j^*}[V_i]$, to evaluate $(y_1^*, y_2^*, ..., y_M^*)$. The latter may lead to information losses, which is a drawback of the decentralization.

(β) The centralized system carries out the optimization in X, whereas the decentralized one in Y_i, $i = 1, ..., M$. The latter leads to a reduction of the process variable assessment time, which is the advantage of the decentralization.

The main problem addressed in this chapter is as follows: Derive a sufficient condition under which the decentralization leads to no loss of information, formalized as

$$x_i^* = (y_1^*, y_2^*, ..., y_M^*), \ \forall i, \tag{3.7}$$

and, consequently,

$$\hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i] = \hat{p}_{x_i^*}[V_i], \ \forall i. \tag{3.8}$$

A solution of this problem is given in Sections 3.4 and 3.5. As it turns out, the sought conditions depend on the properties of the process variable coupling, $P[V_i|V_j]$, $i \neq j$,

$i, j = 1, ..., M$, and on the monotonicity property of the Dempster-Shafer combination rule applied to the set of pmf's $\hat{p}_{y_j}[V_i]$, $i, j = 1, ..., M$.

3.3 Centralized and Decentralized Process Variable Assessment Procedures

Although the techniques utilized here are the same as in Chapter II, they are briefly described below for the sake of clarity.

3.3.1 Centralized case

Assessment of $\hat{p}_{y_i}[V_i]$: If the state $y_i \in Y_i$ has a single non-zero element, the evaluation of $\hat{p}_{y_i}[V_i]$ is carried out based on the data reported by this sensor and its data quality (DQ). The sensor's data and DQ are the inputs to the h-procedure, the steady state of which provides the sought pmf:

$$\hat{p}_{y_i}[V_i = \sigma] = p[S_i = \sigma]DQ + \frac{1 - DQ}{|\Sigma_{V_i}|},\ \sigma \in \Sigma_{V_i},\ i = 1, 2, ..., M, \tag{3.9}$$

where S_i is the random variable characterizing the sensor data; $p[S_i = \sigma]$, $\sigma \in \Sigma_{V_i}$, is its pmf; Σ_{V_i} is the universal set of V_i; and $|\Sigma_{V_i}|$ is the cardinality of Σ_{V_i}. As it follows from (3.9), $\hat{p}_{y_i}[V_i] = p[S_i]$, (i.e., S_i faithfully represents V_i) if $DQ = 1$ and $\hat{p}_{y_i}[V_i] = \frac{1}{|\Sigma_{V_i}|}$ (i.e., $\hat{p}_{y_i}[V_i]$ is uniform and, thus, S_i carries no information about V_i) if $DQ = 0$.

If y_i has more than one non-zero element, for each of them the pmf is evaluated using (3.9) and then the Dempster-Shafer rule is used to combine these pmf's. For instance, if y_i has two non-zero components, resulting in $\hat{p}_{y_i,1}[V_i]$ and $\hat{p}_{y_i,2}[V_i]$, the combined pmf is

$$\hat{p}_{y_i}[V_i = \sigma] = \hat{p}_{y_i,1;y_i,2}[V_i = \sigma] = \frac{\hat{p}_{y_i,1}[V_i = \sigma]\hat{p}_{y_i,2}[V_i = \sigma]}{\sum\limits_{\sigma \in \Sigma_{V_i}} \hat{p}_{y_i,1}[V_i = \sigma]\hat{p}_{y_i,2}[V_i = \sigma]},\ \sigma \in \Sigma_{V_i}. \tag{3.10}$$

When y_i has more than two non-zero components, say $k > 2$ non-zero components, the pmf $\hat{p}_{y_i}[V_i]$ is computed similarly, i.e.,

$$\hat{p}_{y_i}[V_i = \sigma] = \frac{\prod_{j=1}^{k} \hat{p}_{y_i,j}[V_i = \sigma]}{\sum_{\sigma \in \Sigma_{V_i}} \prod_{j=1}^{k} \hat{p}_{y_i,j}[V_i = \sigma]}, \ \sigma \in \Sigma_{V_i}, \ i = 1, 2, ..., M. \tag{3.11}$$

Assessment of $\hat{p}_{y_j}[V_i = \sigma]$: If $\mathbf{SN}_j$ is in the state $y_j \in Y_j$ leading to $\hat{p}_{y_j}[V_j]$, the induced pmf $\hat{p}_{y_j}[V_i]$, $j \neq i$, is calculated using the total probability formula,

$$\hat{p}_{y_j}[V_i] = \sum_{\sigma \in \Sigma_{V_j}} P[V_i | V_j = \sigma] \hat{p}_{y_j}[V_j = \sigma], \ \sigma \in \Sigma_{V_j}, \ y_j \in Y_j, \ i \neq j, \tag{3.12}$$

where $P[V_i|V_j]$, $i \neq j$, is the process variable coupling introduced in Section 3.2.

Assessment of $\hat{p}_x[V_i]$: If each $\mathbf{SN}_j$, $j = 1, ..., M$, is in the state $y_j \in Y_j$, the overall network, $\mathbf{SN}$, is in the state $x = (y_1, y_2, ..., y_M)$ and, therefore, $\hat{p}_x[V_i]$ can be calculated by combining $\hat{p}_{y_i}[V_i]$ and $\hat{p}_{y_j}[V_i]$, $i \neq j$, using the Dempster-Shafer rule:

$$\hat{p}_x[V_i = \sigma] = \hat{p}_{(y_1,y_2,...,y_M)}[V_i = \sigma] = \frac{\prod_{j=1}^{M} \hat{p}_{y_j}[V_i = \sigma]}{\sum_{\sigma \in \Sigma_{V_i}} \prod_{j=1}^{M} \hat{p}_{y_j}[V_i = \sigma]}, \ \sigma \in \Sigma_{V_i}. \tag{3.13}$$

Assessment of $\hat{p}_{x_i^*}[V_i]$: This is carried out using the method of rational controllers, where one controller is assigned to each V_i, $i = 1, 2, ..., M$, with the decision space being X and the penalty function being the entropy of $\hat{p}_x[V_i]$, $x \in X$, $i = 1, 2, ..., M$. These rational controllers were introduced in [17], and shown to perform well in all cyber-physical attack scenarios considered. The application of this method results in the identification of x_i^*, thus leading to $\hat{p}_{x_i^*}[V_i]$.

3.3.2 Decentralized case

Assessment of $\hat{p}_{y_i}[V_i]$: This assessment is carried out in the same manner as in the centralized case.

Assessment of $\hat{p}_{y_i^*}[V_i]$: Here, a rational controller is associated with each subnetwork $\mathbf{SN}_i$, $i = 1, 2, ..., M$, i.e., it operates in the decision space Y_i with the penalty function being the entropy of $\hat{p}_{y_i}[V_i]$, $y_i \in Y_i$. As a result, the state y_i^*, corresponding to the smallest entropy of $\hat{p}_{y_i}[V_i]$, is identified.

Assessment of $\hat{p}_{y_j^*}[V_i]$: This is carried out using the above pmf $\hat{p}_{y_j^*}[V_j]$ and the process variable coupling $P[V_i|V_j]$, $i \neq j$, by applying the total probability formula:

$$\hat{p}_{y_j^*}[V_i] = \sum_{\sigma \in \Sigma_{V_j}} P[V_i|V_j = \sigma]\hat{p}_{y_j^*}[V_j = \sigma],\ \sigma \in \Sigma_{V_j},\ y_j^* \in Y_j,\ i \neq j. \tag{3.14}$$

Assessment of $\hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i]$: This is carried out using the pmf's $\hat{p}_{y_i^*}[V_i]$ and $\hat{p}_{y_j^*}[V_i]$, $i \neq j$, by applying the Dempster-Shafer combination rule:

$$\hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i = \sigma] = \frac{\prod_{j=1}^{M} \hat{p}_{y_j^*}[V_i = \sigma]}{\sum_{\sigma \in \Sigma_{V_i}} \prod_{j=1}^{M} \hat{p}_{y_j^*}[V_i = \sigma]},\ \sigma \in \Sigma_{V_i},\ i = 1, 2, ..., M. \tag{3.15}$$

3.3.3 Assessment entropy

As described above, the assessments of $\hat{p}_{x_i^*}[V_i]$ and $\hat{p}_{y_i^*}[V_i]$, $i = 1, ..., M$, are based on selecting a pmf with the smallest entropy. In Chapter II, the Shannon entropy, [55], defined by

$$I\{p[V]\} = -\sum_{\sigma \in \Sigma_V} p[V = \sigma] \log_{|\Sigma_V|} p[V = \sigma], \tag{3.16}$$

has been used for this purpose. It turns out, however, that the Renyi-2 entropy [56],

$$H_2\{p[V]\} = -\log_{|\Sigma_V|}\left(\sum_{\sigma\in\Sigma_V}\{p[V=\sigma]\}^2\right), \tag{3.17}$$

is more appropriate for the problem at hand. The reason is that, as it is shown in [57], the Renyi-2 entropy is more effective in quantifying the so-called "guesswork" (GW), which is defined as the expected number of trials necessary to guess the outcome of a random variable. Since the assessments in both the centralized and the decentralized cases are conceptually related to GW, and since the Renyi-2 entropy has been shown to be effective in a number of random signal processing problems, [58–60], the current chapter uses $H_2\{p[V]\}$ in both the centralized and the decentralized process variable assessment procedures.

3.4 Condition for Lossless Decentralized Inference Assessment

In this section, we derive a sufficient condition under which no loss of information takes place due to the decentralized inference calculation as compared with the centralized one.

Represent the conditional pmf $P[V_i|V_j]$, $i \neq j$, as a set of column-vectors:

$$P[V_i|V_j] = \left[\mathbf{p}^{(1)}_{V_i|V_j}\ \mathbf{p}^{(2)}_{V_i|V_j}\ \cdots\ \mathbf{p}^{(n_j)}_{V_i|V_j}\right],\ \mathbf{p}^{(1)}_{V_i|V_j}, \mathbf{p}^{(2)}_{V_i|V_j}, ..., \mathbf{p}^{(n_j)}_{V_i|V_j} \in \mathbb{R}^{n_i},\ i \neq j, \tag{3.18}$$

where n_i is the cardinality of Σ_{V_i}. Recall that the components of the $\mathbf{p}_{V_i|V_j}$'s, are on $[0, 1]$ and their sum is 1. Introduce:

Assumption III.1. (a) The 2-norm of all the columns of matrix $P[V_i|V_j]$, $i \neq j$, are the same, i.e.,

$$\|\mathbf{p}^{(1)}_{V_i|V_j}\|_2 = \|\mathbf{p}^{(2)}_{V_i|V_j}\|_2 = \cdots = \|\mathbf{p}^{(n_j)}_{V_i|V_j}\|_2,\ i \neq j. \tag{3.19}$$

(b) The inner products of every pair of columns of matrix $P[V_i|V_j]$, $i \neq j$, are the same,

i.e.,

$$\langle \mathbf{p}_{V_i|V_j}^{(1)}, \mathbf{p}_{V_i|V_j}^{(2)} \rangle = \langle \mathbf{p}_{V_i|V_j}^{(1)}, \mathbf{p}_{V_i|V_j}^{(3)} \rangle = \cdots = \langle \mathbf{p}_{V_i|V_j}^{(n_j-1)}, \mathbf{p}_{V_i|V_j}^{(n_j)} \rangle, \ i \neq j. \tag{3.20}$$

■

While this assumption seems quite formal, its practical implication is as follows:

Lemma III.1. *Under Assumption III.1, if the pmf's $\hat{p}_l[V_j]$ and $\hat{p}_m[V_j]$, $l \neq m$, $l, m \in Y_j$, have equal information about the process variable V_j, then the inferred pmf's $\hat{p}_l[V_i]$ and $\hat{p}_m[V_i]$, $l \neq m$, $l, m \in Y_j$, $i \neq j$, calculated according to* (3.12)*, also have equal information about V_i. In other words,*

$$H_2\{\hat{p}_l[V_j]\} = H_2\{\hat{p}_m[V_j]\} \implies H_2\{\hat{p}_l[V_i]\} = H_2\{\hat{p}_m[V_i]\}, \ l \neq m \in Y_j, \ i \neq j. \tag{3.21}$$

Proof. See the Appendix. ■

Thus, Assumption III.1 guarantees that the quality of induced pmf's remain the same, if the original pmf's are equally informative. This property leads to

Theorem III.1. *Under Assumption III.1, the optimal decentralized inferred pmf $\hat{p}_{y_j^*}[V_i]$, calculated according* (3.14)*, has the same information as the most informative centralized inferred pmf, calculated according to* (3.12)*, i.e.,*

$$H_2\{\hat{p}_{y_j^*}[V_i]\} = \min_{y_j \in Y_j} H_2\{\hat{p}_{y_j}[V_i]\}. \tag{3.22}$$

Proof. See the Appendix. ■

Thus, this theorem provides a sufficient condition under which the decentralized inferred pmf (which requires the pmf evaluated at only $y_j^* \in Y_j$) does not lead to information losses as compared with the centralized inferred pmf assessments (which require the pmf's evaluation at all states $y_j \in Y_j$).

3.5 Condition for Lossless Overall Decentralized Assessment

As mentioned before, the decentralized optimal pmf of all the V_i's are evaluated at the sensor network state $(y_1^*, y_2^*, ..., y_M^*) \in X$. Regarding the centralized system, the optimal pmf of V_i is evaluated at the sensor network state $x_i^* \in X$. Given this situation, a question arises: Under what conditions are the decentralized and the centralized optimal states the same, i.e., $x_i^* = (y_1^*, y_2^*, ..., y_M^*)$, $\forall i$? This question is addressed below.

Recall that the calculation of $\hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i]$ is based on the Dempster-Shafer (D-S) combination rule (3.15). This rule is known to be, in general, non-monotonic [61] in the sense that D-S combination of two pmf's, say, $\hat{p}_1[V]$ and $\hat{p}_2[V]$, may have larger entropy than either of the constituent pmf's. This issue has been investigated in [13], where it has been shown that this does not take place (i.e., the D-S combination rule is, in fact, monotonic) if the constituent pmf's are sufficiently "close" to each other. As it turns out, a condition for the overall lossless decentralization depends on the monotonicity property of D-S rule. Specifically, introduce:

Assumption III.2. The Dempster-Shafer combination rule is monotonic on the set of pmf's $\left\{\hat{p}_{y_j}[V_i]\right\}$, $i, j = 1, 2, ..., M$, in the sense that

$$\begin{aligned} &\text{if } H_2\{\hat{p}_{y_j}[V_i]\} \leq H_2\{\hat{p}_{\bar{y}_j}[V_i]\},\ y_j, \bar{y}_j \in Y_j,\ \forall j, \\ &\text{then } H_2\{\hat{p}_{(y_1, y_2, ..., y_M)}[V_i]\} \leq H_2\{\hat{p}_{(\bar{y}_1, \bar{y}_2, ..., \bar{y}_M)}[V_i]\},\ \forall i. \end{aligned} \tag{3.23}$$

■

This assumption implies the monotonicity mentioned above, as stated by:

Lemma III.2. *Under Assumption III.2,*

$$H_2\{\hat{p}_{(y_1, y_2, ..., y_M)}[V_i]\} \leq \min\left\{H_2\{\hat{p}_{y_1}[V_i]\}, H_2\{\hat{p}_{y_2}[V_i]\}, ..., H_2\{\hat{p}_{y_M}[V_i]\}\right\}, \forall i. \tag{3.24}$$

Proof. See the Appendix. ■

Now, consider the following theorem:

Theorem III.2. *Under Assumptions III.1 and III.2, the centralized optimal state x_i^* is the same for all i, i.e., $x_i^* = x^*$, and, moreover, x^* coincides with the decentralized optimal state $(y_1^*, y_2^*, ..., y_M^*)$. Therefore,*

$$H_2\{\hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i]\} = H_2\{\hat{p}_{x^*}[V_i]\},\ i = 1, 2, ..., M. \tag{3.25}$$

Proof. See the Appendix. ■

Note that although this theorem is not constructive (since Assumption III.2 cannot be verified in a decentralized manner), it nevertheless specifies conditions for a lossless decentralization.

CHAPTER IV

Resilient Control Systems: Model Predictive Control Approach

4.1 Introduction

The resilient control system is intended to calculate plant control inputs based on the plant (or sub-plant) model and the process variable pmf's, $\hat{p}[V_i]$, $i = 1, ..., M$, provided by the resilient monitoring system (see Chapter II). Here, the objective of the control design is to steer the process variable to the desired value in the steady state, while ensuring that it remains in a safe domain in the transients. If the above pmf's were of zero entropy, classical control techniques could be applied. However, when the sensors are under attack, the entropy is non-zero, and new control techniques are necessary. This is because the feedback system to be developed can be neither output-based nor state space-based, but must be "pmf of the output-based" control.

The approach developed here can be briefly described as follows: Let $U_{\rm safe}$ be the control input, which maintains the process variable in the safe domain irrespective of the plant status, and $U_{\rm des}$ be the control input, which is necessary to ensure that the process variable would take the desired value if the process variable pmf had zero entropy. Then the re-

silient control input, $U_{\rm res}$, is defined as $U_{\rm res} = \Delta U_{\rm des} + (1-\Delta)U_{\rm safe}$, where $0 \leq \Delta \leq 1$ is a *weighting factor*, which is selected based on an optimization procedure. Specifically, when the entropy of $\hat{p}[V_i]$ is small, the input $U_{\rm des}$ is suitable for resilient control, and, therefore, the Δ is selected to be close to 1. However, when the entropy of $\hat{p}[V_i]$ is large, the $U_{\rm des}$ may steer the process variable outside the safe domain, and, hence, the Δ is selected to be close to 0. The above mentioned optimization procedure for calculating Δ is similar to those of model predictive control, [1], but is based on the process variable pmf, rather than on the process variable itself.

Although the current chapter presents just the initial results on resilient control systems, we believe that they form a foundation for extensions and future developments in this area of control research.

4.2 Results To-Date in Resilient Controller Design

The architecture of the resilient control system is shown in Figure 4.1, which combines the resilient monitoring system architecture of Figure 2.3 with the pmf-based control, $U_{{\rm res},i}$.

The design of this controller is based on the following: Assume that a sub-plant $\mathbf{G}$ (for the sake of brevity we omit its subscript) is described by the SISO system

$$\begin{array}{rcl} \mathbf{x}_\sigma(n+1) & = & A_\sigma \mathbf{x}_\sigma(n) + B_\sigma U_{\rm res}(n),\ \mathbf{x}_\sigma \in \mathbb{R}^q,\ U_{\rm res} \in \mathbb{R},\ n = 0, 1, ..., \\ \tilde{V}(n) & = & C_\sigma \mathbf{x}_\sigma(n),\ \tilde{V} \in \mathbb{R},\ n = 0, 1, ...,\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}, \end{array} \tag{4.1}$$

where the pair $\{A_\sigma, B_\sigma\}$, $\sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}$, is controllable, and the eigenvalues of A_σ, $\sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}$, are in the interior of the unit circle on the complex plane. Define the resilient control input as follows:

$$U_{\rm res}(n) = \Delta(n) U_{\rm des}(n) + [1 - \Delta(n)] U_{\rm safe}(n),\ n = 0, 1, ..., \tag{4.2}$$

where $U_{\rm safe}(n)$, $U_{\rm des}(n)$, and $\Delta(n)$ are to be determined. This is accomplished below.

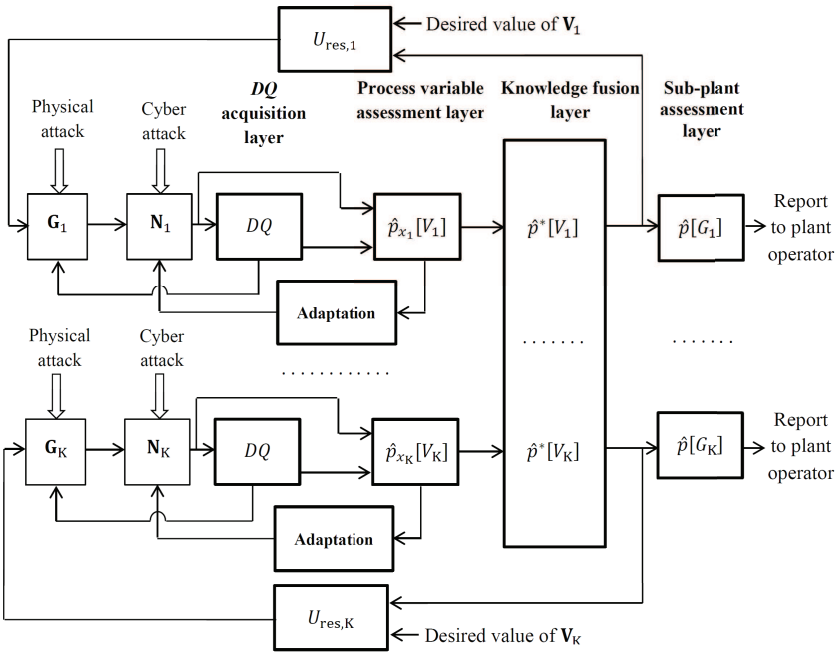


Figure 4.1: Proposed architecture of resilient control system

4.2.1 Calculation of U_{safe}

The value of U_{safe} is selected as an open-loop control input to ensure that the steady state of the process variable is within the safe domain, $[V_{\min}, V_{\max}]$, irrespective of its status. To formalize this, introduce the d.c. gain, α_σ, $\sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$, of the above state space system in the status σ, $\sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$, as

$$\alpha_\sigma = C_\sigma[I - A_\sigma]^{-1}B_\sigma,\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \tag{4.3}$$

and denote the minimum and maximum values of these d.c. gains as

$$\alpha_{\min} = \min\{\alpha_{\mathrm{L}_V}, \alpha_{\mathrm{N}_V}, \alpha_{\mathrm{H}_V}\},\ \alpha_{\max} = \max\{\alpha_{\mathrm{L}_V}, \alpha_{\mathrm{N}_V}, \alpha_{\mathrm{H}_V}\}. \tag{4.4}$$

Introduce the following assumption:

Assumption IV.1. The minimum and maximum values of the d.c. gains of the process

variable are such that

$$\frac{\alpha_{\max}}{\alpha_{\min}} < \frac{V_{\max}}{V_{\min}}, \tag{4.5}$$

where $V_{\min}$ and $V_{\max}$ characterize the safe domain of the process variable.

■

Now, consider the following lemma:

Lemma IV.1. *Under Assumption IV.1, the value of the input U_{safe} can be selected as any constant in the interval*

$$U_{\text{safe}} \in \left[\frac{V_{\min}}{\alpha_{\min}}, \frac{V_{\max}}{\alpha_{\max}}\right]. \tag{4.6}$$

Proof. See the Appendix. ■

Thus, the above lemma implies that under U_{safe} the steady state value of the process variable is in the safe domain, i.e., $U_{\text{safe}}\alpha_\sigma \in [V_{\min}, V_{\max}]$, $\forall \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$.

In a typical feedback control system, the sensor measurements can be used to determine the initial condition, $\mathbf{x}(0)$, of the plant; for instance, this may be accomplished by utilizing an observer. Then, based on the outputs of the observer, a control input can be calculated such that the resulting closed-loop system satisfies both the steady state and the transient performance specifications. However, in the current situation, the sensor measurements may be compromised, and, therefore, the use of an observer may not be appropriate. Thus, we consider the following scenario: Assume that at time $n = 0$, the sub-plant is in the steady state under the input U_{safe}, i.e.,

$$\begin{aligned} &\mathbf{x}_\sigma(0) = [I - A_\sigma]^{-1} B_\sigma U_{\text{safe}},\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \\ &\tilde{V}(0) = C_\sigma \mathbf{x}_\sigma(0),\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \end{aligned} \tag{4.7}$$

where, due to Lemma IV.1, we have $\tilde{V}(0) \in [V_{\min}, V_{\max}]$, irrespective of the status $\sigma \in$

$\{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$. Here, the problem is to determine the specific σ that takes place at $n = 0$, and, based on this information, compute an "appropriate" control input so that the control objectives are met, i.e., the process variable takes the desired value in the steady state, while remaining in the safe domain in the transients. The first part of this problem is solved using the resilient monitoring system, which is assumed to provide the estimate of the pmf of the process variable, $\hat{p}[V(0) = \sigma]$, $\sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$, at time $n = 0$. (Here, the resilient monitoring system may be viewed as the "observer".) The second part of the above problem is addressed below.

4.2.2 Calculation of U_{des}

To define U_{des}, introduce the two-degree of freedom control law, U_σ, which steers the process variable to the desired value, V_{des}, if the actual status of the process variable were σ:

$$U_\sigma(n) = -K_{1,\sigma}\hat{\mathbf{x}}_{\sigma,U_\sigma}(n) + K_{2,\sigma}V_{\mathrm{des}},\ n = 0, 1, 2, ...,\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \tag{4.8}$$

where $\hat{\mathbf{x}}_{\sigma,U_\sigma} \in \mathbb{R}^q$, $\sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$, is the "predicted" state vector of the sub-plant under the input U_σ, $\sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}$, i.e.,

$$\begin{aligned} &\hat{\mathbf{x}}_{\sigma,U_\sigma}(n+1) = A_\sigma\hat{\mathbf{x}}_{\sigma,U_\sigma}(n) + B_\sigma U_\sigma(n),\ n = 0, 1, ...,\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \\ &\hat{\mathbf{x}}_{\sigma,U_\sigma}(0) = \mathbf{x}_\sigma(0),\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}. \end{aligned} \tag{4.9}$$

As it may be observed from this equation, the initial condition, $\hat{\mathbf{x}}_{\sigma,U_\sigma}(0)$, for the prediction is the same as the initial state vector of the plant, $\mathbf{x}_\sigma(0)$ (defined in (4.7)). In the control law (4.8), the $K_{1,\sigma} \in \mathbb{R}^{1\times q}$ is a row vector of feedback gains, selected so that the eigenvalues of $(A_\sigma - B_\sigma K_{1,\sigma})$ are within the unit circle, and the $K_{2,\sigma} \in \mathbb{R}$ is a scalar, selected so that the d.c. gain of the closed loop transfer function, $K_{2,\sigma}C_\sigma\left[zI - (A_\sigma - B_\sigma K_{1,\sigma})\right]^{-1}B_\sigma$, is 1. Clearly, if the actual status of V is σ, the input $U_\sigma(n)$ would steer the process variable to

$V_{\rm des}$. However, since the status of V is unknown, the application of the incorrect U_σ to the plant may lead to a disaster. For example, assume that V is the temperature of the boiler and U is the opening of the fuel valve. In this system, the input $U_{\rm L_V}$ may correspond to increasing of the fuel valve opening, while the inputs $U_{\rm N_V}$ or $U_{\rm H_V}$ may correspond to maintaining or closing, respectively, of the fuel valve opening. Therefore, if the temperature is actually High, but the sensor projects Low (due to an attack), an explosion may occur if the fuel valve opening is further increased (when $U_{\rm L_V}$ is applied). Thus, to alleviate this problem, we synthesize the control input based on the estimated pmf, $\hat{p}[V(0)]$, of the process variable.

Specifically, define $U_{\rm des}$ as the expected value of U_σ with respect to the above pmf, i.e.,

$$\begin{aligned} U_{\rm des}(n) &= \hat{p}[V(0)={\rm L}_V]U_{\rm L_V}(n)+\hat{p}[V(0)={\rm N}_V]U_{\rm N_V}(n) \\ &\quad +\hat{p}[V(0)={\rm H}_V]U_{\rm H_V}(n),\ n=0,1,.... \end{aligned} \tag{4.10}$$

In the above control law, we utilize the pmf $\hat{p}[V(0)]$, rather than the pmf $\hat{p}[V(n)]$, for all time n, since it is assumed that the dynamics of the attacker is much slower than that of the closed loop system, implying that the sensor DQ, which is involved in the $\hat{p}[V]$ evaluation (see Chapter II for details), remains the same. Furthermore, as it may be observed from (4.10), if the entropy of $\hat{p}[V(0)]$ is close to 0, the input $U_{\rm des}(n)$ is suitable for resilient control; however, if the entropy of $\hat{p}[V(0)]$ is close to 1, the input $U_{\rm des}(n)$ may steer the process variable outside the safe domain. To overcome this problem, we define the resilient control input as mentioned in (4.2), i.e., $U_{\rm res}(n)=\Delta(n)U_{\rm des}(n)+[1-\Delta(n)]U_{\rm safe}$, where $0\leq\Delta(n)\leq 1$, $n=0,1,2,...$, is the weighting factor, which reduces the aggressiveness of $U_{\rm des}(n)$. The calculation of this weighting factor is described next.

4.2.3 Calculation of Δ

The goal here is to select Δ such that $U_{\rm res}$ steers the process variable to $V_{\rm des}$ in the steady state, while ensuring that the process variable remains in $[V_{\min}, V_{\max}]$ during the transients.

To accomplish this, we predict the future values of the sub-plant states and the outputs, under the input $U_{\rm res}$, i.e.,

$$\begin{aligned}
\hat{\mathbf{x}}_{\sigma,U_{\rm res}}(n+i) &= A_\sigma^i \hat{\mathbf{x}}_{\sigma,U_{\rm res}}(n) + \sum_{k=1}^{i} A_\sigma^{i-k} B_\sigma U_{\rm res}(n+k-1), \\
\hat{\mathbf{x}}_{\sigma,U_{\rm res}}(0) &= \mathbf{x}_\sigma(0), \\
\hat{V}_{\sigma,U_{\rm res}}(n+i) &= C_\sigma \hat{\mathbf{x}}_{\sigma,U_{\rm res}}(n+i), \\
&\quad n = 0, 1, ...,\ i = 1, 2, ..., N_{\rm p},\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\},
\end{aligned} \tag{4.11}$$

where $N_{\rm p}$ is the duration period of the prediction, and, as before, the initial condition for the prediction, $\mathbf{x}_\sigma(0)$, is defined in (4.7). (The above expression is obtained as the solution of the discrete-time LTI system (4.1).) As described below, the predicted values of the process variable, $\hat{V}_{\sigma,U_{\rm res}}$, are involved in the previously mentioned optimization procedure, which is used to compute Δ.

Based on the definition (4.2), this weighting factor Δ can be considered as a parameter of the input $U_{\rm res}$, i.e.,

$$U_{\rm res}(n) = U_{\rm res}(n; \Delta(n)),\ n = 0, 1, ... \tag{4.12}$$

Similarly, based on the prediction model (4.11), the Δ can also be considered as a parameter of the predicted process variable, i.e.,

$$\hat{V}_{\sigma,U_{\rm res}}(n+1) = \hat{V}_{\sigma,U_{\rm res}}(n+1; \Delta(n)),\ n = 0, 1, ...,\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}. \tag{4.13}$$

Clearly, if these predicted values, $\hat{V}_{\sigma,U_{\rm res}}(n+1; \Delta(n)), \hat{V}_{\sigma,U_{\rm res}}(n+2; \Delta(n+1)), ..., \hat{V}_{\sigma,U_{\rm res}}(n+N_{\rm p}; \Delta(n+N_{\rm p}-1))$, are required to approach and eventually track the desired value, $V_{\rm des}$, the weights $\Delta(n), \Delta(n+1), ..., \Delta(n+N_{\rm p}-1)$ must be selected appropriately. To accomplish this, introduce the following optimization problem, which, as mentioned before, is

based on the ideas of model predictive control:

$$\begin{aligned} \underset{\Delta(n),\Delta(n+1),...,\Delta(n+N_{\rm p}-1)}{\text{minimize}} \quad & \sum_{i=1}^{N_{\rm p}} \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} \frac{1}{2} W_\sigma \left[\hat{V}_{\sigma,U_{\rm res}}(n+i;\Delta(n+i-1)) - V_{\rm des} \right]^2, \\ \text{subject to} \quad & V_{\min} \leq \hat{V}_{\sigma,U_{\rm res}}(n+i;\Delta(n+i-1)) \leq V_{\max}, \\ & n=0,1,...,\ i=1,...,N_{\rm p},\ \sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}, \end{aligned} \tag{4.14}$$

where $\hat{V}_{\sigma,U_{\rm res}}$ is computed using (4.11). The W_σ's, involved in the above penalty function, are selected as

$$W_\sigma = \left(\frac{\hat{p}[V(0)=\sigma]}{\epsilon_W} \right)^{N_W},\ \sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}, \tag{4.15}$$

where $\epsilon_W < 1$ and $N_W \geq 1$ are design parameters. If N_W is selected to be large, then W_σ is small if $\hat{p}[V(0)=\sigma]$ is small (i.e., $\hat{p}[V(0)=\sigma] < \epsilon_W$), and W_σ is large if $\hat{p}[V(0)=\sigma]$ is large (i.e., $\hat{p}[V(0)=\sigma] > \epsilon_W$). The duration period, $N_{\rm p}$, of the prediction must be selected so that it is not too small, which makes the controller response oscillatory, nor too large, which increases the computational complexity of the solution of (4.14).

We assume that the solution of the above problem, (4.14), is feasible for all time n (this is termed as recursive or persistent feasibility, [62]). Given the solution $\Delta^*(n), \Delta^*(n+1), ..., \Delta^*(n+N_{\rm p}-1)$, we utilize only $\Delta^*(n)$ to compute $U_{\rm res}(n)$, while $\Delta^*(n+1), ..., \Delta^*(n+N_{\rm p}-1)$ are discarded (as it is usual in model predictive control). Finally, this $U_{\rm res}(n)$ is applied to the plant.

Since the actual states of the sub-plant cannot be observed, the stopping rule of the above optimization procedure is selected as follows:

$$\|\hat{\mathbf{x}}_{\sigma,U_{\rm res}}(n+1) - \hat{\mathbf{x}}_{\sigma,U_{\rm res}}(n)\| < \epsilon_{\rm stop},\ \forall \sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}, \tag{4.16}$$

where $\epsilon_{\rm stop} << 1$ can be chosen as desired.

Given the control design described here, it may be important to characterize the stability

of the closed loop system and the steady state value of the process variable (if it exists). Assuming that this steady state exists, one of the possible ways to characterize the efficacy of the resilient controller may be to evaluate the pmf of the process variable in the steady state, denoted as $\hat{p}[V_{\rm ss}]$, and determine under what conditions this pmf takes the largest probability in the Normal status. While, in the current work, these issues are not addressed for the case of dynamic plants, below we consider the case of static plants, and provide a sufficient condition under which $\hat{p}[V_{\rm ss}]$, indeed, takes the largest probability in the Normal status.

4.2.4 Characterization of $\hat{p}[V_{\rm ss}]$ properties

Consider the static sub-plant

$$\begin{aligned} x_\sigma(n) &= \alpha_\sigma U_{\rm res}(n),\ x_\sigma \in \mathbb{R},\ n=0,1,\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}, \\ U_{\rm res}(0) &= U_{\rm safe}, \\ \tilde{V}(n) &= x_\sigma(n),\ n=0,1,\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}, \end{aligned} \tag{4.17}$$

where, as before, the process variable $\tilde{V}$ takes values in the interval $[V_{\rm min}, V_{\rm max}]$, with the sub-intervals $[V_{\rm min}, R_1)$, $[R_1, R_2)$, and $[R_2, V_{\rm max})$ specifying the Low, Normal, and High domains, respectively. Denote the actual gain of the sub-plant as $\alpha_{\rm act} \in \{\alpha_{{\rm L}_V}, \alpha_{{\rm N}_V}, \alpha_{{\rm H}_V}\}$, and define the sub-interval $I_{\rm act}$ as

$$I_{\rm act}\left(\tilde{V}(n)\right) := \begin{cases} [V_{\rm min}, R_1), & \text{if } \tilde{V}(n) \in [V_{\rm min}, R_1), \\ [R_1, R_2), & \text{if } \tilde{V}(n) \in [R_1, R_2), \\ [R_2, V_{\rm max}], & \text{if } \tilde{V}(n) \in [R_2, V_{\rm max}], \end{cases} \quad n=0,1. \tag{4.18}$$

Assume that one sensor monitors the process variable, and let the sensor measurements be represented by $\tilde{S}$. Further, assume that the model of the attacker is as follows:

$$\tilde{S}(n) = k_a \tilde{V}(n) + d_a,\ n = 0, 1, \tag{4.19}$$

where k_a and d_a are the gain and the bias, respectively, of the attacker. Based on the data quality acquisition procedure, described in Chapter II, it can be shown that the sensor DQ is a function of the above k_a and d_a, i.e.,

$$DQ = DQ(k_a, d_a). \tag{4.20}$$

Introduce the following assumption:

Assumption IV.2. The attacker gain and bias are such that

$$k_a U_{\text{safe}} \alpha_{\text{act}} + d_a \in I_{\text{act}} \left(\tilde{V}(0) \right). \tag{4.21}$$

■

The implication of this assumption is as follows:

Lemma IV.2. *Under Assumption IV.2, the actual and the estimated pmf's of V at time $n = 0$ take the maximum probability at the same status.*

Proof. See the Appendix. ■

The input U_{res} is calculated as described in Subsections 4.2.1-4.2.3. Clearly, the value of this input at time $n = 1$, i.e., $U_{\text{res}}(1)$, defines the pmf of V at time $n = 1$, i.e., $p[V(1)]$. (Note that since the plant is static, the process variable is in the steady state at $n = 1$. This implies that the pmf $p[V_{\text{ss}}]$ is the same as the pmf $p[V(1)]$.) Thus, to characterize $p[V_{\text{ss}}]$, we have to compute $\hat{p}[V(1)]$, and determine if it takes the maximum probability in the Normal status. To accomplish that, introduce the following definition:

Definition IV.1. The pmf $\hat{p}[V(1)]$ is said to be the *correct permutation* of the pmf $\hat{p}[V(0)]$ if:

(i) both the pmf's are comprised of the same probabilities;

(ii) the pmf $\hat{p}[V(1)]$ takes the largest probability in the Normal status, irrespective of the pmf $\hat{p}[V(0)]$.

■

Since the sub-plant is considered to be static, select the parameter N_{p}, involved in the optimization problem (4.14), as 1. Further, for simplicity, assume that the parameter N_W, also involved in (4.14), is selected as 1. Then,

Theorem IV.1. *Under Assumption IV.2, the plant input U_{res} results in the pmf $\hat{p}[V(1)]$, which is the correct permutation of the pmf $\hat{p}[V(0)]$, if*

$$k_a V_{\mathrm{des}} \alpha_{\mathrm{act}} \left[\frac{\alpha_{\mathrm{L}_V} + \alpha_{\mathrm{N}_V} + \alpha_{\mathrm{H}_V} + [3\alpha_{\mathrm{act}} - \alpha_{\mathrm{L}_V} - \alpha_{\mathrm{N}_V} - \alpha_{\mathrm{H}_V}] DQ(k_a, d_a)}{\alpha^2_{\mathrm{L}_V} + \alpha^2_{\mathrm{N}_V} + \alpha^2_{\mathrm{H}_V} + [3\alpha^2_{\mathrm{act}} - \alpha^2_{\mathrm{L}_V} - \alpha^2_{\mathrm{N}_V} - \alpha^2_{\mathrm{H}_V}] DQ(k_a, d_a)} \right] + d_a \in [R_1, R_2). \quad (4.22)$$

Proof. See the Appendix. ■

This theorem implies that if k_a and d_a are bounded, as characterized by the expressions (4.21) and (4.22), then the resilient control input, indeed, ensures that the process variable takes the Normal status with the maximum probability.

4.3 Example

While the previous section provided sufficient conditions for the efficacy of the developed controller for static systems, in this section, we offer an example showing that the resilient control system works for dynamic systems as well.

Assume that the sub-plant (4.1) is a first order system, i.e., $q = 1$. Further, assume that the parameters involved in the sub-plant, the process variable, and the resilient controller are as follows:

- Sub-plant parameters:
  - $A_{\mathrm{L}_V} = 0.9$, $B_{\mathrm{L}_V} = 0.1$, $C_{\mathrm{L}_V} = 1$;
  - $A_{\mathrm{N}_V} = 0.5$, $B_{\mathrm{N}_V} = 0.75$, $C_{\mathrm{N}_V} = 1$;
  - $A_{\mathrm{H}_V} = 0.67$, $B_{\mathrm{H}_V} = 1$, $C_{\mathrm{H}_V} = 1$.
- Process variable parameters:
  - $V_{\min} = 1$, $V_{\max} = 200$, $R_1 = \frac{200}{3}$, $R_2 = \frac{400}{3}$;
  - $V_{\mathrm{des}} = 100$.
- Resilient controller parameters:
  - $U_{\mathrm{safe}} = 50$;
  - $K_{1,\mathrm{L}_V} = 1.34$, $K_{2,\mathrm{L}_V} = 3$, $K_{1,\mathrm{N}_V} = -4$, $K_{2,\mathrm{N}_V} = -0.134$, $K_{1,\mathrm{H}_V} = 6$, $K_{2,\mathrm{H}_V} = 0.07$;
  - $\epsilon_W = 0.01$, $N_W = 10$, $N_{\mathrm{p}} = 5$.

Assume that the sub-plant is in the steady state if the following stopping rule is satisfied:

$$|\hat{x}_{\sigma,U_{\mathrm{res}}}(n+1) - \hat{x}_{\sigma,U_{\mathrm{res}}}(n)| < 10^{-4},\ \forall \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \tag{4.23}$$

and let N_{ss} denote the time at which the above takes place. (Note that for all scenarios considered below, N_{ss} is determined to be less than 50sec.) As before, assume that the attacker model is characterized by

$$\tilde{S}(n) = k_a \tilde{V}(n) + d_a,\ n = 0, 1, ... \tag{4.24}$$

We evaluate the performance of the resilient controller under nine attack scenarios, described in Tables 4.1 and 4.2. In these tables, the quantity $\overline{\Delta}$ is the average value of

$\Delta^*(n)$, evaluated during the transients, i.e.,

$$\overline{\Delta} = \frac{1}{N_{\rm ss}} \sum_{n=0}^{N_{\rm ss}-1} \Delta^*(n), \tag{4.25}$$

and the quantities k_a, d_a, DQ, $\hat{p}[V(0)]$, and $\hat{p}[V_{\rm ss}]$ are the same as before. The $\overline{\Delta}$ is introduced to characterize the aggressiveness of $U_{\rm res}$ in each of the scenarios considered. In Scenarios 1 - 6, the value of d_a is fixed at 10, and the performance is evaluated for various values of k_a, whereas in Scenarios 7 - 9, the value of k_a is fixed at 0.9, and the performance is evaluated for various values of d_a. In all these scenarios, we assume that the actual subplant status is Low.

As it may be observed from the above tables, the correct permutation of the pmf's, $\hat{p}[V_{\rm ss}]$ and $\hat{p}[V(0)]$, takes place in Scenarios 1,2,3,7, and 8. Clearly, in these scenarios, the attacker's modifications of the sensor measurements are relatively small, and, therefore, the resulting $\hat{p}[V(0)]$ contains a sufficient amount of information about the status of the process variable (since DQ is close or equal to 1). This information is utilized by the resilient controller to steer the process variable into the Normal domain (see Figure 4.2, where the trajectories of $\tilde{V}$ and Δ^* are illustrated for Scenario 3). Regarding Scenarios 4,5,6, and 9, it may be observed that the pmf's $\hat{p}[V_{\rm ss}]$ and $\hat{p}[V(0)]$ are the same. Here, the attacker's modifications of the sensor measurements are relatively large, which results in $\hat{p}[V(0)]$ not containing a sufficient amount of information about the status of the process variable (since DQ is relatively small). Thus, in these scenarios, the resilient control input is not aggressive, which ensures that the process variable is maintained in the same domain, i.e., Low (see Figure 4.3, where the trajectories of $\tilde{V}$ and Δ^* are illustrated for Scenario 5).

Thus, the results obtained here indicate that the model predictive control-based approach to resilient feedback systems can be viewed as a potential solution of the resilient control problem, and should be explored in more details in the future.

Table 4.1: Performance of the resilient controller for Scenarios 1 - 6 ($d_a = 10$)

Scenario	k_a	DQ	$\hat{p}[V(0)]$	$\hat{p}[V_{\text{ss}}]$	$\overline{\Delta}$
1	1	1	$[1, 0, 0]$	$[0, 1, 0]$	1
2	0.95	0.97	$[0.97, 0.015, 0.015]$	$[0.015, 0.97, 0.015]$	0.11
3	0.90	0.85	$[0.9, 0.05, 0.05]$	$[0.05, 0.9, 0.05]$	0.12
4	0.85	0.70	$[0.8, 0.1, 0.1]$	$[0.8, 0.1, 0.1]$	0.14
5	0.58	0.07	$[0.38, 0.31, 0.31]$	$[0.38, 0.31, 0.31]$	0.06
6	0.50	0.02	$[0.34, 0.33, 0.33]$	$[0.34, 0.33, 0.33]$	0

Table 4.2: Performance of the resilient controller for Scenarios 7 - 9 ($k_a = 0.9$)

Scenario	d_a	DQ	$\hat{p}[V(0)]$	$\hat{p}[V_{\text{ss}}]$	$\overline{\Delta}$
7	15	0.85	$[0.9, 0.05, 0.05]$	$[0.05, 0.9, 0.05]$	0.12
8	20	0.85	$[0.9, 0.05, 0.05]$	$[0.05, 0.9, 0.05]$	0.12
9	25	≈ 0	$\approx [1/3, 1/3, 1/3]$	$\approx [1/3, 1/3, 1/3]$	0

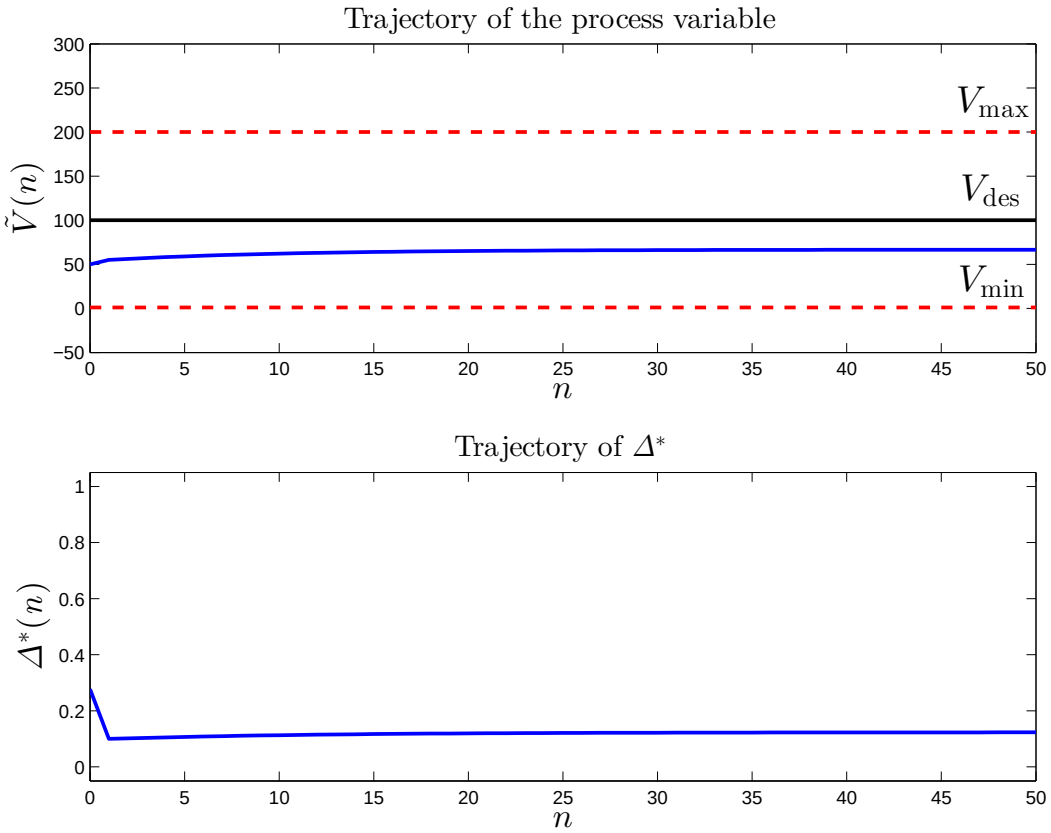


Figure 4.2: Trajectories of $\tilde{V}$ and Δ^* for Scenario 3

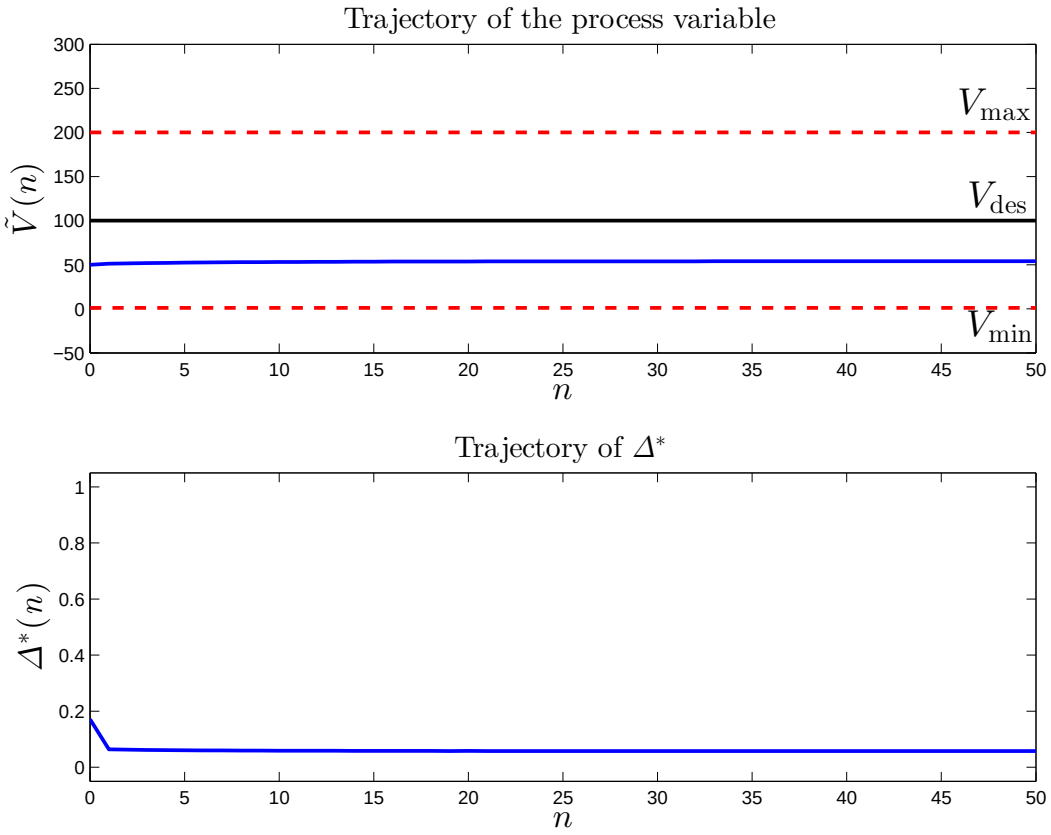


Figure 4.3: Trajectories of $\tilde{V}$ and Δ^* for Scenario 5

CHAPTER V

Actuator/Sensor Health Monitoring and Control Using Synchronous Detection

5.1 Introduction

In Chapter IV, a resilient plant control input is synthesized using the information provided by the resilient monitoring system. In the current chapter, we consider a different approach, namely, the "nominal" feedback controller is modified based on the health assessment of the sensors and actuators (which are assumed to be under attack). As mentioned before, this health assessment is carried out using the method of *synchronous detection*, which is widely applied in communication systems [2].

Resilient feedback systems, considered in this chapter, are feedback control systems that are capable of identifying and mitigating malicious attacks on their sensors and actuators, wherein the attacks are intended to force the plant output to deviate substantially from the reference signal. In the absence of appropriate identification and mitigation strategies, attacks may lead to unwanted consequences, such as damage to the plant. For example, consider the drive system of a Uranium gas enrichment centrifuge, which typically consists of a three-phase AC induction motor, a controller, and a speed sensor [63–67]. Since

this system operates in a closed-loop configuration, an attack on the sensor that forces it to project a 'low' speed may lead to the actual motor speed taking dangerously high values.

We assume that the attacker's actions can be categorized as follows:

- Type 1 attack: The DC gain/s of the sensor or/and actuator is/are modified;
- Type 2 attack: A constant input is projected as an output of the sensor or/and actuator;
- Type 3 attack: A combination of the above two takes place, e.g., the DC gain of the actuator is modified, while a constant input is projected as the output of the sensor.

To illustrate these types of attacks, consider the nominal (non-attacked) feedback control system shown in Fig. 5.1, wherein $K(s)$, $A(s)$, $P(s)$, and $S(s)$ represent the transfer functions of the controller, actuator, plant, and sensor, respectively, and S_0 is the DC gain of the sensor, i.e., $S_0 = \lim_{s\to 0} S(s)$. For this system, a Type 1 attack on the sensor is

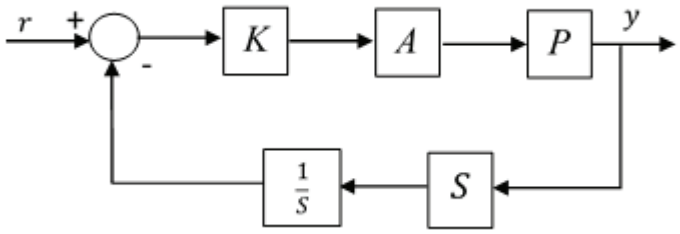


Figure 5.1: Nominal system

depicted in Fig. 5.2, while a Type 2 attack on the actuator is shown in Fig. 5.3.

Given the above model of the attacker, our goal is to devise strategies to identify

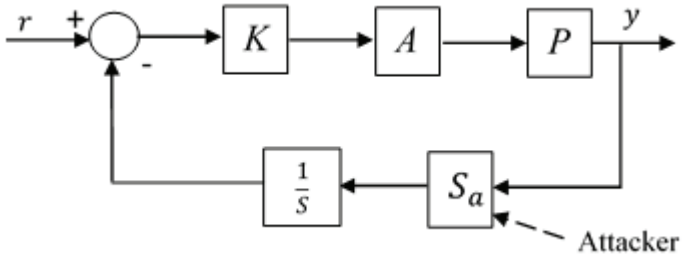


Figure 5.2: Type 1 attack on the sensor, i.e., $\lim_{s\to 0} S_a(s) \neq S_0$

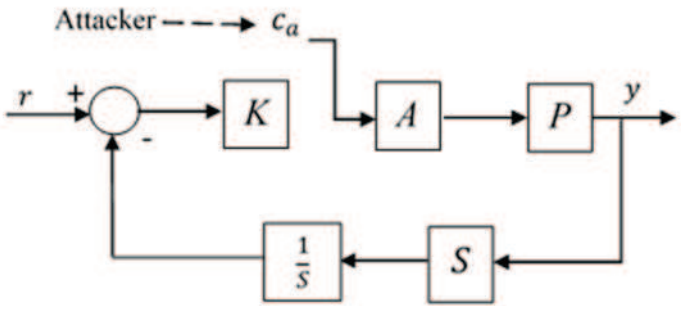


Figure 5.3: Type 2 attack on the actuator (attacker projects a constant input, c_a)

malicious attacks on the control system and ensure that their effects are mitigated as quickly as possible.

The development of the identification and mitigation procedures is carried out under the following assumption:

Assumption V.1. a) The controller, plant, nominal actuator, and nominal sensor are open-loop asymptotically stable, i.e., the poles of the transfer functions $K(s)$, $P(s)$, $A(s)$, and $S(s)$ lie in the open left half plane.

b) The attacked actuator and the attacked sensor are open-loop asymptotically stable.

c) The nominal and the attacked closed-loop systems are asymptotically stable.

■

Under Assumption V.1, and as illustrated in Fig. 5.4, the approach to the identification of attacks comprises of:

- Adding a sinusoidal signal to the reference;
- Multiplying the outputs of the actuator and the $\frac{1}{S_0}$ blocks by the same sinusoidal signal;
- Computing the moving average of the signals resulting from the previous step.

Under the above procedure, the outputs of the moving average blocks, z_1 and z_2, are analyzed from the point of view of their consistency with the nominal values. As explained in

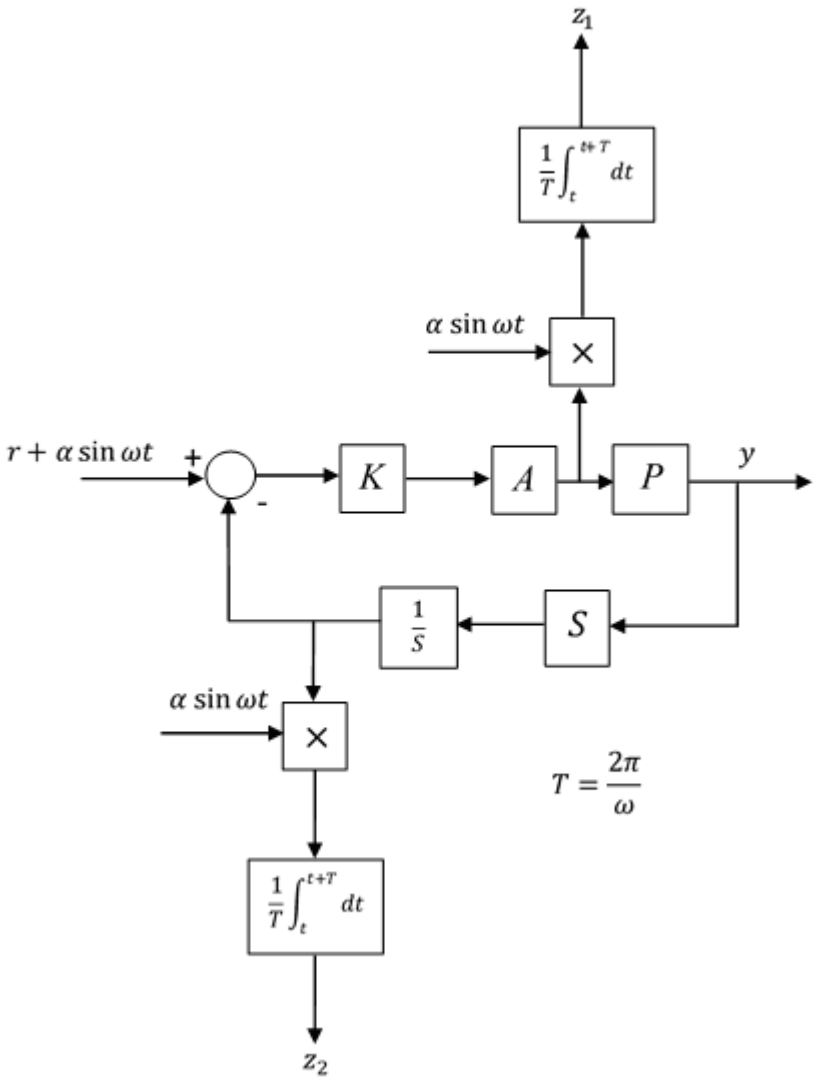


Figure 5.4: Identification of attacks using synchronous detection

details in Section 5.2, attacks on the sensor and the actuator lead to z_1 and z_2 taking steady state values that differ from their nominal ones, which gives rise to the identification of the attacker's actions.

Regarding the mitigation of attacks, this is based on the results of the identification procedure. Specifically, in the case of Type 1 attacks, the DC gains of the controller and the $\frac{1}{S_0}$ block are appropriately modified to ensure that the plant output is close to the reference signal, whereas in the case of Type 2 or Type 3 attacks, operation of the closed loop system is discontinued. These procedures are explained in details in Section 5.3.

Note that if the attacker modifies the gains of the sensor and the actuator such that the closed loop system is unstable, then the signals z_1 and z_2 do not attain the steady state. In this situation, as before, the operation of the closed loop system is discontinued.

The remainder of this chapter is organized as follows: As mentioned previously, the attack identification and mitigation procedures are described in Sections 5.2 and 5.3, respectively. Timing issues are analyzed in Section 5.4. Finally, an example of the application of the developed procedures is presented in Section 5.5.

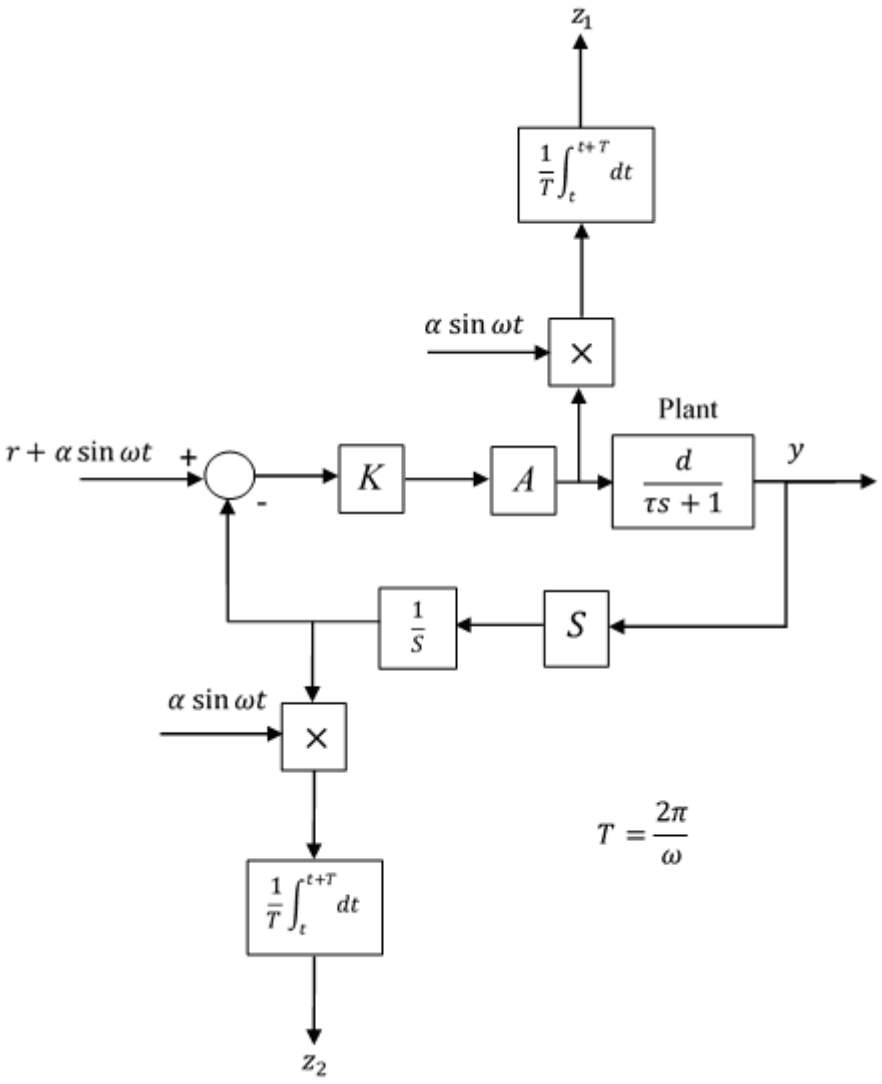


Figure 5.5: Identification of attacks using synchronous detection – Simplified case

5.2 Attack Identification

The development of the techniques in this and the subsequent sections is carried out in terms of the feedback control system shown in Fig. 5.5. The plant's dynamics are assumed to be characterized by the stable first order transfer function,

$$P(s) = \frac{d}{\tau s + 1}, \tag{5.1}$$

where d is the DC gain and τ is the time constant. Further, the controller, sensor, and actuator are assumed to be static, with their gains denoted as K, S, and A, respectively. Note that these assumptions are made in order to simplify the presentation of the material, and that the techniques developed here can be extended to more complex systems (e.g., higher order plants, controllers with dynamics, etc.).

Given the above, the nominal steady state values of the signals z_1 and z_2 can be computed as

$$z_{1,\text{ss}} = \left[\frac{KA(1 + KAd + \omega^2\tau^2)}{\omega^2\tau^2 + (1 + KAd)^2}\right]\frac{\alpha^2}{2}, \tag{5.2}$$

and

$$z_{2,\text{ss}} = \left[\frac{KAd(1 + KAd)}{\omega^2\tau^2 + (1 + KAd)^2}\right]\frac{\alpha^2}{2}, \tag{5.3}$$

respectively. These values are used below to ascertain if an attack has indeed taken place or not.

5.2.1 Identification of Type 1 attack

Assume that both the sensor and the actuator are under a Type 1 attack, i.e., the gains S and A are modified as $S_a > 0$ and $A_a > 0$, respectively. Given this scenario, the steady state

values of z_1 and z_2 can be computed as

$$\begin{aligned} z_{1,\text{ss,a}} &= \left[\frac{KA_a\left(1+KA_ad\frac{S_a}{S}+\omega^2\tau^2\right)}{\omega^2\tau^2+\left(1+KA_ad\frac{S_a}{S}\right)^2}\right]\frac{\alpha^2}{2}, \\ z_{2,\text{ss,a}} &= \left[\frac{KA_ad\frac{S_a}{S}\left(1+KA_ad\frac{S_a}{S}\right)}{\omega^2\tau^2+\left(1+KA_ad\frac{S_a}{S}\right)^2}\right]\frac{\alpha^2}{2}. \end{aligned} \tag{5.4}$$

Using the above equations, the following expressions for S_a and A_a are obtained:

$$\begin{aligned} S_a &= S\left[\frac{4z_{2,\text{ss,a}}-\alpha^2+\sqrt{(4z_{2,\text{ss,a}}-\alpha^2)^2-8z_{2,\text{ss,a}}(1+\omega^2\tau^2)(2z_{2,\text{ss,a}}-\alpha^2)}}{4z_{1,\text{ss,a}}d}\right], \\ A_a &= \frac{2z_{1,\text{ss,a}}}{K[\alpha^2-2z_{2,\text{ss,a}}]}. \end{aligned} \tag{5.5}$$

These expressions are utilized in Section 5.3 to mitigate the effects of Type 1 attacks.

5.2.2 Identification of Type 2 attack

Assuming that a Type 2 attack takes place on the sensor, i.e., a constant input, c_a, is projected by the attacker, the steady state values of z_1 and z_2 can be computed as

$$z_{1,\text{ss,a}} = KA\frac{\alpha^2}{2},\ z_{2,\text{ss,a}} = 0. \tag{5.6}$$

Similarly, under a Type 2 attack on the actuator, we have

$$z_{1,\text{ss,a}} = z_{2,\text{ss,a}} = 0. \tag{5.7}$$

The above expression also applies for the case of simultaneous Type 2 attacks on both the sensor and the actuator.

5.2.3 Identification of Type 3 attack

Assume that a Type 1 attack takes place on the actuator, while a Type 2 attack takes place on the sensor. In this scenario,

$$z_{1,\text{ss,a}} = KA_a\frac{\alpha^2}{2},\ z_{2,\text{ss,a}} = 0. \tag{5.8}$$

Suppose that a Type 2 attack takes place on the actuator, while a Type 1 attack takes place on the sensor. Then,

$$z_{1,\text{ss,a}} = z_{2,\text{ss,a}} = 0. \tag{5.9}$$

Clearly, under attack, the steady state values of the z's are different from the nominal ones. These results are summarized in Table 5.1.

Table 5.1: Steady state values of z_1 and z_2 under various scenarios

Scenario\Signal	$z_{1,\text{ss}}$	$z_{2,\text{ss}}$
Nominal system	$\left[\frac{KA(1+KAd+\omega^2\tau^2)}{\omega^2\tau^2+(1+KAd)^2}\right]\frac{\alpha^2}{2}$	$\left[\frac{KAd(1+KAd)}{\omega^2\tau^2+(1+KAd)^2}\right]\frac{\alpha^2}{2}$
Type 1 attack on S and A	$\left[\frac{KA_a\left(1+KA_ad\frac{S_a}{S}+\omega^2\tau^2\right)}{\omega^2\tau^2+\left(1+KA_ad\frac{S_a}{S}\right)^2}\right]\frac{\alpha^2}{2}$	$\left[\frac{KA_ad\frac{S_a}{S}\left(1+KA_ad\frac{S_a}{S}\right)}{\omega^2\tau^2+\left(1+KA_ad\frac{S_a}{S}\right)^2}\right]\frac{\alpha^2}{2}$
Type 2 attack on S	$KA\frac{\alpha^2}{2}$	0
Type 2 attack on A	0	0
Type 2 attack on S and A	0	0
Type 3 attack: Type 2 attack on S and Type 1 attack on A	$KA_a\frac{\alpha^2}{2}$	0
Type 3 attack: Type 1 attack on S and Type 2 attack on A	0	0

5.3 Attack Mitigation

5.3.1 Mitigation of Type 1 attack

As described in Section 5.2, under a simultaneous Type 1 attack on both the sensor and the actuator, the steady state values of the signals z_1 and z_2 can be used to calculate the values of the gains S_a and A_a. These, in turn, are utilized to compensate for the effects of the attack by modifying the controller K as $K\frac{A}{A_a}$ and the $\frac{1}{S}$ block as $\frac{1}{S_a}$ (see Fig. 5.6).

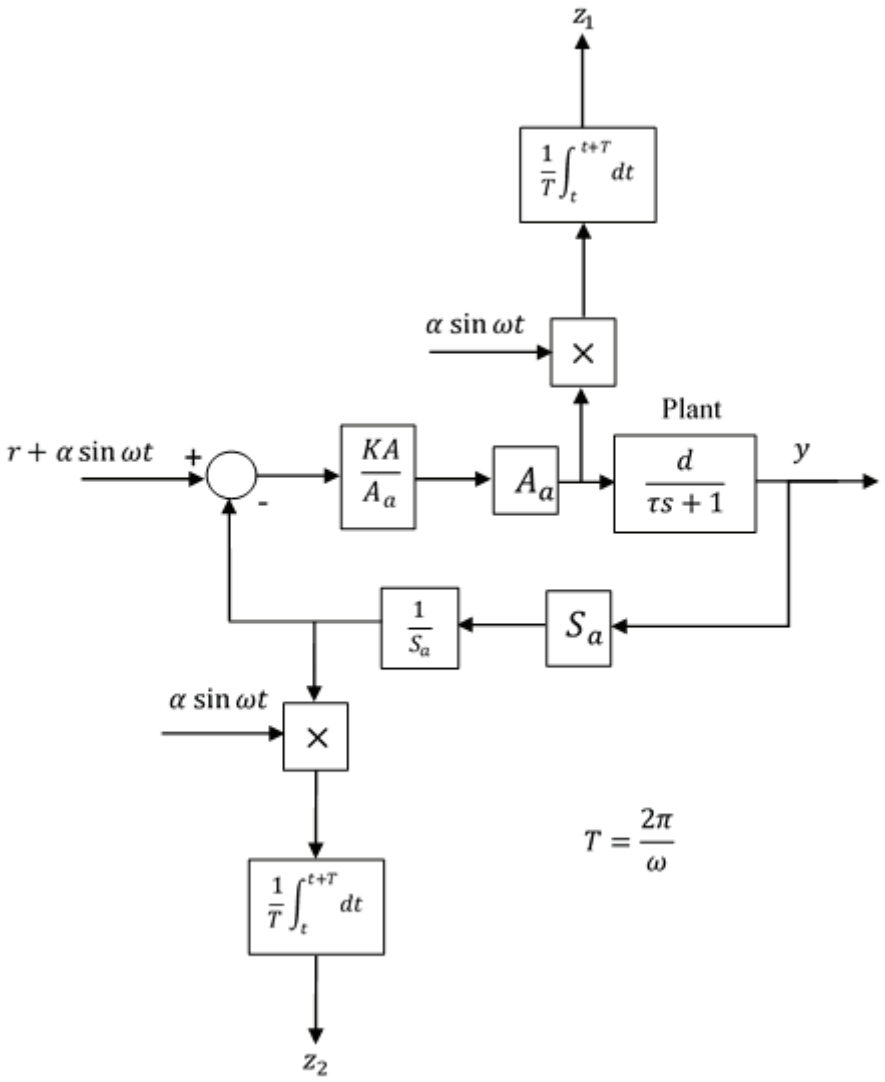


Figure 5.6: Mitigation of Type 1 attack on the sensor and the actuator

5.3.2 Mitigation of Type 2 or Type 3 attack

In the case of either Type 2 or Type 3 attack, it is clear that the feedback loop is disconnected by the attacker. Therefore, this leaves us with no alternative but to discontinue operation of the control system. Normal operation may resume after the attacked component/s is/are either repaired or replaced.

5.4 Timing Issues

In some applications, there may exist a 'critical' time duration, T_c, beyond which it is undesirable for the plant output to be substantially different from the reference. Obviously, it is necessary that the time required to complete the identification and mitigation procedures be less than T_c. Below, we examine the duration of the former under the various types of attacks.

In the case of a Type 1 attack, the transient response of the resilient control system can be partitioned into the following three time intervals:

- Time required for the plant output to go close to the new steady state value, after the attack takes place;
- Time required to calculate the new steady state values of z_1 and z_2, after the above takes place;
- Time required for the plant output to go close to the reference signal, after the identification and mitigation procedures are applied.

We assume that the duration of the first time interval is 3τ, where τ is the time constant of the plant transfer function. As for the second time interval, its duration is T, where T is the time period of the sinusoidal oscillations. Finally, we assume, as before, that the duration of the third time interval is 3τ. Thus, the time, T_{idm}, required to identify and

mitigate a Type 1 attack is

$$T_{\text{idm}} = 6\tau + T. \tag{5.10}$$

Clearly, from the above expression, it is advantageous to choose T as small as possible. However, as explained below, an arbitrarily small T (or arbitrarily large ω) makes it difficult to detect the attack.

Assume that a Type 1 attack takes place on the sensor. Define Δ_{z_1} and Δ_{z_2} as

$$\Delta_{z_1} := |z_{1,\text{ss}} - z_{1,\text{ss,a}}|, \ \Delta_{z_2} := |z_{2,\text{ss}} - z_{2,\text{ss,a}}|, \tag{5.11}$$

and consider, for example, the plot of Δ_{z_2} versus $\omega\tau$, shown in Fig. 5.7. As seen in this figure, it is not desirable to select ω large for plants with large τ, since the value of Δ_{z_2} would be small, hence making it difficult to distinguish between the nominal and attacked scenarios. The underlying reason for this phenomenon is that the plant filters out the high

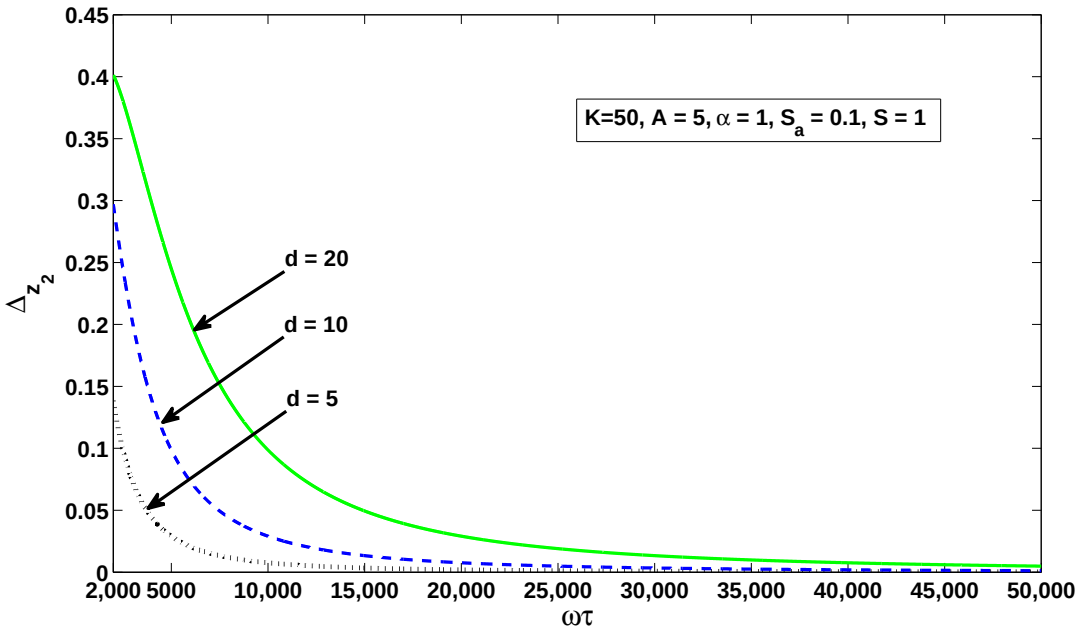


Figure 5.7: Δ_{z_2} vs. $\omega\tau$ under a Type 1 attack on the sensor

frequency sinusoidal signal. Thus, there exists a tradeoff between the choice of frequency of the sinusoidal signal and the difficulty of identifying the attack.

Under a Type 2 attack, since operation of the feedback control system is discontinued after the identification procedure is completed, T_{idm} is given by

$$T_{\text{idm}} = 3\tau + T. \tag{5.12}$$

The above expression applies to the case of a Type 3 attack as well.

To summarize this section, we note that the identification and mitigation procedures developed in this work are effective if $T_{\text{idm}} < T_c$.

5.5 Example: Application to Uranium Enrichment Centrifuge Control System

Consider a three-phase induction motor, whose transfer function between the input voltage and the rotational speed is given by (see [64]):

$$P(s) = \frac{157}{4s+1}. \tag{5.13}$$

Assume that this motor is operated in the closed-loop configuration of Fig. 5.5, with the parameters specified in Table 5.2. Further, assume that an attacker conducts a Type 1 attack on the sensor, with the parameters of the attack provided in Table 5.3. Given these data, the nominal steady state values of z_1 and z_2 can be computed as

$$z_{1,\text{ss}} = 52,\ z_{2,\text{ss}} = 311. \tag{5.14}$$

Similarly, the steady state values of z_1 and z_2 under the attack can be calculated as

$$z_{1,\text{ss},a} = 203,\ z_{2,\text{ss},a} = 307. \tag{5.15}$$

The trajectories of the plant output, y, are illustrated in Fig. 5.8. As seen in this figure, y deviates from the reference signal, r, after the attack takes place at time $t = 15$sec. As described in Section 5.4, the time required by y to reach the new steady state is $3\tau = 12$sec. Further, a duration of $T = 0.06$sec is required to calculate the new steady state values of

Table 5.2: Parameters of the control system

Gains of controller, actuator, and sensor	$K = 20$, $A = 2$, $S = 1$
Value of reference signal	$r = 528$
Amplitude and frequency of sinusoidal signal	$\alpha = 25$, $\omega = 100$

Table 5.3: Parameters of the attack

Attacked DC gain of sensor	$S_a = 0.5$
Time of attack	15sec

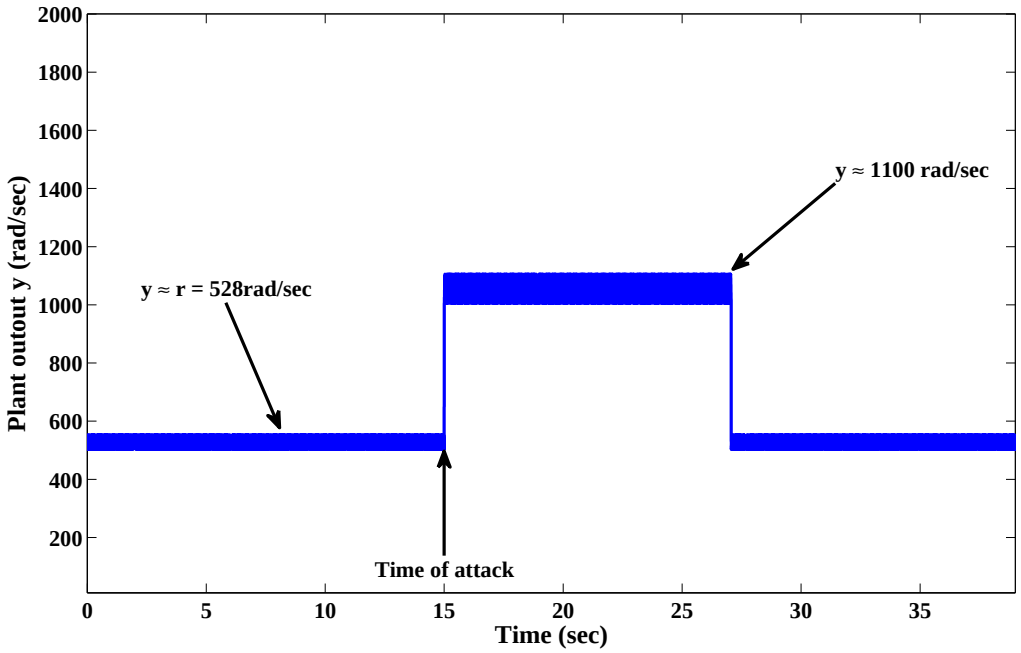


Figure 5.8: Trajectory of the plant output, y

the z's and identify the attack. Thus, the attack is identified at $t = 15 + 3\tau + T = 27.06$sec.

The application of the mitigation procedure at $t = 27.06$sec causes y to begin approaching the reference signal. Finally, after a further 12sec, normal operation of the plant is achieved (at $t = 39.06$sec).

The zoomed trajectories of z_1 and z_2, in the vicinity of $t = 15$sec, are illustrated in Fig. 5.9. The trends of these trajectories can be explained as above.

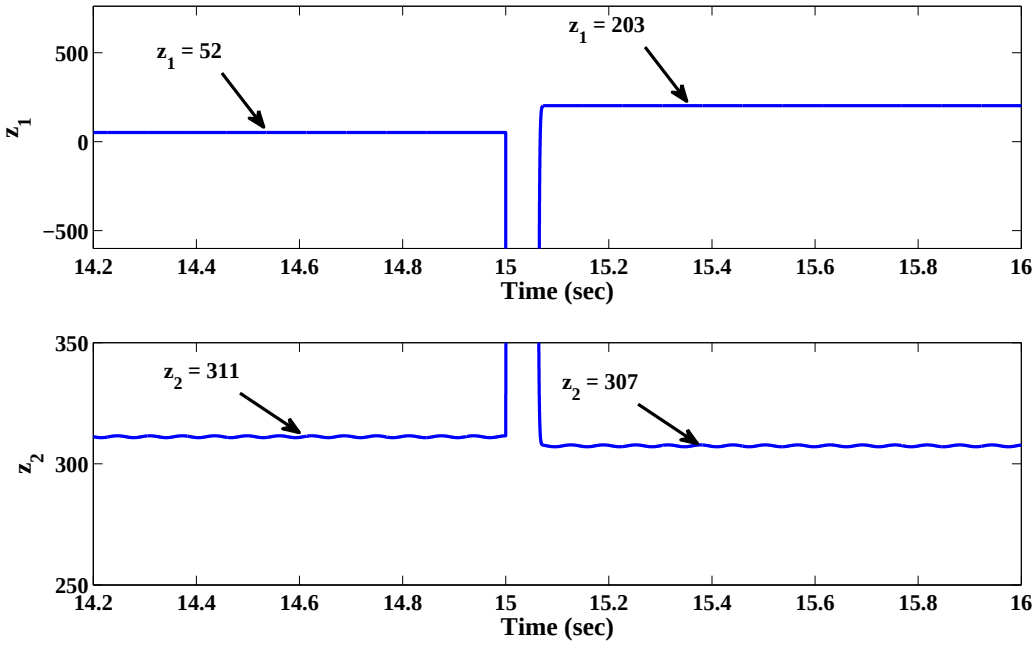


Figure 5.9: Zoomed trajectories of z_1 and z_2, shown in the vicinity of $t = 15$sec

CHAPTER VI

Conclusions and Future Research

6.1 Summary of Results Obtained To-Date

This work designs, analyzes, and evaluates the performance of resilient monitoring and control systems. A brief summary of the key results obtained to-date are as follows:

The development of the resilient monitoring system (RMS) is carried out based on the following five techniques: Data quality acquisition, process variable assessment, plant condition evaluation, sensor network adaptation, and decentralized assessments with inferences, a.k.a., knowledge fusion. Each of these techniques are analyzed rigorously, and are used to design a five-layer RMS architecture. The performance of the resulting RMS is evaluated using a power plant application.

The development of the resilient control system (RCS) is carried out using two different approaches: The model predictive control (MPC) -based approach and the synchronous detection (SD) -based approach. Initial investigations into these approaches show that both of them may be viable for the design of RCS.

Numerous research problems, however, remain open. They are outlined below:

6.2 Problems in Resilient Monitoring Systems

- *Problems related to data quality acquisition:*
  - Investigating efficacy of the probe-based data quality acquisition technique for attackers other than those modifying the expected value of sensor measurements.
  - Improving temporal properties of DQ acquisition. As shown in Chapter II, DQ is acquired in about 5sec. It would be desirable to achieve this an order of magnitude faster. A potential approach is inferring DQ from the transient, rather than the steady state, response of a process variable to the probe.
  - Introducing and investigating other than probe-based DQ acquisition techniques. Perhaps, this could be accomplished by considering inference diagrams of process variables and continually monitoring the level of their satisfaction in the data provided by the sensors.
  - Investigating the possibility of assigning DQ based on a reputation fusion mechanism. Reference [68] introduced a framework for assigning a "reputation" to each sensor $\mathbf{S}_i$, based on the Dempster-Shafer combination of the individual reputations assigned to $\mathbf{S}_i$ by several neighboring sensors. Is there a way to integrate the probe-based technique within the above framework? If so, would the resulting DQ of the sensor be more representative of its actual condition $-$ attacked or operating normally?
- *Problems related to process variable assessment:*
  - Introducing and investigating different than (2.12) models of coupling between the sensor data and process variables. Similarly, investigating different (as compared with the believability (2.11)) effects of DQ on process variable assessment.

- In the current work, the sensor data and DQ's are utilized to assess the process variable pmf's (i.e., h-procedure (2.15), (2.16)) under the assumption that the state of the sensor network remains constant. Are there convergent techniques to accomplish this when the state of the sensor network is non-stationary? If so, the temporal properties of the RMS could be improved substantially.

- *Problems related to sensor network adaptation:*

  - Utilizing other than (2.42) rational controllers. The goal here is to devise rational controllers with faster adaptation rates (see [7] where various types of rational controllers are introduced and analyzed).

  - Introducing and analyzing other than entropy-based penalty functions. Perhaps, there exists a penalty function that would lead to lower uncertainty in process variable assessment than the entropy.

  - Investigating a possibility of associating a rational controller with each sensor of the sensor network. Although this would lead to a non-stationary adaptation environment, it would result, if convergent, in a substantial improvement of adaptation rates.

- *Problems related to decentralized assessments with inferences:*

  - Deriving necessary and sufficient conditions for the optimality of decentralized inferences. At present, only sufficient conditions for the optimality are available (see Chapter III). Do there exist both necessary and sufficient conditions that guarantee the optimality of decentralized inferences?

  - Characterizing the monotonicity of the Dempster-Shafer combination rule. The conditions for lossless decentralization (derived in Chapter III) involve the assumption that the Dempster-Shafer rule is monotonic on the set of process variable pmf's. However, no constructive methods are currently available to verify

if this assumption holds. Developing such methods to verify the monotonicity is an important problem.

- *Problem related to plant assessment:*
  - Investigating a possibility of recursive plant assessment. Because recursive application of the Jeffrey rule may lead to paradoxical results (see Chapter II), in the current work we apply this rule non-recursively, which slows down the plant pmf assessment. So, modifying this rule or developing a new one, which would permit a recursive application, is an important problem.

6.3 Problems in Resilient Control Systems

These problems are divided into two categories: Problems related to the MPC-based approach and problems related to the SD-based approach. They are listed below:

Problems related to MPC-based approach:

- Evaluating the controller from the point of view of stability and performance (e.g., reference tracking and disturbance rejection). Is the closed loop system stable? Is the RCS effective at rejecting disturbances? These are important questions to be addressed.
- Extending the theory to more complex plants, e.g., MIMO plants. At present, the theory is under development for SISO plants. Extending the development to more complex plants is an important problem.
- Application of the approach developed to the power plant model considered in Chapter II.

Problems related to SD-based approach:

- Application of the developed techniques to models of power plants and power grids.

Solutions of these problems will enable designing effective resilient monitoring and control systems for critical infrastructures (e.g., chemical plants, power systems and power grids, computer networks, civil engineering objects) and complex individual plants (e.g., aircraft and space structures).

APPENDICES

APPENDIX A

Proofs of Theorems Stated in Chapter II

A.1 Proof of Theorem II.1, Part 1

The proof consists of the following five steps:

Step 1: Calculate the expected value of the set point $h^*_\sigma(s_{n+1})$, $\sigma \in \Sigma_V$:

Since the sensor measurements are stationary, the expected value of $h^*_\sigma(s_{n+1})$, $\sigma \in \Sigma_V$, is independent of $n \in \mathbb{N}$, and can be denoted as $E\left\{h^*_\sigma(s_{n+1})\right\} = \mu^{h^*}_\sigma$, $\sigma \in \Sigma_V$. This quantity is calculated from (2.16) in the following manner:

$$\mu^{h^*}_\sigma = \beta_{\mathbf{S}} \cdot p[S = \sigma] + \frac{1 - \beta_{\mathbf{S}}}{|\Sigma_V| - 1} \left\{1 - p[S = \sigma]\right\}, \ \sigma \in \Sigma_V. \tag{A.1}$$

Then, from (2.11),

$$\mu^{h^*}_\sigma = DQ_{\mathbf{S}} \cdot p[S = \sigma] + \frac{1 - DQ_{\mathbf{S}}}{|\Sigma_V|}, \ \sigma \in \Sigma_V. \tag{A.2}$$

Step 2: Evaluate $\lim_{n\to\infty} E\left\{h_\sigma(n)\right\}$, $\sigma \in \Sigma_V$:

First, using (2.15), express $h_\sigma(n)$, $\sigma \in \Sigma_V$, in terms of the initial condition, $h_\sigma(0)$,

$\sigma \in \Sigma_V$, and the sequence of set points, $h^*_\sigma(s_n)$, $n = 1, 2, ..., n$, $\sigma \in \Sigma_V$, as follows:

$$h_\sigma(n) = (1-\epsilon_h)^n h_\sigma(0) + \epsilon_h \sum_{i=1}^{n} (1-\epsilon_h)^{n-i} h^*_\sigma(s_i),\ \sigma \in \Sigma_V, \tag{A.3}$$

where $0 < \epsilon_h << 1$ is the step of the h-procedure. Next, take the expected value of both sides of (A.3) to obtain

$$\begin{aligned} E\left\{h_\sigma(n)\right\} &= (1-\epsilon_h)^n h_\sigma(0) + \mu^{h^*}_\sigma \epsilon_h \sum_{i=1}^{n} (1-\epsilon_h)^{n-i},\ \sigma \in \Sigma_V, \\ &= (1-\epsilon_h)^n h_\sigma(0) \\ &\quad + \mu^{h^*}_\sigma \epsilon_h \left(1 + 1 - \epsilon_h + (1-\epsilon_h)^2 + ... + (1-\epsilon_h)^{n-1}\right),\ \sigma \in \Sigma_V. \end{aligned} \tag{A.4}$$

Using the fact that $\epsilon_h < 1$, it can be shown that the limit of (A.4) as $n \to \infty$ is given by

$$\begin{aligned} \lim_{n\to\infty} E\left\{h_\sigma(n)\right\} &= \mu^{h^*}_\sigma \epsilon_h \tfrac{1}{1-(1-\epsilon_h)},\ \sigma \in \Sigma_V, \\ &= \mu^{h^*}_\sigma,\ \sigma \in \Sigma_V. \end{aligned} \tag{A.5}$$

Step 3: Evaluate $\lim_{n\to\infty} E\left\{h^2_\sigma(n)\right\}$, $\sigma \in \Sigma_V$:

Using (A.3), we obtain

$$\begin{aligned} \left\{h_\sigma(n)\right\}^2 &= \left((1-\epsilon_h)^n h_\sigma(0) + \epsilon_h \sum_{i=1}^{n} (1-\epsilon_h)^{n-i} h^*_\sigma(s_i)\right)^2,\ \sigma \in \Sigma_V, \\ &= (1-\epsilon_h)^{2n} h^2_\sigma(0) + \epsilon^2_h \left(\sum_{i=1}^{n} (1-\epsilon_h)^{n-i} h^*_\sigma(s_i)\right)^2 \\ &\quad + 2(1-\epsilon_h)^n \epsilon_h h_\sigma(0) \sum_{i=1}^{n} (1-\epsilon_h)^{n-i} h^*_\sigma(s_i),\ \sigma \in \Sigma_V. \end{aligned} \tag{A.6}$$

Taking the expected value of both sides of (A.6), and applying the limit as $n \to \infty$, we obtain

$$\lim_{n\to\infty} E\left\{h^2_\sigma(n)\right\} = \lim_{n\to\infty} \epsilon^2_h E\left\{\left(\sum_{i=1}^{n} (1-\epsilon_h)^{n-i} h^*_\sigma(s_i)\right)^2\right\},\ \sigma \in \Sigma_V. \tag{A.7}$$

Further, (A.7) can be rewritten as

$$\begin{aligned}\lim_{n\to\infty} E\left\{h_\sigma^2(n)\right\} &= \lim_{n\to\infty}\Bigg\{\epsilon_h^2[v_\sigma^{h^*}-(\mu_\sigma^{h^*})^2]\frac{1-(1-\epsilon_h)^{2n}}{1-(1-\epsilon_h)^2} \\ &\qquad +(\mu_\sigma^{h^*})^2[1-(1-\epsilon_h)^n]^2\Bigg\},\ \sigma\in\Sigma_V,\end{aligned} \tag{A.8}$$

where $v_\sigma^{h^*}$, $\sigma\in\Sigma_V$, denotes the second moment of $h_\sigma^*(s_n)$, $\sigma\in\Sigma_V$, i.e., $v_\sigma^{h^*}:=E\left\{[h_\sigma^*(s_n)]^2\right\}$, $\forall n$, $\sigma\in\Sigma_V$. The limit in (A.8) is evaluated as

$$\lim_{n\to\infty} E\left\{h_\sigma^2(n)\right\} = [v_\sigma^{h^*}-(\mu_\sigma^{h^*})^2]\tfrac{\epsilon_h^2}{1-(1-\epsilon_h)^2}+(\mu_\sigma^{h^*})^2,\ \sigma\in\Sigma_V. \tag{A.9}$$

Since $\frac{\epsilon_h^2}{1-(1-\epsilon_h)^2}=\frac{\epsilon_h}{2-\epsilon_h}$ and ϵ_h is sufficiently small, we have $\frac{\epsilon_h}{2-\epsilon_h}\approx 0$. Therefore,

$$\lim_{n\to\infty} E\left\{h_\sigma^2(n)\right\}\approx(\mu_\sigma^{h^*})^2,\ \sigma\in\Sigma_V. \tag{A.10}$$

Step 4: Evaluate $\lim_{n\to\infty} E\left\{\left(h_\sigma(n)-\mu_\sigma^{h^*}\right)^2\right\}$, $\sigma\in\Sigma_V$:

This quantity can be expressed as $\lim_{n\to\infty} E\left\{h_\sigma^2(n)-(\mu_\sigma^{h^*})^2\right\}$, $\sigma\in\Sigma_V$, which, from (A.10), is close to zero. Therefore,

$$\lim_{n\to\infty} E\left\{\left(h_\sigma(n)-\mu_\sigma^{h^*}\right)^2\right\}\approx 0,\ \sigma\in\Sigma_V. \tag{A.11}$$

Step 5: Use Chebyshev's inequality to obtain the desired result:

From Chebyshev's inequality,

$$\lim_{n\to\infty} P\left(\left|h_\sigma(n)-\mu_\sigma^{h^*}\right|>\alpha\right)<\lim_{n\to\infty}\frac{E\left\{\left(h_\sigma(n)-\mu_\sigma^{h^*}\right)^2\right\}}{\alpha^2},\ \forall\alpha>0,\ \sigma\in\Sigma_V. \tag{A.12}$$

From (A.11), we can conclude that the right hand side of (A.12) is close to zero. Therefore, $\lim_{n\to\infty} P\left(\left|h_\sigma(n)-\mu_\sigma^{h^*}\right|>\alpha\right)\approx 0$, $\forall\alpha>0$, $\sigma\in\Sigma_V$. Moreover, if the recursive state, $h_\sigma(n)$, $\sigma\in\Sigma_V$, is expressed as $h_\sigma(n;\epsilon_h)$, $\sigma\in\Sigma_V$, where ϵ_h is treated as a parameter, the

following can be concluded using the steps described above:

$$\lim_{\epsilon_h \to 0} \lim_{n \to \infty} P\left(\left|h_\sigma(n;\epsilon_h) - \mu_\sigma^{h^*}\right| > \alpha\right) = 0,\ \forall \alpha > 0,\ \sigma \in \Sigma_V. \tag{A.13}$$

This completes the proof of Part 1 of Theorem II.1. ■

A.2 Proof of Theorem II.1, Part 2

The proof is based on the following three lemmas:

Lemma A.1. *Consider the recursive procedure* (2.15), (2.16), (2.18). *Then,*

$$0 \le \lim_{n \to \infty} h_\sigma(n) \le 1,\ \sigma \in \Sigma_V. \tag{A.14}$$

Proof. As it follows from (2.15),

$$\begin{gathered} h_\sigma(n) = w_0(n) h_\sigma(0) + \sum_{i=1}^{n} w_i(n) h_\sigma^*(s_i),\ \sigma \in \Sigma_V, \\ w_0(n) := \prod_{i=1}^{n} [1 - \epsilon_h(i-1)],\ w_i(n) := \epsilon_h(i-1) \prod_{j=i}^{n-1} [1 - \epsilon_h(j)],\ i = 1, 2, ..., n. \end{gathered} \tag{A.15}$$

Thus, $h_\sigma(n) \ge 0$, $\forall n$ and $\forall \sigma$. Also, it can be shown that, due to (2.18),

$$\sum_{i=0}^{n} w_i(n) = 1,\ \lim_{n \to \infty} w_0(n) = 0. \tag{A.16}$$

Therefore,

$$\begin{aligned} \lim_{n \to \infty} h_\sigma(n) &= \lim_{n \to \infty} w_0(n) h_\sigma(0) + \lim_{n \to \infty} \sum_{i=1}^{n} w_i(n) h_\sigma^*(s_i),\ \sigma \in \Sigma_V, \\ &= \lim_{n \to \infty} \sum_{i=1}^{n} w_i(n) h_\sigma^*(s_i) \le \lim_{n \to \infty} \sum_{i=1}^{n} w_i(n),\ \sigma \in \Sigma_V, \end{aligned} \tag{A.17}$$

where the last inequality is due to (2.16). Finally, in view of (A.16), this inequality becomes

$\lim_{n\to\infty} h_\sigma(n) \leq \lim_{n\to\infty}[1 - w_0(n)] = 1$, $\sigma \in \Sigma_V$. ■

Lemma A.2. *Under the assumptions of Theorem II.1, the expected value of the set point,* $h^*_\sigma(s_n)$, $\sigma \in \Sigma_V$, $n \in \mathbb{N}$, *is given by*

$$E[h^*_\sigma(s_n)] = p[S = \sigma]DQ_{\mathbf{S}} + \frac{1 - DQ_{\mathbf{S}}}{|\Sigma_V|},\ \sigma \in \Sigma_V,\ n \in \mathbb{N}. \tag{A.18}$$

Proof. Follows directly from (2.16). ■

Thus, according to this lemma, the expected value of $h^*_\sigma(s_n)$ is independent of $n \in \mathbb{N}$, and can be denoted as $E[h^*_\sigma(s_n)] = \mu_{h^*_\sigma}$, $\sigma \in \Sigma_V$.

To formulate the next lemma, introduce the function

$$f\left(h_\sigma(n)\right) := \frac{1}{2}\left[h^*_\sigma(s_{n+1}) - h_\sigma(n)\right]^2,\ \sigma \in \Sigma_V. \tag{A.19}$$

Lemma A.3. *The unique minimum of* $E\left[f\left(h_\sigma(n)\right)\right]$, $\sigma \in \Sigma_V$, *is attained at*

$$\arg\min_{h_\sigma(n)} E\left[f\left(h_\sigma(n)\right)\right] = \mu_{h^*_\sigma},\ \sigma \in \Sigma_V. \tag{A.20}$$

Proof. Clearly, $E\left[f\left(h_\sigma(n)\right)\right]$ is differentiable and convex in $h_\sigma(n)$ and, therefore, its unique minimum is attained at

$$\frac{\partial}{\partial h_\sigma(n)} E\left[f\left(h_\sigma(n)\right)\right] = 0,\ \sigma \in \Sigma_V. \tag{A.21}$$

Due to (A.19), this expression becomes $h_\sigma(n) - \mu_{h^*_\sigma} = 0$, implying that for any fixed $n \in \mathbb{N}$, the solution of the minimization problem is $h^{\min}_\sigma(n) = \mu_{h^*_\sigma}$, $\sigma \in \Sigma_V$. ■

Proof of Theorem II.1, Part 2: The proof is based on showing that for large n, the recursive procedure (2.15), (2.16), (2.18) solves the aforementioned minimization problem, and, therefore, $h_\sigma(n)$ converges to $\mu_{h^*_\sigma}$, $\sigma \in \Sigma_V$, almost surely.

Since $f\left(h_\sigma(n)\right)$, $\sigma \in \Sigma_V$, is continuously differentiable and convex, there exists a scalar $0 \leq \gamma \leq 1$ such that

$$\begin{aligned} f\left(h_\sigma(n+1)\right) &= f\left(h_\sigma(n)\right) + \left[h_\sigma(n+1) - h_\sigma(n)\right] \left.\frac{\partial f}{\partial h_\sigma(n)}\right|_{h_\sigma(n)=h_\sigma(n)} \\ &+ \left.\frac{[h_\sigma(n+1)-h_\sigma(n)]^2}{2} \frac{\partial^2 f}{\partial h_\sigma^2(n)}\right|_{h_\sigma(n)=h_\sigma(n)+\gamma[h_\sigma(n+1)-h_\sigma(n)]}, \ \sigma \in \Sigma_V. \end{aligned} \tag{A.22}$$

From (A.19) and (2.15), (2.16), we obtain

$$\begin{aligned} f\left(h_\sigma(n+1)\right) &= f\left(h_\sigma(n)\right) - \epsilon_h(n) \left[\frac{\partial f}{\partial h_\sigma(n)}\right]^2 \\ &+ \frac{\epsilon_h^2(n)}{2}[h_\sigma^*(s_{n+1}) - h_\sigma(n)]^2, \ \sigma \in \Sigma_V. \end{aligned} \tag{A.23}$$

Using the summation of both sides of (A.23), we obtain:

$$\begin{aligned} f\left(h_\sigma(n)\right) &= f\left(h_\sigma(0)\right) - \sum_{n=0}^{n-1} \epsilon_h(n) \left[\frac{\partial f}{\partial h_\sigma(n)}\right]^2 \\ &+ \sum_{n=0}^{n-1} \frac{\epsilon_h^2(n)}{2}[h_\sigma^*(s_{n+1}) - h_\sigma(n)]^2, \ \sigma \in \Sigma_V. \end{aligned} \tag{A.24}$$

Now, consider the limit of (A.24) as $n \to \infty$. Since $h_\sigma(n)$ is bounded for all n (see Lemma A.1), the left hand side of the above equation is a finite positive number. Due to the same reason, the term $[h_\sigma^*(s_{n+1}) - h_\sigma(n)]^2$ is bounded for all n, implying that there exists a positive K, such that $[h_\sigma^*(s_{n+1}) - h_\sigma(n)]^2 \leq K$, $\forall n$. Thus,

$$\begin{aligned} \lim_{n\to\infty} f\left(h_\sigma(n)\right) &\leq f\left(h_\sigma(0)\right) - \lim_{n\to\infty} \sum_{n=0}^{n-1} \epsilon_h(n) \left[\frac{\partial f}{\partial h_\sigma(n)}\right]^2 \\ &+ \frac{K}{2} \lim_{n\to\infty} \sum_{n=0}^{n-1} \epsilon_h^2(n), \ \sigma \in \Sigma_V. \end{aligned} \tag{A.25}$$

Observe that since $\sum_{n=0}^{\infty} \epsilon_h^2(n) < \infty$, the last term in the right hand side of (A.25) is bounded. Now, suppose $\frac{\partial f}{\partial h_\sigma(n)}$ does not go to 0 as n tends to ∞. Then the expression $\sum_{n=0}^{\infty} \epsilon_h(n) \left[\frac{\partial f}{\partial h_\sigma(n)}\right]^2$ is unbounded (due to $\sum_{n=0}^{\infty} \epsilon_h(n) = \infty$) and the right hand side of (A.25) becomes $-\infty$. This is a contradiction, since the left hand side is positive and

bounded. Therefore, $\frac{\partial f}{\partial h_\sigma(n)} \to 0$ as $n \to \infty$ almost surely (a.s.).

From the above arguments, $E\left[\frac{\partial f}{\partial h_\sigma(n)}\right] \to 0$ as $n \to \infty$. Furthermore, due to the linearity of expectation, $\frac{\partial}{\partial h_\sigma(n)} E[f(h_\sigma(n))] \to 0$ as $n \to \infty$, implying that the condition (A.21) is satisfied. Therefore, from Lemma A.3, it is clear that $\lim_{n\to\infty} h_\sigma(n) = \mu_{h^*_\sigma}$, $\sigma \in \Sigma_V$, a.s. Finally, using Lemma A.2, we conclude that $\lim_{n\to\infty} h_\sigma(n) = p[S = \sigma]DQ_{\mathbf{S}} + \frac{1-DQ_{\mathbf{S}}}{|\Sigma_V|}$, $\sigma \in \Sigma_V$, a.s. ■

A.3 Proof of Theorem II.2

Since $h_\sigma(n)$ is convergent a.s., for every ϵ, there exists $n_0(\epsilon)$, such that $P\left[|h_{\mathrm{N}_V}(n) - h^{\mathrm{ss}}_{\mathrm{N}_V}| < \epsilon\right] > 1 - \epsilon$, $\forall n > n_0(\epsilon)$. Therefore, for sufficiently large n, equation (2.40) can be rewritten as

$$k_{\mathrm{N}_G}(n+1) = F\left(k_{\mathrm{N}_G}(n)\right) + \mathcal{O}(\epsilon), \tag{A.26}$$

where

$$F\left(k_{\mathrm{N}_G}(n)\right) \;:=\; \left[\frac{a h^{\mathrm{ss}}_{\mathrm{N}_V}}{a k_{\mathrm{N}_G}(n) + [1-a][1-k_{\mathrm{N}_G}(n)]} + \frac{[1-a][1-h^{\mathrm{ss}}_{\mathrm{N}_V}]}{[1-a]k_{\mathrm{N}_G}(n) + a[1-k_{\mathrm{N}_G}(n)]}\right] k_{\mathrm{N}_G}(n),$$

and $\mathcal{O}(\epsilon)$ represents terms of order ϵ. Omitting these terms, equation (A.26) is approximated as

$$k_{\mathrm{N}_G}(n+1) = F\left(k_{\mathrm{N}_G}(n)\right). \tag{A.27}$$

It can be shown that the system (A.27) has three equilibria,

$$k^*_{\mathrm{N}_G} = 1,\; k^{**}_{\mathrm{N}_G} = 0,\; k^{***}_{\mathrm{N}_G} = \frac{h^{\mathrm{ss}}_{\mathrm{N}_V} - a}{1-2a}. \tag{A.28}$$

Based on the perturbation theory ([69]), for ϵ sufficiently small, stability properties of

(A.27) are the same as (A.26). To analyze stability, consider the Jacobians of $F(\cdot)$ at each equilibrium:

$$\begin{aligned} A_1 &= \left.\frac{\partial F}{\partial k_{\mathrm{N}_G}}\right|_{k^{*}_{\mathrm{N}_G}} = \frac{[1-a]^2+[2a-1]h^{\mathrm{ss}}_{\mathrm{N}_V}}{a[1-a]},\ A_2 = \left.\frac{\partial F}{\partial k_{\mathrm{N}_G}}\right|_{k^{**}_{\mathrm{N}_G}} = \frac{a^2+[1-2a]h^{\mathrm{ss}}_{\mathrm{N}_V}}{a[1-a]}, \\ A_3 &= \left.\frac{\partial F}{\partial k_{\mathrm{N}_G}}\right|_{k^{***}_{\mathrm{N}_G}} = \frac{a[1-a]}{h^{\mathrm{ss}}_{\mathrm{N}_V}[1-h^{\mathrm{ss}}_{\mathrm{N}_V}]}. \end{aligned} \tag{A.29}$$

Suppose $h^{\mathrm{ss}}_{\mathrm{N}_V} > 1-a$. Since $0 < a < 0.5$, we have $A_1 < 1$, $A_2 > 1$, and $A_3 > 1$, implying that $k^{*}_{\mathrm{N}_G}$ is asymptotically stable, while $k^{**}_{\mathrm{N}_G}$ and $k^{***}_{\mathrm{N}_G}$ are not. Therefore, $k_{\mathrm{N}_G}(n)$ converges locally to $k^{*}_{\mathrm{N}_G}$ as $n \to \infty$, which proves Part 1 of the theorem. Parts 2 and 3 can be proved similarly. ■

APPENDIX B

Parameters of Simulations Reported in Chapter II

B.1 Parameters of power plant and monitoring system

This subsection provides parameters of the power plant and monitoring system that have been used in simulations reported in Subsection 2.8.2. Note that these parameters are selected for illustration purposes and do not reflect the physical nature of the quantities involved.

B.1.1 Sub-plants, process variables, and sensors

B.1.1.1 Statistical models of the sub-plants

As mentioned in Subsection 2.8.1, these models are defined by conditional probabilities of process variables given the status of a sub-plant $G_i \in \{\mathrm{N}_{G_i}, \mathrm{A}_{G_i}\}$, $i \in \{\mathrm{B}, \mathrm{HT}, \mathrm{RP}, \mathrm{LT}\}$. Accordingly, we quantify these models as follows:

- Boiler: $P[V_1 = \mathrm{N}_{\mathrm{V}_1} | G_{\mathrm{B}} = \mathrm{N}_{\mathrm{G}_{\mathrm{B}}}] = P[V_1 = \mathrm{L}_{\mathrm{V}_1} | G_{\mathrm{B}} = \mathrm{A}_{\mathrm{G}_{\mathrm{B}}}] = 0.95$; all other components of this pmf are 0.05.

- High pressure turbine: $P[V_2 \in \{\mathrm{L}_{(1)\mathrm{V}_2}, \mathrm{N}_{\mathrm{V}_2}\} | G_{\mathrm{HT}} = \mathrm{N}_{\mathrm{G}_{\mathrm{HT}}}] = 0.90$,
  $P[V_2 \in \{\mathrm{VL}_{\mathrm{V}_2}, \mathrm{L}_{(2)\mathrm{V}_2}\} | G_{\mathrm{HT}} = \mathrm{A}_{\mathrm{G}_{\mathrm{HT}}}] = 0.90$; all other components are 0.1.

- Reheat pipe: $P[V_3 \in \{\mathrm{L}_{(1)\mathrm{V}_3}, \mathrm{N}_{\mathrm{V}_3}\} | G_{\mathrm{RP}} = \mathrm{N}_{\mathrm{G}_{\mathrm{RP}}}] = 0.88$,
  $P[V_3 \in \{\mathrm{VL}_{\mathrm{V}_3}, \mathrm{L}_{(2)\mathrm{V}_3}\} | G_{\mathrm{RP}} = \mathrm{A}_{\mathrm{G}_{\mathrm{RP}}}] = 0.91$, $P[V_3 \in \{\mathrm{VL}_{\mathrm{V}_3}, \mathrm{L}_{(2)\mathrm{V}_3}\} | G_{\mathrm{RP}} = \mathrm{N}_{\mathrm{G}_{\mathrm{RP}}}] = 0.12$, and $P[V_3 \in \{\mathrm{L}_{(1)\mathrm{V}_3}, \mathrm{N}_{\mathrm{V}_3}\} | G_{\mathrm{RP}} = \mathrm{A}_{\mathrm{G}_{\mathrm{RP}}}] = 0.09$.

- Low pressure turbine: $P[V_4 \in \{\mathrm{VL}_{(1)\mathrm{V}_4}, \mathrm{L}_{(1)\mathrm{V}_4}, \mathrm{M}_{(1)\mathrm{V}_4}, \mathrm{N}_{\mathrm{V}_4}\} | G_{\mathrm{LT}} = \mathrm{N}_{\mathrm{G}_{\mathrm{LT}}}] = 0.91$,
  $P[V_4 \in \{\mathrm{VL}_{(2)\mathrm{V}_4}, \mathrm{L}_{(2)\mathrm{V}_4}, \mathrm{M}_{(2)\mathrm{V}_4}, \mathrm{H}_{\mathrm{V}_4}\} | G_{\mathrm{LT}} = \mathrm{A}_{\mathrm{G}_{\mathrm{LT}}}] = 0.92$,
  $P[V_4 \in \{\mathrm{VL}_{(2)\mathrm{V}_4}, \mathrm{L}_{(2)\mathrm{V}_4}, \mathrm{M}_{(2)\mathrm{V}_4}, \mathrm{H}_{\mathrm{V}_4}\} | G_{\mathrm{LT}} = \mathrm{N}_{\mathrm{G}_{\mathrm{LT}}}] = 0.09$, and
  $P[V_4 \in \{\mathrm{VL}_{(1)\mathrm{V}_4}, \mathrm{L}_{(1)\mathrm{V}_4}, \mathrm{M}_{(1)\mathrm{V}_4}, \mathrm{N}_{\mathrm{V}_4}\} | G_{\mathrm{LT}} = \mathrm{A}_{\mathrm{G}_{\mathrm{LT}}}] = 0.08$.

B.1.1.2 Models of process variables and sensors

The domains of the process variables and their d.c. gains are specified in Table B.1.

Without loss of generality, we assume that the process variables and the sensor measure-

Table B.1: Domains and d.c. gains of process variables

Process variables	Domains	Values of R's (see (2.2))	d.c. gains (see (2.3))
$\tilde{V}_1$	$[5, 100]$	$R_1 = 50$	$\alpha^{\mathrm{L}}_{\mathbf{v}_1} = 2, \alpha^{\mathrm{N}}_{\mathbf{v}_1} = 2.2$.
$\tilde{V}_2$	$[5, 25]$	$R_1 = 10$, $R_2 = 15$, $R_3 = 20$	$\alpha^{\mathrm{VL}}_{\mathbf{v}_2} = 0.5$, $\alpha^{\mathrm{L}_1}_{\mathbf{v}_2} = 0.6$, $\alpha^{\mathrm{L}_2}_{\mathbf{v}_2} = 0.7$, $\alpha^{\mathrm{N}}_{\mathbf{v}_2} = 0.8$.
$\tilde{V}_3$	$[5, 100]$	$R_1 = 20$, $R_2 = 40$, $R_3 = 50$	$\alpha^{\mathrm{VL}}_{\mathbf{v}_3} = 0.6$, $\alpha^{\mathrm{L}_{(1)}}_{\mathbf{v}_3} = 0.72$, $\alpha^{\mathrm{L}_{(2)}}_{\mathbf{v}_3} = 0.9$, $\alpha^{\mathrm{N}}_{\mathbf{v}_3} = 1.2$.
$\tilde{V}_4$	$[0.1, 20]$	$R_1 = 3$, $R_2 = 6$, $R_3 = 9$, $R_4 = 11$, $R_5 = 13$, $R_6 = 15$, $R_7 = 17$.	$\alpha^{\mathrm{VL}_{(1)}}_{\mathbf{v}_4} = 0.4$, $\alpha^{\mathrm{VL}_{(2)}}_{\mathbf{v}_4} = 0.42$, $\alpha^{\mathrm{L}_{(1)}}_{\mathbf{v}_4} = 0.46$, $\alpha^{\mathrm{L}_{(2)}}_{\mathbf{v}_4} = 0.48$, $\alpha^{\mathrm{M}_{(1)}}_{\mathbf{v}_4} = 0.53$, $\alpha^{\mathrm{M}_{(2)}}_{\mathbf{v}_4} = 0.56$, $\alpha^{\mathrm{N}}_{\mathbf{v}_4} = 0.6$, $\alpha^{\mathrm{H}}_{\mathbf{v}_4} = 0.63$.

ments are Gaussian random variables, $\tilde{V}_i \sim \mathcal{N}\left(\mu_{\tilde{V}_i}, \sigma_{\tilde{V}_i}\right)$ and $\tilde{S}_{ij} \sim \mathcal{N}\left(\mu_{\tilde{S}_{ij}}, \sigma_{\tilde{S}_{ij}}\right)$, $i = 1, 2, 3, 4$, $j = 1, 2$, where the expected values, $\mu_{\tilde{V}_i}$ and $\mu_{\tilde{S}_{ij}}$, are specified in Tables B.2 and B.3, respectively, for all attack scenarios considered in Section 2.8. Regarding the standard deviations of $\tilde{V}_i$ and $\tilde{S}_{ij}$, we assume that they are small enough so that the realizations of these random variables outside of the domains given in Table B.1 may be ignored. Specifically, they are selected as $\sigma_{\tilde{V}_i} = \sigma_{\tilde{S}_{ij}} = 0.01$, $i = 1, 2, 3, 4$, $j = 1, 2$.

Table B.2: Expected values of process variables

Attack scenario	$\mu_{\tilde{V}_1}$	$\mu_{\tilde{V}_2}$	$\mu_{\tilde{V}_3}$	$\mu_{\tilde{V}_4}$
1	80	23	75	16
2	80	23	75	16
3	80	23	44	12.1
4	80	18	76	16
5	30	12	23	10
6	30	12	15	5
7	20	7	10	5

Table B.3: Expected values of sensor measurements

Attack scenario	$\mu_{\tilde{S}_{11}}$	$\mu_{\tilde{S}_{12}}$	$\mu_{\tilde{S}_{21}}$	$\mu_{\tilde{S}_{22}}$	$\mu_{\tilde{S}_{31}}$	$\mu_{\tilde{S}_{32}}$	$\mu_{\tilde{S}_{41}}$	$\mu_{\tilde{S}_{42}}$
1	31	30	22	24	74	74.1	15.8	16.1
2	81	79	22	24	74	74.1	19.2	19.1
3	81	79	22	24	74	74.1	12.2	12.1
4	81	79	22	24	74	74.1	16.1	16.2
5	81	79	12.1	12.2	23	24	16.1	16.2
6	81	79	12.1	12.2	76	75	16.1	16.2
7	81	79	23	22	76	75	16.1	16.2

B.1.2 Parameters of monitoring system

B.1.2.1 Data quality assessment layer

- The amplitudes of the probing signals (2.6) are selected as follows: $A_{\mathbf{V}_1} = 2$, $A_{\mathbf{V}_2} = 0.6$, $A_{\mathbf{V}_3} = 0.7$, and $A_{\mathbf{V}_4} = 0.3$.
- The parameter ϵ, involved in (2.10), is selected as 0.02.
- The $PIC_{\max}$ in (2.10) for the sensors of B, HT, RP, and LT are 0.4, 0.06, 0.08, 0.03, respectively.

B.1.2.2 Process variables assessment layer

- The step size of the h-procedure (2.15) is selected as $\epsilon_h = 0.01$.
- The stopping rule is defined by $|h_\sigma(n+1) - h_\sigma(n)| < 10^{-4}$.

B.1.2.3 Adaptation layer

The parameters involved in (2.42) are selected as follows:

- The level of rationality of the rational controller is selected as $N = 2$.
- The maximum residence time is selected as $T_{\max} = 1$sec.
- The parameter β is chosen as 0.04.

APPENDIX C

Proofs of Lemmas and Theorems Stated in Chapter III

C.1 Proof of Lemma III.1

The proof of Lemma III.1 requires the following notations: Let n_j, $j = 1, 2, ..., M$, be the cardinality of Σ_{V_j}, $j = 1, 2, ..., M$, i.e.,

$$n_j := |\Sigma_{V_j}|,\ j = 1, 2, ..., M. \tag{C.1}$$

Further, let the pmf $\hat{p}_{y_j}[V_j]$, $y_j \in Y_j$, $j = 1, 2, ..., M$, be represented as the column vector $\mathbf{q}_{y_j,V_j} \in \mathbb{R}^{n_j}$, $j = 1, 2, ..., M$, i.e.,

$$\hat{p}_{y_j}[V_j] = \mathbf{q}_{y_j,V_j} := [q^{(1)}_{y_j,V_j}, q^{(2)}_{y_j,V_j}, \cdots, q^{(n_j)}_{y_j,V_j}]^\top,\ y_j \in Y_j,\ j = 1, 2, ..., M, \tag{C.2}$$

where $0 \leq q^{(1)}_{y_j,V_j}, q^{(2)}_{y_j,V_j}, ..., q^{(n_j)}_{y_j,V_j} \leq 1$ and $q^{(1)}_{y_j,V_j} + q^{(2)}_{y_j,V_j} + ... + q^{(n_j)}_{y_j,V_j} = 1$, $y_j \in Y_j$, $j = 1, 2, ..., M$. The inferred pmf, $\hat{p}_{y_j}[V_i]$, $y_j \in Y_j$, $i \neq j,\ i, j = 1, 2, ..., M$, is computed, as before, using the total probability formula, i.e.,

$$\hat{p}_{y_j}[V_i] = \sum_{\sigma \in \Sigma_{V_j}} P[V_i | V_j = \sigma] \hat{p}_{y_j}[V_j = \sigma],\ y_j \in Y_j,\ i \neq j,\ i, j = 1, 2, ..., M, \tag{C.3}$$

and can be similarly represented as the column vector $\mathbf{q}_{y_j,V_i} \in \mathbb{R}^{n_i}$, $i \neq j,\ i,j = 1,2,...,M$:

$$\hat{p}_{y_j}[V_i] = \mathbf{q}_{y_j,V_i} := [q^{(1)}_{y_j,V_i}, q^{(2)}_{y_j,V_i}, \cdots, q^{(n_i)}_{y_j,V_i}]^\top,\ y_j \in Y_j,\ i \neq j,\ i,j = 1,2,...,M, \quad \text{(C.4)}$$

where $0 \leq q^{(1)}_{y_j,V_i}, q^{(2)}_{y_j,V_i}, ..., q^{(n_i)}_{y_j,V_i} \leq 1$ and $q^{(1)}_{y_j,V_i} + q^{(2)}_{y_j,V_i} + ... + q^{(n_i)}_{y_j,V_i} = 1$, $i \neq j,\ i,j = 1,2,...,M$.

As assumed in Assumption III.1, the 2-norms of the columns of the matrix $P[V_i|V_j], i \neq j,\ i,j = 1,2,...,M$, are equal. Let the value of these 2-norms be denoted as $K_{V_i|V_j}$, $i \neq j$, $i,j = 1,2,....,M$, i.e.,

$$K_{V_i|V_j} := \|\mathbf{p}^{(1)}_{V_i|V_j}\|_2 = \|\mathbf{p}^{(2)}_{V_i|V_j}\|_2 = \cdots = \|\mathbf{p}^{(n_j)}_{V_i|V_j}\|_2,\ i \neq j,\ i,j = 1,2,...,M. \quad \text{(C.5)}$$

Similarly, as assumed in Assumption III.1, the angles between all pairs of columns of $P[V_i|V_j], i \neq j,\ i,j = 1,2,...,M$, are equal. Let the value of these angles be denoted as $\theta_{V_i|V_j}$, $i \neq j$, $i,j = 1,2,....,M$, i.e.,

$$\cos\theta_{V_i|V_j} := \frac{\langle \mathbf{p}^{(1)}_{V_i|V_j}, \mathbf{p}^{(2)}_{V_i|V_j} \rangle}{K^2_{V_i|V_j}} = \cdots = \frac{\langle \mathbf{p}^{(n_j)}_{V_i|V_j}, \mathbf{p}^{(n_j-1)}_{V_i|V_j} \rangle}{K^2_{V_i|V_j}},\ i \neq j,\ i,j = 1,2,...,M, \quad \text{(C.6)}$$

where $K_{V_i|V_j}, i \neq j,\ i,j = 1,2,...,M$, is the same as in (C.5).

Introduce the following lemma, which is used to prove Lemma III.1:

Lemma C.1. *Under Assumption III.1,*

$$\|\mathbf{q}_{y_j,V_i}\|_2^2 = K^2_{V_i|V_j}\left[(1-\cos\theta_{V_i|V_j})\|\mathbf{q}_{y_j,V_j}\|_2^2 + \cos\theta_{V_i|V_j}\right],\ i \neq j,\ i,j = 1,2,...,M. \quad \text{(C.7)}$$

Proof of Lemma C.1: Expression (C.3) can be re-written as

$$\mathbf{q}_{y_j,V_i} = q^{(1)}_{y_j,V_j}\mathbf{p}^{(1)}_{V_i|V_j} + q^{(2)}_{y_j,V_j}\mathbf{p}^{(2)}_{V_i|V_j} + ... + q^{(n_j)}_{y_j,V_j}\mathbf{p}^{(n_j)}_{V_i|V_j},\ i \neq j,\ i,j = 1,2,...,M, \quad \text{(C.8)}$$

where, as before, the $\mathbf{p}_{V_i|V_j}$'s, $i \neq j,\ i,j = 1,2,...,M$, are the columns of $P[V_i|V_j], i \neq j,\ \ i,j = 1,2,...,M$. Using the above equation, compute $\|\mathbf{q}_{y_j,V_i}\|_2^2$, $i \neq j,\ \ i,j = 1,2,...,M$, as

$$\begin{aligned}\|\mathbf{q}_{y_j,V_i}\|_2^2 &= [q_{y_j,V_j}^{(1)}]^2\|\mathbf{p}_{V_i|V_j}^{(1)}\|_2^2 + \cdots + [q_{y_j,V_j}^{(n_j)}]^2\|\mathbf{p}_{V_i|V_j}^{(n_j)}\|_2^2 \\ &\quad +2q_{y_j,V_j}^{(1)}q_{y_j,V_j}^{(2)}\langle\mathbf{p}_{V_i|V_j}^{(1)},\mathbf{p}_{V_i|V_j}^{(2)}\rangle \\ &\quad +\cdots+2q_{y_j,V_j}^{(n_j-1)}q_{y_j,V_j}^{(n_j)}\langle\mathbf{p}_{V_i|V_j}^{(n_j-1)},\mathbf{p}_{V_i|V_j}^{(n_j)}\rangle, i \neq j.\end{aligned} \tag{C.9}$$

Substitute (C.5) and (C.6) in the right hand side of (C.9) to get:

$$\begin{aligned}\|\mathbf{q}_{y_j,V_i}\|_2^2 &= K_{V_i|V_j}^2\|\mathbf{q}_{y_j,V_j}\|_2^2 \\ &\quad +K_{V_i|V_j}^2\cos\theta_{V_i|V_j}[2q_{y_j,V_j}^{(1)}q_{y_j,V_j}^{(2)}+\cdots+2q_{y_j,V_j}^{(n_j-1)}q_{y_j,V_j}^{(n_j)}].\end{aligned} \tag{C.10}$$

From the definition of $\mathbf{q}_{y_j,V_j}$, $y_j \in Y_j$, $j = 1,2,...,M$ (see (C.2)), we know that $q_{y_j,V_j}^{(1)} + q_{y_j,V_j}^{(2)} + \cdots + q_{y_j,V_j}^{(n_j)} = 1$, $y_j \in Y_j$, $j = 1,2,...,M$. Square both sides of this expression to get

$$\begin{aligned}&[q_{y_j,V_j}^{(1)}]^2 + \cdots + [q_{y_j,V_j}^{(n_j)}]^2 + 2q_{y_j,V_j}^{(1)}q_{y_j,V_j}^{(2)} + \cdots + 2q_{y_j,V_j}^{(n_j-1)}q_{y_j,V_j}^{(n_j)} = 1,\ j = 1,2,...,M, \\ &\implies \|\mathbf{q}_{y_j,V_j}\|_2^2 + 2q_{y_j,V_j}^{(1)}q_{y_j,V_j}^{(2)} + \cdots + 2q_{y_j,V_j}^{(n_j-1)}q_{y_j,V_j}^{(n_j)} = 1,\ j = 1,2,...,M.\end{aligned} \tag{C.11}$$

Clearly, from the second row of (C.11), we have $2q_{y_j,V_j}^{(1)}q_{y_j,V_j}^{(2)} + \cdots + 2q_{y_j,V_j}^{(n_j-1)}q_{y_j,V_j}^{(n_j)} = 1 - \|\mathbf{q}_{y_j,V_j}\|_2^2$, $y_j \in Y_j$, $j = 1,2,...,M$. Substitute this expression in the right hand side of (C.10) to get $\|\mathbf{q}_{y_j,V_i}\|_2^2 = K_{V_i|V_j}^2\left[(1-\cos\theta_{V_i|V_j})\|\mathbf{q}_{y_j,V_j}\|_2^2 + \cos\theta_{V_i|V_j}\right], y_j \in Y_j, i \neq j,\ i,j = 1,2,...,M$. This completes the proof of this lemma. ■

Proof of Lemma III.1: As assumed in the statement of the lemma, let the pmf's $\mathbf{q}_{y_j,V_j}$, $y_j \in Y_j$, $j = 1,2,...,M$, have the same entropy, i.e.,

$$H_2\{\mathbf{q}_{y_j,V_j}\} = c,\ y_j \in Y_j,\ j = 1,2,...,M,\ 0 \leq c \leq 1, \tag{C.12}$$

where, as before, the entropy is defined as

$$H_2\{\mathbf{q}_{y_j,V_j}\} = -\log_{n_j}(\|\mathbf{q}_{y_j,V_j}\|_2^2),\ y_j \in Y_j,\ j = 1, 2, ..., M. \tag{C.13}$$

The above equation can be re-written as

$$\|\mathbf{q}_{y_j,V_j}\|_2 = n_j^{-\frac{1}{2}H_2\{\mathbf{q}_{y_j,V_j}\}},\ y_j \in Y_j,\ j = 1, 2, ..., M. \tag{C.14}$$

Further, taking into account (C.12), we have

$$\|\mathbf{q}_{y_j,V_j}\|_2 = n_j^{-\frac{c}{2}},\ y_j \in Y_j,\ j = 1, 2, ..., M,\ 0 \leq c \leq 1, \tag{C.15}$$

which implies that the 2-norms of the pmf's $\mathbf{q}_{y_j,V_j}$, $y_j \in Y_j$, $j = 1, 2, ..., M$, are also the same.

As before, the inferred pmf's $\mathbf{q}_{y_j,V_i}$, $y_j \in Y_j$, $i \neq j$, $i, j = 1, 2, ..., M$, are computed using the above pmf's $\mathbf{q}_{y_j,V_j}$, $y_j \in Y_j$, $j = 1, 2, ..., M$, and the total probability formula (see (C.3)). As a result of Lemma C.1, the square of the 2-norm of $\mathbf{q}_{y_j,V_i}$, $y_j \in Y_j$, $i \neq j$, $i, j = 1, 2, ..., M$, can be expressed as

$$\|\mathbf{q}_{y_j,V_i}\|_2^2 = K_{V_i|V_j}^2 \left[(1 - \cos\theta_{V_i|V_j})n_j^{-c} + \cos\theta_{V_i|V_j}\right],\ y_j \in Y_j,\ i \neq j. \tag{C.16}$$

Clearly, the right hand side of the above equation is a constant, and can be denoted as

$$d_{ij} := K_{V_i|V_j}^2 \left[(1 - \cos\theta_{V_i|V_j})n_j^{-c} + \cos\theta_{V_i|V_j}\right],\ i \neq j,\ i, j = 1, 2, ..., M. \tag{C.17}$$

As in (C.13), the entropy of $\mathbf{q}_{y_j,V_i}$, $y_j \in Y_j$, $i \neq j$, $i, j = 1, 2, ..., M$, can be computed as

$$H_2\{\mathbf{q}_{y_j,V_i}\} = -\log_{n_i}(d_{ij}),\ y_j \in Y_j, i \neq j, i, j = 1, 2, ..., M, \tag{C.18}$$

where d_{ij}, $i \neq j$, $i, j = 1, 2, ..., M$, is defined in (C.17). Clearly, the entropies of the inferred pmf's are the same. This completes the proof of this lemma. ■

C.2 Proof of Theorem III.1

As assumed in Section 3.2, the state $y_j^* \in Y_j$, $j = 1, 2, ..., M$, represents the state of the sensor subnetwork $\mathbf{SN}_j$, $j = 1, 2, ..., M$, where the entropy of the pmf of V_j, $j = 1, 2, ..., M$, is minimized, i.e.,

$$H_2\{\mathbf{q}_{y_j^*,V_j}\} < H_2\{\mathbf{q}_{y_j,V_j}\},\ y_j \neq y_j^*,\ y_j \in Y_j,\ j = 1, 2, ..., M. \tag{C.19}$$

Due to the definition of the entropy (C.13), the above expression implies

$$\|\mathbf{q}_{y_j^*,V_j}\|_2 > \|\mathbf{q}_{y_j,V_j}\|_2,\ y_j \neq y_j^*,\ y_j \in Y_j,\ j = 1, 2, ..., M. \tag{C.20}$$

Based on Lemma C.1, the square of the 2-norm of the inferred pmf, $\mathbf{q}_{y_j^*,V_i}$, $i \neq j$, $i, j = 1, 2, ..., M$, can be expressed as

$$\|\mathbf{q}_{y_j^*,V_i}\|_2^2 = K_{V_i|V_j}^2 \left[(1 - \cos\theta_{V_i|V_j})\|\mathbf{q}_{y_j^*,V_j}\|_2^2 + \cos\theta_{V_i|V_j}\right],\ i \neq j. \tag{C.21}$$

Further, due to (C.20), the above equation can be re-written as the following inequality:

$$\|\mathbf{q}_{y_j^*,V_i}\|_2^2 > K_{V_i|V_j}^2 \left[(1 - \cos\theta_{V_i|V_j})\|\mathbf{q}_{y_j,V_j}\|_2^2 + \cos\theta_{V_i|V_j}\right],\ y_j \neq y_j^*,\ i \neq j. \tag{C.22}$$

However, the right hand side of (C.22) equals $\|\mathbf{q}_{y_j,V_i}\|_2^2$, $y_j \in Y_j$, $i \neq j$, $i, j = 1, 2, ..., M$ (see Lemma C.1). Thus,

$$\|\mathbf{q}_{y_j^*,V_i}\|_2 > \|\mathbf{q}_{y_j,V_i}\|_2,\ y_j \neq y_j^*,\ y_j \in Y_j,\ i \neq j,\ i, j = 1, 2, ..., M. \tag{C.23}$$

Due to the definition of the entropy (C.13), the above expression implies $H_2\{\mathbf{q}_{y_j^*,V_i}\} < H_2\{\mathbf{q}_{y_j,V_i}\}$, $y_j \neq y_j^*$, $y_j \in Y_j$, $i \neq j,\ i,j = 1,2,...,M$. This completes the proof of this theorem. ■

C.3 Proof of Lemma III.2

Consider the pmf's $\hat{p}_{y_j}[V_i]$, $y_j \in Y_j$, $i,j = 1,2,...,M$. It can be shown that the entropy of these pmf's takes values in

$$0 \leq H_2\{\hat{p}_{y_j}[V_i]\} \leq 1,\ y_j \in Y_j,\ i,j = 1,2,...,M, \tag{C.24}$$

where the maximum value of $H_2\{\cdot\}$ is attained at the uniform pmf,

$$p_{\text{unif}}[V_i] = \left[\frac{1}{n_i}, \frac{1}{n_i}, \cdots, \frac{1}{n_i}\right]^\top,\ i = 1,2,....,M, \tag{C.25}$$

and the minimum value of $H_2\{\cdot\}$ is attained at the pmf's $[1,0,0,...,0]^\top$, $[0,1,0,...,0]^\top$, ..., $[0,0,...,0,1]^\top$. Thus, we have

$$0 \leq H_2\{\hat{p}_{y_j}[V_i]\} \leq H_2\{p_{\text{unif}}[V_i]\},\ y_j \in Y_j,\ i,j = 1,2,...,M. \tag{C.26}$$

Now, consider the following system of inequalities:

$$\begin{aligned} H_2\{\hat{p}_{y_1}[V_i]\} &\leq H_2\{\hat{p}_{y_1}[V_i]\}, \\ H_2\{\hat{p}_{y_2}[V_i]\} &\leq H_2\{p_{\text{unif}}[V_i]\}, \\ H_2\{\hat{p}_{y_3}[V_i]\} &\leq H_2\{p_{\text{unif}}[V_i]\}, \\ &\vdots \\ H_2\{\hat{p}_{y_M}[V_i]\} &\leq H_2\{p_{\text{unif}}[V_i]\}, \end{aligned} \tag{C.27}$$

where the first of these inequalities is trivially satisfied, while the remaining are due to (C.26). As before, the pmf's $\hat{p}_{y_1}[V_i], \hat{p}_{y_2}[V_i], ..., \hat{p}_{y_M}[V_i]$, involved in the left hand side of

(C.27), can be combined using the Dempster-Shafer rule (see (3.13)) to obtain the pmf $\hat{p}_{(y_1,y_2,...,y_M)}[V_i]$. Further, it can be shown that the Dempster-Shafer combination of the pmf's $\hat{p}_{y_1}[V_i]$, $p_{\text{unif}}[V_i]$, $p_{\text{unif}}[V_i]$,..., $p_{\text{unif}}[V_i]$, involved in the right hand side of (C.27), results in the pmf $\hat{p}_{y_1}[V_i]$. Clearly, due to Assumption III.2, the above arguments imply

$$H_2\{\hat{p}_{(y_1,y_2,...,y_M)}[V_i]\} \leq H_2\{\hat{p}_{y_1}[V_i]\}. \tag{C.28}$$

Similarly, it can be shown that

$$\begin{aligned} H_2\{\hat{p}_{(y_1,y_2,...,y_M)}[V_i]\} &\leq H_2\{\hat{p}_{y_2}[V_i]\}, \\ H_2\{\hat{p}_{(y_1,y_2,...,y_M)}[V_i]\} &\leq H_2\{\hat{p}_{y_3}[V_i]\}, \\ &\vdots \\ H_2\{\hat{p}_{(y_1,y_2,...,y_M)}[V_i]\} &\leq H_2\{\hat{p}_{y_M}[V_i]\}. \end{aligned} \tag{C.29}$$

This completes the proof of this lemma. ■

C.4 Proof of Theorem III.2

Recall that the centralized optimal state $x_i^* \in X$, $i = 1, 2, ..., M$, is the unique minimizer of the penalty function $H_2\{\hat{p}_x[V_i]\}$, $i = 1, 2, ..., M$, $x \in X$ (here x is viewed as the argument of the penalty function). Regarding the decentralized optimal state $(y_1^*, y_2^*, ..., y_M^*) \in X$, consider the following statements:

As assumed in Section 3.2, the pmf's $\hat{p}_{y_i^*}[V_i]$, $i = 1, 2, ..., M$, satisfy the relation

$$H_2\{\hat{p}_{y_i^*}[V_i]\} \leq H_2\{\hat{p}_{y_i}[V_i]\},\ y_i \in Y_i,\ i = 1, 2, ..., M, \tag{C.30}$$

where the equality is attained at $y_i = y_i^*$, $i = 1, 2, ..., M$. Further, as shown in Theorem III.1, the pmf's $\hat{p}_{y_j^*}[V_i]$, $i \neq j$, $i, j = 1, 2, ..., M$, satisfy the relation

$$H_2\{\hat{p}_{y_j^*}[V_i]\} \leq H_2\{\hat{p}_{y_j}[V_i]\},\ y_j \in Y_j,\ i \neq j,\ i, j = 1, 2, ..., M, \tag{C.31}$$

where, as before, the equality is attained at $y_j = y_j^*$, $j \neq i$, $i, j = 1, 2, ..., M$. Clearly, due to Assumption III.2, the above inequalities imply

$$H_2\{\hat{p}_{(y_1^*, y_2^*, ..., y_M^*)}[V_i]\} \leq H_2\{\hat{p}_{(y_1, y_2, ..., y_M)}[V_i]\},\ y_j \in Y_j,\ i, j = 1, 2, ..., M. \qquad \text{(C.32)}$$

Equation (C.32) indicates that the penalty function $H_2\{\hat{p}_x[V_i]\}$, $i = 1, 2, ..., M$, $x \in X$, is minimized at the decentralized optimal state $(y_1^*, y_2^*, ..., y_M^*) \in X$. Furthermore, due to the assumption of uniqueness of the solution of this minimization problem, the decentralized optimal state must be the same as the centralized one, i.e.,

$$(y_1^*, y_2^*, ..., y_M^*) = x_i^*,\ i = 1, 2, ..., M. \qquad \text{(C.33)}$$

This completes the proof of this theorem. ■

APPENDIX D

Proofs of Lemmas and Theorem Stated in Chapter IV

D.1 Proof of Lemma IV.1

The selection of U_{safe} is based on the solution of the following system of inequalities:

$$\begin{aligned} U_{\text{safe}}\alpha_{\min} &\geq V_{\min}, \\ U_{\text{safe}}\alpha_{\max} &\leq V_{\max}. \end{aligned} \tag{D.1}$$

In other words, we require U_{safe} to satisfy the relations $U_{\text{safe}} \geq \frac{V_{\min}}{\alpha_{\min}}$ and $U_{\text{safe}} \leq \frac{V_{\max}}{\alpha_{\max}}$. Clearly, a solution to the above system of inequalities exists due to Assumption (4.5), and, therefore, the U_{safe} is selected as any number in $\left[\frac{V_{\min}}{\alpha_{\min}}, \frac{V_{\max}}{\alpha_{\max}}\right]$. This completes the proof of this lemma. ■

D.2 Proof of Lemma IV.2

From the model (4.17), we know that

$$\tilde{V}(0) = \alpha_{\text{act}} U_{\text{safe}}, \tag{D.2}$$

where, as before, $\alpha_{\rm act}$ is the actual gain of the process variable. Based on the above $\tilde{V}(0)$, the actual status of the process variable at time $n = 0$, denoted as $\sigma_{{\rm act},0}$, can be computed as

$$\sigma_{{\rm act},0} = \begin{cases} {\rm L}_V, & \text{if } \tilde{V}(0) \in [V_{\min}, R_1), \\ {\rm N}_V, & \text{if } \tilde{V}(0) \in [R_1, R_2), \\ {\rm H}_V, & \text{if } \tilde{V}(0) \in [R_2, V_{\max}]. \end{cases} \tag{D.3}$$

Thus, the pmf $p[V(0)]$ is evaluated as

$$\begin{aligned} p[V(0) = \sigma_{{\rm act},0}] &= 1, \\ p[V(0) = \sigma] &= 0,\ \sigma \neq \sigma_{{\rm act},0},\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}. \end{aligned} \tag{D.4}$$

Due to Assumption (4.19), we have $\tilde{S}(0) = k_a \tilde{V}(0) + d_a$. Substitute (D.2) in this expression to get

$$\tilde{S}(0) = k_a U_{\rm safe} \alpha_{\rm act} + d_a. \tag{D.5}$$

Due to Assumption (4.21), the above equation implies

$$\tilde{S}(0) \in I_{\rm act}\left(\tilde{V}(0)\right), \tag{D.6}$$

where $I_{\rm act}$ is, as before, defined in (4.18). Clearly, the pmf $p[S(0)]$ is:

$$\begin{aligned} p[S(0) = \sigma_{{\rm act},0}] &= 1, \\ p[S(0) = \sigma] &= 0,\ \sigma \neq \sigma_{{\rm act},0},\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}. \end{aligned} \tag{D.7}$$

The pmf $\hat{p}[V(0)]$ is evaluated using the steady state of the h-procedure, (2.20), as

$$\hat{p}[V(0) = \sigma] = DQp[S(0) = \sigma] + \frac{1 - DQ}{3},\ \sigma \in \{{\rm L}_V, {\rm N}_V, {\rm H}_V\}. \tag{D.8}$$

Using (D.7), the above expression can be re-written as

$$\begin{aligned} \hat{p}[V(0) = \sigma_{\text{act},0}] &= \tfrac{1+2DQ}{3}, \\ \hat{p}[V(0) = \sigma] &= \tfrac{1-DQ}{3},\ \sigma \neq \sigma_{\text{act},0},\ \sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\}. \end{aligned} \tag{D.9}$$

Clearly, since DQ takes values between 0 and 1, the above pmf takes the maximum probability in the status $\sigma_{\text{act},0}$. This completes the proof of this lemma. ■

D.3 Proof of Theorem IV.1

Select the inputs U_σ, $\sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\}$, as

$$U_\sigma = \frac{V_{\text{des}}}{\alpha_\sigma},\ \sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\}. \tag{D.10}$$

Introduce the following lemma, which is used to prove Theorem IV.1:

Lemma D.1. *The solution of the constrained minimization problem* (4.14) *is given by*

$$\Delta^*(1) = \frac{\sum\limits_{\sigma=\text{L}_V,\text{N}_V,\text{H}_V} W_\sigma D_\sigma [V_{\text{des}} - E_\sigma]}{\sum\limits_{\sigma=\text{L}_V,\text{N}_V,\text{H}_V} W_\sigma D_\sigma^2}, \tag{D.11}$$

where D_σ and E_σ are defined as

$$D_\sigma := \alpha_\sigma [U_{\text{des}} - U_{\text{safe}}],\ E_\sigma := \alpha_\sigma U_{\text{safe}},\ \sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\}. \tag{D.12}$$

Proof. The predicted value of the process variable can be expressed as

$$\hat{V}_{\sigma,U_{\text{res}}}(1; \Delta(1)) = \Delta(1) D_\sigma + E_\sigma,\ \sigma \in \{\text{L}_V, \text{N}_V, \text{H}_V\}, \tag{D.13}$$

where D_σ and E_σ are the same as in (D.12). The constrained minimization problem (4.14) can be re-written as

$$\begin{aligned} \underset{\Delta(1)}{\text{maximize}} \quad & \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} -\frac{1}{2} W_\sigma \left[\hat{V}_{\sigma,U_{\mathrm{res}}}(1;\Delta(1)) - V_{\mathrm{des}}\right]^2, \\ \text{subject to} \quad & V_{\min} \le \hat{V}_{\sigma,U_{\mathrm{res}}}(1;\Delta(1)) \le V_{\max},\ \sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}, \\ & 0 \le \Delta(1) \le 1. \end{aligned} \tag{D.14}$$

Substitute (D.13) in the above constrained maximization problem, and express its Lagrangian as

$$\begin{aligned} L \;=\; & -\tfrac{1}{2} \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} W_\sigma \left[D_\sigma \Delta(1) + E_\sigma - V_{\mathrm{des}}\right]^2 \\ & + \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} \mu_{1,\sigma}[V_{\max} - D_\sigma \Delta(1) - E_\sigma] \\ & + \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} \mu_{2,\sigma}[-V_{\min} + D_\sigma \Delta(1) + E_\sigma] + \mu_3[1 - \Delta(1)] + \mu_4 \Delta(1), \end{aligned} \tag{D.15}$$

where $\mu_{1,\sigma}$, $\sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}$, $\mu_{2,\sigma}$, $\sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}$, μ_3, and μ_4 are the Lagrange multipliers. Apply the Karush-Kuhn-Tucker (KKT) conditions [70] to solve for the candidate optima, $\Delta^*(1)$, $\mu^*_{1,\sigma}$, $\sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}$, $\mu^*_{2,\sigma}$, $\sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}$, μ^*_3, and μ^*_4:

$$\begin{aligned} & - \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} W_\sigma D_\sigma \left[D_\sigma \Delta^*(1) + E_\sigma - V_{\mathrm{des}}\right] - \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} \mu^*_{1,\sigma} D_\sigma \\ & + \sum_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} \mu^*_{2,\sigma} D_\sigma - \mu^*_3 + \mu^*_4 = 0, \end{aligned}$$

$$\begin{aligned} & \mu^*_{1,\sigma}[-V_{\max} + D_\sigma \Delta^*(1) + E_\sigma] = 0,\ \sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}, \\ & \mu^*_{2,\sigma}[V_{\min} - D_\sigma \Delta^*(1) - E_\sigma] = 0,\ \sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}, \\ & \mu^*_3[\Delta^*(1) - 1] = 0,\ \mu^*_4 \Delta^*(1) = 0, \\ & \mu^*_{1,\sigma} \ge 0,\ \mu^*_{2,\sigma} \ge 0,\ \sigma \in \{\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V\}, \\ & \mu^*_3 \ge 0,\ \mu^*_4 \ge 0. \end{aligned} \tag{D.16}$$

The solution of the above system of equations and inequalities is given by

$$\begin{aligned}
\Delta^*(1) &= \frac{\sum\limits_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} W_\sigma D_\sigma [V_{\mathrm{des}} - E_\sigma]}{\sum\limits_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} W_\sigma D_\sigma^2}, \\
\mu_{1,\sigma}^* &= \mu_{2,\sigma}^* = 0,\ \sigma \in \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\}, \\
\mu_3^* &= \mu_4^* = 0.
\end{aligned} \tag{D.17}$$

Clearly, the above $\Delta^*(1)$ satisfies the KKT conditions. Furthermore, it can be shown that the constrained minimization problem (4.14) is convex. Thus, based on these arguments, it can be concluded that the unique solution of (4.14) is, indeed, $\Delta^*(1)$. This completes the proof of this lemma. ■

Proof of Theorem IV.1: Based on Lemma D.1, the input U_{res} at time $n = 1$ can be computed as

$$U_{\mathrm{res}}(1) = V_{\mathrm{des}} \frac{\sum\limits_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} W_\sigma \alpha_\sigma}{\sum\limits_{\sigma=\mathrm{L}_V,\mathrm{N}_V,\mathrm{H}_V} W_\sigma \alpha_\sigma^2}. \tag{D.18}$$

Recall that the value of the process variable and the sensor measurement at time $n = 1$ are

$$\tilde{V}(1) = \alpha_{\mathrm{act}} U_{\mathrm{res}}(1), \tag{D.19}$$

and

$$\tilde{S}(1) = k_a \alpha_{\mathrm{act}} U_{\mathrm{res}}(1) + d_a, \tag{D.20}$$

respectively, where $U_{\rm res}(1)$ is the same as in (D.18). Substitute (D.18) in the right hand side of (D.20) to get

$$\tilde{S}(1) = k_a \alpha_{\rm act} V_{\rm des} \left[\frac{\sum\limits_{\sigma={\rm L}_V,{\rm N}_V,{\rm H}_V} W_\sigma \alpha_\sigma}{\sum\limits_{\sigma={\rm L}_V,{\rm N}_V,{\rm H}_V} W_\sigma \alpha_\sigma^2} \right] + d_a. \tag{D.21}$$

Using the definition of W_σ (see (4.15)), rewrite the above equation as

$$\tilde{S}(1) = k_a \alpha_{\rm act} V_{\rm des} \left[\frac{\sum\limits_{\sigma={\rm L}_V,{\rm N}_V,{\rm H}_V} \hat{p}[V(0)=\sigma] \alpha_\sigma}{\sum\limits_{\sigma={\rm L}_V,{\rm N}_V,{\rm H}_V} \hat{p}[V(0)=\sigma] \alpha_\sigma^2} \right] + d_a. \tag{D.22}$$

Recall that $\hat{p}[V(0)]$, involved in the right hand side of (D.22), is evaluated in (D.9). Thus, re-express (D.22) as

$$\tilde{S}(1) = k_a V_{\rm des} \alpha_{\rm act} \left[\frac{\alpha_{{\rm L}_V} + \alpha_{{\rm N}_V} + \alpha_{{\rm H}_V} + [3\alpha_{\rm act} - \alpha_{{\rm L}_V} - \alpha_{{\rm N}_V} - \alpha_{{\rm H}_V}] DQ(k_a,d_a)}{\alpha_{{\rm L}_V}^2 + \alpha_{{\rm N}_V}^2 + \alpha_{{\rm H}_V}^2 + [3\alpha_{\rm act}^2 - \alpha_{{\rm L}_V}^2 - \alpha_{{\rm N}_V}^2 - \alpha_{{\rm H}_V}^2] DQ(k_a,d_a)} \right] + d_a. \tag{D.23}$$

Clearly, due to Assumption (4.22), we have $\tilde{S}(1) \in [R_1, R_2)$. Therefore, the pmf $p[S(1)]$ is:

$$\begin{aligned} p[S(1) = {\rm N}_V] &= 1, \\ p[S(1) = {\rm L}_V] &= p[S(1) = {\rm H}_V] = 0. \end{aligned} \tag{D.24}$$

Using (D.24), the pmf $\hat{p}[V(1)]$ can be computed as before:

$$\begin{aligned} \hat{p}[V(1) = {\rm N}_V] &= \tfrac{1+2DQ}{3}, \\ \hat{p}[V(1) = {\rm L}_V] &= \hat{p}[V(1) = {\rm H}_V] = \tfrac{1-DQ}{3}. \end{aligned} \tag{D.25}$$

As it may be observed from (D.9) and (D.25), the pmf's $\hat{p}[V(0)]$ and $\hat{p}[V(1)]$ are correctly permuted. This completes the proof of this theorem. ■

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] E. F. Camacho and C. Bordons, *Model predictive control*. Springer, 2007.

[2] A. B. Carlson and P. B. Crilly, *Communication systems*. McGraw-Hill, 2009.

[3] S. Kullback, *Information theory and statistics*. John Wiley and Sons, NY, 1959.

[4] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[5] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.

[6] Y. Peng, S. Zhang, and R. Pan, "Bayesian network reasoning with uncertain evidences," *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, vol. 18, no. 5, pp. 539–564, 2010.

[7] S. M. Meerkov, "Mathematical theory of behavior," *Mathematical Biosciences*, vol. 43, no. 2, pp. 41–106, 1979.

[8] D. D. Siljak and M. B. Vukcevic, "Decentralization, stabilization, and estimation in large-scale systems," *IEEE Trans. on Automatic Control*, vol. AC-21, pp. 363–366, 1976.

[9] M. Ikeda and D. D. Siljak, "Overlapping decompositions, expansions, and contractions of dynamic systems," *Large Scale Systems*, vol. 1, pp. 29–38, 1980.

[10] R. Krtolica and D. D. Siljak, "Suboptimality of decentralized stochastic control and estimation," *IEEE Trans. on Automatic Control*, vol. AC-25, pp. 76–83, 1980.

[11] M. Ikeda, D. D. Siljak, and D. E. White, "Decentralized control with overlapping information sets," *Journal of Optimization Theory and Applications*, vol. 34, no. 2, pp. 279–310, 1981.

[12] D. D. Siljak, *Decentralized control of complex systems*. Dover Publications, NY, 2012.

[13] H. E. Garcia, N. Jhamaria, H. Kuang, W.-C. Lin, and S. M. Meerkov, "Resilient monitoring system: Design and performance analysis," in *Proc. 4th Int. Symp. on Resilient Control Systems*, Boise, ID, USA, Aug. 9-11, 2011, pp. 61–68.

[14] H. E. Garcia, W.-C. Lin, S. M. Meerkov, and M. T. Ravichandran, "Data quality assessment: Modelling and application in resilient monitoring systems," in *Proc. 5th Int. Symp. on Resilient Control Systems*, Salt Lake City, UT, USA, Aug. 14-16, 2012, pp. 124–129.

[15] ——, "Resilient monitoring system for boiler/turbine plant," in *Proc. 6th Int. Symp. on Resilient Control Systems*, San Francisco, CA, USA, Aug. 13-15, 2013, pp. 104–110.

[16] ——, "Resilient plant monitoring system: Design, analysis, and performance evaluation," in *Proc. 52nd IEEE Conf. on Decision and Control*, Florence, Italy, Dec., 2013, pp. 4983–4990.

[17] ——, "Resilient monitoring systems: Architecture, design, and application to boiler/turbine plant," *IEEE Trans. on Cybernetics*, vol. 44, no. 11, pp. 2010–2023, 2014.

[18] H. E. Garcia, S. M. Meerkov, and M. T. Ravichandran, "Combating curse of dimensionality in resilient plant monitoring systems: Overlapping decomposition and knowledge fusion," in *Proc. 52nd Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 1-3, 2014, pp. 637–642.

[19] ——, "Resilient plant monitoring systems: Techniques, analysis, design, and performance evaluation," *Journal of Process Control, accepted for publication*, 2015.

[20] M. Amin, "Toward secure and resilient interdependent infrastructures," *Journal of Infrastructure Systems*, vol. 8, no. 1, pp. 67–75, 2002.

[21] A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Proc. 28th Int. Conf. on Distributed Computing Systems*, Beijing, China, June 17 - 20, 2008, pp. 495–500.

[22] A. M. Madni and S. Jackson, "Towards a conceptual framework for resilience engineering," *IEEE Systems Journal*, vol. 3, no. 2, pp. 181–191, 2009.

[23] C. G. Reiger, D. I. Gertman, and M. A. McQueen, "Resilient control systems: Next generation design research," in *Proc. 2nd Conf. on Human System Interactions*, Catania, Italy, May 21 - 23, 2009, pp. 632–636.

[24] C. G. Reiger, "Notional examples and benchmark aspects of a resilient control system," in *Proc. 3rd Int. Symp. on Resilient Control Systems*, Idaho Falls, ID, USA, Aug. 10 - 12, 2010, pp. 64–71.

[25] C. G. Reiger and K. Villez, "Resilient control system execution agent (ReCoSEA)," in *Proc. 5th Int. Symp. on Resilient Control Systems*, Salt Lake City, UT, USA, Aug. 14 - 16, 2012, pp. 143–148.

[26] S. Amin, A. Cardenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," *Hybrid Systems: Computation and Control*, vol. 5469, pp. 31–45, 2009.

[27] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *Proc. 49th IEEE Conf. on Decision and Control*, Atlanta, GA, USA, Dec., 2010, pp. 5991–5998.

[28] M. Zhu and S. Martinez, "On distributed constrained formation control in operator-vehicle adversarial networks," *Automatica*, vol. 49, no. 12, pp. 3571–3582, 2013.

[29] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[30] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Cyber security of water SCADA systems – Part II: Attack detection using enhanced hydrodynamic models," *IEEE Trans. on Control Systems Technology*, vol. 21, no. 5, pp. 1679–1693, 2013.

[31] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 110–127, 2015.

[32] M. Blanke and J. Schroder, *Diagnosis and fault-tolerant control*. Springer, 2003.

[33] Y. Zhang and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *IFAC Annual Reviews in Control*, vol. 32, no. 2, pp. 229–252, 2008.

[34] H. Noura, D. Theilliol, J. Ponsart, and A. Chamseddine, *Fault-tolerant control systems: Design and practical applications*. Springer, 2009.

[35] S. Song, L. Ling, and C. Manikopoulo, "Flow-based statistical aggregation schemes for network anomaly detection," in *Proc. IEEE Int. Conf. on Networking, Sensing, and Control*, Ft. Lauderdale, FL, USA, Dec., 2006, pp. 786–791.

[36] I. C. Paschalidis and G. Smaragdakis, "Spatio-temporal network anomaly detection by assessing deviations of empirical measures," *IEEE/ACM Trans. on Networking*, vol. 17, no. 3, pp. 685–697, 2009.

[37] A. A. Cardenas, 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 *Proc. 6th ACM Symp. on Information, Computer and Communications Security*, Hong Kong, China, March 2011, pp. 355–366.

[38] S. Sundaram, S. Revzen, and G. Pappas, "A control-theoretic approach to disseminating values and overcoming malicious links in wireless networks," *Automatica*, vol. 48, no. 11, pp. 2894–2901, 2012.

[39] L. K. Carvalho, M. V. Moreira, J. C. Basilio, and S. Lafortune, "Robust diagnosis of discrete-event systems against permanent loss of observations," *Automatica*, vol. 49, no. 1, pp. 223–231, 2013.

[40] J. Wang, D. Rossell, C. Cassandras, and I. Paschalidis, "Network anomaly detection: A survey and comparative analysis of stochastic and deterministic methods," in *Proc. 52nd IEEE Conf. on Decision and Control*, Florence, Italy, Dec. 10-13, 2013, pp. 182–187.

[41] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. on Computer and Communications Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[42] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Limiting false data attacks on power system state estimation," in *Proc. 44th Conf. on Information Sciences and Systems*, Princeton, NJ, USA, Mar. 2010, pp. 1–7.

[43] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. on Smart Grid*, vol. 2, no. 2, pp. 326–333, 2011.

[44] T. Alpcan and T. Basar, *Network Security: A Decision and Game-Theoretic Approach*. Cambridge University Press, 2011.

[45] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 46–65, 2015.

[46] 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, no. 1, pp. 24–45, 2015.

[47] Y. W. Law, T. Alpcan, and M. Palaniswami, "Security games for risk minimization in automatic generation control," *IEEE Trans. on Power Systems*, vol. 30, no. 1, pp. 223–232, 2015.

[48] C. W. Ten, C. C. Liu, and G. Manimaran, "Vulnerability assessment of cybersecurity for SCADA systems," *IEEE Trans. on Power Systems*, vol. 23, no. 4, pp. 1836–1846, 2008.

[49] R. Akella, H. Tang, and B. M. McMillin, "Analysis of information flow security in cyber-physical systems," *Int. Journal of Critical Infrastructure Protection*, vol. 3, no. 2010, pp. 157–173, 2010.

[50] 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," *Int. Journal of Security and Networks*, vol. 6, no. 1, pp. 2–13, 2011.

[51] P. S. Horn, A. J. Pesce, and B. E. Copeland, "A robust approach to reference interval estimation and evaluation," *Clinical Chemistry*, vol. 44, no. 3, pp. 622–631, 1998.

[52] R. M. Gray, *Entropy and information theory*. Springer Verlag, 2013.

[53] P. T. Kabamba, W.-C. Lin, and S. M. Meerkov, "Rational probabilistic deciders - Part I: Individual behavior," *Mathematical Problems in Engineering*, vol. 2007, no. 35897, pp. 1–31, 2007.

[54] ——, "Rational probabilistic deciders - Part II: Collective behavior," *Mathematical Problems in Engineering*, vol. 2007, no. 82184, pp. 1–34, 2007.

[55] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 1, pp. 379–423, 1948.

[56] A. Renyi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, Berkeley, CA, USA, June 1961, pp. 547–561.

[57] C. E. Pfister and W. G. Sullivan, "Renyi entropy, guesswork moments, and large deviations," *IEEE Trans. on Information Theory*, vol. 54, no. 11, pp. 2794–2800, 2004.

[58] S. Aviyente, "Information theoretic measures for quantifying the integration of neural activity," in *Proc. Information Theory and Applications Workshop*, La Jolla, CA, USA, Jan. 2007, pp. 20–26.

[59] X. Yu, X. Pan, W. Yang, and W. Wan, "Audio similarity measure based on Renyi's quadratic entropy," in *Proc. Int. Conf. on Audio Language and Image Processing*, Shanghai, China, Nov. 2010, pp. 722–726.

[60] A. Feltane, G. F. B. Bartels, J. Gaitanis, Y. Boudria, and W. Besio, "Human seizure detection using quadratic Renyi entropy," in *Proc. 6th IEEE EMBS Conf. on Neural Engineering*, San Diego, CA, USA, Nov. 2013, pp. 815–818.

[61] R. R. Yager, "Normalization and representation of non-monotonic knowledge in the theory of evidence," in *Proc. 5th Conf. on Uncertainty in Artificial Intelligence*, Windsor, ON, Canada, Aug. 1989, pp. 394–403.

[62] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Unpublished material, available at: http://www.mpc.berkeley.edu/mpc-course-material, 2014.

[63] "Application basics of operation of three phase induction motors," *Technical report, Rockwell Automation, Aarau, Germany*, pp. 1–59, 1996. [Online]. Available: https://www.rockwellautomation.com/resources/downloads

[64] S. J. Lee, "Multiple simultaneous specifications control design method of a high speed AC induction motor," *MS Thesis, University of Toronto, Canada*, pp. 1–173, 2000.

[65] J. Holtz, "Sensorless control of induction motor drives," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1359–1394, 2002.

[66] R. Langner, "To kill a centrifuge: A technical analysis of what Stuxnet's creators tried to achieve," *Technical report, The Langner Group, Arlington, VA, USA*, pp. 1–36, 2013. [Online]. Available: http://www.langner.com/en/wp-content/uploads

[67] E. Pentzin, "Protecting an industrial AC drive application under cyber sabotage," *M.Sc. Thesis, Aalto University, Finland*, pp. 1–133, 2013.

[68] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava, "Reputation-based framework for high integrity sensor networks," *ACM Trans. on Sensor Networks*, vol. 4, no. 3, pp. 15:1–15:37, 2008.

[69] H. K. Khalil, *Nonlinear Systems*. Prentice Hall, 2002.

[70] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.