Resilient Plant Monitoring Systems: Techniques, Analysis, Design, and Performance Evaluation

H. E. Garcia*, S. M. Meerkov†, and M. T. Ravichandran†

Abstract

Resilient plant monitoring systems are sensor networks that degrade gracefully under malicious attacks on their sensors, causing them to project misleading information. This paper develops techniques to ensure resiliency and illustrates their application using a power plant. Specific techniques developed are: active data quality acquisition, process variable and plant condition assessments, sensor network adaptation, and plant decomposition with knowledge fusion. Based on these techniques, a five-layer resilient monitoring architecture is proposed and analyzed under various cyber-physical attack scenarios. As quantified by Kullback-Leibler divergence, in all scenarios considered, the system offers effective protection against misleading information and identifies the plant conditions - normal or anomalous - in a reliable manner.

I. Introduction

Plant monitoring systems is a relatively new area of control-theoretic research. In this section, we briefly characterize these systems (with emphasis on resiliency), describe a specific scenario addressed, and outline the main techniques developed in this work.

A. What is a resilient plant monitoring system?

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 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

*Idaho National Laboratory, P.O. Box 1625, Idaho Falls, ID 83415-3675, USA. Email: Humberto.Garcia@inl.gov

†Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA. Email: {smm, marutrav}@umich.edu

conditions could be erroneous. This may lead to wrong actions on the part of the operator 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 paper is devoted to developing techniques that can be used to ensure resiliency, analyzing their properties and, on this basis, designing and evaluating 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.

B. Scenario and problem addressed

Briefly, the scenario considered in this paper 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 paper) 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, $\hat{p}[V_i]$, 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 paper 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* [1]).

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

The main 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 [2]). 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* [3] 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* [4].

- 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* [5], 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, 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 above mentioned inferences.

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 the overall architecture and performance evaluation of the resulting monitoring system.

D. Related literature

The literature related to the topic of this paper can be classified into four groups. The first one is devoted to foundational issues, where the problems of resilient monitoring and control are motivated and formulated [6]–[10]. The second group includes publications on control-theoretic methods for attack identification and alleviation, [11]–[13]. 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 which plant status indeed takes place, in spite of the misleading information projected by the sensors.

The third group consists of publications on fault tolerant control, [14]–[16]. 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 communication channels in order to capture anomalous traffic and correlate it with a possible attack, [17]–[19]. In terms of the current work, this implies the identification of DQ. The main tools used here are hypothesis testing and clustering techniques. While the results of [17]–[19] may be useful for resilient plant monitoring, they do not provide methods for process variable and plant condition assessment pursued in the current work.

Our preliminary results on resilient monitoring systems have been reported in conference presentations [20]–[23] and summarized in article [24]. In the current paper, along with reviewing and extending some of these results, we introduce and investigate a decentralized monitoring system based on plant decomposition with knowledge fusion, as a means for combating the curse of dimensionality that mars the performance of the system developed in [20]–[24]. While the decomposition of the plant into sub-plants induces a sensor network decomposition into sub-networks, alleviating thereby the curse of dimensionality, the subsequent knowledge fusion allows for recovering inferences, which are necessary for resiliency. The implementation of this approach necessitates developing a sensor network adaptation technique based on process variable assessments, $\hat{p}[V_i]$, calculating inter- and intra-sub-plant inferences, and designing and investigating the efficacy of a five-layer resilient monitoring system. These developments are described in the current paper, along with an application to monitoring a simplified model of a power plant.

E. Paper outline

The remainder of this paper is structured as follows: Section II addresses the issue of active data quality acquisition. In Section III, the h-procedure and associated techniques for process variable assessment are described. Section IV is devoted to plant pmf assessment. The sensor network adaptation is discussed in Section V, 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 VI. The resulting five-layer monitoring system architecture is presented in Section

VII. An application to a power plant is discussed and investigated by simulations in Section VIII. Finally, the conclusions and directions for future work are given in Section IX. All proofs and the parameters of the power plant model are included in the Appendices.

II. Active Data Quality Acquisition

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

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

Assumption 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

$$\Sigma_V = \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\} \tag{1}$$

with the pmf given by

$$p[V = \mathrm{L}_V] = \int_{V_{\min}}^{R_1} f_{\tilde{V}}(\tilde{v})\, d\tilde{v},\ p[V = \mathrm{N}_V] = \int_{R_1}^{R_2} f_{\tilde{V}}(\tilde{v})\, d\tilde{v},\ p[V = \mathrm{H}_V] = \int_{R_2}^{V_{\max}} f_{\tilde{V}}(\tilde{v})\, d\tilde{v}, \tag{2}$$

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, $\mathbf{U_V}$ (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 $\mu_{\tilde{V}}$):

$$\alpha_{\mathbf{V}} = \begin{cases} \alpha_{\mathbf{V}}^{\mathrm{L}}, & \text{if } \mu_{\tilde{V}} \in [V_{\min}, R_1) \\ \alpha_{\mathbf{V}}^{\mathrm{N}}, & \text{if } \mu_{\tilde{V}} \in [R_1, R_2) \\ \alpha_{\mathbf{V}}^{\mathrm{H}}, & \text{if } \mu_{\tilde{V}} \in [R_2, V_{\max}]. \end{cases} \tag{3}$$

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 (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

$$\Sigma_S = \Sigma_V = \{\mathrm{L}_V, \mathrm{N}_V, \mathrm{H}_V\} \tag{4}$$

with the pmf given by

$$p[S = \mathrm{L}_V] = \int_{V_{\min}}^{R_1} f_{\tilde{S}}(\tilde{s})\, d\tilde{s},\; p[S = \mathrm{N}_V] = \int_{R_1}^{R_2} f_{\tilde{S}}(\tilde{s})\, d\tilde{s},\; p[S = \mathrm{H}_V] = \int_{R_2}^{V_{\max}} f_{\tilde{S}}(\tilde{s})\, d\tilde{s}, \tag{5}$$

where R_1 and R_2 are the same as in (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, $\max\limits_{\sigma \in \Sigma_S} p[S = \sigma]$ may be achieved at L_V, while $\max\limits_{\sigma \in \Sigma_V} p[V = \sigma]$ at 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 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.,

$$u_{\mathbf{V}}(t) = A_{\mathbf{V}} \mathrm{rect}_T(t - t_0). \tag{6}$$

The value of $A_{\mathbf{V}}$ is selected sufficiently small so that $A_{\mathbf{V}} << \min\{[V_{\min}, R_1], [R_1, R_2], [R_2, V_{\max}]\}$. 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:

$$\mu'_{\tilde{S}} - \mu_{\tilde{S}} = A_{\mathbf{V}} \alpha_{\mathbf{V}}(\mu_{\tilde{S}}), \tag{7}$$

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 (3). If the sensor is attacked, (7) does not hold. In order to quantify the severity of the attack, introduce the notion of *probing inconsistency* $(PIC_{\mathbf{S}})$ defined by:

$$PIC_{\mathbf{S}} := \left| (\mu'_{\tilde{S}} - \mu_{\tilde{S}}) - A_{\mathbf{V}} \alpha_{\mathbf{V}} \left(\mu_{\tilde{S}} \right) \right|. \tag{8}$$

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:

$$DQ_{\mathbf{S}} = e^{-F(PIC_{\mathbf{S}})}, \tag{9}$$

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 [21], indicates that a quadratic $F(\,\cdot\,)$ provides better results for subsequent utilization than a linear one. Therefore, we introduce this function as

$$F(PIC_{\mathbf{S}}) := -\frac{\ln \epsilon}{PIC^2_{\max,\mathbf{S}}} PIC^2_{\mathbf{S}}, \tag{10}$$

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 (9) and (10), $\min DQ_{\mathbf{S}} = \epsilon$, which can be viewed as a design parameter.

Expressions (1)-(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.

III. 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 I-C, 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 cannot be applied 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, as well as inferences among the process variables, are introduced.

A. Model of V and S coupling

Introduce the notion of sensor believability:

$$\beta_{\mathbf{S}} = \frac{|\Sigma_V| - 1}{|\Sigma_V|} DQ_{\mathbf{S}} + \frac{1}{|\Sigma_V|}, \tag{11}$$

where $|\Sigma_V|$ is the cardinality of the universal set of V. If, as indicated in (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 2. The coupling between V and S is as follows:

$$\begin{aligned} P[V = \sigma | S &= \sigma] = \beta_{\mathbf{S}}, \\ P[V = \bar{\sigma} | S &= \sigma] = \tfrac{1-\beta_{\mathbf{S}}}{|\Sigma_V|-1}, \end{aligned} \tag{12}$$

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 (12) is used throughout this paper.

B. Process variable pmf assessment using a single sensor

Consider a sensor **S** intended to monitor process variable **V** and assume that Assumption 1 holds. 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

$$\hat{p}[V = \sigma] = \lim_{n \to \infty} P[V = \sigma | s_1, s_2, ..., s_n; DQ_{\mathbf{S}}], \ \forall \sigma \in \Sigma_V. \tag{13}$$

To accomplish this, consider

$$\hat{p}_n[V = \sigma] = P[V = \sigma | s_1, s_2, ..., s_n; DQ_{\mathbf{S}}], \ \forall \sigma \in \Sigma_V, \tag{14}$$

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:

$$h_\sigma(n+1) = h_\sigma(n) + \epsilon_h \left[h_\sigma^*(s_{n+1}) - h_\sigma(n) \right], \ h_\sigma(0) = \frac{1}{|\Sigma_V|}, \ \forall \sigma \in \Sigma_V, \tag{15}$$

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

$$h_\sigma^*(s_{n+1}) = \begin{cases} \beta_{\mathbf{S}}, & \text{if } s_{n+1} = \sigma \\ \frac{1-\beta_{\mathbf{S}}}{|\Sigma_V|-1}, & \text{if } s_{n+1} \neq \sigma, \end{cases} \tag{16}$$

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

$$0 < \epsilon_h << 1, \tag{17}$$

or a function of n monotonically converging to 0 so that

$$0 < \epsilon_h(n) \leq 1, \sum_{n=0}^{\infty} \epsilon_h(n) = \infty, \sum_{n=0}^{\infty} \epsilon_h^2(n) < \infty. \tag{18}$$

As it follows from (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 (15), (16) is referred to as the h-procedure. It can be viewed as a stochastic approximation algorithm [2] with a random set point.

Theorem 1. *Let Assumptions 1 and 2 hold. Then:*

1) *There exists a sufficiently small ϵ_0, such that for all $0 < \epsilon_h < \epsilon_0$, recursive procedure* (15)-(17) *converges in probability as $n \to \infty$ to the following limit:*

$$h_\sigma(n) \xrightarrow{\text{P}} p[S = \sigma]DQ_{\mathbf{S}} + \frac{1 - DQ_{\mathbf{S}}}{|\Sigma_V|}, \ \forall \sigma \in \Sigma_V. \tag{19}$$

2) *Under* (18)*, recursive procedure* (15), (16) *converges to the same limit almost surely.*

Proof: Part 1 is proved in [24]. The proof of Part 2 is given in the Appendix. ■

Thus, according to this 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 (15), (16) is the basis of process variable assessments used throughout this paper.

C. 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.,

$$\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V = \sigma] = \lim_{n\to\infty} P[V = \sigma | s_1^1, ..., s_n^1; DQ_{\mathbf{S}_1}; s_1^2, ..., s_n^2; DQ_{\mathbf{S}_2}], \ \forall \sigma \in \Sigma_V. \tag{20}$$

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 [3]:

$$\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V = \sigma] = \frac{\hat{p}^{\mathbf{S}_1}[V = \sigma]\hat{p}^{\mathbf{S}_2}[V = \sigma]}{\sum_{\sigma} \hat{p}^{\mathbf{S}_1}[V = \sigma]\hat{p}^{\mathbf{S}_2}[V = \sigma]}, \ \forall \sigma \in \Sigma_V. \tag{21}$$

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? To answer this question, let $I\{p[V]\}$ denote the entropy of the pmf $p[V]$ defined as

$$I\{p[V]\} = -\sum_{\sigma\in\Sigma_V} p[V=\sigma]\log_{|\Sigma_V|} p[V=\sigma] \tag{22}$$

and introduce:

Definition 1. A pair of pmf's $\hat{p}^{\mathbf{S}_1}[V]$ and $\hat{p}^{\mathbf{S}_2}[V]$ is *Dempster-Shafer monotonic* (DS-monotonic) if

$$I\{\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V]\} < \min\left[I\{\hat{p}^{\mathbf{S}_1}[V]\}, I\{\hat{p}^{\mathbf{S}_2}[V]\}\right]. \tag{23}$$

■

Thus, only if $\hat{p}^{\mathbf{S}_1}[V]$ and $\hat{p}^{\mathbf{S}_2}[V]$ are DS-monotonic, the combined pmf (21) is beneficial; otherwise, the pmf from one sensor (either $\mathbf{S}_1$ or $\mathbf{S}_2$), having the smallest entropy, should be utilized.

It would be of interest to provide conditions under which $\hat{p}^{\mathbf{S}_1}[V]$ and $\hat{p}^{\mathbf{S}_2}[V]$ are DS-monotonic. At present no general conditions of this type are available. Our numerical study, reported in [20], indicates that if the expected values of $\tilde{S}_1$ and $\tilde{S}_2$ belong to the same outcome of Σ_V and their standard deviations are sufficiently small, then $\hat{p}^{\mathbf{S}_1}[V]$ and $\hat{p}^{\mathbf{S}_2}[V]$ are DS-monotonic; otherwise, $I\{\hat{p}^{\mathbf{S}_1}[V]\}$ or $I\{\hat{p}^{\mathbf{S}_2}[V]\}$ may be smaller than $I\{\hat{p}^{\mathbf{S}_1,\mathbf{S}_2}[V]\}$.

D. Process variable pmf assessment using inferences

Consider a plant characterized by two process variables, $\mathbf{V}_1$ and $\mathbf{V}_2$, monitored by sensors $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively, and operating in accordance with Assumption 1. Denote the universal sets of V_1 as Σ_{V_1} and V_2 as Σ_{V_2}. Assume that these process variables are coupled by conditional pmf's $P[V_1|V_2]$ and $P[V_2|V_1]$. For instance, if $\mathbf{V}_1$ and $\mathbf{V}_2$ are the temperatures of the boiler and turbine, respectively, these conditional pmf's may be of the form

$$P[V_1|V_2] = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \end{bmatrix},\ P[V_2|V_1] = \begin{bmatrix} 1 & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{1}{2} \end{bmatrix}, \tag{24}$$

where the columns represent the states of the condition and the rows that of the random variable itself. Clearly, (24) implies that if $\mathbf{V}_2$ is Normal, $\mathbf{V}_1$ is Normal as well, while if $\mathbf{V}_1$ is Normal, $\mathbf{V}_2$ may be either Normal, or Low, or High. Using these conditional pmf's, the pmf of V_1 (resp., V_2) can be assessed not only by the data and DQ of $\mathbf{S}_1$, (resp., $\mathbf{S}_2$), but also by those of $\mathbf{S}_2$ (resp., $\mathbf{S}_1$). This is important because it offers a possibility of assessing the status of a process variable even if its sensor has $DQ = 0$. To describe this inference procedure, let $\hat{p}^{\mathbf{S}_2}[V_1]$ and $\hat{p}^{\mathbf{S}_2}[V_2]$ denote the pmf's of V_1 and V_2, respectively, evaluated based on the data and DQ of $\mathbf{S}_2$. Obviously, $\hat{p}^{\mathbf{S}_2}[V_2]$ can be evaluated using the h-procedure (15), (16). Then, $\hat{p}^{\mathbf{S}_2}[V_1]$ can be computed using $P[V_1|V_2]$ and the total probability formula:

$$\hat{p}^{\mathbf{S}_2}[V_1] = \sum_{\sigma_2 \in \Sigma_{V_2}} P[V_1|V_2 = \sigma_2]\hat{p}^{\mathbf{S}_2}[V_2 = \sigma_2]. \tag{25}$$

Thus, $\mathbf{V}_1$ is assessed using $\mathbf{S}_2$. Having both $\hat{p}^{\mathbf{S}_2}[V_1]$ and $\hat{p}^{\mathbf{S}_1}[V_1]$, the Dempster-Shafer rule may be used to combine them, if it is beneficial from the point of view of the resulting entropy.

The calculation of $\hat{p}^{\mathbf{S}_1}[V_2]$ is carried out similarly.

IV. Plant pmf Assessment

As mentioned in Subsection I-C, 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 III. Then, $\hat{p}[G]$ can be computed using the following:

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

$$p_0[G] = \left[\frac{1}{3}, \ \frac{1}{3}, \ \frac{1}{3}\right]. \tag{26}$$

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

$$p_0[V_i, G] = P[V_i|G]p_0[G], \ i = 1, 2, ..., M. \tag{27}$$

(c) Calculate the marginal probability:

$$p_0[V_i] = \sum_{G \in \Sigma_G} p_0[V_i, G], \ i = 1, 2, ..., M. \tag{28}$$

(d) Apply the Jeffrey rule [4]:

$$\hat{p}[V_i, G] = p_0[V_i, G]\frac{\hat{p}[V_i]}{p_0[V_i]}, \ i = 1, 2, ..., M. \tag{29}$$

(e) Marginalize to obtain the plant pmf estimate:

$$\hat{p}^{V_i}[G] = \sum_{V_i \in \Sigma_{V_i}} \hat{p}[V_i, G], \ i = 1, 2, .., M. \tag{30}$$

(f) If $M > 1$, combine the pmf's obtained in (30) using the Dempster-Shafer rule:

$$\hat{p}[G = \sigma_{\rm G}] = \frac{\prod_{i=1}^{M} \hat{p}^{V_i}[G = \sigma_{\rm G}]}{\sum_{\sigma_{\rm G}} \prod_{i=1}^{M} \hat{p}^{V_i}[G = \sigma_{\rm G}]}, \ \sigma_{\rm G} \in \Sigma_G. \tag{31}$$

■

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 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 1, defined as follows:

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

$$\hat{p}_n[G], \ \text{where} \ \hat{p}_0[G] = \left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]. \tag{32}$$

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

$$\hat{p}_n[V_i, G] = P[V_i|G]\hat{p}_n[G], \ n = 0, 1, 2, ...; \ i = 1, 2, ..., M. \tag{33}$$

(c) Calculate the marginal probability:

$$\hat{p}_n^G[V_i] = \sum_{G \in \Sigma_G} \hat{p}_n[V_i, G], \ n = 0, 1, 2, ...; \ i = 1, 2, ..., M. \tag{34}$$

(d) Apply the Jeffrey rule:

$$\hat{p}_{n+1}[V_i, G] = \hat{p}_n[V_i, G]\frac{\hat{p}_{n+1}[V_i]}{\hat{p}_n^G[V_i]}, \ n = 0, 1, 2, ...; \ i = 1, 2, ..., M. \tag{35}$$

(e) Marginalize to obtain the plant pmf estimate:

$$\hat{p}_{n+1}^{V_i}[G] = \sum_{V_i \in \Sigma_{V_i}} \hat{p}_{n+1}[V_i, G], \ n = 0, 1, 2, ...; \ i = 1, 2, ..., M. \tag{36}$$

(f) If $M > 1$, combine the pmf's obtained in (36) using the Dempster-Shafer rule:

$$\hat{p}_{n+1}[G=\sigma_{\rm G}] = \frac{\prod_{i=1}^{M} \hat{p}_{n+1}^{V_i}[G=\sigma_{\rm G}]}{\sum_{\sigma_{\rm G}} \prod_{i=1}^{M} \hat{p}_{n+1}^{V_i}[G=\sigma_{\rm G}]},\ n = 0,1,2,...;\ \sigma_{\rm G} \in \Sigma_G. \tag{37}$$

(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:

$$\Sigma_G = \{{\rm N}_G, {\rm A}_G\},\ \Sigma_V = \Sigma_S = \{{\rm N}_V, {\rm A}_V\}. \tag{38}$$

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

$$P[V|G] = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}, \tag{39}$$

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

$$\hat{p}_n[V] = [h_{{\rm N}_V}(n), h_{{\rm A}_V}(n)],\ \hat{p}_n[G] = [k_{{\rm N}_G}(n), k_{{\rm A}_G}(n)], \tag{40}$$

where $h_{{\rm N}_V}(n)$ and $h_{{\rm A}_V}(n)$ are calculated using the h-procedure (15), (16) and $k_{{\rm N}_G}(n)$ and $k_{{\rm A}_G}(n)$ are evaluated using Algorithm 2. To specify the evolution of $k_{{\rm N}_G}(n)$ and $k_{{\rm A}_G}(n)$, substitute (39) and (40) in steps (a)-(e) of this algorithm to obtain

$$k_{{\rm N}_G}(n+1) = \left[\frac{1-a}{C(n)}\right] k_{{\rm N}_G}(n) + \left[\frac{a k_{{\rm N}_G}(n)}{D(n)} - \frac{[1-a]k_{{\rm N}_G}(n)}{C(n)}\right] h_{{\rm N}_V}(n+1), \tag{41}$$

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

$$C(n) := [1-a]k_{{\rm N}_G}(n) + a[1-k_{{\rm N}_G}(n)],\ D(n) := a k_{{\rm N}_G}(n) + [1-a][1-k_{{\rm N}_G}(n)]. \tag{42}$$

Denote the steady state values of $h_{{\rm N}_V}(n)$ and $h_{{\rm A}_V}(n)$, evolving according to the h-procedure (15),(16), as $h^{\rm ss}_{{\rm N}_V}$ and $h^{\rm ss}_{{\rm A}_V}$, respectively. Then, the steady state values of $k_{{\rm N}_G}(n)$ and $k_{{\rm A}_G}(n)$ are quantified as follows:

Theorem 2. *The steady state* $k^{\rm ss}_{{\rm N}_G}$ *of the recursion* (41) *is:*

1) $k^{\rm ss}_{{\rm N}_G} = 1$, *if* $h^{\rm ss}_{{\rm N}_V} > 1-a$;

2) $k^{\rm ss}_{{\rm N}_G} = 0$, *if* $h^{\rm ss}_{{\rm N}_V} < a$;

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

Proof: See the Appendix. ■

This theorem exhibits the paradoxical nature of the recursive Jeffrey rule. Namely, if, for instance, $h^{\text{ss}}_{\text{N}_V} = 0.7$, i.e., $\hat{p}[V] = [0.7, 0.3]$, and $a = 0.4$, then, according to Part 1 of Theorem 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 $h^{\text{ss}}_{\text{N}_V} = 0.3$, 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.

V. Sensor Network Adaptation and Measure of Resiliency

As mentioned in Subsection I-C, 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 VII the latter.

A. 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 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 element, 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 V-D.) The process variable pmf's and the plant pmf assessed in state x of the sensor network are denoted as $\hat{p}_x[V_i]$, $x \in X$, $i = 1, ..., M$, and $\hat{p}_x[G]$, respectively. The goal of the sensor network adaptation is to converge to the state, where the entropy of $\hat{p}_x[G]$ is minimal.

B. Adaptation using a rational controller

As mentioned in Subsection I-C, the adaptation technique used in this work is based on rational controllers introduced in [5] and further developed in [25], [26]. 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:

$$T_x = \begin{cases} T_{\text{max}}, & \text{if } \hat{I}_x(G) \leq \beta \\ \left(\frac{\beta}{\hat{I}_x(G)}\right)^N T_{\text{max}}, & \text{if } \hat{I}_x(G) > \beta, \end{cases} \tag{43}$$

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.,

$$\tau_x = \frac{T_x}{\sum\limits_{x \in X} T_x}. \tag{44}$$

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

$$\bar{p}[G] = \sum_{x \in X} \tau_x \hat{p}_x[G]. \tag{45}$$

It can be shown that if N is sufficiently large, $\bar{p}[G]$ is arbitrarily close to $\arg\min\limits_{x \in X} \hat{I}_x(G)$. 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 (43)-(45) since it is equally applicable to random transitions, which may be necessary in other applications.

C. *Measure of resiliency*

The measure of resiliency employed in this work is based on the Kullback-Leibler divergence [1] of two pmf's, $p_1[G]$ and $p_2[G]$, given by:

$$D\left(p_1[G] || p_2[G]\right) = \sum_{\sigma_G \in \Sigma_G} p_1\left[G = \sigma_G\right] \log_{|\Sigma_G|} \frac{p_1[G = \sigma_G]}{p_2[G = \sigma_G]}. \tag{46}$$

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 (45) 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

$$MR = \frac{D\left(p[G]||p_{\rm nr}[G]\right) - D\left(p[G]||\bar{p}[G]\right)}{D\left(p[G]||p_{\rm nr}[G]\right)}. \tag{47}$$

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 (47). This is carried out in Section VIII for the case of the power plant.

D. 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 II, 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 III and IV, 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, $T_{\rm report} = (2^5)12$sec $\approx$ 6min, whereas in a network of 10 sensors, $T_{\rm report} \approx$ 3hr, which is clearly unacceptable. This curse of dimensionality is the main drawback of the centralized system based on $\hat{p}_x[G]$ adaptation.

VI. Combatting the Curse of Dimensionality: Decentralized system with knowledge fusion

This section provides a method for combatting the curse of dimensionality based on the plant decomposition with knowledge fusion. The development is carried out in terms of a power plant; however, the approach is applicable to other systems as well.

A. Power plant

A simplified model of a power plant is shown in Figure 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 1.

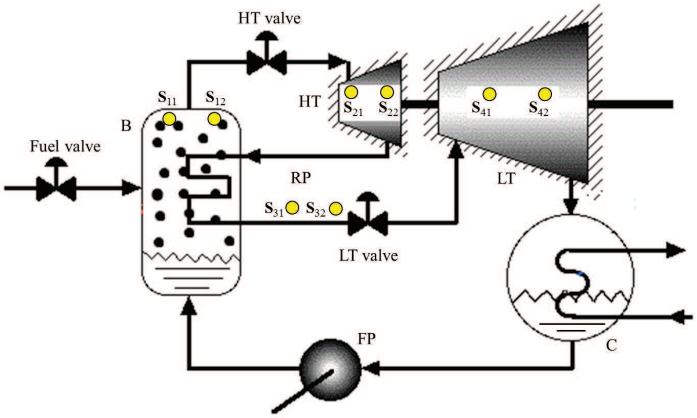
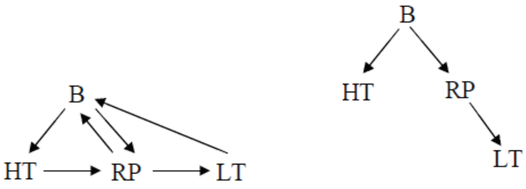


Fig. 1: Schematics of the power plant

Having 8 sensors, the number of network states is 256. Thus, based on the temporal properties discussed in Subsection V-D, a report to the plant operator could be produced in about every 45min. To combat this drawback, a decentralized system could be considered, 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 I-C, on a decentralized system with knowledge fusion and show that under certain conditions such a decomposition, while decreasing the state space of resilient adaptation, leads to no loss in quality of plant condition assessment.

B. 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(a). Assuming, for simplicity, 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 LT condition, the influence LT $\rightarrow$ B can also be ignored. Under these assumptions, the cyclic graph of Figure 2(a) is reduced to the *tree graph* of Figure 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(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 45min, 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 $\bar{p}[G]$.



(a) Cyclic graph representation

(b) Tree graph representation

Fig. 2: Influence diagrams

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 $\hat{p}[G_i]$, $i \in \{\mathrm{B}, \mathrm{HT}, \mathrm{RP}, \mathrm{LT}\}$, become 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 VIII-A, below we describe the knowledge fusion calculations used in this work.

C. 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 II, III, and V. Then, fusion of this information, leading to the sought inferences, is carried out as follows:

1) Inferences for $\mathbf{V}_1$*:*

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

$$\bar{p}_{\mathbf{G}_{\text{LT}}}[V_1] \quad = \sum_{\sigma_3 \in \Sigma_{V_3}} P[V_1|V_3=\sigma_3]\bar{p}_{\mathbf{G}_{\text{LT}}}[V_3=\sigma_3], \tag{48}$$

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

$$\bar{p}_{\mathbf{G}_{\text{LT}}}[V_3] = \sum_{\sigma_4 \in \Sigma_{V_4}} P[V_3|V_4=\sigma_4]\bar{p}_{\mathbf{G}_{\text{LT}}}[V_4=\sigma_4]. \tag{49}$$

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

$$\bar{p}_{\mathbf{G}_{\text{RP}}}[V_1] = \sum_{\sigma_3 \in \Sigma_{V_3}} P[V_1|V_3=\sigma_3]\bar{p}_{\mathbf{G}_{\text{RP}}}[V_3=\sigma_3]. \tag{50}$$

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

$$\bar{p}_{\mathbf{G}_{\text{HT}}}[V_1] = \sum_{\sigma_2 \in \Sigma_{V_2}} P[V_1|V_2=\sigma_2]\bar{p}_{\mathbf{G}_{\text{HT}}}[V_2=\sigma_2]. \tag{51}$$

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

$$\bar{p}_{\mathbf{G}_{\text{B,HT,RP,LT}}}[V_1=\sigma_1] = \frac{\prod\limits_{k=\text{B,HT,RP,LT}} \bar{p}_{\mathbf{G}_k}[V_1=\sigma_1]}{\sum\limits_{\sigma_1 \in \Sigma_{V_1}} \prod\limits_{k=\text{B,HT,RP,LT}} \bar{p}_{\mathbf{G}_k}[V_1=\sigma_1]}. \tag{52}$$

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

$$\bar{p}^*[V_1] = \arg\min\left\{I\left\{\bar{p}_{\mathbf{G}_{\text{B}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\text{HT}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\text{RP}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\text{LT}}}[V_1]\right\}, I\left\{\bar{p}_{\mathbf{G}_{\text{B,HT,RP,LT}}}[V_1]\right\}\right\}. \tag{53}$$

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

D. Accuracy of decentralized systems with knowledge fusion

In this subsection, we address the following question: How much information is lost due to the decentralization with knowledge fusion? Although a complete answer to this question is not available yet, a partial one is provided below by considering a system motivated by the power plant application.

Let the plant $\mathbf{G}$ consist of two process variables $\mathbf{V}_1$ and $\mathbf{V}_2$, each monitored by two sensors, $\{\mathbf{S}_{11}, \mathbf{S}_{12}\}$ and $\{\mathbf{S}_{21}, \mathbf{S}_{22}\}$, respectively. Assume that the universal sets of V_1 and V_2 are

$$\Sigma_{V_1} := \{\mathrm{N}_{\mathrm{V}_1}, \mathrm{A}_{\mathrm{V}_1}\}\,,\ \Sigma_{V_2} := \left\{\mathrm{N}_{\mathrm{V}_2}, \mathrm{A}^{(1)}_{\mathrm{V}_2}, \mathrm{A}^{(2)}_{\mathrm{V}_2}, \mathrm{A}^{(3)}_{\mathrm{V}_2}\right\}, \tag{54}$$

where, as before, N stands for Normal and A for Anomalous. Let $Z_{\mathrm{N}_{\mathrm{V}_1}}$ and $Z_{\mathrm{A}_{\mathrm{V}_1}}$ be the intervals where $\mathbf{V}_1$ is viewed as $\mathrm{N}_{\mathrm{V}_1}$ and $\mathrm{A}_{\mathrm{V}_1}$, respectively, and let $Z_{\mathrm{N}_{\mathrm{V}_2}}$, $Z_{\mathrm{A}^{(1)}_{\mathrm{V}_2}}$, $Z_{\mathrm{A}^{(2)}_{\mathrm{V}_2}}$, and $Z_{\mathrm{A}^{(3)}_{\mathrm{V}_2}}$ be the corresponding intervals for $\mathbf{V}_2$. Similar to (3), assume that the d.c. gains of the process variables are piecewise constant in these intervals. Finally, let the coupling between $\mathbf{V}_1$ and $\mathbf{V}_2$ be characterized by the conditional pmf's

$$P[V_1|V_2] = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},\ P[V_2|V_1] = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}. \tag{55}$$

To combat the curse of dimensionality, introduce two sub-plants, $\mathbf{G}_{\mathrm{I}}$ and $\mathbf{G}_{\mathrm{II}}$, consisting of process variables $\mathbf{V}_1$ and $\mathbf{V}_2$ and their sensors, respectively. Two methods for evaluating the pmf's of V_1 and V_2 are considered below, the centralized and the decentralized with knowledge fusion, and their results are compared.

Centralized assessment (CA): Each sensor is assigned DQ based on the procedure described in Section II. The state space of the overall sensor network is given by

$$X = \{(0000), (1000), (0100), ..., (1111)\}\,, \tag{56}$$

which contains 16 states. Let the sensor network be equipped with a rational controller, whose objective is to minimize the penalty function $\Phi(x)$, specified as

$$\Phi(x) = I\left\{\hat{p}_x[V_1, V_2]\right\},\ x \in X, \tag{57}$$

where the joint pmf $\hat{p}_x[V_1, V_2]$ is computed as

$$\hat{p}_x[V_1, V_2] = P[V_1|V_2]\hat{p}_x[V_2], \tag{58}$$

and $\hat{p}_x[V_2]$ is calculated as described in Section III. Assume that a unique solution of this minimization problem, $\min\limits_{x\in X}\Phi(x)$, exists and is given by

$$x^* = \arg\min_{x\in X}\Phi(x). \tag{59}$$

To solve the above problem, specify the residence time of the controller in state $x \in X$ as

$$T_x = \left(\frac{1}{\Phi(x)}\right)^N, \tag{60}$$

where N is sufficiently large. In this scenario, utilizing the formula similar to (45), the pmf $\bar{p}[V_1, V_2]$ is calculated as $\hat{p}_{x^*}[V_1, V_2]$. Finally, $\bar{p}[V_1, V_2]$ is marginalized to obtain $\bar{p}[V_1]$ and $\bar{p}[V_2]$.

Decentralized assessment with knowledge fusion (DA-KF): Assign DQ to each sensor as before. Decompose the sensor network with state space (56) into two sub-networks, with state spaces $X_{\rm I}$ and $X_{\rm II}$ defined as follows:

$$X_{\rm I} = \{(00)_{\rm I}, (10)_{\rm I}, (01)_{\rm I}, (11)_{\rm I}\}, \; X_{\rm II} = \{(00)_{\rm II}, (10)_{\rm II}, (01)_{\rm II}, (11)_{\rm II}\}. \tag{61}$$

Assume that each sub-network is equipped with a rational controller, whose objective is to minimize the penalty functions

$$\Phi_{\rm I}(x_{\rm I}) = I\left\{\hat{p}_{x_{\rm I}}[V_1]\right\}, \; x_{\rm I} \in X_{\rm I}, \; \Phi_{\rm II}(x_{\rm II}) = I\left\{\hat{p}_{x_{\rm II}}[V_2]\right\}, \; x_{\rm II} \in X_{\rm II}, \tag{62}$$

respectively, where $\hat{p}_{x_{\rm I}}[V_1]$ and $\hat{p}_{x_{\rm II}}[V_2]$ are calculated as in Section III. Assume that unique solutions of these minimization problems exist and are given by

$$x_{\rm I}^* = \arg\min_{x_{\rm I}\in X_{\rm I}}\Phi_{\rm I}(x_{\rm I}), \; x_{\rm II}^* = \arg\min_{x_{\rm II}\in X_{\rm II}}\Phi_{\rm II}(x_{\rm II}). \tag{63}$$

Let the residence time of the rational controllers be specified as in (60). Under this scenario, the pmf's $\bar{p}_{\mathbf{G}_{\rm I}}[V_1]$ and $\bar{p}_{\mathbf{G}_{\rm II}}[V_2]$ are calculated as $\hat{p}_{x_{\rm I}^*}[V_1]$ and $\hat{p}_{x_{\rm II}^*}[V_2]$, respectively. Then, based on the knowledge fusion calculations (48)-(51), the inferences $\bar{p}_{\mathbf{G}_{\rm I}}[V_2]$ and $\bar{p}_{\mathbf{G}_{\rm II}}[V_1]$ are obtained.

To characterize the conditions under which CA and DA-KF result in the same pmf's of process variables, introduce

Definition 2. The pmf's $\bar{p}_{\mathbf{G}_{\rm I}}[V_i]$ and $\bar{p}_{\mathbf{G}_{\rm II}}[V_i]$, $i = 1, 2$, are *max-similar* if

$$\arg\max_{\sigma_i\in\Sigma_{V_i}}\bar{p}_{\mathbf{G}_{\rm I}}[V_i = \sigma_i] = \arg\max_{\sigma_i\in\Sigma_{V_i}}\bar{p}_{\mathbf{G}_{\rm II}}[V_i = \sigma_i], \; i = 1, 2. \tag{64}$$

■

In other words, $\bar{p}_{\mathbf{G}_{\rm I}}[V_1]$ and $\bar{p}_{\mathbf{G}_{\rm II}}[V_1]$ (resp. $\bar{p}_{\mathbf{G}_{\rm I}}[V_2]$ and $\bar{p}_{\mathbf{G}_{\rm II}}[V_2]$) are max-similar if their maxima are attained at the same element of Σ_{V_1} (resp. Σ_{V_2}).

Theorem 3. *Assume that both pairs of pmf's* $\{\bar{p}_{\mathbf{G}_{\text{I}}}[V_1], \bar{p}_{\mathbf{G}_{\text{II}}}[V_1]\}$ *and* $\{\bar{p}_{\mathbf{G}_{\text{I}}}[V_2], \bar{p}_{\mathbf{G}_{\text{II}}}[V_2]\}$ *are DS- monotonic and max-similar. Then,* $(x^*_{\text{I}}, x^*_{\text{II}}) = x^*$*, and the pmf's of* V_1 *and* V_2 *calculated using CA and DA-KF are, respectively, identical.*

Proof: See the Appendix. ■

We hypothesize that this sufficient condition for the efficacy of decentralized systems with knowledge fusion is applicable to more general scenarios than that considered here. A justification of this hypothesis and derivation of more general conditions are topics for future work.

VII. Resilient Monitoring System Architecture

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 3. It consists of four parallel sub-architectures, each corresponding to a sub-plant, $\mathbf{G}_{\text{B}}$, $\mathbf{G}_{\text{HT}}$, $\mathbf{G}_{\text{RP}}$, and $\mathbf{G}_{\text{LT}}$, which could be under a physical attack (or malfunction). The inputs to each sub-architecture are the sensor data provided by the sub-networks $\mathbf{SN}_{\text{B}}$, $\mathbf{SN}_{\text{HT}}$, $\mathbf{SN}_{\text{RP}}$, and $\mathbf{SN}_{\text{LT}}$, 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., $\bar{p}[\text{B}]$, $\bar{p}[\text{HT}]$, $\bar{p}[\text{RP}]$, $\bar{p}[\text{LT}]$.

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 II.
- The process variable assessment layer consists of two parts. The first one represents the evaluation of $\hat{p}_{x_{\text{B}}}[V_1]$ using the methods of Section III. The second part evaluates $\bar{p}_{\mathbf{G}_{\text{B}}}[V_1]$ using the expression (45) applied to the sub-plant (i.e., $\bar{p}_{\mathbf{G}_{\text{B}}}[V_1] = \sum\limits_{x_{\text{B}} \in X_{\text{B}}} \tau_{x_{\text{B}}} \hat{p}_{x_{\text{B}}}[V_1]$, where $\tau_{x_{\text{B}}}$ is the output of the adaptation layer).
- The sub-network adaptation layer operates as described in Section V, but using the entropy of $\hat{p}_{x_{\text{B}}}[V_1]$ as the penalty function.
- The knowledge fusion layer implements the calculations described in Subsection VI-C.
- The sub-plant assessment layer evaluates $\bar{p}[\text{B}]$, $\bar{p}[\text{HT}]$, $\bar{p}[\text{RP}]$, and $\bar{p}[\text{LT}]$ using the technique of Section IV.

The measure of resiliency is evaluated using (47) applied separately to each sub-plant, e.g.,

$$MR_{\text{B}} = \frac{D\left(p[B]||p_{\text{nr}}[B]\right) - D\left(p[B]||\bar{p}[B]\right)}{D\left(p[B]||p_{\text{nr}}[B]\right)}. \tag{65}$$

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

$$\overrightarrow{MR} = \left[MR_{\text{B}},\ MR_{\text{HT}},\ MR_{\text{RP}},\ MR_{\text{LT}}\right]. \tag{66}$$

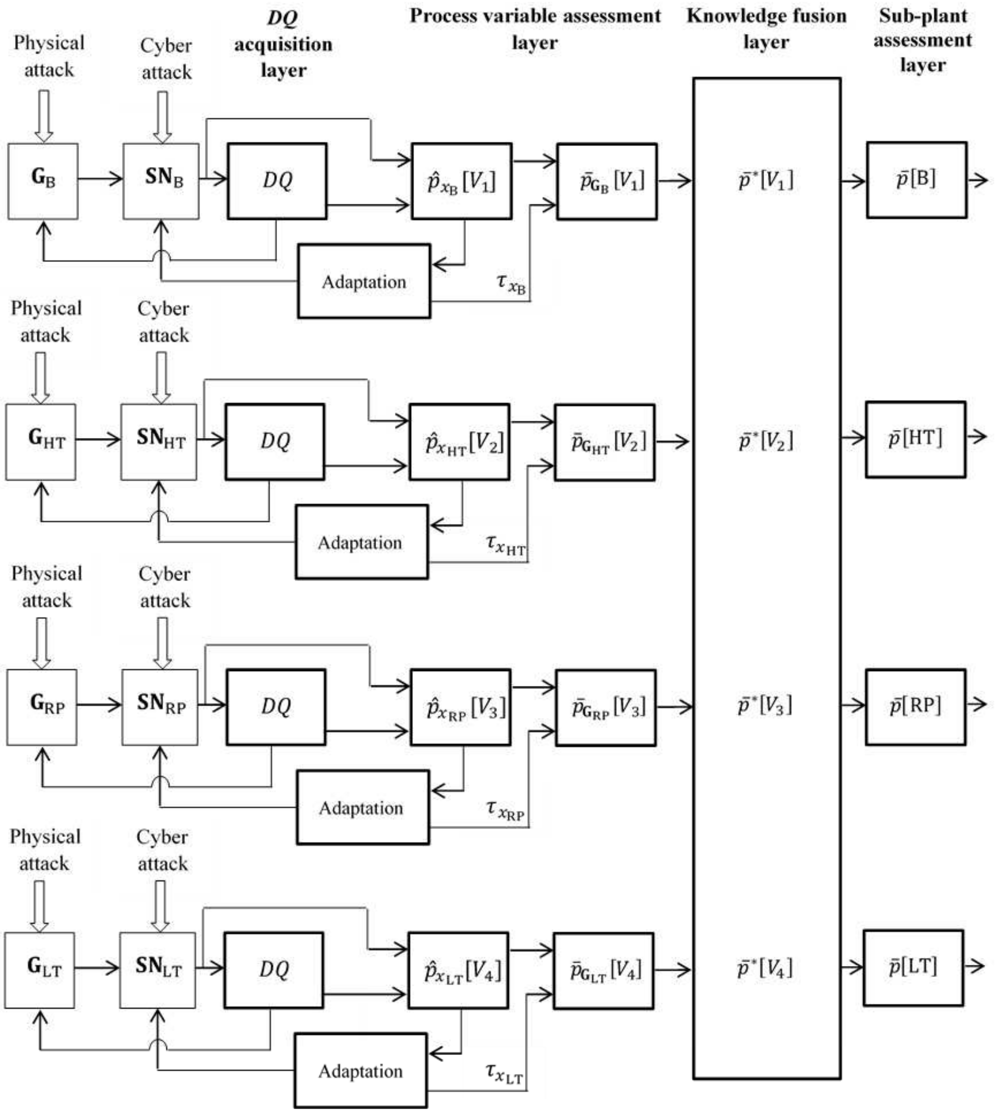

Physical attack
Cyber attack
DQ acquisition layer
Process variable assessment layer
Knowledge fusion layer
Sub-plant assessment layer
$\mathbf{G}_{\mathrm{B}}$
$\mathbf{SN}_{\mathrm{B}}$
DQ
$\hat{p}_{x_{\mathrm{B}}}[V_1]$
$\bar{p}_{\mathbf{G}_{\mathrm{B}}}[V_1]$
$\bar{p}^*[V_1]$
$\bar{p}[\mathrm{B}]$
Adaptation
$\tau_{x_{\mathrm{B}}}$
$\mathbf{G}_{\mathrm{HT}}$
$\mathbf{SN}_{\mathrm{HT}}$
$\hat{p}_{x_{\mathrm{HT}}}[V_2]$
$\bar{p}_{\mathbf{G}_{\mathrm{HT}}}[V_2]$
$\bar{p}^*[V_2]$
$\bar{p}[\mathrm{HT}]$
$\tau_{x_{\mathrm{HT}}}$
$\mathbf{G}_{\mathrm{RP}}$
$\mathbf{SN}_{\mathrm{RP}}$
$\hat{p}_{x_{\mathrm{RP}}}[V_3]$
$\bar{p}_{\mathbf{G}_{\mathrm{RP}}}[V_3]$
$\bar{p}^*[V_3]$
$\bar{p}[\mathrm{RP}]$
$\tau_{x_{\mathrm{RP}}}$
$\mathbf{G}_{\mathrm{LT}}$
$\mathbf{SN}_{\mathrm{LT}}$
$\hat{p}_{x_{\mathrm{LT}}}[V_4]$
$\bar{p}_{\mathbf{G}_{\mathrm{LT}}}[V_4]$
$\bar{p}^*[V_4]$
$\bar{p}[\mathrm{LT}]$
$\tau_{x_{\mathrm{LT}}}$

Fig. 3: Five-layer resilient monitoring system architecture based on decentralization with knowledge fusion

VIII. Application to power plant

In this section, we apply the resilient monitoring system of Figure 3 to the power plant of Figure 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 VIII-A), describe the attack scenarios and the resulting system performance (Subsection VIII-B), and discuss qualitative features of the results obtained (Subsection VIII-C).

A. Sub-plant anomalies and process variable couplings

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 $\Sigma_{V_1} := \{\mathrm{L}_{\mathrm{V}_1}, \mathrm{N}_{\mathrm{V}_1}\}$.

2) High pressure turbine: The anomaly of HT is also the insulation fracture. Taking into account the influence B → 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 $\Sigma_{V_2} := \{\mathrm{VL}_{\mathrm{V}_2}, \mathrm{L}_{(1)\mathrm{V}_2}, \mathrm{L}_{(2)\mathrm{V}_2}, \mathrm{N}_{\mathrm{V}_2}\}$, where VL stands for Very Low, and $\mathrm{L}_{(1)\mathrm{V}_2}$ and $\mathrm{L}_{(2)\mathrm{V}_2}$ indicate Low HT temperature due to B and HT damage, respectively.

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, $V_3 \in \Sigma_{V_3} := \{\mathrm{VL}_{\mathrm{V}_3}, \mathrm{L}_{(1)\mathrm{V}_3}, \mathrm{L}_{(2)\mathrm{V}_3}, \mathrm{N}_{\mathrm{V}_3}\}$.

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 → RP → 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, $V_4 \in \Sigma_{V_4} := \{\mathrm{VL}_{(1)\mathrm{V}_4}, \mathrm{VL}_{(2)\mathrm{V}_4}, \mathrm{L}_{(1)\mathrm{V}_4}, \mathrm{L}_{(2)\mathrm{V}_4}, \mathrm{M}_{(1)\mathrm{V}_4}, \mathrm{M}_{(2)\mathrm{V}_4}, \mathrm{N}_{\mathrm{V}_4}, \mathrm{H}_{\mathrm{V}_4}\}$, where M stands for Medium and H for High.

5) Coupling of process variables: As described in Subsection VI-B, 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:

$$P[V_1|V_2] = P[V_1|V_3] = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}}_{A},\ P[V_2|V_1] = P[V_3|V_1] = \underbrace{\begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \end{bmatrix}}_{B}, \tag{67}$$
$$P[V_3|V_4] = \begin{bmatrix} A & \mathbf{0}_{2\times4} \\ \mathbf{0}_{2\times4} & A \end{bmatrix},\ P[V_4|V_3] = \begin{bmatrix} B & \mathbf{0}_{4\times2} \\ \mathbf{0}_{4\times2} & B \end{bmatrix}.$$

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.

B. 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:

$$\bar{p}[G_{\mathrm{B}}] = [0.95, 0.05],\ \bar{p}[G_{\mathrm{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\mathrm{RP}}] = [0.91, 0.09],\ \bar{p}[G_{\mathrm{LT}}] = [0.92, 0.08], \tag{68}$$

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 V-C) evaluates the pmf of B as $p_{\mathrm{nr}}[G_{\mathrm{B}}] = [0.05, 0.95]$, erroneously indicating that the boiler is damaged. Using (65) and (66), 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:

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\ \bar{p}[G_{\text{LT}}] = [0.49, 0.51], \tag{69}$$

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:

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.12, 0.88],\ \bar{p}[G_{\text{LT}}] = [0.92, 0.08], \tag{70}$$

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:

$$\bar{p}[G_{\text{B}}] = [0.95, 0.05],\ \bar{p}[G_{\text{HT}}] = [0.51, 0.49],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\ \bar{p}[G_{\text{LT}}] = [0.92, 0.08], \tag{71}$$

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 $p_{\text{nr}}[G_{\text{HT}}] = [0.9, 0.1]$, 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

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

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

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

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 L_{V_1} and V_4 being $\text{L}_{(2)\text{V}_4}$. All other sub-plants operate normally, with V_2 being $\text{L}_{(1)\text{V}_2}$ and V_3 being $\text{L}_{(1)\text{V}_3}$. 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:

$$\bar{p}[G_{\text{B}}] = [0.05, 0.95],\ \bar{p}[G_{\text{HT}}] = [0.9, 0.1],\ \bar{p}[G_{\text{RP}}] = [0.91, 0.09],\ \bar{p}[G_{\text{LT}}] = [0.51, 0.49], \tag{72}$$

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

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

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

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

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:

$$\bar{p}[G_{\text{B}}] = [0.05, 0.95], \bar{p}[G_{\text{HT}}] = [0.9, 0.1], \bar{p}[G_{\text{RP}}] = [0.51, 0.49], \bar{p}[G_{\text{LT}}] = [0.5, 0.5], \tag{73}$$

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

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

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

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

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

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

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.

C. 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/or LT are more dangerous than those on B and RP*. This is due to the structure of the conditional probability matrices (67), 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(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" [27], 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).

IX. Conclusion and Future Research

This work provides techniques to ensure resiliency and demonstrates that they are adequate for designing resilient plant monitoring systems. The development is carried out under the assumption that each process variable may be either normal or anomalous, and a cyber-physical attacker shifts sensor measurements so as to project misleading information. In this scenario, we develop a decentralized five-layer monitoring system architecture with knowledge fusion, which, on one hand, alleviates the curse of dimensionality and, on the other hand, allows for calculating inferences necessary for resiliency. Although the development is carried out in terms of a power plant, a similar approach can be used for other critical infrastructure plants, as long as they admit a representation as a set of interrelated sub-plants.

Numerous research problems, however, remain open. These include:

- *Problems related to overall architecture:*
  - The decentralized system with knowledge fusion is based on the reduction of a cyclic influence graph to a tree-graph (see Figure 2). Extending this decomposition to cyclic graphs is important. The approach may be quite similar to that of the present work, and, in this framework, the effect of "circular" coupling among process variables could be investigated.
  - Another important architectural issue is: What are other than decentralization techniques that can effectively combat the curse of dimensionality in resilient monitoring systems? Perhaps, the

overlapping decomposition of [28], [29] could be a productive alternative.

- *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 Subsection V-D, 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.
- *Problems related to process variable assessment:*
  - ○ Introducing and investigating different than (12) models of coupling between the sensor data and process variables. Similarly, investigating different (as compared with the believability (11)) effects of DQ on process variable assessment.
  - ○ Introducing and utilizing other than conditional probability-based coupling (see (24)) among the process variables. This may be based on logical models "if-then", rather than on quantitative ones.
  - ○ In the current work, the sensor data and DQ's are utilized to assess the process variable pmf's (i.e., h-procedure (15), (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 resilient monitoring systems could be improved substantially.
  - ○ Investigating monotonicity properties of Dempster-Shafer rule (21). A sufficient condition for monotonicity is mentioned in Subsection III-C. More general (e.g., necessary and sufficient) conditions would be beneficial for improving the speed of process variable pmf's assessment.
- *Problems related to sensor network adaptation:*
  - ○ Utilizing other than (43) rational controllers. The goal here is to devise rational controllers with faster adaptation rates (see [5] 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 knowledge fusion:*
  - ○ Evaluation of the efficacy of knowledge fusion. This would involve the derivation of more general conditions to quantify, for example, the loss of information due to knowledge fusion calculations.
  - ○ At present, just a rudimentary technique has been used at this layer: a "combination" of process variable pmf's obtained in different sub-architectures (see Figure 3). It would be of interest to investigate fusing information measures other than the pmf's.
- *Problems related to plant assessment:*
  - ○ Investigating a possibility of recursive plant assessment. Because recursive application of the Jeffrey rule may lead to paradoxical result (see Section IV), in the current paper 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.

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

ACKNOWLEDGEMENTS

The University of Michigan students Naman Jhamaria and Heng Kuang are acknowledged for their participation in the initial stages of this research. Support for this research has been provided by the U.S. Department of Energy under DOE Contract DE-AC07-05ID14517 and performed as part of the Instrumentation, Control, and Intelligent Systems (ICIS) initiative at the Idaho National Laboratory.

APPENDIX A
PROOFS

A. Theorem 1

As mentioned in Section III, Part 1 of this theorem is proved in [24]. Below, we prove Part 2. It is based on the following lemmas:

Lemma A.1. *Consider the recursive procedure* (15), (16), (18). *Then,*

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

Proof: As it follows from (15),

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

Thus, $h_\sigma(n) \geq 0$, $\forall n$ and $\forall \sigma$. Also, it can be shown that, due to (18),

$$\sum_{i=0}^{n} w_i(n) = 1,\ \lim_{n\to\infty} w_0(n) = 0. \tag{A.3}$$

Therefore,

$$\begin{aligned} \lim_{n\to\infty} h_\sigma(n) &= \lim_{n\to\infty} w_0(n)h_\sigma(0) + \lim_{n\to\infty}\sum_{i=1}^{n} w_i(n)h_\sigma^*(s_i),\ \sigma \in \Sigma_V, \\ &= \lim_{n\to\infty}\sum_{i=1}^{n} w_i(n)h_\sigma^*(s_i) \leq \lim_{n\to\infty}\sum_{i=1}^{n} w_i(n), \end{aligned} \tag{A.4}$$

where the last inequality is due to (16). Finally, in view of (A.3), 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 1, the expected value of the set point,* $h_\sigma^*(s_n)$, $\sigma \in \Sigma_V$, $n \in \mathbb{N}$, *is given by*

$$E[h_\sigma^*(s_n)] = p[S=\sigma]DQ_{\mathbf{S}} + \frac{1-DQ_{\mathbf{S}}}{|\Sigma_V|},\ \sigma \in \Sigma_V,\ n \in \mathbb{N}. \tag{A.5}$$

Proof: Follows directly from (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

$$f\left(h_\sigma(n)\right) := \frac{1}{2}\left[h_\sigma^*(s_{n+1}) - h_\sigma(n)\right]^2,\ \sigma \in \Sigma_V. \tag{A.6}$$

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

$$\arg\min_{h_\sigma(n)} E\left[f\left(h_\sigma(n)\right)\right] = \mu_{h_\sigma^*},\ \sigma \in \Sigma_V. \tag{A.7}$$

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

$$\frac{\partial}{\partial h_\sigma(n)} E\left[f\left(h_\sigma(n)\right)\right] = 0,\ \sigma \in \Sigma_V. \tag{A.8}$$

Due to (A.6), 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 $h_\sigma^{\min}(n) = \mu_{h_\sigma^*}$, $\sigma \in \Sigma_V$. ■

Proof of Theorem 1, Part 2: The proof is based on showing that for large n, the recursive procedure (15), (16), (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 $0 \le \gamma \le 1$ such that

$$\begin{aligned} f\left(h_\sigma(n+1)\right) = f\left(h_\sigma(n)\right) \quad &+ \quad \left[h_\sigma(n+1) - h_\sigma(n)\right] \left.\tfrac{\partial f}{\partial h_\sigma(n)}\right|_{h_\sigma(n)=h_\sigma(n)} \\ &+ \quad \tfrac{\left[h_\sigma(n+1)-h_\sigma(n)\right]^2}{2} \left.\tfrac{\partial^2 f}{\partial h_\sigma^2(n)}\right|_{h_\sigma(n)=h_\sigma(n)+\gamma\left[h_\sigma(n+1)-h_\sigma(n)\right]},\ \sigma \in \Sigma_V. \end{aligned} \tag{A.9}$$

From (A.6) and (15), (16), we obtain

$$f\left(h_\sigma(n+1)\right) = f\left(h_\sigma(n)\right) - \epsilon_h(n)\left[\frac{\partial f}{\partial h_\sigma(n)}\right]^2 + \frac{\epsilon_h^2(n)}{2}\left[h_\sigma^*(s_{n+1}) - h_\sigma(n)\right]^2,\ \sigma \in \Sigma_V. \tag{A.10}$$

Using the summation of both sides of (A.10), we obtain:

$$f\left(h_\sigma(n)\right) = f\left(h_\sigma(0)\right) - \sum_{n=0}^{n-1}\epsilon_h(n)\left[\frac{\partial f}{\partial h_\sigma(n)}\right]^2 + \sum_{n=0}^{n-1}\frac{\epsilon_h^2(n)}{2}\left[h_\sigma^*(s_{n+1}) - h_\sigma(n)\right]^2,\ \sigma \in \Sigma_V. \tag{A.11}$$

Now, consider the limit of (A.11) 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 $\left[h_\sigma^*(s_{n+1}) - h_\sigma(n)\right]^2$ is bounded for all n, implying that there exists a positive M, such that $\left[h_\sigma^*(s_{n+1}) - h_\sigma(n)\right]^2 \le M$, $\forall n$. Thus,

$$\lim_{n\to\infty} f\left(h_\sigma(n)\right) \le f\left(h_\sigma(0)\right) - \lim_{n\to\infty}\sum_{n=0}^{n-1}\epsilon_h(n)\left[\frac{\partial f}{\partial h_\sigma(n)}\right]^2 + \frac{M}{2}\lim_{n\to\infty}\sum_{n=0}^{n-1}\epsilon_h^2(n),\ \sigma \in \Sigma_V. \tag{A.12}$$

Observe that since $\sum_{n=0}^{\infty}\epsilon_h^2(n) < \infty$, the last term in the right hand side of (A.12) 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.12) 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.8) 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. ■

B. *Proof of Theorem 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 (41) can be rewritten as

$$k_{\mathrm{N}_G}(n+1) = F\left(k_{\mathrm{N}_G}(n)\right) + \mathcal{O}(\epsilon), \tag{A.13}$$

where

$$F\left(k_{\mathrm{N}_G}(n)\right) := \left[\frac{a h^{\mathrm{ss}}_{\mathrm{N}_V}}{a k_{\mathrm{N}_G}(n) + [1-a][1-k_{\mathrm{N}_G}(n)]} + \frac{[1-a][1-h^{\mathrm{ss}}_{\mathrm{N}_V}]}{[1-a]k_{\mathrm{N}_G}(n) + a[1-k_{\mathrm{N}_G}(n)]}\right] k_{\mathrm{N}_G}(n), \tag{A.14}$$

and $\mathcal{O}(\epsilon)$ represents terms of order ϵ. Omitting these terms, equation (A.13) is approximated as

$$k_{\mathrm{N}_G}(n+1) = F\left(k_{\mathrm{N}_G}(n)\right). \tag{A.15}$$

It can be shown that the system (A.15) has three equilibria,

$$k^{*}_{\mathrm{N}_G} = 1,\ k^{**}_{\mathrm{N}_G} = 0,\ k^{***}_{\mathrm{N}_G} = \frac{h^{\mathrm{ss}}_{\mathrm{N}_V} - a}{1-2a}. \tag{A.16}$$

Based on the perturbation theory [30], for ϵ sufficiently small, stability properties of (A.15) are the same as (A.13). To analyze stability, consider the Jacobians of $F(\,\cdot\,)$ at each equilibrium:

$$\begin{aligned} A_1 &= \left.\frac{\partial F}{\partial k_{\mathrm{N}_G}}\right|_{k^{*}_{\mathrm{N}_G}} = \frac{[1-a]^2+[2a-1]h^{\mathrm{ss}}_{\mathrm{N}_V}}{a[1-a]},\ A_2 = \left.\frac{\partial F}{\partial k_{\mathrm{N}_G}}\right|_{k^{**}_{\mathrm{N}_G}} = \frac{a^2+[1-2a]h^{\mathrm{ss}}_{\mathrm{N}_V}}{a[1-a]}, \\ A_3 &= \left.\frac{\partial F}{\partial k_{\mathrm{N}_G}}\right|_{k^{***}_{\mathrm{N}_G}} = \frac{a[1-a]}{h^{\mathrm{ss}}_{\mathrm{N}_V}[1-h^{\mathrm{ss}}_{\mathrm{N}_V}]}. \end{aligned} \tag{A.17}$$

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. ■

C. *Theorem 3*

The proof of this theorem is based on the following five lemmas.

Lemma A.4. *Let $\hat{p}_{x_{\mathrm{I}}}[V_1]$, $x_{\mathrm{I}} \in X_{\mathrm{I}}$, be the pmf of V_1, calculated as mentioned in Subsection VI-D, and $\hat{p}_{x_{\mathrm{I}}}[V_2]$ be the pmf of V_2, calculated using total probability formula:*

$$\hat{p}_{x_{\mathrm{I}}}[V_2] = \sum_{\sigma_1 \in \Sigma_1} P[V_2|V_1=\sigma_1]\hat{p}_{x_{\mathrm{I}}}[V_1=\sigma_1],\ x_{\mathrm{I}} \in X_{\mathrm{I}}, \tag{A.18}$$

where Σ_1 and $P[V_2|V_1]$ are specified by (54) *and* (55)*, respectively. Then,*

$$I\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} = \frac{I\left\{\hat{p}_{x_{\mathrm{I}}}[V_1]\right\}+1}{2},\ x_{\mathrm{I}} \in X_{\mathrm{I}}, \tag{A.19}$$

where $I\{\,\cdot\,\}$ is the entropy.

Proof: Due to the structure of $P[V_2|V_1]$, equation (A.18) can be expressed as $\hat{p}_{x_\text{I}}[V_2 = \text{N}_2] = \hat{p}_{x_\text{I}}[V_2 = \text{A}_{21}] = \frac{1}{2}\hat{p}_{x_\text{I}}[V_1 = \text{N}_1]$ and $\hat{p}_{x_\text{I}}[V_2 = \text{A}_{22}] = \hat{p}_{x_\text{I}}[V_2 = \text{A}_{23}] = \frac{1}{2}\hat{p}_{x_\text{I}}[V_1 = \text{A}_1]$. Consequently, the entropy of $\hat{p}_{x_\text{I}}[V_1]$ can be evaluated as

$$I\left\{\hat{p}_{x_\text{I}}[V_1]\right\} = -\sum_{\sigma_2 \in \Sigma_2} \hat{p}_{x_\text{I}}[V_2 = \sigma_2] \log_2 \hat{p}_{x_\text{I}}[V_2 = \sigma_2] - 1, \tag{A.20}$$

where Σ_2 is defined in (54). Then, applying the change of base formula, $\log_a x = \frac{\log_b x}{\log_b a}$, the right hand side (RHS) of (A.20) becomes $I\left\{\hat{p}_{x_\text{I}}[V_2]\right\} \log_2 4 - 1 = 2I\left\{\hat{p}_{x_\text{I}}[V_2]\right\} - 1$. ■

Lemma A.5. *Let $\hat{p}_x[V_1, V_2]$, $x \in X$, be the joint pmf of V_1 and V_2, calculated as mentioned in Subsection VI-D, and $\hat{p}_x[V_2]$ be the pmf of V_2, obtained by marginalizing $\hat{p}_x[V_1, V_2]$. Then,*

$$I\left\{\hat{p}_x[V_2]\right\} = \frac{3}{2} I\left\{\hat{p}_x[V_1, V_2]\right\}, \; x \in X. \tag{A.21}$$

Proof: The proof of this lemma is similar to that of Lemma A.4. ■

Since $\hat{p}_{x^*}[V_1, V_2]$ and $\hat{p}_{x_\text{I}^*}[V_1]$ have the smallest entropies in the state space of their respective networks (see (59) and (63)), Lemmas A.4 and A.5 indicate that the centralized and the decentralized rational controllers adapt in such a manner that the entropy of the pmf of V_2 is minimized.

Lemma A.6. *Assume that $\hat{p}_{x_\text{I}}[V_2]$ and $\hat{p}_{x_\text{II}}[V_2]$, $x_\text{I} \in X_\text{I}$, $x_\text{II} \in X_\text{II}$, are DS-monotonic and $\hat{p}_{(x_\text{I}, x_\text{II})}[V_2]$ is their concatenation using Dempster-Shafer rule. Then,*

$$\frac{dI\left\{\hat{p}_{(x_\text{I}, x_\text{II})}[V_2]\right\}}{dI\left\{\hat{p}_{x_\text{I}}[V_2]\right\}} > 0 \text{ and } \frac{dI\left\{\hat{p}_{(x_\text{I}, x_\text{II})}[V_2]\right\}}{dI\left\{\hat{p}_{x_\text{II}}[V_2]\right\}} > 0. \tag{A.22}$$

Proof: Introduce notations: $\hat{p}_{x_\text{I}}[V_2] = [p_1, p_2, p_3, p_4]$ and $\hat{p}_{x_\text{II}}[V_2] = [q_1, q_2, q_3, q_4]$. Without loss of generality, assume that $\max\{p_1, p_2, p_3, p_4\} = p_1$ and $\max\{q_1, q_2, q_3, q_4\} = q_1$. The entropy of $\hat{p}_{x_\text{I}}[V_2]$ is

$$I\left\{\hat{p}_{x_\text{I}}[V_2]\right\} = -\sum_{i=1}^{4} p_i \log_4 p_i. \tag{A.23}$$

The differential of $I\left\{\hat{p}_{x_\text{I}}[V_2]\right\}$ is

$$dI\left\{\hat{p}_{x_\text{I}}[V_2]\right\} = \sum_{i=1}^{4} \frac{\partial}{\partial p_i} I\left\{\hat{p}_{x_\text{I}}[V_2]\right\} dp_i, \tag{A.24}$$

where, due to the constraint, $p_1 + p_2 + p_3 + p_4 = 1$,

$$\sum_{i=1}^{4} dp_i = 0. \tag{A.25}$$

Using (A.25), equation (A.24) can be re-written as follows:

$$dI\left\{\hat{p}_{x_\text{I}}[V_2]\right\} = \sum_{i=1}^{3} \left[\frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_4}\right] I\left\{\hat{p}_{x_\text{I}}[V_2]\right\} dp_i. \tag{A.26}$$

Since, as it follows from (A.23), the partial derivative, $\frac{\partial}{\partial p_i} I\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} = -1 - \log_4 p_i$, $i = 1, 2, 3, 4$. From (A.26), we obtain the following differential for the denominator of the first expression of (A.22):

$$dI\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} = \sum_{i=1}^{3} \left[\log_4 \frac{p_4}{p_i}\right] dp_i. \tag{A.27}$$

In a similar manner, it can be shown that the numerator of this expression is given by:

$$\begin{aligned} &dI\left\{\hat{p}_{(x_{\mathrm{I}},x_{\mathrm{II}})}[V_2]\right\} \\ &= \sum_{i=1}^{3} \left[\left(\frac{Dq_4 - p_4 q_4^2}{D^2}\right)\left(1 + \log_4 \frac{p_4 q_4}{D}\right) - \left(\frac{Dq_i - p_i q_i^2}{D^2}\right)\left(1 + \log_4 \frac{p_i q_i}{D}\right)\right] dp_i, \end{aligned} \tag{A.28}$$

where

$$D = \sum_{i=1}^{4} p_i q_i. \tag{A.29}$$

Suppose, $dp_1 > 0$ and $dp_2 = dp_3 = 0$. Then, the RHS of (A.27) becomes $\left[\log_4 \frac{p_4}{p_1}\right] dp_1$, implying that $dI\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} < 0$. Using (A.28), it can also be shown that $dI\left\{\hat{p}_{(x_{\mathrm{I}},x_{\mathrm{II}})}[V_2]\right\} < 0$. Moreover, in all other situations where $dI\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\}$ is less than zero (for instance, when $dp_1 > 0$, $dp_2 > 0$, $dp_3 = 0$, $p_2 > p_4$, and $q_2 > q_4$), it can be shown that $dI\left\{\hat{p}_{(x_{\mathrm{I}},x_{\mathrm{II}})}[V_2]\right\}$ is less than zero as well. These arguments imply that $\frac{dI\left\{\hat{p}_{(x_{\mathrm{I}},x_{\mathrm{II}})}[V_2]\right\}}{dI\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\}} > 0$.

The second expression in (A.22) can be proved in a similar manner. ■

Lemma A.7. *Let x_k be a state in X_k, $k = \mathrm{I}, \mathrm{II}$, with all active sensors being captured. Let the pmf $\hat{p}_{x_k}[V_2]$, $k = \mathrm{I}, \mathrm{II}$, be computed as mentioned before. Finally, let $\epsilon << 1$ be the parameter involved in the data quality exponent* (9), (10). *Then, $I\left\{\hat{p}_{x_k}[V_2]\right\} = 1 - \delta(\epsilon)$, $\delta(\epsilon) \to 0$ as $\epsilon \to 0$, $k = \mathrm{I}, \mathrm{II}$.*

Proof: We prove this lemma for the state $x_{\mathrm{I}} \in X_{\mathrm{I}}$. The proof for $x_{\mathrm{II}} \in X_{\mathrm{II}}$ is similar.

Suppose, $x_{\mathrm{I}} = (10)_{\mathrm{I}}$. Let the pmf of the corresponding active sensor, namely, $\mathbf{S}_{11}$, be $p[S_{11}] = \left[p_{\mathrm{L_B}}^{S_{11}}, p_{\mathrm{N_B}}^{S_{11}}\right]$. Since this sensor is captured, its data quality is assigned as $DQ_{\mathbf{S}_{11}} = \epsilon$ based on the d.c. gain model (3) and the procedure described in Section II. In this situation, the pmf $\hat{p}_{x_{\mathrm{I}}}[V_1]$, calculated using the h-procedure, is as follows (see (15),(16)):

$$\hat{p}_{x_{\mathrm{I}}}[V_1] = \left[p_{\mathrm{L_B}}^{S_{11}} \epsilon + \frac{1-\epsilon}{2}, p_{\mathrm{N_B}}^{S_{11}} \epsilon + \frac{1-\epsilon}{2}\right]. \tag{A.30}$$

The entropy of $\hat{p}_{x_{\mathrm{I}}}[V_1]$ is

$$I\left\{\hat{p}_{x_{\mathrm{I}}}[V_1]\right\} = -a_1(\epsilon)\log_2 a_1(\epsilon) - a_2(\epsilon)\log_2 a_2(\epsilon), \tag{A.31}$$

where

$$\begin{aligned} a_1(\epsilon) &:= p_{\mathrm{L_B}}^{S_{11}}\epsilon + \tfrac{1-\epsilon}{2}, \\ a_2(\epsilon) &:= p_{\mathrm{N_B}}^{S_{11}}\epsilon + \tfrac{1-\epsilon}{2}. \end{aligned} \tag{A.32}$$

Taking into account (A.31) and using Lemma A.4, we have

$$I\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} = \frac{1}{2}\left[1 - a_1(\epsilon)\log_2 a_1(\epsilon) - a_2(\epsilon)\log_2 a_2(\epsilon)\right]. \tag{A.33}$$

As it follows from (A.32), the expression $-a_1(\epsilon)\log_2 a_1(\epsilon) - a_2(\epsilon)\log_2 a_2(\epsilon) \to 1$ as $\epsilon \to 0$.

Differentiating both sides of (A.33) with respect to ϵ, we obtain

$$\frac{d}{d\epsilon} I\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} = \left(p_{\mathrm{L_B}}^{S_{11}} - \frac{1}{2}\right)\log_2 \frac{a_2(\epsilon)}{a_1(\epsilon)}. \tag{A.34}$$

It follows from (A.32) and (A.34) that $\frac{d}{d\epsilon} I\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} \leq 0$, and the equality is attained when $p_{\mathrm{L_B}}^{S_{11}} = \frac{1}{2}$. In other words, the entropy of $\hat{p}_{x_{\mathrm{I}}}[V_2]$ is a decreasing function of ϵ.

Equation (A.33) can be re-expressed as

$$I\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\} = 1 + \frac{-1 - a_1(\epsilon)\log_2 a_1(\epsilon) - a_2(\epsilon)\log_2 a_2(\epsilon)}{2}. \tag{A.35}$$

Denote $\frac{1 + a_1(\epsilon)\log_2 a_1(\epsilon) + a_2(\epsilon)\log_2 a_2(\epsilon)}{2}$ as $\delta(\epsilon)$. It is seen that $\delta(\epsilon) \to 0$ as $\epsilon \to 0$.

Similarly, if x_{I} is either $(01)_{\mathrm{I}}$ or $(11)_{\mathrm{I}}$, it can be shown that $I\left\{\hat{p}_{x_{\mathrm{I}}}[V_2]\right\}$ is equal to $1 - \delta(\epsilon)$, where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. ■

Lemma A.8. *Let the pmf's $\hat{p}_{x_{\mathrm{I}}^*}[V_2]$ and $\hat{p}_{x_{\mathrm{II}}^*}[V_2]$ be DS-monotonic and max-similar. Further, let x_{I} and x_{II} be states in X_{I} and X_{II}, respectively, with their corresponding active sensors being captured. Then, the following is satisfied:*

$$I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} < \min\left[I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}})}[V_2]\right\}, I\left\{\hat{p}_{(x_{\mathrm{I}}, x_{\mathrm{II}}^*)}[V_2]\right\}, I\left\{\hat{p}_{(x_{\mathrm{I}}, x_{\mathrm{II}})}[V_2]\right\}\right]. \tag{A.36}$$

Proof: If the active sensors corresponding to x_{II} are captured, it follows from Lemma A.7 that $\hat{p}_{x_{\mathrm{II}}}[V_2] = \left[\frac{1}{4} + \delta_1(\epsilon), \frac{1}{4} + \delta_2(\epsilon), \frac{1}{4} + \delta_3(\epsilon), \frac{1}{4} + \delta_4(\epsilon)\right]$, where ϵ is the parameter involved in the DQ exponent (9), (10), and $\delta_i(\epsilon) \to 0$ as $\epsilon \to 0$, $i = 1, 2, 3, 4$. Further, let $\hat{p}_{x_{\mathrm{I}}^*}[V_2]$ be expressed as $[p_1, p_2, p_3, p_4]$, where $\sum_{i=1}^{4} p_i = 1$. Consequently, the entropy of the pmf $\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}})}[V_2]$, which is obtained by concatenating $\hat{p}_{x_{\mathrm{I}}^*}[V_2]$ and $\hat{p}_{x_{\mathrm{II}}}[V_2]$ using, as before, Dempster-Shafer rule, can be shown to have the following property:

$$I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}})}[V_2]\right\} = \left[1 - \delta_a(\epsilon)\right] I\left\{\hat{p}_{x_{\mathrm{I}}^*}[V_2]\right\} + \delta_b(\epsilon), \tag{A.37}$$

where

$$\begin{aligned}
\delta_a(\epsilon) &:= \tfrac{D(\epsilon)-\frac{1}{4}}{D(\epsilon)}, \\
\delta_b(\epsilon) &:= -\sum_{i=1}^{4} \frac{\delta_i(\epsilon)p_i}{D(\epsilon)} \log_4 p_i - \sum_{i=1}^{4} \frac{p_i\left[\frac{1}{4}+\delta_i(\epsilon)\right]}{D(\epsilon)} \log_4 \frac{\frac{1}{4}+\delta_i(\epsilon)}{D(\epsilon)}, \\
D(\epsilon) &:= \sum_{i=1}^{4} p_i \left[\frac{1}{4}+\delta_i(\epsilon)\right].
\end{aligned} \tag{A.38}$$

From (A.38), observe that $\delta_a(\epsilon) \to 0$ and $\delta_b(\epsilon) \to 0$ as $\epsilon \to 0$. Using (A.37), the difference between the entropies, $I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}})}[V_2]\right\}$ and $I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\}$, can be expressed as

$$I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}})}[V_2]\right\} - I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} = [1-\delta_a(\epsilon)]\, I\left\{\hat{p}_{x_{\mathrm{I}}^*}[V_2]\right\} - I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} + \delta_b(\epsilon). \tag{A.39}$$

Since the pmf's $\hat{p}_{x_{\mathrm{I}}^*}[V_2]$ and $\hat{p}_{x_{\mathrm{II}}^*}[V_2]$ are DS-monotonic, and ϵ is chosen arbitrarily small, the RHS of (A.39) is greater than zero. This implies that $I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} < I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}})}[V_2]\right\}$.

Using the above approach and Lemma A.7, it can also be shown that $I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} < I\left\{\hat{p}_{(x_1, x_{\mathrm{II}}^*)}[V_2]\right\}$ and $I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} < I\left\{\hat{p}_{(x_1, x_2)}[V_2]\right\}$. ■

Proof of Theorem 3: In this proof, we address the case of V_2; the proof for V_1 is similar.

As mentioned in Subsection VI-D, $\arg\min_{x_{\mathrm{I}} \in X_{\mathrm{I}}} I\{\hat{p}_{x_{\mathrm{I}}}[V_1]\} = x_{\mathrm{I}}^*$, $\arg\min_{x_{\mathrm{II}} \in X_{\mathrm{II}}} I\{\hat{p}_{x_{\mathrm{II}}}[V_2]\} = x_{\mathrm{II}}^*$, and $\arg\min_{x \in X} I\{\hat{p}_x[V_1, V_2]\} = x^*$. The objective here is to show that x^* is the concatenation of x_{I}^* and x_{II}^*, i.e., $(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)$.

Using Lemma A.4 and the assumption that the pmf's $\hat{p}_{(x_{\mathrm{I}}^*, (00)_{\mathrm{II}})}[V_2]$ and $\hat{p}_{((00)_{\mathrm{I}}, x_{\mathrm{II}}^*)}[V_2]$ are DS-monotonic,

$$\begin{aligned}
I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} &< I\left\{\hat{p}_{(x_{\mathrm{I}}, (00)_{\mathrm{II}})}[V_2]\right\}, \ \forall x_{\mathrm{I}} \in X_{\mathrm{I}}, \\
I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} &< I\left\{\hat{p}_{((00)_{\mathrm{I}}, x_{\mathrm{II}})}[V_2]\right\}, \ \forall x_{\mathrm{II}} \in X_{\mathrm{II}}.
\end{aligned} \tag{A.40}$$

Clearly, from Lemma A.5, the minimizer x^* belongs to the set $\{\{X_{\mathrm{I}} \setminus (00)_{\mathrm{I}}\} \times \{X_{\mathrm{II}} \setminus (00)_{\mathrm{II}}\}\}$ (where $\setminus$ is the relative complement and $\times$ is the Cartesian product).

Now, we investigate whether x^* is, indeed, $(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)$ in the following two cases:

- For arbitrary x_{I} and x_{II}, let the pmf's $\hat{p}_{(x_{\mathrm{I}}, (00)_{\mathrm{II}})}[V_2]$ and $\hat{p}_{((00)_{\mathrm{I}}, x_{\mathrm{II}})}[V_2]$ be DS-monotonic. Using Lemma A.6, it can be shown that $I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} < I\left\{\hat{p}_{(x_{\mathrm{I}}, x_{\mathrm{II}})}[V_2]\right\}$.
- Let $\hat{p}_{(x_{\mathrm{I}}, (00)_{\mathrm{II}})}[V_2]$ and $\hat{p}_{((00)_{\mathrm{I}}, x_{\mathrm{II}})}[V_2]$ be non-DS-monotonic due to all active sensors of either or both x_{I} and x_{II} being captured. Then, it can be shown using Lemma A.8 that $I\left\{\hat{p}_{(x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)}[V_2]\right\} < I\left\{\hat{p}_{(x_{\mathrm{I}}, x_{\mathrm{II}})}[V_2]\right\}$.

Thus, these two arguments prove that $x^* = (x_{\mathrm{I}}^*, x_{\mathrm{II}}^*)$. ■

Appendix B
Parameters of power plant and monitoring system

This appendix provides parameters of the power plant and monitoring system that have been used in simulations reported in Subsection VIII-B.

A. Sub-plants, process variables, and sensors

1) Statistical models of the sub-plants: As mentioned in Subsection VIII-A, 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 elements of this pmf are 0.05.
- High pressure turbine: $P[V_2 \in \{\mathrm{L}_{(1)\mathrm{V}_2}, \mathrm{N}_{\mathrm{V}_2}\} | G_{\mathrm{HT}} = \mathrm{N}_{\mathrm{G}_{\mathrm{HT}}}] = P[V_2 \in \{\mathrm{VL}_{\mathrm{V}_2}, \mathrm{L}_{(2)\mathrm{V}_2}\} | G_{\mathrm{HT}} = \mathrm{A}_{\mathrm{G}_{\mathrm{HT}}}] = 0.90$; all other elements 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 $P[V_4 \in \{\mathrm{VL}_{(1)\mathrm{V}_4}, \mathrm{L}_{(1)\mathrm{V}_4}, \mathrm{M}_{(1)\mathrm{V}_4}, \mathrm{N}_{\mathrm{V}_4}\} | G_{\mathrm{LT}} = \mathrm{A}_{\mathrm{G}_{\mathrm{LT}}}] = 0.08$.

2) Models of process variables and sensors: The values of the parameters introduced here are not intended to represent exact physical quantities but, rather, to illustrate the techniques developed in this work.

The domains of the process variables and their d.c. gains (defined in Assumption 1) are specified in Table I.

Without loss of generality, we assume that the process variables and the sensor measurements 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 II and III, respectively, for all attack scenarios considered in Section VIII. 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 I may be ignored. Specifically, they are selected as $\sigma_{\tilde{V}_i} = \sigma_{\tilde{S}_{ij}} = 0.01$, $i = 1, 2, 3, 4$, $j = 1, 2$.

B. Parameters of monitoring system

TABLE I: Domains and d.c. gains of process variables

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

TABLE II: 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 III: 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

1) Data quality assessment layer:

- The amplitudes of the probing signals (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 (10), is selected as 0.02.
- The $PIC_{\max}$ in (10) for the sensors of B, HT, RP, and LT are 0.4, 0.06, 0.08, 0.03, respectively.

2) Process variables assessment layer:

- The step size of the h-procedure (15) is selected as $\epsilon_h = 0.01$.
- The stopping rule is defined by $|h_\sigma(n+1) - h_\sigma(n)| < 10^{-4}$.

3) Adaptation layer: The parameters involved in (43) 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.

References

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

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

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

[4] 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.

[5] S. M. Meerkov, “Mathematical theory of behavior,” *Mathematical Biosciences*, vol. 43, pp. 41–106, 1979.

[6] M. Amin, “Toward secure and resilient interdependent infrastructures,” *Journal of Infrastructure Systems*, vol. 8, pp. 67–75, 2002.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] Y. Zhang and J. Jiang, “Bibliographical review on reconfigurable fault-tolerant control systems,” *IFAC Annual Reviews in Control*, vol. 32, pp. 229–252, 2008.

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] ——, “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.

[23] ——, “Resilient plant monitoring system: Design, analysis, and performance evaluation,” in *Proc. 52nd IEEE Conf. on Decision and Control*, Florence, Italy, Dec., 2013.

[24] ——, “Resilient monitoring systems: Architecture, design, and application to boiler/turbine plant,” *accepted for publication, IEEE Trans. on Cybernetics*, March, 2014.

[25] 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.

[26] ——, “Rational probabilistic deciders - Part II: Collective behavior,” *Mathematical Problems in Engineering*, vol. 2007, no. 82184, pp. 1–34, 2007.

[27] M. A. McQueen and A. Giani, “Known secure sensor measurements for critical infrastructure systems: detecting falsification of system state,” in *Proc. 3rd Int. Conf. on Software Engineering for Resilient Systems*, Geneva, Switzerland, Sept. 29 - 30, 2011, pp. 156–163.

[28] 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.

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

[30] H. K. Khalil, *Nonlinear Systems*. Prentice Hall, 2002.