1-2018

Parameter-Invariant Monitor Design for Cyber Physical Systems

James Weimer
University of Pennsylvania, weimerj@cis.upenn.edu

Radoslav Ivanov
University of Pennsylvania, rivanov@cis.upenn.edu

Sanjian Chen
University of Pennsylvania, sanjian@cis.upenn.edu

Alexander Roederer
University of Pennsylvania, roederer@cis.upenn.edu

Oleg Sokolsky
University of Pennsylvania, sokolsky@cis.upenn.edu

See next page for additional authors

Recommended Citation

James Weimer, Radoslav Ivanov, Sanjian Chen, Alexander Roederer, Oleg Sokolsky, and Insup Lee, "Parameter-Invariant Monitor Design for Cyber Physical Systems", *Proceedings of the IEEE* 106(1), 71-92. January 2018. http://dx.doi.org/10.1109/JPROC.2017.2723847

Proceedings of the IEEE, 106 (1), Jan 2018. pp. 71-92.

Parameter-Invariant Monitor Design for Cyber Physical Systems

Abstract
The tight interaction between information technology and the physical world inherent in Cyber-Physical Systems (CPS) can challenge traditional approaches for monitoring safety and security. Data collected for robust CPS monitoring is often sparse and may lack rich training data describing critical events/attacks. Moreover, CPS often operate in diverse environments that can have significant inter/intra-system variability. Furthermore, CPS monitors that are not robust to data sparsity and inter/intra-system variability may result in inconsistent performance and may not be trusted for monitoring safety and security. Towards overcoming these challenges, this paper presents recent work on the design of parameter-invariant (PAIN) monitors for CPS. PAIN monitors are designed such that unknown events and system variability minimally affect the monitor performance. This work describes how PAIN designs can achieve a constant false alarm rate (CFAR) in the presence of data sparsity and intra/inter system variance in real-world CPS.

To demonstrate the design of PAIN monitors for safety monitoring in CPS with different types of dynamics, we consider systems with networked dynamics, linear-time invariant dynamics, and hybrid dynamics that are discussed through case studies for building actuator fault detection, meal detection in type I diabetes, and detecting hypoxia caused by pulmonary shunts in infants. In all applications, the PAIN monitor is shown to have (significantly) less variance in monitoring performance and (often) outperforms other competing approaches in the literature. Finally, an initial application of PAIN monitoring for CPS security is presented along with challenges and research directions for future security monitoring deployments.

Disciplines
Computer Engineering | Computer Sciences

Comments
Proceedings of the IEEE, 106 (1), Jan 2018. pp. 71-92.

Author(s)
James Weimer, Radoslav Ivanov, Sanjian Chen, Alexander Roederer, Oleg Sokolsky, and Insup Lee

Parameter-Invariant Monitor Design for Cyber Physical Systems

James Weimer, *Member, IEEE,* Radoslav Ivanov, *Student Member, IEEE,* Sanjian Chen, Alexander Roederer, Oleg Sokolsky, *Member, IEEE,* and Insup Lee, *Fellow, IEEE,*

***Abstract*—The tight interaction between information technology and the physical world inherent in Cyber-Physical Systems (CPS) can challenge traditional approaches for monitoring safety and security. Data collected for robust CPS monitoring is often sparse and may lack rich training data describing critical events/attacks. Moreover, CPS often operate in diverse environments that can have significant inter/intra-system variability. Furthermore, CPS monitors that are not robust to data sparsity and inter/intra-system variability may result in inconsistent performance and may not be trusted for monitoring safety and security. Towards overcoming these challenges, this paper presents recent work on the design of parameter-invariant (PAIN) monitors for CPS. PAIN monitors are designed such that unknown events and system variability minimally affect the monitor performance. This work describes how PAIN designs can achieve a constant false alarm rate (CFAR) in the presence of data sparsity and intra/inter system variance in real-world CPS.**

To demonstrate the design of PAIN monitors for safety monitoring in CPS with different types of dynamics, we consider systems with networked dynamics, linear-time invariant dynamics, and hybrid dynamics that are discussed through case studies for building actuator fault detection, meal detection in type I diabetes, and detecting hypoxia caused by pulmonary shunts in infants. In all applications, the PAIN monitor is shown to have (significantly) less variance in monitoring performance and (often) outperforms other competing approaches in the literature. Finally, an initial application of PAIN monitoring for CPS security is presented along with challenges and research directions for future security monitoring deployments.

I. Introduction

The confluence of low-powered low-cost embedded communication, sensing, and actuation technologies with unprecedented computational capabilities stands to revolutionize safety-critical systems and infrastructures. These cyber-physical systems (CPS) have already shaped the way we interact with the world by enabling, for example, smart grids, autonomous driving, medical screening and control, energy efficient buildings, and advanced manufacturing. In the smart grid, networked phasor measurement units (PMUs) have enabled high-fidelity monitoring and control of grid voltage [1], [2], [3], [4], [5]. In medicine, computerized clinical decision support systems (CDSS) can automatically analyze data and provide the corresponding results to clinicians [6], [7], [8], enabling improved patient care and outcomes [9], [10], [11]. In smart buildings, incorporating CPS technologies into existing supervisory control and data acquisition (SCADA) systems has enabled improved building energy management [12], [13], [14], [15]. Similarly, pervasive CPS monitoring technologies have enabled manufacturing systems to quickly identify faults in process control and infrastructures [16], [17], [18]. Recently, data-driven monitoring techniques have been developed for autonomous vehicles to identify potential security threats and critical events [19], [20], [21], [22]

Despite these successes, cyber-physical systems are still prone to failures and vulnerable to attacks, as illustrated in the Maroochy water breach in March 2000 [23], the 2003 northeast blackout in the USA [24] and multiple recent power blackouts in Brazil [25], the SQL Slammer worm attack on the DavisBesse nuclear plant in January 2003 [26], the StuxNet computer worm in June 2010 [27], the recent fatal wrecks involving autonomous cars [28], the recall of medical devices [29], [30], and various industrial security incidents [31]. Moreover, there has been an exponential increase in the number of CPS safety and security incidents in the past decade [32]. Thus, it is imperative that safety and security monitoring techniques be developed for CPS and adopted into their designs.

Complicating this effort, the tight interaction between information technology and the physical world inherent in Cyber-Physical Systems (CPS) can challenge traditional approaches for monitoring safety and security. While new sensing technologies can produce large amounts of data, this data is often *sparse* and may lack rich data describing critical events/attacks. Data sparsity in safety critical CPS is especially troublesome for data-driven techniques, where an insufficiently trained CPS can respond unfavorably to unforeseen events/scenarios (*e.g.,* the 2016 Tesla crash [28]). Moreover, in many applications, data is often corrupted due to sensor noise, unmodeled data artifacts, system interactions, and lossy communication connections. In addition to data sparsity and integrity concerns, CPS often have (significant) intra/inter-system variability. For example, in medical CPS, intra-system variability can result from time-varying physiological effects of a specific patient, while inter-system variability results when a medical device must interact sequentially with multiple patients – each with different physiology.

A fundamental challenge arises in designing CPS monitors robust to data sparsity and inter/intra-system variability since monitors that have significant variability in their performance may not be trusted for monitoring safety and security. Safety-critical CPS performance must provide high performance in all operating scenarios, not just the average scenario. For CPS monitoring results to be useful, they must not only be accurate (as failure might lead to injury or system failure) but also precise. In CPS, a high rate of false positives will fatigue human operators and, over time, can cause them to ignore the

system's recommendations [33]. Trust in a system's results has been found to be a major determinant in the success of monitoring systems [34]. Thus, CPS monitors must produce consistent and predictable results over all operating scenarios. Monitoring systems that do not address these challenges risk becoming another source of information that system operators ignore. This is already apparent in medicine where one of the most pressing issues for the simple clinical decision and support systems in widespread use today. For example, bedside threshold alarm systems are known to generate the large numbers of false positives, which has created the "alarm fatigue" problem, i.e., clinicians ignoring alarms or completely shutting them down [35], [36], [37].

Related Work: Concerns about safety and security of CPS are not new, as the numerous manuscripts on systems attack/fault detection, isolation, and recovery testify – *e.g.,* [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54]. Currently, there exist several approaches to designing monitors ranging from simple sensor thresholding techniques to more complex data-driven learning approaches. One common approach is referred to as a model-based approach, in which physics-based models (e.g., compartmental models [55]) are developed that attempt to capture the relevant processes in the physical world. These techniques perform well when accurate models are developed [56], [57], [58], [59], [50], [48] but they can require extensive training data to tune parameters, especially in settings with large inter-system variability.

When high-fidelity accurate models are not available, researchers employ gray-box [60], [61], [62] or black-box (data-driven) approaches [63], [64], [65], [66], [67], [68]. Data-driven techniques have shown improved performance over competitive approaches when rich/dense training data is available [69], [70]. At the same time, such approaches (including Hidden Markov Models [71]) suffer from the often incorrect assumption that data is generated from a discrete, fixed set of classes that do not change over time – as a result, the performance of these techniques varies greatly across operating scenarios. Continuous density hidden Markov models [72] and Hierarchical Bayesian modeling [73], [74] address the fixed-class assumption and thus have seen success in various CPS applications (*e.g.,* medical CPS [75], [76], [67]), but still provide no individual level guarantees. Recent work on fair learning [77], [78] aims to overcome this shortcoming in data-driven monitoring. It is worth noting that in special cases, such as monitoring malware in CPS communication networks, research have developed *model-free* techniques by exploiting network structure [53], [54].

An alternative approach shown to be robust to data sparsity and inter/intra-system variability is parameter-invariant (PAIN) monitoring [79], [80], [81], [82], [83], [84], [85]. In PAIN monitors, a parameterized model is developed for the purposes of classification – instead of estimating the model parameters, however, PAIN classification provides guaranteed performance (i.e., constant false positive rates) regardless of the parameter values. This approach has been successfully applied in multiple CPS domains with parameterized models such as fault detection in networked systems [79], [80] as well as the development of heating, ventilating and air conditioning systems [86], smart grids [52], and in medical monitoring applications for critical pulmonary shunts [81], [82], hypovolemia [83] and meal-detection in type I diabetics [84], [85].

This paper presents the design methodology for parameter-invariant monitors in safety-critical CPS. The presented design methodology is shown to generalize the recent applications of PAIN monitoring for safety critical and secure CPS. We illustrate the versatility and different features of PAIN monitoring through real-world CPS applications spanning systems with networked dynamics, linear-time invariant (LTI) dynamics, and hybrid dynamics. These applications include: (i) monitoring actuator fault detection in smart buildings that is invariant to changes in the heat-transfer dynamics (*e.g.,* invariant to whether doors and windows are open or closed); (ii) detecting meal ingestion by type 1 diabetic patients that is invariant to the patient's unique physiology; and (iii) predicting hypoxia, caused be a pulmonary shunt, in infants under anesthesia that is invariant to the oxygen diffusion, lung thickness, etc. In all applications, the PAIN monitor is shown to have (significantly) less variance in monitoring performance and (often) outperforms other competing approaches in the literature. An initial application of PAIN monitoring for CPS security is presented along with challenges and research directions for future security monitoring deployments.

In the following, our presentation of PAIN monitoring for CPS is structured as follows. In the subsequent section, we begin by providing an overview of binary hypothesis testing techniques that form the foundation of PAIN monitoring. Section III presents the design of parameter-invariant monitors and discusses its theoretical performance. Utilizing the PAIN design, we then describe multiple monitoring scenarios and discuss them in the context of real-world CPS applications. Networked dynamical systems are discussed in Section IV and utilizes the building fault detection application as an illustrative example. Section V discusses CPS with LTI dynamics and demonstrates a corresponding PAIN monitor for meal detection in type I diabetics. To illustrate the final class of system dynamics, namely hybrid systems, a monitor for critical pulmonary shunts is presented in Section VI. An initial application and future challenges for extending PAIN monitors for CPS security monitoring is discussed in Section VII. The final section provides a discussion and conclusions.

II. Foundations of Binary Hypothesis Testing

Classical binary hypothesis testing concerns the problem of discriminating, based on available measurements, between a *null hypothesis* that captures normal operation (e.g., absence of a fault) and an *event hypothesis* that models the event/alternative scenario (e.g., presence of a fault). To formulate a (binary) hypothesis testing problem, a designer can utilize two models: (i) a noisy *measurement model* relating measurements to underlying parameters; (ii) a *hypothesis model* mapping parameters to potential hypotheses. The noisy measurement model can be represented as a random measurement vector $\boldsymbol{y} \in \mathcal{Y} \subseteq \mathbb{R}^N$ which is drawn from a distribution $\boldsymbol{y} \sim f_{\boldsymbol{\theta},\boldsymbol{\gamma}}(Y)$ that is parametrized by parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^M$

and $\boldsymbol{\gamma} \in \Gamma \subseteq \mathbb{R}^J$, where $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ correspond to *test parameters* and *nuisance parameters*, respectively. The test parameters, $\boldsymbol{\theta}$, are utilized in the hypothesis model such that

$$\mathcal{H}_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{vs.} \quad \mathcal{H}_1 : \boldsymbol{\theta} \in \Theta_1, \tag{1}$$

where, $\mathcal{H}_0$ represents the *null hypothesis*, $\mathcal{H}_1$ denotes the *event hypothesis*, and $\Theta_0, \Theta_1 \subset \Theta$ are the parameter subspaces for the null and event hypotheses, respectively. As their name suggests, the nuisance parameters (denoted by $\boldsymbol{\gamma}$) do not discriminate between the hypotheses (*i.e.,* do not appear in the hypothesis model), but affect the measurement model. Thus, a primary challenge in designing tests for (binary) hypothesis testing problems lies in maximizing test performance in spite of the nuisance parameters.

The performance of a binary test can be quantified in terms of *false positives* and *true positives*. A false positive occurs when the test incorrectly rejects the null hypothesis (*i.e.,* $\phi(\boldsymbol{y}) = 1$ when $\mathcal{H}_0$ is true). Similarly, a *true positive* occurs when the test correctly rejects the null hypothesis (*i.e.,* $\phi(\boldsymbol{y}) = 1$ when $\mathcal{H}_1$ is true). In general, minimizing the false positive rate and maximizing the true positive are competing objectives – decreasing the false positive rate tends to also decrease the true positive rate (and vice versa).

There are two fundamental approaches to hypothesis testing. A Bayesian approach assumes a prior on the likelihood of each hypothesis occurring in the hypothesis model, while a non-Bayesian approach (*e.g.,* frequentist approach) does not utilize a prior distribution assumption on the hypothesis model. While each approach has its benefits and shortcomings, this work concerns CPS safety and security scenarios where the a prior distribution on the hypothesis model is likely unknown. In this scenario, multiple testing techniques exist and can be classified based on parameter subspace assumptions. When a parameter subspace for a hypothesis contains a single element, *i.e.,* $\Gamma = \{\boldsymbol{\gamma}\}$ and $\Theta_0 = \{\boldsymbol{\theta}_0\}$ or $\Theta_1 = \{\boldsymbol{\theta}_1\}$, the corresponding hypothesis is said to be a *simple hypothesis*. Otherwise, the hypothesis is called a *composite hypothesis*. Regardless of whether the hypotheses are simple or composite, the aim in binary hypothesis testing is to design a test that maps the measurement (subspace) to a hypothesis, $\phi : \mathcal{Y} \to \{0, 1\}$, such that $\phi(\boldsymbol{y}) = j$ represents the test's claim that $\mathcal{H}_j$ is true. In the remainder of this section, we overview different types of tests and their performance in the presence of simple and composite hypotheses as motivation for the PAIN test in the following section.

A. Likelihood Ratio Tests

For tests between two simple hypotheses, *i.e.,* $\Gamma = \{\boldsymbol{\gamma}\}$, $\Theta_0 = \{\boldsymbol{\theta}_0\}$ and $\Theta_1 = \{\boldsymbol{\theta}_1\}$, the classic *likelihood ratio test (LRT)* introduced by Neyman and Pearson [87] is often employed. The LRT is constructed by utilizing the *likelihood ratio*,

$$l(\boldsymbol{y}) = \frac{f_{\boldsymbol{\theta}_1,\boldsymbol{\gamma}}(\boldsymbol{y})}{f_{\boldsymbol{\theta}_0,\boldsymbol{\gamma}}(\boldsymbol{y})},$$

which represents the likelihood of the measurement being drawn from the event hypothesis divided by the likelihood of the measurement being drawn from the null hypothesis. To construct a test, a threshold, η, is applied to the likelihood ratio such that

$$\phi_{LRT}(\boldsymbol{y}) = \begin{cases} 0 & \text{if } l(\boldsymbol{y}) \leq \eta \\ 1 & \text{otherwise} \end{cases}, \tag{2}$$

is the likelihood ratio test.

For a large class of distributions, the likelihood ratio has a monotonic (increasing) distribution such that the Neyman-Pearson Lemma [87] ensures the likelihood ratio test is the *uniformly most powerful (UMP)* test. A UMP test is considered the ideal test for binary hypothesis testing and ensures that for every false positive rate (*i.e.,* size), the corresponding true positive rate (*i.e.,* power) is larger than any other test with the same (or less) false positive rate. More formally, a candidate test, ϕ, is UMP with false alarm rate $\alpha \in [0, 1]$ if and only if the false alarm rate equals α (*i.e.,* $P[\phi(\boldsymbol{y}) = 1 \mid \mathcal{H}_0] = \alpha$) and for any other test, ϕ', such that has no greater false alarm rate (*i.e.,* $P[\phi'(\boldsymbol{y}) = 1 \mid \mathcal{H}_0] \leq \alpha$), the candidate test has at least the same true positive rate (*i.e.,* $P[\phi(\boldsymbol{y}) = 1 \mid \mathcal{H}_1] \geq P[\phi'(\boldsymbol{y}) = 1 \mid \mathcal{H}_1]$). Unfortunately, UMP tests often do not exist for composite hypotheses since it requires a test to maximize power (*i.e.,* maximize the true positive rate) for a specific size (*i.e.,* false alarm rate) over all parameters, $\boldsymbol{\theta} \in \Theta$.

B. Generalized Likelihood Ratio Tests (GLRT)

In cases where the null hypothesis and/or the event hypothesis is composite, the *generalized likelihood ratio test* (GLRT) extends the notion of the likelihood ratio test by utilizing the ratio of the maximum likelihood under each hypothesis,

$$\hat{l}(\boldsymbol{y}) = \frac{\max\limits_{\boldsymbol{\theta}\in\Theta_1,\boldsymbol{\gamma}\in\Gamma} f_{\boldsymbol{\theta},\boldsymbol{\gamma}}(\boldsymbol{y})}{\max\limits_{\boldsymbol{\theta}\in\Theta_0,\boldsymbol{\gamma}\in\Gamma} f_{\boldsymbol{\theta},\boldsymbol{\gamma}}(\boldsymbol{y})},$$

and constructing a threshold test similar to the likelihood ratio test,

$$\phi_{GLRT}(\boldsymbol{y}) = \begin{cases} 0 & \text{if } \hat{l}(\boldsymbol{y}) \leq \eta \\ 1 & \text{otherwise} \end{cases}. \tag{3}$$

One attractive feature of the GLRT is that when both hypotheses are simple, the test is equivalent to the LRT; thus, when the LRT is a UMP test, the GLRT is also a UMP test. Moreover, when the generalized likelihood ratio has a monotonic distribution, the Karlin-Rubin Theorem [88] states that the GLRT is UMP. In general, a UMP test for composite hypothesis testing only exists in a few special cases (*e.g.,* simple hypotheses). A benefit of the GLRT is that when the estimated value of the parameter $\boldsymbol{\theta}$ under the true hypothesis is accurately estimated, the GLRT asymptotically approaches the performance of a test that has access to the unknown parameters [89]. However, as the maximum likelihood estimate of the parameter deviates for the actual value, the performance can degrade rapidly.

C. Maximally Invariant Tests

In CPS, it is entirely possible (if not likely) that the maximum likelihood estimate of the nuisance parameters deviate

from their true values. This deviation can be caused by a number of effects, including too little data and/or poor distribution/model assumptions. In these situations, an attractive alternative is to design a *maximally invariant test*. Maximally invariant tests attempt to remove the effect of the nuisance parameters from the test decision. Modeling the effect of the nuisance parameters on the measurements as a group of measurement transformations, $\mathcal{G} \subset \mathcal{Y}^{\mathcal{Y}}$, allows us to capture how changing the nuisance parameters can change the measurement. We provide examples of such groups in the next section in our presentation of the parameter-invariant test.

A test statistic, $t(\boldsymbol{y})$, is invariant to the group of transformations, $\mathcal{G}$, if and only if

$$\forall g \in \mathcal{G}, \quad t(\boldsymbol{y}) = t(g(\boldsymbol{y}))$$

where, in words, we say the statistic is invariant to the group of transformation if the statistic has the same value regardless of the unknown parameters. There are many invariant statistics (e.g. $t(\boldsymbol{y}) = 0$ is trivially invariant to all parameters, but is useless as a test statistic.). Thus, we wish (if possible) to choose an invariant statistic that is *maximally invariant*, namely the statistic which only removes the effect of the unknown nuisance parameters. This concept is captured mathematically by the following implication: for any two measurement vectors $\boldsymbol{y}, \boldsymbol{y}' \in \mathcal{Y}$,

$$t(\boldsymbol{y}) = t(\boldsymbol{y}') \longrightarrow \exists g \in \mathcal{G}, \quad \boldsymbol{y} = g(\boldsymbol{y}').$$

In words, a statistic is maximally invariant if it is invariant to the unknown nuisance parameters *and* if the information removed by the invariant statistics can be fully explained by a change in the unknown nuisance parameters. Similar to the LRT and GLRT, this statistic can be utilized to design a (maximally) invariant test

$$\phi_{MI}(\boldsymbol{y}) = \begin{cases} 0 & \text{if } t(\boldsymbol{y}) \leq \eta \\ 1 & \text{otherwise} \end{cases}, \tag{4}$$

where, as long as the maximally invariant statistic preserves the original testing dichotomy, the maximally invariant test will be invariant to $\mathcal{G}$ (*i.e.,* $\forall g \in \mathcal{G}, \ \phi(g(\boldsymbol{y})) = \phi(\boldsymbol{y})$). Additionally, an optimal maximally invariant test is said to be the *uniformly most powerful invariant (UMPI)* test, which has a direct relation to the UMP test. Specifically, a maximally invariant test to $\mathcal{G}$, namely ϕ, is UMPI if for any other test, ϕ', also invariant to $\mathcal{G}$, ϕ is UMP. Similar to the case for the UMP test, a direct consequence of Neyman-Pearson Lemma [90] and the Karlin-Rubin Theorem [88], if a maximally invariant statistic $t : \mathcal{Y} \to \mathcal{T}$ has a monotone likelihood ratio, then the corresponding likelihood ratio test, $\phi : \mathcal{T} \to \{0, 1\}$, is UMPI.

Realizing a maximally invariant test is not always possible. Maximally invariant statistics do not yield constant false alarm rates for hypothesis testing problems where the test parameter sets contain more than a single value (*i.e.,* the test parameters are composite). In these scenarios, maximally invariant statistics can have wide ranging performance. Especially in systems where the hypotheses can be potentially incorrect. In these scenarios, we wish to accept/reject the null hypothesis for the right reason – and not just because there is a discrepancy between the real-world and the assumed models.

III. Parameter-Invariant Monitor Design

In many CPS monitoring applications, designing robust monitors in the presence of unknown or uncertain physical processes presents a fundamental challenge. In general, unknown physical processes can affect both the normal scenario (null hypothesis) and the potential event (event hypothesis), resulting in a problem where both the null and event hypotheses are composite. In this section, we present the design of parameter-invariant monitors to robustly address real-world uncertainty. A PAIN monitor consists of a test of near maximally invariant statistics. In a CPS monitoring context, the PAIN monitor removes the effect of the unknown nuisance parameters (using maximally invariant statistics) which do not reliably differentiate the hypotheses – *i.e.,* are unnecessary for monitoring. Theoretically, PAIN designs can achieve a constant false alarm rate (CFAR) regardless of parameter uncertainty. Practically, PAIN designs have provided near-constant false alarm rates in many CPS applications spanning networks [80], smart buildings [91], and medicine [92], [82], [83].

In this section, we present the design of parameter-invariant monitors in three parts. In the following subsection, we discuss the modeling of physical processes for parameter invariant monitoring and present a general form of the measurement and hypotheses models. Next, in Section III-B, we formulate maximally invariant statistics for the nuisance parameters in the measurement model. The final subsection utilizes the maximally invariant statistics to construct a parameter-invariant test and discusses its sequential implementation. Before presenting PAIN monitor design, we briefly summarize the notation employed in the following.

Notation: We write $\mathbb{R}_{\geq 0}$ to be the set of non-negative real numbers. For a matrix, $\boldsymbol{F}$, we write $\boldsymbol{F}^{-1}$, $\boldsymbol{F}^{\top}$, and $|\boldsymbol{F}|$ to be the inverse, transpose, and determinant of $\boldsymbol{F}$, respectively. Additionally, we write $|\boldsymbol{F}|_c$ to be the number of columns of $\boldsymbol{F}$ and denote the *column space* of $\boldsymbol{F}$ as $\langle \boldsymbol{F} \rangle$ – *i.e.,* $\langle \boldsymbol{F} \rangle$ denotes the range that can be modeled as a linear combination of the columns of $\boldsymbol{F}$. For completeness, we write $\langle \boldsymbol{F} \rangle^{\perp}$ to denote the *nullspace* of $\boldsymbol{F}^{\top}$. For a random variable, $\boldsymbol{n}$, we write $\mathbb{E}[\boldsymbol{n}]$ to be the expected value of $\boldsymbol{n}$.

A. Model Development for PAIN Monitoring

Consistent with the hypothesis testing foundations in Section II, designing a PAIN monitor also requires a measurement model and hypothesis model. In general, the measurement model captures the effects of the nuisance and test parameters on the measurements, while the hypothesis model describes the test parameters under each hypothesis. However, in CPS, relating models of the physical world (and their corresponding parameters) to measurement and hypothesis models (having nuisance and test parameters) can be challenging conceptually. In this section, we present a general form for the measurement and hypothesis models utilized in PAIN monitoring. In later sections, we will show how the general form of both the measurement and hypothesis models can be realized for different real-world CPS monitoring applications.

1) Measurement Model: In PAIN monitoring, we assume the measurement vector, $\boldsymbol{y}$, can be written as a linear model of M (potentially unknown) parameters, $\boldsymbol{p} \in \mathbb{R}^{M-1}$ and $\sigma \in \mathbb{R}_{\geq 0}$, according to

$$\begin{aligned} \boldsymbol{y} =& \boldsymbol{H}\boldsymbol{p} + \sigma\boldsymbol{n} = \begin{bmatrix} \boldsymbol{F} & \boldsymbol{G}_0 & \boldsymbol{G}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \theta_0\boldsymbol{\rho}_0 \\ \theta_1\boldsymbol{\rho}_1 \end{bmatrix} + \sigma\boldsymbol{n} \\ =& \boldsymbol{F}\boldsymbol{\mu} + \theta_0\boldsymbol{G}_0\boldsymbol{\rho}_0 + \theta_1\boldsymbol{G}_1\boldsymbol{\rho}_1 + \sigma\boldsymbol{n} \end{aligned} \tag{5}$$

where, $\boldsymbol{H} = [\boldsymbol{F}, \boldsymbol{G}_0, \boldsymbol{G}_1]$ is a known *signal matrix* whose columns correspond to potential signals and $\boldsymbol{n}$ is a random noise. In this model, there are no restrictions on what the parameters represent physically. For instance, the parameters can represent physical parameters (*e.g.,* metabolic rate [82]), but could also represent lumped parameters that have no explicit physical world interpretation (*e.g.,* coefficients of a transfer function [83], [91]). In PAIN monitoring, we assume that $\boldsymbol{\mu} \in \Gamma_\mu = \mathbb{R}^{|\boldsymbol{F}|_c}$, $\boldsymbol{\rho}_i \in \Gamma_{\rho,i} = \{\boldsymbol{\rho} \in \mathbb{R}^{|\boldsymbol{G}_i|_c} \mid \|\boldsymbol{\rho}\| = 1\}$, and $\sigma \in \Gamma_\sigma = \{\sigma \in \mathbb{R} \mid \sigma > 0\}$ are nuisance parameters, and $\theta_0, \theta_1 \in \mathbb{R}_{\geq 0}$ are test parameters. For completeness, this means the nuisance parameters, $\boldsymbol{\gamma}$, and test parameters, $\boldsymbol{\theta}$, introduced in Section II correspond to

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\rho}_0 \\ \boldsymbol{\rho}_1 \\ \sigma \end{bmatrix} \quad \text{and} \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}.$$

Observing that the test parameter θ_i can be interpreted as a *magnitude* since it can take any non-negative value (*i.e.,* $\theta_i \in \mathbb{R}_{\geq 0}$). Similarly, the nuisance parameter $\boldsymbol{\rho}_i$ can be interpreted as a *direction* since it is any $|\boldsymbol{G}_i|_c$-dimensional vector that has unit length (*i.e.,* $\boldsymbol{\rho}_i \in \Gamma_{\rho,i}$). Consequently, the product of these parameters represents any vector of the same dimension as $\boldsymbol{\rho}_i$. Thus, the measurement model can capture any linear combination of signals (*i.e.,* the columns of $\begin{bmatrix} \boldsymbol{F} & \boldsymbol{G}_0 & \boldsymbol{G}_1 \end{bmatrix}$), since the corresponding parameters $\boldsymbol{p}$ – made up of $\boldsymbol{\mu}$, $\theta_0\boldsymbol{\rho}_0$, and $\theta_1\boldsymbol{\rho}_1$ – can all be arbitrary vectors of appropriate dimensions. It will be shown through different applications in later sections that the model in Equation 5 is capable of accurately capturing the dynamics of many real world CPS since it can represent any *unknown linear combination of known signals* – plus noise.

While the parameters in $\boldsymbol{p}$ denote a linear combination of known signals, the parameter σ multiplies the noise vector, $\boldsymbol{n}$. The combined effect of σ and $\boldsymbol{n}$ can have multiple interpretations. Similar to θ_i, σ can be thought of as an unknown magnitude and $\boldsymbol{n}$ as a random direction. In this interpretation, we claim that the magnitude of the noise is a nuisance while the direction is a random variable. More commonly, if the noise is an i.i.d. with zero mean (*i.e.,* $\mathbb{E}[\boldsymbol{n}] = \boldsymbol{0}$), then σ represents the covariance of the noise. This interpretation can lead to nice theoretical results for some common distributions, *e.g.,* Gaussian noise. Regardless of the exact interpretation, σ accounts for the measurement model error. Thus, when some linear combination of the signals accurately captures the measurement, σ is small. When the model is inaccurate, σ is larger. In general, knowing the accuracy of a model in a real-world CPS may be unreasonable, thus in PAIN monitoring we refer to σ as the magnitude of the model error and treat it as a nuisance parameter. In general, when interpreting $\sigma\boldsymbol{n}$ as a *model error*, we observe that any system (linear or non-linear) can be modeled by Equation 5; however, the magnitude of the model error (*i.e.,* σ) increases as model accuracy decreases. The effect of this relation will be discussed in terms of the performance of the PAIN monitor.

2) Hypothesis Model: As identified in the previous section, the test parameters in PAIN monitoring correspond to θ_0 and θ_1. Consequently, we state the hypothesis model used in PAIN monitoring as

$$\mathcal{H}_0 : \theta_0 > 0, \theta_1 = 0 \quad \text{vs.} \quad \mathcal{H}_1 : \theta_0 = 0, \theta_1 > 0 \tag{6}$$

where, the hypotheses are both composite. The interpretation of the hypothesis model is that under the null hypothesis, $\mathcal{H}_0$, the signals corresponding to the columns of $\boldsymbol{G}_1$ do not affect the measurements. Similarly, under the even hypothesis, $\mathcal{H}_1$, signals denoted by the columns of $\boldsymbol{G}_0$ do not affect the measurements. Restricting the hypothesis model in PAIN monitoring to the form of Equation 6, often requires a co-design between the measurement model and hypothesis model. For example, when testing whether the system is in one mode or another (possibly corresponding to a normal model versus an attack/fault mode), the matrix $\boldsymbol{G}_0$ can represent the signals that can possibly affect the measurements when no fault/attack is present, while $\boldsymbol{G}_1$ may represent potential fault/attack signals. In this scenario, under the null hypothesis (no attack/fault), the test parameter θ_1 equals zero and denotes that none of the fault/attack signals in $\boldsymbol{G}_1$ affect the measurement. Similarly, in the event hypothesis (corresponding to an attack/fault), the measurements are not affected by the nominal signals in $\boldsymbol{G}_0$, since θ_0 is equal to zero. Moreover, we note that $\boldsymbol{F}$ in Equation 5 denotes the signals that can affect the measurement under both the null and event hypothesis (if they exist). Depending on the application, the specific signals contained in $\boldsymbol{F}$, $\boldsymbol{G}_0$, and $\boldsymbol{G}_1$ may change; however, in PAIN monitoring we restrict the corresponding measurement model and hypothesis model to the form of Equations 5 and 6, respectively. This co-design is discussed in the context of various CPS monitoring applications in later sections.

B. Maximally Invariant Statistics

In this section, we develop statistics that are maximally invariant to the nuisance parameters in the measurement model in Equation 5. To achieve this, for each nuisance parameter (*i.e.,* μ, ρ_i, and σ), we discuss its effect in words, then formalize the effect as a *group of transformations* induced upon the measurements by changing the nuisance parameters. These groups induce *orbits* within the measurement space such that all measurements on a unique orbit can be achieved by (only) changing the nuisance parameters. For simplicity in the following discussion, we make some additional assumptions on the structure of the measurement model

- The columns of $\begin{bmatrix} \boldsymbol{F} & \boldsymbol{G}_0 & \boldsymbol{G}_1 \end{bmatrix}$ are orthonomal

Any pair of measurement and hypothesis models having the form of Equation 5 and Equation 6, respectively, can

be restated such that the new measurement and hypothesis models still have the form of Equation 5 and Equation 6 *and* this assumption holds *while* preserving the dichotomy of the original testing problem. We note that this transformation can be automated (*e.g.,* see [82], [85]).

In the following, we first present maximally invariant statistics for each nuisance parameter independently, then compose the individual statistics (under the above assumption) to generate maximally invariant statistics for the combined effect of the nuisance parameters. To illustrate the effects and the nuisance parameters and motivate their corresponding maximally invariant statistics, we make use of Figure 1.

- **Maximal Invariance to $\boldsymbol{\mu}$:** Given the measurement vector $\boldsymbol{y}$ in Figure 1a, any change in the nuisance parameter only translates the measurement in column space of $\boldsymbol{F}$ (*i.e.,* in $\langle \boldsymbol{F} \rangle$). This is illustrated in Figure 1a by assuming $\hat{\boldsymbol{\mu}}$ denotes a change in the nuisance parameter such that a resulting change in the measurement is denoted as $\boldsymbol{y} + \boldsymbol{F}\hat{\boldsymbol{\mu}}$. This nuisance parameter-induced translation is referred to as a *subspace bias*, and can arbitrarily translate the portion of the measurement vector in the column space of $\boldsymbol{F}$. Formally, the set of all subspace biases induced by changing the nuisance parameter $\boldsymbol{\mu}$ is captured by the group of transformations,

  $$\mathcal{G}_{bias} = \left\{ g \;\middle|\; g(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{F}\hat{\boldsymbol{\mu}},\ \hat{\boldsymbol{\mu}} \in \Gamma_{\mu}, \right\}$$

  where, each $g \in \mathcal{G}_{bias}$ represents a directional bias induced by a change in the nuisance parameter $\boldsymbol{\mu}$.
  We observe that the projection of the original measurement and the biased measurement onto the nullspace of $\boldsymbol{F}^{\top}$ – denoted by the red vector in Figure 1a – are the same, regardless of the induced bias. Thus, the nullspace of $\boldsymbol{F}^{\top}$ represents the lumped region of the measurement space where only changing the value of $\boldsymbol{\mu}$ will have no effect on the measurement. By exploiting this property, we can project the measurement vector onto the nullspace such that changing $\boldsymbol{\mu}$ will have no effect on the resulting projection. Mathematically we write this projected measurement as

  $$t_{\mu}(\boldsymbol{y}) = \left(\boldsymbol{I} - \boldsymbol{F}\boldsymbol{F}^{\top}\right)\boldsymbol{y}, \tag{7}$$

  where, $\boldsymbol{I} - \boldsymbol{F}\boldsymbol{F}^{\top}$ is the projection onto the null space of $\boldsymbol{F}^{\top}$. It is well known (and shown in [93]) that $t_{\mu}(\boldsymbol{y})$, is maximally invariant $\mathcal{G}_{bias}$ corresponding to a subspace bias.
- **Maximal Invariance to $\boldsymbol{\rho}_i$:** Given the measurement vector $\boldsymbol{y}$ in Figure 1b, any change in the nuisance parameter $\boldsymbol{\rho}_i$ rotates the measurement in column space of $\boldsymbol{G}_i$ (*i.e.,* in $\langle \boldsymbol{G}_i \rangle$). This effect can be explained in two steps. First, since $\boldsymbol{G}_i$ is an orthonormal basis and $\boldsymbol{\rho}_i \in \Gamma_{\rho,i}$ always has unit length, then the $\boldsymbol{G}_i\boldsymbol{\rho}_i$ will always have unit magnitude (*i.e.,* $\forall \boldsymbol{\rho}_i \in \Gamma_{\rho,i},\ \|\boldsymbol{G}_i\boldsymbol{\rho}_i\| = 1$). Second, if changing the parameter always results in a vector with unit magnitude, then the change is a *rotation* (*i.e.,* a change in direction only). Formally, the set of all rotations induced by changing the nuisance parameter $\boldsymbol{\rho}_i$ is stated by letting $\mathcal{R}$ be the set of rotation matrices,

  $$\mathcal{R} = \left\{ \boldsymbol{R} \;\middle|\; \boldsymbol{R}^{-1} = \boldsymbol{R}^{\top},\ |\boldsymbol{R}| = 1 \right\},$$

  and writing

  $$\mathcal{G}_{rotate,i} = \left\{ g \;\middle|\; \begin{array}{l} g(\boldsymbol{y}) = \left(\boldsymbol{P} + \boldsymbol{G}_i\boldsymbol{R}\boldsymbol{G}_i^{\top}\right)\boldsymbol{y}, \\ \boldsymbol{P} = \boldsymbol{I} - \boldsymbol{G}_i\boldsymbol{G}_i^{\top},\ \boldsymbol{R} \in \mathcal{R} \end{array} \right\}$$

  to be the group defining orbits within the measurement space explained by rotation in $\boldsymbol{G}_i$.
  In Figure 1b, we observe that a rotation in the space of $\boldsymbol{G}_i$ does not affect its magnitude (as illustrated by the red radius of the resulting circle). Moreover, a rotation in the space of $\boldsymbol{G}_i$ does not affect the measurement in the nullspace of $\boldsymbol{G}_i^{\top}$ (as illustrated by the vertical red line in Figure 1b). Mathematically we write this pair of statistics as

  $$\hat{t}_{\rho,i}(\boldsymbol{y}) = \begin{bmatrix} t_{\rho,i}(\boldsymbol{y}) \\ \bar{t}_{\rho,i}(\boldsymbol{y}) \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{G}_0^{\top}\boldsymbol{y}\| \\ \left(\boldsymbol{I} - \boldsymbol{G}_0\boldsymbol{G}_0^{\top}\right)\boldsymbol{y} \end{bmatrix} \tag{8}$$

  where $t_{\rho_i}(\boldsymbol{y})$ represents the magnitude of the measurement in the space of $\boldsymbol{G}_i$ and $\bar{t}_{\rho_i}(\boldsymbol{y})$ denotes the projection of the measurement onto the nullspace of $\boldsymbol{G}_i^{\top}$. Recalling the maximally invariant statistic for $\boldsymbol{\mu}$ (*i.e.,* $t_{\mu}(\boldsymbol{y})$) we observe that $\bar{t}_{\rho,i}(\boldsymbol{y})$ is maximally invariant to a bias in the column space of $\boldsymbol{G}_i$. This fact will come in useful when discussing the PAIN test in the next subsection. As with the subspace bias, it is shown in [93] that the statistic, $\hat{t}_{\rho,i}(\boldsymbol{y})$, is maximally invariant to rotation in the column space of $\boldsymbol{G}_i$, namely $\mathcal{G}_{rotate,i}$.
- **Maximal Invariance to σ.** To normalize the unknown noise scaling factor, σ, requires us to multiply the measurement model in Equation 5 by σ^{-1}. This effect is illustrated in Figure 1c, where scaling $\boldsymbol{y}$ by σ^{-1} results in a vector in the same direction, but a different magnitude – *i.e.,* $\frac{\boldsymbol{y}}{\sigma}$. Thus, we say σ induces an unknown *scaling* of the measurement space. Formally, the set of all scalings induced by changing the nuisance parameter σ is captured by the group of transformations,

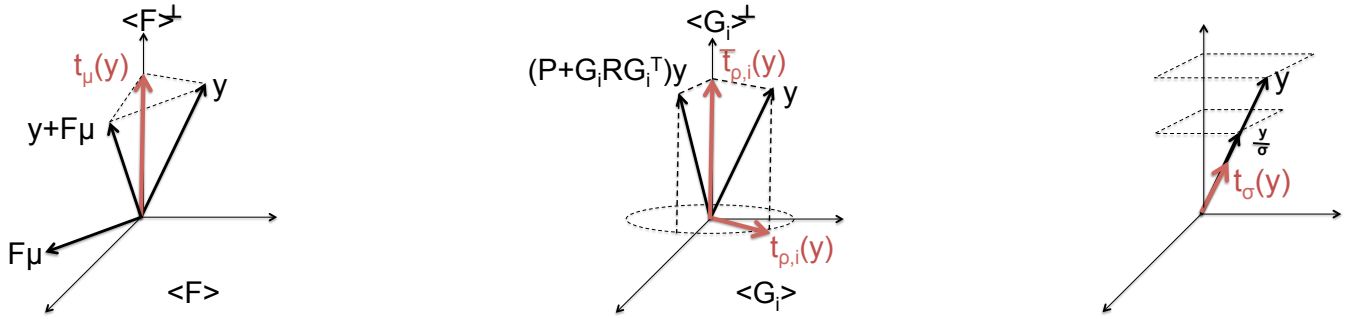
  $$\mathcal{G}_{scale} = \left\{ g \;\middle|\; g(\boldsymbol{y}) = \frac{1}{\sigma}\boldsymbol{y},\ \sigma \in \mathbb{R}_{\geq 0} \right\}$$

  which represents the group defining orbits within the measurement space that vary by scale. Since the direction of the measurement is unaffected by scaling, we consider the statistic

  $$t_{\sigma}(\boldsymbol{y}) = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \tag{9}$$

  where it is discussed in [93] that $t_{\sigma}(\boldsymbol{y})$ is, in fact, maximally invariant to $\mathcal{G}_{scale}$.

In the above, we discussed the effects of the nuisance parameters individually and developed maximally invariant statistics for each nuisance parameters. Namely, for the measurement model in Equation 5, $t_{\mu}(\boldsymbol{y})$ is maximally invariant to the subspace bias induced by $\boldsymbol{\mu} \in \Gamma_{\mu}$, $\hat{t}_{\rho,i}(\boldsymbol{y})$ provides invariance to the rotation induced by $\boldsymbol{\rho}_i \in \Gamma_{\rho,i}$, and $t_{\sigma}(\boldsymbol{y})$ is maximally invariant to the scaling needed to normalize the measurement model against $\sigma \in \Gamma_{\sigma}$. Observing that the



(a) Subspace Bias in $\langle \boldsymbol{F} \rangle$ (b) Rotation in $\langle \boldsymbol{G}_i \rangle$ (c) Scaling by $\frac{1}{\sigma}$

Fig. 1: Types of transformations included in the transformation group $\mathcal{G}$.

nuisance parameters can simultaneously affect the measurements, we need a statistic that is maximally invariant to their combined effect. Formally, their combined effect is defined as the composition of each individual transformation group, namely

$$\mathcal{G} = \left\{ g_\mu \circ g_{\rho,0} \circ g_{\rho,1} \circ g_\sigma \;\middle|\; \begin{array}{l} g_\mu \in \mathcal{G}_{bias},\ g_{\rho,0} \in \mathcal{G}_{rotate,0}, \\ g_\sigma \in \mathcal{G}_{scale},\ g_{\rho,1} \in \mathcal{G}_{rotate,1} \end{array} \right\}$$

Exploiting the orthogonality of $\boldsymbol{F}$, $\boldsymbol{G}_0$ and $\boldsymbol{G}_1$, we can write a maximally invariant statistic to $\mathcal{G}$ as

$$t(\boldsymbol{y}) = \begin{bmatrix} t_0(\boldsymbol{y}) \\ t_1(\boldsymbol{y}) \\ \bar{t}(\boldsymbol{y}) \end{bmatrix} = \begin{bmatrix} t_{\rho,0} \circ t_\sigma \circ \bar{t}_{\rho,1} \circ t_\mu \\ t_{\rho,1} \circ t_\sigma \circ \bar{t}_{\rho,0} \circ t_\mu \\ t_\sigma \circ \bar{t}_{\rho,0} \circ \bar{t}_{\rho,1} \circ t_\mu \end{bmatrix}$$

where, the individual maximally invariant statistics are (mostly) applied in the opposite order as the group of transformations such that $t(\boldsymbol{y})$ is maximally invariant to $\mathcal{G}$. The exception to this structure occurs when applying $\boldsymbol{t}_{\rho,0}$ and $\boldsymbol{t}_{\rho,1}$, which are applied last since they each reduce the N-dimensional measurement space to a scalar value. The maximally invariant statistic, $t(\boldsymbol{y})$, developed in this section is utilized in the following subsection to introduce the PAIN test.

C. Parameter-Invariant Hypothesis Testing

This section presents the PAIN test for evaluating the hypothesis model in Equation 6 invariant to the nuisance parameters in the measurement model in Equation 5. We begin by observing the statistic, $t(\boldsymbol{y})$, in the previous subsection is maximally invariant to the nuisance parameters in the measurement model in Equation 5. Therefore, any test for the hypothesis model based on the maximally invariant statistic, $t(\boldsymbol{y})$ will also be invariant to the nuisance parameters. Before proceeding in our discussion, we recall that $t(\boldsymbol{y})$ represents a concatenation of multiple statistics, namely $t_0(\boldsymbol{y})$, $t_1(\boldsymbol{y})$, and $\bar{t}(\boldsymbol{y})$. We observe that $t_0(\boldsymbol{y})$ is invariant to θ_1 since, $t_0(\boldsymbol{y})$ is composed of $\bar{t}_{\rho,1}(\boldsymbol{y})$ which is invariant to any bias in the column space of $\boldsymbol{G}_1$, including $\theta_1 \boldsymbol{\rho}_1$ in Equation 5. Similarly, $t_1(\boldsymbol{y})$ is invariant to θ_0 since $t_1(\boldsymbol{y})$ is composed of $\bar{t}_{\rho,0}(\boldsymbol{y})$ which is invariant to any bias in the column space of $\boldsymbol{G}_0$, including $\theta_0 \boldsymbol{\rho}_0$ in Equation 5. Lastly, we note that $\bar{t}(\boldsymbol{y})$ is invariant to all the test parameters (*i.e.,* both $\boldsymbol{\theta}_0$ and θ_1) since it contains both $\bar{t}_{\rho,1}(\boldsymbol{y})$ and $\bar{t}_{\rho,0}(\boldsymbol{y})$. This implies that $\bar{t}(\boldsymbol{y})$ does not discriminate between the hypotheses (since it is invariant to the test parameters). Thus, we can re-write the hypothesis model in Equation 6 based on $t_0(\boldsymbol{y})$ and $t_1(\boldsymbol{y})$ as

$$\mathcal{H}_0 : t_0(\boldsymbol{y}) > 0,\ t_1(\boldsymbol{y}) = 0 \quad \text{vs.} \quad \mathcal{H}_1 : t_0(\boldsymbol{y}) = 0,\ t_1(\boldsymbol{y}) > 0.$$

Another interpretation of this updated hypothesis testing problem is that the two statistics test the hypotheses independently. The first statistic, $t_0(\boldsymbol{y})$, should be (near) zero if the event hypothesis is true and will vary based on θ_0 if the null hypothesis is true. Conversely, the second statistic, $t_1(\boldsymbol{y})$ should be (near) zero if the null hypothesis is true and will vary based on θ_0 if the event hypothesis is true. These properties provide the opportunity to design multiple tests using the statistics t_0 and t_1. For example, a threshold test on $t_0(\boldsymbol{y})$ can (for some distributions) yield a constant false alarm rate (CFAR) detector. Similarly, a threshold test on $t_1(\boldsymbol{y})$ can (for some distributions) yield a detector that can achieve a minimum probability of detection.

When designing monitors for CPS, a number of errors can be introduced. The measurement model in Equation 5 may not be entirely accurate (or altogether wrong). This can result in large model errors that decrease both $t_0(\boldsymbol{y})$ and $t_1(\boldsymbol{y})$ resulting a low powered test (*i.e.,* unlikely to detect events). Similarly, due to the complexity of CPS, it is possible that neither the null or event hypothesis accurately describes the system. In this scenario, both statistics $t_0(\boldsymbol{y})$ and $t_1(\boldsymbol{y})$ will tend to increase – indicating that both hypotheses are rejected. Thus, a PAIN monitor utilizes both statistics and not only monitors for when to alarm, but also tests for low power and inaccurate models. Specifically, a PAIN monitor makes a decision according to Table I, where η_0 and η_1 denote the test thresholds.

TABLE I: Test Decision Space for Alarm System

	$t_0(\boldsymbol{y}) \geq \eta_0$	$t_0(\boldsymbol{y}) < \eta_0$
$t_1(\boldsymbol{y}) \geq \eta_1$	**Warning** (inaccurate model)	**No alarm**
$t_1(\boldsymbol{y}) < \eta_1$	**Alarm**	**Warning** (low power)

In Table I, the parameter invariant test makes a definitive decision (i.e., alarm or no alarm) when both statistic tests agree.

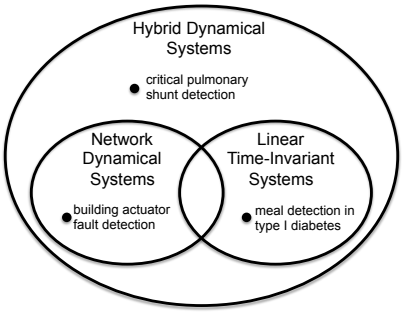


Fig. 2: Overview of PAIN monitoring in CPS applications.

When the tests do not agree, the parameter-invariant alarm generates one of two types of warnings. An indecision warning occurs when neither test rejects its assumption, indicating that there is not enough power in the test to disambiguate between the two hypotheses to the level of accuracy specified. A model inaccuracy warning occurs when both tests reject their assumptions, indicating neither model accurately describes the measurements. The benefit of this two-sided testing approach is that the event (or null) hypothesis won't be accepted/rejected just because there wasn't enough power or because the models were inaccurate. This is a similar concept to the *reject option* in the machine learning literature [94].

In special cases, such when $\boldsymbol{n}$ is an i.i.d. Gaussian, the PAIN monitor has some nice theoretical results. Specifically, the test thresholds η_0 and η_1 can be chosen such that

$$P\left[t_0(\boldsymbol{y}) \geq \eta_0\right] \leq \alpha \quad \text{and} \quad P\left[t_1(\boldsymbol{y}) \geq \eta_1\right] \leq \beta$$

where α and β are the probability of false alarm (type I error) and probability of miss (type II error), respectively. While achieving these theoretical results requires restrictive assumptions on the distribution of the noise in the measurement model that may not hold in CPS applications, practically, the PAIN monitors have shown near-constant false alarm rates in multiple CPS application domains.

D. Application of PAIN monitoring in CPS

The versatility of the PAIN monitoring approach is illustrated by its use in multiple CPS applications. An overview of different PAIN applications is provided in Figure 2 that spans systems whose dynamics are modeled using network models, linear time-invariant models, and hybrid models. In the following sections, we describe each modeling framework, discuss the use of each modeling approach in various CPS applications, and illustrate the development of a PAIN monitor for a representative application. Specifically, in the next section CPS with network dynamics are considered and a PAIN monitor is described for actuator fault detection and diagnosis in a smart building. Section V discusses modeling physical dynamics as linear time invariant (LTI) models and presents a PAIN meal detector for type I diabetic patients. Lastly, Section VI presents modeling physical dynamics as a (hybrid) mode-switching systems and demonstrates the concept using PAIN detector for critical pulmonary shunts in infants.

IV. Networked Dynamical Systems

Networked dynamical systems represent systems whose state dynamics that depend on the interaction of a certain state (or node) with its neighbors. Many CPS have real-world processes can be described (either fully or in part) by such dynamics, including power grid dynamics [95], [96], gas diffusion dynamics used in environmental monitoring [97], the wireless communication protocols [98], distributed multi-agent control applications (*e.g.,* robotic swarms) [99], thermal energy storage in smart buildings [100], and the spreading and mitigation of malware attacks [101], [102]. In these applications, many of which are safety-critical and/or security focused, fast and accurate detection of system faults is necessary. When undetected, system faults and/or attacks can lead to several flavors of detriments: from mild inconveniences in HVAC systems (poor air quality) to disruptive ripple effects in power systems (extended blackouts).

Traditional approaches to fault detection in networked dynamical systems utilize a model of the networked dynamics to construct monitors. A general model for the networked dynamics considers a system with J interconnected nodes for which there exists an underlying interconnection graph, $\mathcal{G}(\mathcal{V}, \mathcal{E})$, between the J nodes, where $\mathcal{V} = \{1, \ldots, J\}$ is the vertex set, with $i \in \mathcal{V}$ corresponding to node i, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set of the graph. The undirected edge $\{i, j\}$ is incident on vertices i and j if the dynamics of nodes i and j are interdependent. The neighborhood of node i is defined as $\mathcal{N}_i = \{j \in \mathcal{V} \mid \{i, j\} \in \mathcal{E}\}$. The networked dynamics can be described by a discrete-time system (representing potentially a discretization of a continuous-time networked system),

$$x_i(k+1) = \sum_{j=1}^{J} a_{ij}(k) x_i(k) + \sum_{\ell=1}^{L} b_{i\ell} u_\ell(k) \qquad (10)$$

where, L denotes the number of input signals, each written as $u_l(k)$, and b_{il} denotes the effect of the l-th input to the i-th node, and

$$a_{ij}(k) = \begin{cases} 1 - \sum_{j \in \mathcal{N}_i} \frac{w_{ij}}{m_i} & \text{if } i = j \\ \frac{w_{ij}}{m_i} & \text{if } j \in \mathcal{N}_i, \quad i \neq j \\ 0 & \text{otherwise} \end{cases}$$

describe the inter-node networked dynamics. In most physical systems, w_{ij} represents an *impedance* and thus, $w_{ij} = w_{ji}$. For example, the impedance introduce by a wall separating two thermal air masses is often the same regardless of the thermal flow direction [100]. Additionally, m_i represents the effective *mass* of a specific node and represents a node's resistance to change – a node with large mass is less effected by a neighboring node with a smaller mass. We note that in some applications, the dynamics include additional states beyond the networked dynamics presented in the above model. For example, the swing equation governing power grid dynamics includes a second order integrator at each node [103]. Thus, depending on the application, the networked dynamic model in Equation 10 may require additional states (with corresponding dynamics) in each node.

While fault detection algorithms can theoretically undoubtedly benefit from the knowledge of accurate models, practical constraints often these models are often parametrized by unknown environmental variables. For instance, in power transmission networks, the resistance of the transmission lines varies with temperature and icing conditions, while in building automation, the heat transfer between air-masses varies with the humidity, external temperature, and the opening or closing of windows and doors. Under these uncertain conditions, fault detection algorithms that cascade parameter-estimation with hypothesis testing (*e.g.,* the GLRT) can yield significantly varying results based on the accuracy of the estimated parameters, see, e.g., [104, Example 1, page 46]. However, testing approaches (*e.g.,* PAIN monitors) that are invariant to the underlying model parameters are designed to have the same baseline performance independent of the unknown parameters. Utilizing the invariance property allows PAIN monitors to detect faults in networked systems when network interactions are unknown (or untrusted).

To illustrate how the PAIN monitor can be applied to systems with networked dynamics, the remainder of this section overviews a recent application of PAIN monitors for fault detection and diagnostics in smart buildings.

A. Fault Detection and Diagnostics in Smart Buildings

Heating, ventilation and cooling (HVAC) are known to be the largest consumer of energy in buildings, accounting for 40% and 47% of the national energy consumption of the U.S. [105] and U.K. [106], respectively. Due to this high usage, energy-efficient HVAC system operations has been thrust to the forefront of world-wide research agendas. Recently, several researchers have studied how to improve the control of HVAC systems by deploying more embedded sensors to monitor temperature, humidity, and CO_2 levels [107], using information about occupant behavior [108], [109], [110], and improving the modeling and control approaches [111], [112], [113], [114], [115], [116]. In smart building applications, undetected HVAC sensor and actuator failures can result in poor temperature, poor air quality management, and a reduction in potential energy savings. However, HVAC Fault Detection and Diagnostic (FDD) schemes which result in unpredictable or erratic performance can deter building managers from investigating potential failures and utilizing the system altogether. For these reasons, technological development of FDD schemes tailored for HVAC systems is paramount and has received much research interest in the recent years [117], [118], [119], [120], [121].

The study of HVAC FDD systems has only been investigated since the late 1980s, with a particular interest in identifying low-cost, timely, and accurate methods for detecting actuator faults. A thorough review of approaches to HVAC actuator fault detection, diagnostics, and prognostics prior to 2006 is provided in [117], [118]. In general, approaches to HVAC actuator fault detection can be classified as either hardware-based or software-based solutions [117]. The hardware-based solutions introduce additional smart components strictly for the purposes of actuator fault detection and provide accurate detection capabilities; however, hardware solutions are far more expensive to both deploy and maintain than software-based approaches, and are much more difficult to reconfigure with the introduction of additional smart-actuator devices [118]. Moreover, the inclusion of additional hardware has the added drawback of further increasing the complexity of the HVAC system itself. Most software-based actuator FDD approaches are attractive in theory, but suffer from either a reliance on unknown (and difficult to learn) physical models or system-specific detector design specifications [119], [117], [118], [121].

To overcome the challenges in software-based actuator FDD, we consider a PAIN monitor designed to detect actuator failures with minimal knowledge of the building's physical dynamics. A common failure in HVAC systems occurs when the actuator "stick" and no longer changes its set point, despite controller requests. This type of actuator failure can occur in any position. For example, a valve can be stuck fully open, fully shut, or at any intermediate setting. The remainder of this subsection presents a PAIN monitor for detecting actuator faults invariant to known physical parameters of the building (*i.e.,* windows open/closed, wall thickness/material, etc.).

1) A PAIN monitor for HVAC actuator faults: To develop a PAIN monitor for detecting faults in HVAC actuators requires a physical model describing the dynamics. Building dynamics are commonly modeled using networked dynamics where each node in Equation 10 represents a *zone* that is assumed to have a uniform temperature. The interactions between zones is governed by the impedance of the plane separating the zones (*i.e.,* w_{ij} representing a wall) and the volume of air contained in the zone (*i.e.,* m_i). In buildings, the volume corresponding to a zone is constant (and easy to calculate); however, the interactions between zones can be difficult to model and can change drastically. For instance, two adjacent rooms separated by a door will interact very differently depending on whether the door is open or closed. Adding additional hardware to monitor whether the door is open or closed can be expensive (for large buildings), unsightly, and adds additional points of failure (and attack surfaces). Thus, we aim to design a PAIN monitor that can *detect actuator faults invariant to the unknown inter-zone interactions*.

To develop a measurement model, we assume that actuator ℓ receives a known actuation signal, $s_\ell(k)$. If the actuator is working correctly, corresponding to the null hypothesis, then $u_\ell(k) = s_\ell(k)$ in Equation 10. However, if the actuator is 'stuck', corresponding to the event hypothesis, then $u_\ell(k) = c_\ell$ for some arbitrary value of c_ℓ. For appropriate $\boldsymbol{S}_\ell(k)$ and $\boldsymbol{b}_\ell$, where $\boldsymbol{S}_\ell(k)$ are known actuation signals and $\boldsymbol{b}_\ell$ are corresponding gain parameters, the networked dynamics in Equation 10 can be written (assuming $\sigma\boldsymbol{n}$ represents the model error/noise) as

$$\begin{aligned} x(k+1) &= A(k)x(k) + \boldsymbol{S}_\ell(k)\boldsymbol{b}_\ell + B_\ell u_\ell(k) + \sigma n(k) \\ y(k) &= x(k), \end{aligned}$$

where $A(k) = [a_{ij}(k)]$, $B_\ell = [b_{i\ell}]$, $x(k) = [x_1(k), \ldots, x_J(k)]^\top$, $u(k) = [u_1, \ldots, u_L]^\top$, and $y(k)$ denotes the measurements. To design a PAIN monitor, we observe that the networked dynamics have a natural invariant, namely,

given $\boldsymbol{m}^\top = [m_1, \ldots, m_J]$ then $\boldsymbol{m}^\top A(k) = \mathbf{1}^\top$ for all time k – which is directly observable from Equation 10. Thus, utilizing the vector of air masses, $\boldsymbol{m}^\top$ and its corresponding nullspace, $\boldsymbol{P}_m = \boldsymbol{I} - \frac{\boldsymbol{m}\boldsymbol{m}^\top}{\boldsymbol{m}^\top \boldsymbol{m}}$, we can write the time-concatenated measurement model, in the PAIN general form of Equation 5, as

$$\boldsymbol{y} = \begin{bmatrix} y(1) - \boldsymbol{m}\frac{\mathbf{1}^\top y(0)}{\boldsymbol{m}^\top \boldsymbol{m}} \\ \vdots \\ y(k) - \boldsymbol{m}\frac{\mathbf{1}^\top y(k-1)}{\boldsymbol{m}^\top \boldsymbol{m}} \end{bmatrix}, \; \boldsymbol{G}_1 = \begin{bmatrix} \boldsymbol{I} s_\ell(0) \\ \vdots \\ \boldsymbol{I} s_\ell(k-1) \end{bmatrix}$$

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{P}_m & & \boldsymbol{S}_\ell(0) \\ & \ddots & \vdots \\ & \boldsymbol{P}_m & \boldsymbol{S}_\ell(k-1) \end{bmatrix}, \; \boldsymbol{G}_0 = \begin{bmatrix} \boldsymbol{I} \\ \vdots \\ \boldsymbol{I} \end{bmatrix}$$

with nuisance parameters, σ and

$$\boldsymbol{\mu} = \begin{bmatrix} A(0)x(0) \\ \vdots \\ A(k-1)x(k-1) \\ \boldsymbol{b}_\ell \end{bmatrix}, \quad \boldsymbol{\rho}_0 = \frac{B_\ell c_\ell}{\|B_\ell c_\ell\|}, \quad \boldsymbol{\rho}_1 = \frac{B_\ell}{\|B_\ell\|},$$

and test parameters (under each hypothesis)

$$\mathcal{H}_0 : \theta_0 = 0, \; \theta_1 = \|B_\ell\| \quad \text{vs.} \quad \mathcal{H}_1 : \theta_0 = \|B_\ell c_\ell\|, \; \theta_1 = 0$$

We note that $\boldsymbol{y}$, $\boldsymbol{F}$, $\boldsymbol{G}_0$, and $\boldsymbol{G}_1$ are all known matrices and that the test results in a measurement model and hypothesis model that satisfy the general PAIN models in Equation 5 and Equation 6, respectively. Thus, a PAIN monitor to detect actuator faults invariant to the unknown inter-zone interactions can be achieved by following the procedure outlined in Section III.

2) Experimental Evaluation and Discussion: The evaluation of the actuator fault detector was performed in the KTH HVAC testbed [122]. To evaluate the parameter-invariant detector performance, multiple experiments were performed utilizing room A225 in Figs. 3 and 4 as the test room. In this evaluation, we aim to detect whether there is an actuator failure in the air conditioning system, namely whether the fresh air vent (actuator ST901 in Fig. 4) is stuck in a given position. To emulate an actuator failure, we hold the vent in a constant position and assume that the actuator signal cycles between fully open and fully closed (as a diagnostic signal) on 15 minute intervals for 3 hours. We evaluate the detector performance in terms of missed alarm rate and false alarm rate, as denoted by P_{MA} and P_{FA} in Table II, in scenarios when the window in room A225 is open and closed.

For comparison, the PAIN monitor was compared with two model-based approaches utilizing a likelihood ratio test, as described in Section II. The first model-based approach is designed assuming the windows in room A225 are open, while the second model-based approach assumes the windows are closed. Both of these models are described in [86]. For evaluation purposes, we aim to minimize the probability of false alarm while ensuring a 10% probability of miss. For the model-based monitors, we select the likelihood test thresholds such that a probability of miss is 10% is achieved when the model is correct. The motivation behind designing a detector that achieves a constant probability of miss (rather than a constant probability of false alarm) is discussed in [91]. The results of the experimental evaluation are provided in Table II.

Fig. 3: KTH HVAC testbed at the second floor of the Q-building at KTH. Each of the five rooms considered contain sensors and actuators used for HVAC control. Additional sensors are located in the corridor and outside of the building.

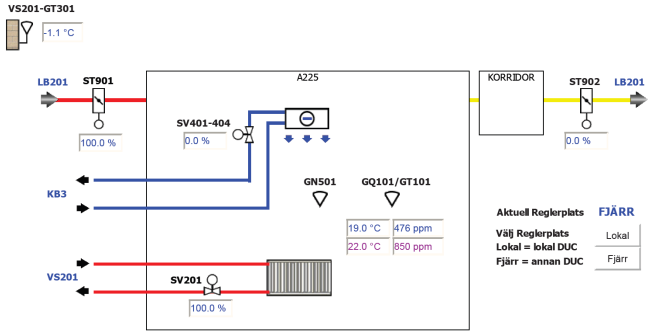


Fig. 4: The HVAC system components in room A225, the Automatic Control experimental lab. Various sensors and actuators are available allowing for the control of ventilation and heating.

TABLE II: Building Actuator FDD Results

Approach	Window State	P_{MA} (%)	P_{FA} (%)
PAIN	open	10.1	5.1
	closed	9.8	0.2
	varying	9.9	3.1
Model-based (window open)	open	10.0	4.0
	closed	18.2	0.1
	varying	14.1	2.9
Model-based (window closed)	open	15.1	3.1
	closed	10.0	0.2
	varying	12.1	3.0

In Table II, we observe that the PAIN monitor achieves nearly identical missed alarm rates regardless of whether the window is open, closed, or switching between open and closed (*i.e.,* varying). In contrast, both model-based approaches have varying missed alarm rates depending on the window state. These results illustrate that while it may be possible to design a model-based monitor that (slightly) outperforms the PAIN monitor when the model is correct, the performance of the model-based approach can vary significantly if the model is inaccurate. Moreover, these results are achieved without adapt-

ing the PAIN monitor to the specific testing scenario; thus, illustrating the robustness of the PAIN monitoring approach to changes in the underlying networked dynamical system.

V. Linear Time Invariant Systems

Linear time invariant (LTI) systems represent systems where the relationship between the inputs and the outputs is a linear map that does not vary with time. Modeling the dynamics of a CPS as an LTI system is common in many monitoring applications, since LTI systems serve as good approximations for most real-world systems, especially over *small enough* time windows [123]. LTI models arise in include, compartment modeling in medicine and biology [124], [125], robotic dynamics [126], [127], electric motor models [128], and advection diffusion processes [129]. In addition, LTI systems represent a generalization for many of the networked dynamical systems in the previous subsection [97], [95], [99].

LTI models for CPS come in many forms including continuous/discrete-time state space models [130], autoregressive moving average models with exogenous inputs (ARMAX) [131], and difference equations [132]. In general, LTI models of the physical world consist of a set of differential (or difference) equations in which the state variables have physical meaning (*e.g.,* vehicle velocity, room temperature, amount of oxygen in blood) and the equations represent the lumped interactions/interconnections between different state variables,

$$\dot{x} = Ax + Bu, \quad y = Cx \tag{11}$$

where, x represents the system *state*, u denotes the *inputs*, y corresponds to the *outputs*, and A, B, and C capture the linear dynamics. Many variations of the LTI model exist, including stochastic versions with process and measurement noise. A benefit of state-space models is that they can be designed to arbitrary accuracy through a natural grouping (or ungrouping) of physical effects. For example, air mass interaction through walls in smart buildings can be captured using a simple resistor-capacitor model or by modeling the different materials that make up the wall. Similarly, in the glucose-insulin physiological system one can model the cumulative effect of carbohydrate ingestion as a single first-order differential equation representing collectively the stomach, intestine, and blood compartments as the *glucose system*, or one can model each of the three physical compartments individually as interactive systems.

The risk in using a linear model to describe real world dynamics is that it may not be accurate. While it is likely that *some* linear model can sufficiently capture the dynamics of a CPS, it is unlikely that the same linear model is accurate for all dynamics. Thus, the model error introduced by using LTI systems (with estimated or assumed parameters) to design CPS monitors can lead to poor monitor performance in safety-critical scenarios. Moreover, while accurate model parameters are useful for distinguishing between the null and event hypotheses, they are often unnecessary. For instance, to test a vehicle's braking system, an autonomous driver may trigger the brakes and observe a sudden decrease in speed. Knowledge or estimation of the brake pad friction coefficient is unnecessary to identify that triggering the brakes resulted in a sudden decrease in speed (*i.e.,* the braking system is working). While this example is somewhat trivial, it does illustrate that knowledge of system parameters is (sometimes) unnecessary to test a hypothesis.

To illustrate how the PAIN monitor can be applied to systems with LTI models, the remainder of this section overviews a recent application of PAIN monitors for meal detection in type I diabetic patients.

A. Meal Detection in Type 1 Diabetic Patients

Type 1 diabetics depend on daily insulin infusion or injection to keep their glucose level within an acceptable range. Too much insulin can cause life-threatening hypoglycemia (low glucose levels) and too little insulin can cause nerve-damaging hyperglycemia (high glucose levels) [133]. Ingested carbohydrates from meals cause a major disturbance to blood glucose levels. Thus, all diabetics must carefully titrate insulin doses for every meal so that post-meal hyperglycemia is effectively controlled, while avoiding administering too much insulin and risking hypoglycemia. Currently, many type 1 diabetics use continuous glucose monitors (CGMs) and wearable insulin pumps. These devices allow the user to manually input the time and estimated carb count of each ingested meal into the device, which then calculates a suggested insulin dose. Unfortunately, self-reported meal information is known to be inherently unreliable [134].

Several meal detection strategies exist in the literature. Dassau et al. proposes voting-based meal detector that tracks the glucose rate-of-change (RoC) estimated by different methods (including Kalman filtering) and announces a meal when three out of the four RoC estimates cross their thresholds [134]. Using similar Kalman filtering techniques, Lee et al. develop a meal detector that announces a meal based on thresholding glucose RoC and estimates the meal size by feeding the filtered glucose RoCs into a finite response filter [135]. Harvey et al. recently propose a meal detection algorithm that announces meals based on a two-stage CGM filtering process and a RoC criteria [136]. Cameron et al. use a simple glucose model to match estimated glucose trajectories assuming no meals to CGM residuals as a means to detect meals [137]. Turksoy et al. simultaneously aim to estimate physiological variables and model parameters to provide accurate meal detection and estimation [138]. In terms of performance, the aforementioned detectors require balancing a tradeoff between false alarm rate, detection rate, and detection delay –which can vary significantly due to non-meal factors, such as exercise [139], stress [140], and depletion of insulin-on-board [141]. Additionally, the shape-matching meal detector [137] requires personalization to a subject's time-varying physiology (*e.g.,* insulin sensitivity) and tends to have a long average detection delay [142], while the adaptive technique [138] provides no guarantee that the physiological parameter estimates converge to their true value.

To improve patient safety, more dependable meal detection methods are necessary; ideally, meal events could be detected

directly from physiologic data, freeing the patient from having to manually input meals. In the following sections we formulate the meal detection problem in the PAIN general form and discuss its performance with respect to existing meal detectors in the literature.

1) A PAIN monitor for Meal Detection: To develop a PAIN monitor for meal-detection in type I diabetes requires a model of insulin-glucose physiology. There are multiple ways to model insulin-glucose physiology ranging from low-fidelity *minimal* models [125] to a high-fidelity *maximal* models such as the FDA-accepted UVa/Padova Type 1 Diabetes Mellitus Metabolic Simulator (T1DMS) [143]. To enable the design of a PAIN monitor, a minimal model denoted by a fifth-order order LTI system that describes the glucose-insulin kinetics [144] under carbohydrate ingestion [145] and subcutaneous insulin injections [146], [147], written as

$$\underbrace{\begin{bmatrix}\dot{G}(t)\\ \dot{m}(t)\\ \dot{g}(t)\\ \dot{I}(t)\\ \dot{x}(t)\end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} \approx \underbrace{\begin{bmatrix}p_1 & 1 & & p_2 & \\ & \frac{-1}{t_G} & \frac{1}{t_G} & & \\ & & \frac{-1}{t_G} & & \\ & & & -k_e & \frac{k_a}{V_d}\\ & & & & -k_a\end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix}G(t)\\ m(t)\\ g(t)\\ I(t)\\ x(t)\end{bmatrix}}_{\boldsymbol{x}(t)} + \underbrace{\begin{bmatrix}p_3\\ \\ \frac{A_G}{t_G}d(t)\\ \\ u(t)\end{bmatrix}}_{\boldsymbol{u}(t).} \tag{12}$$

In Equation 12, $\boldsymbol{x}$ represents the physiological state vector, where $G(t)$ represents the plasma glucose, $x(t)$ and $I(t)$ is the insulin in the subcutaneous compartment and plasma, $g(t)$ denotes glucose in the lumped digestive track compartment, and $m(t)$ represents the rate of plasma glucose appearance due to meals. Injected insulin is represented as $u(t)$ while meal carbohydrate inputs correspond to $d(t)$. The parameters p_1, p_2, and p_3 are unspecified lumped physiological parameters, k_a and k_e are rate parameters, V_d is the insulin volume, and A_G and t_G represent the unspecified carbohydrate bioavailability and maximum glucose rate of appearance, respectively – all assumed to be unknown. The details of the model are provided in [85], [84].

By modeling the CGM measurement as $y = G$, the application of standard discretization and z-transform techniques, results in a LTI discrete-time model (specifically an ARMAX model) which approximates the measurement at time step k using a weighted sum of past measurements and inputs, namely

$$y(k) = \sum_{j=1}^{5} a_j y(k-j) + b_{3,j}\frac{A_G}{t_G}d(k-j) + b_{5,j}u(k-j) + c + \sigma n(k).$$

In the above equation, $\sigma n(k)$ denotes the model error, while $\boldsymbol{a} = [a_1, \ldots, a_5]^\top$, $\boldsymbol{b}_n = [b_{n,1}, \ldots b_{n,5}]^\top$, and $c = p_3 \sum_{j=1}^{5} b_{1,j}$ represent unknown patient-specific lumped physiological coefficients that correspond to the z-domain transfer function coefficients relate to the LTI model in Equation 12.

By hypothesizing that meals can occur in at most one of two windows, $\mathcal{S}_0, \mathcal{S}_1 \in \{0, \ldots, k\}$ such that $\mathcal{S}_0 \cap \mathcal{S}_1 = \emptyset$. Writing $\boldsymbol{e}_k$ to be a unit vector with a single unit entry in the k-the element, and using $\boldsymbol{e}_k$ to write $\boldsymbol{E}_k = \begin{bmatrix} \boldsymbol{e}_{k+1} & \ldots & \boldsymbol{e}_{k+5} \end{bmatrix}$. We write $\bar{\boldsymbol{b}}_{m,n}(k) = \frac{A_G}{t_G} d_{k-\mathcal{S}_m[n]} \boldsymbol{b}_3$

We can write the time-concatenated measurement model, in the PAIN general form of Equation 5, as $\boldsymbol{y} = [y(5), \ldots y(k)]^\top$,

$$\boldsymbol{F} = \begin{bmatrix} y(4) & \ldots & y(0) & u(4) & \ldots & u(0) & 1 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ y(k-1) & \ldots & y(k-5) & u(k-1) & \ldots & u(k-5) & 1 \end{bmatrix},$$

$$\boldsymbol{G}_0 = \begin{bmatrix} \boldsymbol{E}_{\mathcal{S}_0[1]} & \ldots & \boldsymbol{E}_{\mathcal{S}_1[|\mathcal{S}_0|]} \end{bmatrix},$$

$$\boldsymbol{G}_1 = \begin{bmatrix} \boldsymbol{E}_{\mathcal{S}_1[1]} & \ldots & \boldsymbol{E}_{\mathcal{S}_1[|\mathcal{S}_1|]} \end{bmatrix}$$

with nuisance parameters, σ, $\boldsymbol{\mu} = [\boldsymbol{a}^\top, \boldsymbol{b}_5^\top, c]^\top$, and

$$\boldsymbol{\rho}_0 = \frac{\begin{bmatrix} \bar{\boldsymbol{b}}_{0,1}(k) \\ \vdots \\ \bar{\boldsymbol{b}}_{0,|\mathcal{S}_0|}(k) \end{bmatrix}}{\left\| \begin{bmatrix} \bar{\boldsymbol{b}}_{0,1}(k) \\ \vdots \\ \bar{\boldsymbol{b}}_{0,|\mathcal{S}_0|}(k) \end{bmatrix} \right\|}, \quad \boldsymbol{\rho}_1 = \frac{\begin{bmatrix} \bar{\boldsymbol{b}}_{1,1}(k) \\ \vdots \\ \bar{\boldsymbol{b}}_{1,|\mathcal{S}_1|}(k) \end{bmatrix}}{\left\| \begin{bmatrix} \bar{\boldsymbol{b}}_{1,1}(k) \\ \vdots \\ \bar{\boldsymbol{b}}_{1,|\mathcal{S}_1|}(k) \end{bmatrix} \right\|},$$

and test parameters (under each hypothesis)

$$\mathcal{H}_0 : \theta_0 = 0 \quad , \quad \theta_1 = \left\| \begin{bmatrix} \bar{\boldsymbol{b}}_{1,1}(k) \\ \vdots \\ \bar{\boldsymbol{b}}_{1,|\mathcal{S}_1|}(k) \end{bmatrix} \right\|$$

$$\mathcal{H}_1 : \theta_0 = \left\| \begin{bmatrix} \bar{\boldsymbol{b}}_{0,1}(k) \\ \vdots \\ \bar{\boldsymbol{b}}_{0,|\mathcal{S}_0|}(k) \end{bmatrix} \right\|, \quad \theta_1 = 0$$

We note that $\boldsymbol{y}$, $\boldsymbol{F}$, $\boldsymbol{G}_0$, and $\boldsymbol{G}_1$ are all known matrices and that the test results in a measurement model and hypothesis model that satisfy the general PAIN models in Equation 5 and Equation 6, respectively. Thus, a PAIN monitor to detect meals can be achieved by following the procedure outlined in Section III.

2) Experimental Evaluation and Discussion: The PAIN-based detector is evaluated against three existing meal-detection algorithms: the Dassau (et al.) detector [134], Harvey (et al.) detector [136], and Lee (and Bequette) detector [135]. A complete description of the clinical data set and evaluation criteria are provided in [85]. In the evaluation, each detector is tuned to an operating point of approximately 2 false alarms per day such that the detection rates in Table III result.

TABLE III: Operating Points of the Four Detectors on the Clinical Data set Used for Analysis

Approach	Detection rate (%)	False alarms per day
PAIN [85]	86.9	2.01
Dassau [134]	74.1	1.99
Lee [135]	73.4	1.99
Harvey [136]	79.4	1.97

The experimental evaluation shows that the PAIN-based detector significantly improves the detection performance when compared with the other three detectors. Figure 5 compares the performance variability, in terms of false alarm and detection rates, of each meal detector on different patients in the

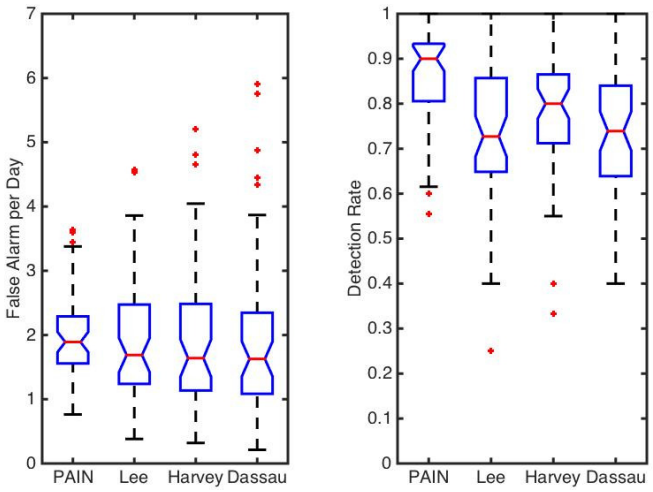


Fig. 5: Distributions of the four detectors false alarm and detection rates on different patients in the clinical data set used for analysis. The dots corresponds to outliers in patient performance for each detector.

data set. The box-plots represent the performance of each detector on the majority of the patients, while the dots indicate corresponding outliers. These results provide a measure of the consistency of detection performance at the individual level, that is, whether a detector can perform particularly bad on any patient. In Figure 5, the PAIN monitor detects at least 55% of all reported meals and never has a false alarm rate greater than 3.7 false alarms per day. In sharp contrast to the PAIN-based detector, all other three detectors miss significantly more meals (both on average and worst case), and have false alarm rates with higher variances and higher worst-case values. For example, compared with the Harvey detector, the PAIN-based detector reduces the number of missed detections by 36% without increasing the false alarm rate in the experimental evaluation. The performance distribution over the patients validates the unique strength of the PAIN-based detector: it is designed to be invariant to differences in patients physiological parameters and thereby achieves low variance detection performance across a real population.

VI. Hybrid Dynamical Systems

Hybrid dynamical systems represent a class of systems that exhibit both continuous and discrete dynamic behavior. The continuous behavior is often described by differential equations (potentially LTI), while the discrete behavior is captured by a corresponding state machine. A hybrid system is a general formulation that encompasses many other classes of systems including networked systems and LTI systems. Often, a discrete state in a hybrid system will correspond to an *unsafe* or *error* condition. Due to its generality as a model, hybrid systems monitors have been developed in many CPS applications [127], [126], [148], [149], [150] and can capture both the LTI and networked dynamics models of the previous sections.

In general, the state of a hybrid system is defined by the values of the continuous variables and a discrete mode. The state changes either continuously, according to a flow condition, or discretely according to a control graph. Continuous flow is permitted as long as so-called invariants hold, while discrete transitions can occur as soon as given jump conditions are satisfied. Discrete transitions may be associated with events. Thus, the model of a hybrid system can be written as

$$\mathcal{S} = \langle \mathcal{X}, \mathcal{Q}, \mathcal{X}_{init}, \mathcal{X}_{inv}, \mathcal{F}(\mathcal{P}), T \rangle ,$$

where $\mathcal{X}$ represents the continuous states, $\mathcal{Q}$ denotes the discrete modes, $\mathcal{X}_{init} \in \mathcal{R}_{\mathcal{X}}$ specifies the initial condition space, $\mathcal{F}(\mathcal{P})$ captures the flows parameterized by a vector $\mathcal{P} \in \mathcal{R}_P$, $\mathcal{X}_{inv}$ identifies invariants mapping modes to flows, and T relates the transitions between modes. An output $y = \phi(t; \mathcal{X}_{init})$ denotes the measurement, with $\phi(t, \mathcal{X}_{init})$ describing the measurement at time t, having evolved from initial condition $\mathcal{X}_{init}$.

The risk in using discrete mode transitions as monitors is that the underlying continuous models may be inaccurate. As in the networked and LTI systems, the model error introduced by the underlying continuous dynamic model (*e.g.,* LTI model), can lead to poor monitor performance in safety-critical scenarios. Moreover, while accurate model parameters are necessary to accurately describe the continuous dynamics, if the discrete dynamics define the null and event hypotheses, then they are not necessary for monitoring. For instance, a vehicle transmission has a specific system response relating engine torque to vehicle velocity when in each unique gear, but varies based upon road grade and surface conditions (*e.g.,* icy or dry). When the vehicle's speed drops below a certain value, the automobile should shift to a lower gear to prevent undue stress on the vehicle transmission. Monitoring a change in the transmission represents monitoring for a mode switch in the dynamics between vehcile speed and engine torque (or RPMs). As in the LTI systems, we observe that switching between two modes can be monitored without knowing the specific dynamics of each mode, and only monitoring whether a change in the dynamics occurs.

To illustrate how the PAIN monitor can be applied to hybrid systems, the remainder of this section overviews a recent application of PAIN monitors for detecting critical pulmonary shunts in infants.

A. Critical Pulmonary Shunt Detection in Infants

During surgery, blood O_2 content is perhaps the most closely monitored physiological variable, as values that are too low can lead to organ failure (*e.g.,* brain damage), and values that are too high can cause atelectasis (*i.e.,* collapse of the lungs). Pulmonary shunts, which occur when a patient is breathing with only one lung, can cause dangerous drops in O_2 levels. Shunts can be caused by a physical disorder, such as pulmonary edema, or may occur when the patient is under mechanical ventilation (e.g., when the perioperative lung is not ventilated). Infants are especially vulnerable to shunts because they have underdeveloped lungs. In these patients, critical shunts are common (e.g., caused by small shifts in the

endotraceal tube), thereby often leading to dangerously low levels of O_2 content.

Despite its importance, blood O_2 content is challenging to monitor, as it cannot currently be measured non-invasively or in real time. Instead, clinicians must monitor proxy variables. One popular proxy is the hemoglobin oxygen saturation in the peripheral capillaries, denoted by S_pO_2. While it is a good non-invasive measure of the O_2 content in the location at which it is measured (*e.g.,* a fingertip), S_pO_2 is a delayed measure of the O_2 content in other parts of the body (*e.g.,* the arteries), as blood takes time to circulate.

To improve patient safety, earlier detection of critical pulmonary shunts is necessary. In the following, we formulate the critical shunt detection problem in the PAIN general form and discuss its performance with respect to monitoring the S_pO_2.

1) A PAIN monitor for Critical Pulmonary Shunts: To develop a PAIN monitor for detecting critical pulmonary shunts requires models of the blood-oxygen dynamics in the presence and absence of a pulmonary shunt. Since no established models exist, we elect to model the dynamics under each scenario using *compartment models*. In general, compartment models (discussed in [55]) tend to balance the trade-off between model accuracy and usefulness for monitoring. Compartment models of the physical world consist of a (potentially switching) set of differential or difference equations in which the state variables have physical meaning (*e.g.,* amount of oxygen in blood) and the equations represent the lumped interactions/interconnections between different state variables in compartments. A simplified schematic model of the gas partial pressures in the circulatory and respiratory systems is illustrated in Figure 6, where the details of the model are contained in our previous work [81], [82].

Following [82], we can construct a state-space model of the system (not shown in this paper due to space constraints) based on the schematic model in Figure 6, and after some algebraic manipulation, the time-series measurement can be modeled as

$$\boldsymbol{y} = \theta_0 \boldsymbol{G}_0 \rho_0 + \theta_1 \boldsymbol{G}_1 \rho_1 + \sigma \boldsymbol{n}$$

where $\boldsymbol{G}_0$ corresponds to the model in the *absence of a shunt* (*i.e.,* null hypothesis) as illustrated in Figure 6aand $\boldsymbol{G}_1$ corresponds to the model in the *presence of a shunt* (*i.e.,* event hypothesis) as illustrated in Figure 6b. This model has corresponding nuisance parameters, σ, and

$$\boldsymbol{\rho}_0 = \frac{\begin{bmatrix} \alpha \\ \mu\alpha \end{bmatrix}}{\left\| \begin{bmatrix} \alpha \\ \mu\alpha \end{bmatrix} \right\|}, \quad \boldsymbol{\rho}_1 = \frac{\begin{bmatrix} \alpha \\ \mu\alpha \end{bmatrix}}{\left\| \begin{bmatrix} \alpha \\ \mu\alpha \end{bmatrix} \right\|}$$

with test parameters (under each hypothesis),

$$\mathcal{H}_0 : \theta_0 = 0, \theta_1 = \left\| \begin{bmatrix} \alpha \\ \mu\alpha \end{bmatrix} \right\|$$
$$\mathcal{H}_1 : \theta_0 = \left\| \begin{bmatrix} \alpha \\ \mu\alpha \end{bmatrix} \right\|, \theta_1 = 0.$$

Here we observe that the test parameters serve to indicate which model $\boldsymbol{G}_0$ or $\boldsymbol{G}_1$ explains the measurements under each hypothesis. For example, under the null hypothesis, θ_1 is always zero and θ_0 is non-zero, which indicates that $\boldsymbol{G}_1\boldsymbol{\rho}_1$ (*i.e.,* the model in the presence of a shunt) does not affect the measurement.

2) Experimental Evaluation and Discussion: To evaluate the performance of the PAIN monitor, real patient data from lung lobectomy surgeries performed on infants is utilized. A complete description of the clinical data set and evaluation criteria are provided in [82]. In the evaluation, each detector is tuned to an operating point of approximately 1 false alarm per hour such that the detection rates in Table IV result. For comparison, the PAIN monitor is compared with a standard change detection technique, namely, the cumulative sum control chart (CUSUM) detector. In order to develop the CUSUM detector, a maximum likelihood estimate of the model parameters for each patient were estimated (similar to the GLRT approach described in Section II) and then the detector algorithm from [151][Ch. 8.10] was employed.

TABLE IV: Operating Points of the PAIN Monitor and CUSUM Detector on the Clinical Data set Used for Analysis

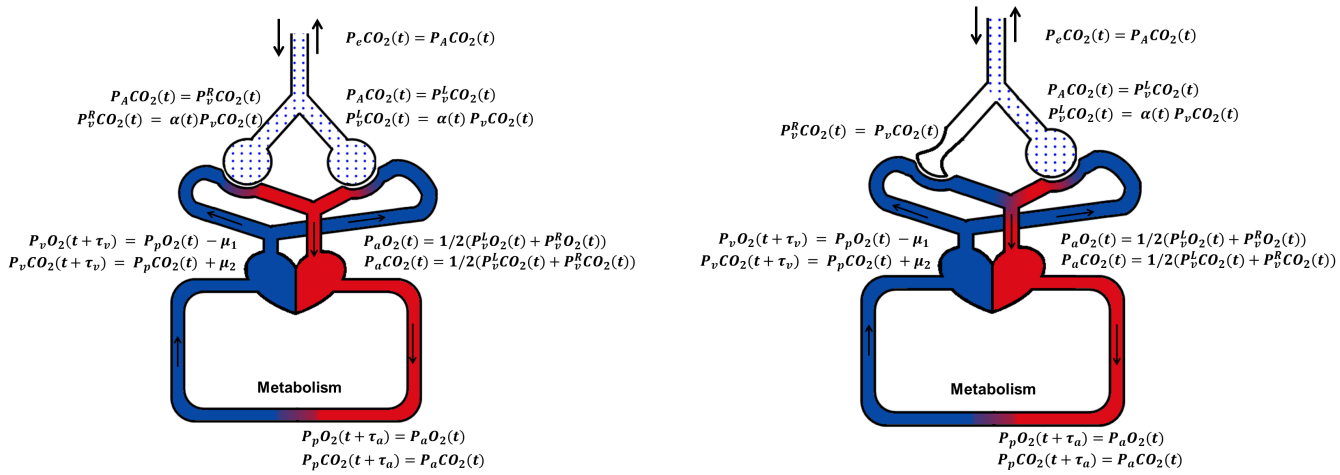
Approach	Detection rate (%)	False alarms per hour
PAIN [82]	88	1.0
CUSUM [151]	18	1.0

As discussed in detail in [82], the PAIN monitor is able to accurately predict 88 % of critical pulmonary shunts about 90 seconds before a drop in the S_pO_2 occurs while only having 1 false alarm per hour. Moreover, as can be seen in Table IV, the PAIN monitor greatly outperforms the CUSUM detector. There are two main reasons for this difference. First, the model in Figure 6 captures general trends but is a poor predictor of the future that makes it unsuitable for model-predictive and estimation-based techniques. Second, it is difficult to obtain good parameter estimates in the presence of noisy and missing measurements.

To further illustrate the performance disparity, Figure 5 compares the performance variability, in terms of false alarm rate, of each shunt detector on different patients in the data set. The box-plots represent the performance of each detector on the majority of the patients, while the red dots indicate corresponding outliers. These results provide a measure of the consistency of detection performance at the individual level, that is, whether a detector can perform particularly bad on any patient. In Figure 7, the PAIN monitor never has a false alarm rate greater than 3.5 false alarms per hour. In sharp contrast to the PAIN-based detector, the CUSUM detector misses significantly more critical events and has a false alarm rate with higher variances and higher worst-case values. The performance distribution over the patients reaffirms the previous results that illustrate the unique strength of the PAIN-based detector: it is designed to be invariant to differences in patients physiological parameters and thereby achieves low variance detection performance across a real population.

VII. Extensions to CPS Security

As described in Section I, we have witnessed a significant increase in the number of security related incidents in cyber-



(a) System without a shunt. (b) System with a shunt.

Fig. 6: Model of the respiratory and cardiovascular partial pressures with and without a shunt.

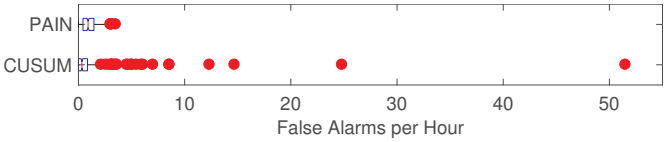


Fig. 7: Distributions of the PAIN and CUSUM detectors false alarm rates on different patients in the clinical data set used for analysis. The dots corresponds to outliers in patient performance for each detector.

physical systems in recent years. These incidents have seriously raised security awareness in CPS, where there is a tight coupling of computation and communication substrates with sensing and actuation components. However, the complexity and heterogeneity of this next generation of safety-critical, networked and embedded control systems have challenged the existing design methods in which security is usually consider as an afterthought.

Therefore, many researchers have begun to consider attacks against the control system as the primary function of CPS, where the attacker can (1) take over a sensor and supply wrong or untimely sensor readings, or (2) disrupt actuation. These attacks manifest themselves to the controller as malicious interference signals, and the defenses against them have to be introduced in the control design phase. Specifically, resilience against these attacks is built into the low-level and supervisory control algorithms. This approach have attracted a lot of attention, with several efforts focused on the use of control-level and monitoring techniques, which exploit a model of the normal system behavior, for attack-detection and identification in CPS (*e.g.,* [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52]). A common feature to the design of these approaches is their reliance on an model of normal system (possibly with bounded or stochastic noise). However, this may not be feasible in all applications and scenarios (as illustrated by the examples in Section IV-VI).

The following subsection introduces preliminary results on attack detection using PAIN monitors. The final subsection overviews challenges associated with attack detection using PAIN monitors and provides insight to potential research directions for overcoming these challenges.

A. Attack Detection using PAIN Monitors

The development of monitoring techniques that can accurately detect attacks with minimal reliance on underlying dynamical models can improve the penetration of CPS security techniques. Here, the ability of PAIN monitors to provide constant monitoring accuracy invariant to (possibly maliciously altered) parameters, can prove useful. This utility has already been demonstrated in [152], where a PAIN monitor has been developed for detecting attacks in systems with redundant sensors. In CPS, utilizing multiple correlated sensors is a well established method of providing better estimates of control variables and model parameters; however, when attacked, a sensor can be used as a means to maliciously alter a CPS. Thus, to ensure safe performance requires securing the sensory data. A PAIN monitor can be utilized to identify inconsistent sensors by observing that for a known $\boldsymbol{C}$, redundant sensor measurements can be modeled as

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{e} + \boldsymbol{n}$$

where, $\boldsymbol{x}$ denotes an unknown system state that affects the measurements, $\boldsymbol{n}$ denotes a random noise (note that we do not include σ as in the previous models), and $\boldsymbol{e}$ represents a potential attack vector. We say that a measurement in $\boldsymbol{y}$ is under attack if the corresponding element of $\boldsymbol{e}$ is non-zero. Thus, an attack detector aims to distinguish between when *no* measurement is under attack (*i.e.,* $\|\boldsymbol{e}\| = 0$) versus when *some* measurement is under attack (*i.e.,* $\|\boldsymbol{e}\| > 0$). A PAIN monitor can be designed for this problem by writing

$$\boldsymbol{F} = \boldsymbol{C}, \quad \boldsymbol{G}_0 = 0, \quad \boldsymbol{G}_1 = \boldsymbol{I}$$

and defining $\mu = \boldsymbol{x}$ and $\boldsymbol{\rho}_1 = \frac{\boldsymbol{e}}{\|\boldsymbol{e}\|}$ as nuisance parameters with corresponding test parameter(s) under each hypothesis,

$$\mathcal{H}_0 : \theta_1 = 0 \quad \text{vs.} \quad \mathcal{H}_1 : \theta_1 = \|\boldsymbol{e}\|,$$

where, θ_0 need not be defined since $\boldsymbol{G}_0 = 0$. Moreover, in the noiseless case (*i.e.,* $\boldsymbol{n} = 0$), if $\boldsymbol{C}$ and $\boldsymbol{e}$ are structured such that there always exists a subset of unattacked measurements that can be used to exactly reconstruct the unknown state $\boldsymbol{x}$ (a property refered to as S-sparse observability in [39]), then the PAIN monitor perfectly detects attacks [152] – *i.e.,* detects all attacks with no false alarms – invariant to the unknown state $\boldsymbol{x}$. We note that this result is not exclusive to PAIN attack monitors, as many other techniques can also provide perfect detection in the absence of noise (*e.g.,* [39], [46], [47], [48], [49], [50]).

B. Challenges and Future Research Directions

While the PAIN monitor can provide perfect detection capabilities in the absence of noise, bounding the attack detection performance of PAIN monitors in the presence of noise remains an open challenge. Bounding this performance is further complicated when measurements (and as a potential consequence the attack effect) comprise the $\boldsymbol{F}$, $\boldsymbol{G}_0$, and $\boldsymbol{G}_1$ matrices – as is common in networked dynamical systems, LTI systems, and hybrid dynamical systems. In this scenario, the models themselves can potentially be altered by a malicious attacker. While overcoming this challenge remains an open research area, solutions will likely utilize a collection of possibly application dependent techniques, ranging from bounding the potential attacker capabilities, to the co-design of security monitors and control systems in a CPS. In the remainder of this section, we discuss potential approaches for overcoming tampered models using system robustness, redundancy, and trust.

When the model noise is bounded, we say that a CPS exhibits robustness. Robustness is commonly used in CPS security to monitor whether the system is misbehaving. In safety-critical CPS this equates to monitoring for when the measurements significantly different from their estimates. Thus, it is likely that PAIN monitors can be designed for attack detection that leverage system robustness to ensure security; however, bounding the performance of PAIN attack monitoring remains challenging due to the highly non-linear nature of PAIN monitors.

When multiple measurements (or functions of measurements) provide similar information, we say that a CPS exhibits redundancy. Recently, measurement redundancy has been utilized by resilient state estimators [39], [46] to identify attacked sensors and actuators. Utilizing sensor redundancy by either designing separate monitors for individual sensors or switching the roles of sensors in the model may yield monitors that can accurate detect attacks. However, as in resilient state estimation, the primary challenge with utilizing redundancy in PAIN attack monitoring lies in searching all possible combinations of secure and insecure sensors to detect an attack [46]. Moreover, utilizing redundancy alone can not secure systems with (in general) more than half of the measurements attacked; thus, to provide maximum attack detection will ultimately require complementing redundancy techniques with some other approach (*e.g.,* robustness and/or trust).

Another approach to detect attacks in CPS utilizes a notion of *trust*. Trust represents the belief that a measurement/component/algorithm is likely to be correct. Recent work on behavioral-trust management has utilized Bayesian techniques to capture temporal behavioral (*i.e.,* fault/attack) correlations in the networked control setting [153]. Applying the concept of trust to PAIN attack detection appears promising since clean measurements can be identified that can be used to correlate with untrusted measurements. While quantified trust can be utilized to overcome the limitations of redundancy techniques, it potentially introduces new vulnerabilities – an attacker may try to manipulate our trust in a measurement. Limiting the risk of trust manipulation in CPS remains an open research problem where application-specific solutions are likely to prevail.

While robustness, redundancy, and trust provide three separate approaches for improving PAIN monitoring to support attack detection in real CPS, it is likely that some combination of these approaches will prove superior. By enabling PAIN monitoring for attack detection, requirements on accurate models may be avoided. This is especially true in (medical) cyber-physical systems where accurate (physiological) models often do not exist.

VIII. Conclusions

Monitoring for critical events in CPS presents several challenges, that include sparse data, inter/intra-system variance, and the fact that trust in monitors is brittle. Towards overcoming these challenges, this paper generalizes recent work on the design of PAIN monitors for CPS. Theoretically, PAIN designs can achieve a constant false alarm rate (CFAR) regardless of parameter uncertainty. Practically, the utility and robustness of PAIN monitors are illustrated through real-world CPS case studies including actuator fault detection and diagnosis in a smart buildings, meal detector for type I diabetic patients, and detection of critical pulmonary shunts in infants. These applications span CPS with networked dynamics, LTI dynamics, and hybrid dynamics. These applications illustrate how the general theory of PAIN monitor construction can be simplified for specific classes of system dynamics, which give rise to more specific methodologies. In all applications, the PAIN monitor is shown to have (significantly) less variance in monitoring performance and (often) outperforms other competing approaches in the literature. An initial extension of PAIN monitoring to detecting attacks in CPS is discussed and future research challenges identified.

References

[1] R. Platon, J. Martel, N. Woodruff, and T. Y. Chau, "Online fault detection in pv systems," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1200–1207, 2015.

[2] H. Jiang, J. J. Zhang, W. Gao, and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2947–2956, 2014.

[3] E. Fadel, V. C. Gungor, L. Nassef, N. Akkari, M. A. Malik, S. Almasri, and I. F. Akyildiz, "A survey on wireless sensor networks for smart grid," *Computer Communications*, vol. 71, pp. 22–33, 2015.

[4] R. Mitchell and R. Chen, "Behavior-rule based intrusion detection systems for safety critical smart grid applications," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1254–1263, 2013.

[5] K. Koziy, B. Gou, and J. Aslakson, "A low-cost power-quality meter with series arc-fault detection capability for smart grid," *IEEE Transactions on Power Delivery*, vol. 28, no. 3, pp. 1584–1591, 2013.

[6] R. Rinott, B. Carmeli, C. Kent, Y. Maman, Y. Rubin, and N. Slonim, "Utilizing assigned treatments as labels for supervised machine learning in clinical decision support," *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*, p. 493, 2012.

[7] R. A. Greenes, *Clinical Decision Support: The Road to Broad Adoption: Second Edition*, R. A. Greenes, Ed., 2014.

[8] J. A. Osheroff, E. A. Pifer, J. M. Teich, D. F. Sittig, and R. A. Jenders, "Improving outcomes with clinical decision support: an implementer's guide." HIMSS, 2005.

[9] D. L. Hunt, R. B. Haynes, S. E. Hanna, and K. Smith, "Effects of Computer-Based Clinical Decision Support Systems on Physician Performance and Patient Outcomes," *Jama*, vol. 280, no. 15, p. 1339, Oct. 1998.

[10] A. X. Garg, N. K. J. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. J. Devereaux, J. Beyene, J. Sam, and R. B. Haynes, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review." *Jama*, vol. 293, no. 10, pp. 1223–38, Mar. 2005.

[11] B. Chaudhry, J. Wang, and S. Wu, "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care," *Annals of internal medicine*, vol. 144, pp. 742–752, 2006.

[12] J. Kleissl and Y. Agarwal, "Cyber-physical energy systems: Focus on smart buildings," in *Proceedings of the 47th Design Automation Conference*. ACM, 2010, pp. 749–754.

[13] N. Batista, R. Melício, J. Matias, and J. Catalão, "Photovoltaic and wind energy systems monitoring and building/home energy management using zigbee devices within a smart grid," *Energy*, vol. 49, pp. 306–315, 2013.

[14] B. Spencer, "A study on building risk monitoring using wireless sensor network mica mote," in *First International Conference on Structural Health Monitoring and Intelligent Infrastructure, Japan*, 2003, pp. 353–363.

[15] L. Gurgen, O. Gunalp, Y. Benazzouz, and M. Gallissot, "Self-aware cyber-physical systems and applications in smart buildings and cities," in *Proceedings of the Conference on Design, Automation and Test in Europe*. EDA Consortium, 2013, pp. 1149–1154.

[16] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.

[17] E. Gilabert and A. Arnaiz, "Intelligent automation systems for predictive maintenance: A case study," *Robotics and Computer-Integrated Manufacturing*, vol. 22, no. 5, pp. 543–549, 2006.

[18] P.-Y. Chen, S. Yang, and J. A. McCann, "Distributed real-time anomaly detection in networked industrial sensing systems," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3832–3842, 2015.

[19] R. Loureiro, S. Benmoussa, Y. Touati, R. Merzouki, and B. O. Bouamama, "Integration of fault diagnosis and fault-tolerant control for health monitoring of a class of mimo intelligent autonomous vehicles," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 30–39, 2014.

[20] A. Serna and B. Marcotegui, "Detection, segmentation and classification of 3d urban objects using mathematical morphology and supervised learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 93, pp. 243–255, 2014.

[21] P. McDaniel, N. Papernot, and Z. B. Celik, "Machine learning in adversarial settings," *IEEE Security & Privacy*, vol. 14, no. 3, pp. 68–72, 2016.

[22] M. Jordan and T. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[23] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *International Conference on Critical Infrastructure Protection*. Springer, 2007, pp. 73–82.

[24] J. Minkel, "The 2003 northeast blackout–five years later," *Scientific American*, vol. 13, 2008.

[25] J. P. Conti, "The day the samba stopped [power blackouts]," *Engineering & Technology*, vol. 5, no. 4, pp. 46–47, 2010.

[26] S. Kuvshinkova, "Sql slammer worm lessons learned for consideration by the electricity sector," *North American Electric Reliability Council*, vol. 1, no. 2, p. 5, 2003.

[27] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[28] E. Ackerman, "Fatal tesla self-driving car crash reminds us that robots arent perfect," *IEEE-Spectrum*, vol. 1, 2016.

[29] D. Tong and F. Chung, "Recall after total intravenous anaesthesia due to an equipment misuse," *Canadian journal of anaesthesia*, vol. 44, no. 1, p. 73, 1997.

[30] W. H. Maisel, M. O. Sweeney, W. G. Stevenson, K. E. Ellison, and L. M. Epstein, "Recalls and safety alerts involving pacemakers and implantable cardioverter-defibrillator generators," *Jama*, vol. 286, no. 7, pp. 793–799, 2001.

[31] G. Richards, "Hackers vs slackers-[control security]," *Engineering & Technology*, vol. 3, no. 19, pp. 40–43, 2008.

[32] G. Loukas, *Cyber-physical attacks: A growing invisible threat*. Butterworth-Heinemann, 2015.

[33] A. S. Kesselheim, K. Cresswell, S. Phansalkar, D. W. Bates, and A. Sheikh, "Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation." *Health affairs (Project Hope)*, vol. 30, no. 12, pp. 2310–7, Dec. 2011.

[34] G. L. Alexander, "Issues of Trust and Ethics in Computerized Clinical Decision Support Systems," *Nursing Administration Quarterly*, vol. 30, no. 1, pp. 21–29, Jan. 2006.

[35] T. Clark, Y. David, M. Baretich, and T. Bauld, "Impact of clinical alarms on patient safety," ACCE Healthcare Technology Foundation, Tech. Rep., 2006.

[36] R. Koppel, A. Cohen, B. Abaluck, A. R. Localio, S. E. Kimmel, and B. L. Strom, "Role of computerized physician order entry systems in facilitating medication errors," *Journal of the American Medical Association*, vol. 293, no. 10, pp. 1197–1203, 2005.

[37] S. N. Weingart, M. Toth, D. Z. Sands, M. D. Aronson, R. B. Davis, and R. S. Phillips, "Physicians' decisions to override computerized drug alerts in primary care." *Archives of internal medicine*, vol. 163, no. 21, pp. 2625–31, Nov. 2003.

[38] R. S. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 90–95, 2011.

[39] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[40] S. Sundaram, M. Pajic, C. N. Hadjicostis, R. Mangharam, and G. J. Pappas, "The wireless control network: Monitoring for malicious behavior," in *Decision and Control (CDC), 2010 49th IEEE Conference on*. IEEE, 2010, pp. 5979–5984.

[41] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 1854–1859.

[42] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2014.

[43] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems*, vol. 35, no. 1, pp. 93–109, 2015.

[44] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber–physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[45] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *American Control Conference (ACC), 2013*. IEEE, 2013, pp. 3344–3349.

[46] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *ICCPS'14: ACM/IEEE 5th International Conference on Cyber-Physical Systems (with CPS Week 2014)*. IEEE Computer Society, 2014, pp. 163–174.

[47] Y. Shoukry, A. Puggelli, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Sound and complete state estimation for linear dynamical systems under sensor attacks using satisfiability modulo theory solving," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 3818–3823.

[48] Y. Shoukry, P. Nuzzo, N. Bezzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state reconstruction in differentially flat systems under sensor attacks using satisfiability modulo theory solving," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 3804–3809.

[49] M. Pajic, P. Tabuada, I. Lee, and G. J. Pappas, "Attack-resilient state estimation in the presence of noise," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 5827–5832.

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

[51] A. Gupta, C. Langbort, and T. Basar, "Optimal control in the presence of an intelligent jammer with limited actions," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, dec. 2010, pp. 1096 –1101.

[52] J. Weimer, S. Kar, and K. H. Johansson, "Distributed detection and isolation of topology attacks in power networks," in *International Conference on High Confidence Networked Systems*, 2012, pp. 65–72.

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

[54] J. Wang and I. C. Paschalidis, "Statistical traffic anomaly detection in time-varying communication networks," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 2, pp. 100–111, 2015.

[55] C. Cobelli and E. Carson, *Introduction to modeling in physiology and medicine*. Academic Press, 2008.

[56] D. Noble, "Modeling the heart–from genes to cells to the whole organ," *Science*, vol. 295, no. 5560, pp. 1678–1682, 2002.

[57] H. G. Mond and A. Proclemer, "The 11th world survey of cardiac pacing and implantable cardioverter-defibrillators: Calendar year 2009–a world society of arrhythmia's project," *Pacing and Clinical Electrophysiology*, vol. 34, no. 8, pp. 1013–1027, 2011.

[58] P. Bogdan, S. Jain, K. Goyal, and R. Marculescu, "Implantable pacemakers control and optimization via fractional calculus approaches: A cyber-physical systems perspective," in *Proceedings of the Third International Conference on Cyber-Physical Systems*, 2012, pp. 23–32.

[59] C. Cobelli, E. Renard, and B. Kovatchev, "Artificial pancreas: past, present, future," *Diabetes*, vol. 60, no. 11, pp. 2672–2682, 2011.

[60] H. J. Tulleken, "Grey-box modelling and identification using physical knowledge and bayesian techniques," *Automatica*, vol. 29, no. 2, pp. 285–308, 1993.

[61] T. Bohlin and S. F. Graebe, "Issues in nonlinear stochastic grey box identification," *International Journal of Adaptive Control and Signal Processing*, vol. 9, no. 6, pp. 465–490, 1995.

[62] N. R. Kristensen, H. Madsen, and S. B. Jørgensen, "Parameter estimation in stochastic grey-box models," *Automatica*, vol. 40, no. 2, pp. 225–237, 2004.

[63] A. Burgos, A. Goñi, A. Illarramendi, and J. Bermúdez, "Real-time detection of apneas on a pda," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 4, pp. 995–1002, 2010.

[64] V. A. Convertino, S. L. Moulton, G. Z. Grudic *et al.*, "Use of advanced machine-learning techniques for noninvasive monitoring of hemorrhage," *Journal of Trauma-Injury, Infection, and Critical Care*, vol. 71, no. 1, pp. S25–S32, 2011.

[65] T. M. Mitchell, R. Hutchinson, M. A. Just *et al.*, "Classifying instantaneous cognitive states from fmri data," in *AMIA Annual Symposium Proceedings*, 2003, pp. 465–469.

[66] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 1, pp. 1–12, 2010.

[67] S. Saria, D. Koller, and A. Penn, "Learning individual and population level traits from clinical temporal data," in *Proceedings of Neural Information Processing Systems*, 2010, pp. 1–9.

[68] S. Saria, A. Rajani, J. Gould, D. Koller, and A. Penn, "Integration of early physiological responses predicts later illness severity in preterm infants," *Science translational medicine*, vol. 2, no. 48, pp. 48–65, Sep. 2010.

[69] A. E. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford, "Machine learning and decision support in critical care," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 444–466, 2016.

[70] M. Stacey and C. McGregor, "Temporal abstraction in intelligent clinical data analysis: A survey," *Artificial intelligence in medicine*, vol. 39, no. 1, pp. 1–24, 2007.

[71] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[72] F. Sha and L. Saul, "Large margin hidden Markov models for automatic speech recognition," *Advances in neural information processing systems*, pp. 1249—-1256, 2006.

[73] R. E. Kass and D. Steffey, "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)," *Journal of the American Statistical Association*, vol. 84, no. 407, pp. 717–726, 1989.

[74] A. Gelman, *Bayesian data analysis*, 1995.

[75] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Nonparametric Bayesian Learning of Switching Linear Dynamical Systems," *Proceedings of Neural Information Processing Systems*, vol. 21, 2008.

[76] E. Fox, E. B. Sudderth, M. I. Jordan, and S. Alan, "Bayesian Nonparametric Inference of Switching Dynamic Linear Models," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1569–1585, 2011.

[77] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ACM, 2012, pp. 214–226.

[78] R. S. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork, "Learning fair representations." *ICML (3)*, vol. 28, pp. 325–333, 2013.

[79] J. Weimer, D. Varagnolo, and K. H. Johansson, "Distributed model-invariant detection of unknown inputs in networked systems," in *International Conference on High Confidence Networked Systems*, 2013, pp. 127–134.

[80] J. Weimer, D. Varagnolo, M. Stankovic, and K. Johansson, "Parameter-invariant detection of unknown inputs in networked systems," in *Conference on Decision and Control*, 2013, pp. 4379–4384.

[81] R. Ivanov, J. Weimer, A. Simpao, M. Rehman, and I. Lee, "Early detection of critical pulmonary shunts in infants," in *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems (ICCPS)*, 2015, pp. 110–119.

[82] R. Ivanov, J. Weimer, A. F. Simpao, M. A. Rehman, and I. Lee, "Prediction of critical pulmonary shunts in infants," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 6, pp. 1936–1952, Nov 2016.

[83] A. Roederer, J. Weimer, J. DiMartino, J. Gutsche, and I. Lee, "Robust monitoring of hypovolemia in intensive care patients using photoplethysmogram signals," in *Proceedings of the 37th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. EMBC, 2015.

[84] S. Chen, J. Weimer, M. R. Rickels, A. Peleckis, and I. Lee, "Towards a model-based meal detector for type i diabetics," in *Medical Cyber-Physical Systems Workshop 2015*, 2015.

[85] J. Weimer, S. Chen, A. Peleckis, M. R. Rickels, and I. Lee, "Physiology-invariant meal detection for type 1 diabetes," *Diabetes Technology & Therapeutics*, vol. 18, no. 10, pp. 616–624, 2016.

[86] J. Weimer, S. A. Ahmadi, J. Araujo, F. M. Mele, D. Papale, I. Shames, H. Sandberg, and K. H. Johansson, "Active actuator fault detection and diagnostics in HVAC systems," in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ser. BuildSys '12. New York, NY, USA: ACM, 2012, pp. 107–114. [Online]. Available: http://doi.acm.org/10.1145/2422531.2422551

[87] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Royal Society of London Philosophical Transactions Series A*, vol. 231, pp. 289–337, 1933.

[88] S. Karlin and H. Rubin, "The theory of decision procedures for distributions with monotone likelihood ratio," *The Annals of Mathematical Statistics*, pp. 272–299, 1956.

[89] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, Inc., New York, 1968.

[90] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," in *Breakthroughs in Statistics*. Springer, 1992, pp. 73–108.

[91] J. Weimer, J. Araujo, M. Amoozadeh, S. A. Ahmadi, H. Sandberg, and K. H. Johansson, "Parameter-invariant actuator fault diagnostics in cyber-physical systems with application to building automation," in *Control of Cyber-Physical Systems*. Springer, 2013, pp. 179–196.

[92] J. Weimer, R. Ivanov, A. Roederer, S. Chen, and I. Lee, "Parameter-invariant design of medical alarms," *Design Test, IEEE*, vol. 32, no. 5, pp. 9–16, Oct 2015.

[93] L. L. Scharf and C. Demeure, *Statistical Signal Processing*. Addison-Wesley Publishing Company, 1991.

[94] C. M. Bishop, "Pattern recognition," *Machine Learning*, 2006.

[95] S. Bolognani, A. Carron, A. Di Vittorio, D. Romeres, L. Schenato, and S. Zampieri, "Distributed multi-hop reactive power compensation in smart micro-grids subject to saturation constraints," in *51st IEEE Conference on Decision and Control*, 2012, pp. 785–790.

[96] J. Machowski, J. Bialek, and J. Bumby, *Power system dynamics: stability and control*. John Wiley & Sons, 2011.

[97] J. Weimer, B. Sinopoli, and B. Krogh, "An approach to leak detection using wireless sensor networks at carbon sequestration sites," *International Journal of Greenhouse Gas Control*, vol. 9, pp. 243–253, 2012.

[98] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," in *IEEE International Symposium on Intelligent Control, MCCA*, 2005, pp. 719–724.

[99] M. Andreasson, D. V. Dimarogonas, H. Sandberg, and K. H. Johansson, "Distributed control of networked dynamical systems: Static feedback, integral action and consensus," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1750–1764, 2014.

[100] B. Zalba, J. M. Marin, L. F. Cabeza, and H. Mehling, "Review on thermal energy storage with phase change: materials, heat transfer analysis and applications," *Applied thermal engineering*, vol. 23, no. 3, pp. 251–283, 2003.

[101] M. Khouzani and S. Sarkar, "Maximum damage battery depletion attack in mobile sensor networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2358–2368, 2011.

[102] M. Khouzani, S. Sarkar, and E. Altman, "Maximum damage malware attack in mobile wireless networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 20, no. 5, pp. 1347–1360, 2012.

[103] J. J. Grainger and W. D. Stevenson, *Power system analysis*. McGraw-Hill, 1994.

[104] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.

[105] U. DoE, "Energy efficiency trends in residential and commercial buildings," *US Department of Energy, Washington, DC Available at: http://apps1. eere. energy. gov/buildings/publications/pdfs/corporate/bt_stateindustry. pdf*, 2008.

[106] J. Dimitropoulos*, L. C. Hunt, and G. Judge, "Estimating underlying energy demand trends using uk annual data," *Applied Economics Letters*, vol. 12, no. 4, pp. 239–244, 2005.

[107] Y. Kim, T. Schmid, M. B. Srivastava, and Y. Wang, "Challenges in resource monitoring for residential spaces," in *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM, 2009, pp. 1–6.

[108] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse, "The smart thermostat: using occupancy sensors to save energy in homes," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2010, pp. 211–224.

[109] V. L. Erickson, M. Á. Carreira-Perpiñán, and A. E. Cerpa, "Observe: Occupancy-based system for efficient reduction of hvac energy," in *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*. IEEE, 2011, pp. 258–269.

[110] Y. Agarwal, B. Balaji, S. Dutta, R. Gupta, and T. Weng, "Duty-cycling buildings aggressively: The next frontier in hvac control," in *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, april 2011, pp. 246 –257.

[111] A. Marchiori and Q. Han, "Distributed wireless control for building energy management?" in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ser. BuildSys '10. New York, NY, USA: ACM, 2010, pp. 37–42. [Online]. Available: http://doi.acm.org/10.1145/1878431.1878441

[112] J. Ma, J. Qin, T. Salsbury, and P. Xu, "Demand reduction in building energy systems based on economic model predictive control," *Chemical Engineering Science*, vol. 67, no. 1, pp. 92 – 100, 2012, ¡ce:title¿Dynamics, Control and Optimization of Energy Systems¡/ce:title¿. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0009250911005240

[113] H. A. I. S. Goyal and P. Barooah, "Zone level control algorithms based on occupancy information for energy efficient buildings," in *American Control Conference*, 2012.

[114] Y. Ma, A. Kelman, A. Daly, and F. Borrelli, "Predictive control for energy efficient buildings with thermal storage: Modeling, stimulation, and experiments," *Control Systems, IEEE*, vol. 32, no. 1, pp. 44 –64, feb. 2012.

[115] F. Oldewurtel, A. Parisio, C. N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari, "Use of model predictive control and weather forecasts for energy efficient building climate control," *Energy and Buildings*, vol. 45, no. 0, pp. 15 – 27, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378778811004105

[116] T. X. Nghiem, M. Behl, R. Mangharam, and G. J. Pappas, "Scalable scheduling of building control systems for peak demand reduction," in *American Control Conference*, 2012.

[117] S. Katipamula and M. R. Brambley, "Methods for fault detection, diagnostics, and prognostics for building systems - a review, part i," *HVAC&R Research*, vol. 11(1), p. 325, Jan. 2005.

[118] ——, "Methods for fault detection, diagnostics, and prognostics for building systems - a review, part ii," *HVAC&R Research*, vol. 11(2), p. 169187, Apr. 2005.

[119] N. Fernandez, M. Brambley, S. Katipamula, H. Cho, J. Goddard, and L. D. b., "Self correcting hvac controls project final report pnnl-19074," Pacific Northwest National Laboratory, Richland, WA., Tech. Rep., 2009.

[120] N. Djuric and V. Novakovic, "Review of possibilities and necessities for building lifetime commissioning," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 2, pp. 486 – 492, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032107001591

[121] L. Jagemar and D. Olsson, "The epbd and continuous commisioning," CIT Energy Mangement AB, Goteborg, Sweden, Tech. Rep., Oct. 2007.

[122] G. Pattarello, L. Wei, A. Ebadat, B. Wahlberg, and K. H. Johansson, "The kth open testbed for smart hvac control," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–2.

[123] P. M. Bentler and T. Dijkstra, "Efficient estimation via linearization in structural models," *Multivariate analysis VI*, pp. 9–42, 1985.

[124] D. H. Anderson, *Compartmental modeling and tracer kinetics*. Springer Science & Business Media, 2013, vol. 50.

[125] R. N. Bergman, "Toward physiological understanding of glucose tolerance: minimal-model approach," *Diabetes*, vol. 38, no. 12, pp. 1512–1527, 1989.

[126] A. Mandow, J. L. Martinez, J. Morales, J. L. Blanco, A. Garcia-Cerezo, and J. Gonzalez, "Experimental kinematics for wheeled skid-steer mobile robots," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. IEEE, 2007, pp. 1222–1227.

[127] K. Sreenath, N. Michael, and V. Kumar, "Trajectory generation and control of a quadrotor with a cable-suspended load-a differentially-flat hybrid system," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4888–4895.

[128] P. C. Sen, "Electric motor drives and control-past, present, and future," *IEEE Transactions on Industrial Electronics*, vol. 37, no. 6, pp. 562–575, 1990.

[129] L. A. Rossi, B. Krishnamachari, and C.-C. Kuo, "Distributed parameter estimation for monitoring diffusion phenomena using physical models," in *Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*. IEEE, 2004, pp. 460–469.

[130] J. Durbin and S. J. Koopman, "Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 1, pp. 3–56, 2000.

[131] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[132] J. D. Hamilton, *Time series analysis*. Princeton university press Princeton, 1994, vol. 2.

[133] A. D. Association *et al.*, "Hypoglycemia (low blood glucose)," *URL http://www. diabetes. org/living-with-diabetes/treatment-and-care/blood-glucose-control/hypoglycemia-low-blood. html*, 2011.

[134] E. Dassau, B. W. Bequette, B. A. Buckingham, and F. J. Doyle, "Detection of a meal using continuous glucose monitoring implications for an artificial β-cell," *Diabetes care*, vol. 31, no. 2, pp. 295–300, 2008.

[135] H. Lee and B. W. Bequette, "A closed-loop artificial pancreas based on model predictive control: Human-friendly identification and automatic meal disturbance rejection," *Biomedical Signal Processing and Control*, vol. 4, no. 4, pp. 347–354, 2009.

[136] R. A. Harvey, E. Dassau, H. Zisser, D. E. Seborg, and F. J. Doyle, "Design of the glucose rate increase detector a meal detection module for the health monitoring system," *Journal of diabetes science and technology*, p. 1932296814523881, 2014.

[137] F. Cameron, G. Niemeyer, and B. A. Buckingham, "Probabilistic evolving meal detection and estimation of meal total glucose appearance," *Journal of diabetes science and technology*, vol. 3, no. 5, pp. 1022–1030, 2009.

[138] K. Turksoy and A. Cinar, "Real-time insulin bolusing for unannounced meals using cgm measurements," *IFAC-PapersOnLine*, vol. 48, no. 20, pp. 219–224, 2015.

[139] J. Grimm, "Exercise in type 1 diabetes," *Exercise and sport in diabetes*, pp. 25–43, 2005.

[140] S. E. Capes, D. Hunt, K. Malmberg, P. Pathak, and H. C. Gerstein, "Stress hyperglycemia and prognosis of stroke in nondiabetic and diabetic patients a systematic overview," *Stroke*, vol. 32, no. 10, pp. 2426–2432, 2001.

[141] P. Reichard, B.-Y. Nilsson, and U. Rosenqvist, "The effect of long-term intensified insulin treatment on the development of microvascular complications of diabetes mellitus," *New England Journal of Medicine*, vol. 329, no. 5, pp. 304–309, 1993.

[142] K. J. Guelfi, T. W. Jones, and P. A. Fournier, "The decline in blood glucose levels is less with intermittent high-intensity compared with moderate exercise in individuals with type 1 diabetes," *Diabetes Care*, vol. 28, no. 6, pp. 1289–1294, 2005.

[143] C. Dalla Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, and C. Cobelli, "The UVA/PADOVA Type 1 Diabetes Simulator new features," *Journal of diabetes science and technology*, vol. 8, no. 1, pp. 26–34, 2014.

[144] R. N. Bergman, Y. Z. Ider, C. R. Bowden, and C. Cobelli, "Quantitative estimation of insulin sensitivity." *American Journal of Physiology-Endocrinology And Metabolism*, vol. 236, no. 6, p. E667, 1979.

[145] R. Gillis, C. C. Palerm, H. Zisser, L. Jovanovic, D. E. Seborg, and F. J. Doyle, "Glucose estimation and prediction through meal responses using ambulatory subject data for advisory mode model predictive control," *Journal of diabetes science and technology*, vol. 1, no. 6, pp. 825–833, 2007.

[146] G. Nucci and C. Cobelli, "Models of subcutaneous insulin kinetics. a critical review," *Computer methods and programs in biomedicine*, vol. 62, no. 3, pp. 249–257, 2000.

[147] T. Kobayashi, S. Sawano, T. Itoh, K. Kosaka, H. Hirayama, and Y. Kasuya, "The pharmacokinetics of insulin after continuous subcutaneous infusion or bolus subcutaneous injection in diabetic patients," *Diabetes*, vol. 32, no. 4, pp. 331–336, 1983.

[148] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Proceedings of the 47th Design Automation Conference*. ACM, 2010, pp. 731–736.

[149] E. A. Lee, "Cps foundations," in *Design Automation Conference (DAC), 2010 47th ACM/IEEE*. IEEE, 2010, pp. 737–742.

[150] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber–physical systems," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 13–28, 2012.

[151] F. Gustafsson and F. Gustafsson, *Adaptive filtering and change detection*. Citeseer, 2000, vol. 1.

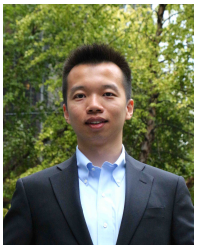
[152] J. Weimer, N. Bezzo, M. Pajic, G. J. Pappas, O. Sokolsky, and I. Lee, "Resilient parameter-invariant control with application to vehicle cruise control," in *Control of Cyber-Physical Systems*. Springer, 2013, pp. 197–216.

[153] S. Sundaram, J. Chang, K. K. Venkatasubramanian, C. Enyioha, I. Lee, and G. J. Pappas, "Reputation-based networked control with data-corrupting channels," in *Proceedings of the 14th international conference on Hybrid systems: computation and control*. ACM, 2011, pp. 291–300.

James Weimer received a B.S. degree in Electrical Engineering from Purdue University, West Lafayette in 2005 and M.S. and Ph.D. degrees in Electrical and Computer Engineering from Carnegie Mellon University in 2007 and 2010, respectively. He is currently a Research Assistant Professor in the Department of Information and Computer Science at the University of Pennsylvania and was previously a Postdoctoral Researcher in the Department of Automatic Control at the Royal Institute of Technology KTH, Stockholm and the PRECISE center at the University of Pennsylvania. His research interests focus on the stochastic and formal methods for cyber-physical systems with application to medical devices/monitors, networked systems, building energy management, and smart grids.

Radoslav Ivanov received his B.A. in Computer Science and Economics from Colgate University, New York and is currently pursuing a doctoral degree in the Computer and Information Science Department at the University of Pennsylvania. His main research focus is on the security and control of cyber-physical systems (CPS), in particular automotive and medical CPS. He is also interested in predictive and retrospective analysis of medical patient data.

Alexander Roederer received his B.S. in Computer Science and Mathematics from the University of Miami, his M.S.E. and Ph.D. in Computer and Information Science from the University of Pennsylvania. His research interests include the application of machine learning to high-frequency, multi-source physiologic data to develop clinical decision support systems.



Sanjian Chen received his B.E. degree from Tsinghua University, Beijing and M.S.E. and Ph.D. in Computer and Information Science from the University of Pennsylvania His research interests include model-based analysis of cyber-physical systems with application to medical and automotive systems. He received IEEE RTSS 2012 Best Paper Award.

Oleg Sokolsky is a research associate professor of computer and information science at the University of Pennsylvania. His research interests include the application of formal methods to the development of cyber-physical systems, architecture modeling and analysis, specification-based monitoring, as well as software-safety certification. He received the Ph.D. degree in computer science from Stony Brook University, New York.

Insup Lee is the Cecilia Fitler Moore Professor of Computer and Information Science and serves as the Director of the PRECISE Center at the University of Pennsylvania. He holds a secondary appointment in the Department of Electrical and Systems Engineering. His research interests include cyber-physical systems, real-time and embedded systems, runtime assurance and verification, trust management, and high-confidence medical systems. He received a PhD in Computer Science from the University of Wisconsin, Madison. He is IEEE fellow. He received IEEE TC-RTS Outstanding Technical Achievement and Leadership Award in 2008.