4-2018

Context-Aware Detection in Medical Cyber-Physical Systems

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

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

Insup Lee
University of Pennsylvania, lee@cis.upenn.edu

9th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS) (part of CPS Week 2018), Porto, Portugal, April 11 - 13, 2018.

Context-Aware Detection in Medical Cyber-Physical Systems

Abstract
This paper considers the problem of incorporating context in medical cyber-physical systems (MCPS) applications for the purpose of improving the performance of MCPS detectors. In particular, in many applications additional data could be used to conclude that actual measurements might be noisy or wrong (e.g., machine settings might indicate that the machine is improperly attached to the patient); we call such data context. The first contribution of this work is the formal definition of context, namely additional information whose presence is associated with a change in the measurement model (e.g., higher variance). Given this formulation, we developed the context-aware parameter-invariant (CA-PAIN) detector; the CA-PAIN detector improves upon the original PAIN detector by recognizing events with noisy measurements and not raising unnecessary false alarms. We evaluate the CA-PAIN detector both in simulation and on real-patient data; in both cases, the CA-PAIN detector achieves roughly a 20-percent reduction of false alarm rates over the PAIN detector, thus indicating that formalizing context and using it in a rigorous way is a promising direction for future work.

Disciplines
Computer Engineering | Computer Sciences

Comments
9th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS) (part of CPS Week 2018), Porto, Portugal, April 11 - 13, 2018.

Context-Aware Detection in Medical Cyber-Physical Systems

Radoslav Ivanov
Department of Computer and Information Science
University of Pennsylvania
Email: rivanov@seas.upenn.edu

James Weimer
Department of Computer and Information Science
University of Pennsylvania
Email: weimerj@seas.upenn.edu

Insup Lee
Department of Computer and Information Science
University of Pennsylvania
Email: lee@cis.upenn.edu

***Abstract*—This paper considers the problem of incorporating context in medical cyber-physical systems (MCPS) applications for the purpose of improving the performance of MCPS detectors. In particular, in many applications additional data could be used to conclude that actual measurements might be noisy or wrong (e.g., machine settings might indicate that the machine is improperly attached to the patient); we call such data context. The first contribution of this work is the formal definition of context, namely additional information whose presence is associated with a change in the measurement model (e.g., higher variance). Given this formulation, we developed the context-aware parameter-invariant (CA-PAIN) detector; the CA-PAIN detector improves upon the original PAIN detector by recognizing events with noisy measurements and not raising unnecessary false alarms. We evaluate the CA-PAIN detector both in simulation and on real-patient data; in both cases, the CA-PAIN detector achieves roughly a 20-percent reduction of false alarm rates over the PAIN detector, thus indicating that formalizing context and using it in a rigorous way is a promising direction for future work.**

I. Introduction

Recent advances in medical device technologies have led to an explosion in the number and kind of medical devices available. The large amounts of digital data collected by these devices [7], [16] provide great opportunities for developing Medical Cyber-Physical Systems (MCPS) in order to improve health outcomes and reduce costs [20]. Such systems would aid clinicians in multiple ways, ranging from providing prompts to clinicians (in case they are focused on the patient and not looking at the monitors), to alerting clinicians of unsuspected events (by processing time-series data and discovering trends over a long period of time, e.g., in pulmonary shunt detection [12], [13] and hypovolemia detection [22]), to providing fully closed-loop systems (e.g., the artificial pancreas project [6]).

Building reliable MCPS with guarantees presents multiple challenges, however. The first obstacle is that data is often missing or wrong. Furthermore, there are multiple artifacts during a patient's stay in a hospital (e.g., the patient is moving or clinicians are checking the mechanical ventilator for leaks) that introduce noise in measurement data or render it useless altogether (*e.g.,* pulse oximeter falls off the patient). Thirdly, human physiology varies greatly with patients, which means that MCPS need to be robust to very different model parameters and patient reactions. Finally, the safety-criticality of medical applications poses strict requirements on the performance of MCPS such that they must both have very good average performance and provide guarantees for each individual as well.

In order to address the last two challenges, namely guaranteed performance and robustness to physiological parameters, in previous work we developed the Parameter-Invariant (PAIN) detector for detection of critical patient events in real time [32], [31]. By using PAIN statistics that are unaffected by the specific values of physiological parameters (e.g., metabolic rate, diffusion rate), the PAIN detector provides maximal invariance to the nuisance physiological parameters, thus resulting in (near) constant false alarm rates across different patients. The PAIN detector has been shown to have very good detection performance in multiple medical applications such as pulmonary shunt detection [12], [13], meal detection in type I diabetes [30], and detection of hypovolemic shock [22]. Despite its robustness, however, the PAIN detector is still affected by bad data, as caused by technical issues (e.g., machine not properly connected) or other unmodeled events.

To address the shortcomings of the PAIN detector, in this work we make use of other available (context) data in an operating room/intensive care unit (OR/ICU) in order to judge when measurements might be noisy or inaccurate. For example, a typical anesthesia machine [1] outputs a range of context data in addition to standard measurements (e.g., machine settings, clinician inputs). If context data is not within normal ranges, one can conclude that the machine is likely not operating according to specification and that the measurements that it provides should be treated with caution.[1] Thus, in this paper we aim to augment the PAIN detector with knowledge of (formally defined) context in an effort to reduce some of the technical alarms that might be raised in an OR/ICU.

There is a significant body of literature in the MCPS detection/estimation space. At a high level, works can be

This work was supported in part by the NSF CNS-1505799, CNS-1652544 and CNS-1505701 grants, and the Intel-NSF Partnership for Cyber-Physical Systems Security and Privacy.

[1] A similar notion of context can be naturally used in other domains as well, e.g., autonomous vehicles, as discussed in Section VII.

classified into three classes: white-box (model-based), grey-box, and black-box (data-driven). When good models are available, one might be able to develop approaches with strong theoretical guarantees and with good results [3], [18]. Alternatively, compartmental models are also developed that result in a grey-box setting with lumped-parameter models; examples include the cardiac [28] and insulin-glucose systems [14]. Finally, machine learning approaches have also been developed looking for trends as revealed by the data instead of building first-principles models [17], [19], [23], [24]. All of the above approaches, however, require either good physiological models or rich training data, which are not available in many medical applications. The PAIN detector alleviates some of these challenges by providing guaranteed performance regardless of specific patient parameters and without requiring training data. At the same time, as argued above, the PAIN detector might perform poorly in scenarios with bad data as caused by unmodeled artifacts (e.g., an improperly connected machine). While it is possible to use standard anomaly detection techniques to identify bad data points [29], it is challenging to develop detectors that take this information into account (e.g., by using predictions instead of measurements) in scenarios with uncertain models.

The notion of using (discrete) contextual information has also been explored in different CPS domains, at different levels of generality and formality. In detection and estimation, works have been developed for handling both discrete and continuous data (e.g., target tracking [15], quantized measurements [21], Kalman filtering with intermittent measurements [27]) but these works are either too general (thus resulting in approximation algorithms) or too specific, i.e., they do not consider context to be additional data but rather a function of available measurements. In robotics, researchers have used contextual cues for the purpose of simultaneous localization and mapping [2], [4], [5], [33], although these approaches do not provide a general definition of context. In prior work, we developed the context-aware filter, in which context measurements are modeled as binary measurements with a known probability given the state [9]; this approach was applied to the problem of estimating blood oxygen content during surgery [11]. One limitation of this work is that obtaining the distribution of context measurements often requires significant domain expertise and/or training data that may not be available.

To address the limitations of related work, in this paper we combine the notions of context and parameter invariance. We define a context measurement as binary context data that is correlated with inaccurate measurements (*e.g.,* an unexpected machine setting is a context measurement that indicates the machine might be providing noisy or altogether meaningless measurements). Using this framework, we develop the context-aware parameter-invariant (CA-PAIN) detector that still provides guaranteed performance regardless of physiological parameters but is also aware of when the available measurements might be noisy. In particular, when a context measurement is present, the detector adapts and treats the measurements during that time as having an unknown high variance.

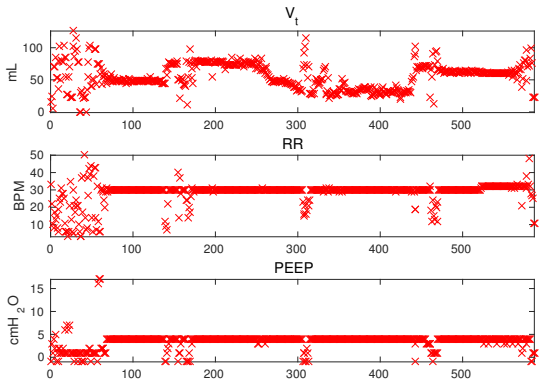


Fig. 1: Typical missing/bad data patterns over time in a surgery case at CHOP. V_t denotes tidal volume (measured in milliliters), RR denotes respiratory rate (measurement in beats per minute, BPM), and PEEP denotes positive end-expiratory pressure (measured in centimeters of water). Missing measurements are set to -1, i.e., they lie on the lower border of each graph.

To achieve both goals of the CA-PAIN detector (namely, improve performance during bad data periods while maintaining the parameter invariance guarantee), a total of three PAIN statistics are generated: 1) a PAIN statistic assuming all measurements have the same variance, 2) a PAIN statistic obtained only by measurements with low variance, and 3) a PAIN statistic obtained only by using measurements with high variance. We derive a decision rule based on these statistics such that alarms are only raised when all statistics provide evidence in favor of the same hypothesis. It is important to emphasize that the CA-PAIN detector is also invariant to the actual value of both normal and high variances; thus we make no assumptions on the values of the respective variances.

The CA-PAIN detector is evaluated both in simulation and on real data. We simulated multiple patients with different (unknown) parameters; at each time, if a context measurement is present, a high-variance noise is added to the actual (continuous) measurements. We observe on average a 20-percent improvement due to context, *i.e.,* for similar detection rates, the CA-PAIN detector's false alarm rate is roughly 20 percent lower than that of the vanilla PAIN detector.

For real-data evaluation, we use data obtained from the Children's Hospital of Philadelphia (CHOP) during lung lobectomy surgeries on infants.[2] Since lobectomies often require one-lung ventilation (thus introducing a pulmonary shunt, i.e., a blockage of one of the bronchi), patients often experience desaturation events (low-blood-oxygen events) that are especially dangerous in infants. Thus, we aim to detect such desaturation events before they occur and alert clinicians in real time. The PAIN detector was originally applied to this problem with good performance but raised multiple false alarms during

[2]A lung lobectomy is the incision of a cystic lesion on a lung.

bad-data events [12], [13]; such events were often caused by an improperly connected anesthesia machine (e.g., when clinicians are checking for leaks).

To improve the performance of the PAIN detector, we identified a context measurement that is correlated with bad-data events. As illustrated in Figure 1, missing positive end-expiratory pressure (PEEP) or peak-inspiratory pressure (PIP) values (which are inputs set by clinicians) are correlated with greater noise in actual measurements since they are provided by the same machine and are not recorded if the machine is not properly connected. Consequently, we applied the CA-PAIN detector by using missing PIP/PEEP as context measurements. The CA-PAIN detector results in a significant reduction in false alarms, eliminating about 20 percent of false alarms on average. Equivalently, for the same false alarm rate, the CA-PAIN results in an addition 5% of life-critical detections.

In summary the contributions of this paper are three-fold: 1) the formalization of context in the MCPS detection setting; 2) the development of the CA-PAIN detector that is invariant to patient physiology and is aware of the quality of available measurements through the use of context; 3) the evaluation of the CA-PAIN detector both in simulation and on real-patient data from CHOP.

The remainder of this paper is organized as follows. Section II introduces the background terminology and notation, and Section III presents the problems addressed in this paper. A formal definition of context is provided in Section IV, whereas Section V presents the design of the CA-PAIN detector. Section VI contains the detector's evaluation, while Sections VII and VIII provide a discussion and concluding remarks, respectively.

II. Background

Before presenting the precise problem formulation, in this section we provide a brief introduction to the PAIN detector, as it pertains to clinical monitoring applications.

A. Overview of Detection Approaches in MCPS

In the PAIN detection setting, the problem is to distinguish between one of two hypotheses: 1) the null that captures normal operation (e.g., patient is in a safe state) or 2) the alternative that models the detection scenario (e.g., patient is in an unsafe state). Given a parameterized model for each hypothesis, the problem is to determine from available measurements which hypothesis is true, regardless of the specific values of the parameters. Mathematically, we assume the measurements $\boldsymbol{y} \in \mathbb{R}^N$ were drawn from a distribution $\boldsymbol{y} \sim f_{\boldsymbol{\theta}}$ that is parameterized by the patient-specific parameters $\boldsymbol{\theta}$. The hypothesis test is thus a test of parameters:

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

where $\mathcal{H}_0$ is the null hypothesis, $\mathcal{H}_1$ is the alternative, and Θ_0 and Θ_1 are corresponding parameter sets associated with each hypothesis.

The performance of any detector ϕ is evaluated using false positive and true positive rates (also known as false alarms and true alarms). A false positive occurs when $\phi(\boldsymbol{y}) = 1$ when $\mathcal{H}_0$ is true whereas a true positive occurs when $\phi(\boldsymbol{y}) = 1$ when $\mathcal{H}_1$ is true. In general, these metrics introduce a trade-off such that one is improved at the expense of the other. Note that achieving a good balance between the two metrics is only possible in cases where the sets Θ_0 and Θ_1 do not have significant overlap. Thus, one of the major challenges in hypothesis testing (especially in MCPS where models are usually only partially known) is to develop models that lead to well separated hypotheses. Once the hypothesis test is developed, the goal is to build a detector that has the highest-possible true positive rate (given the separation of the two hypotheses) for any false alarm rate – this is known as a Universally Most Powerful (UMP) test. Thus, a UMP test would allow system designers to pick a desired false alarm rate depending on the application and achieve the highest-possible corresponding true positive rate.

If the parameters were known, say $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ under each hypothesis respectively, then one can build a UMP test by computing the likelihood ratio test:

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

and comparing the result to a predefined threshold in order to determine whether to raise an alarm on not.

In the medical setting, patient parameters are almost never known and are difficult to measure (e.g., measuring the diffusion rate would require knowing the exact geometry of the lung); in addition, parameters tend to change over time. In such cases, it might be possible to estimate parameters from data and perform a generalized likelihood ratio test by comparing the likelihoods of the estimated parameters under each hypothesis:

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

Generalized likelihood ratio tests work well in applications with sufficient high-quality data where good parameter estimates can be computed. However, due to the inferior quality and often small amount of medical data available, obtaining good parameter estimates is challenging in the MCPS setting. In such a scenario one might develop a parameter-invariant test, *i.e.,* a test that generates a statistic that is invariant to the specific values of the parameters. Invariance is formally captured using groups; we say that a statistic $t(\boldsymbol{y})$ is invariant to a group of transformations $\mathcal{G}$ if and only if

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

This means that the statistic's value does not change even if some transformation g is applied to the measurements first. Intuitively, groups capture all the behaviors that might change across patients and affect the detector's performance. Specific groups are discussed in the next subsection.

Ideally, one would like to develop a maximally invariant statistic, i.e., a statistic that is invariant only to $\mathcal{G}$ and nothing

else (note that $t(\boldsymbol{y}) = 0$ is invariant to any group but it is useless). Formally, this is defined as:

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

i.e., the maximally invariant statistic naturally creates an equivalence relation on the orbits in $\mathcal{G}$.

Maximally invariant statistics are useful because they preserve as much information as possible while also being invariant to $\mathcal{G}$. They can be used to design maximally invariant tests:

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

where η is some chosen threshold. Usually, the group $\mathcal{G}$ is chosen so that it captures the transformations under $\mathcal{H}_0$ – this ensures that $t(\boldsymbol{y})$ has the same distribution under $\mathcal{H}_0$ under transformations in $\mathcal{G}$; in turn, this means that the test ϕ_{MI} can be designed to have a constant false alarm rate regardless of the parameter values. The true alarm rate of ϕ_{MI} will vary with the parameter values. While it might be possible to design a universally most powerful invariant (UMPI) test (i.e., a test that achieves the highest true alarm rate for any chosen level of false alarm rate) under certain circumstances, this cannot be done in general. Yet, parameter-invariant tests have been shown to work well in practice.

B. Parameter Invariant Tests for Linear Systems

In multiple medical detection applications (namely, all MCPS detectors we have developed [12], [13], [22], [30], [31]), the detection problem often reduces to a matched-subspace test:

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{y} = \boldsymbol{F}_0\boldsymbol{\theta}_0 + \sigma\boldsymbol{n} \\ \mathcal{H}_1 &: \boldsymbol{y} = \boldsymbol{F}_1\boldsymbol{\theta}_1 + \sigma\boldsymbol{n}, \end{aligned} \tag{3}$$

where $\boldsymbol{F}_0 \in \mathbb{R}^{N\times k}$ and $\boldsymbol{F}_1 \in \mathbb{R}^{N\times p}$ are known matrices that determine the subspace of the measurements' mean under each hypothesis, $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are unknown (vector) parameters that determine the coordinates of the measurements' mean under the respective hypothesis, σ is an unknown scale, and $\boldsymbol{n}$ is zero-mean noise.[3]. Since the unknown parameters can take on any values, the test in (3) is essentially a test of which subspace (the column space of $\boldsymbol{F}_0$ or $\boldsymbol{F}_1$) the measurements are more likely to belong to.

In order to develop a test that is invariant to the unknown parameters in (3) under $\mathcal{H}_0$, we ask for invariance to the group $\mathcal{G}_0$ of bias in the column space of $\boldsymbol{F}_0$ (introduced by $\boldsymbol{\theta}_0$), scale (introduced by σ) and rotation in the space of $\boldsymbol{F}_1$. We ask for invariance to the rotation of $\boldsymbol{F}_1$ in an effort to test whether the magnitude of $\boldsymbol{\theta}_1$ is non-zero (*i.e.,* $\|\boldsymbol{\theta}_1\| = 0$ vs. $\|\boldsymbol{\theta}_0\| = 0$), where $\|\boldsymbol{\theta}_1\| > 0$ would indicate that the data comes from the model under $\mathcal{H}_1$ rather than $\mathcal{H}_0$. $\mathcal{G}_0$ can be concisely written as:

$$\mathcal{G}_0 = \left\{ g_\mu \circ g_\rho \circ g_\sigma \;\middle|\; \begin{array}{l} g_\mu \in \mathcal{G}_{bias,0}, \\ g_\rho \in \mathcal{G}_{rotate,1}, \\ g_\sigma \in \mathcal{G}_{scale} \end{array} \right\}$$

where

$$\mathcal{G}_{bias,0} = \left\{ g \mid g(\boldsymbol{y}) = \boldsymbol{y} + \boldsymbol{F}_0\mu, \mu \in \mathbb{R}^k \right\}$$

$$\mathcal{G}_{rotate,1} = \left\{ g \;\middle|\; \begin{array}{l} g(\boldsymbol{y}) = \left(\boldsymbol{P}_1 + \boldsymbol{U}_1\boldsymbol{R}\boldsymbol{U}_1^\top\right)\boldsymbol{y}, \\ \boldsymbol{P}_1 = \boldsymbol{I} - \boldsymbol{F}_1(\boldsymbol{F}_1^\top\boldsymbol{F}_1)^{-1}\boldsymbol{F}_1^\top \\ \boldsymbol{R}^{-1} = \boldsymbol{R}^\top,\ \det(\boldsymbol{R}) = 1 \end{array} \right\}$$

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

and $\boldsymbol{U}_1$ comes from the singular value decomposition of $\boldsymbol{F}_1 = \boldsymbol{U}_1\boldsymbol{\Lambda}\boldsymbol{V}_1^\top$ for some $\boldsymbol{V}_1$.

Given $\mathcal{G}_0$, it is possible to design a maximally invariant statistic as follows [25]:

$$r_0(\boldsymbol{y}) = \frac{\boldsymbol{y}^\top\boldsymbol{P}_{\bar{\boldsymbol{F}}_1}\boldsymbol{y}/rank(\bar{\boldsymbol{F}}_1)}{\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_{\bar{\boldsymbol{F}}_1})\boldsymbol{y}/nullity(\bar{\boldsymbol{F}}_1)}, \tag{4}$$

where $\bar{\boldsymbol{F}}_1 = \left(\boldsymbol{I} - \boldsymbol{F}_0(\boldsymbol{F}_0^\top\boldsymbol{F}_0)^{-1}\boldsymbol{F}_0^\top\right)\boldsymbol{F}_1$, $\boldsymbol{P}_{\bar{\boldsymbol{F}}_1} = \bar{\boldsymbol{F}}_1\left(\bar{\boldsymbol{F}}_1^\top\bar{\boldsymbol{F}}_1\right)^{-1}\bar{\boldsymbol{F}}_1^\top$, $rank(\bar{\boldsymbol{F}}_1)$ denotes the rank of $\bar{\boldsymbol{F}}_1$, and $nullity(\bar{\boldsymbol{F}}_1)$ is the dimension of the null space of $\bar{\boldsymbol{F}}_1$. If the noise $\boldsymbol{n}$ is Gaussian, then under $\mathcal{H}_0$, $r_0(\boldsymbol{y})$ is distributed according to an F-distribution with $rank(\bar{\boldsymbol{F}}_1)$ numerator and $nullity(\bar{\boldsymbol{F}}_1)$ denominator degrees of freedom, respectively. **Regardless of whether $\boldsymbol{n}$ is Gaussian**, a maximally invariant test can be designed using $r_0(\boldsymbol{y})$ that has a constant false alarm rate for any values of the unknown parameters. Finally, note that one can also create a statistic $r_1(\boldsymbol{y})$ for testing $\mathcal{H}_1$ against $\mathcal{H}_0$ by following the same steps and replacing $\boldsymbol{F}_0$ with $\boldsymbol{F}_1$; such a two-sided test could be used in scenarios where multiple models could match the event hypothesis (e.g., comorbidities) and would allow us not to raise unnecessary alarms.

III. Problem Statement

This section presents the context-aware parameter-invariant detection problem. The problem has two parts – 1) context formulation and 2) a corresponding parameter-invariant test. The next subsection states the former first.

A. Context Definition

As described in the introduction, context is intuitively defined as additional information that is not directly used as measurement data but is related to the measurements in an unknown but structured way. In this paper, we are specifically interested in binary context that informs the system that the available measurements might be inaccurate. As argued in Section I and as illustrated in Figure 1, examples of such context measurements include known properties of MCPS (*e.g.,* missing PEEP/PIP measurements in the data recorded from an anesthesia machine are correlated with inaccurate tidal volume measurements since the machine might be temporarily improperly connected).

[3]It is customary to assume that $\boldsymbol{n}$ has a Gaussian distribution because in that case it is possible to compute the distribution of the resulting maximally invariant statistic. However, even if $\boldsymbol{n}$ is not Gaussian (as is likely), we would still obtain a maximally invariant test assuming that $\boldsymbol{n}$ is zero-mean (if the mean is non-zero, one could estimate it from data and reduce the problem to the zero-mean problem by subtracting the estimated mean)

Thus, the first problem addressed in this paper is how to define context measurements. Note that such a definition need not be tied to the parameter-invariant detector but rather the opposite: a reasonable definition should be formulated to capture the notion of context first, which can then be utilized in a corresponding hypothesis test. In this sense, such a test may or may not be solved using a PAIN detector.

Problem. *The first problem addressed in this paper is how to define context, with respect to a hypothesis testing problem, in a general manner in order to capture the intuition that context measurements are correlated with more/less noisy measurements.*

B. Context-Aware Parameter-Invariant Detector

Once context is defined for a hypothesis testing problem, a corresponding hypothesis test can be formulated. This test will also contain unknown parameters, namely all parameters contained in the original hypothesis test in (3) plus any extra parameters that might be introduced by the definition of context. Similar to other detectors in the medical setting, our goal would still be to build a PAIN detector that provides a guaranteed level of false alarm rate regardless of the values of the model parameters or of context parameters.

To approach this problem, one would need to first identify a group of transformations induced by the hypothesis testing problem. Once such a group is defined, a maximally-invariant statistic needs to be constructed that will result in a corresponding maximally-invariant test.

Problem. *The second problem statement is to develop the CA-PAIN detector,* i.e., *develop a maximally-invariant test to the group of transformations induced by the context-aware hypothesis testing problem.*

IV. Context Definition

This section presents some considerations when defining context before presenting the actual context definition and the resulting modified hypothesis test.

A. Frequentist vs. Bayesian Approaches

As usual with definitions, the definition of context presents a variant of the frequentist vs. Bayesian dichotomy. On the one hand, with a Bayesian definition, one can capture the precise probabilistic relation between context and the system state and thereby obtain strong theoretical results (e.g., an unbiased state estimator). Such an approach was taken in our prior work on context-aware estimation [9], [10] where we assume that context measurements are binary measurements with a known probability given the state; given this assumption, we developed a closed-form context-aware filter. The context-aware filter was successfully applied to the problem of estimating the patient's blood oxygen content during surgery [11].

On the other hand, the Bayesian approach often requires expert knowledge or rich training data in order to acquire the prior relation between context and the measurements. In applications where this may be challenging one might adopt a frequentist approach instead. In this setting, we do not make any prior assumptions about the distribution of context but rather rely on data in order to make conclusions about the occurrence of context. The detection setting is also naturally suited for a frequentist approach because it is possible to obtain good results without knowledge of certain parameters, as discussed in Section II. In contrast, this is not true for estimation problems where good parameter estimates are a necessary condition for good estimation performance.

B. Context Formulation

Given the considerations in the previous subsection, we take a frequentist approach and make minimal assumptions about the relationship between context and the actual measurements. In other words, our definition aims to only capture the fact that the presence of context measurements increases a measurement's variance – we do not assume anything about the magnitude of that variance nor about the frequency of context measurements.

Formally, we assume each context measurement b_k^i at time k is a binary variable indicating whether context i is present (*e.g.,* PEEP is missing or not missing). The vector $\boldsymbol{b}_k = [b_k^1, b_k^2, \dots]^\top$ contains all context measurements at time k. It is important to emphasize that we do not make any assumptions about the distribution or likelihood of $\boldsymbol{b}_k$ at any given time. Since the presence of each b_k^i suggests that the measurements might be more noisy at that time, under $\mathcal{H}_0$ we capture this formally with an **unknown** scale:

$$y_k = \begin{cases} \mu_{0,k} + \sigma_1 n_k & \text{if } \exists i \ b_k^i = 1 \\ \mu_{0,k} + \sigma_0 n_k & \text{if } \forall i \ b_k^i = 0 \end{cases}, \tag{5}$$

where $\boldsymbol{y} = y_{1:N}$, $\boldsymbol{n} = n_{1:N}$, $\boldsymbol{F}_0\boldsymbol{\theta}_0 = \mu_{0,1:N}$, and N is the total number of measurements as before. A similar relationship for the measurement could be obtained under $\mathcal{H}_1$.

Having defined context, we can now state the modified hypothesis test. The test is the same as the one in (3) except that some measurements might have different variances depending on whether context is present or not. Thus, the hypothesis test for $\boldsymbol{y}$ can be succinctly written as:

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{y} = \boldsymbol{F}_0\boldsymbol{\theta}_0 + \sigma_0\boldsymbol{D}_0\boldsymbol{n} + \sigma_1\boldsymbol{D}_1\boldsymbol{n} \\ \mathcal{H}_1 &: \boldsymbol{y} = \boldsymbol{F}_1\boldsymbol{\theta}_1 + \sigma_0\boldsymbol{D}_0\boldsymbol{n} + \sigma_1\boldsymbol{D}_1\boldsymbol{n}, \end{aligned} \tag{6}$$

where $\boldsymbol{D}_0$ is a diagonal selection matrix such that $[D_0]_{jj} = 1$ if $\forall i, b_j^i = 0$ and $[D_0]_{jj} = 0$, otherwise; $\boldsymbol{D}_1$ is similarly defined such that $\boldsymbol{D}_0 + \boldsymbol{D}_1 = \boldsymbol{I}$.[4] Note that (6) simplifies to the standard hypothesis test in (3) when either $\boldsymbol{D}_0$ or $\boldsymbol{D}_1$ is the identity matrix.

V. Design of the Context-Aware Parameter-Invariant Detector

Given the context definition presented in Section IV, this section develops the CA-PAIN detector for the hypothesis testing problem in (6).

[4]Note that to simplify notation we use only two values for the noise variance, i.e., all context measurements result in the same high variance, if present. The framework in this paper could be straightforwardly extended to handle the case with multiple variance values by introducing more $\boldsymbol{D}_i$'s.

A. Group of Transformations Induced by Nuisance Parameters

To construct a CA-PAIN detector, we first need to specify the group that we ask for invariance to. Similar to other PAIN detectors, this group needs to capture the transformations induced by the *nuisance parameters* in the hypothesis testing problem. This will ensure that, if an invariant statistic is used to build the detector, then the false alarm rate will be invariant to the specific values of the nuisance physiological parameters.

The new group of transformations induced under $\mathcal{H}_0$, namely $\mathcal{G}_{0,C}$, is similar to $\mathcal{G}_0$; the only difference is that it also contains the multi-scale terms:

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

where $\mathcal{G}_{bias,0}$ and $\mathcal{G}_{rotate,1}$ are as defined in Section II, and $\mathcal{G}_{multi-scale}$ is defined as follows:

$$\mathcal{G}_{multi-scale} = \{g \mid g(\boldsymbol{y}) = (\alpha_0 \boldsymbol{D}_0 + \alpha_1 \boldsymbol{D}_1)\, \boldsymbol{y}, \alpha_0, \alpha_1 \in \mathbb{R}_{\geq 0}\}. \tag{7}$$

The group $\mathcal{G}_{1,C}$ for testing $\mathcal{H}_1$ against $\mathcal{H}_0$ in a two-sided test can be constructed in a similar manner.

B. CA-PAIN Statistics

This subsection presents the CA-PAIN statistics used in this paper. First note that finding a maximally-invariant statistic to $\mathcal{G}_{0,C}$ is challenging. The different scaling terms essentially induce an additional rotation between their respective spaces but it is difficult to identify the invariance that is preserved across all transformations in the group (in a single-scale setting, such an invariance is direction such that a statistic that normalizes the measurement vector to unit length is maximally invariant to the group of single scaling). It remains future work to identify a maximally invariant statistic for this group, if one exists.

As a result of the challenge described in the previous paragraph, in this paper we utilize near-maximally invariant statistics. In particular, we split the measurement vector $\boldsymbol{y}$ into two, $\boldsymbol{y}_0 = \boldsymbol{D}_0\boldsymbol{y}$ and $\boldsymbol{y}_1 = \boldsymbol{D}_1\boldsymbol{y}$. For each $\boldsymbol{y}_i$, we construct the statistic $t_{0i} := r_0(\boldsymbol{y}_i)$ as defined in (4). Note that t_{0i} are invariant to $\mathcal{G}_{0,C}$ but are not maximal. By definition, each t_{0i} is also maximally invariant to a corresponding subgroup of $\mathcal{G}_0$ (by projecting the group into the space spanned by $\boldsymbol{D}_i$). The benefit of using t_{0i} (and t_{1i}, respectively) is that they are independent statistics (because different measurements are used to generate them); this means that we can compute their joint likelihood (e.g., in the case of Gaussian noise) and can construct tests that make decisions only when both statistics (independently) agree, i.e., provide evidence in favor of the same hypothesis.

C. Decision Space

Without obtaining a maximally invariant statistic to $\mathcal{G}_{0,C}$, we cannot construct a new maximally invariant test. However, the information provided by the CA-PAIN statistics, i.e., the t_{0i} and t_{1i} statistics, can still be useful, especially when they agree. Thus, we construct the CA-PAIN detector by augmenting the original PAIN detector with the information provided by the CA-PAIN statistics. In particular, the CA-PAIN detector uses the information provided by the CA-PAIN statistics to silence some of the alarms of the PAIN detector in case the two sets of statistics provide information in favor of opposite hypotheses. The entire decision space of the CA-PAIN detector is detailed below.

The PAIN detector uses a two-sided test to make a decision. For any chosen thresholds η_0 and η_1, the PAIN detector outputs one of three decisions:

1) raise an alarm if $r_0(\boldsymbol{y}) > \eta_0$ and $r_1(\boldsymbol{y}) \leq \eta_1$ (i.e., they agree in favor of $\mathcal{H}_1$),
2) raise a warning if $r_0(\boldsymbol{y}) > \eta_0$ and $r_1(\boldsymbol{y}) > \eta_1$ or if $r_0(\boldsymbol{y}) \leq \eta_0$ and $r_1(\boldsymbol{y}) \leq \eta_1$ (i.e., they disagree),
3) remain silent if $r_0(\boldsymbol{y}) \leq \eta_0$ and $r_1(\boldsymbol{y}) > \eta_1$ (i.e., they agree in favor of $\mathcal{H}_0$).

The CA-PAIN statistics can be used in a similar fashion to make a decision. Since t_{00} and t_{01} are independent (and so are t_{10} and t_{11}), we can compute (assuming the noise $\boldsymbol{n}$ is Gaussian) the value of the joint cumulative distribution function (CDF) of both t_{00} and t_{01}, call it R_0, and of both t_{10} and t_{11}, call it R_1. **Note that the Gaussian assumption is not constraining** – if the CA-PAIN statistics are not F-distributed, the CDF comparison will not yield the targeted false alarm rate but the test will still be invariant to the nuisance parameters. With this in mind, for any chosen thresholds χ_0 and χ_1, the three decisions made using the CA-PAIN statistics are as follows:

1) raise an alarm if $R_0 > \chi_0$ and $R_1 \leq \chi_1$ (i.e., they agree in favor of $\mathcal{H}_1$),
2) raise a warning if $R_0 > \chi_0$ and $R_1 > \chi_1$ or if $R_0 \leq \chi_0$ and $R_1 \leq \chi_1$ (i.e., they disagree),
3) remain silent if $R_0 \leq \chi_0$ and $R_1 > \chi_1$ (i.e., they agree in favor of $\mathcal{H}_0$).

Finally, we construct the CA-PAIN detector by combining the decision of both sets of statistics. In particular, we use the information provided by the CA-PAIN statistics only to silence the PAIN detector, never to raise additional alarms (this is done if PAIN statistics agree in favor of $\mathcal{H}_1$ but the CA-PAIN statistics agree in favor of $\mathcal{H}_0$). The rationale behind this choice is the following: if the PAIN detector raised an alarm during a noisy-data scenario but the CA-PAIN statistics (which decouple high-variance from low-variance measurements) all provide information in favor of $\mathcal{H}_0$, then most likely the PAIN detector raised a false alarm due to the difference in variances (which does not match the model under the null hypothesis). Note that we do not change the PAIN detector's decision in the opposite case because if no PAIN alarm was raised during a high-variance scenario, then the alarm raised by the CA-PAIN statistics is likely to be false.

Tables I, II and III contain a summary of the entire decision space. Table III summarizes the final decision made by the

TABLE I: Test Decision Space for PAIN detector.

	$r_0(\boldsymbol{y}) > \eta_0$	$r_0(\boldsymbol{y}) \leq \eta_0$
$r_1(\boldsymbol{y}) > \eta_1$	D_{PAIN} = **Warning**	D_{PAIN} = **No alarm**
$r_1(\boldsymbol{y}) \leq \eta_1$	D_{PAIN} = **Alarm**	D_{PAIN} = **Warning**

TABLE II: Test Decision Space for CA-PAIN statistics.

	$R_0 > \chi_0$	$R_0 \leq \chi_0$
$R_1 > \chi_1$	$D_{CA-PAIN}$ = **Warning**	$D_{CA-PAIN}$ = **No alarm**
$R_1 \leq \chi_1$	$D_{CA-PAIN}$ = **Alarm**	$D_{CA-PAIN}$ = **Warning**

TABLE III: Test Decision Space for CA-PAIN detector.

	D_{PAIN} = **No alarm**	D_{PAIN} = **Warning**	D_{PAIN} = **Alarm**
$D_{CA-PAIN}$ = **No alarm**	**No Alarm**	**Warning**	**Warning**
$D_{CA-PAIN}$ = **Warning**	**No Alarm**	**Warning**	**Alarm**
$D_{CA-PAIN}$ = **Alarm**	**Warning**	**Warning**	**Alarm**

CA-PAIN detector.

Note that the threshold values are chosen depending on the application and the desired trade-off between a false alarm rate and a detection rate. Since the true probability distribution is never known in practice (and is likely not Gaussian), it is impossible to a priori map threshold values to false alarm rates. At the same time, it is important to emphasize that whatever threshold is chosen, the CA-PAIN detector produces the same false alarm rate regardless of the specific parameter values.

VI. Evaluation

This section evaluates the CA-PAIN detector developed in Section V. We evaluate the performance both in simulation and on real-patient data. Both types of evaluations are based on the critical pulmonary shunt detection problem that was originally used to build the PAIN detector [12], [13]. The following subsection first presents the shunt detection problem before providing both types of evaluations in the subsequent subsections.

A. Prediction of Critical Pulmonary Shunts in Infants

Blood O_2 content is one of the most closely monitored physiological variables during surgery, as too low values can lead to organ failure (*e.g.,* brain damage), and too high values can cause atelectasis (*i.e.,* collapse of the lungs). Oxygen content can drop dangerously low if the patient is experiencing a pulmonary shunt, i.e., the patient is breathing with only one lung. Shunts occur frequently during surgeries with one-lung ventilation; one such example is a lung lobectomy, i.e., the incision of cystic lesion from the patient's lung. Infants are especially vulnerable to shunts because they have underdeveloped lungs. In these patients, breathing with one lung may not supply enough O_2 to the body, thus leading to quick drops in the overall O_2 content.

Monitoring blood O_2 is challenging, however, as it cannot be currently measured non-invasively or in real time. Instead, clinicians monitor the hemoglobin oxygen saturation in the peripheral capillaries, S_pO_2, and use it as a proxy to estimating the overall content. While S_pO_2 is a good measure of the O_2 content in the location at which it is measured (usually a fingertip, or the foot in small infants), it 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.

In order to predict drops in a patient's O_2 content before they are observed through low S_pO_2, in prior work we developed a PAIN detector that utilizes other pulmonary measurements that are provided by the anesthesia machine, namely the partial pressures of O_2 and carbon dioxide (CO_2), tidal volume, and respiratory rate [12], [13]. The PAIN detector is based on a physiological model describing the circulation of O_2 and CO_2 around the cardiovascular and pulmonary systems; the model has the same form as defined in (3):

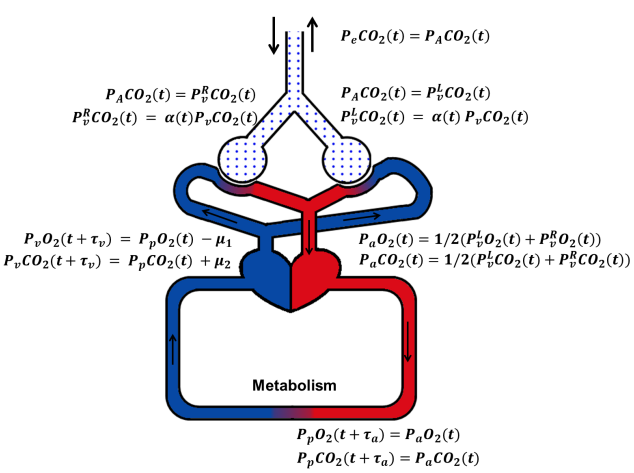
$$\mathcal{H}_0 : \boldsymbol{y} = \boldsymbol{F}_0\boldsymbol{\theta}_0 + \sigma\boldsymbol{n}$$
$$\mathcal{H}_1 : \boldsymbol{y} = \boldsymbol{F}_1\boldsymbol{\theta}_1 + \sigma\boldsymbol{n},$$

where the measurements are the partial pressure of expired CO_2 (denoted by E_tCO_2), whereas the matrices $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$ are constructed using past E_tCO_2 measurements as well as the patient's tidal volume (denoted by V_t) and respiratory rate (denoted by RR). The detector is run in a sliding window fashion (with a window size of 34), so that at each time $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$ contain 34 past measurements. Please consult [12], [13] for a detailed explanation of the physiological model; a short summary is also provided in Figure 2.

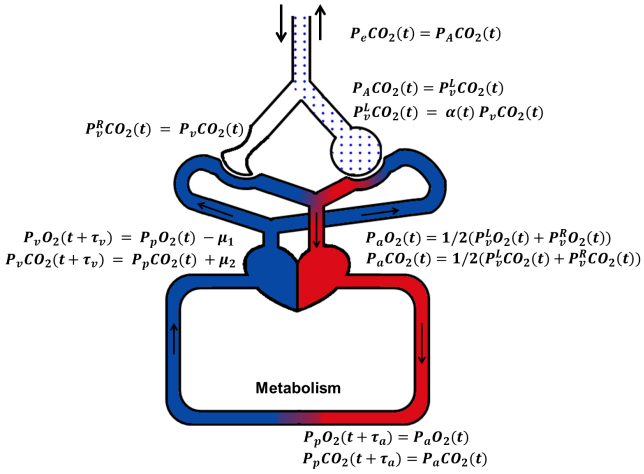
Although the PAIN detector achieved very good performance on the shunt detection problem (achieving about 87% detection rate while introducing on average 1 false alarm per hour), it also produced multiple false alarms caused by bad data. These false alarms were often correlated with missing measurements that are otherwise provided by the anesthesia machine such as the positive end-expiratory pressure (PEEP) and the peak inspiratory pressure (PIP). In the original PAIN detector [12], [13], simple rules are utilized to deal with such noisy-data scenarios, *i.e.,* silence the detector during windows with multiple missing PEEP/PIP values such as the one illustrated in Figure 1. The shortcoming of using simple rule-based detectors is that it blindly pre-determines how context will be handled without considering the data, thus potentially sacrificing valuable testing power for detecting critical pulmonary shunts.

B. Simulation Evaluation

To illustrate the potential benefit of adding context, we first evaluate the CA-PAIN detector in simulations. We simulated 200 patients who did not experience a shunt (*i.e.,* any alarm during those cases would be a false alarm) and 200 patients who experienced a shunt during the case. The physiological model that was used to simulate the patients is the same as



(a) System without a shunt.



(b) System with a shunt.

Fig. 2: Model of the respiratory and cardiovascular partial pressures with and without a shunt.

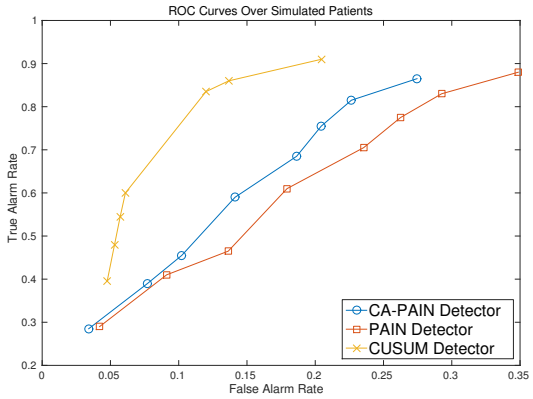


Fig. 3: ROC curves over simulated patients.

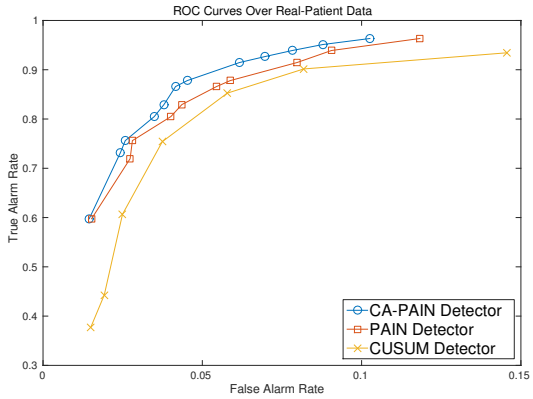


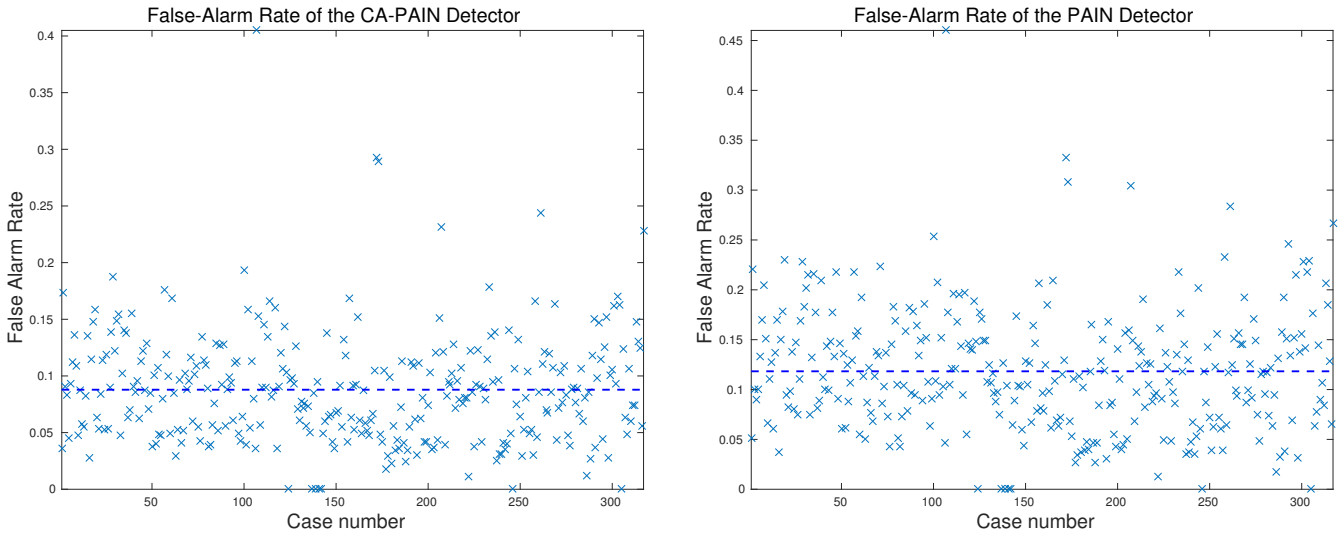
Fig. 4: ROC curves over real-patient data.

the real model described in the previous section and developed in our prior work [12], [13]. Different parameter values were used for different patients such that a wide range of possible patients is covered.

We evaluate the detector's false alarm rate by computing the average false alarm rate over the 200 patients who did not experience an event. The detection rate is computed as the number of good detections over the total number of events; a good detection is any detection made during the period starting five minutes before the event and going up to two minutes after the event.[5] Context was added in a similar fashion to real data: missing PEEP/PIP measurements were introduced with 10% probability at each time; each missing measurement resulted in a five-time increase in the measurement variance, i.e., from 1 to 5.

Figure 3 presents the receiver operating characteristic (ROC) curves achieved by the CA-PAIN and PAIN detectors on simulated data. For better evaluation, we also compare the two detectors with the cumulative sum control chart (CUSUM) detector, which is a standard change detection technique; each patient's parameters were estimated using the expectation maximization technique described in [26]; the detector algorithm was then borrowed from Chapter 8.10 of [8]. In order to obtain different points on the curve, the alarm thresholds η_i and χ_i, $i \in \{0, 1\}$, were varied between 0.0001 and 0.1. ROC curves provide a holistic evaluation of a detector's performance by showing different operating points. A detector is qualitatively better than another one if its ROC curve tends to the upper left corner of the curve since this means higher true alarm rates (and lower false negative rates, equivalently) for the same false alarm rates. As can be observed from Figure 3, the CA-PAIN detector outperforms the PAIN detector for all detection rates, and the reduction in false alarm rates is around 20% on average, especially in the upper section in the curve, which is likely to contain a desired operating point. In this controlled simulated setting, the CUSUM detector outperforms both PAIN detectors since it is able to obtain good parameter estimates and benefits from the model's good predictive power.

[5]After consultation with our clinician collaborators, we established that alarms during this period would be deemed clinically useful.



(a) False alarm rates for each patient as produced by the CA-PAIN detector.

(b) False alarm rates for patient as produced by the PAIN detector.

Fig. 5: False alarm rates for each patient as produced by the two PAIN detectors for a chosen operating point with similar detection performance. Dashed lines denote the average false alarm rate across all patients.

C. Real-Data Evaluation

This subsection provides real-data evaluation of the CA-PAIN detector. The data was collected during lung lobectomy surgeries at CHOP during the period 2005-14. The dataset contains a very diverse population and is a great application for the CA-PAIN detector; it consists of 484 children ranging from a-few-days to several-years old. 167 of these cases had shunts; the remaining 317 cases did not experience one-lung ventilation, i.e., any alarms during those cases would be considered false. A case typically lasts around two to three hours, although some were much longer due to complications; considering that vital signs were sampled every 15 seconds, each case generates around several hundred measurements in total. Note that some cases were eliminated since the event occurred too early or during periods with no data; overall, we retain 82 cases with events. It is important to emphasize that the original PAIN detector only used 62 cases – 20 cases were excluded because the event occurred during missing PEEP/PIP events; due to its formal treatment of context, the CA-PAIN detector is able to make a decision during those cases with the same detection performance as during the other events.

We evaluate the CA-PAIN detector in a similar fashion to the simulation setting. The detector's false alarm rate is the average false alarm rate over all cases with no shunts. The detection rate is the number of good detections divided by the total number of events (with good detections defined in the same way). Figure 4 presents the results, where alarm thresholds were varied between 0.0001 and 0.1. We observe similar patterns to the simulation setting. The CA-PAIN detector outperforms the PAIN detector at all detection rates, and once again the reduction in false alarm rates for all operating points above 80% is at least 20%. Equivalently, the CA-PAIN detector correctly detects an additional 5% of life-critical events without increasing the false alarm rate. Finally, the CUSUM detector performs poorly as caused by wrong parameter estimates and by the model's weak predictive power. This result demonstrates the power of PAIN detectors: although they have weaker theoretical performance on average, their guarantees make them very suitable for scenarios with inaccurate models and patient variability.

To further evaluate the CA-PAIN detector, we also inspect the distribution of false alarms for a specific operating point from Figure 4. Figure 5 shows the distribution of the false alarm rates of both detectors for operating points with similar detection performance. As is clear from the figures, the CA-PAIN detector improves several of the cases with highest false alarm rates; in particular, there are only six cases with a false alarm rate higher than 20% under the CA-PAIN detector, whereas there are more than 20 such cases under the PAIN detector. Furthermore, upon closer inspection, the two biggest outliers both have almost all PEEP/PIP values missing – such a scenario invalidates the benefit of context because the CA-PAIN detector reduces to the PAIN detector (i.e., all measurements are considered to have the same variance). At the same time, **both detectors result in relatively uniformly spread false alarm rates**, with very few outliers, which is one of the most appealing features of the PAIN detector family.

VII. Discussion

This section provides a discussion on the benefit of using context in the MCPS domain and in general CPS applications.

For the application considered in this paper, incorporating context results roughly in a 20-percent reduction of false alarm rates for high-detection-rate operating points. Equivalently, by lowering the alarm threshold, the CA-PAIN detector leads to a 5-percent increase in the number of life-critical detections without increasing the false alarm rate. There are two ob-

servations that are worth emphasizing about this result. The first is that the CA-PAIN statistics used in this work are not maximally invariant, so the results are likely to improve somewhat with the addition of better (*i.e.,* more maximal) invariant statistics. The second point is that context provides coarse information, so it cannot be expected to turn a mediocre detector into a great detector. It is likely that the biggest benefit of context is improving worst cases – e.g., if a patient absolutely does not match the detector's model, then it might be possible to use context to warn clinicians accordingly.

We also emphasize that context is a very general concept that can be applied to many other CPS domains. One example domain might be autonomous vehicles – these systems have multiple information sources and many opportunities for extracting context. A simple context application would be to use a humidity sensor to detect a fog – if a fog is detected, the system can conclude that its object detection algorithm (based on camera measurements) will be more noisy (and possibly useless) as long as the vehicle is moving through the fog.

VIII. Conclusion

This paper presented the problem of context-aware detection in the MCPS domain. We formalized the notion of context as additional information that, if present, results in a change in the measurement model (e.g., higher variance). Based on this definition, we developed the CA-PAIN detector that utilizes both parameter-invariant and context-aware statistics. The CA-PAIN detector was evaluated both in simulation and on real-patient data, and it was shown to outperform the original PAIN detector at all operating points. The immediate future work would be to better understand the challenges associated with obtaining a maximally-invariant context-aware statistic. More globally, a main avenue for future work is formulating context in a general CPS setting and understanding its benefits and limitations.

References

[1] The Dräger Apollo Anesthesia Machine. http://www.draeger.com/sites/enus_us/Pages /Hospital/Apollo.aspx.

[2] N. R. Ahmed, E. M. Sample, and M. Campbell. Bayesian multicategorical soft data fusion for human–robot collaboration. *Robotics, IEEE Transactions on*, 29(1):189–206, 2013.

[3] D. Arney, M. Pajic, J. Goldman, I. Lee, R. Mangharam, and O. Sokolsky. Toward patient safety in closed-loop medical device systems. In *Proceedings of the 1st International Conference on Cyber-Physical Systems*, pages 139–148, 2010.

[4] N. Atanasov, M. Zhu, K. Daniilidis, and G. Pappas. Semantic localization via the matrix permanent. In *Robotics: Science and Systems*, 2014.

[5] A. N. Bishop and B. Ristic. Fusion of spatially referring natural language statements with random set theoretic likelihoods. *Aerospace and Electronic Systems, IEEE Transactions on*, 49(2):932–944, 2013.

[6] C. Cobelli, E. Renard, and B. Kovatchev. Artificial pancreas: past, present, future. *Diabetes*, 60(11):2672–2682, 2011.

[7] Frost & Sullivan. Drowning in Big Data? Reducing Information Technology Complexities and Costs For Healthcare Organizations. Technical report, 2012.

[8] F. Gustafsson. *Adaptive filtering and change detection*, volume 1. Wiley New York, 2000.

[9] R. Ivanov, N. Atanasov, M. Pajic, G. Pappas, and I. Lee. Robust estimation using context-aware filtering. In *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 590–597, 2015.

[10] R. Ivanov, N. Atanasov, M. Pajic, J. Weimer, G. Pappas, and I. Lee. Continuous estimation using context-dependent discrete measurements. In *IEEE Transactions on Automatic Control*. Accepted.

[11] R. Ivanov, N. Atanasov, J. Weimer, M. Pajic, A. Simpao, M. Rehman, G. J. Pappas, and I. Lee. Estimation of blood oxygen content using context-aware filtering. In *Proceedings of the 7th International Conference on Cyber-Physical Systems*, page 28. IEEE Press, 2016.

[12] 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)*, pages 110–119, 2015.

[13] 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*, 24(6):1936–1952, Nov 2016.

[14] L. Magni, D. M. Raimondo, L. Bossi, C. Dalla Man, G. De Nicolao, B. Kovatchev, and C. Cobelli. Model predictive control of type 1 diabetes: an in silico trial. *Journal of diabetes science and technology*, 1(6):804–812, 2007.

[15] R. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House, Inc., 2007.

[16] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbes, C. Roxburgh, and A. H. Byers. Big data : The next frontier for innovation, competition, and productivity. Technical Report June, 2011.

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

[18] M. Pajic, R. Mangharam, O. Sokolsky, D. Arney, J. Goldman, and I. Lee. Model-driven safety analysis of closed-loop medical systems. *IEEE Transactions on Industrial Informatics*, 10(1):3–16, 2014.

[19] 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*, 40(1):1–12, 2010.

[20] W. Raghupathi and V. Raghupathi. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1):3, 2014.

[21] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis. Soi-kf: Distributed kalman filtering with low-cost communications using the sign of innovations. *IEEE Transactions on Signal Processing*, 54(12):4782–4795, 2006.

[22] 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.

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

[24] S. Saria, A. K. Rajani, J. Gould, et al. Integration of early physiological responses predicts later illness severity in preterm infants. *Science translational medicine*, 2(48):48ra65, Sept. 2010.

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

[26] R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis*, 3(4):253–264, 1982.

[27] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry. Kalman filtering with intermittent observations. *Automatic Control, IEEE Transactions on*, 49(9):1453–1464, 2004.

[28] N. A. Trayanova. Whole-heart modeling applications to cardiac electrophysiology and electromechanics. *Circulation Research*, 108(1):113–128, 2011.

[29] A. Wald. *Sequential analysis*. Courier Corporation, 1973.

[30] J. Weimer, S. Chen, A. Peleckis, M. R. Rickels, and I. Lee. Physiology-invariant meal detection for type 1 diabetes. *Diabetes Technology & Therapeutics*, 18(10):616–624, 2016.

[31] J. Weimer, R. Ivanov, S. Chen, A. Roederer, O. Sokolsky, and I. Lee. Parameter-invariant monitor design for cyber physical systems. *Proceedings of the IEEE*, PP(99):1–22, 2017.

[32] J. Weimer, R. Ivanov, A. Roederer, S. Chen, and I. Lee. Parameter-invariant design of medical alarms. *IEEE Design & Test*, 32(5):9–16, 2015.

[33] K. Wyffels and M. Campbell. Negative information for occlusion reasoning in dynamic extended multiobject tracking. *Robotics, IEEE Transactions on*, 31(2):425–442, 2015.