2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Robust Sequential Detection of Non-stealthy Sensor Deception Attacks in an Artificial Pancreas System

Fatih Emre Tosun and André Teixeira

Abstract— **This paper considers deterministic sensor deception attacks in closed-loop insulin delivery. Since the quality of decision-making in control systems heavily relies on accurate sensor measurements, timely detection of attacks is imperative. To this end, we consider a model-based anomaly detection scheme based on Kalman filtering and sequential change detection. In particular, we derive the minimax robust CUSUM and Shewhart tests that minimizes the worst-case mean detection delay and maximizes the instant detection rate, respectively. As a byproduct of our analysis, we show that the notorious χ^2 test shares an interesting optimality property with the two-sided Shewhart test. Finally, we show that one-sided sequential detectors can significantly improve sensor anomaly detection for preventing overnight hypoglycemia which can be fatal.**

I. Introduction

Sequential change detection (SCD) has a wide variety of applications including anomaly detection in control systems. SCD involves detecting an abrupt change in the distribution of a random observation sequence at an unknown point in time. The goal is then to detect the change as early as possible with respect to a certain criterion subject to a specified false alarm constraint. In this work, we consider state-of-the-art SCD methods for detecting non-stealthy false data injection (FDI) attacks on the glucose sensor of an automated insulin delivery system termed the artificial pancreas (AP).

AP systems automatically regulate the blood glucose (BG) levels in individuals with type 1 diabetes (T1D) by means of a portable insulin infusion pump. The required insulin rate is determined based on real-time glucose measurements provided by the continuous glucose monitor (CGM). Due to closed-loop operation, the safety and efficacy of an AP are highly dependent on accurate CGM readings. Hence, it is of utmost importance to detect erroneous readings and mitigate their effect in a timely manner. The bulk of the previous work on detecting CGM failures utilizes data-driven methods that solely depend on sensor data such as support vector machines [1], wavelets [2], and principal component analysis [3]. In [4], the authors proposed a hybrid physiological model and data-driven scheme for the detection and diagnosis of CGM failures. The fundamental difference between the aforementioned works and ours is that they consider natural sensor failures. In contrast, we consider an FDI attack staged by a smart adversary who can exploit the vulnerabilities present in system dynamics.

Model-based anomaly detection in control systems may be divided into two subsequent steps: residual generation and evaluation [5]. In this work, we consider a stochastic linear time-invariant (LTI) model for the AP system. The Kalman filter is a common approach for residual generation in LTI systems, where the filter innovations are selected as the residual signal due to their well-known statistics [6]. The detection of deterministic FDI attacks can then naturally be formulated as an SCD problem where the pre-change (i.e., before the attack) distribution of the residual is known. However, it is not possible to know the post-change distribution without the knowledge of the attack sequence.

In detection theory, there are two main approaches to deal with uncertain post-change distributions: adaptive and minimax detectors [7]. In particular, adaptive detectors such as the generalized likelihood ratio test aim to reduce the uncertainty by estimating the unknown parameters. However, their implementation can be prohibitively complex or unfeasible in real-time embedded systems such as the AP. A minimax detector, on the other hand, aims to guarantee a certain worst-case performance without parameter estimation. Under mild regularity conditions, a minimax detector is a simple detector that is optimal under the least favorable post-change distribution [8]. This approach is better suited for anomaly detection in the AP due to its low complexity.

For residual evaluation, two well-known SCD procedures are the cumulative sum (CUSUM) and Shewhart tests [9]. In particular, the CUSUM test minimizes the worst-case mean detection delay while the Shewhart test maximizes the probability of instant detection for a given false alarm rate (FAR) [10]. The Shewhart test may be preferable over its arguably more popular competitor CUSUM test for deception attacks that are stealthy in the transient phase but become non-stealthy in the steady state. Such examples include replay attacks [11] and constant bias injection attacks if the plant has an integrator [12].

In this work, we define the uncertainty set of possible post-change distributions based on a minimal assumption on the attack (Assumption 1). The main contributions of this paper are summarized as follows:

- We establish the minimax robust optimality of the Shewhart test for detecting mean shifts in Gaussian processes with fixed variance (Proposition 2 and 3).
- We show that the optimality of the Shewhart test is preserved even when the true change parameter does not match the presumed change parameter for imple-

This work was supported by the Swedish Research Council under grant 2018-04396, and by the Swedish Foundation for Strategic Research.

Fatih Emre Tosun is with the Department of Electrical Engineering, Uppsala University, 751 03 Uppsala, Sweden. email: fatihemre.tosun@angstrom.uu.se

André Teixeira is with the Department of Information Technology, Uppsala University, 751 03 Uppsala, Sweden. email: andre.teixeira@it.uu.se

mentation as opposed to the CUSUM test. Furthermore, we establish the equivalence of the two-sided Shewhart and χ^2 tests (Proposition 4). The implication is that the seemingly naive χ^2 test enjoys an interesting optimality property in that it maximizes the instant detection probability at the attack onset for a desired FAR.
- We demonstrate that one-sided SCD tests are superior to their two-sided counterparts in preventing overnight hypoglycemia.

The rest of the paper is organized as follows. Section II presents the SCD formulation of deterministic FDI attacks on the CGM. Section III lays out the theoretical foundations and presents the minimax robust SCD problems for attack detection. Section IV presents the corresponding solutions. Section V demonstrates the effectiveness of the proposed simple detection scheme via numerical simulations. Section VI concludes the paper.

Notation: $\mathbb{N}, \mathbb{R}$ denote the set of natural and real numbers, respectively. R^n denotes the n-dimensional Euclidean space. A^T denotes the matrix transpose of A. $\mathcal{N}(\mu, \mathbf{\Sigma})$ denotes the Gaussian distribution with mean μ and (co)variance $\mathbf{\Sigma}$. The short-hand notations for an ordered sequence and the ReLU function are, respectively as, $x_j^k \triangleq (x_j,\ x_{j+1},\ \ldots,\ x_k)$, $(x)^+ \triangleq \max\{0,\ x\}$.

II. PROBLEM FORMULATION

Since the paper focuses on sensor attacks, we consider the following autonomous discrete-time LTI system:

$$\begin{aligned} x_{k+1} &= Fx_k + w_k \\ y_k &= Hx_k + v_k + a_k \end{aligned} \tag{1}$$

where $k \in \mathbb{N}$ is the discrete-time index, $x_k \in \mathbb{R}^n$ is the state vector, $y_k \in \mathbb{R}$ is the measured output, and $a_k \in \mathbb{R}$ is the FDI on the sensor. The state and the output are driven by mutually independent white Gaussian noises $w_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$, respectively.

We employ the Kalman filter for residual generation. For ease of exposition, we assume the filter has already reached the steady-state. Then, the filter dynamics may be written as:

$$\hat{x}_{k+1} = (F - KH)\hat{x}_k + Ky_k \tag{2}$$

$$K = FPH^TS^{-1} \tag{3}$$

$$S = HPH^T + R \tag{4}$$

$$P = F(P - KHP)F^T + Q \tag{5}$$

where $\hat{x}_k \triangleq \mathbb{E}[x_k \mid y_0^{k-1}]$ is the one-step prediction of x_k, K is the filter gain, and P is the steady-state estimation error covariance that solves the algebraic Riccati equation (5).

The Kalman filter innovation sequence is the difference between the measured and predicted output as follows:

$$r_k \triangleq y_k - H\hat{x}_k. \tag{6}$$

In the absence of anomalies (e.g., FDI), it is well-known that $r_k \sim \mathcal{N}(0, S)$ [13]. On the other hand, when the attack sequence a_k is deterministic, the innovation sequence r_k follows a Gaussian distribution with a time-varying mean with the same (co)variance. The attacked residual mean μ_k evolves as follows:

$$\begin{aligned} \tilde{x}_{k+1}^a &= (F - KH)\tilde{x}_k^a - Ka_k,\ \tilde{x}_0^a = 0 \\ \mu_k &= H\tilde{x}_k^a + a_k \end{aligned} \tag{7}$$

where $\tilde{x}_k^a$ denotes the isolated contribution of the attack to the state estimation error $\tilde{x}_k \triangleq x_k - \hat{x}_k$. The linearity of the Kalman filter is exploited while deriving (7). Hence, non-stealthy FDI detection is an SCD problem of the form:

$$\begin{aligned} H_0 &: r_k \sim \mathcal{N}(0,\ S) \text{ for } 0 \le k \\ H_1 &: r_k \sim \mathcal{N}(0,\ S) \text{ for } 0 \le k < v \\ &\quad r_k \sim \mathcal{N}(\mu_k,\ S) \text{ for } k \ge v \end{aligned} \tag{8}$$

where v is the change point. The null hypothesis H_0 defines the nominal behavior whereas the alternative hypothesis H_1 implies that an anomaly has occurred in the observation sequence r_k. As can be seen from (8), H_1 is rather rich in parameters, which is the major challenge in attack detection. So as to make the detection problem tractable, we introduce a simplifying yet reasonable assumption.

Assumption 1. *For a given attack sequence a_k, the corresponding change in the magnitude of the residual mean is bounded from below such that $|\mu_k| \ge m$.*

In the sequel, we explain how to tackle (8) under this assumption through minimax robust SCD.

III. SEQUENTIAL CHANGE DETECTION

In this section, we formally define SCD procedures as well as their operating characteristics. In order to facilitate understanding, we present the following key definitions [9].

Definition 1 (*Simple and Composite Hypotheses*)**.** *A hypothesis is said to be simple if all parameters of the underlying distribution are specified, and is said to be composite otherwise.*

In other words, the parameter space of a composite hypothesis has at least 2 elements.

Definition 2 (*Stopping Time*)**.** *A stopping time with respect to a random sequence $(X_n)_{n\ge1}$ is a random variable T such that for each discrete-time instant n, the event $\{T = n\}$ belongs to the sigma-algebra generated by X_1^n.*

To put it simply, T does not depend on any future observation $X_{j>n}$. An SCD procedure is simply a stopping time on an observation sequence:

$$T = \inf\{n \ge 1 : d(X_1^n) > \tau\} \tag{9}$$

where $d(\cdot)$ is the test statistic which is a causal function of observations, τ is the decision threshold, and $\inf\{\emptyset\} = \infty$. Thus, T is the number of observations taken until an alarm is raised for the first time.

The stopping time concept is analogous to the impulse response of LTI systems in that it uniquely defines the characteristics of SCD procedures. Thus, the SCD theory is essentially the study of designing optimal stopping times.

The optimality is defined with respect to certain detection performance measures given a false alarm constraint.

For notational brevity, let $\mathbb{P}_v(T)$ be the probability measure on T when the change point is v, and $\mathbb{E}_v[T]$ is the corresponding expectation. Consequently, $\mathbb{P}_\infty(T)$ and $\mathbb{E}_\infty[T]$ are convenient short-hand notations for when no change is present. In the absence of *a priori* information on v, a common and natural metric to quantify the FAR is the average frequency of false alarms as follows [10]:

$$FAR(T) \triangleq 1/\mathbb{E}_\infty[T]. \tag{10}$$

Hence, we define the feasible set for the SCD procedures as:

$$\mathbf{C}_\alpha \triangleq \{T : 0 < FAR(T) \leq \alpha \leq 1\}. \tag{11}$$

The detection delay $(T-v)^+$ is a random variable. Thus, one can either opt to minimize the mean detection delay or maximize the probability of detection within a certain window of length $N+1$. The first can be formulated as the following optimization problem:

$$\begin{array}{ll} \text{minimize} & ADD_v(T) \triangleq \mathbb{E}_v[T-v \mid T \geq v] \\ \text{subject to} & T \in \mathbf{C}_\alpha. \end{array} \tag{12}$$

while the latter can be formulated as:

$$\begin{array}{ll} \text{maximize} & \mathbb{P}_v(T = v+N \mid T \geq v) \\ \text{subject to} & T \in \mathbf{C}_\alpha, \end{array} \tag{13}$$

Clearly, detection performance depends heavily on the unknown change point v. Unfortunately, an SCD procedure that solves (12) or (13) uniformly over all $v \geq 1$ does not exist [9]. Instead, one can employ a minimax approach, that is, optimize with respect to the worst-case scenario. For the first problem, we resort to Lorden's measure of worst-case detection delay [14]:

$$J_d(T,\, \mu_k) \triangleq \sup_{v \geq 1} \left\{ \operatorname{ess\,sup} \mathbb{E}_v \left[(T-v)^+ \mid r_0^{v-1} \right] \right\}. \tag{14}$$

It is basically the maximal mean detection delay conditioned on the worst admissible realizations. As for the second problem, we consider the minimal detection probability originally proposed by Moustakides [15]:

$$J_p(T,\, \mu_k, N) \triangleq \inf_{v \geq 1} \mathbb{P}_v(T = v+N \mid T \geq v). \tag{15}$$

Let $\mathcal{U} \subset \mathbb{R}$ denote the uncertainty set, that is, the set of all values that μ_k can take. Then, we seek the minimax robust SCD procedures that solve the following optimization problems:

$$\begin{array}{ll} \text{minimize} & \sup J_d(T,\, \mu_k) \\ T \in \mathbf{C}_\alpha & \mu_k \in \mathcal{U} \end{array} \tag{16}$$

$$\begin{array}{ll} \text{maximize} & \inf J_p(T,\, \mu_k,\, 0) \\ T \in \mathbf{C}_\alpha & \mu_k \in \mathcal{U} \end{array} \tag{17}$$

In (17), we restrict to the special case when $N = 0$ as no results for the general case $N > 0$ have been reported to our best knowledge. Next, we present the state-of-the-art algorithms for these problems.

A. One-sided Sequential Change Detection

First, let us consider a simple SCD problem with $\mathcal{U} = \{n\}$ where $n \in \mathbb{R}$ is a known constant. Let f_0 and f_1 be the pre- and post-change distributions such that

$$f_0 \triangleq \mathcal{N}(0,\, S),\; f_1 \triangleq \mathcal{N}(n,\, S). \tag{18}$$

The log-likelihood ratio (LLR) of the observation sequence r_k with respect to f_0 and f_1 read as:

$$\ell_{k+1} = log \frac{f_1(r_k)}{f_0(r_k)} = \frac{n r_k - 0.5n^2}{S}. \tag{19}$$

The first procedure we shall present is the Shewhart test, which is a repeated Neyman-Pearson test as follows [9]:

$$T_{sh} = inf\{k \geq 1 : \ell_k > \tau_{sh}\} \tag{20}$$

where T_{sh} denotes the corresponding stopping time. Despite being the earliest SCD procedure, its optimality properties were relatively recently discovered [15]. Pertinent to our work, this test solves (17) when the threshold τ_{cs} is chosen to ensure $FAR(T_{sh}) = \alpha$, or equivalently $\mathbb{P}_\infty(\ell_k > \tau_{sh}) = \alpha$ [16]. In other words, it maximizes the worst-case *instant* detection probability for a given FAR.

Next, we present the CUSUM test, which minimizes the worst-case mean detection delay for a given FAR:

$$T_{cs} = \inf\{k \geq 1 : g_k = (g_{k-1} + \ell_k)^+ > \tau_{cs}\},\, g_0 = 0. \tag{21}$$

Similarly, to solve (16), τ_{cs} must be chosen to ensure $FAR(T_{cs}) = \alpha$ [17]. To determine the value of τ_{cs}, one can either numerically solve the Fredholm integral equations or use Monte Carlo simulations to approximate it [9]. In the next subsection, we consider a natural extension of this simple SCD problem to its composite counterpart to handle bidirectional changes in μ_k.

B. Two-sided Sequential Change Detection

Now suppose we know the magnitude but not the direction of change such that $\mathcal{U} = \{-m,\, m\}$. A simple and intuitive solution is to run two tests for $n = m$ and $n = -m$ in parallel. The LLRs corresponding to the positive and negative changes, respectively, read as:

$$\ell^+_{k+1} = \frac{m r_k - 0.5m^2}{S},\; \ell^-_{k+1} = \frac{-m r_k - 0.5m^2}{S}. \tag{22}$$

Then, the two-sided Shewhart test is defined as:

$$T_{sh2} = \inf\{k \geq 1 : \ell^+_k > \tau_{sh2} \text{ or } \ell^-_k > \tau_{sh2}\}. \tag{23}$$

When $FAR(T_{sh2}) = \alpha$, or equivalently $\mathbb{P}_\infty(\ell^+_k > \tau_{sh2}) = \alpha/2$, it enjoys the same optimality property as the one-sided Shewhart test [16].

Similarly, the two-sided CUSUM test is defined as:

$$\begin{aligned} T_{cs2} &= \inf\{k \geq 1 : g^+_k > \tau_{cs2} \text{ or } g^-_k > \tau_{cs2}\} \\ g^+_{k+1} &= (g^+_k + \ell^+_k)^+, \quad g^+_0 = 0 \\ g^-_{k+1} &= (g^-_k + \ell^-_k)^+, \quad g^-_0 = 0. \end{aligned} \tag{24}$$

This test is asymptotically optimal as $FAR(T_{cs2}) \to 0$ [9]. To our best knowledge, whether the two-sided CUSUM test is globally optimal remains an open problem.

IV. Robust Sequential Change Detection

In this section, we show that when assumption 1 holds, the Shewhart and CUSUM tests tuned for $|\mu_k| = m$ are minimax optimal.

Proposition 1. *Consider the robust SCD problem (16). The one-sided CUSUM test (21) with $FAR(T_{cs}) = \alpha$ is a solution for the uncertainty sets $\mathcal{U} = \{\mu_k : \mu_k \geq n > 0\}$ and $\mathcal{U} = \{\mu_k : \mu_k \leq n < 0\}$.*

Proof: Please see Theorem III.2 in [8]. ■

This result is in line with the intuition that the CUSUM test tuned with respect to the minimum change in magnitude should be worst-case optimal since larger deviations should get detected even faster. Recalling the discussion in Section III-B, the following corollary is easily seen:

Corollary 1.1. *The two-sided CUSUM (24) is asymptotically optimal as for $\mathcal{U} = \{\mu_k : |\mu_k| \geq m\}$ as $FAR(T_{cs2}) \to 0$ in the sense of (16).*

Next, we prove the robust optimality of the Shewhart test in the sense of (17). We begin by introducing the relationship between saddle points and minimax theory.

Definition 3 (*Saddle point*)**.** *The pair $(a^* \in \mathcal{A}, b^* \in \mathcal{B})$ is a saddle point for $f(a,b)$ if $f(a^*,b) \leq f(a^*,b^*) \leq f(a,b^*)$ for all $a \in \mathcal{A}$ and $b \in \mathcal{B}$.*

Lemma 1 (*Minimax optimality* [18])**.** *Suppose we have the following optimization problem:*

$$\begin{array}{ll} \text{maximize} & \inf f(a,b) \\ a \in \mathcal{A} & b \in \mathcal{B}. \end{array}$$

The pair (a^*, b^*) is a solution to this problem if and only if it is a saddle point for $f(\cdot)$.

Proposition 2. *Consider the minimax robust SCD problem (17). The one-sided Shewhart test (20) with $FAR(T_{sh}) = \alpha$ is a solution for the uncertainty sets $\mathcal{U} = \{\mu_k : \mu_k \geq n > 0\}$ and $\mathcal{U} = \{\mu_k : \mu_k \leq n < 0\}$.*

Proof: Let T^n_{sh} denote the stopping time (20) which satisfies $FAR(T_{sh}) = \alpha$. In what follows, we prove $(T^n_{sh}, n, 0)$ is a saddle point for $J_p(T,\, \mu_k,\, 0)$. More precisely, we show

$$J_p(T^{'},\, n,\, 0) \leq J_p(T^n_{sh},\, n,\, 0) \leq J_p(T^n_{sh},\, \mu_k,\, 0)$$

where $T^{'}$ is any stopping time in $\mathbf{C}_\alpha$. Then, the statement of the proposition follows by the virtue of Lemma 1.

As discussed in Section III-A, the left-hand side inequality was proved in [16]. Thus, it suffices to prove the right-hand side inequality. Using $\mathbb{E}[r_k] = \mu_k$ and (19), the LLR sequence is a Gaussian variable as follows:

$$\ell_k \sim \mathcal{N}\left((n\mu_k - 0.5n^2)/S,\, n/\sqrt{S}\right). \tag{25}$$

The right tail probability for $X \sim \mathcal{N}(\mu,\, \sigma^2)$ is equal to

$$\mathbb{P}(X > \tau) = 0.5erfc((\tau - \mu)/\sqrt{2}\sigma) \tag{26}$$

where $erfc(\cdot)$ is the complementary error function. Thus, the instant detection probability of T^n_{sh2} becomes:

$$\begin{aligned} J_p(T^n_{sh},\, \mu_k,\, 0) &= \mathbb{P}_v(\ell_k > \tau_{sh}) \\ &= 0.5erfc\left(\frac{S\tau_{sh} - n\mu_k + 0.5n^2}{n\sqrt{2S}}\right) \end{aligned} \tag{27}$$

Since $erfc(\cdot)$ is monotonically decreasing, (27) attains its minimum at $\mu_k = n$ for both $\mu_k \geq n > 0$ and $\mu_k \leq n < 0$. ■

Next, we establish a similar result for the two-sided Shewhart test.

Proposition 3. *Consider the minimax robust SCD problem (17). The two-sided Shewhart test (23) with $FAR(T_{sh}) = \alpha$ is a solution for the uncertainty set $\mathcal{U} = \{\mu_k : |\mu_k| \geq m\}$.*

Proof: Let T^m_{sh2} denote the stopping time (23) which satisfies $FAR(T_{sh2}) = \alpha$. Then, we show the following holds:

$$J_p(T^{'},\, m,\, 0) \leq J_p(T^m_{sh2},\, m,\, 0) \leq J_p(T^m_{sh2},\, |\mu_k|,\, 0).$$

As should be clear from the previous discussion, proving right-hand side inequality completes the proof. Using (22) and (23), the two-sided Shewhart detection rule may be rewritten as:

$$|r_k| > \frac{S\tau_{sh2}}{m} + 0.5m \iff \frac{r_k^2}{S} > S\left(\frac{\tau_{sh2}}{m} + \frac{0.5m}{S}\right)^2 \triangleq \tau_{\chi^2}. \tag{28}$$

The test statistic r_k^2/S under attack is a one-degree-of-freedom non-central χ^2 distributed random variable with a noncentrality parameter of μ_k^2/S. Consequently, the instant detection probability of T^m_{sh2} read as:

$$J_p(T^m_{sh2},\, |\mu_k|,\, 0) = \mathbb{P}_v(r_k^2/S > \tau_{\chi^2}) = Q_{0.5}(|\mu_k|/\sqrt{S},\, \tau_{\chi^2}) \tag{29}$$

where $Q_v(a,b)$ denotes the generalized Marcum $Q-$function of real order $v > 0$, which is strictly increasing in a [19]. Thus, as in the unidirectional case, (29) attains its minimum value at $|\mu_k| = m$. ■

Propositions 2 and 3 establish the minimax robust optimality of the Shewhart test for detecting mean shifts in Gaussian processes with fixed variance. They verify our intuition that larger deviations should get detected with higher probability. The next proposition shows that this test enjoys a remarkable robustness property.

Proposition 4. *For any fixed FAR of α, the one-sided Shewhart test (20) is equivalent to the simpler test $r_k > \tau$ where $\mathbb{P}_\infty(r_k > \tau) = \alpha$. Similarly, the two-sided Shewhart test (23) is equivalent to the simpler χ^2 test: $r_k^2/S > \tau_{\chi^2}$ where $\mathbb{P}_\infty(r_k^2/S > \tau_{\chi^2}) = \alpha$.*

Proof: By definition, we have $\alpha = P_\infty(r_k > \tau)$. From (26), we get $\tau = \sqrt{2S}\, erfc^-(2\alpha)$. Similarly, for the one-sided Shewhart test, we have $\alpha = P_\infty(\ell_k > \tau_{sh})$ which yields the following relation between τ and τ_{sh}:

$$(S\tau_{sh} + 0.5n^2)/n = \sqrt{2S}\, erfc^-(2\alpha) = \tau. \tag{30}$$

The above equation clearly shows $r_k > \tau \iff l_k > \tau_{sh}$. The proof of Proposition 3 readily establishes the equivalence of the two-sided Shewhart and χ^2 tests. ■

Please note that the detection rules $r_k > \tau$ and $r_k^2/S > \tau_{\chi^2}$ are independent of the post-change distributions. Hence, the optimality of the Shewhart test is insensitive to the true value of the change parameter μ_k. In contrast, the optimality of the CUSUM test holds only when the true change parameter is equal to the presumed value [9].

Remark. *The proof of Proposition 4 is constructed for the scalar case since this paper focuses on the AP where the only measured quantity is BG levels. However, changing to vector notation, the equivalence of the two-sided Shewhart and χ^2 tests can easily be extended to multiple output systems.*

V. Numerical Simulations and Discussion

In this section, we perform numerical simulations to illustrate the analytical results obtained in the previous section and to demonstrate an application in the context of the AP.

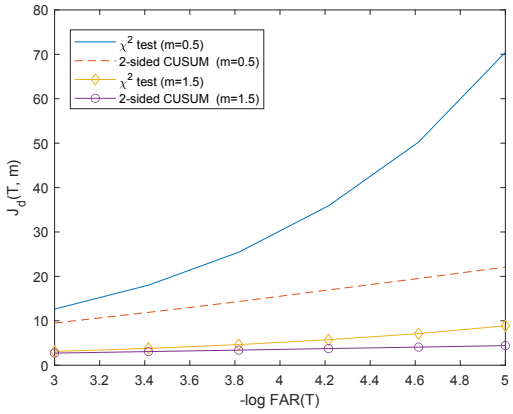
A. Performance Measures

Based on Propositions 3 and 4, a natural question may arise: Should we always employ the χ^2 test to detect bidirectional changes in the mean of white Gaussian processes? The answer is no, because the χ^2 test is optimal with respect to the instant detection criterion which is too stringent. In particular, the maximum achievable instant detection rate becomes so low that this performance measure is no longer practical for the cases where the magnitude of change and/or FAR is too low. To illustrate this, we consider a scenario where the pre-change distribution is $\mathcal{N}(0,\ 1)$ and the post-change distribution is $\mathcal{N}(m,\ 1)$. Then, we compute the two performance measures via Monte Carlo simulations: $J_d(T, m)$ as in (14) and $J_p(T, m, 0)$ as in (15) for the χ^2 and CUSUM tests with a moderate and a large value of m.

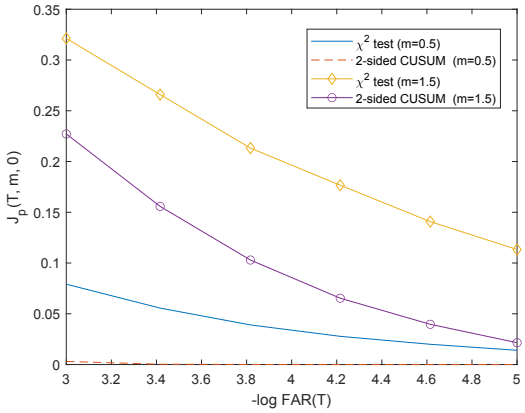
The results are reported in Fig. 1a and 1b, respectively. As stipulated by Proposition 4, the χ^2 test has a higher instant detection rate than that of the CUSUM test at all times as shown in Fig 1b. However, when $m = 0.5$ and the FAR is 0.5 %, the instant detection rate of the χ^2 test is less than 3% while the detection delay is as high as 70 samples as shown in Fig 1a. Clearly, this detection performance is not acceptable for most applications. On the other hand, when $m = 1.5$, the detection delays of the χ^2 and the CUSUM tests are comparable. Thus, for sufficiently large changes, the χ^2 detector may be preferable over the celebrated CUSUM detector as it maximizes the probability of detection right at the attack onset at the expense of a slightly higher worst-case mean detection delay.

B. Overnight Hypoglycemia Prevention

In this subsection, we show an application of the proposed model-based anomaly detection scheme for overnight hypoglycemia prevention. Hypoglycemia is a state of having too low BG levels which can be lethal. Automatic suspension of insulin delivery before entering hypoglycemia is the first step of the six-step pathway toward a fully automated AP



(a) Worst-case mean detection delay versus -log(FAR)



(b) Worst-case instant detection rate versus -log(FAR)

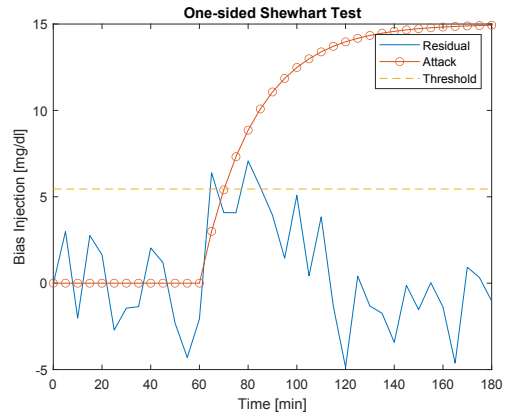
Fig. 1: Performance comparisons of the χ^2 and two-sided CUSUM tests for a small and a large change.

proposed by the Juvenile Diabetes Research Foundation [20]. We consider the model developed by Bequette to predict the future sensor glucose trajectory with the aid of a Kalman filter [21]. The model parameters are summarized as follows:

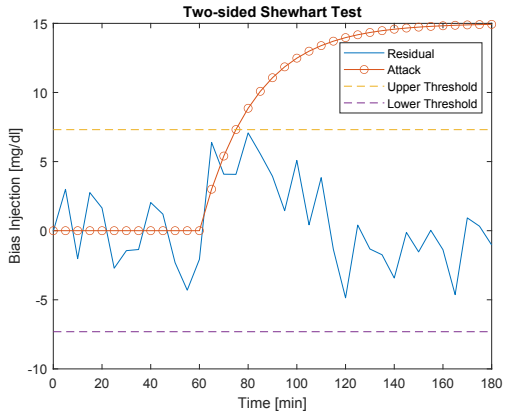
$$F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\ Q = \begin{pmatrix} 0 & 0 \\ 0 & 0.01 \end{pmatrix},\ H = \begin{pmatrix} 1 & 0 \end{pmatrix},\ R = 4$$

Based on this simple model, an automatic insulin pump shut-off algorithm has been developed and tested with 15 real patients in a hospital setting [20]. The algorithm successfully prevented hypoglycemia in 11 patients. Three of the four unsuccessful cases were due to a positive bias in the CGM. Had these sensor failures been detected earlier, the algorithm could have also prevented hypoglycemia in those cases.

Please note that F has eigenvalues at unity which can be exploited by the attacker. In particular, a constant bias injection attack is steady-state stealthy in the sense that $\lim_{k\to\infty} \mu_k = 0$ if the plant has an integrator [12]. Moreover, this holds for any linear observer not just for the Kalman filter. Clearly, in this case, bias injection is a good strategy for an FDI attack since an arbitrarily large bias can be injected without risking detection so long as it is not detected in the transient phase of the filter response. Thus, we assume the attacker designs a slowly increasing bias injection attack to



(a) One-sided Shewhart test



(b) Two-sided Shewhart test

Fig. 2: Residual evaluation with one- and two-sided Shewhart tests for detecting a bias injection attack.

remain stealthy during the transients. In this case, the attack sequence reads as follows:

$$a_{k+1} = \beta a_k + (1-\beta)\bar{a} \tag{31}$$

where β determines the speed of injection, $\bar{a}$ determines the final value of the injected bias. In the simulations, the attack starts at 30 min with the following parameter values $\bar{a} = 15$ mg/dl and $\beta = 0.2$. The attack will be detectable only for a short period of time since the effect of the injected bias on the residual statistics quickly vanishes. Since the Shewhart test maximizes the instant detection rate, it is more suitable for detecting the attack during transients. We apply one- and two-sided Shewhart tests on the innovations with an FAR of $\alpha = 1\%$. Fig. 2 demonstrates the outcomes of a particular realization of the simulations for this scenario. The one-sided test was able to detect the attack instantly while the two-sided test missed detection.

VI. Conclusions and Future Work

In this work, we investigated a common model-based anomaly detection scheme against non-stealthy sensor deception attacks in AP systems. The detection scheme is based on Kalman filtering and state-of-the-art minimax robust SCD. We established an interesting robust optimality property of the celebrated χ^2 detector. In a sense, the χ^2 test maximizes the best-case detection delay (i.e., instant detection) while the CUSUM test minimizes the worst-case delay. Finally, we advocated the use of a one-sided Shewhart test for preventing overnight hypoglycemia. In future work, we plan to study robust sequential detection in the presence of nuisance parameters and provide more insights into the tradeoff between robust detectability and the impact of the attack.

References

[1] Y. Leal, L. Gonzalez-Abril, C. Lorencio, J. Bondia, and J. Vehi, "Detection of correct and incorrect measurements in real-time continuous glucose monitoring systems by applying a postprocessing support vector machine," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1891–1899, 2013.

[2] Q. Shen, S. J. Qin, and K. J. Doniger, "Online dropout detection in subcutaneously implanted continuous glucose monitoring," in *Proceedings of the 2010 American Control Conference*, 2010, pp. 4373–4378.

[3] Y. Leal, M. Ruiz, C. Lorencio, J. Bondia, L. Mujica, and J. Vehi, "Principal component analysis in combination with case-based reasoning for detecting therapeutically correct and incorrect measurements in continuous glucose monitoring systems," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 603–614, 2013.

[4] K. Turksoy, A. Roy, and A. Cinar, "Real-time model-based fault detection of continuous glucose sensor measurements," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1437–1445, 2016.

[5] S. X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2013.

[6] H. Sandberg, V. Gupta, and K. H. Johansson, "Secure networked control systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 445–464, 2022.

[7] M. Fauß, A. M. Zoubir, and H. V. Poor, "Minimax robust detection: Classic results and recent advances," *IEEE Transactions on signal Processing*, vol. 69, pp. 2252–2283, 2021.

[8] J. Unnikrishnan, V. V. Veeravalli, and S. P. Meyn, "Minimax robust quickest change detection," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1604–1614, 2011.

[9] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[10] L. Xie, S. Zou, Y. Xie, and V. V. Veeravalli, "Sequential (quickest) change detection: Classical results and new directions," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 2, pp. 494–514, 2021.

[11] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2009, pp. 911–918.

[12] J. Milošević, T. Tanaka, H. Sandberg, and K. H. Johansson, "Analysis and mitigation of bias injection attacks against a kalman filter," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8393–8398, 2017.

[13] B. D. Anderson and J. B. Moore, *Optimal filtering*. Courier Corporation, 2012.

[14] G. Lorden, "Procedures for reacting to a change in distribution," *The annals of mathematical statistics*, pp. 1897–1908, 1971.

[15] G. V. Moustakides, "Multiple optimality properties of the shewhart test," *Sequential Analysis*, vol. 33, no. 3, pp. 318–344, 2014.

[16] M. Pollak and A. M. Krieger, "Shewhart revisited," *Sequential Analysis*, vol. 32, no. 2, pp. 230–242, 2013.

[17] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *the Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.

[18] D. Bertsekas, *Convex optimization theory*. Athena Scientific, 2009, vol. 1.

[19] Y. Sun, Baricz, and S. Zhou, "On the monotonicity, log-concavity, and tight bounds of the generalized marcum and nuttall q-functions," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1166–1186, 2010.

[20] B. W. Bequette, F. Cameron, B. A. Buckingham, D. M. Maahs, and J. Lum, "Overnight hypoglycemia and hyperglycemia mitigation for individuals with type 1 diabetes: how risks can be reduced," *IEEE Control Systems Magazine*, vol. 38, no. 1, pp. 125–134, 2018.

[21] B. W. Bequette, "Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms," *Journal of diabetes science and technology*, vol. 4, no. 2, pp. 404–418, 2010.