Journal of Process Control 135 (2024) 103162

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Quickest detection of bias injection attacks on the glucose sensor in the artificial pancreas under meal disturbances

Fatih Emre Tosun [a,*], André M.H. Teixeira [b], Mohamed R.-H. Abdalmoaty [c], Anders Ahlén [a], Subhrakanti Dey [a]

[a] Department of Electrical Engineering, Uppsala University, Uppsala, 751 03, Sweden
[b] Department of Information Technology, Uppsala University, Uppsala, 751 03, Sweden
[c] Automatic Control Laboratory, ETH Zurich, Physikstrasse 3, Zurich, 8092, Switzerland

ARTICLE INFO

Keywords:
Type 1 diabetes mellitus
Artificial pancreas
Quickest change detection
Control-theoretic security
Sensor deception attack

ABSTRACT

Modern glucose sensors deployed in closed-loop insulin delivery systems, so-called artificial pancreas use wireless communication channels. While this allows a flexible system design, it also introduces vulnerability to cyberattacks. Timely detection and mitigation of attacks are imperative for device safety. However, large unknown meal disturbances are a crucial challenge in determining whether the sensor has been compromised or the sensor glucose trajectories are normal. We address this issue from a control-theoretic security perspective. In particular, a time-varying Kalman filter is employed to handle the sporadic meal intakes. The filter prediction error is then statistically evaluated to detect anomalies if present. We compare two state-of-the-art online anomaly detection algorithms, namely the χ^2 and CUSUM tests. We establish a robust optimal detection rule for unknown bias injections. Even if the optimality holds only for the restrictive case of constant bias injections, we show that the proposed model-based anomaly detection scheme is also effective for generic non-stealthy sensor deception attacks through numerical simulations.

1. Introduction

Type 1 diabetes (T1D) is an autoimmune disease in which insulin secretion is lost due to the self-destruction of pancreatic beta cells. In 2021, a study estimated that there were 8.4 million individuals worldwide affected by T1D [1]. Furthermore, this number was projected to increase up to 17.4 million in 2040. Thus, treatment of T1D is of paramount importance. Patients with T1D require exogenous insulin administration to maintain their blood glucose (BG) levels in a safe range. This is achieved by either multiple daily injections of long-acting insulin or by continuous infusion of rapid-acting insulin via a portable pump. The latter offers more flexibility to patients in their social lives and also enables tighter control of the BG levels [2]. Moreover, thanks to the advances in wearable medical devices and wireless communication technologies, insulin pump therapy can be conveniently automated. In fact, the concept of automated insulin delivery systems termed the artificial pancreas (AP), has been an active research endeavor since the 1960s [3]. Finally, in 2016, the U.S. Food and Drug Administration (FDA) approved a commercial AP for the first time [4]. The AP is essentially a control system consisting of three main components: an insulin pump as the actuator, a continuous glucose monitor (CGM) as the sensor, and an embedded controller. The control objective is to maximize the time spent in the normoglycemic range, which is typically 70–140 mg/dl for preprandial (fasting) glucose, and less than 180 mg/dl for postprandial glucose [5]. The pump dispenses a proper amount of insulin to regulate the BG levels as dictated by the closed-loop controller. The infusion rate is determined based on real-time CGM readings.

Modern AP systems use wireless communication technologies such as Bluetooth low energy to exchange data between their components. While wireless communication renders the design of a portable AP feasible, it also introduces vulnerabilities to cyber threats. In 2011, a study showed that both passive and active attacks on insulin pumps and glucose sensors were possible with public-domain information and easily accessible off-the-shelf hardware [6]. At a security conference in 2012, a white hat hacker hijacked an insulin pump from as far as 300 ft away with the aid of a high-gain antenna to boost the scanner range [7]. These early examples of cyberattacks were staged when the targeted systems lacked basic network security measures such as authentication and encryption. Network security measures are clearly necessary, but not sufficient to completely eliminate the risk of cyber threats. In fact, Medtronic Inc., a leading diabetes device company, had to issue cybersecurity-related safety notifications to their users

* Corresponding author.
E-mail address: fatihemre.tosun@angstrom.uu.se (F.E. Tosun).

https://doi.org/10.1016/j.jprocont.2024.103162
Received 19 October 2023; Received in revised form 22 December 2023; Accepted 8 January 2024
Available online 13 January 2024

in 2019 and 2022, which are also reported on the FDA's site [8]. Fortunately, the FDA is aware of no cyberattacks that targeted real AP users. Nevertheless, engineers must constantly strive to minimize the risk of cyber threats.

Complementary to the network security measures such as encryption [9], intelligent anomaly detection algorithms that exploit the input–output history as well as dynamical model knowledge enhance the security of cyber–physical systems (CPS). In particular, the field of control-theoretic security for CPS has emerged and attracted increasing research attention over the last decade. So far as the practical applications are concerned, the existing literature in this field is mainly focused on power, transportation, and industrial process infrastructures [10]. However, we believe that biomedical CPS such as the AP would also greatly benefit from a control-theoretic security perspective.

The range of cyberattack strategies against CPS is rather vast. In this work, we focus on sensor bias injection attacks, where the adversary injects a constant bias into the compromised sensor readings. This type of attack requires limited model knowledge as opposed to stealthy false data injection (FDI) attacks [11]. Due to ease of implementation, bias injections are arguably more likely than more sophisticated FDI attacks to [12]. Moreover, positive bias injections can drive the patient into a hypoglycemic coma (i.e., too low BG levels), which can be fatal, if not detected timely [13].

The intelligence and malicious intent of the adversary constitute the fundamental difference between FDI attacks and natural sensor failures [14]. The latter problem was addressed by some notable work for a general class of CGM failures from a process fault detection perspective [15–17]. However, such fault detection methods cannot adequately address carefully orchestrated FDI attacks since faults are random and not engineered by a smart entity. For instance, to make the attack harder to detect, the adversary can exploit the added uncertainty during meal ingestion by choosing the attack onset accordingly [13].

Typical defense strategies against sensor attacks include secure state estimation schemes that exploit sensor redundancy [18] and machine learning-based methods [19]. The former is not suitable for AP systems since currently, the only measurable physiological variable is the BG. The latter requires a significant amount of data collection in both the attack-free and the attack scenarios. Moreover, they are prone to relatively large detection delays as they process data in batches. Consequently, their effectiveness is hindered by the necessity of a sufficiently fast sampling rate. An ideal sensor anomaly detector for the AP must be able to timely detect attacks in the presence of non-persistent meal disturbances without sensor redundancy.

In this work, we formulate the challenge of detecting bias injections on the CGM under meal uncertainty as a statistical change detection problem. We use a time-varying Kalman filter as a software sensor to monitor the state of the AP. The filter uses a physiological model for BG dynamics and CGM readings to estimate the unmeasured states. The filter innovations are evaluated to decide whether an anomaly has occurred. For statistical evaluation, we consider two well-known online anomaly detectors: the χ^2 and cumulative sum (CUSUM) tests. We compare and study these detectors through the lens of the quickest change detection (QCD) framework, and derive a robust optimal detection procedure for constant bias injection attacks. Through simulations, we show that the proposed method is suboptimal but effective against generic non-stealthy FDI attacks including slow bias injections.

The rest of the article is organized as follows. Section 2 presents the preliminaries and the problem formulation. The proposed online anomaly detection scheme is introduced in Section 3. Section 4 delves into the implementations of the χ^2 and CUSUM detectors within the formalism of QCD theory. The efficacy of the proposed method is demonstrated through numerical simulations in Section 5. Finally, Section 6 summarizes the key findings of our study and provides concluding remarks.

2. Preliminaries

This section presents the necessary preliminaries for a thorough understanding of this paper. The first subsection presents the mathematical notation used throughout this paper. The second subsection presents the physiological model for BG dynamics. The final subsection presents the mathematical formulation of the attack detection problem.

2.1. Notation

For the convenience of the reader, the notation is grouped into the following categories:

Algebra: The sets of natural and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$, respectively. The n-dimensional Euclidean space is denoted by $\mathbb{R}^n$. The absolute value operator is denoted by $|\cdot|$. The operator $\lfloor x \rceil$ rounds x to the nearest integer. A positive semi-definite matrix $\mathbf{\Sigma}$ is denoted by $\mathbf{\Sigma} \succeq \mathbf{0}$. The symbol $\mathbf{0}$ means a vector/matrix of zeros of appropriate dimension.

Probability: The symbols $\mathbb{P}(\cdot)$ and $\mathbb{P}(\cdot \mid \cdot)$ denote the probability measure and conditional probability, respectively. Similarly, $(\mathbb{E}[\cdot \mid \cdot])$ $\mathbb{E}[\cdot]$ denotes the (conditional) expectation. A Gaussian variable X with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$ is denoted by $X \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$.

Miscellaneous: The symbols $\triangleq$ and $\equiv$ mean "defined as" and "identical to", respectively. The infimum (greatest lower bound) and supremum (least upper bound) of the set A are denoted by $\inf A$ and $\sup A$, respectively.

2.2. Physiological model for BG dynamics

In this work, we consider the Medtronic virtual patient (MVP) model for BG dynamics in T1D [20]. It is a low-order control-relevant model whose validation was performed against independent clinical data [21]. In control systems terminology, the insulin infusion rate is the control input and meal intakes are disturbances to BG regulation. The MVP model consists of the following set of ordinary differential equations:

$$\begin{aligned}
\frac{dI_{sc}(t)}{dt} &= -\frac{I_{sc}(t)}{\tau_1} + \frac{U_{sc}(t)}{\tau_1 C_I} \\
\frac{dI_p(t)}{dt} &= -\frac{I_p(t)}{\tau_2} + \frac{I_{sc}(t)}{\tau_2} \\
\frac{dI_e(t)}{dt} &= -p_2\, I_e(t) + p_2\, S_I\, I_p(t) \\
\frac{dG(t)}{dt} &= -(GEZI + I_e(t))\, G(t) + EGP + R_a(t) \\
\frac{dG_{sc}(t)}{dt} &= -\frac{G_{sc}(t)}{\tau_s} + \frac{G(t)}{\tau_s}.
\end{aligned} \tag{1}$$

The control input is denoted by $U_{sc}(t)$ (mIU/min) with IU being the international unit for insulin. The variables $I_{sc}(t)$ and $I_p(t)$ denote the insulin levels (mIU/L) in the subcutaneous (SC) tissue and in plasma, respectively. The interchangeable time constants τ_1 and τ_2 determine the rate of insulin transport from the SC tissue to plasma, and C_I is the insulin clearance rate. The variable $I_e(t)$ (1/min) defines the remote insulin effect. The parameter p_2 is the reciprocal of the insulin action time constant, S_I is the insulin sensitivity, $GEZI$ is the glucose effectiveness at zero insulin, and EGP is the endogenous glucose production rate. The controlled variable $G(t)$ (mg/dl) is the BG level while the measured variable $G_{sc}(t)$ (mg/dl) is the SC glucose level with τ_s being the sensor time constant. The term $R_a(t)$ (mg/dl/min) is the rate of glucose appearance following a meal intake whose dynamics are given by the following linear two-compartment model:

$$\begin{aligned}
\frac{dD(t)}{dt} &= -\frac{1}{\tau_m} D(t) + C_h(t) \\
\frac{dR_a(t)}{dt} &= -\frac{1}{\tau_m} R_a(t) + \frac{1}{\tau_m^2 V_G} D(t)
\end{aligned} \tag{2}$$

where $C_h(t)$ (g/min) is the meal intake, $D(t)$ (g) is the glucose mass in the first gut compartment, and τ_m is the corresponding time constant.

Typically, meals are assumed to be ingested instantly [22]. Thus, $C_h(t)$ is modeled as an impulse train as:

$$C_h(t) = \sum_{i\in\mathbb{N}} c_i \delta(t - t_i) \tag{3}$$

where t_i (min) is the time instant of the ith meal intake, c_i (g) is the amount of the carbohydrate (CHO) consumed at t_i, and $\delta(\cdot)$ is the Dirac delta function.

2.3. Problem setup

In this section, we formulate sensor anomaly detection as a statistical change detection problem. In our context, an anomaly is defined as any deviation of the controlled CPS (i.e., the AP) from the nominal behavior. In this work, we consider model-based anomaly detection that involves two subsequent steps: residual generation and evaluation [23]. As the name suggests, a model-based detection scheme utilizes model knowledge and input–output history to detect anomalous behavior in the system.

We employ a discrete-time (DT) Kalman filter as the residual generator since the CGM provides measurements at discrete sampling instants. Designing a Kalman filter requires a linear model. However, the MVP model is nonlinear and in continuous time. Hence, we derive an approximate DT linear model as follows. Let $k \in \mathbb{N}$ be the discrete-time index, and h be the sampling period with $t = kh$. Then, the following DT linear system approximates the combined dynamics of (1) and (2):

$$\begin{aligned} x[k+1] &= Ax[k] + Bu[k] + MC_h[k] + w[k] \\ y[k] &= Cx[k] + v[k] + a[k] \end{aligned} \tag{4}$$

where $x[k] \in \mathbb{R}^7$ is the state at time k, $u[k] \in \mathbb{R}$ is the known control input, and $C_h[k] \in \mathbb{R}$ is the partially known meal disturbance from the user-provided meal announcement as explained in the next subsection. We introduce the process noise $w[k] \in \mathbb{R}^7$ to account for modeling errors (e.g., due to linearization). The output $y[k] \in \mathbb{R}$ is the real-time CGM readings which are corrupted by the inherent sensor noise $v[k] \in \mathbb{R}$ as well as a possible FDI $a[k] \in \mathbb{R}$. We assume that $w[k]$, $v[k]$, and the initial state $x[0]$ are mutually independent random variables with $w[k] \sim \mathcal{N}(\mathbf{0}, Q \succeq \mathbf{0})$, $v[k] \sim \mathcal{N}(0, R > 0)$, and $x[0] \sim \mathcal{N}(\mathbf{0}, \Sigma)$. The linearization is made around the fasting equilibrium with basal insulin delivery rate. The appendix presents a detailed explanation of the linearization and discretization procedures.

2.3.1. Meal announcements

A meal intake is characterized by the size and time of the consumed CHO as in (3). We assume the time of each meal intake is accurately reported by the patient. However, the meal size estimate is subject to error [24]. The meal size can be estimated by manual CHO counting or with the aid of a dedicated computer vision-based app such as GoCARB [25]. Alternatively, instead of inputting the exact CHO amount, the user may be prompted to select from a three- or four-scale meal size [22]. In any case, a meal announcement may be formulated as:

$$\hat{C}_h[k] = \sum_{i\in\mathbb{N}} \hat{c}_i \, \delta_{k,k_i}, \; k_i = \lfloor t_i/h \rceil \tag{5}$$

where $\hat{c}_i$ (g) is the ith meal size estimate, k_i is the corresponding discrete-time index, and δ_{k,k_i} is the Kronecker delta function defined as follows:

$$\delta_{k,k_i} = \begin{cases} 0 & \text{if } k \neq k_i \\ 1 & \text{if } k = k_i. \end{cases} \tag{6}$$

In this setting, the round-off error regarding the time of meal onset is neglected. The actual meal size c_i is an unknown deterministic variable. For simplicity, we model the meal estimation error as a normal variable as follows:

$$\hat{c}_i = c_i + \tilde{c}_i, \; \tilde{c}_i \sim \mathcal{N}(0, \sigma_m^2) \tag{7}$$

where $\tilde{c}_i$ is the error in estimating the size of the ith meal, σ_m is the corresponding standard deviation.

2.3.2. Attack detection as a hypothesis testing problem

A constant bias injection on the CGM is modeled as:

$$a[k] = \bar{a}\, H[k - k_a] \tag{8}$$

where $H[\cdot]$ denotes the discrete unit step function, k_a is the attack start time, and $\bar{a}$ is the amount of injected bias. These two attack parameters are unknown but deterministic. The attack detection may be formulated as a binary hypothesis testing problem where the null and alternative hypotheses are, respectively, as follows:

H_0 : System is operating normally (i.e., $a[k] \equiv 0$)

H_1 : System is under attack (i.e., $a[k] \not\equiv 0$).

The probability laws corresponding to H_0 and H_1 are required to implement a statistical test, and derived in the next section.

3. Proposed detection scheme

In this section, we present the proposed model-based detection scheme for FDI attacks as depicted in Fig. 1. We consider a slightly modified version of the standard Kalman filter to handle erroneous meal announcements. Let us define a time-varying process noise $Q'[k] \triangleq Q + M\sigma_m^2 M^T \delta_{k,k_i}$ to include the effect of the meal uncertainty in state estimation. Let $\mathcal{Y}[k] \triangleq \{(u[i], \hat{C}_h[i], y[i]) : 0 \le i \le k\}$ be the set of input–output observations up to time k. Then, the Kalman filter equations read as:

$$\begin{aligned} \hat{x}[k+1] &= (A - K[k]C)\hat{x}[k] + Bu[k] \\ &\quad + M\hat{C}_h[k] + K[k]y[k] \end{aligned} \tag{9}$$

$$K[k] = AP[k]C^T(CP[k]C^T + R)^{-1} \tag{10}$$

$$P[k+1] = AP[k]A^T - K[k]CP[k]A^T + Q'[k] \tag{11}$$

where $\hat{x}[k] \triangleq \mathbb{E}[x[k] \mid \mathcal{Y}[k-1]]$ is the one-step predictor of the state $x[k]$ in (4), $P[k] = \mathbb{E}[(x[k] - \hat{x}[k])(x[k] - \hat{x}[k])^T]$ is the associated error covariance, and $K[k]$ is the filter gain [26]. The recursions start from $\hat{x}[0] = \mathbf{0}$ and $P[0] = \Sigma$.

The difference between the measured and the predicted output by the Kalman filter is called the innovation, and is defined as follows:

$$z[k] \triangleq y[k] - C\hat{x}[k]. \tag{12}$$

In the absence of anomalies, that is $a[k] \equiv 0$, $z[k]$ is a zero-mean white Gaussian sequence with standard deviation $\sigma_r[k] = \sqrt{CP[k]C^T + R}$ [27]. The innovation $z[k]$ is a non-stationary random sequence since $\sigma_r[k]$ is time-varying. To circumvent this issue, we normalize $z[k]$ with its variance as follows:

$$r[k] = z[k]/\sigma_r[k]. \tag{13}$$

The standardized innovation sequence $r[k]$ is independent and identically distributed (IID), and chosen as the residual signal for anomaly detection.

When H_0 holds true, we have already established that $r[k] \sim \mathcal{N}(0, 1)$. Similarly, when H_1 holds true, $r[k]$ simply assumes a Gaussian with the same variance but a time-varying mean for any additive deterministic FDI attack. To derive the mean of $r[k]$ under such attacks, consider the following residual dynamics:

$$\begin{aligned} \tilde{x}[k+1] &= (A - K[k]C)\tilde{x}[k] - M\tilde{C}_h[k] + w[k] \\ &\quad - K[k](v[k] + a[k]) \\ z[k] &= C\tilde{x}[k] + v[k] + a[k] \\ r[k] &= z[k]/\sigma_r[k] \end{aligned} \tag{14}$$

where $\tilde{x}[k] \triangleq x[k] - \hat{x}[k]$ is the state estimation error, and $\tilde{C}_h[k] \triangleq \sum \tilde{c}_i \, \delta_{k,k_i}$ is the meal estimation error. Next, we exploit the linearity of the Kalman filter by invoking the superposition principle. To this end, let us decompose $\tilde{x}[k]$ into two parts as $\tilde{x}[k] = \tilde{x}_t[k] + \tilde{x}_f[k]$ where

$$\tilde{x}_t[k] \triangleq \tilde{x}[k]|_{a[k]=0}, \; \tilde{x}_f[k] \triangleq \tilde{x}[k]|_{\substack{\tilde{C}_h[k]=v[k]=0 \\ w[k]=\mathbf{0}}}. \tag{15}$$

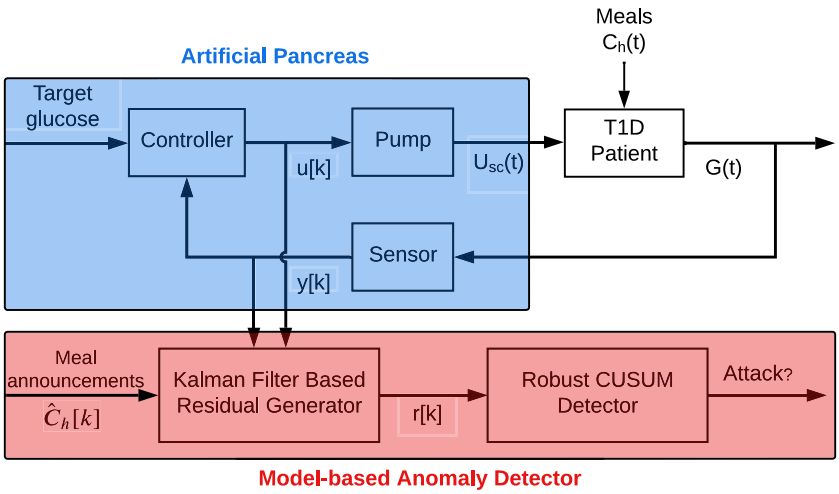


Fig. 1. A schematic diagram of the proposed anomaly detection scheme.

In particular, $\tilde{x}_t[k]$ is the state estimation error under no attack while $\tilde{x}_f[k]$ is the isolated contribution of the attack to $\tilde{x}[k]$. The effect of $\tilde{x}_f[k]$ on the residual measurements is manifested by the following dynamics:

$$\begin{aligned}\tilde{x}_f[k+1] &= (A - K[k]C)\tilde{x}_f[k] - K[k]a[k] \\ r_f[k] &= (C\tilde{x}_f[k] + a[k])/\sigma_r[k].\end{aligned} \tag{16}$$

Here, $r_f[k]$ is simply the deviation of the residual from the true noisy measurements at time k. Thus; when H_1 holds true, the residual becomes $r[k] \sim \mathcal{N}(r_f[k], 1)$.

For ease of theoretical exposition, a steady-state analysis of the attacked residual is in order. The steady-state error covariance $\bar{P}$ is obtained by solving the following DT algebraic Riccati equation:

$$\bar{P} = A[\bar{P} - \bar{P}C^T(C\bar{P}C^T + R)^{-1}C\bar{P}]A^T + Q. \tag{17}$$

The corresponding steady-state filter gain $\bar{K}$ is equal to:

$$\bar{K} = A\bar{P}C^T(C\bar{P}C^T + R)^{-1}. \tag{18}$$

Let $\bar{x}_f$ and $\bar{r}_f$ be the corresponding steady-state values of $\tilde{x}_f[k]$ and $r_f[k]$. By definition, they are computed by solving (16) for $\tilde{x}_f[k+1] = \tilde{x}_f[k] = \bar{x}_f$. Thus, we get:

$$\bar{x}_f = (A - \bar{K}C - I)^{-1}\bar{K}\bar{a} \tag{19}$$

$$\Gamma = \frac{C(A - \bar{K}C - I)^{-1}\bar{K} + 1}{\sqrt{C\bar{P}C^T + R}} \tag{20}$$

$$\bar{r}_f = \Gamma\bar{a} \tag{21}$$

where Γ is a scaling factor for the residual-mean under a constant bias injection.

Since $A - \bar{K}C$ is Schur stable, the inverse of $(A - \bar{K}C - I)$ uniquely exists. Due to filter dynamics, it takes a few iterations before the attacked residual-mean $r_f[k]$ converges to its steady-state value of $\bar{r}_f$. Nevertheless, the transients are neglected in the subsequent analysis as the Kalman filter is known to converge quickly. Consequently, the detection of bias injections on the CGM is formulated as a sequential (online) change detection problem as follows:

$$\begin{aligned}H_0 &: r[i] \sim \mathcal{N}(0,\ 1) \text{ for } 0 \le i \le k \\ H_1 &: r[i] \sim \mathcal{N}(0,\ 1) \text{ for } 0 \le i < k_a \\ &\quad\ r[i] \sim \mathcal{N}(\bar{r}_f, 1) \text{ for } k_a \le i \le k.\end{aligned} \tag{22}$$

This is a classical problem of detecting a shift in the mean of a Gaussian sequence with known variance and unknown change point which is in our case the attack start time k_a. The major challenge here is the lack of knowledge regarding the change parameter $\bar{r}_f$, which is a linear function of the injected bias $\bar{a}$ as in (21). Since it is not realistic to know $\bar{a}$ beforehand, a conceivable way to address this issue is to employ a χ^2 detector which only requires the probability distribution corresponding to H_0. However, despite the simplicity, negligence of the alternative hypothesis H_1 may result in significant performance loss in detection delay as shown in Section 5.

If $\bar{r}_f$ is (assumed to be) known, H_1 is a simple hypothesis just as H_0. On the contrary, H_1 is a composite hypothesis if $\bar{r}_f$ belongs to a set with at least two elements. Thus, we call (22) a composite change detection problem when $\bar{r}_f$ is not fully known. In detection theory, there are two major paradigms to handle composite change detection problems: adaptive and minimax [28]. Adaptive detectors such as the generalized likelihood ratio test aim to detect the change by estimating the unknown $\bar{r}_f$. However, they are computationally infeasible in real-time unless a sliding-window approach is used. However, this results in performance loss since only partial data history is used [29]. On the other hand, a minimax detector aims to guarantee a certain performance under the worst-case scenario. More precisely, a minimax detector is tuned according to the least favorable value of $\bar{r}_f$ instead of estimating it. The upside of the minimax approach is that it is of the same complexity as a simple change detection problem where $\bar{r}_f$ is known. Moreover, it admits a recursive solution; hence, all data history can be exploited with minimal computational burden.

In this work, we propose to employ a minimax robust two-sided CUSUM detector as summarized in Algorithm 1. As explained in detail in Section 4.2.3, the detector is tuned to be sensitive to the maximum tolerable bias in CGM readings. Due to the symmetry, the detector is equally sensitive to the positive and negative biases. In essence, we treat the unknown change parameter $\bar{r}_f$ as a tuning knob that adjusts the trade-off between the detector's sensitivity to the noise and detection performance. As evident from (21), it is a function of both the system dynamics and the injected bias $\bar{a}$. This approach ensures robust detection for a wide range of attack parameters.

4. Quickest change detection theory

In this section, we present some key notions regarding statistical hypothesis testing, more specifically in the context of QCD. In particular, QCD algorithms aim to detect abrupt changes in the statistical properties of a random process as quickly as possible after the unknown change time whilst satisfying a specified false alarm constraint. We restrict the discussion to the case where the observations before and after the change are IID. We denote the pre- and post-change distributions by f_0 and f_1, respectively. We begin by introducing some key definitions.

Algorithm 1 Minimax robust two-sided CUSUM test

Initialize: $k = 0,\ g = 0,\ g_1 = 0,\ g_2 = 0,\ \tau,\ m$
Require: $\tau > 0,\ m > 0$
Input: $r[k]$ ▷ Kalman filter innovation sequence
while $g < \tau$ **do**
  $l^+ \leftarrow mr[k] - 0.5m^2$ ▷ See (36)
  $l^- \leftarrow -mr[k] - 0.5m^2$ ▷ See (36)
  $g_{cs}^+ \leftarrow \max(0,\ g_{cs}^+ + l^+)$ ▷ See (38)
  $g_{cs}^- \leftarrow \max(0,\ g_{cs}^- + l^-)$ ▷ See (39)
  $g \leftarrow \max(g_1,\ g_2)$ ▷ See (37)
  $k \leftarrow k + 1$
end while
Output: raise an alarm

Definition 1 (*Log-likelihood Ratio*)**.** The log-likelihood ratio (LLR) of a random variable X with respect to f_1 and f_0 is given by:

$$\ell(X_n) \triangleq \log \frac{f_1(X_n)}{f_0(X_n)}. \tag{23}$$

It is a key quantity for optimal hypothesis testing [29].

Definition 2 (*Kullback–Leibler Divergence*)**.** The Kullback–Leibler divergence (KLD) between two distributions f_1 and f_0 is defined as

$$D(f_1 \parallel f_0) \triangleq \int_{-\infty}^{\infty} f_1(x) \log \frac{f_1(x)}{f_0(x)} dx \geq 0. \tag{24}$$

The information-theoretic notion of KLD was proposed as a measure of the dissimilarity between f_1 and f_0 in the seminal work of Kullback and Leibler [30]. In particular, $D(f_1 \parallel f_0)$ is 0 only when $f_1(x) = f_0(x)$, and positive otherwise. It is a key quantity in characterizing the performance of QCD algorithms [31] as well as the stealthiness of FDI attacks in stochastic CPS [32].

Definition 3 (*Stopping Time*)**.** A stopping time on a random sequence $(X_n)_{n\geq 1}$ is a random variable T such that for each discrete-time instant n, the event $\{T = n\}$ belongs to the σ-algebra generated by $(X_1, \dots, X_n)$.

To put it simply, T depends only on the information available up to and including time n, but not on any future information. An online anomaly detector may be conveniently defined in terms of a stopping time on its residual sequence as follows:

$$T = \inf\{k \geq 1 : g[k-1] > \tau\} \tag{25}$$

where $g[k]$ is the test statistic at time k, which is a causal function of $r[k]$ and τ is the decision threshold. Thus, the stopping time of a detector is a positive integer-valued random variable that gives the number of residual measurements taken until an alarm is triggered for the first time. The stopping time is a quintessential notion for quantifying the operating characteristics of a detector including the detection delay as shall be explained now.

For notational brevity, let $\mathbb{E}_{k_a}[T]$ be the mean value of T when the change point is k_a. Consequently, let $\mathbb{E}_\infty[T]$ denote the mean of T when the change occurs at infinity, or equivalently when no attack is present. In particular, $\mathbb{E}_\infty[T]$ is the average time between false alarms which should ideally be as large as possible. From now on, we refer to this quantity as the false alarm interval (FAI). When no prior distribution on k_a is available, the following constraint set is used for the QCD algorithms of interest [31]:

$$\mathbf{C}_\gamma = \{T : \mathbb{E}_\infty[T] \geq \gamma > 1\}. \tag{26}$$

The parameter γ denotes the minimum acceptable FAI. Next, let us define the detection delay as $T - k_a$, which is clearly a random variable. Thus, we consider the average detection delay (ADD) as follows:

$$ADD_{k_a}(T) \triangleq \mathbb{E}_{k_a}[T - k_a \mid T > k_a]. \tag{27}$$

Please note that in the present setting, $T - k_a = 1$ implies instant detection of the attack. Ideally, we wish to find the stopping time in the set $\mathbf{C}_\gamma$ that minimizes (27) uniformly over all possible change points $k_a \geq 0$. However, such procedures do not exist [29]. Instead, one can resort to a minimax (i.e., worst-case) approach. To this end, we consider the worst-case (maximal) ADD as originally proposed by Pollak [33]:

$$SADD(T) \triangleq \sup_{k_a \geq 0} \mathbb{E}_{k_a}\left[(T - k_a) \mid T > k_a\right]. \tag{28}$$

Hence, we seek to find an optimal detector T^* such that:

$$SADD(T^*) = \inf_{T \in \mathbf{C}_\gamma} SADD(T). \tag{29}$$

In words, an optimal detection rule minimizes the maximal ADD within the feasible set $\mathbf{C}_\gamma$. Unfortunately, finding T^* proves to be intractable in most cases, but certain QCD algorithms such as the CUSUM test are second-order asymptotically (i.e., as $\gamma \to \infty$) optimal [31]. The exact definition of second-order optimality is highly technical and beyond the scope of this paper. Instead, hereafter, we shall colloquially refer to it as nearly optimal. Based on the theoretical foundations laid out above, we now delve into the online anomaly detectors explored in this study.

4.1. χ^2 Test

The χ^2 test is a general-purpose, widely-used anomaly detector to monitor CPS due to its simplicity [34]. It is implemented as follows:

$$T_{\chi^2}^N = \inf\{k \geq 1 : g_{\chi^2}^N[k-1] = \sum_{i=k-N}^{k-1} (r[i])^2 > \tau\} \tag{30}$$

where $N \geq 1$ is the window size, $T_{\chi^2}^N$ is the corresponding stopping time and $r[i < 0] = 0$. When $N = 1$, the χ^2 detector (30) is said to be stateless (or memoryless) and stateful otherwise. Since $r[k]$ is a sequence of independent standard normal variables, $g_{\chi^2}^N[k]$ follows the χ^2 distribution with N degrees-of-freedom, hence the name. The threshold τ is typically chosen to guarantee a false alarm probability of α as follows:

$$\mathbb{P}(g_{\chi^2}^N[k] > \tau \mid H_0) = \alpha. \tag{31}$$

The value for τ can easily be obtained by solving (31) with the aid of statistical software or a distribution table.

To make a meaningful comparison between the χ^2 and CUSUM tests, we must use the same metric for the false alarm constraint, namely the FAI. In general, there is no direct relation between α and $\mathbb{E}_\infty[T_{\chi^2}^N]$. Only in the stateless case, they are reciprocals as $\mathbb{E}_\infty[T_{\chi^2}^1] = 1/\alpha$. In the stateful case, the test statistic $g_{\chi^2}^N[k]$ is a random walk, and finding the value of τ ensuring $\mathbb{E}_\infty[T_{\chi^2}^N] = \gamma$ involves deriving and solving complicated integral equations. One can instead use Monte Carlo (MC) method to compute τ which is arguably simpler and more intuitive.

4.2. CUSUM test

The CUSUM test exploits the full history of measurements as well as the knowledge of the post-change distribution, as opposed to the χ^2 test. The rationale behind this algorithm is to exploit the different behavior of the LLR before and after the change. The following relations between the LLR and the KLD are easily derived from (23) and (24):

$$\mathbb{E}_\infty[\ell_k] = -D(f_0 \parallel f_1) < 0,\ \mathbb{E}_{k_a}[\ell_k] = D(f_1 \parallel f_0) > 0 \tag{32}$$

where ℓ_k is the LLR at time k. As can be seen from (32), ℓ_k has a negative mean before the change and a positive mean after the change. Hence, computing the cumulative sum $\sum \ell_k$ should be informative about whether a change has occurred. More precisely, $\sum \ell_k$ is most likely to attain its minimum at the change point.

In (22), the pre-change distribution f_0 is the standard Gaussian whereas the post-change distribution f_1 is a Gaussian with mean $\bar{r}_f$ and unit variance. In the following subsections, we present three variations of this algorithm.

4.2.1. One-sided CUSUM test

This is the most basic version of the algorithm. Suppose we have a simple change detection problem with $\bar{r}_f = n$ where n is a known constant. Then, the LLR sequence of the residual measurements reads as:

$$\ell_{k+1} = n(r[k] - 0.5n) \text{ for } k \geq 0. \tag{33}$$

The one-sided CUSUM test may be implemented recursively as follows:

$$T_{cs} = \inf\{k \geq 1 : g_{cs}[k] = \max\{0,\ g_{cs}[k-1] + \ell_k\} > \tau\} \tag{34}$$

with the recursion starting from $g_{cs}[0] = 0$. The recursive nature of (34) enables efficient real-time implementation of the test [31]. When τ is selected to ensure $\mathbb{E}_\infty[T_{cs}] = \gamma$, T_{cs} is nearly optimal in the sense of (29) with the following asymptotic relationship [29]:

$$SADD(T_{cs}) \approx \gamma / D(f_1 \parallel f_0) \text{ as } \gamma \to \infty. \tag{35}$$

It is also exactly optimal with respect to Lorden's more pessimistic measure of worst detection delay which we do not consider in this work [31]. Therefore, we believe the CUSUM test among all other known QCD algorithms is the most suitable choice for this work.

4.2.2. Two-sided CUSUM test

Now, suppose we only know the magnitude but not the sign of $\bar{r}_f$ such that $\bar{r}_f \in \{-m, m\}$ with $m > 0$ being the magnitude. Then, the LLRs corresponding to the positive and negative changes, respectively, read as:

$$\ell^+_{k+1} = m(r[k] - 0.5m),\ \ell^-_{k+1} = -m(r[k] + 0.5m). \tag{36}$$

The two-sided CUSUM test is simply two one-sided tests running in parallel as follows:

$$T_{cs2} = \inf\{k \geq 1 : g^+_{cs}[k] > \tau \text{ or } g^-_{cs}[k] > \tau\} \tag{37}$$

where

$$g^+_{cs}[k] = \max\{0,\ g^+_{cs}[k-1] + \ell^+_k\},\ g^+_{cs}[0] = 0 \tag{38}$$

$$g^-_{cs}[k] = \max\{0,\ g^-_{cs}[k-1] + \ell^-_k\},\ g^-_{cs}[0] = 0. \tag{39}$$

Similarly, when τ is selected to ensure $\mathbb{E}_\infty[T_{cs2}] = \gamma$, the test is nearly optimal in the sense of (29) [29].

4.2.3. Minimax robust two-sided CUSUM test

The optimality properties of the one- and two-sided CUSUM tests hold only when the presumed change parameter is equal to the true change parameter which is seldom the case. In general, the true change parameter belongs to a so-called uncertainty set. The minimax robust CUSUM test is then simply the ordinary CUSUM test where the presumed change parameter is equal to the worst change parameter. The worst change parameter is the one which renders the post-change distribution least favorable for detection [35]. This approach ensures robustness to the unknown change parameter by minimizing the worst-case delay among all values of $\bar{r}_f$ within the uncertainty set.

Finding the worst change parameter is not always possible, but luckily, in our case, it is straightforward. Suppose, we know only the minimum magnitude of change such that the uncertainty set is $\mathcal{U} = \{\bar{r}_f \in \mathbb{R} : |\bar{r}_f| \geq m\}$. Then, the worst change parameter is $\pm m$ due to Theorem III.2 in [35]. Hence, we propose to employ the two-sided CUSUM detector (37) with m being the minimum change magnitude to which we wish to be sensitive. This value can be determined from (21) for a given bias $\bar{a}$. If the true change magnitude $|\bar{r}_f|$ turns out to be greater than m, the attack will get detected even faster as it should be intuitively clear. Please note that smaller bias injections with $|\bar{r}_f| < m$ are also detectable albeit with a delay. The idea is to design a detector with maximal FAI by tolerating the increased detection delay for less harmful attacks. We emphasize that m is a tuning parameter for the detector rather than an absolute bound on the true change parameter.

Table 1
The MVP model parameters with their numerical values identified for a certain subject.

Parameter	Value	Unit
C_I	2.01	[L/min]
τ_1	49	[min]
τ_2	47	[min]
p_2	$1.06\ 10^{-2}$	[min^{-1}]
S_I	$8.11\ 10^{-4}$	[L/mIU/min]
GEZI	$2.2\ 10^{-3}$	[min^{-1}]
EGP	1.33	[mg/dl/min]
V_G	253	[dl]
τ_m	50	[min]
τ_s	10	[min]

5. Numerical simulations

In this section, we present the numerical simulations conducted to demonstrate the efficacy of the proposed detection scheme as well as the theoretical discussion in the previous section.

5.1. In silico experimental design

Throughout the rest of the section, we use the MVP simulator as described by (1) and (2) with the parameter values in Table 1 [20]. The simulator corresponds to the *T1D Patient* block in Fig. 1. The Kalman filter processes the CGM measurements based on the DT dynamical model (4) whose numerical values of the system matrices can easily be calculated by referring to Table 1 and the appendix. Similar to [36], we set Q as a diagonal matrix with the following entries: $(10^{-6},\ 10^{-6},\ 10^{-6},\ 0.5,\ 10^{-6},\ 0,\ 0)$ and $R = 60\ (mg/dl)^2$. We select the maximum tolerable bias as $15\ (mg/dl)$. The desired FAI is set to 300 samples, which amounts to slightly less than one false alarm per day assuming a sampling period $h = 5$ minutes. We employ a DT-PID controller with a filtered derivative term as follows:

$$e[k] = y[k] - G_0 \tag{40}$$

$$D[k] = \frac{NT_d}{h + NT_d} D[k-1] + \frac{K_p T_d}{h + NT_d} (e[k] - e[k-1]) \tag{41}$$

$$u[k] = \bar{U}_{sc} + K_p \left(e[k] + \frac{h}{T_i} \sum_{j=1}^{k} e[j] \right) + D[k]. \tag{42}$$

The target glucose level G_0 is set to 100 mg/dl, and the corresponding basal insulin rate $\bar{U}_{sc}$ is given in (A.6). The controller parameters are: $K_p = 0.2$ (mIU/min)/(mg/dl), $T_i = 90$ (min), $T_d = 60$ (min), and $N = 0.1$.

5.2. Detection performance in the absence of meals

We compare the χ^2(30) and two-sided CUSUM (37) detectors without meals under constant bias injections. Fig. 2 plots the operating characteristics of these detectors under two scenarios as computed by MC simulations. In the first scenario, the presumed change parameter is equal to the true change parameter with $\bar{a} = 15$ (mg/dl). The results for this scenario are illustrated by the top three curves. In this case, the optimal and minimax robust CUSUM detectors are identical by definition. The stateless χ^2 detector suffers from a large detection delay. The stateful χ^2 detector with a sliding window length of 5 samples greatly reduces the detection delay, yet it is still too large. The delay can be further reduced by increasing N, but with a higher computational cost in terms of memory and arithmetic operations. However, the CUSUM detector will outperform the χ^2 detector for any arbitrarily large N with much less computational cost.

In the second scenario, we have a larger bias injection with $\bar{a} = 30$ (mg/dl). In this case, the maximal ADD of the minimax robust CUSUM detector is lower than that of the first scenario. Moreover,

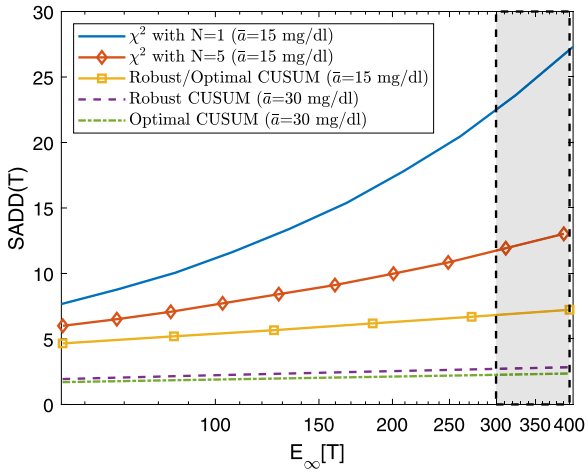


Fig. 2. The trade-off plots between the FAI and the SADD for the χ^2 and CUSUM detectors under a small and a large bias injection attack. The shaded area represents the desired FAIs.

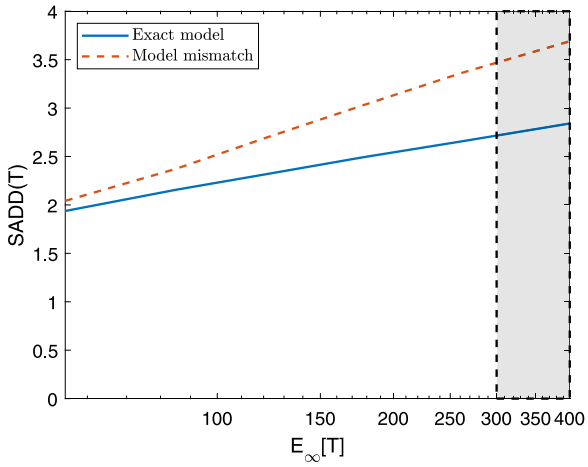


Fig. 3. The trade-off plots between the FAI and the SADD for the robust CUSUM detector with the exact model and model mismatch.

the performance degradation is insignificant for the desired FAIs as depicted by the bottom two curves. The bottom-most curve refers to the hypothetical scenario where the attack parameter $\bar{a}$ is known exactly, and the optimal CUSUM detector is used. The gap between the curves of the minimax robust and optimal CUSUM detectors may be thought of as the price to pay for the loss of information about the true change parameter.

Thus far, we have assumed perfect model knowledge. Next, we investigate the robustness of the proposed method to model mismatch. Since the insulin sensitivity can vary up to 30% over the day, we select S_I as the perturbed variable for our investigation [37].

In particular, we simulate the "real" BG dynamics with the values in Table 1, but overestimate S_I by 30% in the Kalman filter equations. The results are reported Fig. 3. The maximal ADD is slightly increased due to the discrepancy between the statistics of the theoretical and observed residual. However, the performance degradation is acceptable in spite of the large parameter uncertainty.

5.3. Detection performance in the presence of meals

In this subsection, we present an illustrative example to show the efficacy of the proposed method under partially known meal disturbances. The simulations were performed over a 24 h period with 3 meals taken at 5 h, 13 h, and 18 h. To assess the robustness of our method to meal uncertainties, two distinct scenarios were considered. The first scenario incorporated a small variance of $\sigma_m^1 = 10$ grams for the meal estimates, while the second scenario incorporated a larger variance of $\sigma_m^2 = 30$ grams. The meal intakes were identical and equal to 75 g.

Table 2
Monte Carlo estimates of the ADD (in samples) of detectors for $\gamma = 300$, $\sigma_m^1 = 10$ g and $\sigma_m^2 = 30$ g.

Detector	ADD for σ_m^1	ADD for σ_m^2
χ^2 with N = 1	7.35	10.93
χ^2 with N = 5	4.69	7.18
Robust CUSUM	2.31	4.38

In order to be stealthier, the attacker slowly injects the bias using a first-order low-pass filter as follows:

$$a[k+1] = \bar{a}H[k-k_a] + (1-\beta)a[k],\ a[0] = 0 \tag{43}$$

where β is a parameter that determines the rate of bias injection and is taken as 0.1. The attacker makes the attack onset k_a coincide with the third meal intake time so as to exploit the extra uncertainty stemming from the unknown meal size. The attack sequence is depicted in Fig. 4(c).

We perform closed-loop MC simulations of the proposed detection algorithm for this attack scenario. The ADDs of the χ^2 and CUSUM detectors for both small and large meal uncertainties are reported in Table 2. The CUSUM detector has quite satisfactory performance as the attack is detected soon after its onset. The stateful χ^2 detector performs reasonably well against the large bias injection with small meal uncertainty. However, its detection performance under large meal uncertainty is not as good. The detection delay of the stateless χ^2 detector is unacceptably large even for the small meal variance.

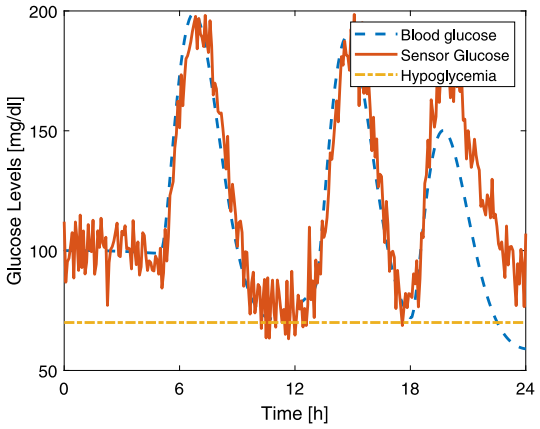
Fig. 4 shows a representative outcome from the simulations for the small meal variance. In particular, Fig. 4(a) plots the blood and sensor glucose trajectories. Before the attack starts at 18 h, the difference between the blood and sensor glucose values is only due to the sensing delay and measurement noise. Thankfully, the two-sided minimax robust CUSUM test is able to detect the attack well before hypoglycemia is achieved as depicted in Fig. 4(d).

Fig. 4(b) plots the meal disturbance trajectories. The time-varying Kalman filter gradually corrects the meal disturbance estimate, which is initially computed from the meal announcement by the user, as glucose measurements come in. The filter estimate of the meal disturbance gets worse during the FDI attack which is reflected as an anomaly in the CUSUM test statistics. This shows that the time-varying Kalman filter can successfully handle erroneous meal estimates.

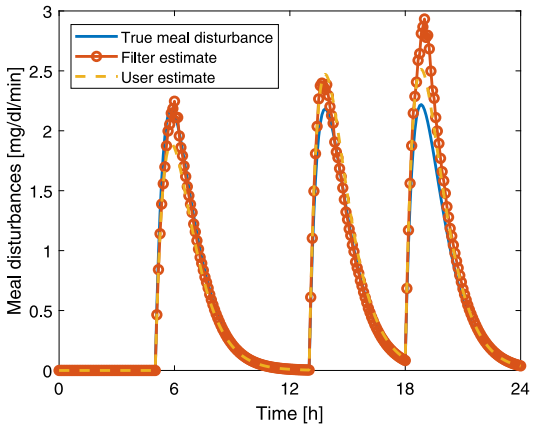
6. Conclusion

In this work, we considered deterministic FDI attacks on the CGM deployed in an AP under partially known meal disturbances. Our problem formulation was generic enough to address natural additive sensor faults, as well. We proposed a model-based detection scheme that is effective, robust, and easy to implement. A time-varying Kalman filter was used to handle the sporadic meal disturbances with known meal intake times. The standardized Kalman filter innovation sequence was chosen as the residual signal due to its amenability to statistical evaluation. We derived the worst-case optimal detection rule against constant bias injection attacks, namely the minimax robust two-sided CUSUM test. We also empirically showed the robustness of our approach to model mismatch.

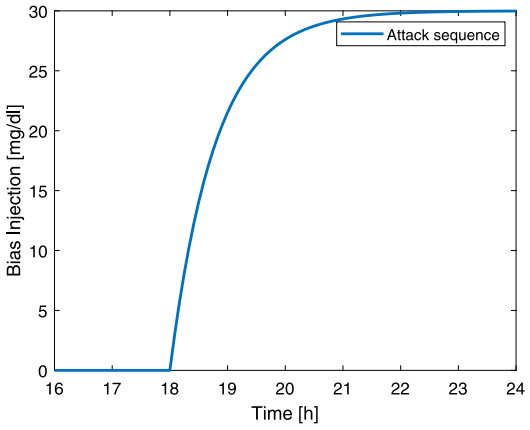
In future work, we aim to characterize the trade-off between the impact and stealthiness of sensor deception attacks. We also plan to address the problem of distinguishing unannounced meals from a sensor attack as well as intraday variation in physiological parameters.



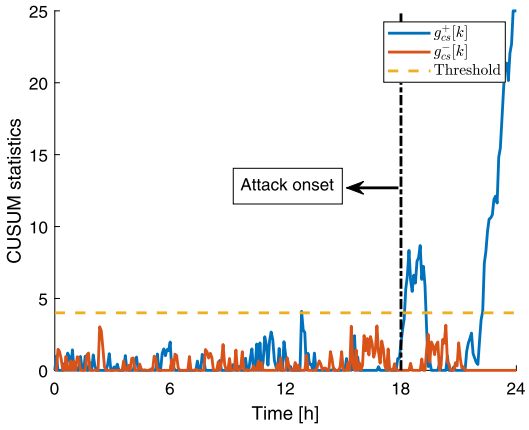
(a) The blood and sensor glucose trajectories



(b) Meal disturbances and estimates



(c) The attack sequence starting from 18h



(d) Residual evaluation with the two-sided CUSUM test

Fig. 4. A representative sample of the closed-loop simulation of the proposed anomaly detection scheme.

CRediT authorship contribution statement

Fatih Emre Tosun: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing. **André M.H. Teixeira:** Writing – original draft, Supervision. **Mohamed R.-H. Abdalmoaty:** Methodology, Software. **Anders Ahlén:** Supervision, Writing – original draft, Writing – review & editing. **Subhrakanti Dey:** Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Fatih Emre Tosun reports financial support was provided by Swedish Research Council. Fatih Emre Tosun reports financial support was provided by Swedish Foundation for Strategic Research.

Data availability

No data was used for the research described in the article.

Acknowledgments

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

Appendix

This section derives the approximate linear DT dynamics in (4), which is used for the design of the Kalman filter. Let the state and input vectors be defined, respectively, as:

$$x_g(t) = \begin{pmatrix} I_{sc}(t) & I_p(t) & I_e(t) & G(t) & G_{sc}(t) \end{pmatrix}^T$$
$$u_g(t) = \begin{pmatrix} U_{sc}(t) & EGP & R_a(t) \end{pmatrix}^T$$

Let $f(x_g(t), u_g(t)) \triangleq \dot{x}_g(t)$ which is simply the system of equations on the right-hand side of (1). Let $(\bar{x}_g, \bar{u}_g)$ be the fasting (i.e., $R_a(t) \equiv 0$) equilibrium point around which the linearization is made. Then, the following holds by definition:

$$f(\bar{x}_g, \bar{u}_g) = \mathbf{0} \tag{A.1}$$

which can easily be solved by back substitution as follows. From the last scalar equation in (A.1) we get

$$\bar{G}_{sc} = \bar{G} \tag{A.2}$$

where the bar superscript denotes the steady-state values. Here, $\bar{G}$ is the target glucose level which is the set-point for the controller. For any properly calibrated sensor, the sensor glucose levels G_{sc} must converge to the actual blood glucose levels G as dictated by this equality.

By the same token, we obtain the following steady-state equalities for the remaining states and the control input:

$$\bar{I}_e = \frac{EGP}{\bar{G}} - GEZI \tag{A.3}$$

$$\bar{I}_p = \frac{\bar{I}_e}{S_I} \tag{A.4}$$

$$\bar{I}_p = \bar{I}_{sc} \tag{A.5}$$

$$\bar{U}_{sc} = \bar{I}_{sc}\, C_I \tag{A.6}$$

where $\bar{U}_{sc}$ is basal insulin delivery rate. Thus, the equilibrium point may be expressed as:

$$(\bar{x}_g,\ \bar{u}_g) = ([\bar{I}_{sc}\ \ \bar{I}_p\ \ \bar{I}_e\ \ \bar{G}\ \ \bar{G}_{sc}]^T,\ [\bar{U}_{sc}\ \ EGP\ \ 0]^T). \tag{A.7}$$

Now that we have computed the operating point for linearization, we introduce the following deviation variables:

$$\Delta x_g(t) \triangleq x_g(t) - \bar{x}_g$$
$$\Delta u(t) \triangleq U_{sc}(t) - \bar{U}_{sc}$$
$$\Delta u_g(t) \triangleq u_g(t) - \bar{u}_g = [\Delta u(t)\ \ 0\ \ R_a(t)]^T.$$

Noting (A.1), the first order Taylor series approximation of $f(\bar{x}_g, \bar{u}_g)$ around $(\bar{x}_g\ \bar{u}_g)$ reads as:

$$f(x_g, u_g) \approx \left.\frac{\partial f}{\partial x_g}\right|_{\substack{x_g=\bar{x}_g\\ u_g=\bar{u}_g}} \Delta x_g(t) + \left.\frac{\partial f}{\partial u_g}\right|_{\substack{x_g=\bar{x}_g\\ u_g=\bar{u}_g}} \Delta u_g(t) \tag{A.8}$$

From (A.8) and the identity $\Delta\dot{x}_g = \dot{x}_g$, the linearized MVP model reads as:

$$\begin{aligned}\Delta\dot{x}_g(t) &\approx A^c \Delta x_g(t) + B_u^c \Delta u(t) + B_d^c R_a(t)\\ y(t) &= C^c \Delta x_g(t)\end{aligned} \tag{A.9}$$

where the system matrices are

$$A^c = \left.\frac{\partial f}{\partial x_g}\right|_{\substack{x_g=\bar{x}_g\\ u_g=\bar{u}_g}} \tag{A.10}$$

$$= \begin{pmatrix} -1/\tau_1 & 0 & 0 & 0 & 0\\ 1/\tau_2 & -1/\tau_2 & 0 & 0 & 0\\ 0 & p_2 S_I & -p_2 & 0 & 0\\ 0 & 0 & -\bar{G} & -(GEZI + \bar{I}_e) & 0\\ 0 & 0 & 0 & 1/\tau_s & -1/\tau_s \end{pmatrix} \tag{A.11}$$

$$B_u^c = \left.\frac{\partial f}{\partial U_{sc}}\right|_{\substack{x_g=\bar{x}_g\\ U_{sc}=\bar{U}_{sc}}} = \begin{pmatrix} 1/\tau_1 C_I & 0 & 0 & 0 & 0 \end{pmatrix}^T \tag{A.12}$$

$$B_d^c = \frac{\partial f}{\partial R_a} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix}^T \tag{A.13}$$

$$C^c = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.14}$$

The dynamics in (A.9) are in continuous-time, and hence depicted by the c superscript. Please note that the constant input term EGP has vanished, but it indirectly affects the linearized dynamics through (A.3).

Next, we derive an equivalent discrete-time model for (A.9). We use different discretization methods for the insulin-glucose and meal dynamics due to the different nature of inputs applied to these systems. This is necessitated to minimize the numerical errors due to discretization with a relatively large sampling time.

A.1. Insulin-glucose dynamics

The Zero-Order Hold (ZOH) discretization method provides an exact match between the continuous- and discrete-time systems in the time domain for staircase (i.e., piece-wise constant) inputs. Let $x_g[k] = \Delta x_g(t)\big|_{t=kh}$, $u[k] = \Delta u(t)|_{t=kh}$, and $R_a[k] = R_a(t)\big|_{t=kh}$ with h being the sampling period. Then, the ZOH discrete equivalent of the system (A.9) reads as:

$$\begin{aligned}x_g[k+1] &= A_g x_g[k] + B_u^d u[k] + B_d^d R_a[k]\\ y[k] &= C_g x_g[k]\end{aligned} \tag{A.15}$$

The system matrices are as follows:

$$A_g = e^{A^c h},\ \ B_u^d = \int_0^h e^{A^c t}\, B_u^c\, dt$$

$$B_d^d = \int_0^h e^{A^c t}\, B_d^c\, dt,\ \ C_g = C^c.$$

A.2. Meal disturbance

Meal disturbance dynamics are given by (2). We naturally select the state vector as $x_m(t) = [D(t)\ R_a(t)]^T$, and the output as $R_a(t)$. With this choice, the state space equations of the meal subsystem read as

$$\begin{aligned}\dot{x}_m(t) &= A_m^c x_m(t) + B_m^c C_h(t)\\ R_a(t) &= C_m^c x_m(t)\end{aligned} \tag{A.16}$$

with

$$A_m^c = \begin{pmatrix} -1/\tau_m & 0\\ 1/\tau_m^2 V_G & -1/\tau_m \end{pmatrix},\ B_m^c = \begin{pmatrix} 1\\ 0 \end{pmatrix},\ C_m^c = \begin{pmatrix} 0 & 1 \end{pmatrix}. \tag{A.17}$$

Similar to the ZOH discretization, the impulse invariant method provides an exact match for impulsive inputs. The impulse invariant discretization of (A.16) reads as follows:

$$\begin{aligned}x_m[k+1] &= A_m x_m[k] + B_m C_h[k]\\ R_a[k] &= C_m x_m[k]\end{aligned} \tag{A.18}$$

where $x_m[k] = x_m(t)\big|_{t=kh}$ and the system matrices are $A_m = e^{A_m^c h}$, $B_m = A_m B_m^c$, $C_m = C_m^c$.

A.3. Augmented model

One can easily merge (A.15) and (A.18) to study BG dynamics in a single system as follows:

$$\begin{aligned}x[k+1] &= Ax[k] + Bu[k] + MC_h[k]\\ y[k] &= Cx[k]\end{aligned} \tag{A.19}$$

where the augmented state is $x[k] \triangleq \begin{pmatrix} x_g[k]\\ x_m[k] \end{pmatrix}$, and the system matrices are

$$A = \begin{pmatrix} A_g & B_d^d C_m\\ \mathbf{0} & A_m \end{pmatrix},\ B = \begin{pmatrix} B_u^d\\ \mathbf{0} \end{pmatrix} \tag{A.20}$$

$$M = \begin{pmatrix} \mathbf{0}\\ B_m \end{pmatrix},\ C = \begin{pmatrix} C_g & \mathbf{0} \end{pmatrix}. \tag{A.21}$$

The dynamical model (4) is then simply obtained by adding process and sensor noise to (A.19).

References

[1] G.A. Gregory, T.I. Robinson, S.E. Linklater, F. Wang, S. Colagiuri, C. de Beaufort, K.C. Donaghue, D.J. Magliano, J. Maniam, T.J. Orchard, et al., Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: A modelling study, Lancet Diabetes Endocrinol. 10 (10) (2022) 741–760.

[2] C. Berget, L.H. Messer, G.P. Forlenza, A clinical overview of insulin pump therapy for the management of diabetes: Past, present, and future of intensive therapy, Diabetes Spectrum: Publ. Am. Diabetes Assoc. 32 (3) (2019) 194.

[3] B. Kovatchev, Automated closed-loop control of diabetes: The artificial pancreas, Bioelectron. Med. 4 (1) (2018) 14.

[4] U.S. Food and Drug Administration, The artificial pancreas device system, 2018, URL https://www.fda.gov/medical-devices/consumer-products/artificial-pancreas-device-system. (Last Accessed: 16 February 2023).

[5] D. Care, Diabetes: Standards of medical, care in diabetes—2022, Diabetes Care 45 (1) (2022) S113–S124.

[6] Chunxiao Li, A. Raghunathan, N.K. Jha, Hijacking an insulin pump: Security attacks and defenses for a diabetes therapy system, in: 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services, 2011, pp. 150–156.

[7] D.C. Klonoff, Cybersecurity for connected diabetes devices, J. Diabetes Sci. Technol. 9 (5) (2015) 1143–1147.

[8] US Food and Drug Administration, Cybersecurity], 2022, URL https://www.fda.gov/medical-devices/digital-health-center-excellence/cybersecurity. (Last Accessed: 16 February 2023).

[9] H. Weng, C. Hettiarachchi, C. Nolan, H. Suominen, A. Lenskiy, Ensuring security of artificial pancreas device system using homomorphic encryption, Biomed. Signal Process. Control 79 (2023) 104044.

[10] S.M. Dibaji, M. Pirani, D.B. Flamholz, A.M. Annaswamy, K.H. Johansson, A. Chakrabortty, A systems and control perspective of CPS security, Annu. Rev. Control 47 (2019) 394–411.
[11] A. Teixeira, D. Pérez, H. Sandberg, K.H. Johansson, Attack models and scenarios for networked control systems, in: Proceedings of the 1st international conference on High Confidence Networked Systems, 2012, pp. 55–64.
[12] A. Teixeira, K.C. Sou, H. Sandberg, K.H. Johansson, Secure control systems: A quantitative risk management approach, IEEE Control Syst. Mag. 35 (1) (2015) 24–45.
[13] F. Emre Tosun, A. Teixeira, A. Ahlén, S. Dey, Detection of bias injection attacks on the glucose sensor in the artificial pancreas under meal disturbance, in: 2022 American Control Conference, ACC, 2022, pp. 1398–1405.
[14] A.A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, S. Sastry, Attacks against process control systems: Risk assessment, detection, and response, in: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, 2011, pp. 355–366.
[15] C. Zhao, Y. Fu, Statistical analysis based online sensor failure detection for continuous glucose monitoring in type I diabetes, Chemometr. Intell. Lab. Syst. 144 (2015) 128–137.
[16] K. Turksoy, A. Roy, A. Cinar, Real-time model-based fault detection of continuous glucose sensor measurements, IEEE Trans. Biomed. Eng. 64 (7) (2016) 1437–1445.
[17] X. Yu, M. Rashid, J. Feng, N. Hobbs, I. Hajizadeh, S. Samadi, M. Sevil, C. Lazaro, Z. Maloney, A. Cinar, Fault detection in continuous glucose monitoring sensors for artificial pancreas systems, IFAC-PapersOnLine 51 (18) (2018) 714–719, 10th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2018.
[18] S. Mishra, Y. Shoukry, N. Karamchandani, S.N. Diggavi, P. Tabuada, Secure state estimation against sensor attacks in the presence of noise, IEEE Trans. Control Netw. Syst. 4 (1) (2016) 49–59.
[19] S. Chen, Z. Wu, P.D. Christofides, Cyber-attack detection and resilient operation of nonlinear processes under economic model predictive control, Comput. Chem. Eng. 136 (2020) 106806.
[20] S.S. Kanderian, S. Weinzimer, G. Voskanyan, G.M. Steil, Identification of intraday metabolic profiles during closed-loop glucose control in individuals with type 1 diabetes, J. Diabetes Sci. Technol. (2009).
[21] S.S. Kanderian, S.A. Weinzimer, G.M. Steil, The identifiable virtual patient model: Comparison of simulation and clinical closed-loop study results, J. Diabetes Sci. Technol. 6 (2) (2012) 371–379.
[22] A. El Fathi, M.R. Smaoui, V. Gingras, B. Boulet, A. Haidar, The artificial pancreas and meal control: An overview of postprandial glucose regulation in type 1 diabetes, IEEE Control Syst. Mag. 38 (1) (2018) 67–85.
[23] S.X. Ding, Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools, Springer Science & Business Media, 2013.
[24] N. Resalat, J. El Youssef, R. Reddy, J. Castle, P.G. Jacobs, Adaptive tuning of basal and bolus insulin to reduce postprandial hypoglycemia in a hybrid artificial pancreas, J. Process Control 80 (2019) 247–254.
[25] D. Rhyner, H. Loher, J. Dehais, M. Anthimopoulos, S. Shevchik, R.H. Botwey, D. Duke, C. Stettler, P. Diem, S. Mougiakakou, et al., Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: A comparative study, J. Med. Internet Res. 18 (5) (2016) e5567.
[26] R. Diversi, R. Guidorzi, U. Soverini, Kalman filtering in extended noise environments, IEEE Trans. Automat. Control 50 (9) (2005) 1396–1402.
[27] B.D. Anderson, J.B. Moore, Optimal Filtering, Courier Corporation, 2012.
[28] M. Fauß, A.M. Zoubir, H.V. Poor, Minimax robust detection: Classic results and recent advances, IEEE Trans. Signal Process. 69 (2021) 2252–2283.
[29] A. Tartakovsky, I. Nikiforov, M. Basseville, Sequential Analysis: Hypothesis Testing and Changepoint Detection, CRC Press, 2014.
[30] S. Kullback, R.A. Leibler, On information and sufficiency, The Ann. Math. Stat. 22 (1) (1951) 79–86.
[31] L. Xie, S. Zou, Y. Xie, V.V. Veeravalli, Sequential (quickest) change detection: Classical results and new directions, IEEE J. Sel. Areas Inf. Theory 2 (2) (2021) 494–514.
[32] Q. Zhang, K. Liu, A.M.H. Teixeira, Y. Li, S. Chai, Y. Xia, An online Kullback-Leibler divergence-based stealthy attack against cyber–physical systems, IEEE Trans. Automat. Control (2022) 1–8.
[33] M. Pollak, Optimal detection of a change in distribution, Ann. Statist. (1985) 206–227.
[34] H. Sandberg, V. Gupta, K.H. Johansson, Secure networked control systems, Annu. Rev. Control Robot. Autonom. Syst. 5 (2022) 445–464.
[35] J. Unnikrishnan, V.V. Veeravalli, S.P. Meyn, Minimax robust quickest change detection, IEEE Trans. Inform. Theory 57 (3) (2011) 1604–1614.
[36] Z. Mahmoudi, F. Cameron, N.K. Poulsen, H. Madsen, B.W. Bequette, J.B. Jørgensen, Sensor-based detection and estimation of meal carbohydrates for people with diabetes, Biomed. Signal Process. Control 48 (2019) 12–25.
[37] A. Haidar, The artificial pancreas: How closed-loop control is revolutionizing diabetes, IEEE Control Syst. Mag. 36 (5) (2016) 28–47.