Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 58-4 (2024) 508–513

Kullback–Leibler Divergence-Based Tuning of Kalman Filter for Bias Injection Attacks in an Artificial Pancreas System

Fatih Emre Tosun * **André Teixeira** ** **Anders Ahlén** * **Subhrakanti Dey** *

* *Department of Electrical Engineering, Uppsala University, 751 03 Uppsala, Sweden (email: {fatihemre.tosun, anders.ahlen, subhrakanti.dey}@angstrom.uu.se).*

** *Department of Information Technology, Uppsala University, 751 03 Uppsala, Sweden (email: andre.teixeira@it.uu.se)*

Abstract: This paper considers constant bias injection attacks on the glucose sensor deployed in an artificial pancreas system. The main challenge with such apparently simple attacks is that they are detectable for only a limited duration if the system is linear and has an integrator. More formally put, such attacks are steady-state stealthy. To address this issue, we propose a method to design a bias-sensitive Kalman filter based on the Kullback–Leibler divergence metric. The resulting filter outperforms the nominal Kalman filter for attack detection as illustrated by numerical simulations on a realistic model.

Copyright © 2024 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: Filtering and change detection, artificial pancreas, Kullback–Leibler divergence, cyber–physical systems security, sensor deception attack.

1. INTRODUCTION

In healthy individuals, blood glucose levels (BGL) are regulated by the secretion of insulin and glucagon hormones from the pancreas. However, patients with type 1 diabetes (T1D) require external insulin administration for survival due to loss of insulin secretion caused by the autoimmune destruction of pancreatic beta cells (Atkinson et al., 2014). The major goal in treating T1D is to maintain the BGL in a safe range (70-180 mg/dl) most of the time (Care, 2022). Successful management of T1D involves meticulous tasks of regularly checking BGL and adjusting insulin doses as needed. In particular, insufficient insulin administration leads to hyperglycemia (high BGL) which causes health complications such as retinopathy (blindness) and nephropathy (kidney failure) in the long term (Bequette et al., 2018). At the other extreme, hypoglycemia (low BGL) is a short-term risk that causes dizziness, seizure, coma, and even death (Turksoy et al., 2018).

To alleviate the burden of self-managing T1D and achieve tighter control of BGL, the concept of automated insulin delivery termed the artificial pancreas (AP) has been an active research endeavor, especially after the advent of continuous glucose monitors (CGM) in the mid-2000s (Haidar, 2016). In particular, CGM are minimally invasive sensor technology that measure the interstitial glucose levels in the subcutaneous tissue (Vashist, 2013). Interstitial glucose levels are a good proxy for BGL albeit with a certain delay (Cengiz and Tamborlane, 2009). The control algorithm determines the appropriate dose of insulin to be injected by the pump based on the real-time glucose readings.

Modern CGM are wearable glucose sensors that communicate over wireless networks such as Bluetooth® (Saltzstein, 2020). While wireless networks grant portability and flexibility for AP systems, they can also be a portal for cyber-attacks if not protected adequately (Chunxiao Li et al., 2011). To this end, complementary to network security measures such as authentication and encryption, we take a control-theoretic security approach for the safety of AP systems. In particular, we propose a model-based detection scheme against bias injection attacks on the CGM. Despite its apparent simplicity, a constant bias injection attack is steady-state stealthy for linear systems with integrators, i.e., one or more poles on the unit disk (Miloševič et al., 2017). Moreover, the attack impact can be arbitrarily large unless detected during the transients (Teixeira et al., 2015). Hence, in order to ensure the safety of the AP user, it is imperative to detect and mitigate these attacks as early as possible.

In control systems, model-based anomaly detection involves two subsequent steps: residual generation and evaluation (Ding, 2013). A residual is the difference between the measured and predicted outputs. It is expected to have a low magnitude during normal operation. The idea of utilizing Kalman filters as residual generators in linear stochastic systems dates back to the seminal paper by (Mehra and Peschon, 1971). To date, the celebrated Kalman filter is still arguably the most common choice in such systems (Sandberg et al., 2022). In this work, we propose a method to tune the Kalman filter for enhanced anomaly detection by maximizing the statistical gap between the normal and attacked residual measurements. This gap is quantified by the following information-theoretical metric (Basseville et al., 1993):

2405-8963 Copyright © 2024 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2024.07.269

Definition 1. (Kullback-Liebler divergence). Let p and q be two random variables (possibly vectors) with joint probability densities $f_p(\cdot)$ and $f_q(\cdot)$, respectively. Then, the Kullback-Liebler divergence (KLD) between p and q is

$$D(p \parallel q) \triangleq \int_{\{\zeta | f_p(\zeta)>0\}} f_p(\zeta) \log \frac{f_p(\zeta)}{f_q(\zeta)} d\zeta. \tag{1}$$

The KLD is a non-negative measure of dissimilarity between two distributions. It is equal to zero only when the distributions are identical, i.e., $D(p_k \| q_k) = 0 \Leftrightarrow f_p = f_q$. It is also used to quantify the stealthiness level of false data injection (FDI) attacks in cyber-physical control systems (Bai et al., 2017). Hence, maximizing the KLD between the normal and attacked residual distributions amounts to maximizing detectability. The novelty of our approach lies in the fact that the Kalman gain for optimal attack detection is not necessarily the same as the one for optimal state estimation. To put it in another way, we propose to tune the filter to be more sensitive to bias injection attacks at the expense of a higher estimation error.

The rest of the paper is outlined as follows. In Section 2, we lay out the problem formulation. Section 3 presents the proposed residual generation method based on Kalman filtering. Section 4 illustrates the efficacy of the proposed method by numerical simulations. We conclude the paper in Section 5 with possible extensions.

Notation: The sets of natural, real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$, respectively. The space of n-dimensional Euclidean vectors is denoted by $\mathbb{R}^n$. The superscript ^{T} implies matrix transposition. The identity matrix of order n is denoted by I_n. The set of symmetric positive semi-definite matrices of order n is denoted by $\mathbb{S}^n_+$. For an $M \in \mathbb{S}^n_+$, $\sqrt{M}$ denotes its Cholesky decomposition. A normal variable X with mean μ and (co)variance Σ is denoted as $X \sim \mathcal{N}(\mu, \Sigma)$.

2. PROBLEM SETUP

We assume the AP system is modeled by the following discrete-time (DT), linear time-invariant (LTI) dynamics:

$$\begin{aligned} x[k+1] &= Ax[k] + Bu[k] + w[k] \\ y_m[k] &= Cx[k] + v[k] + a[k] \end{aligned} \tag{2}$$

where $x[k] \in \mathbb{R}^n$ is the system state, $u[k] \in \mathbb{R}^m$ is the known input (i.e., the insulin infusion rate), $w[k] \in \mathbb{R}^n$ is the process noise, $y_m[k] \in \mathbb{R}$ is the BGL measurement, $v[k] \in \mathbb{R}$ is the sensor noise and $a[k] \in \mathbb{R}$ is the FDI attack at time $k \in \mathbb{N}$. The stochastic disturbances $w[k]$ and $v[k]$ are independent white Gaussian processes with covariance $Q \in \mathbb{S}^n_+$ and variance $R > 0$, respectively.

We assume the pair (A, C) is detectable and $(A, \sqrt{Q})$ is stabilizable which are common regularity conditions in feedback systems (Gustafsson, 2000). For residual generation, we use an LTI filter as follows:

$$\hat{x}[k+1] = (A - LC)\hat{x}[k] + Bu[k] + Ly[k] \tag{3}$$

$$\hat{y}_m[k] = C\hat{x}[k] \tag{4}$$

where L is the filter gain, $\hat{x}[k]$ is the state estimate and $\hat{y}_m[k]$ is the output prediction at k. Let $\tilde{x}[k] = x[k] - \hat{x}[k]$ be the state estimation error and $r_m[k] = y_m[k] - \hat{y}_m[k]$ be the residual measurement. Then, the residual dynamics may be expressed as:

$$\begin{aligned} \tilde{x}[k+1] &= (A - LC)\tilde{x}[k] + w[k] - L(v[k] + a[k]) \\ r_m[k] &= C\tilde{x}[k] + v_k + a[k] \end{aligned} \tag{5}$$

The error dynamics are driven by two sources, namely, the natural stochastic disturbances $w[k]$, and the deterministic FDI attack $a[k]$. Exploiting the superposition principle, we decompose $\tilde{x}[k]$ as $\tilde{x}[k] = \tilde{x}_s[k] + \tilde{x}_a[k]$ where:

$$\tilde{x}_s[k] \triangleq \tilde{x}[k]|_{a[k]=0}\,,\ \tilde{x}_a[k] \triangleq \tilde{x}[k]|_{w[k]=0,\, v[k]=0}\,. \tag{6}$$

We also decompose $r_m[k]$ into $r_s[k]$ and $r_a[k]$ in the same manner. Thus, the isolated contribution of the FDI attack to the residual is manifested by the following dynamics:

$$\begin{aligned} \tilde{x}_a[k+1] &= (A - LC)\tilde{x}_a[k] - La[k],\ \tilde{x}_a[0] = 0 \\ r_a[k] &= C\tilde{x}_a[k] + a[k] \end{aligned} \tag{7}$$

From linear stochastic systems theory, we know that $\tilde{x}_s[k] \sim \mathcal{N}(0, \Sigma)$ where Σ solves the following DT Lyapunov equation (Glad and Ljung, 2018):

$$\Sigma = (A - LC)\Sigma(A - LC)^T + Q + LRL^T. \tag{8}$$

The estimation error under a deterministic FDI attack is also a Gaussian with the same covariance but with a time-varying mean depending on the attack sequence. Thus, the residual statistics for the unattacked and attacked cases become $r_s[k] \sim \mathcal{N}(0, \Sigma_r)$ and $r_m[k] \sim \mathcal{N}(r_a[k], \Sigma_r)$, respectively where $\Sigma_r = C\Sigma C^T + R$.

The formulation up to now is valid for any generic deterministic FDI attack. However, in this work, we restrict our analysis to constant bias injection attacks as follows:

$$a[k] = \bar{a}\ H[k - k_a] \tag{9}$$

where $\bar{a}$ is the amount of injected bias, k_a is the attack start time and $H[k]$ is the unit step function. To maximize the detectability, ideally, we would like to find the optimal filter gain L that maximizes $D(r_s[k] \| r_m[k])$ uniformly over all times $k \geq 0$ and stabilizes (5). However, such an L does not exist, except for some possible degenerate cases. In the present case, $D(r_s[k] \| r_m[k]) = r_a[k]^2/\Sigma_r$ which is basically the signal-to-noise ratio (SNR). For the sake of theoretical exposition, without loss of generality, we set $k_a = 0$ and $\bar{a} = 1$. Thus, from (7) we get:

$$r_a[k] = 1 - C \sum_{\tau=0}^{k-1} (A - LC)^{k-\tau-1} L. \tag{10}$$

As seen from (10), the KLD number $D(r_s[k] \| r_m[k])$ is time-varying; thus the optimal L in general will be different at each time instant k. A conceivable approach instead is to choose a particular time instant at which the KLD is maximized.

Proposition 1. The nominal Kalman filter maximizes the KLD at the attack onset $k = k_a = 0$.

Proof. From (7), we get $D(r_s[0] \| r_m[0]) = 1/\Sigma_r$. Thus, maximizing $D(r_s[0] \| r_m[0])$ is equivalent to minimizing the residual variance Σ_r for which the Kalman filter is the solution (Glad and Ljung, 2018). □

In this work, we propose to maximize the KLD at the next time instance after the attack instead of at the onset. The rationale behind this idea is to slow down the rate at which the filter step response decays to zero so that the attack remains detectable for longer. Thus, we formulate the problem of finding the optimal residual generator as follows:

Problem 1.

$$\max_{L\in\mathbb{R}^n} D(r_a[1] \parallel r_m[1]) = \frac{(1-CL)^2}{C\Sigma(L)C^T+R} \quad \text{s.t. } \rho(A-LC)<1 \text{ and (8)} \tag{11}$$

where $\rho(\cdot)$ is the spectral radius, that is, the largest magnitude of all eigenvalues. The notation $\Sigma(L)$ is used to emphasize the implicit dependence of the error covariance on the filter gain through (8). The effectiveness of this seemingly *ad hoc* approach is illustrated in Section 4.

For residual evaluation, we consider the stateless χ^2 test as follows:

$$g[k] = r_m[k]^2/\Sigma_r > \tau. \tag{12}$$

The test statistic $g[k]$ follows a one-degree-of-freedom χ^2 distribution. The decision threshold τ is chosen according to a desired false alarm probability (FAP) of α as follows:

$$\alpha = \mathbb{P}(g[k] > \tau \mid a[k] = 0). \tag{13}$$

The value for τ may be computed by using a statistical software package or a distribution table. The reason for choosing such a simple test for residual evaluation is twofold. First, the focus of the paper is on residual generation rather than evaluation. Second, this test is optimal in the sense of maximizing the instantaneous detection probability for detecting mean changes in Gaussian sequences with fixed variance (Tosun and Teixeira, 2023).

3. DESIGN OF RESIDUAL GENERATOR

This section proposes a principled way to design a residual generator in light of the discussion in the previous section. In general, Problem 1 is intractable as the objective function to maximize is not concave, and it has a nonlinear inequality constraint. Nevertheless, a solution can be obtained by grid search. However, it is not obvious how to select the search grid. Moreover, grid search suffers from the curse of dimensionality (Hutter et al., 2019). A further complication exists for higher-order systems as there is no closed-form expression for the stability constraint for $n \geq 3$. To circumvent this issue, we resort to a different approach to deal with a more tractable problem. Namely, we tune a Kalman filter for enhanced detection in lieu of its *raison d'etre* of optimal state estimation. More precisely, the objective function in Problem 1 remains the same, but the decision variable L is now restricted to be a Kalman gain as explained in the sequel. Consequently, the stability constraint can be removed since $A-LC$ is a stable matrix for any Kalman gain (Glad and Ljung, 2018).

We assume the system (2) to be perfectly known. The Kalman filter gain is obtained by solving the following DT algebraic Riccati equation:

$$P = A[P - PC^T(CPC^T + R')^{-1}CP]A^T + Q \tag{14}$$

where P is the filter covariance, and R' is the sensor variance assumed by the filter, which is the tuning parameter in our case. The corresponding Kalman gain reads as:

$$L = APC^T(CPC^T + R')^{-1} \tag{15}$$

Please note that the filter error covariance P is in general not equal to the true error covariance Σ. Rather, it is a quantity needed for the computation of the filter gain L. Only for the nominal Kalman filter with $R' = R$, it holds that $P = \Sigma$ and Σ_r attains its minimum. However, our goal is to maximize the SNR in (11). Thus, using R' as the tuning parameter, we seek the L gain that maximizes the SNR for bias detection as follows:

Algorithm 1 Grid Search for Tuning the Kalman Filter

Data: The system matrices (A, B, C) and the grid values for the filter sensor variance $R_j^{'}$
1: **Initialize:** $j = 0$
2: **for each** $R_j^{'}$ **do** ▷ for a particular grid value of $R^{'}$
3: **Set:** $R^{'} = R_j^{'}$
4: Compute P by solving (14)
5: $L_j = APC^T(CPC^T + R')^{-1}$ ▷ from (15)
6: Compute $\Sigma(L_j)$ by solving (8)
7: Store L_j in the set $\mathcal{L}$
8: $j = j + 1$
9: **end for**
10: **Return:** $L = \arg\max_{L\in\mathcal{L}} D(r_a[1] \parallel r_m[1])$

Problem 2.

$$\max_{R'>0} \frac{(1-CL(R^{'}))^2}{C\Sigma(L(R'))C^T+R} \quad \text{s.t. (8), (14) and (15)} \tag{16}$$

In other words, we parameterize a subset of the stabilizing filters over the filter variance $R^{'}$. This is a simpler problem than the first one because not only the stability constraint is eliminated, but also the decision variable R' is a scalar regardless of the system order n. This greatly reduces the search space and makes it scalable to higher-order systems. The range of the search grid may then be the set $[\epsilon,\ \delta R]$, where ϵ is an arbitrarily small number and δ is some scaling factor for the filter variance.

At first glance, (16) seems too hard to solve due to nonlinear equality constraints. However, it can easily be converted into an unconstrained problem by direct substitution since efficient solvers for (8) and (14) are readily available (e.g., Matlab commands *dlyap* and *idare*, respectively). We summarize the proposed solution method to Problem 2 in Algorithm 1. Even though this method yields merely a suboptimal solution to Problem 1, the attack detection capability of the resulting filter may be significantly improved compared to the nominal Kalman filter as demonstrated in the next section.

4. NUMERICAL SIMULATIONS

In this section, we demonstrate the enhanced detectability of bias injection attacks due to the proposed residual generation method. To this end, we provide a real-world numerical example within the context of AP systems. In particular, we consider a recent continuous-time LTI model developed to describe the blood glucose dynamics, which has undergone clinical validation (Magdelaine et al., 2015). The authors suggest that the integrator in their model enables a longer-term fit and a more accurate glycemic equilibrium characteristic. The model equations without the meal disturbance read as follows:

$$\frac{d}{dt}x(t) = \begin{pmatrix} 0 & -k_{si} & 0 \\ 0 & 0 & 1 \\ 0 & -\frac{1}{T_u^2} & -\frac{2}{T_u} \end{pmatrix} x(t) + \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ \frac{\theta_2}{T_u^2} & 0 \end{pmatrix} \begin{pmatrix} u_c(t) \\ \theta_1 \end{pmatrix} \tag{17}$$

Table 1. Model parameters for patient IF2

Parameter	Value	Meaning
θ_1 [mg/dl/min]	0.162	Net endogenous glucose production
k_{si} [mg/IU/min]	197	Insulin sensitivity
T_u [min]	122	Time constant for insulin absorption
θ_2 [min/dl]	0.059	Static gain for insulin absorption

with $x(t) = (G(t)\ I(t)\ dI(t)/dt)^T$. In particular, $G(t)$ [mg/dl] is the output (i.e., BGL), $I(t)$ [IU/dl] is the plasma insulin level, and IU is the international unit for human insulin. The control input $u_c(t)$ [IU/min] is the insulin injection rate. The parameter values identified for a certain patient are reported in Table 1.

We simulate (17) with the values in Table 1 as a sampled-data system with the sampling time $h = 5$ min, a typical rate for CGM. The zero-order hold discretization of (17) yields the following system matrices:

$$A = \begin{pmatrix} 1 & -984.7 & -2396 \\ 0 & 0.9992 & 4.799 \\ 0 & -3.22 \times 10^{-4} & 0.9205 \end{pmatrix}, \ B = \begin{pmatrix} -0.016 & 5 \\ 4.82 \times 10^{-5} & 0 \\ 1.90 \times 10^{-5} & 0 \end{pmatrix}$$

$$C = (1\ 0\ 0)\,.$$

Since we simulate a deterministic continuous-time system, we select a small process noise covariance as $Q = 10^{-12} I_3$. We set $R = 30$ to allow a glucose sensor uncertainty of 10 mg/dl (Bequette et al., 2018). The grid search for Problem 2 yields the optimum at $R' = 115$ as shown in Fig. 1. In the sequel, we denote the nominal Kalman gain with $R' = 30$ as K_0, and the bias-sensitive Kalman gain with $R' = 115$ as K_1.

Next, to illustrate the superiority of the bias-sensitive Kalman filter, we consider a realistic attack scenario. To this end, we employ a PD controller of the following form:

$$e[k] = y_m[k] - G_0 \tag{18}$$

$$P[k] = K_p e[k] \tag{19}$$

$$D[k] = \frac{NT_d}{h + NT_d} D[k-1] + \frac{K_p T_d}{h + NT_d}(e[k] - e[k-1]) \tag{20}$$

$$u_c[k] = u_b + P[k] + D[k] \tag{21}$$

where G_0 [mg/dl] is the set-point, K_p [(IU/min)/(mg/dl)] is the proportional gain, T_d [min] is the derivative time constant, N is the filter coefficient for the derivative term, and u_b [IU/min] is the basal (steady-state) insulin rate. The target glucose level is set to $G_0 = 85$. Following the guidelines in (Huyett et al., 2015), we choose the following control parameters: $K_p = 0.0015$, $T_d = 20$, $N = 0.1$ and $u_b = 0.014$.

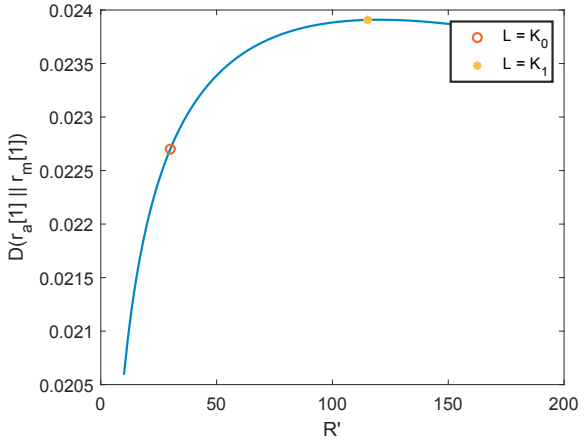


Fig. 1. Grid search over R' for solving Problem 2

We assume the attacker uses a first-order low pass filter to adjust the stealthiness level of the bias injection as follows:

$$a_{k+1} = (1 - \beta)\, a_k + \beta\, \bar{a}\, H[k - k_a]. \tag{22}$$

In particular, β determines the speed and stealthiness of the bias injection. For $\beta = 1$, (22) reduces to the constant step function (9), which is the least stealthy bias injection. To simulate a slow bias injection, we set $\beta = 0.1$. Let $\bar{a} = 15$ so that the controller will be forced to cross the hypoglycemic threshold of 70 mg/dl. As shown in Figs. 2a and 2b, K_1 yields a higher KLD than K_0 at all times except at $k = 0$ in both cases.

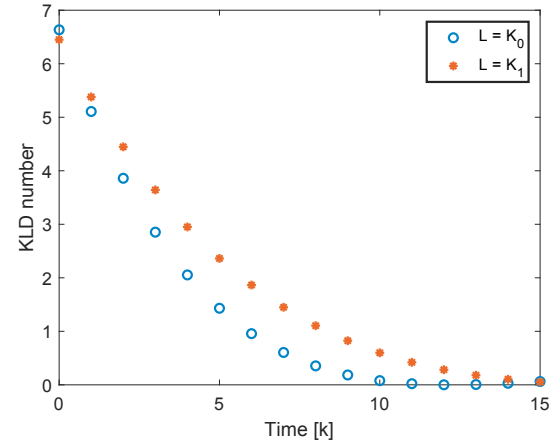
The KLD number is hard to interpret as is. To put the numbers in Figs. 2a and 2b into perspective, we compute the corresponding detection probabilities using the χ^2 test (13) with $\alpha = 0.05$. The probabilities are computed via Monte Carlo simulations and reported in Figs. 2c and 2d. The KLD number and the detection probability follow the same trend as expected. The bias-sensitive Kalman filter results in enhanced detectability for a longer time window. The difference is even more dramatic for the slow and stealthier bias injection as shown in Fig. 2d.

Finally, we show a representative sample from simulations with the slow bias injection scenario in Fig. 3. The attack parameters remain the same except now the starting time is 1 h. The attack is detected at a time much earlier than when the hypoglycemic threshold is crossed. As expected, the bias-sensitive Kalman filter K_1 detects the attack earlier than its competitor K_0.

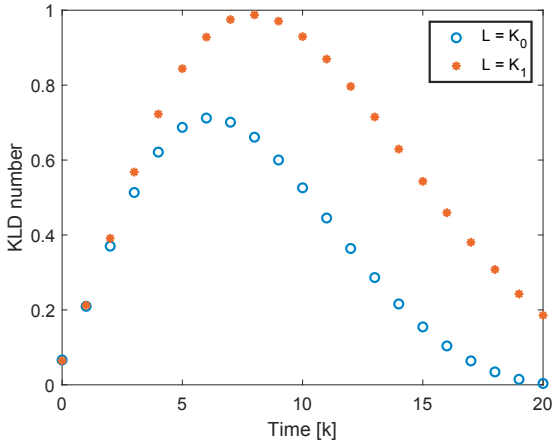
Remark 1. It is important to note that, in principle, the attacker may perform an arbitrarily slow bias injection attack which results in practically complete stealthiness in the sense that $D(r_s[k] \| r_m[k]) \approx 0$ at all times. This is true for any linear residual generator including the global solution to Problem 1. However, such attacks are harder to stage as they require to access and modify the sensor readings for a longer period before causing serious harm. Thus, while the proposed method makes it harder for the attacker to be stealthy, it is not an absolute solution for the vulnerability to bias injections. A conceivable remedy for this issue could be to raise an alarm when the insulin injection rate $u_c(t)$ differs by a certain factor from the basal rate u_b without a good reason such as a meal intake.

5. CONCLUSION

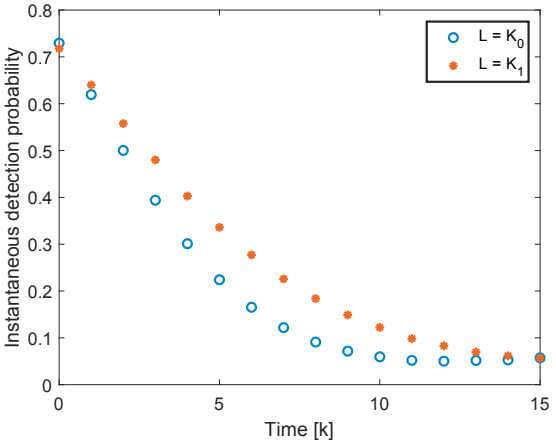
In this paper, we proposed a novel residual generation method based on Kalman filtering and the KLD metric for enhancing the detection of bias injection attacks. An AP system is used as an illustrative example to show the effectiveness of the proposed method. For future work, we plan to study the analytical properties of the solution to Problem 2 to come up with a more efficient algorithm than a simple grid search. We also plan to extend our study to multiple output systems using more efficient grid search algorithms such as with Bayesian optimization.



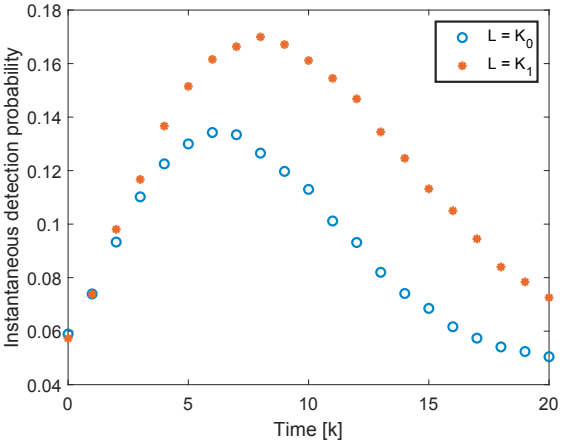
(a) KLD versus time for $\beta = 1$



(b) KLD versus time for $\beta = 0.10$

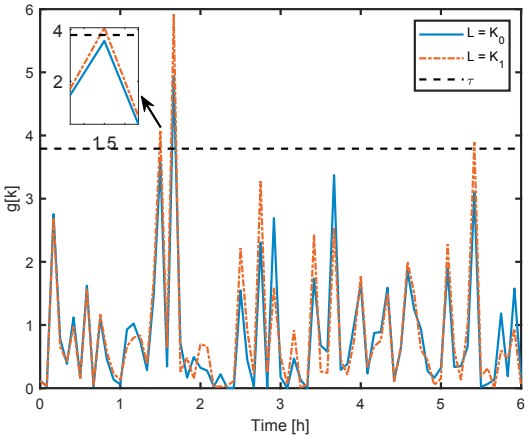


(c) Detection probability versus time for $\beta = 1$

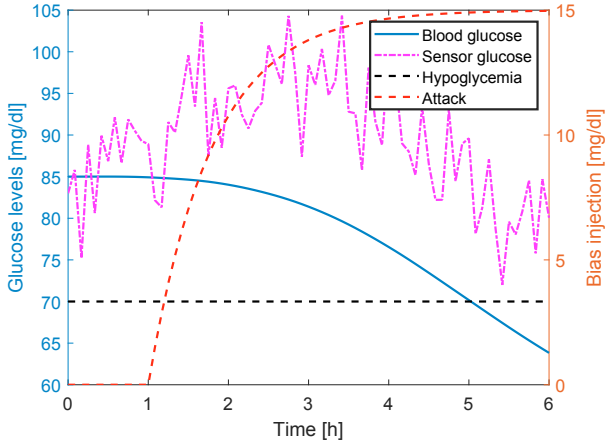


(d) Detection probability versus time for $\beta = 0.10$

Fig. 2. The top panel shows the KLD plots of the filters K_0 and K_1 for a step and a slow bias injection attack. The bottom panel shows the attack detection probability using the χ^2 test with an FAP of 5%. The bias-sensitive Kalman filter greatly increases the detection probability, especially for the slow bias injection.



(a) The χ^2 test statistics of the residuals generated by K_0 and K_1 with an FAP of 5%.



(b) The blood and sensor glucose levels versus time are plotted on the left axis. The attack sequence is plotted on the right axis.

Fig. 3. The right panel plots the residual evaluation using the χ^2 test with the residual generators K_0 and K_1. The left panel plots the trajectories of the blood and sensor glucose along with the slow bias injection attack sequence. The bias injection becomes stealthy after a certain while. While both the nominal and bias-sensitive Kalman filters were able to detect the slow bias injection before the attack became stealthy, the latter detected it earlier.

ACKNOWLEDGEMENTS

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

REFERENCES

Atkinson, M.A., Eisenbarth, G.S., and Michels, A.W. (2014). Type 1 diabetes. *The Lancet*, 383(9911), 69–82.

Bai, C.Z., Pasqualetti, F., and Gupta, V. (2017). Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs. *Automatica*, 82, 251–260.

Basseville, M., Nikiforov, I.V., et al. (1993). *Detection of abrupt changes: theory and application*, volume 104. prentice Hall Englewood Cliffs.

Bequette, B.W., Cameron, F., Buckingham, B.A., Maahs, D.M., and Lum, J. (2018). Overnight hypoglycemia and hyperglycemia mitigation for individuals with type 1 diabetes: how risks can be reduced. *IEEE Control Systems Magazine*, 38(1), 125–134.

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

Cengiz, E. and Tamborlane, W.V. (2009). A tale of two compartments: interstitial versus blood glucose monitoring. *Diabetes technology & therapeutics*, 11(S1), S–11.

Chunxiao Li, Raghunathan, A., and Jha, N.K. (2011). 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*, 150–156. doi: 10.1109/HEALTH.2011.6026732.

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

Glad, T. and Ljung, L. (2018). *Control theory*. CRC press.

Gustafsson, F. (2000). *Adaptive filtering and change detection*, volume 1. Wiley New York.

Haidar, A. (2016). The artificial pancreas: How closed-loop control is revolutionizing diabetes. *IEEE Control Systems Magazine*, 36(5), 28–47.

Hutter, F., Kotthoff, L., and Vanschoren, J. (2019). *Hyperparameter Optimization*. Springer International Publishing.

Huyett, L.M., Dassau, E., Zisser, H.C., and Doyle III, F.J. (2015). Design and evaluation of a robust pid controller for a fully implantable artificial pancreas. *Industrial & engineering chemistry research*, 54(42), 10311–10321.

Magdelaine, N., Chaillous, L., Guilhem, I., Poirier, J.Y., Krempf, M., Moog, C.H., and Le Carpentier, E. (2015). A long-term model of the glucose–insulin dynamics of type 1 diabetes. *IEEE Transactions on Biomedical Engineering*, 62(6), 1546–1552. doi: 10.1109/TBME.2015.2394239.

Mehra, R.K. and Peschon, J. (1971). An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica*, 7(5), 637–640.

Miloševič, J., Tanaka, T., Sandberg, H., and Johansson, K.H. (2017). Analysis and mitigation of bias injection attacks against a kalman filter. *IFAC-PapersOnLine*, 50(1), 8393–8398.

Saltzstein, W. (2020). Bluetooth wireless technology cybersecurity and diabetes technology devices. *Journal of diabetes science and technology*, 14(6), 1111–1115.

Sandberg, H., Gupta, V., and Johansson, K.H. (2022). Secure networked control systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 5, 445–464.

Teixeira, A., Shames, I., Sandberg, H., and Johansson, K.H. (2015). A secure control framework for resource-limited adversaries. *Automatica*, 51, 135–148.

Tosun, F.E. and Teixeira, A. (2023). Robust sequential detection of non-stealthy sensor deception attacks in an artificial pancreas system. In *2023 IEEE 62nd Conference on Decision and Control (CDC)*. IEEE. Accepted.

Turksoy, K., Littlejohn, E., and Cinar, A. (2018). Multimodule, multivariable artificial pancreas for patients with type 1 diabetes: regulating glucose concentration under challenging conditions. *IEEE Control Systems Magazine*, 38(1), 105–124.

Vashist, S.K. (2013). Continuous glucose monitoring systems: a review. *Diagnostics*, 3(4), 385–412.