CUSUM and Chi-squared Attack Detection of Compromised Sensors

Carlos Murguia and Justin Ruths

***Abstract*— A vector-valued model-based cumulative sum (CUSUM) procedure is proposed for identifying falsified sensor measurements. To fulfill a desired detection performance, given the system dynamics, we derive tools for tuning the CUSUM procedure. We characterize the state degradation that a stealthy attacker can induce to the system while remaining undetected by the detection procedure. We quantify the advantage of using a dynamic detector (CUSUM), which leverages the history of the state, over a static detector (chi-squared) which uses a single measurement at a time. Simulation experiments are presented to illustrate the performance of the detection scheme.**

I. Introduction

During the past half-century, scientific and technological advances have greatly improved the performance of control systems. From heating/cooling devices in our homes, to cruise-control in our cars, to robotics in manufacturing centers. However, these new technologies have also led to vulnerabilities of some our most critical infrastructures - e.g., power, water, transportation. Advances in communication and computing power have given rise to adversaries with enhanced and adaptive capabilities. Depending on their resources, attackers may deteriorate the functionality of systems while remaining undetected. Therefore, designing efficient attack detection schemes and attack-robust control systems is of key importance for guaranteeing the safety and proper operation of critical systems. Tools from sequential analysis and fault detection have to be adapted to deal with the systematic, strategic, and persistent nature of attacks. These new challenges have attracted the attention of many researchers in the control and computer science communities [1]-[7].

Most of the current work on security of control systems has focused on *static* detection procedures (either bad-data or chi-squared detectors), which identify anomalies based on a single measurement at a time [2]-[4]. Nevertheless, a complete and clear characterization of how features of the system (e.g., system matrices, control/estimator gains, noise, sampling) affect the performance of the detector (e.g., state deviation, false alarm rate) is not completely addressed in the literature. There are some papers in this direction already. For instance, in [5],[6], the authors relate detector properties to estimation errors for a class of scalar linear stochastic systems. They quantify how much the attacker can degrade the estimate of the state while remaining undetected. In the same spirit, the authors in [1],[2] study how the attacker propagates its effect through the control structure to degrade the system dynamics while remaining stealthy with respect to the detection mechanism. There is only a small amount of literature considering the use of dynamic change detection procedures such as the Sequential Probability Ratio Test (SPRT) or the Cumulative Sum (CUSUM) [8], which employ measurement history, in the context of security of Cyber-Physical Systems (CPS) [7]. Dynamic detectors present an appealing alternative to the aforementioned static procedures. Using measurement history provides extra degrees of freedom for improving the performance of our attack detection strategies; in particular, against low amplitude persistent attacks.

This paper addresses the problem of detecting compromised sensors in Linear Time-Invariant (LTI) systems subject to sensor/actuator noise, using chi-squared static and CUSUM dynamic detectors. Standard Kalman filters are proposed to estimate the state of the physical process; then, these estimates are used to construct distance measures between sensor measurements and the estimated outputs coming from the Kalman filter. These distance measures are accumulated such that if its accumulation is more than expected an alarm is triggered indicating a possible compromised sensor. This is the well-known change detection procedure referred to as CUSUM. In this manuscript, we propose a quadratic form in the residual variables (the differences between sensor observations and estimated outputs) as distance measure. We provide systematic tools for tuning the CUSUM procedure given the system dynamics, the Kalman filter, the stochastic properties of the distance measure, and a desired false alarm rate. In particular, sufficient conditions for mean square boundedness of the CUSUM sequence are derived when it is driven by the quadratic form of the residuals. Then, using a Markov a chain approximation of the CUSUM sequence, we give a procedure for selecting the *decision threshold* such that a desired *false alarm rate* is satisfied. For a class of stealthy attacks, we characterize the impact of the attack sequence on the system dynamics when the vector-valued CUSUM is deployed for attack detection. We use the well-known chi-squared procedure (which is a quadratic distance measure compared to a threshold) as a benchmark to compare the performance of the CUSUM. In order to do so, we also provide tools for tuning the chi-squared procedure to achieve a desired false alarm rate.

In our previous work [9], we have started analyzing these ideas for the sensor-wise case, i.e., when there is dedicated detector on each sensor (or on any sensor we want to include

This work was supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2014NCR-NCR001-40) and administered by the National Cybersecurity R&D Directorate.

C. Murguia and J. Ruths are with the Engineering Systems and Design Pillar, Singapore University of Technology and Design. emails: murguia_rendon@sutd.edu.sg, justinruths@sutd.edu.sg

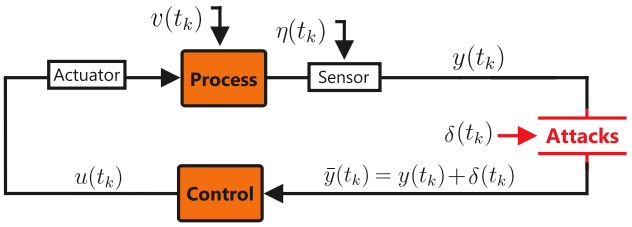

Fig. 1. Cyber-physical system under attacks on the sensor measurements.

in the detection scheme). We have compared the CUSUM performance against scalar bad-data procedures. Here, we present a generalization to a vector-valued detection scheme with an aggregate detector. We also compare the performance of the CUSUM against vector-valued chi-squared detectors. While on the surface these papers have similar structure and build on a common set of fundamental results, they represent different original and self-contained results. In particular, the proofs of these results (the full exposition of which is limited due to space) require different tools to tackle.

II. System Description & Attack Detection

We study LTI stochastic systems of the form:

$$\begin{cases} x(t_{k+1}) = Fx(t_k) + Gu(t_k) + v(t_k), \\ y(t_k) = Cx(t_k) + \eta(t_k), \end{cases} \tag{1}$$

with sampling time-instants $t_k, k \in \mathbb{N}$, state $x \in \mathbb{R}^n$, measured output $y \in \mathbb{R}^m$, control input $u \in \mathbb{R}^l$, matrices F, G, and C of appropriate dimensions, and i.i.d. multivariate zero-mean Gaussian noises $v \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^m$ with covariance matrices $R_1 \in \mathbb{R}^{n\times n}$, $R_1 \geq 0$ and $R_2 \in \mathbb{R}^{m\times m}$, $R_2 \geq 0$, respectively. The initial state $x(t_1)$ is assumed to be a zero-mean Gaussian random vector with covariance matrix $R_0 \in \mathbb{R}^{n\times n}$, $R_0 \geq 0$. The processes $v(t_k)$, $k \in \mathbb{N}$ and $\eta(t_k)$, $k \in \mathbb{N}$ and the initial condition $x(t_1)$ are mutually independent. At the time-instants $t_k, k \in \mathbb{N}$, the output of the process $y(t_k)$ is sampled and transmitted over a communication channel. The received output $\bar{y}(t_k)$ is used to compute control actions $u(t_k)$ which are sent back to the process, see Fig. 1. The complete control-loop is assumed to be done instantaneously, i.e., the sampling, transmission, and arrival time-instants are supposed to be equal. In this paper, we focus on attacks on sensor measurements. That is, in between transmission and reception of sensor data, an attacker may replace the signals coming from the sensors to the controller, see Fig. 1. After each transmission and reception, the attacked output $\bar{y}$ takes the form:

$$\bar{y}(t_k) := y(t_k) + \delta(t_k) = Cx(t_k) + \eta(t_k) + \delta(t_k), \tag{2}$$

where $\delta(t_k) \in \mathbb{R}^m$ denotes additive sensor attacks. Define $x_k := x(t_k)$, $u_k := u(t_k)$, $v_k := v(t_k)$, $\bar{y}_k := \bar{y}(t_k)$, $\eta_k := \eta(t_k)$, and $\delta_k := \delta(t_k)$. Using this notation, the attacked system is written as follows

$$\begin{cases} x_{k+1} = Fx_k + Gu_k + v_k, \\ \bar{y}_k = Cx_k + \eta_k + \delta_k. \end{cases} \tag{3}$$

A. Kalman Filter

In order to estimate the state of the process, an estimator with the following structure is proposed:

$$\hat{x}_{k+1} = F\hat{x}_k + Gu_k + L_k\big(\bar{y}_k - C\hat{x}_k\big), \tag{4}$$

with estimated state $\hat{x}_k \in \mathbb{R}^n$, $\hat{x}_1 = E[x(t_1)]$, where $E[\cdot]$ denotes expectation, and gain matrix $L_k \in \mathbb{R}^{n\times m}$. Define the estimation error $e_k := x_k - \hat{x}_k$. In the Kalman filter, the matrix L_k is designed to minimize the covariance matrix $P_k := E[e_k e_k^T]$ (in the absence of attacks). If the pair (F, C) is detectable, the covariance matrix converges to steady state in the sense that $\lim_{k\to\infty} P_k = P$ exists [10]. We assume that the system has reached steady state before an attack occurs. Then, the estimation of the random sequence $x_k, k \in \mathbb{N}$ can be obtained by the estimator (4) with P_k and L_k in steady state. It can be verified that, if $R_2 + CPC^T$ is positive definite, the following estimator gain

$$L_k = L := \big(FPC^T\big)\big(R_2 + CPC^T\big)^{-1}, \tag{5}$$

leads to the minimal steady state covariance matrix P, with P given by the solution of the algebraic Riccati equation:

$$FPF^T - P + R_1 = FPC^T(R_2 + CPC^T)^{-1}CPF^T. \tag{6}$$

The reconstruction method given by (4)-(6) is referred to as the steady state Kalman Filter, cf. [10].

B. Residuals and Hypothesis Testing

The main idea behind fault detection theory is the use of an estimator to forecast the evolution of the system. If the difference between what it is measured and the estimation is larger than expected, there may be a fault in or attack on the system. In this paper, we use the steady state Kalman filter introduced in the previous section as our estimator. Define the residual sequence $r_k, k \in \mathbb{N}$ as

$$r_k := \bar{y}_k - C\hat{x}_k = Ce_k + \eta_k + \delta_k, \tag{7}$$

which evolves according to the following difference equation

$$\begin{cases} e_{k+1} = \big(F - LC\big)e_k - L\eta_k - L\delta_k + v_k, \\ r_k = Ce_k + \eta_k + \delta_k. \end{cases} \tag{8}$$

If there are no attacks, the mean of the residual is

$$E[r_{k+1}] = CE[e_{k+1}] + E[\eta_{k+1}] = \mathbf{0}_{m\times 1}, \tag{9}$$

and the covariance is

$$\Sigma := E[r_{k+1}r_{k+1}^T] = CPC^T + R_2. \tag{10}$$

It is assumed that $\Sigma \in \mathbb{R}^{m\times m}$ is positive definite (a standard assumption that guarantees that the Kalman filter converges). For this residual, we identify two hypothesis to be tested: $\mathcal{H}_0$ the *normal mode* (no attacks) and $\mathcal{H}_1$ the *faulty mode* (with attacks). Then, we have

$$\mathcal{H}_0 : \begin{cases} E[r_k] = \mathbf{0}_{m\times 1}, \\ E[r_k r_k^T] = \Sigma, \end{cases} \qquad \mathcal{H}_1 : \begin{cases} E[r_k] \neq \mathbf{0}_{m\times 1}, \\ E[r_k r_k^T] \neq \Sigma, \end{cases}$$

where $\mathbf{0}_{m\times 1}$ denotes an m-dimensional vector composed of only zeros. In this manuscript, we mainly focus on the CUSUM procedure [11] for examining the residual and subsequently detecting attacks. However, for comparison, we also present results about the so-called *chi-squared* change detection procedure (which is widely used in the literature).

C. Distance Measures and Cumulative Sum (CUSUM)

The input to the CUSUM procedure is a *distance measure*, i.e., a measure of how deviated the estimator is from the actual system. In this work, we assume there is a vector-valued detection scheme with an aggregate detector, i.e., there is a single detector monitoring the complete vector of residuals. We propose the following quadratic form as distance measure:

$$z_k := r_k^T \Sigma^{-1} r_k. \tag{11}$$

Note that, if there are no attacks, $r_k \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Hence, $\delta_k = \mathbf{0}$ implies that z_k follows a χ^2-*distribution* with m degrees of freedom [12] (note that m is the dimension of r_k) and

$$E[z_k] = m \text{ and } \text{var}[z_k] = 2m. \tag{12}$$

For a given *distance measure* $z_k \in \mathbb{R}$, the CUSUM of Page [11] is written as:

CUSUM: $S_1 = 0$,

$$\begin{aligned} &S_k = \max(0, S_{k-1} + z_k - b), && \text{if } S_{k-1} \leq \tau, \\ &S_k = 0 \ \text{ and } \ \tilde{k} = k-1, && \text{if } S_{k-1} > \tau. \end{aligned} \tag{13}$$

Design parameters: bias $b \in \mathbb{R}_{>0}$ and threshold $\tau \in \mathbb{R}_{>0}$.
Output: alarm time(s) $\tilde{k}$.

The idea is that the test sequence S_k accumulates the distance measure z_k and alarms are triggered when S_k exceeds the threshold τ. The test is reset to zero each time S_k becomes negative or larger than τ. If z_k is an independent nonnegative sequence (which is our case) and b is not sufficiently large, the CUSUM sequence S_k grows unbounded until the threshold τ is reached, no matter how large τ is set. In order to prevent these drifts, inevitably yielding false alarms, the bias b must be selected properly based on the properties of the distance measure. Once the the bias is chosen, the threshold τ must be selected to fulfill a required *false alarm rate* $\mathcal{A}^*$ (see Section III-B).

III. CUSUM-Tuning

We have already mentioned that too small a bias can lead to unbounded growth of the CUSUM test sequence. At the same time, too large a bias may hide the effect of the attacker and provide more opportunity for sensor attacks to influence the system while still remaining undetected. In what follows, we provide tools for selecting these parameters given the statistical properties of the distance measure z_k introduced in (12). First, we provide sufficient conditions on the bias b such that, in the absence of attacks, the sequence S_k of the CUSUM remains bounded (independent of the reset due to τ) in mean-squared sense. This is important because it avoids false alarms due to the divergence of S_k. Secondly, we characterize the false alarm rate of the CUSUM in terms of b and τ given a *desired* false alarm rate.

A. Boundedness

First, we introduce the following concept of boundedness of stochastic processes, cf. [13], followed by sufficient conditions for boundedness of the CUSUM sequence. Let $E_b[a]$ denote the conditional expectation of a given b.

Definition 1 *The sequence S_k, $k \in \mathbb{N}$ is said to be bounded in mean square, if*

$$\sup_{k \in \mathbb{N}} E_{S_1}\left[S_k^2\right] < \infty,$$

is satisfied, i.e., the second moment of S_k is finite.

Theorem 1 *Consider the discrete-time process* (3) *and the steady state Kalman filter* (4)-(6)*. Assume that there are no attacks to the system, i.e., $\delta_k = \mathbf{0}$. Let the* CUSUM (13) *with bias $b \in \mathbb{R}_{>0}$ and threshold $\tau \in \mathbb{R}_{>0}$ be driven by the distance measure $z_k = r_k^T \Sigma^{-1} r_k$, $k \in \mathbb{N}$ with residual sequence $r_k \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Then, for $b > \bar{b} := m$, the* CUSUM *sequence S_k, $k \in \mathbb{N}$ is bounded in mean square sense independent of the threshold τ.*

The proof is omitted here due to the page limit. The result stated in Theorem 1 implies that for $b > \bar{b}$, the second moment (and hence the first) of the sequence S_k, $k \in \mathbb{N}$ does not diverge. Consequently, we avoid false alarms due to intrinsic unboundedness of the CUSUM sequence. Note that if the bias b is selected greater than but close to $\bar{b}$, small changes in the distance measure z_k would lead to divergence of S_k. Therefore, the smaller the bias, the higher the sensitivity against changes in (or uncertain characterization of) the residual signals.

B. False Alarms

Once the bias is selected such that boundedness of $E[S_k^2]$ is guaranteed, the next step is to select the threshold τ to fulfill a desired false alarm rate. The occurrence of an alarm in the CUSUM when there are no attacks to the CPS is referred to as a false alarm. Let $\mathcal{A} \in (0, 1)$ denote the *false alarm rate* for the CUSUM procedure defined as the expected proportion of observations which are false alarms [14],[15]. Define the *run length* $\mathcal{K}$ of the CUSUM (13) as the number of iterations needed such that $S_{\mathcal{K}} > \tau$ (with no attacks), i.e.,

$$\mathcal{K} := \inf\{k \geq 1 : S_k > \tau\}. \tag{14}$$

The expected value of $\mathcal{K}$ is known in the literature as the *Average Run Length* (ARL). The ARL is related to $\mathcal{A}$ by

$$\mathcal{A} = 1/\text{ARL}, \tag{15}$$

see [14],[15] for details. Then, for a given $b > \bar{b}$, the problem of selecting τ to satisfy a desired false alarm rate $\mathcal{A}^*$ can be reformulated as the problem of selecting τ such that

$$\text{ARL} = 1/\mathcal{A}^*. \tag{16}$$

To determine a pair (b, τ) satisfying (16), an expression for the ARL $= E[\mathcal{K}]$ is required but, in general, its exact evaluation is analytically intractable [16]. The problem of approximating the ARL for CUSUM procedures has been addressed by many authors during the last decades [17]-[19]. Particularly, accurate numerical methods have been proposed by, for instance, [18]-[19]. These methods rely on two main techniques, namely Markov chain and integral equation approaches. Both methods give accurate predictions of the ARL (see [18] for a comparison); however, we find

the Markov chain approach more constructive and easier to implement. In particular, in this manuscript, we use the result of Evans and Brook [20]. With this result, we outline a procedure for selecting the threshold τ given the bias b and a required false alarm rate $\mathcal{A}^*$.

For $b > \bar{b}$ and $\tau \in \mathbb{R}_{>0}$, consider the sequence S_k generated by the CUSUM procedure (13) driven by the distance measure $z_k = r_k^T \Sigma^{-1} r_k$, $r_k \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Given the recursive nature of the CUSUM procedure and independence of r_k the sequence S_k forms a Markov chain taking values in $\mathbb{R}_{\geq 0}$ [21]. By discretizing the probability distribution of the distance measure, it is possible to subdivide the CUSUM sequence S_k into a finite set of partitions. The idea is to approximate the continuous scheme by a Markov chain having $N+1$ states labeled as $\{E_0, E_1, \ldots, E_N\}$, where E_N is absorbing. Then, the probability that the chain remains in the same state at the next step should correspond to the case when S_k does not change in value by more than a small amount, say $\frac{1}{2}\Delta_S$, i.e., the next distance measure z_k does not differ from the bias b by more than $\frac{1}{2}\Delta_S$. The constant Δ_S determines the width of the grouping interval involved in the discretization of the probability distribution of z_k. The interval width $\frac{1}{2}\Delta_S$ must be selected such that the probability of jumping from E_j, $j \in \{0, \ldots, N-1\}$ to the absorbing state E_N is approximately equal to the probability that the CUSUM sequence S_k jumps beyond the threshold τ from a position $S_{k-1} \in [0, \tau)$ which corresponds approximately to the state E_j. This requirement is satisfied by taking

$$\Delta_S := \frac{2\tau}{2N-1}, \tag{17}$$

see [20] for details. Then, the transition probabilities from a starting state E_j, $j = 0, \ldots, N-1$, can be determined from the probability distribution of $z_k - b = r_k^T \Sigma^{-1} r_k - b$, as follows:

$$\begin{aligned}
\text{pr}(E_j \to E_0) &= \text{pr}(z_k - b \leq -j\Delta_S + \frac{1}{2}\Delta_S), \\
\text{pr}(E_j \to E_N) &= \text{pr}((N-j)\Delta_S - \frac{1}{2}\Delta_S < z_k - b), \\
\text{pr}(E_j \to E_\nu) &= \text{pr}(z_k - b \leq (\nu - j)\Delta_S + \frac{1}{2}\Delta_S) \\
&\quad - \text{pr}(z_k - b < (\nu - j)\Delta_S - \frac{1}{2}\Delta_S).
\end{aligned}$$

Note that $\text{pr}(E_0 \to E_N) = \text{pr}(z_k - b > \tau)$. The system forms a Markov chain whose transition matrix can be constructed from the probability distribution of z_k, given b and τ. Define $T_\chi := \text{pr}(z_k - b \leq \chi\Delta_S + \frac{1}{2}\Delta_S)$ and $p_\chi := \text{pr}(\chi\Delta_S - \frac{1}{2}\Delta_S < z_k - b \leq \chi\Delta_S + \frac{1}{2}\Delta_S)$. Then, the Markov transition matrix $\mathcal{P} \in \mathbb{R}^{(N+1)\times(N+1)}$ is:

$$\mathcal{P} := \begin{pmatrix}
T_0 & p_1 & p_2 & \ldots & p_{N-1} & 1 - T_{N-1} \\
T_{-1} & p_0 & p_1 & \ldots & p_{N-2} & 1 - T_{N-2} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
T_{-j} & p_{1-j} & p_{2-j} & \ldots & p_{N-1-j} & 1 - T_{N-1-j} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
T_{1-N} & p_{2-N} & p_{3-N} & \ldots & p_0 & 1 - T_0 \\
0 & 0 & 0 & \ldots & 0 & 1
\end{pmatrix}. \tag{18}$$

To compute the transition probabilities of $\mathcal{P}$, we need the Cumulative Distribution Function (CDF) of the shifted distance measure $z_k - b = r_k^T \Sigma^{-1} r_k - b$. If there are no attacks, $r_k \sim \mathcal{N}(\mathbf{0}, \Sigma)$; therefore $z_k - b$ follows a shifted χ^2-distribution with CDF given by

$$F_{z_k - b}(x) := \begin{cases} \text{P}\left(\frac{m}{2}, \frac{x+b}{2}\right), & \text{for } x \geq -b, \\ 0, & \text{for } x < -b, \end{cases} \tag{19}$$

where $\text{P}(\cdot, \cdot)$ denotes the regularized lower incomplete gamma function. Then, the entries of $\mathcal{P}$ are given by:

$$\begin{cases} p_\chi = F_{z_k - b}(\chi\Delta_S + \frac{\Delta_S}{2}) - F_{z_k - b}(\chi\Delta_S - \frac{\Delta_S}{2}), \\ T_\chi = F_{z_k - b}\left(\chi\Delta_S + \frac{\Delta_S}{2}\right). \end{cases} \tag{20}$$

Next, having defined the transition matrix $\mathcal{P}$ of the approximated Markov chain, we can compute an approximation $\tilde{\mathcal{A}}$ of the false alarm rate $\mathcal{A}$ based on the result in [20], equation (15), and (17)-(20). Let $\mathbf{1}_{N\times 1}$ denote an N-dimensional vector composed of only ones and $\text{col}(\mu_1, \ldots, \mu_N)$ stand for the column vector composed of the elements $\mu_1, ..., \mu_N$.

Theorem 2 *Assume that there are no attacks to the system and let the* CUSUM (13) *with bias $b > \bar{b} = m$ and threshold $\tau \in \mathbb{R}_{>0}$ be driven by the distance measure $z_k = r_k^T \Sigma^{-1} r_k$, $k \in \mathbb{N}$ with residual sequence $r_k \sim \mathcal{N}(\mathbf{0}, \Sigma)$. For a finite number of partitions $N \in \mathbb{N}$, define $\mathcal{R} \in \mathbb{R}^{N\times N}$ as the matrix obtained from the transition matrix $\mathcal{P} \in \mathbb{R}^{(N+1)\times(N+1)}$* (17)-(20) *by removing its last row and column and*

$$\mu := (I_N - \mathcal{R})^{-1} \mathbf{1}_{N\times 1} = \text{col}(\mu_1, \ldots, \mu_N). \tag{21}$$

Then, the false alarm rate $\mathcal{A} = 1/\text{ARL}$ is approximately given by $\tilde{\mathcal{A}} := \mu_1^{-1}$. Moreover, as $N \to \infty$, $\tilde{\mathcal{A}} \to \mathcal{A}$, i.e., $\lim_{N\to\infty} \tilde{\mathcal{A}} = \mathcal{A}$.

The proof is omitted here due to the page limit.

Remark 1 *By construction, the entries of $\mathcal{R}$ are non-negative and its row sums are less than one. Therefore, by Gershgorin circle theorem, $\rho(\mathcal{R}) < 1$, where $\rho(\cdot)$ denotes spectral radius; therefore, the matrix $(I_N - \mathcal{R})$ is invertible.*

Remark 2 *Theorem 2 provides a tool for approximating the false alarm rate $\mathcal{A}$ of the* CUSUM *procedure for given bias b and threshold τ. In particular, for a given $b > \bar{b}$, it provides a map $\mathcal{S} : \mathbb{R}_{>0} \to (0, 1)$ from the threshold τ to the approximated false alarm rate $\tilde{\mathcal{A}}$, i.e., $\tau \mapsto \mathcal{S}(\tau)$, $\tilde{\mathcal{A}} = \mathcal{S}(\tau)$. Given that $F_{z_k - b}(z)$ is a continuous function for all $z \in \mathbb{R}$, it can be proved that $\mathcal{S}$ is a continuous map for all $\tau \in \mathbb{R}_{>0}$. Then, using Theorem 2, simple bisection methods can be employed to determine the threshold $\tau = \tau^* \in \mathbb{R}_{>0}$ required to satisfy a desired false alarm rate $\mathcal{A}^*$ for given $b > \bar{b}$.*

IV. Chi-squared Tuning

In this work, we use the chi-squared approach as a benchmark to compare the performance of the CUSUM. Consider again the residual r_k and its covariance matrix Σ.

Chi-squared procedure:

$$\text{If } \; z_k = r_k^T \Sigma^{-1} r_k > \alpha, \;\; \tilde{k} = k. \tag{22}$$

Design parameter: threshold $\alpha \in \mathbb{R}_{>0}$.
Output: alarm time(s) $\tilde{k}$.

The idea is that alarms are triggered if z_k exceeds the threshold α. Similar to the CUSUM procedure, the parameter α is selected to satisfy a required false alarm rate $\mathcal{A}^*$.

Theorem 3 *Assume that there are no attacks to the system and consider the chi-squared procedure* (22) *with threshold* $\alpha \in \mathbb{R}_{>0}$*,* $r_k \sim \mathcal{N}(\mathbf{0}, \Sigma)$*. Let* $\alpha = \alpha^* := 2\mathrm{P}^{-1}(\frac{m}{2}, 1 - \mathcal{A}^*)$*, where* $\mathrm{P}^{-1}(\cdot,\cdot)$ *denotes the inverse regularized lower incomplete gamma function, then* $\mathcal{A} = \mathcal{A}^*$*.*

V. State Degradation

In this section, we assess the performance of the CUSUM procedure by quantifying the effect of the attack sequence δ_k on the state of the system when the CUSUM is used to detect anomalies. In particular, we characterize, for a class of *stealthy attacks*, the largest deviation of the expectation of the state due to the attack sequence. More precisely, we derive upper bounds on the expected value of the state given the system dynamics, the control strategy, the attack sequence, and the parameters of the CUSUM. Also, for the same class of attacks, we quantify the largest deviation of the expectation of the state when using the chi-squared procedure and then compare it with the one obtained with the CUSUM.

A. Feedback Controller

Consider dynamic output feedback controllers of the form:

$$u_k := K\hat{x}_k, \tag{23}$$

where $\hat{x}_k \in \mathbb{R}^n$ is the state of the Kalman filter (4)-(6) and $K \in \mathbb{R}^{l\times n}$ denotes the control matrix. Assume that the pair (F, G) is stabilizable, then the matrix K can be selected such that $(F + GK)$ is Schur. The closed-loop system (3), (4)-(6),(23) can be written in terms of the estimation error $e_k = x_k - \hat{x}_k$ as follows

$$\begin{cases} x_{k+1} = (F + GK)x_k - GKe_k + v_k, \\ e_{k+1} = \big(F - LC\big)e_k - L\delta_k - L\eta_k + v_k. \end{cases} \tag{24}$$

Note that the attack sequence δ_k directly affects the estimation error dynamics, whereas the effect of the attack on the system dynamics is through the interconnection term GKe_k.

B. Stealthy Attacks

Here, we quantify the damage that the attacker may induce to the system in the *worst-case scenario* while remaining undetected by the detection procedure. To this end, it is assumed that the attacker has perfect knowledge of the system dynamics, the Kalman Filter, control inputs, measurements, and detection procedure (either CUSUM or chi-squared). It is further assumed that all the sensors can be compromised by the attacker at each time step. In particular, we are interested in attack sequences δ_k that can induce changes in the system dynamics while remaining undetected by the detection procedure. This class of attacks is known in the literature as *stealthy attacks* [7], [2], [1]. First, consider the chi-squared procedure (22) and write z_k in terms of the estimation error e_k, then:

$$z_k = (Ce_k + \eta_k + \delta_k)^T\Sigma^{-1}(Ce_k + \eta_k + \delta_k). \tag{25}$$

By assumption, the attacker has access to $y_k = Cy_k + \eta_k$. Moreover, given its perfect knowledge of the Kalman filter, the adversary can compute the estimated output $C\hat{x}_k$ and then construct $Ce_k + \eta_k$. For a given chi-squared threshold α, define $\bar{\alpha} := \{\bar{\alpha} \in \mathbb{R}^m | \bar{\alpha}^T\bar{\alpha} = \alpha\}$; take, for instance, $\bar{\alpha} = \mathrm{col}(\sqrt{\frac{\alpha}{m}}, \sqrt{\frac{\alpha}{m}}, \ldots, \sqrt{\frac{\alpha}{m}})$ or $\bar{\alpha} = \mathrm{col}(\sqrt{\alpha}, 0, \ldots, 0)$. Then, it follows that

$$\delta_k = -Ce_k - \eta_k + \Sigma^{\frac{1}{2}}\bar{\alpha} \rightarrow z_k = \alpha, \tag{26}$$

where $\Sigma^{\frac{1}{2}}$ denotes the symmetric squared root matrix of Σ, is a feasible attack sequence given the capabilities of the attacker. These attacks maximize the damage to the CPS by immediately saturating and maintaining z_k at the threshold α. The expectation of the closed-loop system under the attack (26) is given by

$$\begin{cases} E[x_{k+1}] = (F + GK)E[x_k] - GKE[e_k], \\ E[e_{k+1}] = FE[e_k] - L\Sigma^{\frac{1}{2}}\bar{\alpha}. \end{cases} \tag{27}$$

Note that if $\rho[F] > 1$, then $|E[e_k]|$, and also $|E[x_k]|$ due to the interconnection, diverge to infinity as $k \rightarrow \infty$ [10]. That is, the attack sequence (26) destabilizes the system if $\rho[F] > 1$. If $\rho[F] \leq 1$, then $|E[e_k]|$, may or may not diverge to infinity depending on algebraic and geometric multiplicities of the eigenvalues with unit modulus of F (a known fact from stability of LTI systems [10]).

Proposition 1 *Consider the process* (3)*, the Kalman filter* (4)-(6)*, the controller* (23)*, and the chi-squared procedure* (22)*. Let the sensors be attacked by the stealthy attack sequence* (26)*. Then, if* $\rho[F] < 1$*, it is satisfied that* $\lim_{k\rightarrow\infty} |E[x_k]| = \gamma_{\chi^2}$*, where*

$$\gamma_{\chi^2} := \left\| (I - F - GK)^{-1}GK(I - F)^{-1}L\Sigma^{\frac{1}{2}}\bar{\alpha} \right\|.$$

Next, consider the CUSUM procedure and write (13) in terms of the estimation error e_k:

$$S_k = \max\left(0, S_{k-1} + \left(\Sigma^{-\frac{1}{2}}(Ce_k + \eta_k + \delta_k)\right)^2 - b\right), \tag{28}$$

if $S_{k-1} \leq \tau$; and $S_k = 0$, if $S_{k-1} > \tau$. As with the chi-squared procedure, we look for attack sequences that immediately saturate and then maintain the CUSUM statistic at the threshold $S_k = \tau$. Assume that the attack starts at some $k = k^* \geq 2$ and $S_{k^*-1} \leq \tau$, i.e., the attack does not start immediately after a false alarm. For given threshold τ and bias b, define $\bar{\tau} := \{\bar{\tau} \in \mathbb{R}^m | \bar{\tau}^T\bar{\tau} = \tau + b - S_{k-1}\}$ and $\tilde{b} := \{\tilde{b} \in \mathbb{R}^m | \tilde{b}^T\tilde{b} = b\}$. Consider the attack sequence:

$$\delta_k = \begin{cases} -Ce_k - \eta_k + \Sigma^{\frac{1}{2}}\bar{\tau}, & k = k^*, \\ -Ce_k - \eta_k + \Sigma^{\frac{1}{2}}\tilde{b}, & k > k^*. \end{cases} \tag{29}$$

Note that the attacker can only induce this sequence by knowing S_{k^*-1} exactly, i.e., the value of the CUSUM sequence one step before the attack. This is a strong assumption since it represents a real-time quantity that is not communicated over the communication channel. Even if the opponent has access to the parameters of the CUSUM, (b, τ), given the stochastic nature of the residuals, the attacker would need to know the complete history of observations (from when the CUSUM was started) to be able to reconstruct S_{k^*-1} from data. This is an inherent security advantage in favor of the CUSUM over static detectors like the bad-data or chi-squared. Nevertheless, for evaluating the worst case scenario, we assume that the attacker has access to S_{k^*-1}.

Without loss of generality, assume $k^* = 2$. By construction, $E[x_i] = E[e_i] = \mathbf{0}$, $i = 1, 2$; then, the expectation of closed-loop system under the attack sequence (29) is written as: $E[x_3] = \mathbf{0}$, $E[e_3] = -L\Sigma^{\frac{1}{2}}\bar{\tau}$, and

$$\begin{cases} E[x_{k+1}] = (F + GK)E[x_k] - GKE[e_k], \\ E[e_{k+1}] = FE[e_k] - L\Sigma^{\frac{1}{2}}\tilde{b}, \end{cases} \tag{30}$$

for $k > k^* = 2$.

Proposition 2 *Consider the process* (3)*, the Kalman filter* (4)-(6)*, the controller* (23)*, and the CUSUM procedure* (13)*. Let the sensors be attacked by the stealthy attack sequence* (29)*. Then, if* $\rho[F] < 1$*, it is satisfied that* $\lim_{k\to\infty} |E[x_k]| = \gamma_{\text{CS}}$*, where*

$$\gamma_{\text{CS}} := \left\| (I - F - GK)^{-1} GK (I - F)^{-1} L\Sigma^{\frac{1}{2}}\tilde{b} \right\|.$$

The proofs of both Proposition 1 and Proposition 2 are omitted here due to the page limit.

C. Detector Comparison

For stealthy attacks, we have derived upper bounds on the steady state value of $|E[x_k]|$ for both the chi-squared and CUSUM procedures provided that $\rho(F) < 1$. There are some aspects of these bounds to be highlighted. Note that the τ-dependent term in the attack sequence (29) does not affect $|E[x_k]|$ in steady state. This is because $\Sigma^{\frac{1}{2}}\bar{\tau}$ is only induced at $k = k^*$; it follows that, since $\rho(F) < 1$, the contribution of $\bar{\tau}$ to $E[x_k]$ exponentially decreases to zero as $k \to \infty$.

For comparison, let $\tilde{b} \in \text{Im}[\bar{\alpha}]$, i.e., $\tilde{b} = c\bar{\alpha}$ for some $c \in \mathbb{R}$. Then, $\gamma_{\text{CS}} = |c|\gamma_{\chi^2}$; therefore, $\gamma_{\text{CS}} < \gamma_{\chi^2}$, if and only if $|c| < 1$. Moreover, by construction, $\tilde{b} = c\bar{\alpha} \to \tilde{b}^T\tilde{b} = b = c^2\bar{\alpha}^T\bar{\alpha} = c^2\alpha$; hence, $c = \pm\sqrt{b/\alpha}$ and $|c| < 1 \leftrightarrow b < \alpha$. That is, if $b < \alpha$, under the same class of stealthy attacks, the CUSUM procedure leads to smaller steady state deviations of $|E[x_k]|$ than the chi-squared procedure.

In general, to increase the chances of attack detection, it is desired to select b as close as possible to $\bar{b}$ in Theorem 1. It follows that $b \approx \bar{b} = m$. On the other hand, according to Theorem 3, α must be selected as $\alpha = \alpha^* = 2\text{P}^{-1}(\frac{m}{2}, 1 - \mathcal{A}^*)$ to fulfill a desired false alarm rate $\mathcal{A}^*$. In this case, we want to select $\mathcal{A}^*$ close to zero, such that there are only a few false alarms. Let $\mathcal{A}^* \in [0.01, 0.1]$ and $m = 2$, i.e., false alarms between 1% and 10% and two dimensional outputs. Then, $\alpha = 2\text{P}^{-1}(\frac{m}{2}, 1 - \mathcal{A}^*) \in [4.60, 9.21]$ and, for $b \approx \bar{b} = 2$, $0.45 \lessapprox |c| \lessapprox 0.65$. This implies that for the same class of attacks and $\mathcal{A}^* \in [0.01, 0.1]$, the chi-squared procedure leads to around two times larger upper bounds than the CUSUM. Actually, for having $\alpha = b \to |c| = 1$, it is necessary to allow for a rate of $\mathcal{A}^* = 0.63$, which is high for practical purposes. For the CUSUM procedure, the threshold τ is selected to fulfill the desired $\mathcal{A}^*$. Given that there are no exact closed-form expressions to relate τ and $\mathcal{A}^*$ (we only have a numeric approximation in Theorem 2), it is not possible to tell exactly how large τ needs be to satisfy $\mathcal{A}^*$. However, as already mentioned, the contribution of $\bar{\tau}$ to $E[x_k]$ vanishes exponentially, i.e., independent of how large τ is, its contribution to the upper bound of $|E[x_k]|$ is zero in steady state.

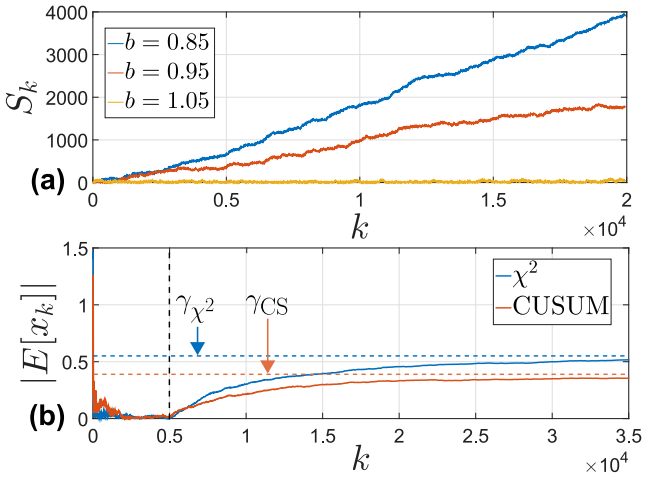

Fig. 2. (a) Cusum evolution for different values of bias b. (b) Degradation of $|E[x_k]|$ due to stealthy attacks. Attacks start at $k = 5000$.

VI. Simulation Experiments

Consider the closed-loop system (3),(4)-(6),(23) with matrices as given in (31). First, assume no-attacks, i.e., $\delta_k = \mathbf{0}$, and consider the CUSUM procedure (13) with distance measure $z_k = r_k^T\Sigma^{-1}r_k$ and residual sequence (8). According to Theorem 1, the bias b must be selected larger than $\bar{b} = m = 2$ to ensure mean square boundedness of S_k independent of the threshold τ. Figure 2a depicts the evolution of the CUSUM for $b \in \{0.85\bar{b}, 0.95\bar{b}, 1.05\bar{b}\}$ and $k \in [1, 20000]$. For the purpose of illustrating this unbounded growth, we have omitted the reset procedure of the CUSUM. Note that the bound for b is tight, small deviations from $\bar{b}$ lead to (boundedness) unboundedness of S_k.

Next, for desired false alarm rates $\mathcal{A}^* \in \{0.25, 0.10, 0.02\}$, we compute the corresponding thresholds $\tau = \tau^*$ using Theorem 2 and Remark 2. For these thresholds, in Table 1, we present the actual false alarm rate $\mathcal{A}$ (obtained by simulation) and the desired $\mathcal{A}^*$. Note that the difference between $\mathcal{A}$ and $\mathcal{A}^*$ is less that 6% in all cases. It actually seems that our thresholds lead to $\mathcal{A} < \mathcal{A}^*$; however, further analysis is required to corroborate this relation.

Finally, in Fig. 2b, we present the evolution of $|E[x_k]|$ when both the chi-squared and the CUSUM are deployed for attack detection and the attack sequences δ_k are given by the stealthy attacks introduced in (26) and (29), respectively, with $\bar{\alpha} = \sqrt{\alpha}\bar{\delta}$, $\tilde{b} = \sqrt{b}\bar{\delta}$, $\bar{\tau} = \sqrt{\tau + b - S_{k-1}}\bar{\delta}$, and $\bar{\delta} = \text{col}(1, 0, \ldots, 0) \in \mathbb{R}^m$. For the CUSUM, we select $b = 1.15\bar{b} = 2.30$ and $\tau = \tau^* = 2.7468$ such that $\mathcal{A} \approx \mathcal{A}^* = 0.10$ (see Table 1). Likewise, we select $\alpha = \alpha^* = 2\text{P}^{-1}(\frac{2}{2}, 1 - 0.10) = 4.6051$ such that, according to Theorem 3, $\mathcal{A} \approx \mathcal{A}^* = 0.10$. The attack is induced at $k = 5 \times 10^3$. Note that, as stated by Proposition 1 and Proposition 2, given that $\rho(F) < 1$, $\lim_{k\to\infty} |E[x_k]| = \gamma_{CS} = 0.4026$ and $\lim_{k\to\infty} |E[x_k]| = \gamma_{\chi^2} = 0.5697$. Moreover, as mentioned in V-C, we expect that the CUSUM leads to a smaller deviation on $|E[x_k]|$ because $b < \alpha$ and $\tilde{b} \in \text{Im}[\bar{\alpha}]$. This is what we see in Fig. 2.

VII. Conclusion

In this paper, for a class of discrete-time stochastic linear systems, we have characterized a model-based CUSUM procedure for identifying compromised sensors. In particular, Kalman filters have been proposed to estimate the state of the physical process; then, these estimates have been used to construct residual variables (between sensor measurements and estimations) which drive the CUSUM procedure. Using stability results for stochastic systems and Markov chain approximations of the CUSUM sequence, we derived systematic tools for tuning the CUSUM procedure such that mean square boundedness of the CUSUM sequence is guaranteed and the desired false alarm rate is fulfilled. For a class of stealthy attacks, we have characterized the performance of the proposed CUSUM procedure in terms of the effect that the attack sequence can induce on the system dynamics. Then, we have compared this performance against the one obtained using chi-squared procedure.

$$\begin{cases} F = \begin{pmatrix} 0.84 & 0.23 \\ -0.47 & 0.12 \end{pmatrix}, G = \begin{pmatrix} 0.07 \\ 0.23 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, K = \begin{pmatrix} -1.85 & -0.96 \end{pmatrix}, L = \begin{pmatrix} 0.25 & 0.17 \\ -0.18 & -0.07 \end{pmatrix}, \\ R_1 = \begin{pmatrix} 0.45 & -0.11 \\ -0.11 & 0.20 \end{pmatrix}, R_0 = R_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 1.65 & 0.40 \\ 0.40 & 1.46 \end{pmatrix}. \end{cases} \tag{31}$$

TABLE I

Simulation Experiments. Results of Theorem 2 and Remark 2.

	$\mathcal{A}^* = 0.25$		$\mathcal{A}^* = 0.10$		$\mathcal{A}^* = 0.02$	
$b/\bar{b}$	τ^*	$\mathcal{A}$ (Sim.)	τ^*	$\mathcal{A}$ (Sim.)	τ^*	$\mathcal{A}$ (Sim.)
1.05	0.71	0.20	3.12	0.09	10.18	0.02
1.15	0.49	0.20	2.74	0.09	8.62	0.02
2.00	–	–	0.61	0.09	4.18	0.02

References

[1] F. Pasqualetti, F. Drfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, 2013.

[2] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2010, pp. 5967–5972.

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

[4] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding sensor outputs for injection attacks detection," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, 2014, pp. 5776–5781.

[5] C.-Z. Bai and V. Gupta, "On kalman filtering in the presence of a compromised sensor: Fundamental performance bounds," in *American Control Conference (ACC), 2014*, 2014, pp. 3029–3034.

[6] C. Z. Bai, F. Pasqualetti, and V. Gupta, "Security in stochastic control systems: Fundamental limitations and performance bounds," in *American Control Conference (ACC), 2015*, 2015, pp. 195–200.

[7] A. Cárdenas, S. Amin, Z. Lin, Y. Huang, C. Huang, and 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.

[8] F. Gustafsson, *Adaptive Filtering and Change Detection*. West Sussex, Chichester, England: John Wiley and Sons, LTD, 2000.

[9] C. Murguia and J. Ruths, "Characterization of a cusum model-based sensor attack detector," in *Decision and Control (CDC), 2016 55th IEEE Conference on*, 2016, (Submitted).

[10] K. J. Aström and B. Wittenmark, *Computer-controlled Systems (3rd Ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.

[11] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[12] M. Ross, *Introduction to Probability Models, Ninth Edition*. Orlando, FL, USA: Academic Press, Inc., 2006.

[13] T. Tarn and R. Yona, "Observers for nonlinear stochastic systems," *Automatic Control, IEEE Transactions on*, vol. 21, pp. 441–448, 1976.

[14] B. Adams, W. Woodall, and C. Lowry, "The use (and misuse) of false alarm probabilities in control chart design," *Frontiers in Statistical Quality Control 4*, pp. 155–168, 1992.

[15] C. van Dobben de Bruyn, *Cumulative sum tests : theory and practice*. London : Griffin, 1968.

[16] R. A. Khan, "Wald's approximations to the average run length in cusum procedures," *Journal of Statistical Planning and Inference*, vol. 2, pp. 63 – 77, 1978.

[17] C. Park, "A corrected wiener process approximation for cusum arls," *Sequential Analysis*, vol. 6, pp. 257–265, 1987.

[18] C. Champ and S. Rigdon, "A comparison of the markov chain and the integral equation approaches for evaluating the run length distribution of quality control charts," *Communications in Statistics - Simulation and Computation*, vol. 20, pp. 191–204, 1991.

[19] W. Woodall, "The distribution of the run length of one-sided cusum procedures for continuous random variables," *Technometrics*, vol. 25, pp. 295–301, 1983.

[20] D. A. E. D. Brook, "An approach to the probability distribution of cusum run length," *Biometrika*, vol. 59, no. 3, pp. 539–549, 1972.

[21] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.