The Performance and Limitations of ϵ-Stealthy Attacks on Higher Order Systems

Enoch Kung, Subhrakanti Dey, and Ling Shi

***Abstract*—In a cyber-physical system, security problems are of vital importance as the failure of such system can have catastrophic effects. Detection methods can be employed to sense the existence of an attack. In a previous study of an attack on the controller while avoiding detection in scalar systems under a certain control assumption, the notion of ϵ-stealthiness was introduced and the strength of ϵ-stealthy attacks was fully characterized. We generalize to the vector system and prove the cases in which we show that the limitations of ϵ-stealthy attack do not extend, in the sense that ϵ-stealthy can inflict damage of arbitrary magnitude to a vector system.**

***Index Terms*—Cyber-physical systems, detection, security.**

I. Introduction

In a wireless cyber-physical system (CPS), remote estimation plays a vital role in approximating the system state. However, this set up is open to many forms of potential cyber or physical attacks. Therefore, it is essential for one to devise an accurate estimation method as well as study the effects of attack on a given system. This note will be about the latter.

The security of a CPS is and will continue to be a central topic of study. Communication, transportation, and utility networks are just a few examples of vital systems to modern society. Wireless communication increases the scale of the CPS cheaply but exposes the system to unconventional problems. The compromising of security such as the case of the Maroochy Water Breach [5] and the SQL Slammer worm attack on the nuclear plant [6] emphasizes the importance to study CPS security.

In a typical control system, a plant, sensor and estimator require constant communication, while exposed to natural or malignant sources of corruption. Attackers choose the form and the placement of the attack based on their own ability and purpose. For example, a denial-of-service attack [7] simply prevents a packet of information from successfully transmitting, decreasing the estimation quality. An attacker may also replace the transmitted packet with malicious information [4] further leading the system astray. Therefore, methods of defense must be developed to ensure an acceptable degree of system performance.

One line of defense is to detect the presence of an attack. A detection policy is a protocol in which an estimator decides whether the received data is corrupted. For instance, with a false data injection attack with multiple sensors, a sensor network can verify the validity of one sensor's data by its neighboring sensors [2], [3]. Detection imposes on the attacker a tradeoff; it must maximize its attack but remain invisible.

In [1], a scalar control system under a control assumption is considered in which an attacker may alter the control input constructed by the estimator. The authors then introduced the notion of ϵ-stealthiness; the attacker remains undetected if the corrupted output does not differ from a reference exceeding ϵ under KL divergence. Under such a detection, the paper provides an ϵ-stealthy attack which maximizes the average error covariance. Because it is much more practical to study vector systems, a worthwhile task is to explore these results in a system of higher dimensions and study if, and the extent to which, these results carry over.

In this work, we consider the control system in [1] in a multivariate setting, where an attacker may launch an ϵ-stealthy attack. Two scenarios are analyzed. One is when the state and output variables are equal in length and the other is when the state vector is longer than the output vector. In these two cases, the effect of ϵ-stealthy attacks are different. The tradeoff between the magnitude of damage inflicted to the system and the stealthiness of such attack is only evident in the former case, but not the latter. The main contributions are as follows:

1) In the former scenario, we will provide an upper bound to the average covariance achievable by an ϵ-stealthy attack. An ϵ-stealthy attack is constructed that achieves our upper bound. Along the way, the relationship between this upper bound and the system parameters, which was not evident in the scalar case, is made explicit.
2) In the latter case, we construct a stealthy attack which is capable of setting the average covariance to be arbitrarily large, thereby proving that there exists no upper bound to the average covariance similar to the one in the previous case.

The note is organized as follows. In Section II, the problem will be formulated after a brief summary of the concepts in [1]. Section III presents the main results, along with their proofs. A numerical simulation of the results is given in Section IV. The conclusion is given in the end.

Notations: Throughout this note, $\mathbb{R}^{m\times n}$ represents the set of $m \times n$ real matrices. Let X' be the transpose of X and $\mathcal{N}(\mu, \Sigma)$ a Gaussian distribution with mean μ and covariance Σ. Denote the set of $n \times n$ symmetric matrices by S^n, the set of positive semi-definite matrices by S^n_+, and the set of positive definite matrices by S^n_{++}. Suppose $X \in S^n_{++}$, then $X^{1/2}$ is the positive definite square root of X. Define also the function $\delta(x)$ to be the greater solution to the equation $\delta(x) = 2x + 1 + \log \delta(x)$.

II. Preliminaries

A. Kalman Filter

Consider the system

$$x_{k+1} = Ax_k + Bu_k + w_k$$
$$y_k = Cx_k + v_k$$

Manuscript received January 20, 2016; revised April 27, 2016; accepted April 29, 2016. Date of publication May 9, 2016; date of current version January 26, 2017. This work was supported by RGC General Research Fund 16210015. Recommended by Associate Editor M. L. Corradini.

E. Kung and L. Shi are with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (e-mail: ekung@ust.hk; eesling@ust.hk).

S. Dey is with the Department of Engineering Sciences, Uppsala University, 751 21 Uppsala, Sweden (e-mail: Subhrakanti.Dey@signal.uu.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2016.2565379

where $A, B \in \mathbb{R}^{n\times n}, x_k, u_k, w_k \in \mathbb{R}^n, C \in \mathbb{R}^{m\times n}$, and $y_k, v_k \in \mathbb{R}^m$. The noise variables w_k, v_k are independent and follow the distributions $\mathcal{N}(0, Q)$ and $\mathcal{N}(0, R)$, respectively. The initial state x_0 is a zero-mean Gaussian variable that is independent of w_k and v_k. It is assumed that (A, C) is observable and $(A, Q^{1/2})$ is controllable. These parameters are known to both the estimator and the attacker.

The matrix B is assumed, similar to [1], to be invertible. Furthermore, for simplicity, B can be considered to be I. This does not hinder the results, and will only slightly affect the form of the constructed attack in Theorem III.3.ii and in Theorem III.4.

The problem of estimating the state x_k given the output vector y_k is solved by the Kalman filter. The Kalman filter is a set of equations from which an estimate $\hat{x}_k$ can be obtained such that the error covariance between $\hat{x}_k$ and x_k is minimized; this estimate is named the minimum mean-squared error (MMSE) estimate. Writing

$$\begin{aligned}\hat{x}_k &= \mathbb{E}[x_k|y_1,\dots,y_k]\\ P_k &= \mathbb{E}\left[(x_k-\hat{x}_k)(x_k-\hat{x}_k)'|y_1,\dots,y_k\right]\end{aligned}$$

the Kalman filter are as follows:

$$\begin{aligned}\hat{x}_{k+1} &= A\hat{x}_k + K_k(y_k - CA\hat{x}_k) + u_k\\ P_{k+1|k} &= AP_kA' + Q\\ K_{k+1} &= P_{k+1|k}C'[CP_{k+1|k}C' + R]^{-1}\\ P_{k+1} &= (I - K_kC)P_{k+1|k}\end{aligned}$$

where K_k is the Kalman gain.

By the observability and controllability conditions mentioned, the terms $\{K_k\}$ and $\{P_k\}$ converges exponentially to a steady-state Kalman gain K and error covariance P, respectively. Hence, we may assume steady-state has been achieved. The steady-state covariance P is positive semi-definite solution to $g \circ h(X) = X$, where

$$\begin{aligned}g(X) &= X - XC'[CXC' + R]^{-1}CX\\ h(X) &= AXA' + Q\end{aligned}$$

and $K = h(P)C'[Ch(P)C' + R]^{-1}$.

B. Attack Model

We employ a model in which the attacker corrupts the control vector by altering it arbitrarily. Denote by I_k to be the attacker's information set at time k. The set I_k must satisfy:

1) $u_k \in I_k$ for all k;
2) $I_k \subset I_{k+1}$;
3) I_k is independent of all noise $\{v_i\}_1^\infty$ and $\{w_i\}_1^\infty$.

The system dynamics can be written as

$$\begin{aligned}\tilde{x}_{k+1} &= A\tilde{x}_k + \tilde{u}_k + w_k\\ \tilde{y}_k &= C\tilde{x}_k + v_k\end{aligned} \tag{1}$$

assuming C is full rank, and the estimation equation is

$$\hat{x}_{k+1} = A\hat{x}_k + Kz_k + u_k. \tag{2}$$

The term $z_k = y_k - C\hat{x}_k \sim \mathcal{N}(0, CPC' + R)$ is the innovation. In absence of an attack, this estimation is the MMSE estimate. However, since the plant is now influenced by the corrupted control, the estimation is no longer optimal.

The sub-optimal estimator leads to a higher estimation error, and it is the objective of the attack to maximize this error, which is quantified by the performance metric

$$J = \limsup_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \mathbf{tr}\tilde{P}_i$$

where $\tilde{P}_i = \mathbb{E}[(\tilde{x}_k - \hat{x}_k)(\tilde{x}_k - \hat{x}_k)']$. This metric is the average error covariance over an infinite time horizon.

C. Stealth Model

We will give a brief outline of ϵ-stealthiness as described in [1]. The estimator, not knowing the presence nor the style of attack performed upon it, may establish a detection policy based on the output vectors $\{y_i\}_1^k$ to raise alarms when there is evidence of an attacker's presence.

At time k, the estimator obtains the outputs $\{y_1,\dots,y_k\}$, which it uses to perform a hypothesis test on the two hypotheses

$$\begin{aligned}H_0 &: \text{Attack does not exist}\\ H_1 &: \text{Attack exists.}\end{aligned}$$

Writing

$$\begin{aligned}P_k(H_1|H_0) &= p_k^{FA}\ \text{(False Alarm)}\\ P_k(H_1|H_1) &= p_k^{D}\ \text{(Detection)}\end{aligned}$$

the definition of ϵ-stealthiness is introduced in the following.

Definition II.1 (ϵ-Stealthiness) [1]: For $0 < \delta < 1$, an attack is ϵ-stealthy if for any detector that satisfies $0 < 1 - p_k^D \le \delta$

$$\limsup_{k\to\infty} -\frac{1}{k}\log\left(p_k^{FA}\right) \le \epsilon.$$

It is proven in the paper that this is equivalent to the following condition.

Definition II.2: A sequence of attacks $\{u_k\}$ is ϵ-stealthy for the resulting innovations $\{\tilde{z}_1^k\}$

$$\limsup_{k\to\infty} \frac{1}{k} D\left(\tilde{z}_1^k \| z_1^k\right) \le \epsilon$$

where $D(\tilde{z}_1^k \| z_1^k)$ is the KL divergence between $\tilde{z}_1^k$ and z_1^k, i.e.,

$$D\left(\tilde{z}_1^k \| z_1^k\right) = \int_{-\infty}^{\infty} \log\frac{f_{\tilde{z}}\left(z_1^k\right)}{f_z\left(z_1^k\right)} f_{\tilde{z}}\left(z_1^k\right) dz_1^k.$$

The KL divergence describes the "difference" between two distributions. Here, the ϵ parameter acts as a degree of tolerance for the difference between the $\tilde{z}_1^k$, which is the innovation corrupted by attack, and z_1^k, which is the innovation if not under attack. Hence the attacker avoids being detected if its attack can retain this difference under ϵ.

D. Problem Setup

Given the system (1) and estimator (2), the goal is to find the maximum of J that can be afflicted on the system by an ϵ-stealthy attack. This question is answered for the scalar case $m = n = 1$ in [1], where an upper bound to J is calculated and a stealthy attack is constructed that reaches this bound.

In this note, we continue to look at the optimization

$$\max J = \max_{u_1^\infty} \lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \mathbf{tr}\tilde{P}_i \text{ where } \{u_1^\infty\} \text{ is } \epsilon - \text{stealthy}$$

It will be shown, however, that higher order systems are not as tame and similar results do not always hold. In particular, if $m < n$, the ϵ-stealthy criterion does not limit the power of the attack, in the sense that there exists an ϵ-stealthy attack that can make J arbitrarily large. The answer provided in [1] can be extended nicely only for the case where $m = n$.

III. Results

As oppose to the scalar case, the vector system can be divided into two cases: $m = n$ and $m < n$. The case in which $m > n$ is in fact a subcase of $m = n$. The former case is solved by providing an upper bound of J and the optimal ϵ-stealthy attack that reaches the bound. For the latter case, we will prove that the attacker can design an ϵ-stealthy attack of arbitrary power.

A. $m = n$

We begin the analysis by defining several terms

$$\tilde{P}_i = \mathbb{E}\left[(\tilde{x}_i - \hat{x}_i)(\tilde{x}_i - \hat{x}_i)'\right], \quad \tilde{\Theta}_i = C\tilde{P}_i C' + R$$
$$\Sigma = CPC' + R, \qquad D_k = \frac{1}{k} D\left(\tilde{z}_1^k \| z_1^k\right)$$
$$\mathcal{U}_i = \mathbb{E}[\tilde{u}_i - u_i].$$

Remark: $\tilde{\Theta}_i$ is not the covariance of the distribution of $\tilde{z}_i$ because $\tilde{\Theta}_i = \mathbb{E}[\tilde{z}_i \tilde{z}_i']$ and it is not assumed that the mean of $\tilde{z}_i$ is 0, unlike the Kalman innovation z_i.

An attack is stealthy if the innovation it produces, $\{\tilde{z}_1^k\}$, satisfies

$$\lim_{k\to\infty} \frac{1}{k} D\left(\tilde{z}_1^k \| z_1^k\right) = \lim_{k\to\infty} D_k \le \epsilon.$$

The term D_k can be expanded, after considerable calculation, to be

$$D_k = -\frac{1}{k} h\left(\tilde{z}_1^k\right) + \frac{1}{2}\log\left((2\pi)^n |\Sigma|\right) + \frac{1}{2k}\sum_{i=1}^{k} \mathbf{tr}\, \Sigma^{-1}\tilde{\Theta}_i. \quad (3)$$

Here, $h(\tilde{z}_1^k)$ is the differential entropy of $\tilde{z}_1^k$. We may then obtain the inequality

$$\begin{aligned}
\frac{1}{2k}\sum_{i=1}^{k} \mathbf{tr}\, \Sigma^{-1}\tilde{\Theta}_i &= D_k + \frac{1}{k} h\left(\tilde{z}_1^k\right) - \frac{1}{2}\log\left((2\pi)^n|\Sigma|\right) \\
&\le D_k + \frac{1}{k}\sum_{i=1}^{k} h(\tilde{z}_i) - \frac{1}{2}\log\left((2\pi)^n|\Sigma|\right) \\
&\le D_k + \frac{1}{2k}\sum_{i=1}^{k} \log\left[(2\pi e)^n |\tilde{\Sigma}_i|\right] \\
&\quad - \frac{1}{2k}\sum_{i=1}^{k} \log\left((2\pi)^n|\Sigma|\right) \\
&= D_k + \frac{n}{2} + \frac{1}{2}\log\left(\prod_{i=1}^{k} |\Sigma^{-1}\tilde{\Sigma}_i|\right)^{\frac{1}{k}} \\
&\le D_k + \frac{n}{2} + \frac{1}{2}\log\left(\prod_{i=1}^{k} |\Sigma^{-1}\tilde{\Theta}_i|\right)^{\frac{1}{k}}.
\end{aligned}$$

The final inequality is justified as follows. By the equation $\mathbb{E}[(x-\mu)(x-\mu)'] = \mathbb{E}[xx'] - \mu\mu'$, the covariance of $\tilde{z}_i$ is $\tilde{\Sigma}_i = \tilde{\Theta}_i - C\mathcal{U}_i\mathcal{U}_i'C'$. Since $C\mathcal{U}_i\mathcal{U}_i'C'$ is positive semidefinite, the inequality $\tilde{\Theta}_i \ge \tilde{\Sigma}_i$ holds. Furthermore, both $\tilde{\Theta}_i$ and $\tilde{\Sigma}_i$ are positive definite, hence $|\tilde{\Theta}_i| \ge |\tilde{\Sigma}_i|$. The inequality follows by the fact that log is an increasing function. The others are results of the subadditivity of differential entropy and the maximum entropy theorem [1].

It is a known fact that the eigenvalues of $\Sigma^{-1}\tilde{\Theta}_i$ are equivalent to those of $\Sigma^{-(1/2)}\tilde{\Theta}_i\Sigma^{-(1/2)}$. Denote the eigenvalues of $\Sigma^{-(1/2)}\tilde{\Theta}_i\Sigma^{-(1/2)}$ by λ_{ij}. Translating the above inequality using these eigenvalues, we have

$$\frac{1}{2k}\sum_{i,j}\lambda_{ij} \le D_k + \frac{n}{2} + \frac{1}{2k}\sum_{i,j}\log\lambda_{ij}. \quad (4)$$

The following two lemmas are useful tools in the proof of the the main result.

Lemma III.1: If $\{\lambda_{ij}\}$ satisfies the equality of (4), then there exists $\{\epsilon_{ij}\}$ such that

$$\lambda_{ij} = \delta(\epsilon_{ij}) \text{ and } \frac{1}{k}\sum_{i,j}\epsilon_{ij} = D_k.$$

Proof: If equality of (4) holds, then it is not possible for all i, j to satisfy

$$\frac{1}{2}\lambda_{ij} > \frac{D_k}{n} + \frac{1}{2} + \frac{1}{2}\log\lambda_{ij}.$$

Without losing generality, assume that

$$\frac{1}{2}\lambda_{11} \le \frac{D_k}{n} + \frac{1}{2} + \frac{1}{2}\log\lambda_{11}.$$

Since it is also known that

$$\frac{1}{2}\lambda_{11} \ge \frac{1}{2} + \frac{1}{2}\log\lambda_{11}$$

there must exist $\epsilon_{11} \in [0, D_k/n]$ such that

$$\frac{1}{2}\lambda_{11} = \epsilon_{11} + \frac{1}{2} + \frac{1}{2}\log\lambda_{11}$$

that is, $\lambda_{11} = \delta(\epsilon_{11})$. This can be subtracted from both sides of (4) to get

$$\frac{1}{2k}\sum_{(i,j)\neq(1,1)}\lambda_{ij} = (D_k - \epsilon_{11}) + \left(\frac{n}{2} - \frac{1}{2k}\right) + \frac{1}{2k}\sum_{(i,j)\neq(1,1)}\log\lambda_{ij}.$$

Repeat the same reasoning to obtain subsequent values of ϵ_{ij} such that $\lambda_{ij} = \delta(\epsilon_{ij})$. Plugging this back to (4) so that equality is satisfied, then

$$\begin{aligned}
D_k + \frac{n}{2} + \frac{1}{2k}\sum_{i,j}\log\lambda_{ij} &= \frac{1}{2k}\sum_{i,j}\lambda_{ij} = \frac{1}{2k}\sum_{i,j}\delta(\epsilon_{ij}) \\
&= \frac{1}{2k}\sum_{i,j}\left[2\epsilon_{ij} + 1 + \log\delta(\epsilon_{ij})\right] \\
&= \frac{1}{2k}\sum_{i,j}[2\epsilon_{ij} + 1 + \log\lambda_{ij}].
\end{aligned}$$

Canceling terms on both sides results in

$$kD_k = \sum_{i,j}\epsilon_{ij}.$$

□

Lemma III.2: Let $X = (x_{ij}) \in S_{++}^n$ with its eigenvalues $s_1 \ge \cdots \ge s_n \ge 0$ and $Y \in S^n$ with eigenvalues $y_1 \ge \cdots \ge y_n$. Then

$$\mathbf{tr}(XY) \le s_1 y_1 + \cdots + s_n y_n.$$

Proof: By the symmetry of Y, there exists orthogonal Q such that $Y = Q\bar{Y}Q'$, where $\bar{Y} = \text{diag}\ (y_1, \ldots, y_n)$. Then $\mathbf{tr}(XY) = \mathbf{tr}(Q'XQ\bar{Y})$. Since $Q'XQ$ is positive semi-definite, it can be assumed without loss of generality that Y is already diagonal.

Suppose x_k for $k = 1, \ldots, n$ are the diagonal elements of X in decreasing order. By the rearrangement inequality

$$\mathbf{tr}(XY) = x_{11}y_1 + \cdots + x_{nn}y_n \le x_1 y_1 + \cdots + x_n y_n.$$

Furthermore, as X is symmetric, by the Schur-Horn Theorem [9], the eigenvalues of X majorizes its diagonal, that is

$$\sum_{j=1}^{r} x_j \le \sum_{j=1}^{r} s_j, \quad r = 1, 2, \ldots, n.$$

Then

$$\begin{aligned}\mathbf{tr}(XY) &\le x_1y_1+\cdots+x_ny_n\\ &= y_n\left(\sum_{j=1}^{i}x_j\right)+\sum_{i=1}^{n-1}(y_i-y_{i+1})\left(\sum_{j=1}^{i}x_j\right)\\ &\le y_n\left(\sum_{j=1}^{i}s_j\right)+\sum_{i=1}^{n-1}(y_i-y_{i+1})\left(\sum_{j=1}^{i}s_j\right)\\ &= s_1y_1+\cdots+s_ny_n.\end{aligned}$$

This lemma shows that the maximization of $\mathbf{tr}(XY)$ rests solely on properly selecting eigenvalues. □

The following theorem is the main result of this section.

Theorem III.3: Suppose $m=n$ and C is invertible, its inverse denoted by E. Then

i)

$$J=\lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\mathbf{tr}\tilde{P}_i\le\mathbf{tr}P+\sum_{j=1}^{n}s_j\left(\delta_j^*(\epsilon)-1\right).$$

where s_j are the eigenvalues of $\Sigma^{1/2}E'E\Sigma^{1/2}$ with $s_1\ge\cdots\ge s_n$ and $\delta_j^*(\epsilon)$ are defined by

$$\frac{s_j}{s_1}=\frac{1-\frac{1}{\delta_j^*}}{1-\frac{1}{\delta_1^*}}\text{ and }2\epsilon=\sum_{j=1}^{n}\delta_j^*-\log\delta_j^*-1.$$

ii) There exists an attack $\tilde{u}_1^\infty$ that the resulting $\tilde{P}_i$ achieves the upper bound.

Proof of I: The proof of the theorem will be carried out in the following way as we aim to bound $\mathbf{tr}\tilde{P}_i$. Suppose that it is possible for the bounding of $\mathbf{tr}\tilde{P}_i$ to be transformed to the bounding of $\mathbf{tr}(XY)$, where $X\in S_{++}^n$ is fixed by the system, and $Y\in S^n$ is free to design under constraint. Then the upper bound is dependent solely on the eigenvalues of X and an appropriately selected eigenvalue of Y by the second lemma. The first lemma then reveals the set of eigenvalues we can choose from. From this set of eigenvalues, Lagrange multipliers will be carried out to find the optimal eigenvalues that yields the desired bound.

To transform the maximization problem, we notice that

$$\begin{aligned}\mathbf{tr}\tilde{P}_i &= \mathbf{tr}P+\mathbf{tr}\left(E(\tilde{\Theta}_i-\Sigma)E'\right)\\ &= \mathbf{tr}P+\mathbf{tr}\left(\Sigma^{\frac{1}{2}}E'E\Sigma^{\frac{1}{2}}\left(\Sigma^{-\frac{1}{2}}\tilde{\Theta}_i\Sigma^{-\frac{1}{2}}-I\right)\right).\end{aligned}\tag{5}$$

The first term is constant, hence we want to bound the second term. The matrix $\Sigma^{1/2}E'E\Sigma^{1/2}\in S_{++}^n$ and $\Sigma^{-(1/2)}\tilde{\Theta}_i\Sigma^{-(1/2)}-I\in S^n$, hence Lemma III.2. can be applied. The former term is fixed by system parameters, whereas the second term includes $\tilde{\Theta}_i$, which is determined by the attacker.

If we represent their eigenvalues as $\{s_j\}$ and $\{\tilde{\Lambda}_i^j\}$, respectively, listed in decreasing order, the lemma yields

$$\mathbf{tr}\left(\Sigma^{\frac{1}{2}}E'E\Sigma^{\frac{1}{2}}\left(\Sigma^{-\frac{1}{2}}\tilde{\Theta}_i\Sigma^{-\frac{1}{2}}-I\right)\right)\le\tilde{\Lambda}_i^1s_1+\cdots+\tilde{\Lambda}_i^ns_n.\tag{6}$$

The remaining task is to maximize (6) through a correct selection of $\{\tilde{\Lambda}_i^j\}$ while satisfying (4), i.e.,

$$-\frac{n}{2}-\frac{1}{2k}\sum_{i,j}\log\left[\left(\tilde{\Lambda}_i^j+1\right)\right]+\frac{1}{2}\frac{1}{k}\sum_{i,j}\left[1+\tilde{\Lambda}_i^j\right]\le D_k.\tag{7}$$

The expression on the left side is an increasing function for positive $\tilde{\Lambda}_i^j$, meaning that if the inequality is strict, one may increase any $\tilde{\Lambda}_i^j$, which in turn increases (6). In other words, the optimal choice of $\tilde{\Lambda}_i^{j*}$ must satisfy equality of (7), paving the way for us to use Lemma III.1.

The lemma guarantees the existence of ϵ_{ij} such that

$$1+\tilde{\Lambda}_i^j=\delta(\epsilon_{\mathrm{ij}})\text{ and }\frac{1}{k}\sum_{i,j}\epsilon_{\mathrm{ij}}=D_k.$$

With these new terms, the maximization of (6) becomes

$$\max\sum_{i,j}s_j\left(\delta(\epsilon_{\mathrm{ij}})-1\right)\text{ subject to }\frac{1}{k}\sum_{i,j}\epsilon_{\mathrm{ij}}=D_k.\tag{8}$$

Recall that $\delta(x)$ solves $\delta(x)=2x+1+\log\delta(x)$, so the constraint can be rephrased as

$$\frac{1}{2k}\sum_{i,j}\delta(\epsilon_{\mathrm{ij}})-1-\log\delta(\epsilon_{\mathrm{ij}})=D_k$$

and instead of using ϵ_{ij} as our variables, we can take $\delta(\epsilon_{\mathrm{ij}})$ to be the variables. Naturally, Lagrange multipliers can be employed to solve for optimal values $\delta^*(\epsilon_{\mathrm{ij}})$. Solving

$$\nabla\sum_{i,j}s_j\left(\delta(\epsilon_{\mathrm{ij}})-1\right)=\eta\nabla\left(\frac{1}{2}\sum_{i,j}\delta(\epsilon_{\mathrm{ij}})-1-\log\delta(\epsilon_{\mathrm{ij}})\right)$$

gives us the equation

$$(s_1,\ldots,s_n)=\eta\left(\frac{1}{2}\left[1-\frac{1}{\delta(\epsilon_{11})}\right],\ldots,\frac{1}{2}\left[1-\frac{1}{\delta(\epsilon_{kn})}\right]\right).$$

Thus the optimal values $\delta^*(\epsilon_{\mathrm{ij}})$ must satisfy the equations

$$\frac{s_j}{s_1}=\frac{1-\frac{1}{\delta^*(\epsilon_{\mathrm{ij}})}}{1-\frac{1}{\delta^*(\epsilon_{11})}}\text{ for all }i,j$$

$$2D_k=\sum_{j=1}^{n}\delta^*(\epsilon_{\mathrm{ij}})-\log\delta^*(\epsilon_{\mathrm{ij}})-1$$

Note that these optimal values $\delta^*(\epsilon_{\mathrm{ij}})$ are dependent on D_k and j but not on i, hence they can be denoted by $\delta_j^*(D_k)$.

This results in the upper bound of (5)

$$\mathbf{tr}\tilde{P}_i=\mathbf{tr}P+\mathbf{tr}\left(Q_i'\Sigma^{\frac{1}{2}}E'E\Sigma^{\frac{1}{2}}Q_i\tilde{\Lambda}_i\right)\le\mathbf{tr}P+\sum_{j=1}^{n}s_j\left(\delta_j^*(D_k)-1\right)$$

which immediately leads to

$$\begin{aligned}J &= \lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\mathbf{tr}\tilde{P}_i\\ &\le \lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}\left[\mathbf{tr}P+\sum_{j=1}^{n}s_j\left(\delta_j^*(D_k)-1\right)\right]\\ &= \lim_{k\to\infty}\mathbf{tr}P+\sum_{j=1}^{n}s_j\left(\delta_j^*(D_k)-1\right)\\ &= \mathbf{tr}P+\sum_{j=1}^{n}s_j\left(\delta_j^*(\lim_{k\to\infty}D_k)-1\right)\\ &\le \mathbf{tr}P+\sum_{j=1}^{n}s_j\left(\delta_j^*(\epsilon)-1\right).\end{aligned}$$

□

Clearly, this proof extends the results from [1] because in the scalar case, the eigenvalue of $\Sigma^{1/2}E'E\Sigma^{1/2}$ equals σ_z^2/c^2 and $\delta_j^*(\epsilon)$ is

simply $\delta(\epsilon)$. Then, given that $\sigma_z^2 = c^2 P + r$

$$\begin{aligned}\mathbf{tr}P + s_1\left(\delta_j^*(\epsilon) - 1\right) &= P + \frac{\sigma_z^2}{c^2}\left(\delta(\epsilon) - 1\right)\\ &= \delta(\epsilon)P + \frac{\left(\delta(\epsilon) - 1\right)r}{c^2}.\end{aligned}$$

This is the result of [1].

Proof of II: In this proof, we first propose an attack. Then we verify that it indeed reaches the upper bound stated in Theorem III.3. and is ϵ-stealthy.

Define Q to be the matrix that diagonalizes $\Sigma^{1/2}E'E\Sigma^{1/2}$, i.e.,

$$Q'\Sigma^{\frac{1}{2}}E'E\Sigma^{\frac{1}{2}}Q = S = \text{diag}\,(s_1, \ldots, s_n).$$

Also, let $\{\zeta_k\}$ be a sequence of Gaussian random variables of distribution $\mathcal{N}(0, E\Sigma^{1/2}Q\Lambda Q'\Sigma^{1/2}E')$, where $\Lambda = \text{diag}\,(\delta_1^*(\epsilon) - 1, \ldots, \delta_n^*(\epsilon) - 1)$. Define an attack by

$$\tilde{u}_k = u_k - (A - KC)\zeta_{k-1} + \zeta_k \tag{9}$$

with $\zeta_1 = 0$. To facilitate the verification of this construction's validity, define an intermediate process x_k^s defined by

$$x_{k+1}^s = Ax_k^s + K\left(y_k - Cx_k^s\right) + \tilde{u}_k$$

with $x_1^s = 0$. This yields the MMSE estimate of the state x_k, in particular, $\mathbb{E}[(x_k^s - x_k)(x_k^s - x_k)^\top] = P$. If $e_k = \hat{x}_k - x_k^s$, we have

$$\begin{aligned}\tilde{z_k} &= y_k - C\hat{x}_k = (y_k - Cx_k^s) - Ce_k\\ e_{k+1} &= (A - KC)e_k + (A - KC)\zeta_{k-1} - \zeta_k.\end{aligned} \tag{10}$$

Given $e_1 = 0$, the solution to the second recursion is $e_k = -\zeta_{k-1}$.

It can now be verified that

$$\begin{aligned}\mathbf{tr}\ \tilde{P}_i &= \mathbb{E}\left[\left(\hat{x}_i - x_i^s\right)^\top\left(\hat{x}_i - x_i^s\right)\right] + \mathbb{E}\left[\left(x_i^s - x_i\right)^\top\left(x_i^s - x_i\right)\right]\\ &\quad + 2\mathbb{E}\left[\left(\hat{x}_i - x_i^s\right)^\top\left(x_i^s - x_i\right)\right]\\ &= \mathbf{tr}P + \mathbf{tr}\ \mathbb{E}\left[e_k e_k^\top\right] = \mathbf{tr}P + \mathbf{tr}S\Lambda\\ &= \mathbf{tr}P + \sum_{j=1}^{n} s_j\left(\delta_j^*(\epsilon) - 1\right)\end{aligned}$$

and consequently

$$J = \lim_{k\to\infty}\ \mathbf{tr}\ \tilde{P}_i = \mathbf{tr}P + \sum_{j=1}^{n} s_j\left(\delta_j^*(\epsilon) - 1\right)$$

which is our stated upper bound.

It remains to show that this attack is ϵ-stealthy. By (10), it is immediate that $\tilde{z}_i \sim \mathcal{N}(0, \Sigma^{1/2}[I + Q\Lambda Q']\Sigma^{1/2})$. Since $\tilde{\Sigma}_i = \Sigma^{1/2}[I + Q\Lambda Q']\Sigma^{1/2}$, a quick calculation shows

$$\begin{aligned}\mathbf{tr}(\Sigma^{-1}\tilde{\Sigma}_i) &= \mathbf{tr}(I + \Lambda) = \sum_{j=1}^{n}\delta_j^*(\epsilon)\\ |\Sigma^{-1}\tilde{\Sigma}_i| &= |I + \Lambda| = \prod_{j=1}^{n}\delta_j^*(\epsilon).\end{aligned}$$

Plugging this into (3), noting that the differential entropy $(1/k)h(\tilde{z}_1^k) = (1/2)\log(2\pi e)^n\left(\prod_{i=1}^k|\tilde{\Sigma}_i|\right)^{1/k}$ if $\tilde{z}_1^k$ is Gaussian, we have

$$\begin{aligned}&-\frac{n}{2} - \frac{1}{2}\log\left(\prod_{i=1}^{k}|\Sigma^{-1}\tilde{\Sigma}_i|\right)^{\frac{1}{k}} + \frac{1}{2k}\sum_{i=1}^{k}\mathbf{tr}\Sigma^{-1}\tilde{\Sigma}_i\\ &\qquad = -\frac{n}{2} - \frac{1}{2}\sum_{j=1}^{n}\log\delta_j^*(\epsilon) + \frac{1}{2}\sum_{j=1}^{n}\delta_j^*(\epsilon) = \epsilon.\end{aligned}$$

If B is a general invertible matrix, then (9) would be rewritten as

$$B\tilde{u}_k = Bu_k - (A - KC)\zeta_{k-1} + \zeta_k.$$

The attack would take the form

$$\tilde{u}_k = u_k - B^{-1}(A - KC)\zeta_{k-1} + B^{-1}\zeta_k.$$

□

Finally to settle the case $n < m$, suppose $C \in \mathbb{R}^{m\times n}$ and full rank. Intuitively, if C is one-to-one, then all of the information in the state variable should roughly be encoded into the output, hence should not be any different from the case when C is square. In detail, there exists an invertible matrix row operation $\bar{C} \in \mathbb{R}^{m\times m}$ such that

$$C = \bar{C}\begin{bmatrix} I \\ 0 \end{bmatrix}$$

which when substituted into the system equations renders

$$y_k = Cx_k + v_k = \bar{C}\begin{bmatrix} x_k \\ 0 \end{bmatrix} + v_k.$$

Since $\bar{C}$ is a square, full rank matrix, the results obtained in this section extends to this scenario.

B. $m < n$

In this section, it will be shown that the detection employed by the estimator is not effective in a vector system against ϵ-stealthy attacks in the sense that there exist an ϵ-stealthy attack that can arbitrarily increase J.

Theorem III.4: Let $m < n$ and assume that C is full rank. There exists an attack $\tilde{u}_1^\infty$ such that from its produced error covariance $\{\tilde{P}_i\}$, the performance metric J can be arbitrarily large.

Proof: Note that by the surjectivity of C, there exists an invertible matrix $\bar{C}$ such that $C = [I \quad 0]\bar{C}$. Suppose we find a $\tilde{\Sigma}$ satisfying (4), then writing

$$\bar{C}\tilde{P}_i\bar{C}' = \begin{bmatrix}\bar{P}_i^1 & \bar{P}_i^2\\ \bar{P}_i^3 & \bar{P}_i^4\end{bmatrix} \tag{11}$$

will result in the equality

$$C\tilde{P}_iC' = \bar{P}_i^1 = \tilde{\Sigma} - R.$$

This means that the other submatrices are degrees of freedom which may cause J to diverge. For example, choose $\bar{P}_i^2 = \bar{P}_i^3 = 0$ and $\bar{P}_i^4 = \alpha I$ for some α. Then

$$\begin{aligned}\mathbf{tr}\left(\begin{bmatrix}\bar{P}_i^1 & 0\\ 0 & \alpha I\end{bmatrix}\right) &= \mathbf{tr}\ \bar{C}\tilde{P}_i\bar{C}'\\ &= \mathbf{tr}\ \tilde{P}_i\bar{C}'\bar{C}\end{aligned}$$

which by the inequality $\mathbf{tr}AB \le \mathbf{tr}(A)\mathbf{tr}(B)$ for positive definite matrices A, B [8]

$$\mathbf{tr}\ \tilde{P}_i \ge \frac{\mathbf{tr}\left(\begin{bmatrix}\bar{P}_i^1 & 0\\ 0 & \alpha I\end{bmatrix}\right)}{\mathbf{tr}\ \bar{C}'\bar{C}}. \tag{12}$$

With $\bar{P}_i^1$ and $\bar{C}'\bar{C}$ being constant for a fixed ϵ and C, it is straightforward to see that the selection of α can arbitrarily increase the term $\lim_{k\to\infty}(1/k)\sum_{i=1}^k\mathbf{tr}\tilde{P}_i$.

With this strategy in mind, let $\zeta_k \sim \mathcal{N}(0, Z)$, where Z satisfies

$$\bar{C}Z\bar{C}' = \begin{bmatrix}\left(\delta\left(\frac{\epsilon}{n}\right) - 1\right)\Sigma & 0\\ 0 & \beta I\end{bmatrix}.$$

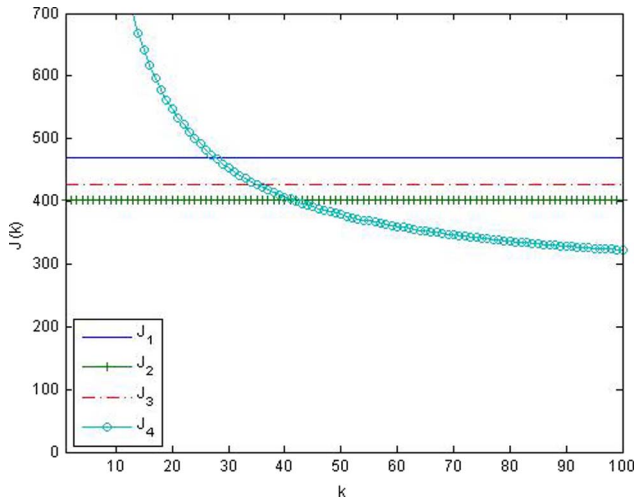

Fig. 1. J_1: upper bound; J_2: $\epsilon_{\rm ij} = \epsilon/n$; J_3: $\epsilon_{\rm ij} = 2j\epsilon/n(n+1)$; J_4: $\epsilon_{\rm ij}$ is random.

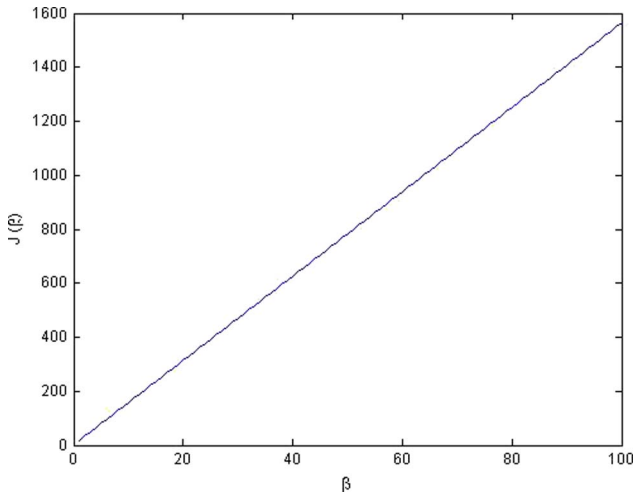

Fig. 2. Average covariance described in part B dependent on β.

As with the proof of Theorem III.3., define the attack

$$\tilde{u}_k = u_k - (A - KC)\zeta_{k-1} + \zeta_k$$

or if $B \neq I$

$$\tilde{u}_k = u_k - B^{-1}(A - KC)\zeta_{k-1} + B^{-1}\zeta_k.$$

It can now be verified that

$$\begin{aligned}\mathbf{tr}\,\tilde{P}_i &= \mathbf{tr}\,\mathbb{E}\left[e_k e_k^\top\right] + \mathbf{tr}\,P \\ &\geq \frac{1}{\mathbf{tr}\,\bar{C}'\bar{C}}\left[\mathbf{tr}\,\left(\delta\left(\frac{\epsilon}{n}\right)-1\right)\Sigma + (n-m)\beta\right] + \mathbf{tr}\,P. \qquad (13)\end{aligned}$$

Since β can be arbitrarily large, so can $\lim_{k\to\infty}(1/k)\sum_{i=1}^{k}\mathbf{tr}\tilde{P}_i$.

The final requirement is to prove that this attack is ϵ-stealthy. By (10), it follows that $\tilde{z}_k \sim \mathcal{N}(0, \delta(\epsilon/n)\Sigma)$, that is, $\tilde{\Sigma}_i = \delta(\epsilon/n)\Sigma$. So

$$\frac{1}{k}D\left(\tilde{z}_1^k \| z_1^k\right) = \frac{n}{2}\delta\left(\frac{\epsilon}{n}\right) - \frac{n}{2} - \frac{n}{2}\log\delta\left(\frac{\epsilon}{n}\right) = \frac{n}{2}\left(2\frac{\epsilon}{n}\right) = \epsilon.$$

Hence our constructed attack $\tilde{u}_k$ is ϵ-stealthy. □

IV. Numerical Results

The two results will be illustrated by numerical examples. It can be seen that when $m = n$, no other average covariances obtained from stealthy attacks can exceed the stated bound. As for the second result, the average covariance resulting from the proposed attack in the previous section is shown to increase indefinitely with β.

For Fig. 1, we used

$$A = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},\ C = \begin{bmatrix}3 & 4\\ 1 & 1\end{bmatrix}$$

$$Q = R = \begin{bmatrix}0.6 & 0\\ 0 & 0.3\end{bmatrix},\ \epsilon = 0.1.$$

The dependent variable is the term $(1/k)\sum_{i=1}^{k}\mathbf{tr}\tilde{P}_i$, which conveniently can be denoted by $J(k)$. Aside from the upper bound, the other average covariances are obtained by choosing $\{\epsilon_{\rm ij}\}$ satisfying the constraint

$$\lim_{k\to\infty}\frac{1}{k}\sum_{i,j}\epsilon_{\rm ij} = \epsilon.$$

J_2 is obtained by taking $\epsilon_{\rm ij}$ to be constant; J_3 is obtained by letting $\epsilon_{i1} = 2\epsilon/(n(n+1))$ and $\epsilon_{i2} = 2\epsilon_{i1}, \epsilon_{i3} = 3\epsilon_{i1}, \ldots, \epsilon_{\rm in} = n\epsilon_{i1}$; J_4 is obtained by randomly selecting $\epsilon_{\rm ij}$ such that the constraint is satisfied. As $k \to \infty$, J_1 is the highest.

The reason that J_4 is greater than J_1 for smaller values of k is that the optimal values $\{\delta_j^*\}$ for the optimization problem (8) when $D_k = \epsilon$ is not the optimal values for other choices of D_k. Therefore, it is possible that the optimal values for a certain $D_k \neq \epsilon$ can achieve a higher average covariance at time k than $\{\delta_j^*\}$. However, the assumption that D_k tends to ϵ as k increases implies that $J_4(k)$ must sink below our upper bound as $k \to \infty$ which is evident in this example.

For $m < n$, the parameters are

$$A = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix},\ C = \begin{bmatrix}3 & 4\end{bmatrix}$$

$$Q = \begin{bmatrix}0.6 & 0\\ 0 & 0.3\end{bmatrix},\ R = 0.6,\ \epsilon = 0.1.$$

Then by the proposed attack in the previous section, taking

$$Z = \bar{C}^{-1}\begin{bmatrix}\left(\delta\left(\frac{\epsilon}{n}\right)-1\right)\Sigma & 0\\ 0 & \beta I\end{bmatrix}(\bar{C}')^{-1}$$

we acquire

$$\lim_{k\to\infty}\frac{1}{k}\sum_{i}\tilde{P}_i = \mathbf{tr}Z + \mathbf{tr}P$$

which can be denoted by $J(\beta)$. By simple observation, this value increases linearly by β; this is shown in the Fig. 2.

V. Conclusion and Future Work

In the framework of ϵ-stealthiness, we aim to study the estimation performance under a stealthy attack. In this work, we specify in higher dimensions the situation where results carry from [1] and when they do not hold. In the vector case, one can see the interplay between system parameters with greater clarity. The results further shows that in the more practical setting, with short output vectors and long state vectors, the ϵ-stealthiness detection method is ineffective in preventing an attack.

The next step naturally is to consider the defence against ϵ-stealthiness when $m < n$. The objective of the controller would be to expose the attacker, if present, by maximizing the KL divergence, while the attacker attempts the opposite. In this respect, the problem can be formulated as a two-person infinite horizon dynamic game between the controller and the attacker. This and other defence mechanisms are beyond the scope of this note and will be investigated in future work.

REFERENCES

[1] C. Z. Bai, F. Pasqualetti, and V. Gupta, "Security in stochastic control systems: Fundamental limitations and performance bounds," in *Proc. Amer. Control Conf.*, Chicago, IL, USA, Jul. 1–3, 2015, pp. 195–200.

[2] F. Ye, H. Luo, S. Lu, and L. Zhang, "Statistical en-route filtering of injected false data in sensor networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2004, vol. 4, pp. 839–850.

[3] V. Shukla and D. Qiao, "Distinguishing data transience from false injection in sensor networks," in *Proc. IEEE 4th Annu. Commun. Soc. Conf. Sens., Mesh Ad Hoc Commun. Netw.*, 2007, pp. 41–50.

[4] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. Comput. Commun. Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[5] J. Slay and M. Miller, "Lessons learned from the Maroochy water breach," *Critical Infrastruct. Protect.*, vol. 253, pp. 73–82, 2007.

[6] S. Kuvshinkova, "SQL slammer worm lessons learned for consideration by the electricity sector," *North Amer. Elect. Reliab. Council*, Jun. 2003.

[7] S. Amin, A. Cárdenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," *Hybrid Syst.: Comput. Control*, vol. 5469, pp. 31–45, Apr. 2009.

[8] X. Yang, X. Yang, and K. L. Teo, "A matrix trace inequality," *J. Math. Anal. Appl.*, vol. 263, pp. 327–333, 2001.

[9] "Uber eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie, Sitzungsber," *Berl. Math. Ges.*, vol. 22, pp. 9–20, 1923.

[10] S. H. Ahmed, G. Kim, and D. Kim, "Cyber physical system: Architecture, applications and research challenges," in *Wireless Days, IFIP*, Valencia, Spain, Nov. 2013, pp. 1–5.