# Secure Dynamic State Estimation via Local Estimators

Yilin Mo*, Emanuele Garone†

***Abstract*— We consider the problem of estimating the state of a linear time-invariant Gaussian system using $m$ sensors, where a subset of the sensors can potentially be compromised by an adversary. We prove that under mild assumptions, we can decompose the optimal Kalman estimate as a weighted sum of local state estimates, each of which is derived using only the measurements from a single sensor. We then propose a convex optimization based approach, instead of the weighted sum approach, to combine the local estimate into a more secure state estimate. Our proposed estimator coincides with the Kalman estimator with certain probability when all sensors are benign and is stable when less than half of the sensors are compromised. Numerical simulations are provided to illustrate the performance of the proposed state estimation scheme.**

## I. Introduction

The increasing use of networked embedded sensors to monitor and control critical infrastructures provides potential malicious agents with the opportunity to disrupt their operations by corrupting sensor measurements. Supervisory Control And Data Acquisition (SCADA) systems, for example, run a wide range of safety critical plants and processes, including manufacturing, water and gas treatment and distribution, facility control and power grids. A wide variety of motivations exists for launching an attack on such kind of systems, ranging from financial reasons, e.g., reducing the electricity bill, all the way to terrorism, e.g., threatening the life of possibly an entire population by controlling electricity and other life-critical resources. A successful attack to such kind of systems may significantly hamper the economy, the environment, and may even lead to the loss of human life. The first-ever SCADA system malware (called Stuxnet) was detected in July 2010 and rose significant concerns about SCADA system security [1], [2]. The recent Ukraine power plant hack provides a clear example of the catastrophic outcomes of a successful attack on SCADA systems. The research community has acknowledged the importance of addressing the challenge of designing secure detection, estimation and control systems [3].

The problem of detecting and isolating faulty sensors has been well studied over the past decades. Bad data detection and identification techniques have been widely used in large scaled systems such as power grid [4]. While such approaches are very successful in detecting and removing random failures, they are not effective against an intelligent adversary. Liu et al. [5] illustrate how an adversary can inject a stealthy input into the measurements to change the state estimation, without being detected by the bad data detector. Sandberg et al. [6] consider how to find a sparse stealthy input, which enables the adversary to launch an attack with a minimum number of compromised sensors. Kim et al. [7] studied a so-called framing attack that can misled the bad data detector to mistakenly remove critical measurements, without which the network is unobservable. Xie et al. [8] further illustrate that stealthy integrity attacks on state estimation can lead to a financial gain in the electricity market for the adversary.

For dynamical system, detecting malicious components via fault detection and isolation based methods has also been extensively studied [9], [10]. However, pinpointing the exact set of the malicious components is in general a computationally hard problem, as it either involves generating a residue filter for every possible set of malicious sensor [9] or solving an $L_0$ minimization problem [10], both of which are combinatorial in nature.

Another area of research is the design of state estimators that can tolerate a small portion of the sensory data being altered. For static estimation problem, robust estimators, e.g M-estimator, L-estimator, and R-estimator, have also been extensively studied in the literature [11], [12], [13]. Usually, the robustness is measured by breakdown points [14], [15] or influence functions [16]. Mo and Sinopoli [17] propose an estimator that has minimum mean squared error against the worst-case attacks.

However, it must be remarked that the problem of designing a secure state estimator for a dynamical system is much more challenging. Fig 1 illustrates the information flow of a standard Kalman filter. It is worth noticing that the bias injected by an adversary can accumulate in the state estimation and that the adversary can potentially exploit this fact to introduce a large or even unbounded estimation error [18], [19].

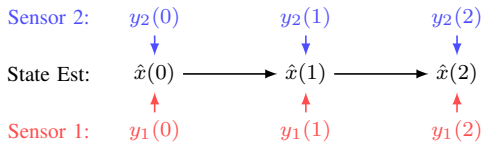


Fig. 1. The information flow of the Kalman Filter.

To address this problem, Fawzi et al.[10] propose to use a moving horizon approach. In other words, the estimator will only use the measurements from time $k-T+1$ to time $k$ to estimate the current state $x(k)$, which effectively reduces the dynamic state estimation problem into a static estimation problem. This approach is further generalized by Pajic et al. [20], [21] to systems subject to random or

*: Yilin Mo is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Email: ylmo@ntu.edu.sg

†: Emanuele Garone is with the Control and System Analysis department, Universite Libre de Bruxelles, Brussels, Belgium . Email: egarone@ulb.ac.be

bounded noise. The main merit of this approach is that the static estimation problem can be solved efficiently using $\ell_1$ relaxation by exploiting the sparseness of the bias injected by the adversary. However, the sensory data before time $k-T$ are discarded in the moving horizon approach, which may result in a degradation of the estimation performance.

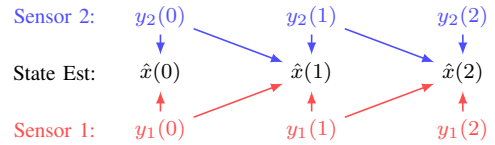


Fig. 2. The information flow of the estimator proposed in [10], [20], [21] with window size $T=2$. Notice that the state estimate $\hat{x}(k)$ only depends on $y_i(k-1)$ and $y_i(k)$.

In this paper, we consider the problem of designing a secure state estimator for a linear time-invariant Gaussian system, subject to up to $p$ compromised sensors. The set of the malicious sensors is assumed to be fixed over time. The structure of our estimate is illustrated in Fig 3 and is described below:

1) For each sensor $i$, we construct a local state estimator, which leverages all the historical measurements from itself to derive a local state estimate.
2) The current global state estimate can then be computed based solely on the *current* local state estimates using a secure fusion scheme.

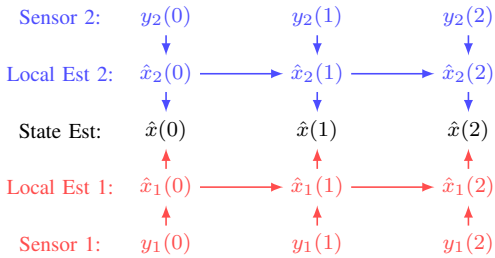


Fig. 3. The information flow of the proposed filter.

The main merits of our approach are twofold:

1) The historical sensory data are stored in the local state estimate and are never discarded. As a result, when the system is not under the attack, we can recover the optimal Kalman estimate with certain probability.
2) Since we assume that the set of the compromised sensors is fixed, there are at most $p$ corrupted local state estimate. Notice that computing the current state estimation based on the local state estimation is a static problem. As a result, we can use $\ell_1$ based method to generate a stable state estimate.

The rest of the paper is organized as follows: In Section II the setting of the problem is introduced with all the relevant notation. In Section III, we prove that the Kalman estimator can be decomposed as a linear combination of local estimators. A convex optimization based approach is proposed in Section IV to derive a more secure state estimate from the local estimates. Section V extends our results to non-observable cases. The performance of the proposed estimator is illustrated via a numerical example in Section VI and finally Section VII concludes the paper.

## II. Problem Formulation

In this paper, we consider a secure dynamic state estimation problem. Consider the linear time-invariant system:

$$x(k+1) = Ax(k) + w(k), \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the state, $w(k) \sim \mathcal{N}(0, Q)$ are i.i.d. Gaussian process noise with zero mean and covariance matrix $Q > 0$. The initial state $x(0) \sim \mathcal{N}(0, \Sigma)$ is assumed to be zero mean Gaussian and is independent from the noise process $\{w(k)\}$.

It is assumed that $m$ sensors are measuring the system and that the measurement from the $i$th sensor is:

$$y_i(k) = C_i x(k) + v_i(k) + a_i(k), \tag{2}$$

where $y_i(k) \in \mathbb{R}$ and $v_i(k)$ is Gaussian measurement noise. The scalar $a_i(k)$ denotes the bias injected by an adversary. Clearly, for a benign sensor $i$, $a_i(k) = 0$ for all $k$ while for a compromised sensor $i$, $a_i(k)$ can be arbitrary. We further assume that the set of compromised sensor remains constant over time.

By defining the aggregated vectors

$$y(k) \triangleq \begin{bmatrix} y_1(k) \\ \vdots \\ y_m(k) \end{bmatrix}, C(k) \triangleq \begin{bmatrix} C_1(k) \\ \vdots \\ C_m(k) \end{bmatrix},$$
$$a(k) \triangleq \begin{bmatrix} a_1(k) \\ \vdots \\ a_m(k) \end{bmatrix}, v(k) \triangleq \begin{bmatrix} v_1(k) \\ \vdots \\ v_m(k) \end{bmatrix}, \tag{3}$$

we can rewrite (2) as

$$y(k) = Cx(k) + v(k) + a(k). \tag{4}$$

We assume that $v(k) \sim \mathcal{N}(0, R)$ with $R > 0$ is i.i.d and independent of the noise process $\{w(k)\}$ and the initial condition $x(0)$. Without loss of generality, we assume $(A, C)$ observable[1].

If all sensors are benign, i.e., $a(k) = 0$ for all $k$, the optimal state estimator is the classical Kalman filter:

$$\hat{x}(k) = \hat{x}(k|k-1) + K(k)\left[y(k) - C\hat{x}(k|k-1)\right],$$
$$P(k) = P(k|k-1) - K(k)CP(k|k-1),$$

where

$$\hat{x}(k+1|k) = A\hat{x}(k),\ P(k+1|k) = AP(k)A^T + Q,$$
$$K(k) = P(k|k-1)C^T(CP(k|k-1)C^T + R)^{-1},$$

with initial condition

$$\hat{x}(0|-1) = 0,\ P(0|-1) = \Sigma.$$

[1]In the case $(A, C)$ is not observable, we can always perform a Kalman decomposition and only consider the observable space.

Since the system is observable, it is well known that the estimation error covariance matrices $P(k)$ and the gain $K(k)$ will converge to

$$P \triangleq \lim_{k\to\infty} P(k), P_+ = APA^T + Q \tag{5}$$

$$K \triangleq P_+C^T(CP_+C^T + R)^{-1}. \tag{6}$$

Since the control system typically will be running for an extended period of time, we can assume that the Kalman filter is at convergence, or equivalently that $\Sigma = P$, and thus the Kalman filter reduces to the following fixed-gain linear estimator:

$$\hat{x}(k+1) = (A - KCA)\hat{x}(k) + Ky(k+1). \tag{7}$$

For reasons that will be clearer soon, we will denote with $K_i$ the $i$th column vector of the matrix $K = [K_1, \ldots, K_m]$. Accordingly, (6) can be rewritten as

$$\hat{x}(k+1) = (A - KCA)\hat{x}(k) + \sum_{i=1}^{m} K_i y_i(k+1). \tag{8}$$

The goal of this paper is to propose an algorithm able to estimate the state in such a way that: 1) if no sensor is compromised, i.e. $a_i(t) = 0$ for all $i$ and for all $t$, the estimate coincides with certain probability to the same estimate obtained using the Kalman filter (7); 2) if less then half of the sensors are compromised, it still gives a stable estimate of the state.

To achieve this goal, two results are presented. In the next section, it is shown that, under mild hypothesis, the estimate of the Kalman can be written as a linear combination of estimates generated by a set of local estimators. Then, in Section IV, a secure fusion scheme is proposed to replace the linear fusion scheme.

## III. Decomposition of Kalman Filter using Local Estimate

In this section, we propose a method to decompose the Kalman estimate (7) into a linear combination of local state estimates. Our goal is to generate $m$ local estimators of the form:

$$\hat{x}_i(k+1) = (A - L_iC_iA)\hat{x}_i(k) + L_iy_i(k+1). \tag{9}$$

and local state estimate $\hat{x}_i(k)$, $i = 1, \ldots, m$, such that:

1) Each local estimator is stable, i.e., $A - L_iC_iA$ is strictly stable;
2) The Kalman estimate $\hat{x}$ can be recovered as a linear combination of $\hat{x}_i(k)$, i.e.,

$$\hat{x}(k) = F_1\hat{x}_1(k) + \ldots + F_m\hat{x}_m(k).$$

To this end, we will make the following assumptions throughout this section:

1) $A$ is invertible.
2) $A - KCA$ has $n$ distinct eigenvalues. Moreover, $A - KCA$ and $A$ do not share any eigenvalue.
3) $(A, C_i)$ is observable for each $C_i$.

**Remark 1.** *Notice that since we assume that $(A, C)$ is observable, then the invertibility of $A$ implies that $(A, CA)$ is also observable. Hence, we can freely assign the poles of $A - KCA$ by choosing a proper gain $K$. As a result, if for the optimal Kalman gain $K$, $A - KCA$ does not satisfy assumption 2, then we can perturb the gain matrix $K$ to enforce the second condition, which will only result in a small estimation performance loss if the perturbation is small.*

**Remark 2.** *The third assumption is quite strong but, as we will show in Section V, it is possible to relax it. However, in order to cope with the space limits, we have chosen to keep this assumption throughout Section III and IV.*

Since $A - KCA$ has distinct eigenvalues, it can be diagonalized as:

$$A - KCA = V\Lambda V^{-1}. \tag{10}$$

As a result, we can rewrite (8) as

$$\left[V^{-1}\hat{x}(k+1)\right] = \Lambda\left[V^{-1}\hat{x}_k\right] + \sum_{i=1}^{m} V^{-1}K_iy_i(k+1). \tag{11}$$

For sensor $i$, since $(A, C_i)$ is observable [22], we can compute an $L_i$, such that $A - L_iC_iA$ shares the same eigenvalues as $A - KCA$. Now consider the following stable[2] estimator:

$$\hat{x}_i(k+1) = (A - L_iC_iA)\hat{x}_i(k) + L_iy_i(k+1). \tag{12}$$

Similarly we can diagonalize $A - L_iC_iA$ as $A - L_iC_iA = V_i\Lambda V_i^{-1}$, and rewrite (12) as

$$\left[V_i^{-1}\hat{x}_i(k+1)\right] = \Lambda\left[V_i^{-1}\hat{x}_i(k)\right] + V_i^{-1}L_iy_i(k+1). \tag{13}$$

The following lemma characterizes the vector $V_i^{-1}L_i$,

**Lemma 1.** *Suppose that $A - KCA$ and $A$ do not share any eigenvalue, then all entries of vector $V_i^{-1}L_i$ is non-zero.*

*Proof.* The proof is reported in the appendix for the sake of legibility. □

Since $V_i^{-1}L_i$ does not contain zero entries, we can find a diagonal matrix $\Lambda_i$ such that

$$V^{-1}K_i = \Lambda_iV_i^{-1}L_i. \tag{14}$$

As a result, if we multiply the LHS and RHS of (13) by $\Lambda_i$, we get

$$\left[\Lambda_iV_i^{-1}\hat{x}_i(k+1)\right] = \Lambda\left[\Lambda_iV_i^{-1}\hat{x}_i(k)\right] + V^{-1}K_iy_i(k+1), \tag{15}$$

where we use the fact that $\Lambda$ and $\Lambda_i$ are commutative since they are both diagonal. Hence, if we sum (15) for all $i = 1, \ldots, m$ and compare it with (11), we can conclude that

$$\hat{x}(k) = \sum_{i=1}^{m} F_i\hat{x}_i(k), \tag{16}$$

where $F_i = V\Lambda_iV_i^{-1}$. We will call (16) as a linear fusion scheme, since the Kalman estimate is recovered as a linear

[2]Since $A - KCA$ is stable, $A - L_iC_iA$ will also be stable since they share the same eigenvalues.

combination of local estimates. The following lemma characterizes a very interesting property of $F_i$ matrices.

**Lemma 2.** *Suppose that $A$ and $A - KCA$ do not share eigenvalues, then the $F_i$ matrices satisfy the following equation:*

$$\sum_{i=1}^{m} F_i = I. \tag{17}$$

*Proof.* The proof is reported in the appendix for the sake of legibility. □

**Remark 3.** *It is worth noticing that we do not necessarily need to implement the local estimator on the sensor side, since we can let our centralized estimator implement the filter equation* (12) *for each sensor and then combine them via* (16)*. However, depending on the application, it may be advantageous to implement the local estimator on the sensor side to reduce the computational burden of the central estimator.*

### *A. A least square interpretation for* (16)

In this subsection, we show that the linear fusion scheme (16) can be interpreted as a a least square problem, which will be used later to derive a secure fusion scheme. To this end, let us define the state estimation error of the $i$th local estimation as $e_i(k)$, i.e.,

$$e_i(k) = x(k) - \hat{x}_i(k),$$

which satisfies the following recursive equation:

$$\begin{aligned} e_i(k+1) = (A - L_iC_iA)e_i(k) + (I - L_iC_i)w(k) \\ - L_iv_i(k) - L_ia_i(k). \end{aligned} \tag{18}$$

Let us define $\mu_i(k), \nu_i(k)$ as follows:

$$\begin{aligned} \mu_i(k+1) &= (A - L_iC_iA)\mu_i(k) + (I - L_iC_i)w(k) - L_iv_i(k), \\ \nu_i(k+1) &= (A - L_iC_iA)\nu_i(k) - L_ia_i(k). \end{aligned} \tag{19}$$

One can view $\mu_i(k)$ as the error of the local estimate caused by noise and $\nu_i(k)$ as the error caused by the bias injected by the adversary. By linearity,

$$e_i(k) = \mu_i(k) + \nu_i(k).$$

Let us define $\tilde{\mu}(k) \in \mathbb{R}^{mn}, \tilde{A} \in \mathbb{R}^{mn \times mn}$ as

$$\tilde{\mu}(k) \triangleq \begin{bmatrix} \mu_1(k) \\ \vdots \\ \mu_m(k) \end{bmatrix}, \tilde{A} \triangleq \begin{bmatrix} A - L_1C_1A & & \\ & \ddots & \\ & & A - L_mC_mA \end{bmatrix}. \tag{20}$$

Similarly we can define $\tilde{e}(k)$ and $\tilde{\nu}(k)$ by stacking $e_i(k)$ and $\nu_i(k)$ as a big vector respectively.

It is easy to prove that the covariances of the following vectors is

$$\begin{aligned} &\operatorname{Cov}\left(\begin{bmatrix} I - L_1C_1 \\ \vdots \\ I - L_mC_m \end{bmatrix} w(k)\right) \\ &= \begin{bmatrix} I - L_1C_1 \\ \vdots \\ I - L_mC_m \end{bmatrix} Q \begin{bmatrix} I - L_1C_1 \\ \vdots \\ I - L_mC_m \end{bmatrix}^T, \end{aligned} \tag{21}$$

and

$$\operatorname{Cov}\left(\begin{bmatrix} L_1v_1(k) \\ \vdots \\ L_mv_m(k) \end{bmatrix}\right) = \left(\begin{bmatrix} L_1 \\ \vdots \\ L_m \end{bmatrix} \begin{bmatrix} L_1 \\ \vdots \\ L_m \end{bmatrix}^T\right) \circ (R \otimes \mathbf{1}_{n\times n}), \tag{22}$$

where $\circ$ denotes element-wise matrix multiplication, $\otimes$ is the Kronecker product and $\mathbf{1}_{n\times n}$ is an all one matrix of size $n \times n$. Now let us define $\tilde{Q} \in \mathbb{R}^{mn\times mn}$ to be

$$\tilde{Q} \triangleq \operatorname{Cov}\left(\begin{bmatrix} I - L_1C_1 \\ \vdots \\ I - L_mC_m \end{bmatrix} w(k)\right) + \operatorname{Cov}\left(\begin{bmatrix} L_1v_1(k) \\ \vdots \\ L_mv_m(k) \end{bmatrix}\right). \tag{23}$$

As a result, we know that $\tilde{\mu}_i(k)$ will be Gaussian distributed and its covariance satisfies the following Lyapunov equation:

$$\operatorname{Cov}\tilde{\mu}(k+1) = \tilde{A}\operatorname{Cov}\tilde{\mu}(k)\tilde{A}^T + \tilde{Q}. \tag{24}$$

Finally, define $\tilde{W}$ as the fix point[3] of (24), i.e.,

$$\tilde{W} = \tilde{A}\tilde{W}\tilde{A}^T + \tilde{Q}. \tag{25}$$

Consider the following optimization problem:

$$\begin{aligned} \underset{\check{x}(k),\check{e}(k)}{\text{minimize}} \quad & \frac{1}{2}\check{e}(k)^T\tilde{W}^{-1}\check{e}(k) \\ \text{subject to} \quad & \begin{bmatrix} \hat{x}_1(k) \\ \vdots \\ \hat{x}_m(k) \end{bmatrix} = H\check{x}(k) + \check{e}(k), \end{aligned} \tag{26}$$

where $H \triangleq \begin{bmatrix} I & \cdots & I \end{bmatrix}^T \in \mathbb{R}^{mn\times n}$. This problem can be interpreted as the problem of finding an estimate $\check{x}(k)$ that minimizes a weighted least square of the error with the local estimates $\hat{x}_i(k)$, where the weighting matrix is related with the covariance of the error of the local estimates.

The following theorem, establishes the connection between the linear fusion scheme (16) and the least-square problem (26).

**Theorem 1.** *The solution of the least-square problem* (26) *is given by*

$$\begin{aligned} \check{x}(k) &= \hat{x}(k) = \sum_{i=1}^{m} F_i\hat{x}_i(k), \\ \check{e}(k) &= (I - H\begin{bmatrix} F_1 & \dots & F_m \end{bmatrix})\tilde{e}(k). \end{aligned}$$

[3] $\tilde{W}$ is well defined since all $A - L_iC_iA$ matrices are strictly stable

*Proof.* The proof is reported in the appendix for the sake of legibility. □

**Remark 4.** *Notice that the results and the proofs presented in this section are purely algebraic. Therefore, the result can be easily generalized to other linear fixed-gain estimators and other noise models.*

It is worth noticing that the linear fusion scheme (16) (or equivalently the least-square problem (26)) is not secure in the sense that if sensor $i$ is compromised, then the adversary can manipulate $\hat{x}_i(k)$ by injecting the bias $a_i(k)$ into the measurements $y_i(k)$. Therefore, the adversary can potentially change the Kalman estimate arbitrarily. In the next section, to address the security challenges, we modify (26) by adding an $\ell_1$ penalty to ensure the stability of the state estimation in the presence of malicious sensors.

## IV. Secure Information Fusion

In this section, we propose a secure way to compute the state estimation based on the local estimations.

Notice that the error $e_i(k)$ can be decomposed as the error caused by the noise $\mu_i(k)$ and the error caused by the bias injected by the adversary $\nu_i(k)$. As a result, we propose the following secure fusion scheme based on LASSO [23]:

$$\begin{aligned} \underset{\check{x}_s(k),\check{\mu}(k),\check{\nu}(k)}{\text{minimize}} \quad & \frac{1}{2}\check{\mu}(k)^T\tilde{W}^{-1}\check{\mu}(k)+\gamma\|\check{\nu}(k)\|_1 \qquad (27)\\ \text{subject to} \quad & \hat{x}_i(k)=\check{x}_s(k)+\check{\mu}_i(k)+\check{\nu}_i(k),\ \forall i, \end{aligned}$$

where $\check{x}_s(k)$ is the secure state estimation. $\gamma$ is a constant chosen by the system operator. $\check{\mu}(k)$, $\check{\nu}(k)$ are defined as:

$$\check{\mu}(k)\triangleq\begin{bmatrix}\check{\mu}_1(k)\\ \vdots\\ \check{\mu}_m(k)\end{bmatrix},\ \check{\nu}(k)\triangleq\begin{bmatrix}\check{\nu}_1(k)\\ \vdots\\ \check{\nu}_m(k)\end{bmatrix}.$$

We now have the following lemma characterizing the solution of the optimization problem:

**Lemma 3.** *Let $\check{x}_s(k)$, $\check{\mu}(k)$, $\check{\nu}(k)$ be the minimizer for the optimization problem* (27)*. Let $\check{x}(k),\check{e}(k)$ be the minimizer for the least-square problem* (26)*. Then the following statements hold:*

1) *The following inequality holds:*

$$\|\tilde{W}^{-1}\check{\mu}(k)\|_\infty\le\gamma. \qquad (28)$$

2) *If $\|\tilde{W}^{-1}\check{e}(k)\|_\infty\le\gamma$, then*

$$\check{x}_s(k)=\check{x}(k)=\hat{x}(k),\ \check{\mu}(k)=\check{e}(k),\ \check{\nu}(k)=0.$$

*Proof.* We will first prove (28). Assume the opposite, i.e.,

$$\|\tilde{W}^{-1}\check{\mu}(k)\|_\infty>\gamma.$$

Therefore, we can find a vector $\zeta$, such that

$$\zeta^T\tilde{W}^{-1}\check{\mu}(k)>\gamma,$$

with $\|\zeta\|_1=1$. Clearly, for any $\alpha>0$, $\check{x}_s(k),\check{\mu}(k)-\alpha\zeta$ and $\check{\nu}(k)+\alpha\zeta$ will also be a feasible solution for (27). The corresponding cost function can be calculated as

$$\begin{aligned}&\frac{1}{2}\left(\check{\mu}(k)-\alpha\zeta\right)^T\tilde{W}^{-1}\left(\check{\mu}(k)-\alpha\zeta\right)+\gamma\|\check{\nu}(k)+\alpha\zeta\|_1\\ &\le\frac{1}{2}\check{\mu}(k)^T\tilde{W}^{-1}\check{\mu}(k)+\gamma\|\check{\nu}(k)\|_1\\ &+\alpha\left(\gamma\|\zeta\|_1-\zeta^T\tilde{W}^{-1}\check{\mu}(k)\right)+\frac{1}{2}\alpha^2\zeta^T\tilde{W}^{-1}\zeta.\end{aligned}$$

Notice that

$$\gamma\|\zeta\|_1-\zeta^T\tilde{W}^{-1}\check{\mu}(k)<0.$$

Hence, for small enough $\alpha$, we have

$$\begin{aligned}&\frac{1}{2}\left(\check{\mu}(k)-\alpha\zeta\right)^T\tilde{W}^{-1}\left(\check{\mu}(k)-\alpha\zeta\right)+\gamma\|\check{\nu}(k)+\alpha\zeta\|_1\\ &<\frac{1}{2}\check{\mu}(k)^T\tilde{W}^{-1}\check{\mu}(k)+\gamma\|\check{\nu}(k)\|_1,\end{aligned}$$

which contradicts with the optimality of $\check{x}_s(k)$, $\check{\mu}(k)$ and $\check{\nu}(k)$. Therefore, (28) must hold.

The second statement can be proved using KKT conditions and the detailed proof is omitted due to space limit. □

We now consider two scenarios:

1) All sensors are benign and the system is operating normally.
2) $p$ sensors are compromised.

The following two theorems characterize the performance of the secure fusion scheme (27) for each scenario:

**Theorem 2.** *Suppose that all the sensors are benign, i.e., $a(k)=0$ for all $k$. The secure state estimate $\check{x}_s(k)$ equals the optimal Kalman estimate $\hat{x}(k)$ if the following inequality holds:*

$$\left\|\bar{W}^{-1}(I-H\begin{bmatrix}F_1 & \cdots & F_m\end{bmatrix})\tilde{e}(k)\right\|_\infty\le\gamma. \qquad (29)$$

*Proof.* This theorem is a direct consequence of Lemma 3 and Theorem 1. □

**Remark 5.** *If all sensors are benign, the local estimation error $e_i(k)$ will be zero mean Gaussian distributed and*

$$\lim_{k\to\infty}\mathrm{Cov}(\tilde{e}(k))=\tilde{W}.$$

*If the system is operating long enough, we have $\mathrm{Cov}(\tilde{e}(k))\approx\tilde{W}$ and we can compute the probability that the inequality* (29) *holds, i.e., the probability that the secure state estimate equals to the optimal Kalman estimate.*

We now consider the second scenario, where $p$ sensors are compromised. Before stating the main theorem, let us define the following operator: $f_i: R\times R\times\cdots\times R\to R$, such that $f_i(\beta_1,\ldots,\beta_m)$ equals to the $i$th smallest element in the set $\{\beta_1,\ldots,\beta_m\}$. Assuming that $e_1,\ldots,e_m\in\mathbb{R}^n$ are vectors. With slightly abuse of notations, we define $f_i(e_1,\ldots,e_m)$ as a vector where each of its entry is the $i$th smallest element among the corresponding entries in $e_1,\ldots,e_m$. We further define $f_{i+1/2}=(f_i+f_{i+1})/2$.

**Theorem 3.** *Suppose that $p < m/2$ sensors are compromised, then the error of the secure state estimate is bounded by*

$$f_{(m+1)/2-p}(\mu_1(k),\dots,\mu_m(k)) - \frac{\gamma}{\|\tilde{W}^{-1}\|_\infty} \le x(k) - \check{x}(k) \le f_{(m+1)/2+p}(\mu_1(k),\dots,\mu_m(k)) + \frac{\gamma}{\|\tilde{W}^{-1}\|_\infty}, \quad (30)$$

*where $\|\tilde{W}^{-1}\|_\infty$ is the induced infinity norm of $\tilde{W}^{-1}$.*

*Proof.* Due to the space limit, only the sketch of the proof is provided. First, one can prove that $\check{x}(k)$ and $\check{\mu}(k)$ must satisfies the following relationship:

$$\check{x}(k) = f_{\frac{m+1}{2}}(\hat{x}_1(k) - \check{\mu}_1(k),\dots,\hat{x}_m(k) - \check{\mu}_m(k)). \quad (31)$$

We know that $\hat{x}_1(k) = x(k) - \mu_i(k) - \nu_i(k)$ and $\|\check{\mu}(k)\|_\infty \le \gamma/\|\tilde{W}\|_\infty$. Furthermore, since at most $p$ $\nu_i(k)$ are non-zero, we have

$$\begin{aligned} &f_{(m+1)/2-p}(\mu_1(k),\dots,\mu_m(k)) \\ &\le f_{(m+1)/2}(\mu_1(k)+\nu_1(k),\dots,\mu_m(k)+\nu_m(k)) \\ &\le f_{(m+1)/2+p}(\mu_1(k),\dots,\mu_m(k)). \end{aligned}$$

Combining with (31), we can prove (30). □

**Remark 6.** *Notice that by Theorem 2, increasing $\gamma$ will increase the likelihood that the secure estimation equals the Kalman estimation during normal operation. On the other hand, by Theorem 3, a large $\gamma$ can potentially result in a large estimation error when the system is under attack.*

## V. Extension

In this section, we relax the assumption that the system is observable for every single sensor. Assume that the system is not fully observable by the $i$th sensor, i.e., $(A, C_i)$ is not observable. Consider the following recursive equation:

$$\hat{\xi}_i(k+1) = \Lambda\hat{\xi}_i(k) + \mathbf{1}_n y_i(k), \quad (32)$$

where $\mathbf{1}_n \in \mathbb{R}^{n\times 1}$ is an all-one vector and $\Lambda$ is defined in (10). Let us define the matrix $G_i$ as

$$G_i \triangleq \begin{bmatrix} C_iA(A-\lambda_1 I)^{-1} \\ \vdots \\ C_iA(A-\lambda_n I)^{-1} \end{bmatrix},$$

where $\lambda_i$ are the $i$th eigenvalues of $\Lambda$. Notice that the inverse of $A-\lambda_i I$ is well defined since $A$ does not share eigenvalues with $\Lambda$. The following theorem establishes the connection between $\hat{\xi}_i(k)$ and the state $x(k)$.

**Theorem 4.** *Let $\epsilon_i(k) \triangleq G_i x(k) - \hat{\xi}_i(k)$, then*

$$\epsilon_i(k+1) = \Lambda\epsilon_i(k) + (G_i - \mathbf{1}_n C_i)w(k) - \mathbf{1}_n v_i(k+1).$$

*In other words, $\hat{\xi}_i(k)$ is a stable estimate of $G_i x(k)$.*

*Proof.* By definition, we have

$$\begin{aligned} \epsilon_i(k+1) &= G_i x(k+1) - \hat{\xi}_i(k+1) \\ &= G_i Ax(k) + G_i w(k) - \Lambda\hat{\xi}_i(k) \\ &\quad - \mathbf{1}_n(C_iAx(k) + C_iw(k) + v_i(k+1)) \\ &= (G_iA - \mathbf{1}_nC_iA)x(k) - \Lambda\hat{x}i_i(k) \\ &\quad + (G_i - \mathbf{1}_nC_i)w(k) - \mathbf{1}_nv_i(k+1). \end{aligned}$$

By the definition of $G_i$, one can prove that $G_iA - \mathbf{1}_nC_iA = \Lambda G_i$. Therefore,

$$\begin{aligned} \epsilon_i(k+1) &= \Lambda G_i x(k) - \Lambda\hat{\xi}_i(k) \\ &\quad + (G_i - \mathbf{1}_nC_i)w(k) - \mathbf{1}_nv_i(k+1), \end{aligned}$$

which concludes the proof. □

**Remark 7.** *It is worth noticing that if $(A, C_i)$ is not fully observable, then $G_i$ will not be full rank. In fact, by Cayley-Hamilton theorem $(A-\lambda_i I)^{-1}$ can be written as a linear combination of $I$, $A,\dots,A^{n-1}$. As a result, each row vector of $G_i$ will belong to the observable space of $(A, C_i)$.*

Now we can choose $F_i$ as

$$F_i = V\,\mathrm{diag}(V^{-1}K_i),$$

where $V$ is defined in (10) and $\mathrm{diag}(V^{-1}K_i)$ is an $n\times n$ diagonal matrix with the $j$th diagonal entry equals to the $j$th entry of the vector $V^{-1}K_i$. Comparing (32) and (11), we can prove that:

$$\hat{x}(k) = \sum_{i=1}^{m} F_i\hat{\xi}_i(k). \quad (33)$$

We can further rewrite the linear fusion scheme (33) as the solution of a least-square problem and create a secure fusion scheme by adding the $\ell_1$ penalty term similar to (27). The detailed derivation is omitted due to space limit.

## VI. Numerical Example

In this section, we demonstrate our proposed secure estimation via a numerical example. We assume the following parameters for our system:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},\ C = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix},\ Q = I,\ R = I.$$

The optimal steady state Kalman gain $K$ and estimation covariance $P$ matrices are given by

$$K = \begin{bmatrix} 0.223 & 0.399 & 0.135 \\ 0.083 & -0.259 & 0.253 \end{bmatrix}, P = \begin{bmatrix} 0.311 & -0.088 \\ -0.088 & 0.171 \end{bmatrix}.$$

The corresponding $A-KCA$ matrix has eigenvalues at $0.2242$ and $-0.1324$. As a result, we can derive the $L_i$ matrices to ensure that $A - L_iC_iA$ shares the same eigenvalues with $A-KCA$ as:

$$L_1 = \begin{bmatrix} 0.4393 \\ 0.5310 \end{bmatrix},\ L_2 = \begin{bmatrix} 0.4393 \\ -0.5310 \end{bmatrix},\ L_3 = \begin{bmatrix} 0.4393 \\ 0.2655 \end{bmatrix}.$$

We consider two scenarios:

1) all sensors are benign;

2) the first sensor is under the attack and $a_1(k) = 100$ for all $k$.

We compute the empirical Mean Squared Error (MSE) of the secure estimator for each scenarios and for different choices of $\gamma$. Notice that when all sensors are benign, the optimal Kalman estimator has an MSE equals to $\mathrm{tr}(P) = 0.482$. As a result, we define the relative MSE as the MSE divided by $0.482$. Fig 4 illustrates the relative MSE of the proposed secure estimator versus $\gamma$. It can be seen that when $\gamma \geq 2$, the secure estimator achieves roughly the same estimation performance as the optimal Kalman estimator under normal operation. On the other hand, if sensor 1 is malicious, then the MSE achieves the minimum at around $\gamma = 1.8$.

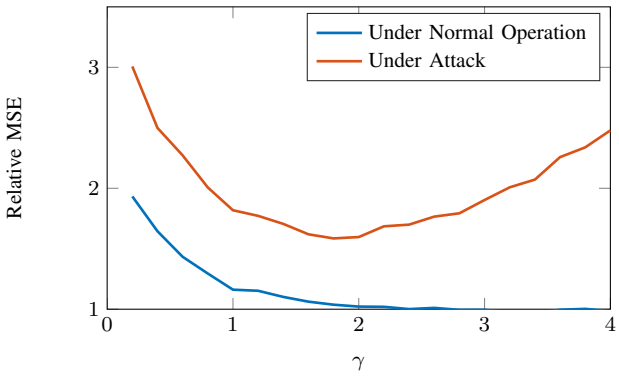


Fig. 4. The relative MSE of the secure estimator v.s. different choices of $\gamma$. The blue line indicates the relative MSE when all sensors are benign, while the red line indicates the relative MSE when sensor 1 is malicious.

## VII. Conclusion

We consider the problem of estimating the state of a linear time-invariant Gaussian system using $m$ sensors, where some of the sensors can potentially be compromised by an adversary. Under mild assumptions, we prove that we can decompose the optimal Kalman estimate as a weighted sum of local state estimates. We then propose a convex optimization based approach to combine the local estimate into a more secure state estimate. Numerical example illustrates that our secure estimator achieves good performance under both normal operation and attack scenarios.

## Appendix I
## Proof of Lemma 1

*Proof of Lemma 1.* Notice that we can write $V_i^{-1}$ as

$$V_i^{-1} = \begin{bmatrix} \zeta_1 & \cdots & \zeta_n^T \end{bmatrix}^T.$$

where each $\zeta_j$ is a left eigenvector of $A - L_iC_iA$, i.e., $\zeta_j^T(A - L_iC_iA) = \lambda_j\zeta_j^T$. Suppose that the $j$th entry of $V_i^{-1}L_i$ is zero, which implies that $\zeta_j^T L_i = 0$. Therefore, we have

$$\lambda_j\zeta_j^T = \zeta_j^T(A - L_iC_iA) = \zeta_j^T A - \zeta_j^T L_iC_iA = \zeta_j^T A,$$

which indicates that $\zeta_j$ is also a left eigenvector of $A$ with eigenvalue $\lambda_j$. However, this is impossible since we assume that $A$ does not share any eigenvalue with $A - KCA$ (or $A - L_iC_iA$). □

## Appendix II
## Proof of Lemma 2

*Proof of Lemma 2.* Since $F_i = V\Lambda_i V_i^{-1}$, to prove that $\sum_{i=1}^m F_i = I$, we only need to prove

$$S \triangleq \sum_{i=1}^m \Lambda_i V_i^{-1} = V^{-1}. \tag{34}$$

By (14), $K_i = V\Lambda_i V_i^{-1}L_i$. Therefore,

$$A - KCA = A - \sum_{i=1}^m K_iC_iA = A - V\sum_{i=1}^m \Lambda_i V_i^{-1}L_iC_iA. \tag{35}$$

On the other hand, since $A - L_iC_iA = V_i\Lambda V_i^{-1}$, we know that

$$\Lambda_i V_i^{-1}C_iA = \Lambda_i V_i^{-1}A - \Lambda\Lambda_i V_i^{-1}, \tag{36}$$

where we use the fact that $\Lambda_i$ and $\Lambda$ are commutative. Combining (35) and (36), we have

$$A - KCA = A - VSA + V\Lambda S. \tag{37}$$

By (10), we know that $A - KCA = V\Lambda V^{-1}$. Hence,

$$\begin{aligned} 0 &= A - VSA + V\Lambda S - V\Lambda V^{-1} \\ &= V\left[(V^{-1} - S)A - \Lambda(V^{-1} - S)\right]. \end{aligned}$$

Since $V$ is invertible, we have $(V^{-1} - S)A = \Lambda(V^{-1} - S)$, which implies that the $j$th row vector $\zeta_j^T$ of $V^{-1} - S$ satisfies

$$\zeta_j^T A = \lambda_j\zeta_j^T.$$

However, $A$ does not share eigenvalue with $\Lambda$. Therefore, all row vectors $\zeta_j$ must be 0, which proves that $V^{-1} = S$. □

## Appendix III
## Proof of Theorem 1

Before proving the theorem we will prove the following lemma that will be used in the proof of Theorem 1.

**Lemma 4.** *Let $K$ be the steady state Kalman gain defined in* (6)*. For any $L$, such that $A - LCA$ is strictly stable, the following Lyapunov equation holds:*

$$\begin{aligned} P &= (A - KCA)P(A - LCA)^T \\ &\quad + (I - KC)Q(I - LC)^T + KRL^T, \end{aligned} \tag{38}$$

*where $P$ is defined in* (5)*.*

*Proof.* Let us rewrite the RHS of (38) as

$$\begin{aligned} \text{RHS} &= (A - KCA)PA^T + (I - KC)Q \\ &\quad + \left[KR - (A - KCA)PA^TC^T - (I - KC)QC^T\right]L^T \end{aligned}$$

Thus, Lemma 4 is equivalent to

$$P = (A - KCA)PA^T + (I - KC)Q,$$

and

$$0 = KR - (A - KCA)PA^TC^T - (I - KC)QC^T,$$

which can be proved using the definition of $K$ and $P$ matrices. □

At this point we can prove Theorem 1:

*Proof of Theorem 1.* Let us first rewrite $\tilde{W}$ matrix in a block diagonal form:

$$\tilde{W} = \begin{bmatrix} \tilde{W}_{11} & \cdots & \tilde{W}_{1m} \\ \vdots & \ddots & \vdots \\ \tilde{W}_{m1} & \cdots & \tilde{W}_{mm} \end{bmatrix},$$

where each $\tilde{W}_{ij} \in \mathbb{R}^{n\times n}$. As a result, by (25), we know that $\tilde{W}_{ij}$ satisfies:

$$\begin{aligned}\tilde{W}_{ij} &= (A - L_iC_iA)\tilde{W}_{ij}(A - L_jC_jA)^T \\ &\quad + (I - L_iC_i)Q(I - L_jC_j)^T + r_{ij}L_iL_j^T,\end{aligned}$$

where $r_{ij}$ is the element of the matrix $R$ on $i$th row and $j$th column. Since $F_i(A-L_iC_iA)F_i^{-1} = A-KCA$ and $F_iL_i = K_i$, it is easy to prove the following recursive equation holds:

$$\begin{aligned}F_i\tilde{W}_{ij} &= (A - KCA)F_i\tilde{W}_{ij}(A - L_jC_jA)^T \\ &\quad + (F_i - K_iC_i)Q(I - L_jC_j)^T + r_{ij}K_iL_j^T,\end{aligned}$$

Therefore, let $\tilde{S}_j = \sum_{i=1}^m F_i\tilde{W}_{i,j}$, we can conclude that $\tilde{S}_j$ satisfies the following recursive equation

$$\tilde{S}_j = (A - KCA)\tilde{S}_j(A - L_jC_jA)^T \tag{39}$$

$$+ \sum_{i=1}^m (F_i - K_iC_i)Q(I - L_jC_j)^T + \sum_{i=1}^m r_{ij}K_iL_j^T. \tag{40}$$

By Lemma 2, we know that $\sum_{i=1}^m (F_i - K_iC_i) = I - KC$. Furthermore, define matrix $\mathcal{L}_j \in \mathbb{R}^{n\times m}$ as an all zero matrix except the $j$th column to be $L_j$, i.e.,

$$\mathcal{L}_j \triangleq \begin{bmatrix}0 & \dots & 0 & L_j & 0 \dots & 0\end{bmatrix}$$

One can prove that

$$L_jC_j = \mathcal{L}_jC,\ \sum_{i=1}^m r_{ij}K_iL_j^T = KR\mathcal{L}_j^T.$$

As a result, (40) can be simplified as

$$\begin{aligned}\tilde{S}_j &= (A - KCA)\tilde{S}_j(A - \mathcal{L}_jCA)^T \\ &\quad + (I - KC)Q(I - \mathcal{L}_jC)^T + KR\mathcal{L}_j^T.\end{aligned}$$

Hence, by Lemma 4, $\tilde{S}_j = P$ for all $j = 1, \dots, m$, which implies that

$$\begin{bmatrix}F_1 & \cdots & F_m\end{bmatrix}\tilde{W} = PH^T. \tag{41}$$

On the other hand, it is easy to show that the optimal solution of (26) is given by

$$\check{x}(k) = (H^T\tilde{W}^{-1}H)^{-1}H^T\tilde{W}^{-1}\begin{bmatrix}\hat{x}_1(k) \\ \vdots \\ \hat{x}_m(k)\end{bmatrix}.$$

By (41),

$$H^T\tilde{W}^{-1} = P^{-1}\begin{bmatrix}F_1 & \cdots & F_m\end{bmatrix},$$

$$H^T\tilde{W}^{-1}H = P^{-1}\sum_{i=1}^m F_i = P^{-1}.$$

Therefore, $\check{x}(k) = \sum_{i=1}^m F_i\hat{x}_i(k) = \hat{x}(k)$. Similarly, one can prove the relationship between $\check{e}(k)$ and $\tilde{e}(k)$. □

## REFERENCES


[1] T. M. Chen, "Stuxnet, the real start of cyber warfare? [editor's note]," *IEEE Network*, vol. 24, no. 6, pp. 2–3, 2010.

[2] D. P. Fidler, "Was stuxnet an act of war? decoding a cyberattack," *IEEE Security & Privacy*, vol. 9, no. 4, pp. 56–59, 2011.

[3] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proc. Conf. Hot Topics in Security*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1–6.

[4] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[5] Y. Liu, M. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. Computer and Commun. Security*, 2009.

[6] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems*, 2010.

[7] J. Kim, L. Tong, and R. J. Thomas, "Data framing attack on state estimation," *IEEE J. Selected Areas in Commun.*, vol. 32, no. 7, pp. 1460–1470, Jul. 2014.

[8] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[9] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: a system theoretic approach," *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 90–104, Jan 2010.

[10] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[11] S. Kassam, H. V. Poor *et al.*, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, 1985.

[12] R. A. Maronna, D. R. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. NJ: Wiley, 2006.

[13] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. NJ:Wiely, 2009.

[14] F. R. Hampel, "A general qualitative definition of robustness," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887–1896, 1971.

[15] D. L. Donoho and P. J. Huber, "The notion of breakdown point," *A Festschrift for Erich L. Lehmann*, pp. 157–184, 1983.

[16] F. R. Hampel, "The influence curve and its role in robust estimation," *J. the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[17] Y. Mo and B. Sinopoli, "Secure estimation in the presence of integrity attacks," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1145–1151, April 2015.

[18] ——, "False data injection attacks in cyber physical systems," in *First Workshop on Secure Control Systems*, 2010.

[19] ——, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Transactions on Automatic Control*, vol. PP, no. 99, pp. 1–1, 2015.

[20] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *Proc. ACM/IEEE Int. Conf. Cyber-Physical Systems*, Apr. 2014, pp. 163–174.

[21] M. Pajic, P. Tabuada, I. Lee, and G. J. Pappas, "Attack-resilient state estimation in the presence of noise," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 5827–5832.

[22] J. Kautsky, N. K. Nichols, and P. Van Dooren, "Robust pole assignment in linear state feedback," *International Journal of Control*, vol. 41, no. 5, pp. 1129–1155, 1985.

[23] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.