

Secure Estimation based Kalman Filter for Cyber-Physical Systems against Adversarial Attacks

Young Hwan Chang*, Qie Hu*, Claire J. Tomlin

Abstract—Cyber-physical systems (CPSs) are found in many applications such as power networks, manufacturing processes, and air and ground transportation systems. Maintaining security of these systems under cyber attacks is an important and challenging task, since these attacks can be erratic and thus difficult to model. Secure estimation problems study how to estimate the true system states when measurements are corrupted and/or control inputs are compromised by attackers. The authors in [1] proposed a secure estimation method when the set of attacked nodes (sensors, controllers) is fixed. In this paper, we extend these results to scenarios in which the set of attacked nodes can change over time. We formulate this secure estimation problem into the classical error correction problem [2] and we show that accurate decoding can be guaranteed under a certain condition. Furthermore, we propose a combined secure estimation method with our proposed secure estimator above and the Kalman Filter (KF) for improved practical performance. Finally, we demonstrate the performance of our method through simulations of two scenarios where an unmanned aerial vehicle is under adversarial attack.

Index Terms—Cyber-physical systems, Error correction, Secure estimation

I. INTRODUCTION

Cyber-physical systems (CPSs) consist of physical components such as actuators, sensors and controllers that communicate with each other over a network. For example, we consider a scenario in which a group of unmanned aerial vehicles (UAV) is flying in a formation, and each UAV continuously sends information such as its position to other vehicles wirelessly. Or we can consider each UAV continuously sends information to a remote control center. Although communication networks are often protected by security measures, cyber attacks could still take place when a malicious attacker obtains unauthorized knowledge of the access key. The attacker can either jam a communication link preventing information from being sent or received [3], or it can spoof sensor readings and send erroneous control signals to actuators [4]. For CPS systems, cyber attacks not only compromise information but can also cause damage in the physical process. This presents new challenges and thus demands new strategies and algorithms [5].

Researchers have studied the problem of CPS under attack from different points of view. Each work for security problems

of CPS relies on specific assumptions about attack model and it is rarely the case that one estimator/detector can protect against all possible attacks. From the attacker's point of view, there has been work on how to design optimal attack signals for different control systems and applications [6]–[9]. From the controller's point of view, researchers have studied how to detect attacks [10], [11] and how to accurately estimate the states and control the system when it is under attack. One approach to design secure estimation and control algorithms for a CPS-under-attack is to model the attack signal as process or measurement noise. This approach is adopted in robust control and filtering methods. For example, H_2 and H_∞ controllers are designed so that the system continues to function properly when it is subjected to bounded model uncertainty or disturbance [12]. Similarly, stochastic controllers [13] and filtering methods such as residual filters [14] and the Kalman Filter (KF) [4], [15] model the attack signal as noise or disturbance that follows a certain probabilistic distribution. Therefore these controllers may fail to detect attacks that are poorly modeled by a noise process, such as adversarial attacks.

As a result, researchers formulated algorithms that treat attack signals more explicitly. In game theory, for example, the controller and attacker are players with competing goals in a game, where the attacker is trying to maximize some cost of loss and the controller is trying to minimize it [16]–[20]. This formulation requires assumptions about the attacker's possible strategies. Nevertheless, attacks are usually erratic and the assumption that they can be described by a specific model or that the controller is aware of such a model may be unjustified. In [13], the authors studied a hybrid controller consisting of different controllers, each designed to protect against a specific type of attack signal.

More recently, there has been work on secure estimation where attack signals can be arbitrary and unbounded, but the set of attacked nodes is fixed [1]. In [1], the authors studied secure estimation of a linear time invariant system when the set of attacked nodes does not change over time, and they proposed a novel method by formulating the system under attack as an estimation problem. They also showed how one can increase the number of correctable errors (i.e., improve the system's resilience against attacks). Similarly, in [21], the authors extended the work in [1] by relaxing the assumption of having an exact system model, however they kept the assumption of a fixed set of attacked nodes. Furthermore, in [7], the authors studied attacks on both sensors and actuators of a stochastic linear dynamical system. They assumed that the system used statistical analysis on the residual between actual measurements and those predicted by a KF to determine

Y. H. Chang is with the Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR 97201 USA (e-mail: chanyo@ohsu.edu).

Q. Hu, C.J Tomlin are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: {qiehu, tomlin}@eecs.berkeley.edu).

*These authors contributed equally

whether it was under attack. The authors concluded that attacks on actuators alone can only cause bounded estimation error without being detected; whereas sensor attacks can cause unbounded estimation error, eventually driving it to infinity as time goes to infinity, while being undetected.

In this paper, we are interested in the case in which the set of attacked nodes can change over time. We focus on secure estimation and control of systems under sensor attack, because this type of attack is relatively easy to perform and thus particularly interesting. Take UAVs for example: actuators are installed onboard with hardwired local feedback loops, which are unlikely to be corrupted by adversarial attacks. Communications with external sources, on the other hand, are much more vulnerable. This includes GPS position measurements and communications between a UAV and a remote control center for UAV traffic management. An attacker that wants to disrupt the information injects false position measurements (Man-In-The-Middle (MITM) attack [22]). If it is aware that the UAV uses a decoder designed against fixed attacked nodes [1], then it can attack an additional measurement at random from time to time, so that such a decoder would fail. Thus in general, a malicious agent who is aware that the system employs a decoder designed against fixed attacked nodes, can purposely attack different nodes at different time steps so that such a decoder would fail.

A. Contributions

One of the key contributions of our paper is that **we relax the assumption of fixed attacked nodes in previous works** [1] [21], hence allow attacked nodes to change between different time steps. For example, let $e^{(t)}$ denote the attack/error vector at time t . In other words, if node i is not attacked at time t then the i -th component of $e^{(t)}$ is zero, otherwise, it is nonzero. The assumption of fixed attacked nodes in previous works means that the support of $e^{(t)}$ is fixed for $t \in \{0, \dots, T-1\}$. As an illustration, consider a system with two corrupted nodes and construct an error matrix using the error vectors $e^{(t)}$ as

$$(e^{(0)}, e^{(1)}, \dots, e^{(T-1)}) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ * & * & \cdots & * \\ 0 & 0 & \cdots & 0 \\ * & * & \cdots & * \end{bmatrix}$$

where $*$ denotes a nonzero component (i.e., attack or corruption). Observe that the set of nonzero rows corresponds to the set of attacked nodes. **By leveraging this fact, the secure estimator in [1] detects attack signals by minimizing the number of nonzero rows in the above error matrix.**

On the other hand, we study the case in which the set of attacked nodes can change over time:

$$(e^{(0)}, e^{(1)}, \dots, e^{(T-1)}) = \begin{bmatrix} * & 0 & \cdots & * \\ * & * & \cdots & 0 \\ 0 & 0 & \cdots & * \\ 0 & * & \cdots & 0 \end{bmatrix}.$$

Note that the set of nonzero rows no longer corresponds to the set of attacked nodes. **Thus, the row support minimization method used in [1] cannot be applied to detect attack**

signals in our setting. Instead, we propose an l_1 -optimization-based estimator by vertically stacking the error vectors and then, applying the classical error correction technique [2]. Therefore, note that our decoder is different from the one proposed in [1]. We prove that, when the set of attacked nodes can change between different time points, the maximum number of sensor attacks that can be corrected is still the same as when the set of attacked nodes are fixed [1]. This is not a straightforward result, since the assumption of fixed attacked nodes is removed and a different secure estimator is used.

In addition, we show that under a certain condition, our secure estimation problem is equivalent to the classical error correction problem [2]. In order to guarantee accurate decoding in the latter problem, a suitable coding matrix is usually chosen *a priori*. However, in our problem, the coding matrix consists of system matrices and therefore, cannot be chosen arbitrarily. Thus, instead of using the usual condition known as Restricted Isometric Property [2], we provide a more practical way, using pole placement, to guarantee the existence of an accurate decoder in our setting. Furthermore, we show the connection between a secure estimation problem and an error correction problem when $T > 1$ (in [1], the authors only discussed this connection for when $T = 1$). We then use results from [1] and [2] to propose a computationally efficient secure estimator. As in [1], we do not assume that the attack signal follows any model, therefore our method works for any type of attack.

In Theorem 1, we give a theoretical condition for perfect recovery of system states when the number of attacked nodes is less than the maximum allowable limit. Although this condition may not always be satisfied in practice, simulation results show that our proposed estimator would still recover the correct states with high probability. Simulation results further show that the above theoretical condition is often too conservative in practice, and can be relaxed.

Finally, we propose to combine our secure estimator with a KF to further improve its practical performance. The KF does not only filter out occasional estimation errors by the secure estimator, but also noisy measurements. We demonstrate the effective of our combined estimator using two examples of UAVs under adversarial attack.

B. Organization of the Paper

This paper is organized as follows. Section II gives an overview of secure estimation for CPS when attacked nodes are fixed, as well as compressive sensing and error correction. Section III formulates the problem of secure estimation when attacked nodes can change over time and connects it to the case when attacked nodes are fixed. Section IV explains our proposed secure estimation algorithm and assesses its practical performance using a numerical example. We also describe how to combine it with a KF to improve its performance in practice. Finally we present two more realistic numerical examples of UAVs subject to adversarial attack in Section V. In this last section, we also characterize how pole placement can be used to trade-off control and secure estimation, and show that sensor fusion can improve practical performance. In this paper, the terms ‘estimator’ and ‘decoder’ are used interchangeably.

C. Notation

- $|\text{supp}(x)|$ denotes the support of vector x , i.e., the number of nonzero components in x .
- $\|x\|_{l_1} := \sum_{i=1}^n |x_i|$ where $x \in \mathbb{R}^n$. Note that this is not the same as $|\text{rowsupp}(\cdot)|$ defined in [1].
- For a matrix $M \in \mathbb{R}^{m \times n}$, $\mathcal{N}(M) = \{x \in \mathbb{R}^n | Mx = 0\}$ represents the null space of M . $\mathcal{R}(M)$ denotes the range space of M , and is defined as the set of all possible linear combinations of its column vectors.

II. OVERVIEW

A. Secure Estimation for Fixed Attacked Nodes [1]

Consider a linear dynamical system in the presence of attacks:

$$\begin{aligned} x^{(t+1)} &= Ax^{(t)} \\ y^{(t)} &= Cx^{(t)} + e^{(t)} \end{aligned} \quad (1)$$

where $x^{(t)} \in \mathbb{R}^n$ represents the state of the system at time $t \in \mathbb{N}$, $y^{(t)} \in \mathbb{R}^p$ is the output of the sensors at time t and $e^{(t)} \in \mathbb{R}^p$ represents attack signals injected by malicious agents at the sensors.

In [1], the authors proposed an elegant state estimation algorithm against adversarial attacks, where the attack signals can be unbounded and do not have to follow any particular model, but they assumed that the set of attacked nodes K does not change over time. More precisely, if $K \subset \{1, \dots, p\}$ is the set of nodes that were attacked, then we have for all t , $\text{supp}^1(e^{(t)}) \subset K$. Then, they formulated the problem of reconstructing the initial state $x^{(0)}$ of the plant from the corrupted observations $(y^{(t)})_{t=0, \dots, T-1}$ as follows:

Definition 1. ([1]) q errors are correctable after T steps by the decoder $\mathcal{D} : (\mathbb{R}^p)^T \rightarrow \mathbb{R}^n$ if for any $x^{(0)} \in \mathbb{R}^n$, any $K \subset \{1, \dots, p\}$ with $|K| \leq q$, and any sequence of vectors $e^{(0)}, \dots, e^{(T-1)}$ in \mathbb{R}^p such that $\text{supp}(e^{(t)}) \subset K$, we have $\mathcal{D}(y^{(0)}, \dots, y^{(T-1)}) = x^{(0)}$ where $y^{(t)} = CA^t x^{(0)} + e^{(t)}$ for $t = 0, \dots, T-1$.

Proposition 1. ([1]) Let $T \in \mathbb{N} \setminus \{0\}$. The following are equivalent:

- There is a decoder that can correct q errors after T steps;
- For all $z \in \mathbb{R}^n \setminus \{0\}$, $|\text{supp}(Cz) \cup \text{supp}(CAz) \cup \dots \cup \text{supp}(CA^{T-1}z)| > 2q$.

The authors then proposed the optimal decoder:

$$\mathcal{D}_0(y^{(0)}, y^{(1)}, \dots, y^{(T-1)}) = \arg \min_{\hat{x} \in \mathbb{R}^n} \|Y^{(T)} - \Phi^{(T)} \hat{x}\|_{l_0} \quad (2)$$

where the linear map $\Phi^{(T)} : \mathbb{R}^n \rightarrow \mathbb{R}^{p \times T}$ is defined by $x \mapsto [Cx \mid CAx \mid \dots \mid CA^{T-1}x]$, $Y^{(T)} = [y^{(0)} \mid y^{(1)} \mid \dots \mid y^{(T-1)}] \in \mathbb{R}^{p \times T}$, and the l_0 “norm” of M is the number of nonzero rows in M as follows [1]:

$$\|M\|_{l_0} \triangleq |\text{rowsupp}(M)| = |\{i \in \{1, \dots, p\} | M_i \neq 0\}| \quad (3)$$

¹If f is any real-valued or vector-valued function on a topological space X , the support of f , denoted by $\text{supp}(f)$, is the closure of the set points where f is nonzero: $\text{supp}(f) = \{x \in X | f(x) \neq 0\}$.

where M_i represents the i -th row of M . The cardinality of the row support of $(Y^{(T)} - \Phi^{(T)} \hat{x})$ used in Equation (2) is based on the assumption of fixed attacked nodes, because in this case, this cardinality is equal to the number of corrupted nodes. The l_0 norm in Equation (2) can be approximated by an l_1 norm to give a convex decoder and thus making it computationally feasible.

We are interested in generalizing this result to scenarios where the attacked nodes can change over time. More precisely, the set of attacked nodes is allowed to change at every time step. The decoder in Equation (2) would not be applicable in these scenarios because the cardinality of the row support of $(Y^{(T)} - \Phi^{(T)} \hat{x})$ no longer represents the number of attacked nodes. Consider the example of attacking a system with n nodes numbered $\{0, 1, \dots, n-1\}$, and an attacker corrupts node i at time t , where i is the remainder of t/i . In other words, the attacked node is cycled through the set of all nodes. The number of nodes attacked at any time step is one, however the cardinality of the matrix $(Y^{(T)} - \Phi^{(T)} \hat{x})$ is equal to T . Thus, the cardinality of the row support in Equation (2) does not help in estimating \hat{x} any more. One may argue that the decoder in [1] can be used, with $T = 1$, to solve a secure estimation problem where the attacked nodes can change over time. However, when $T = 1$, the number of correctable errors can not be large. From Proposition 1 (i.e. Proposition 2 in [1]), “the decoder can correct q errors after $T = 1$ step if and only if for all $z \in \mathbb{R}^n \setminus \{0\}$, $|\text{supp}(Cz)| > 2q$ ”. As an example, for any $C \in \mathbb{R}^{p \times n}$ where $p < n$, C has a nontrivial null space, hence there exists $z \in \mathbb{R}^n \setminus \{0\}$ such that $|\text{supp}(Cz)| = 0$; in other words, zero errors are correctable. Therefore, the decoder in [1] cannot be easily extended to the case where the attacked nodes can change over time.

In this paper, we propose a new secure estimation method inspired by the error correction problem [2] that relaxes the assumption of fixed attacked nodes in [1], under certain conditions, and therefore generalizes the results in [1] to the case where the set of attacked nodes can change over time. Our algorithm can be formulated as an l_1 -optimization, hence it is computationally efficient. Before we describe this technique, we first give a brief overview about compressive sensing and error correction in the following section.

B. Overview: Compressive Sensing and the Error Correction Problem [2]

1) *Compressed Sensing:* Sparse solutions $x \in \mathbb{R}^n$, are sought to the following problem:

$$\min_x \|x\|_0 \text{ subject to } b = Ax \quad (4)$$

where $b \in \mathbb{R}^m$ are the measurements, $A \in \mathbb{R}^{m \times n}$ ($m \ll n$) is a sensing matrix and $\|x\|_0$ denotes the number of nonzero elements of x . The following lemma provides a sufficient condition for a unique solution to (4) [2]:

Lemma 1. ([23]) If the sparsest solution to (4) has $\|x\|_0 = q$, $m \geq 2q$ and all subsets of $2q$ columns of A are full rank, then the solution is unique.

Proof. Suppose the solution is not unique, hence there exists $x_a \neq x_b$ such that $Ax_1 = b$ and $Ax_2 = b$ where $\|x_1\|_0 = \|x_2\|_0 = q$. Then $A(x_1 - x_2) = 0$ and $x_1 - x_2 \neq 0$. Since $\|x_1 - x_2\|_0 \leq 2q$ and all $2q$ columns of A are full rank (i.e. linearly independent), it is impossible to have $x_1 - x_2 \neq 0$ that satisfies $A(x_1 - x_2) = 0$ (contradiction). \square

2) *The Error Correction Problem [2]:* Consider the classical error correction problem: $y = Cx + e$ where $C \in \mathbb{R}^{p \times n}$ is a coding matrix ($p > n$) and assumed to be full rank. We wish to recover the input vector $x \in \mathbb{R}^n$ from corrupted measurements y . Here, e is an arbitrary and unknown sparse error vector. To reconstruct x , note that it is obviously sufficient to reconstruct the vector e since knowledge of $Cx + e$ together with e gives Cx , and consequently x since C has full rank [2]. In [2], the authors construct a matrix F which annihilates C on the left, i.e. $FCx = 0$ for all x . Then, they apply F to the output y and obtain

$$\tilde{y} = F(Cx + e) = Fe \quad (5)$$

Thus, the decoding problem can be reduced to that of reconstructing a sparse vector e from the observations $\tilde{y} = Fe$. Therefore, by Lemma 1, if all subsets of $2q$ columns of F are full rank, then we can reconstruct any e whose $|\text{supp}(e)| \leq q$. We refer to a decoder that can correct q errors as a q -error-correcting decoder.

3) *Equivalence between the two programs l_0 and l_1 :* In [2], the authors mentioned that the computational intractability of the l_0 -program led researchers to develop alternatives, and a frequently discussed approach considers a similar program in the l_1 norm which goes by the name of Basis Pursuit. Motivated by the problem of finding sparse decompositions of special signals in the field of mathematical signal processing, a series of beautiful and ground breaking works [24]–[27] showed exact equivalence between the two programs l_0 and l_1 . Therefore, the l_0 norm in (4) can be approximated by an l_1 norm to give a convex decoder and is therefore computationally feasible. We will discuss in more detail the condition required to ensure accurate decoding using l_1 -optimization in Section IV.

C. Connection between Secure Estimation and Error Correction when $T = 1$

It is interesting to note the connection between the conditions for there to exist a q -error-correcting decoder in a secure estimation problem (Proposition 1) and an error correction problem (Lemma 1 and Section II-B2). We consider the case $T = 1$ in this section and cover the general case for $T > 1$ in Section III.

Proposition 2. *Given $C \in \mathbb{R}^{p \times n}$ where $p > n$ and C is full column rank, the following are equivalent for there to exist a q -error-correcting decoder:*

- (i) *for all $z \in \mathbb{R}^n \setminus \{0\}$, $|\text{supp}(Cz)| > 2q$;*
- (ii) *all subsets of $2q$ columns of F are linearly independent where $\mathcal{N}(F) = \mathcal{R}(C)$.*

Proof. (i) \implies (ii): Suppose there exist $2q$ columns of F which are linearly dependent. Then, there exists $e_0 \neq 0$ such

that $Fe_0 = 0$ where $|\text{supp}(e_0)| \leq 2q$. Since $\mathcal{N}(F) = \mathcal{R}(C)$, there exists z such that $e_0 = Cz$ (i.e., $e_0 \in \mathcal{R}(C)$). Then, $|\text{supp}(Cz)| = |\text{supp}(e_0)| \leq 2q$ (contradiction).

(ii) \implies (i): We again resort to contradiction. Suppose that there exists $z \neq 0$ such that $|\text{supp}(Cz)| \leq 2q$. Let L_1 and L_2 be two disjoint subsets of $\{1, \dots, p\}$ with $|L_1| \leq q$ and $|L_2| \leq q$ such that $L_1 \oplus L_2 = \text{supp}(Cz)$ (such L_1 and L_2 exist since $|\text{supp}(Cz)| \leq 2q$). Let $e_1 = Cz|_{L_1}$ be the vector obtained from Cz by setting all the components outside of L_1 to 0, and similarly let $e_2 = -Cz|_{L_2}$ (i.e., $e_1 \neq e_2$). Then we have $Cz = e_1 - e_2$ with $\text{supp}(e_1) \subset L_1$ and $\text{supp}(e_2) \subset L_2$ with $|L_1| \leq q$ and $|L_2| \leq q$. Now, consider $y = Cz + e$

$$\begin{aligned} \tilde{y} &= Fy = F(Cz + e) = F(e_1 - e_2 + e) = F(e_1 - e_2) + Fe \\ &= Fe \implies F(e_1 - e_2) = 0 \end{aligned}$$

The last equality is due to $\mathcal{N}(F) = \mathcal{R}(C)$ (i.e. $FC = 0$). Since all subsets of $2q$ columns of F are linearly independent, $e_1 - e_2 = 0$ (contradiction). \square

Note that in Proposition 2, (i) is the condition in Proposition 1 when $T = 1$ and (ii) is the condition given in Lemma 1 and Section II-B2. Not surprisingly, when $T = 1$ (i.e., no dynamics), the secure estimation problem for the case when attacked nodes are fixed is identical to the scenario when attacked nodes can change over time. In the following section, we show the connection between the secure estimation problem, when attacked nodes can change, and the error correction problem when $T > 1$.

III. SECURE ESTIMATION VIA ERROR CORRECTION

A. Problem Setting

Consider the linear control system as follows:

$$\begin{aligned} x^{(t+1)} &= A_o x^{(t)} + Bu^{(t)} \\ y^{(t)} &= Cx^{(t)} + e^{(t)} \end{aligned} \quad (6)$$

where $x^{(t)}$, $y^{(t)}$ and $e^{(t)}$ are as defined in Equation (1), and $u^{(t)} \in \mathbb{R}^m$ represents the control inputs.

In [1], the authors showed that given a system (A_o, C) , we can increase its number of correctable errors using feedback control. More specifically, for a given B , design a matrix G so that the pair $(A_o + BG, C)$ is resilient against a large number of attacks, while satisfying other design constraints. Practically, this represents the following scenario: a physical system possesses a local control loop that has direct access to the state of the plant and that can control the evolution of the physical system. This is reasonable if the sensors are connected to the local controller through a wired link that is not subject to external attacks. Also, as part of the overall plant, a higher-level supervisory and monitoring system receives measurements from the sensors through wireless and vulnerable communication links that are subject to attacks [1].

Motivated by this, in this paper, we assume that the local control loop implements a state feedback law of the form $u^{(t)} = Gx^{(t)}$. The resulting closed loop system matrix is $(A_o + BG)$. To simplify notation, we will use A to denote $(A_o + BG)$. Without loss of generality, if $B = 0$, we have the open-loop system: $x^{(t+1)} = A_o x^{(t)}$ and $y^{(t)} = Cx^{(t)} + e^{(t)}$.

B. Reformulation into an Error Correction Problem

The problem that we consider in this section is to reconstruct the initial state $x^{(0)}$ of the plant from the corrupted observations $(y^{(t)})$ where $t = 0, \dots, T-1$. Let $E_{q,T}$ denote the set of error vectors $[e^{(0)}; \dots; e^{(T-1)}] \in \mathbb{R}^{p \cdot T}$ where each $e^{(t)}$ satisfies $|\text{supp}(e^{(t)})| \leq q \leq p$. Note that the set of attacked nodes can change over time.

$$\begin{aligned} Y &\triangleq \begin{bmatrix} y^{(0)} \\ y^{(1)} \\ \vdots \\ y^{(T-1)} \end{bmatrix} = \begin{bmatrix} Cx^{(0)} + e^{(0)} \\ CAx^{(0)} + e^{(1)} \\ \vdots \\ CA^{T-1}x^{(0)} + e^{(T-1)} \end{bmatrix} \\ &= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix} x^{(0)} + E_{q,T} \triangleq \Phi x^{(0)} + E_{q,T} \end{aligned} \quad (7)$$

where $Y \in \mathbb{R}^{p \cdot T}$ is the set of corrupted measurements and $\Phi \in \mathbb{R}^{p \cdot T \times n}$ represents the observability matrix of the closed loop system. We assume that $\text{rank}(\Phi) = n$. This is a reasonable assumption because if not, the system is unobservable and thus we cannot determine $x^{(0)}$ even if there was no attack ($E_{q,T} = 0$).

- Open-loop case ($B = 0$): A full column rank condition represents the pair (A_o, C) being observable. In other words, if not, one cannot reconstruct $x^{(0)}$ even if there were no errors in the measurements $y^{(0)}, \dots, y^{(T-1)}$ [1].
- State-feedback case: Since the state-feedback may affect the observability of a system (even though the pair (A_o, C) is observable), we have to satisfy $\text{rank}(\Phi) = n$ for the closed-loop system with state-feedback.

Note that the closed-loop system with state-feedback is controllable if and only if the open-loop system is controllable. However, state-feedback may affect the observability of a system.

We now present two methods, inspired by error correction techniques [2] [23], to estimate $x^{(0)}$, and show that they are equivalent. The first method determines the error vector $E_{q,T}$ first, and then solves for $x^{(0)}$. Consider the QR decomposition of $\Phi \in \mathbb{R}^{p \cdot T \times n}$,

$$\Phi = [Q_1 \quad Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 \quad (8)$$

where $[Q_1 \quad Q_2] \in \mathbb{R}^{p \cdot T \times p \cdot T}$ is orthogonal, $Q_1 \in \mathbb{R}^{p \cdot T \times n}$, $Q_2 \in \mathbb{R}^{p \cdot T \times (p \cdot T - n)}$ and $R_1 \in \mathbb{R}^{n \times n}$ is a rank- n upper triangular matrix. Pre-multiply Equation (7) by $[Q_1 \quad Q_2]^\top$

$$\begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix} Y = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x^{(0)} + \begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix} E_{q,T} \quad (9)$$

We can now solve for $E_{q,T}$ using the second block row:

$$\tilde{Y} \triangleq Q_2^\top Y = Q_2^\top E_{q,T} \quad (10)$$

where $Q_2^\top \in \mathbb{R}^{(p \cdot T - n) \times p \cdot T}$. From Lemma 1, Equation (10) has a unique, $s(\leq q \cdot T)$ -sparse solution if all subsets of $2s(\leq 2q \cdot T)$ columns of Q_2^\top are full rank (this is a reasonable

assumption if $(p \cdot T - n) \geq 2s = 2q \cdot T$). Thus, we consider solving the following l_1 -minimization problem

$$\hat{E}_{q,T} = \arg \min_E \|E\|_{l_1} \quad \text{subject to } \tilde{Y} = Q_2^\top E \quad (11)$$

Given $\hat{E}_{q,T}$, we can then solve for $x^{(0)}$ from the first block row of Equation (9):

$$x^{(0)} = R_1^{-1} Q_1^\top (Y - \hat{E}_{q,T}) \quad (12)$$

The second method recovers $x^{(0)}$ from the corrupted data Y directly by solving the following l_1 -minimization problem [2]:

$$\min_x \|Y - \Phi x\|_{l_1} \quad (13)$$

Lemma 2. $x^{(0)}$ is the unique solution of (13) if and only if $\hat{E}_{q,T}$ is the unique solution of (11).

Proof. (By [2]) Observe that on one hand, since $Y = \Phi x^{(0)} + E_{q,T}$ and we may decompose $x = x^{(0)} + v$, hence

$$(13) \Leftrightarrow \min_v \|E_{q,T} - \Phi v\|_{l_1}$$

On the other hand, the constraint $Q_2^\top E = \tilde{Y} = Q_2^\top E_{q,T}$ means that $E = E_{q,T} - \Phi v$ for some $v \in \mathbb{R}^n$ and, therefore,

$$\begin{aligned} (11) &\Leftrightarrow \min_v \|E\|_{l_1}, \quad E = E_{q,T} - \Phi v \\ &\Leftrightarrow \min_v \|E_{q,T} - \Phi v\|_{l_1} \end{aligned}$$

Thus, (11) and (13) are equivalent programs [2]. \square

Even though we are interested in the state $x^{(0)}$ and not necessarily the error vectors $E_{q,T}$, Lemma 2 states that if the attack vectors cannot be uniquely determined from (11), then we cannot estimate $x^{(0)}$ uniquely from (13). [1] also mentioned this notion: the existence of a decoder that can correct q errors is equivalent to saying that the map, $[\Phi \quad \mathbf{I}_{q \cdot T}] : \mathbb{R}^n \times E_{q,T} \rightarrow (\mathbb{R}^p)^T$ has an inverse for the first n components of its domain where $Y = [\Phi \quad \mathbf{I}_{q \cdot T}] \begin{bmatrix} x^{(0)} \\ E_{q,T} \end{bmatrix}$ since the attack vectors are uniquely determined by $x^{(0)}$ and the $y^{(t)}$'s, i.e., $e^{(t)} = y^{(t)} - CA^t x^{(0)}$.

Next we provide the condition for there to exist a unique solution to (13):

Proposition 3. $x^{(0)}$ is the unique solution of (13) if all subsets of $2s$ columns of Q_2^\top are linearly independent and Φ is full column rank. Also, this condition is equivalent to $|\text{supp}(\Phi z)| > 2s = 2(q \cdot T)$ for all $z \in \mathbb{R}^n \setminus \{0\}$.

Proof. By Lemma 1, 2 and Proposition 2 and noting that by definition $\mathcal{N}(Q_2^\top) = \mathcal{R}(\Phi)$. \square

C. Connection to Secure Estimation with Fixed Attacked Nodes

It may be interesting to note the connection between the conditions for the existence of a q -error-correcting decoder when the attacked nodes are fixed (Proposition 1) and when

the attacked nodes can change over time (Proposition 3): $\forall z \in \mathbb{R}^n \setminus \{0\}$,

$$(i) |\text{supp}(Cz) \cup \text{supp}(CAz) \cup \dots \cup \text{supp}(CA^{T-1}z)| > 2q$$

(the set of attacked nodes is fixed)

$$(ii) |\text{supp}(\Phi z)| = \sum_{i=0}^{T-1} |\text{supp}(CA^i z)| > 2q \cdot T$$

(the set of attacked nodes can change)

(14)

We have shown in Proposition 2 that when $T = 1$, (i) and (ii) are equivalent. When $T > 1$ (i.e., with dynamics), in order to connect conditions (i) and (ii) in (14), we make use of the following lemma:

Lemma 3. Assume A has n distinct non-zero eigenvalues ($|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0$) and $T \geq n$. Then, the following are equivalent:

$$(i) \forall z \in \mathbb{R}^n \setminus \{0\},$$

$$|\text{supp}(Cz) \cup \text{supp}(CAz) \cup \dots \cup \text{supp}(CA^{T-1}z)| > 2q$$

$$(ii) \forall v_i \in \mathbb{R}^n \text{ where } Av_i = \lambda_i v_i \text{ (i.e., eigenvector of } A),$$

$$|\text{supp}(Cv_i)| > 2q$$

Proof. Refer to the proof in [1]. \square

By using the above lemma, in order to validate condition (i) (i.e., the condition to ensure accurate decoding), we can simply check condition (ii) for all eigenvectors. Motivated by this, we prove the following result:

Theorem 1. Let $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{p \times n}$. Assume that (A, C) is observable and A has n distinct magnitude and non-zero eigenvalues (i.e., $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0$). Define:

- $s_i \triangleq |\text{supp}(Cv_i)|$, where v_i is an eigenvector of A ,
- $\mathcal{S} \triangleq \{s_1, s_2, \dots, s_n\}$,
- \mathcal{S}_m is a subset of \mathcal{S} with m elements.

Choose T such that $T \geq \max\{T_2, \dots, T_n\}$ where $T_m > \frac{(m-2) \cdot p + \min \mathcal{S}_m}{\max \mathcal{S}_m - 2q}$ for all subsets \mathcal{S}_m . Then, the following are equivalent (when $m = 1$, we have shown that (i) and (ii) in Equation (14) are equivalent where $T = 1$):

$$(i) \forall v_i \in \mathbb{R}^n \text{ where } Av_i = \lambda_i v_i, |\text{supp}(Cv_i)| > 2q$$

$$(ii) \forall v_i \in \mathbb{R}^n \text{ where } Av_i = \lambda_i v_i, |\text{supp}(\Phi v_i)| > 2q \cdot T$$

$$(iii) \forall z \in \mathbb{R}^n \setminus \{0\}, |\text{supp}(\Phi z)| > 2q \cdot T$$

(i.e. condition (ii) in Equation (14))

In order to prove Theorem 1, we make use of lemmas 6, 7 and proposition 6 (see Appendix):

Proof. (Proof of Theorem 1)

First, it is simple to prove that (i) and (ii) are equivalent ($\because |\text{supp}(\Phi v_i)| = \sum_{k=0}^{T-1} |\text{supp}(CA^k v_i)| = \sum_{k=0}^{T-1} |\text{supp}(\lambda_i^k Cv_i)| = T \cdot |\text{supp}(Cv_i)|$).

Second, we want to show that (ii) and (iii) are equivalent. The direction (iii) \implies (ii) is trivial, since (ii) is a specific case of (iii) with $z = v_i$. The other direction is more complex. Note that A is diagonalizable, therefore its eigenvectors form a basis for \mathbb{R}^n . Now consider the decomposition of z in the eigenbasis of A , i.e. $z = \sum_{i=1}^n \alpha_i v_i$ with $\alpha_i \neq 0$ for at least one i .

1) $m = 1$: Suppose there exists $z \in \mathbb{R}^n \setminus \{0\}$ such that $|\text{supp}(\Phi z)| \leq 2q \cdot T$. Without loss of generality, let $\alpha_i \neq 0$ and $\alpha_j = 0$ for all $j \neq i$, then, $2q \cdot T \geq |\text{supp}(\Phi z)| = |\text{supp}(\alpha_i \Phi v_i)| = |\text{supp}(\Phi v_i)|$ (contradiction, $\because \forall v_i, |\text{supp}(\Phi v_i)| > 2q \cdot T$).

2) $m = 2$: By Lemma 6.

3) $m \geq 3$: By Lemma 7 and Proposition 6.

We need to choose T to satisfy the worst case for any m such that $n \geq m \geq 2$. Thus, if T is chosen according to Theorem 1, then (ii) and (iii) are also equivalent. \square

Lemma 4. Assume that the pair (A_o, B) is controllable. Then the closed-loop system with state feedback is controllable and thus, there exists G such that the eigenvalues of the closed-loop matrix $A (= A_o + BG)$, i.e., $\lambda_1, \dots, \lambda_n$ can be arbitrarily located on the complex plane.

Proposition 4. Assume that $\text{rank}(\Phi) = n$, the pair (A_o, B) is controllable, the closed-loop matrix $A (= A_o + BG)$ has n non-zero eigenvalues with distinct magnitudes and T is chosen as in Theorem 1. Then, the condition for secure estimation of q -errors when the set of attacked nodes is fixed ((i) in (14)) is the same as the condition for when the set of attacked nodes can change over time ((ii) in (14)).

Proof. By Proposition 2, Lemmas 3 and 4 and Theorem 1. \square

D. Discussion: Connection between Secure Estimation and Error Correction

Note that the condition on T to ensure accurate decoding is different depending on whether the attacked nodes are fixed or changing. As mentioned previously, this is because when the set of attacked nodes changes over time, the connection between the row support of the error matrix and the corrupted nodes is lost, therefore more time steps are required to ensure equivalence between conditions (ii) and (iii) in Theorem 1. In the proof, we consider the case where A has n nonzero eigenvalues with distinct magnitudes (i.e. all eigenvalues are real). However, this is the worst case scenario. Our simulation results have shown that the lower bound on T given in Theorem 1 can be relaxed significantly, in practice (Section IV-B. Practical Performance).

What if A has complex eigenvalues? For instance, assume that $A \in \mathbb{R}^{3 \times 3}$ ($n = 3$) and it has one pair of complex conjugate eigenvalues and one real eigenvalue, i.e., $\lambda_1, \lambda_2 (= \bar{\lambda}_1), \lambda_3$ where $\lambda_1, \lambda_2 \in \mathbb{C}$, $\lambda_3 \in \mathbb{R}$ and $\bar{\cdot}$ represents the complex conjugate. We denote $v_1 = x + iy$, $v_2 = \bar{v}_1 = x - iy$ and $v_3 = w$ where $x, y, w \in \mathbb{R}^n$ are linearly independent. Any $z \in \mathbb{R}^n$ can be represented by a linear combination of n independent vectors in \mathbb{R}^n , i.e., $z = \alpha v_1 + \bar{\alpha} v_2 + \beta w = 2\text{Re}(\alpha v_1) + \beta w = \alpha_1 x + \alpha_2 y + \beta w$ where $\alpha \in \mathbb{C}$, $\beta \in \mathbb{R}$ and $\alpha_1 = \text{Re}(\alpha) \in \mathbb{R}$ and $\alpha_2 = -\text{Im}(\alpha) \in \mathbb{R}$. Therefore, the same results for real eigenvalues applies. In other words, if $|\text{supp}(Cx)| = |\text{supp}(Cy)| = |\text{supp}(Cw)| = p$ and $T > \frac{\sum_{k=0}^{T-1} s_{r,123}^k}{p-2q}$, then $|\text{supp}(\Phi z)| > 2q \cdot T$ where $s_{r,123}^k$ represents the number of cancelled support of linear combinations of x, y and w at time step k .

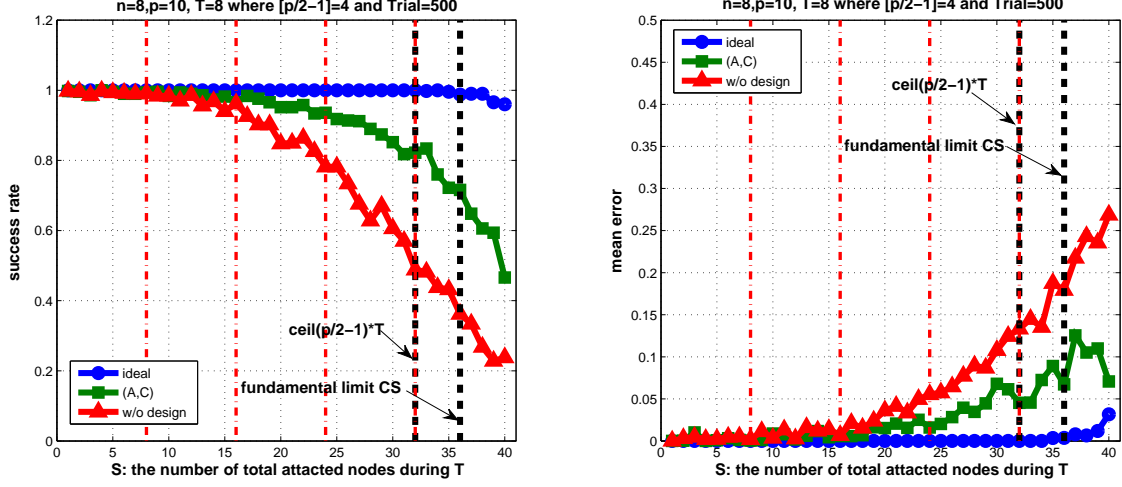


Fig. 1. Success rate and mean error of l_1 decoder on different systems (ideal coding matrix, designed state feedback and poorly designed system with $n = 8$, $p = 10$ and $T = n$, where black dot lines show the fundamental limit for dynamical systems and ideal coding matrix case respectively. We see that as the number of attacked nodes increase, success rate decreases. Also, by designing state feedback gain properly, we improve success rate and decrease mean error.

IV. DESIGN THE OPTIMAL DECODER

To ensure accurate decoding using the error correction method [2], the coding matrix must satisfy suitable conditions named Restricted Isometry Properties (RIP). In general, these conditions are extremely difficult to check. Thus, in practice, Theorem 1.4 [2] is almost always used to design the coding matrix *a priori*. This theorem states that a coding matrix whose entries are sampled from independent and identical distributions satisfies the RIP condition with overwhelming probability. However, it is hard to choose such a matrix *a priori* since our coding matrix is structurally constrained: as shown in Equation (7), Φ consists of CA^i where $i = \{0, \dots, T-1\}$. In this Section, we use Lemma 3 and the results from Proposition 4 in [1] to design a matrix Φ with good RIP, using state feedback. The decoder can then be formulated as an l_1 -minimization problem.

A. Decoder Design

Since conditions (ii) and (iii) in Theorem 1 are equivalent, condition (ii) can be used to design a state feedback controller such that the closed system can achieve accurate decoding. More specifically, choose an adequate control law such that:

- each row of C is not identically zero,
- the closed-loop system matrix A has n distinct magnitude and non-zero eigenvalues ($|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0$),
- $\forall v_i \in \mathbb{R}^n$ where $Av_i = \lambda v_i$, $|\text{supp}(Cv_i)| > 2q$.

Without loss of generality, the first condition holds. For example, if there exists a zero row in C , we can simply remove that row from C without changing the system's behavior. The second condition is required for equivalence in Theorem 1 and the third condition is used for good RIP.

Next, we prove that the theoretical limit of the maximum number of attacked nodes (which can change over time) that can be corrected is $\lceil p/2 - 1 \rceil$, under the condition that $e^{(t)}$

satisfies $|\text{supp}(e^{(t)})| \leq q$ for all t . In [1], the authors proved that the maximum number of correctable errors for fixed attacked nodes is also $\lceil p/2 - 1 \rceil$. Therefore, we can relax the assumption of fixed attacked nodes, without compromising the maximum number of correctable errors. In addition, we show that if $e^{(t)}$ can appear in an arbitrary fashion, then it is possible that more errors can be corrected at a certain time step.

Lemma 5. *Let $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{p \times n}$ and assume that (A, C) is observable and A has n non-zero eigenvalues with distinct magnitudes. Then, when the attacked nodes can change over time, the number of correctable errors after $T = \max\{n, T^*\}$ steps is maximal and equal to $\lceil p/2 - 1 \rceil$, where T^* is the lower bound on T from Theorem 1.*

Proof. By Proposition 3, if all subsets of $2q \cdot T$ columns of $Q_2^\top \in \mathbb{R}^{(p \cdot T - n) \times p \cdot T}$ are linearly independent, then $x^{(0)}$ is the unique solution of (13) and hence q non-fixed errors are correctable. Since the number of rows of Q_2^\top is smaller than the number of columns of Q_2^\top and all subsets of $2q \cdot T$ columns of Q_2^\top are linearly independent, we must have $p \cdot T - n \geq 2q \cdot T$ to satisfy the condition of Proposition 3. Rearranging this gives $p/2 - n/(2T) \geq q$. Because $p, q \in \mathbb{N}$, hence after $T = n$ steps, $p/2 - 1/2 \geq q$, i.e., the maximal number of errors that could be corrected is $q = \lfloor p/2 - 1/2 \rfloor = \lceil p/2 - 1 \rceil$. Finally, since the pair (A, C) and the value of T satisfy the conditions in Theorem 1, then this many errors are correctable. \square

In [1], the authors prove the maximum number of correctable fixed attacked nodes using (Lebesgue) measure zero. However, we prove the maximum number of correctable nodes, when they can change over time, using the dimensionality of Q_2^\top and Theorem 1.

Remark 1. *Suppose $e^{(t)}$ can appear in an arbitrary fashion, in other words, $e^{(t)}$ does not satisfy $|\text{supp}(e^{(t)})| \leq q$ for all t . For example, no error occurs at $t = 1$ but $2q$ errors occur*

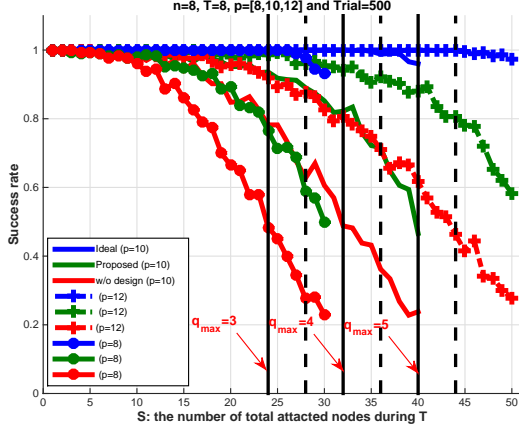


Fig. 2. Success rate and mean error of l_1 decoder on three different systems (ideal coding matrix, designed state feedback and poorly designed system with $n = 8$ and $T = n$) with different $p = [8, 10, 12]$. Black solid lines show the fundamental limit for dynamical systems and black dashed lines show the fundamental limit for the ideal coding matrix case. We see that as the number of attacked nodes increases, the success rate decreases. Also, by designing the state feedback gain properly, we improve success rate and decrease mean error.

at $t = 2$. Let $k = |\text{supp}(E_{q,T})|$ denote the total number of errors in T steps, then $x^{(0)}$ is the unique solution of (13) if all subsets of $2k$ columns of Q_2^\top are linearly independent and Φ is full rank.

Therefore, it is possible that more errors can be corrected at a certain time step. However, the total number of correctable errors in T steps is still $q \cdot T$. Note that this is a more general problem setup than that in [1].

B. Practical Performance

We show the performance of our proposed decoding algorithm using an arbitrary system. We consider the l_1 decoder on a system with $n = 8$, $p = 10$ and $T = n$ where A is chosen arbitrarily (i.e., the entries of A are chosen as independent and identically distributed random variables) and C is chosen such that every row of C has only one nonzero component (i.e., each row of C is not identically zero). This setting is more reasonable than a random system [1], i.e., i.i.d. Gaussian entries since such a matrix would have good RIP. For different numbers of attacked nodes, we test the decoder on 500 different trials with different systems and initial conditions. The initial conditions $x^{(0)}$ were randomly generated from the standard Gaussian distribution and the set of attacked nodes were chosen at random and could change over time. Since we increase the number of total attacked nodes during T steps by 1 as shown in Figure 1, we first distribute $q = \lfloor S/T \rfloor$ attacks arbitrary for each time $e^{(t)}$ and randomly distribute the remaining $(S - q \cdot T)$ attacks. For example, in Figure 1, for $S = 20$, we have at least 2 attacks for

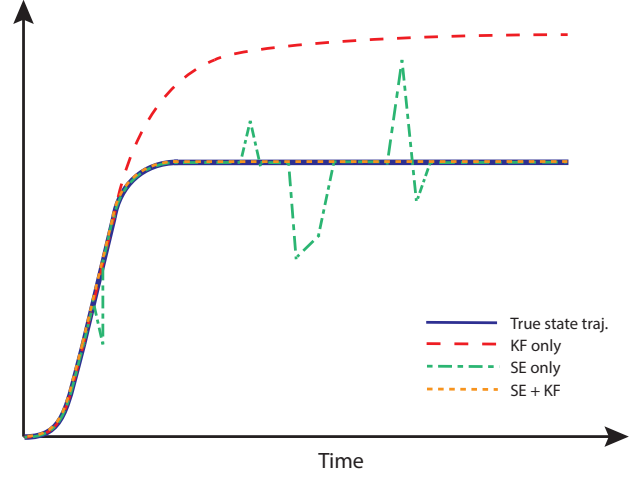


Fig. 3. Illustrative comparison of different estimation methods (KF only, secure estimator only, secure estimator with a KF). KF on its own fails to estimate the true state as attack signal is non-Gaussian. Secure estimation only method correctly estimates system state most of the time but occasionally has large estimation errors. The combination of secure estimator and KF can track the true state trajectory perfectly.

each time $e^{(i)}$ where $i = \{0, \dots, 7\}$ and distribute the remaining 4 attacks arbitrary.

In order to compare with coding matrix with good RIP, we illustrate the performance of such matrices together. Note that since our coding matrix Φ is constructed by stacking CA^i in Equation (7) where $i = \{0, \dots, T-1\}$, there could be limitations on having good restricted isometry constants caused by structural constraints. Fundamental limits of the l_1 -optimization for both secure estimation and ideal coding are also shown in Figure 1.

Figures 1 and 2 show the practical performance (success rate) of our proposed l_1 decoder on three different systems with $n = 8$, $p = 10$ and $T = n$. Note that even though a very small value of $T (= n)$ is used, compared to that dictated by Theorem 1, our proposed decoder's performance is still quite good. This demonstrates that the lower bound on T given in Theorem 1 is conservative, and can be relaxed in practice.

In Figure 2, we see that as S increases, more measurements (p) were needed to correctly recover the state of the system. In practice, this can be done by fusing different types of measurements and sensors together. For example, consider a simplified UAV dynamics where $x = [p_x, p_y, p_z, v_x, v_y, v_z]^\top$, $p(\cdot)$'s represent positions and $v(\cdot)$'s represent velocities. The observation equations of the IMU and GPS are as follows:

$$C_{IMU} = [\mathbf{0}_{3 \times 3} \quad \mathbf{I}], C_{GPS} = [\mathbf{I} \quad \mathbf{0}_{3 \times 3}] \quad (15)$$

By combining measurements from both IMU and GPS, we can increase the types of measurements (p), $C = [C_{IMU}; C_{GPS}] \in \mathbb{R}^{p \times n}$

C. Secure Estimation in conjunction with a Kalman Filter

Consider the state estimation problem for a system under attack, with the following dynamics:

$$\begin{aligned} x^{(t+1)} &= Ax^{(t)} + Bu^{(t)} + w^{(t)} \\ y^{(t)} &= Cx^{(t)} + e^{(t)} + v^{(t)} \end{aligned} \quad (16)$$

where $x^{(t)}$, $y^{(t)}$, $u^{(t)}$ and $e^{(t)}$ are as defined in Equation (6); w and v are zero mean i.i.d. Gaussian process and measurement noises, respectively.

One option is to model the attack signal as noise and apply a standard KF. More specifically, one would define a new measurement noise vector $\tilde{v}^{(t)} = e^{(t)} + v^{(t)}$, so that the measurement equation becomes $y^{(t)} = Cx^{(t)} + \tilde{v}^{(t)}$. A KF would then take the inputs $u^{(t)}$ and the corrupted measurements $y^{(t)}$ to estimate the states [7]. However, KFs are derived using the assumption that both process and measurement noises are zero mean i.i.d. white Gaussian processes, but attack signals are usually erratic and may be poorly modeled by Gaussian processes [7]. For example, in GPS spoofing attacks, attack signals are often structured to resemble normal GPS signals or can be genuine GPS signals captured elsewhere. Therefore, when the system is subjected to adversarial attacks that are poorly modeled by Gaussian processes, a KF is expected to fail to recover the true states. Figure 3 gives an illustrative example where an attack signal that increases linearly with time is injected into the measurements of state x_i . The red dashed line shows a plausible estimated state trajectory from a KF.

On the other hand, our proposed secure estimation method does not assume that the attack signal follows any particular model. Nevertheless, when the number of attacked nodes are close to the theoretical limit of correctable errors, our secure estimator would occasionally fail to perfectly recover the states (Figures 1 and 2). The green dashed line in Figure 3 depicts a possible estimation result in such a scenario: the estimated state trajectory follows the true trajectory most of the time, but with occasional deviations.

Hence, to improve the performance, we propose to combine our secure estimator with a KF as follows:

Algorithm 1 Combined secure estimator with KF

```

1: Initialize the KF
2: for each  $t$  do
3:   if  $t \geq T$  then
4:     Estimate the attack signal at time  $t$ ,  $\hat{e}^{(t)}$ , using
       secure estimator
5:   else
6:     Set  $\hat{e}^{(t)} = 0$ 
7:   end if
8:   Form a new measurement equation:  $\tilde{y}^{(t)} = Cx^{(t)} + \tilde{v}^{(t)}$ , where  $\tilde{y}^{(t)} = y^{(t)} - \hat{e}^{(t)}$  and  $\tilde{v}^{(t)} = e^{(t)} - \hat{e}^{(t)} + v^{(t)}$ 
9:   Apply standard KF using  $u$  and  $\tilde{y}$ 
10: end for

```

The intuition is that the secure estimator acts as a pre-filter for the KF, so that $\tilde{v}^{(t)}$ is close to a zero mean i.i.d. Gaussian

process even when the attack signal, $e^{(t)}$, is not. More specifically, the secure estimator usually perfectly recovers $e^{(t)}$, thus $e^{(t)} - \hat{e}^{(t)} = 0$ and $\tilde{v}^{(t)} = v^{(t)}$. What happens when the secure estimator fails? Equation (7) shows that the estimated state at time t , $\hat{x}^{(t)}$, does not directly depend on the estimated state at another time point $\hat{x}^{(\tau)}$ ($t \neq \tau$). As a result, when the secure estimator fails, its estimation error, $e^{(t)} - \hat{e}^{(t)}$, appears to be quite random. Putting these together: $\tilde{v}^{(t)} = e^{(t)} - \hat{e}^{(t)} + v^{(t)}$ is closer to a zero mean white Gaussian process than the original attack signal $e^{(t)}$. This improves the KF's performance.

Finally, the *if* statement in Algorithm 1 ensures that the secure estimator always has access to T past measurements, as required by Theorem 1. A more realistic example illustrating these behaviors is shown in Figure 8.

V. NUMERICAL EXAMPLES

In this section, we demonstrate our method through simulation of two more realistic examples where a UAV is under adversarial attack.

A. UAV Model

We consider a quadrotor with the following dynamics:

$$\begin{aligned} x^{(t+1)} &= A_0 x^{(t)} + Bu^{(t)} + k + w^{(t)} \\ y^{(t)} &= Cx^{(t)} + e^{(t)} + v^{(t)} \end{aligned} \quad (17)$$

where $x = [p_x, v_x, \theta_x, \dot{\theta}_x, p_y, v_y, \theta_y, \dot{\theta}_y, p_z, v_z]^T$ is the state vector. p_x , p_y and p_z represent the quadrotor's position along the x , y and z axis, respectively. v_x , v_y and v_z represent its velocities. θ_x and θ_y are the pitch and roll angles respectively, $\dot{\theta}_x$ and $\dot{\theta}_y$ are their corresponding angular velocities. $u = [\theta_{r,x}, \theta_{r,y}, F]^T$ is the input vector: $\theta_{r,i}$ is the reference pitch or roll angle, and F is the commanded thrust in the vertical direction. $y = [\tilde{p}_x, \tilde{p}_y, \tilde{p}_z]^T$ represents compromised position measurements from the GPS under attack signal e . w and v represent process and measurement noise respectively. k is a constant vector which represents gravitational effects, and can be dropped without loss of generality because we can always subtract it out in u . $A_\theta^{i,j}$ refers to the ij -th entry of the subsystem matrix of the discretized rotational dynamics A_θ , and B_θ^i refers to the i -th entry of the input-to-state map B_θ for the discretized rotational dynamics. T_s is the discrete time step, g is the gravitational acceleration, m is the mass of the quadrotor and K_T is a thrust coefficient. Further details about this model and its derivation can be found in [28].

B. Pole Placement

Assume that the UAV uses the state feedback control law $u^{(t)} = Gx^{(t)}$, where G is the feedback matrix which can be designed. If the pair (A_0, B) is controllable, then we can choose G to place the closed loop poles anywhere in the complex plane. For example, we can design a Linear Quadratic Regulator (LQR), such a controller would be optimal with respect to a certain quadratic cost function, but may not have good secure estimation properties. For instance, we evaluate whether the sufficient condition for q -error correction (i.e., $|\text{supp}(\Phi v_i)| > 2q \cdot T$ for all i) holds for both an LQR

$$\begin{aligned}
A_0 &= \begin{bmatrix} 1 & T_s & g \cdot \frac{T_s^2}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & gT_s & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_\theta^{11} & A_\theta^{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_\theta^{21} & A_\theta^{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & T_s & g \cdot \frac{T_s^2}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & g \cdot T_s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_\theta^{11} & A_\theta^{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_\theta^{21} & A_\theta^{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & T_s \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ B_\theta^1 & 0 & 0 \\ B_\theta^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & B_\theta^1 & 0 \\ 0 & B_\theta^2 & 0 \\ 0 & 0 & \frac{k_T \cdot T_s^2}{2m} \\ 0 & 0 & \frac{k_T \cdot T_s}{m} \end{bmatrix}, \\
C_I &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}
\end{aligned} \tag{18}$$

controller and an alternative controller, with modified closed loop poles. Figure 4 shows that the alternative controller satisfies the condition in Theorem 1, but the LQR controller does not: $|\text{supp}(\Phi v_i)| < 2q \cdot T$ for $i = 1, 2, 9$ and 10 . Although the alternative controller is less optimal with respect to the cost function, it's better at estimating attack signals and hence the states, as shown in Figures 5 and 6. This implies that we can use feedback control to design the closed loop poles of a system to achieve our desired trade-off between control and secure estimation performance.

Although it might be hard to do this in a systematic way, in our experience, small perturbations to the poles, such as making them more spread out and reducing their magnitudes, often improve the controller's estimation performance. Although this is a heuristic method, it is relatively easy to carry out in order to satisfy the conditions in Theorem 1; whereas for the error correction method [2], checking whether a coding matrix satisfies RIP is extremely difficult.

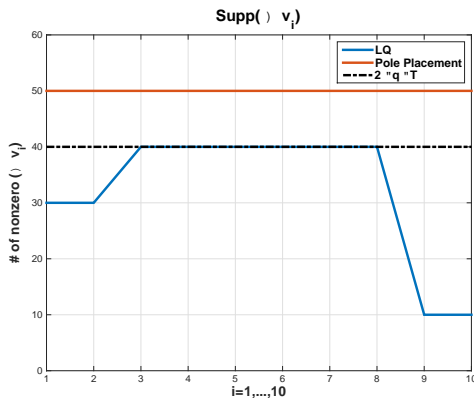


Fig. 4. Evaluation of sufficient condition ($|\text{supp}(\Phi v_i)| > 2q \cdot T$) for all closed loop eigenvectors v_i where closed loop eigenvalues are designed by LQR and closed loop poles can be placed properly to improve attack signal estimation: modified system (pole placement) satisfies the condition in Lemma 3 for all eigenvectors but system designed by LQR does not satisfy the condition.

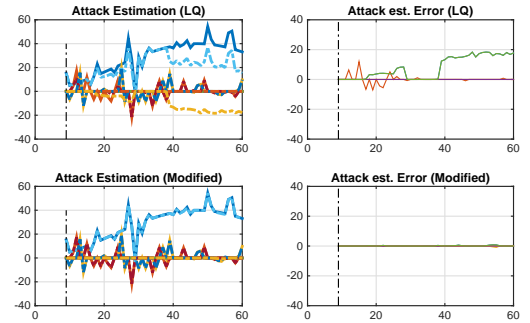


Fig. 5. True attack signal, estimated attacked signal and estimation error of 2 cases of secure estimation (closed loop eigenvalues designed by LQR, closed loop eigenvalues modified to improve attack estimation). In the left column, solid lines represent true attack signals, dashed lines are estimations. In the right column, difference between the true attack signals and their estimates in each case is plotted.

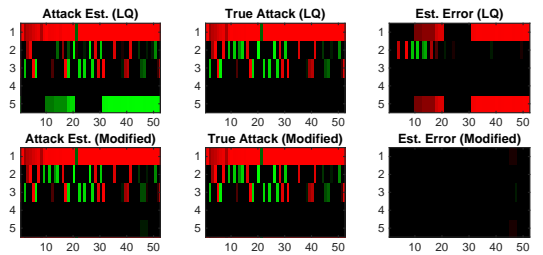


Fig. 6. Estimated attack signal, true attack signal and estimation error of 2 cases of secure estimation (closed loop eigenvalues designed by LQR, closed loop eigenvalues modified to improve attack estimation). Left column shows estimated attack signals. Middle column shows true attack signal. Right column shows estimation error. Each row corresponds to one type of measurement. Red pixels indicate positive values, green pixels indicate negative values and black indicates zero.

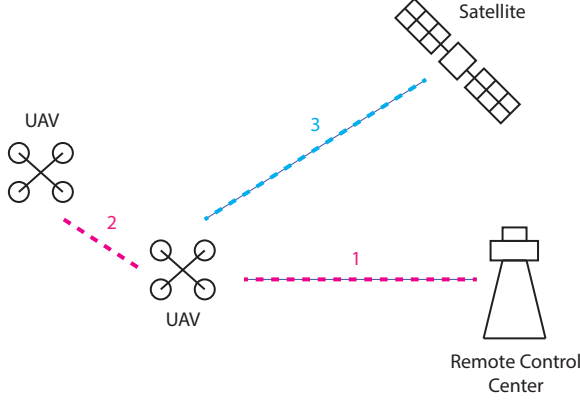


Fig. 7. Different communication channels that could be subject to adversarial attacks.

C. UAV under Adversarial Attack

1) *MITM Attack in Communication with a Remote Control Center or in UAV Formation:* In this section we illustrate and compare the performance of three state estimation methods, namely 1) KF only, 2) secure estimation only, and 3) secure estimation combined with KF. On February 15, 2015, the Federal Aviation Administration proposed to allow routine use of certain small, non-recreational UAVs in today's aviation system [29]. Thus in the near future, we may see UAVs such as Amazon Prime Air [30] and Google Project Wing vehicles [31] sharing the airspace. In order to manage this UAV traffic, we may imagine a scenario in which each UAV periodically sends measurements such as its position and velocity wirelessly to a remote control center, which then estimates the vehicle's trajectory, for collision avoidance for example. The communications link between the UAV and the control center could be subject to MITM attacks in which a malicious agent spoofs the information being sent and/or received (Channel 1 in Figure 7). Similar attack scenarios could arise in UAV formation: for formation control, individual UAVs receive information from other UAVs wirelessly in order to estimate other vehicles' positions, and this communications channel is attacked (Channel 2 in Figure 7).

Assume that the attacker spoofs the position measurements in order to deceive the receiver that the UAV is deviating in the x -direction. In the UAV traffic management scenario, this could cause the control center to initiate unnecessary collision avoidance procedures; in the UAV formation example, this

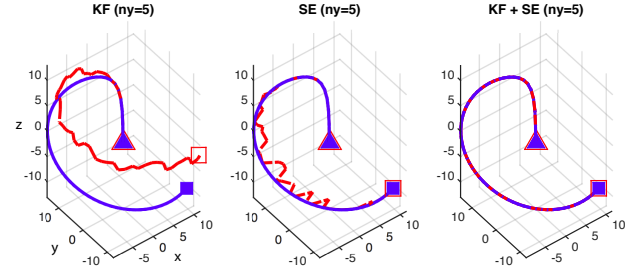


Fig. 8. Estimated UAV trajectory by three methods (KF only, Secure estimator only, Secure estimator with KF) under MITM attack. Solid blue line is the true UAV trajectory. Red dotted lines represent estimated trajectory by each method.

could break up the formation. The malicious agent injects a continuous and increasing signal in the x -position measurements. To make the estimation task harder for the receiver, the agent also injects a random Gaussian noise to an additional measurement, and the choice of this node can change at each time step. Figure 8 shows the simulation results. The true trajectory of the UAV (solid blue line) starts from the position marked by the blue triangle and ends at the position marked by the blue square. KF fails to filter out the attack signal in the x -position measurements as it is highly non-Gaussian, and estimates a trajectory (dashed red line) which significantly differs from the true trajectory. Although the secure estimator correctly estimates some portions of the trajectory and the final position of the vehicle, it produces spontaneous errors in the x direction. Finally the combined method with secure estimation and KF perfectly recovers the true path of the UAV.

2) *GPS Spoofing:* We then study adversarial attacks in the navigation system. Assume that the UAV uses a Linear Quadratic Gaussian (LQG) controller to follow a desired path, $x_r^{(t)}$, designed by LQ control. In other words, a KF takes compromised and noisy measurements $y^{(t)}$ and outputs a state estimate $\hat{x}^{(t)}$, which is then used for state feedback control: $u^{(t)} = G(\hat{x}^{(t)} - x_r^{(t)})$, where G is the feedback gain matrix.

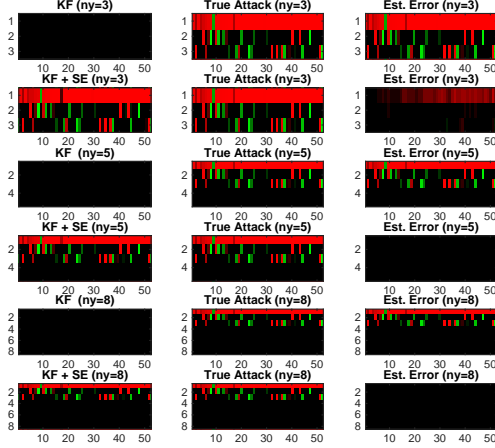


Fig. 9. Estimated attack signal, true attack signal and estimation error of different cases (KF using 3, 5 and 8 different measurements respectively, KF with secure estimation using 3, 5 and 8 different measurements respectively). Left column shows estimated attack signals. Middle column shows true attack signal. Right column shows estimation error. Each row corresponds to one type of measurement. Red pixels indicate positive values, green pixels indicate negative values and black indicates zero.

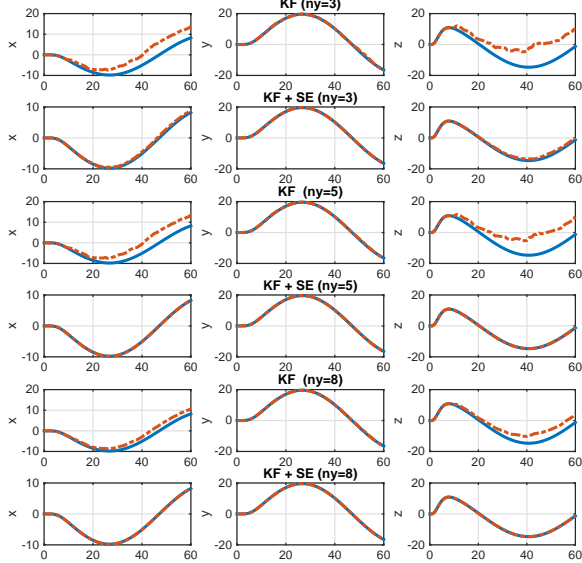


Fig. 10. Desired and actual UAV trajectory of different cases (KF using 3, 5 and 8 different measurements respectively, KF with secure estimation using 3, 5 and 8 different measurements respectively). Blue solid lines represent desired x , y and z trajectories, red dashed lines are actual UAV trajectories.

Note that in Section V-C1, the feedback controller had direct access to the state measurements $x^{(t)}$, therefore the actual vehicle trajectory is unaffected by attacks.

We first assume that the vehicle only uses GPS for navigation, i.e., $y = [\tilde{p}_x, \tilde{p}_y, \tilde{p}_z]^T$ where \tilde{p}_i 's are noisy GPS position measurements. Consider the scenario where an attacker wants to deviate the UAV's actual trajectory from its desired path, thus preventing it from reaching its destination. He injects a sinusoidal signal to \tilde{p}_x (Channel 3 in Figure 7). In addition he injects a Gaussian noise to a randomly chosen measurement node at each time step. Figures 9 and 10 show that KF cannot filter out the attack signal in this case (KF, $n_y = 3$), thus gives incorrect state estimates $\hat{x}^{(t)}$, which then causes the true vehicle trajectory (red dashed line) to deviate significantly from its desired path (solid blue line) as shown in Figure 11. As proposed in Section IV-C, we then combine secure estimation with KF. We can see from Figures 9 and 10 (KF + SE, $n_y = 3$) that its attack signal estimation is significantly more accurate with only a small estimation error for p_x and p_z . Therefore the UAV can follow its planned path much more closely (Figure 11).

We now illustrate the effect of sensor fusion. In addition to GPS signals, IMU can be used to measure velocities, pitch and roll angles. Increased types of measurements improves attack estimation by the secure estimation method as shown in Figures 9 and 10. On the other hand, it has no affect on KF's estimation accuracy (KF $n_y = 5$, KF $n_y = 8$). Figure 11 shows when 5 and 8 types of different measurements are available, the combined secure estimation with KF method enables the UAV to completely filter out the attack signal and perfectly follow its original planned trajectory (KF + SE $n_y = 5$, KF + SE $n_y = 8$).

The proposed secure estimator is designed based on the assumption that the number of compromised sensors is no more than $\lceil p/2 - 1 \rceil$. On the other hand, if more sensors are corrupted, then we can no longer guarantee perfect decoding. Therefore, as Figure 11 shows, the number of measurements, p , can be increased when designing the secure estimator, in order to protect the system against a larger number of attacked sensors.

VI. CONCLUSION

In this paper, we consider the problem of secure estimation for CPS under adversarial attacks. Unlike the assumption of fixed attack nodes in [1], we allow the attack nodes to change from time to time but we can still estimate these attack signal accurately under a certain condition. In addition, we propose a secure estimation based KF that is computationally efficient and makes no assumptions about the attack signal model. We demonstrate, through numerical examples, that our proposed secure estimator outperforms standard KF. Furthermore, we illustrate practical applications of secure estimation in UAVs under adversarial attacks. This is important not only for today's aviation system but also delivery systems with drones in the near future.

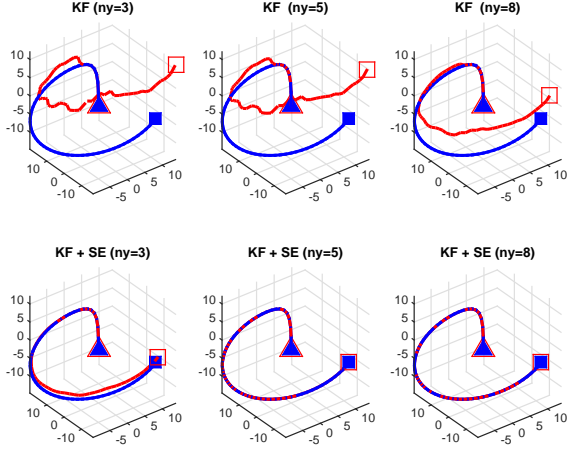


Fig. 11. Desired and actual UAV trajectory in different cases (KF using 3, 5 and 8 different measurements respectively, KF with secure estimation using 3, 5 and 8 different measurements respectively). Blue solid lines represent the desired trajectory. Red dash lines represent actual UAV trajectory under adversarial attack.

APPENDIX

(Proof of Theorem 1)

In the following lemmas and proposition, we assume the following:

- 1) $A \in \mathbb{R}^{n \times n}$ has n distinct magnitude and non-zero eigenvalues ($|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0$, $\lambda_i \neq 0$), (A, C) is observable.
- 2) $\forall v_i \in \mathbb{R}^n$ where $Av_i = \lambda_i v_i$ (i.e., eigenvector of A), $|\text{supp}(Cv_i)| > 2q$

In addition, we define the following notations: $\mathbf{v} \triangleq [v_1 \ v_2 \ \dots \ v_n] \in \mathbb{R}^{n \times n}$, $\Lambda \triangleq \text{diag}\{\lambda_1, \dots, \lambda_n\} \in \mathbb{R}^{n \times n}$ and $\psi_i \triangleq [Cv_i; \lambda_i Cv_i; \lambda_i^2 Cv_i; \dots; \lambda_i^{T-1} Cv_i]$. Since A is diagonalizable, then $\text{rank}(\mathbf{v}) = n$. Thus the eigenvectors of A form a basis for \mathbb{R}^n and any $z \in \mathbb{R}^n$ can be expressed in the eigenbasis of A , i.e., $z = \sum_{i=1}^n \alpha_i v_i = \mathbf{v} \cdot \alpha$, where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$.

We will first consider a simple case ($m = 2$ and $m = 3$) and then generalize the results to $m > 3$.

Lemma 6. ($m = 2$) Consider $z = \sum_{i=1}^2 \alpha_i v_i$ (i.e., $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, $\alpha_j = 0, \forall j \geq 3$):

- 1) If the i -th row of $C\mathbf{v}\Lambda^k\alpha = 0$ at time step k where $c_i^\top v_1 \neq 0$ and $c_i^\top v_2 \neq 0$, then the i -th row of $C\mathbf{v}\Lambda^l\alpha \neq 0$ for all $l \in \{0, \dots, T-1\}$ where $l \neq k$.
- 2) If $T > \frac{\min\{s_1, s_2\}}{\max\{s_1, s_2\} - 2q}$, then $|\text{supp}(\Phi z)| > 2q \cdot T$.

Proof. (1) (suppose not) $\mathbf{e}_i^\top C\mathbf{v}\Lambda^k\alpha = c_i^\top \mathbf{v}\Lambda^k\alpha = 0$ where

$$C = \begin{bmatrix} c_1^\top \\ c_2^\top \\ \vdots \\ c_p^\top \end{bmatrix}. \text{ We have } c_i^\top \mathbf{v}\Lambda^l\alpha = c_i^\top \mathbf{v}\Lambda^k\alpha = 0,$$

$$c_i^\top [v_1 \ v_2] \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} \lambda_1^l \\ \lambda_2^l \end{bmatrix} = c_i^\top [v_1 \ v_2] \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} \lambda_1^k \\ \lambda_2^k \end{bmatrix} = 0$$

We can reformulate this as follows:

$$\begin{bmatrix} 1 & \gamma^l \\ 1 & \gamma^k \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} c_i = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where $\gamma = \frac{\lambda_2}{\lambda_1}$. Since A has n non-zero and distinct magnitude eigenvalues, $|\gamma| \neq 1$ and $\alpha_1 \neq 0, \alpha_2 \neq 0$, thus we must have $v_1^\top c_i = v_2^\top c_i = 0$ (contradiction, since the i -th row of Cv_1 and Cv_2 are not zero by assumption).

(2) Let L_1, L_2, L_{12} be three disjoint subsets of $\{1, \dots, p\}$ such that $L_1 = \text{supp}(Cv_1) \cap \text{supp}(Cv_2)^c$, $L_2 = \text{supp}(Cv_2) \cap \text{supp}(Cv_1)^c$, and $L_{12} = \text{supp}(Cv_1) \cap \text{supp}(Cv_2)$ where the superscript c represents the complement. Then, $\text{supp}(Cv_1) = L_1 \oplus L_{12}$, $\text{supp}(Cv_2) = L_2 \oplus L_{12}$, $s_1 \triangleq |\text{supp}(Cv_1)| = |L_1 \oplus L_{12}| > 2q$, $s_2 \triangleq |\text{supp}(Cv_2)| = |L_2 \oplus L_{12}| > 2q$ and $s_{12} \triangleq |L_{12}| \leq \min\{s_1, s_2\}$. Also, possible cancellations only occur in the subset L_{12} by definition.

$$\begin{aligned} |\text{supp}(\Phi z)| &= |\text{supp}(\alpha_1 \psi_1 + \alpha_2 \psi_2)| \\ &= T \cdot (s_1 - s_{12}) + T \cdot (s_2 - s_{12}) \\ &\quad + \sum_{k=0}^{T-1} (s_{12} - s_{r,12}^k) \\ &= T \cdot (s_1 + s_2 - s_{12}) - \sum_{k=0}^{T-1} s_{r,12}^k \\ &\geq T \cdot (s_1 + s_2 - \min\{s_1, s_2\}) - \sum_{k=0}^{T-1} s_{r,12}^k \end{aligned}$$

where $s_{r,12}^k$ is the number of cancelled support in L_{12} at time step k . More specifically, $i \in s_{r,12}^k$ if $c_i^\top v_1 \neq 0$ and $c_i^\top v_2 \neq 0$, but $c_i^\top \mathbf{v}\Lambda^k\alpha = 0$. From (1), we have the followings:

$$\begin{aligned} s_{r,12}^0 &\leq |L_{12}| \\ s_{r,12}^1 &\leq |L_{12}| - s_{r,12}^0 \\ s_{r,12}^2 &\leq |L_{12}| - s_{r,12}^0 - s_{r,12}^1 \\ &\vdots \\ s_{r,12}^{T-1} &\leq |L_{12}| - \sum_{k=0}^{T-2} s_{r,12}^k \end{aligned}$$

Thus, $\sum_{k=0}^{T-1} s_{r,12}^k \leq |L_{12}| \leq \min\{s_1, s_2\}$, and

$$|\text{supp}(\Phi z)| \geq T \cdot \max\{s_1, s_2\} - \min\{s_1, s_2\} > 2q \cdot T. \quad \square$$

Lemma 7. ($m = 3$) Consider $z = \sum_{i=1}^3 \alpha_i v_i$ where $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, $\alpha_3 \neq 0$ and $\alpha_i = 0$ for $i = \{4, 5, \dots, n\}$.

- 1) $\sum_{k=0}^{T-1} s_{r,123}^k \leq 2 \cdot s_{123}$ where $s_{123} = |L_{123}| = |\text{supp}(Cv_1) \cap \text{supp}(Cv_2) \cap \text{supp}(Cv_3)|$.
- 2) If $T > \frac{p + \min\{s_1, s_2, s_3\}}{\max\{s_1, s_2, s_3\} - 2q}$, then $|\text{supp}(\Phi z)| > 2q \cdot T$.

Proof. (1) Claim: the i -th row of $C\mathbf{v}\Lambda^k\alpha$ can be zero at most 2 times over T time steps.

(suppose not) $c_i^\top \mathbf{v}\Lambda^d\alpha = c_i^\top \mathbf{v}\Lambda^e\alpha = c_i^\top \mathbf{v}\Lambda^f\alpha = 0$ ($d \neq e \neq f$):

$$\underbrace{\begin{bmatrix} \lambda_1^d & \lambda_2^d & \lambda_3^d \\ \lambda_1^e & \lambda_2^e & \lambda_3^e \\ \lambda_1^f & \lambda_2^f & \lambda_3^f \end{bmatrix}}_{\text{nonsingular}} \begin{bmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{bmatrix} \mathbf{v}^\top c_i = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(We can prove the non-singularity of this matrix using elementary row operations and showing it is full rank:

$$\begin{bmatrix} 1 & \gamma_{12}^d & \gamma_{13}^d \\ 1 & \gamma_{12}^e & \gamma_{13}^e \\ 1 & \gamma_{12}^f & \gamma_{13}^f \end{bmatrix} \sim \begin{bmatrix} 1 & \gamma_{12}^d & \gamma_{13}^d \\ 0 & \gamma_{12}^e - \gamma_{12}^d & \gamma_{13}^e - \gamma_{13}^d \\ 0 & \gamma_{12}^f - \gamma_{12}^d & \gamma_{13}^f - \gamma_{13}^d \end{bmatrix} \\ \sim \begin{bmatrix} 1 & \gamma_{12}^d & \gamma_{13}^d \\ 0 & 1 & \frac{\gamma_{13}^e - \gamma_{13}^d}{\gamma_{12}^e - \gamma_{12}^d} \\ 0 & 1 & \frac{\gamma_{13}^f - \gamma_{13}^d}{\gamma_{12}^f - \gamma_{12}^d} \end{bmatrix} \sim \begin{bmatrix} 1 & \gamma_{12}^d & \gamma_{13}^d \\ 0 & 1 & \frac{\gamma_{13}^e - \gamma_{13}^d}{\gamma_{12}^e - \gamma_{12}^d} \\ 0 & 0 & \frac{\gamma_{13}^f - \gamma_{13}^d}{\gamma_{12}^f - \gamma_{12}^d} - \frac{\gamma_{13}^e - \gamma_{13}^d}{\gamma_{12}^e - \gamma_{12}^d} \end{bmatrix}$$

where $\frac{\gamma_{13}^f - \gamma_{13}^d}{\gamma_{12}^f - \gamma_{12}^d} - \frac{\gamma_{13}^e - \gamma_{13}^d}{\gamma_{12}^e - \gamma_{12}^d} \neq 0$)

Therefore $\mathbf{v}^\top c_i = 0$ (contradiction). A similar derivation as in Lemma 6 then shows $\sum_{k=0}^{T-1} s_{r,123}^k \leq 2 \cdot s_{123}$.

(2) Consider Cv_1, Cv_2, Cv_3 and Φz :

$\text{supp}(Cv_1)$	$\text{supp}(Cv_2)$	$\text{supp}(Cv_3)$
L_1	$\mathbf{0}$	$\mathbf{0}$
$\mathbf{0}$	L_2	$\mathbf{0}$
$\mathbf{0}$	$\mathbf{0}$	L_3
L_{12}	L_{12}	$\mathbf{0}$
$\mathbf{0}$	L_{23}	L_{23}
L_{13}	$\mathbf{0}$	L_{13}
L_{123}	L_{123}	L_{123}

Without loss of generality, assume $s_3 \geq s_2 \geq s_1$ (recall $s_1 = |L_1 \oplus L_{12} \oplus L_{13} \oplus L_{123}| = |L_1| + s_{12} + s_{13} + s_{123}$):

$$\begin{aligned} |\text{supp}(\Phi z)| &= T \cdot (s_1 - s_{12} - s_{13} - s_{123}) \\ &\quad + T \cdot (s_2 - s_{12} - s_{23} - s_{123}) \\ &\quad + T \cdot (s_3 - s_{23} - s_{13} - s_{123}) \\ &\quad + \sum_{k=0}^{T-1} \left((s_{12} - s_{r,12}^k) + (s_{23} - s_{r,23}^k) \right. \\ &\quad \left. + (s_{13} - s_{r,13}^k) + (s_{123} - s_{r,123}^k) \right) \\ &= T \cdot (s_3 + s_1 + s_2 - s_{12} - s_{13} - s_{23} - 2 \cdot s_{123}) \\ &\quad - \sum_{k=0}^{T-1} (s_{r,12}^k + s_{r,23}^k + s_{r,13}^k + s_{r,123}^k) \\ &\geq T \cdot s_3 - p - s_{123} \geq T \cdot s_3 - p - \min\{s_1, s_2, s_3\} \\ &> 2q \cdot T \end{aligned}$$

where $\sum_{k=0}^{T-1} (s_{r,12}^k + s_{r,23}^k + s_{r,13}^k) \leq s_{12} + s_{23} + s_{13} \leq p - s_{123}$, $\sum_{k=0}^{T-1} s_{r,123}^k \leq 2 \cdot s_{123}$ and $s_{123} \leq \min\{s_1, s_2, s_3\}$ and note that $T > \frac{p + \min\{s_1, s_2, s_3\}}{\max\{s_1, s_2, s_3\} - 2q}$. \square

Proposition 5. Consider m eigenvector combinations ($n \geq m \geq 2$). Then, the total number of cancellations over T

time steps satisfies $\sum_{k=0}^{T-1} s_{r,12\dots m}^k \leq (m-1) \cdot s_{12\dots m}$ where $s_{12\dots m} = |\text{supp}(Cv_1) \cap \dots \cap \text{supp}(Cv_m)|$.

Proof. Claim: the i -th row of $C\mathbf{v}\Lambda^k\alpha$ can be zero at most $(m-1)$ times over T times steps. (suppose not)

$$\begin{bmatrix} \lambda_1^d & \lambda_2^d & \dots & \lambda_m^d \\ \lambda_1^e & \lambda_2^e & \dots & \lambda_m^e \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^r & \lambda_2^r & \dots & \lambda_m^r \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_m \end{bmatrix} \mathbf{v}^\top c_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ = \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ \lambda_1^{e-d} & \lambda_2^{e-d} & \dots & \lambda_m^{e-d} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{r-d} & \lambda_2^{r-d} & \dots & \lambda_m^{r-d} \end{bmatrix}}_{\text{an } m \times m \text{ alternant matrix}} \cdot \begin{bmatrix} \lambda_1^d \cdot \alpha_1 & 0 & \dots & 0 \\ 0 & \lambda_2^d \cdot \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_m^d \cdot \alpha_m \end{bmatrix} \mathbf{v}^\top c_i$$

Since the first matrix is an alternant matrix, we can show that it is a nonsingular matrix by using an alternant determinant and a Vandermonde matrix [32]. Therefore $\mathbf{v}^\top c_i = 0$ (contradiction). A similar derivation as in Lemma 6 then shows $\sum_{k=0}^{T-1} s_{r,12\dots m}^k \leq (m-1) \cdot s_{12\dots m}$. \square

Proposition 6. Consider $z = \sum_{i=1}^m \alpha_i v_i$ where $\alpha_i \neq 0$ for $i = \{1, \dots, m\}$ and $n \geq m \geq 2$. If $T > \frac{(m-2) \cdot p + \min S_m}{\max S_m - 2q}$ where $S_m \triangleq \{s_1, s_2, \dots, s_m\}$, then $|\text{supp}(\Phi z)| > 2q \cdot T$.

Proof.

$$\begin{aligned} |\text{supp}(\Phi z)| &\geq T \cdot \max\{s_i\} - 1 \cdot \underbrace{(s_{12} + s_{13} + \dots + s_{m-1,m})}_{\text{correspond to } {}_m C_2} \\ &\quad - 2 \cdot \underbrace{(s_{123} + s_{124} + \dots + s_{m-2,m-1,m})}_{\text{correspond to } {}_m C_3} \\ &\quad - \dots - (m-1) \cdot \underbrace{s_{12\dots m}}_{\text{correspond to } {}_m C_m} \\ &\geq T \cdot \max\{s_i\} - \underbrace{(s_{12} + \dots + s_{12\dots m})}_{\leq p} \\ &\quad - 1 \cdot (s_{123} + s_{124} + \dots + s_{m-2,m-1,m}) \\ &\quad - \dots - (m-2) \cdot s_{12\dots m} \\ &\geq T \cdot \max\{s_i\} - (m-2) \cdot p - s_{12\dots m} \\ &\geq T \cdot \max\{s_i\} - (m-2) \cdot p - \min\{s_i\} \\ &> 2q \cdot T \end{aligned}$$

\square

ACKNOWLEDGMENT

This work was supported by the NSF CPS project ActionWebs under grant number 0931843, NSF CPS project FORCES under grant number 1239166.

REFERENCES

- [1] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *Automatic Control, IEEE Transactions on*, vol. 59, no. 6, pp. 1454–1467, June 2014.
- [2] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.
- [3] V. D. Gligor, "A note on denial-of-service in operating systems," *IEEE Transactions on Software Engineering*, vol. 10, no. 3, pp. 320 – 324, May 1984.
- [4] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," *Preprints of the 1st Workshop on Secure Control Systems*, 2010.
- [5] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: towards survivable cyber-physical systems," *28th international conference on distributed computing systems workshop*, pp. 495–500, June 2008.
- [6] O. Kosut, J. Liyan, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid Special Issue on Cyber, Physical, and System Security for Smart Grid*, vol. 2, no. 4, pp. 645–658, 2012.
- [7] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," *American Control Conference*, 2013.
- [8] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 13, May 2011.
- [9] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. Sastry, "Cyber security analysis of state estimators in electric power systems," *49th IEEE Conference on Decision and Control*, pp. 5991 – 5998, December 2010.
- [10] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and fault-tolerant control*, 2nd ed. Berlin, Germany: Springer, 2006.
- [11] M.-A. Massoumnia, G. C. Verghese, and A. S. Willsky, "Failure detection and identification," *IEEE Transactions on Automatic Control*, vol. 34, no. 3, pp. 316–321, March 1989.
- [12] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. Englewood, Cliffs, NJ, USA: Prentice-Hall, 1998.
- [13] C. Kwon and I. Hwang, "Hybrid robust controller design: Cyber attack attenuation for cyber-physical systems," *52nd IEEE Conference on Decision and Control*, 2013.
- [14] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical attacks in power networks: models, fundamental limitations and monitor design," *50th IEEE conference on decision and control and european control conference*, pp. 2195 – 2201, December 2011.
- [15] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Combating false data injection attacks in smart grid using kalman filter," *International Conference on Computing, Networking and Communications*, pp. 16–20, February 2014.
- [16] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A survey of game theory as applied to network security," *43rd Hawaii International Conference on System Sciences*, 2010.
- [17] A. Gupta, C. Langbort, and T. Basar, "Optimal control in the presence of an intelligent jammer with limited actions," *49th IEEE Conference on Decision and Control*, pp. 1096 – 1101, December 2010.
- [18] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Basar, and J.-P. Hubaux, "Game theory meets network security and privacy," *ACM Computing Surveys*, vol. 45, no. 3, June 2013.
- [19] A. Gueye, V. Marbukh, and J. C. Walrand, "Towards a metric for communication network vulnerability to attacks: A game theoretic approach," *3rd International ICST Conference on Game Theory for Networks*, May 2012.
- [20] M. Fei, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," *52nd IEEE Conference on Decision and Control*, pp. 1854 – 1859, December 2013.
- [21] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, "Robustness of attack-resilient state estimators," in *Cyber-Physical Systems (ICCPs), 2014 ACM/IEEE International Conference on*, 2014, pp. 163–174.
- [22] D. Welch and S. Lathrop, "Wireless security threat taxonomy," *IEEE Systems, man and cybernetics society information assurance workshop*, pp. 76 – 83, June 2003.
- [23] D. Hayden, Y. H. Chang, J. Goncalves, and C. Tomlin, "Compressed sensing for network reconstruction," *arXiv:1411.4095*, 2014.
- [24] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization," *Proc. Natl. Acad. Sci.*, pp. 2197–2202, 2003.
- [25] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of r^n bases," *Information Theory, IEEE Transactions on*, vol. 48, no. 9, pp. 2558–2567, 2002.
- [26] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *Information Theory, IEEE Transactions on*, vol. 49, no. 12, pp. 3320–3325, 2003.
- [27] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [28] P. Bouffard, "On-board model predictive control of a quadrotor helicopter: Design, implementation, and experiments," University of California Berkeley, <http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-241.html>, Technical Report UCB/EECS-2012-241, December 2012.
- [29] "Press release – DOT and FAA propose new rules for small unmanned aircraft systems," http://www.faa.gov/news/press_releases/news_story.cfm/?newsId=18295, accessed: 2015-02-15.
- [30] "Amazon Prime Air," <http://www.amazon.com/b?node=8037720011>.
- [31] "Google project wing," <http://www.theatlantic.com/technology/archive/2014/08/inside-g/> accessed: 2014-08-28.
- [32] D. Kalman, "The generalized vandermonde matrix," *Mathematics Magazine*, pp. 15–21, 1984.