ICAS 2016
30th Congress of the International Council of the Aeronautical Sciences
DCC, Daejeon, Korea ; September 25-30, 2016

Secure Estimation for Unmanned Aerial Vehicles against Adversarial Cyber Attacks

Qie Hu[1,*], Young Hwan Chang[2,*], and Claire J. Tomlin[1]

[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA

[2]Department of Biomedical Engineering, Oregon Health and Science University, OR 97239 USA

[*]These authors contributed equally

Keywords: Secure Estimation, UAV, Cyber Attacks, Error Correction

***Abstract*— *In the coming years, usage of Unmanned Aerial Vehicles (UAVs) is expected to grow tremendously. Maintaining security of UAVs under cyber attacks is an important yet challenging task, as these attacks are often erratic and difficult to predict. Secure estimation problems study how to estimate the states of a dynamical system from a set of noisy and maliciously corrupted sensor measurements. The fewer assumptions that an estimator makes about the attacker, the larger the set of attacks it can protect the system against. In this paper, we focus on sensor attacks on UAVs and attempt to design a secure estimator for linear time-invariant systems based on as few assumptions about the attackers as possible. We propose a computationally efficient estimator that protects the system against arbitrary and unbounded attacks, where the set of attacked sensors can also change over time. In addition, we propose to combine our secure estimator with a Kalman Filter for improved practical performance and demonstrate its effectiveness through simulations of two scenarios where an UAV is under adversarial cyber attack.*

I. Introduction

The already widespread use of Unmanned Aerial Vehicles (UAVs) is expected to continue to grow at a tremendous rate over the next few years [1]. Civilian applications of UAVs, such as cargo delivery [2], [3], infrastructure surveillance [4], and agricultural applications [5], can provide great benefits to society.

However, UAVs may be vulnerable to a variety of cyber attacks. For example, to manage the increased UAV traffic, each UAV may periodically send its position measurements wirelessly to a remote traffic management center. Similarly, two UAVs may exchange position and velocity information in a collaborative collision avoidance procedure. These communication links could be subject to Man-In-The-Middle (MITM) attacks in which a malicious agent spoofs the information being sent and/or received. Successful attacks can lead to collisions of vehicles, economic loss and bodily damage. Therefore, maintaining the security of UAVs under such cyber attacks is an important but also challenging task, as attacks are often erratic and difficult to model.

Secure estimation problems study how to estimate the states of a dynamical system from a set of noisy and maliciously corrupted sensor measurements. In designing such estimators, it is desirable to make as few assumptions about the attackers as possible. This is because it is very difficult, if not impossible, to predict the behavior of attackers, and when an attack signal violates the assumptions of a secure estimator, then this estimator would fail to detect the attack.

Researchers have studied various approaches to securing general cyber-physical systems, each based on a different set of assumptions about the attacker. For example, the authors in [6], [7] assume that the attack signal would follow certain probabilistic distributions and then design filters for detection of such attacks. In [8], [9], [10], [11], [12], the authors use the game theory framework, where the controller and attacker are players with competing goals in

a game. Attackers are assumed to adopt specific strategies that maximize a certain cost and the controller or estimator is designed to minimize such a cost. More recently, Fawzi *et al.* proposed in [13] a secure estimation method for arbitrary attacks, with a limiting assumption that the set of attacked sensors do not change with time.

In this paper, we focus on sensor attacks on UAVs and attempt to design a secure estimator for linear time-invariant (LTI) systems based on as few assumptions about the attackers as possible. First, we do not assume that the attack signals follow any stochastic distributions, and thus our proposed estimator works for arbitrary and unbounded attacks. Second, we allow the set of attacked sensors to change over time. The only assumption we make is that the number of attacked sensors is sparse.

We formulate this secure estimation problem into the classical error correction problem, from which we propose an l_1-optimization based estimator that is computationally efficient. In addition, we prove the maximum number of sensor attacks that can be corrected with our estimator and propose a practical method for estimator design that guarantees accurate decoding. Finally, to improve the estimator's practical performance, we propose to combine our secure estimator with a Kalman Filter (KF), and demonstrate its effectiveness using two examples of UAVs under adversarial cyber attacks.

II. Classical Error Correction: A Review

A. Compressed Sensing

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

$$\min_x \|x\|_0 \text{ subject to } b = Ax \tag{1}$$

where $b \in \mathbb{R}^m$ are the measurements, and $A \in \mathbb{R}^{m\times n}$ $(m \ll n)$ is a sensing matrix. $\|x\|_0$ denotes the number of nonzero elements of x. The following lemma provides a sufficient condition for a unique solution to (1).

Lemma 1: ([14]) If the sparsest solution to (1) has $\|x\|_0 = q$ and $m \geq 2q$ and all subsets of $2q$ columns of A are full rank, then the solution is unique.

Proof: Suppose the solution is not unique. Therefore, there exists $x_1 \neq x_2$ 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 $A(x_1 - x_2) = 0$. This contradicts the assumption. ■

B. The Error Correction Problem [14]

Consider the classical error correction problem: $y = Cx + e$ where $C \in \mathbb{R}^{l\times n}$ is a coding matrix $(l > n)$ 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 [14]. In [14], 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

$$\tilde{y} = F(Cx + e) = Fe. \tag{2}$$

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 such that $\|e\|_0 \leq q$.

III. Secure Estimation

A. Problem Formulation

Consider the LTI system as follows:

$$\begin{aligned} x(k+1) &= A_o x(k) + Bu(k) \\ y(k) &= Cx(k) + e(k), \end{aligned} \tag{3}$$

where $x(k) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^p$ and $u(k) \in \mathbb{R}^m$ are the states, measurements and control inputs at time step k. $e(k) \in \mathbb{R}^p$ represents the attack signal at time k. Our goal is to reconstruct the initial state $x(0)$ of the plant from the corrupted observations $y(k)$'s where $k = 0, ..., T-1$.

The attack vector $e(k)$ is such that if the ith sensor is attacked at time k, then $e_i(k)$, the ith element of $e(k)$ is nonzero, otherwise $e_i(k) = 0$. We assume that the attack signal can be arbitrary and unbounded. In addition, we assume that the

set of attacked sensors can change over time. As illustrated by the following example, if 2 sensors are attacked at each time step, we can have sensors 1 and 3 attacked at time step 0, sensors 2 and 3 attacked at time 1, and so on:

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

where $*$ denotes a nonzero component (i.e., an attack or corruption).

Furthermore, assume that a local control loop implements secure state feedback and is not subject to attack: $u(k) = Gx(k)$. In the case of UAVs, this corresponds to using measurements from onboard, hardwired sensors such as Inertial Measurement Units (IMU) for autopilot and trajectory following. The resulting closed loop system is:

$$\begin{aligned} x(k+1) &= Ax(k) \\ y(k) &= Cx(k) + e(k), \end{aligned} \tag{4}$$

where the closed loop system matrix $A = A_o + BG$.

Finally, we define the number of correctable attacks/errors as follows:

Definition 1: When the set of attacked sensors/nodes can change over time, q errors are correctable after T steps by the estimator/decoder $\mathcal{D} : (\mathbb{R}^p)^T \to \mathbb{R}^n$ if for any $x(0) \in \mathbb{R}^n$ and any sequence of vectors $e(0), ..., e(T-1)$ in $\mathbb{R}^p$ such that $|\mathsf{supp}(e(k))| \leq q$, we have $\mathcal{D}(y(0), ..., y(T-1)) = x(0)$ where $y(k) = CA^k x(0) + e(k)$ for $k = 0, ..., T-1$.

B. Methodology

Let $E_{q,T}$ denote the set of error vectors $\begin{bmatrix} e(0); & \dots & ; e(T-1) \end{bmatrix} \in \mathbb{R}^{p \cdot T}$ where each $e(k)$ satisfies $\|e(k)\|_0 \leq q \leq p$.

$$\begin{aligned} Y \triangleq \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(T-1) \end{bmatrix} &= \begin{bmatrix} Cx(0) + e(0) \\ CAx(0) + e(1) \\ \vdots \\ CA^{T-1}x(0) + e(T-1) \end{bmatrix} \\ &= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-1} \end{bmatrix} x(0) + E_{q,T} \triangleq \Phi x(0) + E_{q,T} \end{aligned} \tag{5}$$

where $Y \in \mathbb{R}^{p \cdot T}$ is a collection of corrupted measurements over T time steps and $\Phi \in \mathbb{R}^{p \cdot T \times n}$ represents an observability-like matrix of the system. Here, we need to assume that $\mathrm{rank}(\Phi) = n$; otherwise, the system is unobservable and we cannot determine $x(0)$ even if there is no attack (i.e., $E_{q,T} = 0$).

Inspired by the error correction techniques proposed in [14] and [15], we first determine the error vector $E_{q,T}$, and then solve for $x(0)$. Consider the QR decomposition of $\Phi \in \mathbb{R}^{p \cdot T \times n}$,

$$\Phi = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1 \tag{6}$$

where $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \in \mathbb{R}^{p \cdot T \times p \cdot T}$ 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-multiplying (5) by $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^\top$ gives:

$$\begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix} Y = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x(0) + \begin{bmatrix} Q_1^\top \\ Q_2^\top \end{bmatrix} E_{q,T}. \tag{7}$$

We can compute $E_{q,T}$ by using the second block row:

$$\tilde{Y} \triangleq Q_2^\top Y = Q_2^\top E_{q,T} \tag{8}$$

where $Q_2^\top \in \mathbb{R}^{(p \cdot T - n) \times p \cdot T}$. From Lemma 1, (8) has a unique, s-sparse solution (where $s \leq q \cdot T$) if all subsets of $2s$ columns (at most $2q \cdot T$ columns) of $Q_2^\top$ are full rank. Clearly, this is a reasonable assumption if $(p \cdot T - n) \geq 2q \cdot T$. Therefore, we consider solving the following l_1-minimization problem:

$$\hat{E}_{q,T} = \arg\min_E \|E\|_{l_1} \text{ s.t. } \tilde{Y} = Q_2^\top E \tag{9}$$

Now, given the vector $\hat{E}_{q,T}$, we can compute $x(0)$ from the first block row of (7) as follows:

$$x(0) = R_1^{-1} Q_1^\top (Y - \hat{E}_{q,T}) \tag{10}$$

The following lemma provides the conditions under which the solution to (10) exists and is unique.

Lemma 2: $x(0)$ is the unique solution if $|\mathsf{supp}(\Phi z)| > 2s = 2(q \cdot T)$ for all $z \in \mathbb{R}^n \backslash \{0\}$.

Proof: We first prove the claim C1: if $|\mathsf{supp}(\Phi z)| > 2s = 2(q \cdot T)$ for all $z \in \mathbb{R}^n \backslash \{0\}$ then all subsets of $2s$ columns of $Q_2^\top$ are full rank. Then by Lemma 1 and noting that by definition the null space of $Q_2^\top$ equals the column space of Φ, we have $x(0)$ is the unique solution.

Proof of C1 by contradiction: Suppose there exist $2s$ columns of $Q_2^\top$ that are linearly dependent. Then, there exists $E_0 \neq 0$ such that $Q_2^\top E_0 = 0$ where $|\mathsf{supp}(E_0)| \leq 2s$. Since the null space of $Q_2^\top$ equals the column space of Φ, there exists z such that $E_0 = \Phi z$ (i.e., E_0 is in the column space of Φ). Then, $|\mathsf{supp}(\Phi z)| = |\mathsf{supp}(E_0)| \leq 2s$ (contradiction). ■

The sufficient condition, provided in Lemma 2, for the existence of a unique solution to (10) is hard to check as it requires satisfiability of the condition for all $z \in \mathbb{R}^n \backslash \{0\}$. In the following Theorem, we prove an equivalent, yet simple-to-check, sufficient condition that only needs to be verified for the eigenvectors of A.

Theorem 1: Let $A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{p\times n}$. Assume that C is full rank, (A, C) is observable and A has n distinct positive eigenvalues such that $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_n$. Define:

- $s_i \triangleq |\mathsf{supp}(Cv_i)|$, where v_i is an eigenvector of A,
- $\mathcal{S} \triangleq \{s_1, s_2, \cdots, s_n\}$,
- For every $m \in \{2, \ldots, n\}$, let $\mathcal{S}_m$ be any subset of $\mathcal{S}$ with m elements, define $T_{\mathcal{S}_m} \triangleq \frac{(m-2)\cdot p + \min \mathcal{S}_m}{\max \mathcal{S}_m - 2q}$. Then T_m is such that $T_m > T_{\mathcal{S}_m}$ for all subsets $\mathcal{S}_m$, i.e. all subsets of m elements from the set $\mathcal{S}$.

Choose T such that $T \geq \max\{T_2, \cdots, T_n\}$. Then, the following are equivalent:

$$
\begin{aligned}
&(i)\ \forall v_i \in \mathbb{R}^n \text{ where } Av_i = \lambda_i v_i, \\
&\quad |\mathsf{supp}(Cv_i)| > 2q \\
&(ii)\ \forall v_i \in \mathbb{R}^n \text{ where } Av_i = \lambda_i v_i, \\
&\quad |\mathsf{supp}(\Phi v_i)| > 2q \cdot T \\
&(iii)\ \forall z \in \mathbb{R}^n \backslash \{0\}, |\mathsf{supp}(\Phi z)| > 2q \cdot T
\end{aligned}
$$

Proof: Interested readers are referred to the proof for Theorem 1 in our archived paper [16]. ■

Theorem 1 states that if the feedback system and the secure estimator are designed such that all the conditions in the theorem are satisfied, then our proposed secure estimator can guarantee accurate correction of q errors by checking the following very simple condition:

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

C. Number of Correctable Errors

Given that the set of attacked nodes can change over time and $e(k)$ satisfies $|\mathsf{supp}(e(k))| \leq q$ for all k, we prove in Proposition 1 (see below) that the maximum number of correctable errors (as defined in Definition 1) by our decoder is $\lceil p/2 - 1 \rceil$, where p is the number of measurements.

Proposition 1: Let $A_0 \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$ and assume that the pair (A_0, B) is controllable, C is full rank and each row of C is not identically zero. Then there exists a finite set $F \subset \mathbb{R}_+$ such that for any choice of n numbers $\lambda_1, \cdots, \lambda_n \in \mathbb{R}_+ \backslash F$ such that $0 < \lambda_1 < \cdots < \lambda_n$, there exists $G \in \mathbb{R}^{m\times n}$ such that:

- The eigenvalues of the closed-loop matrix $A\ (= A_0 + BG)$ are $\lambda_1, \cdots, \lambda_n$.
- If the pair (A, C) is observable, then the number of correctable errors for the pair (A, C) is maximal after $T = \max\{n, T^*\}$ time steps and is equal to $\lceil p/2 - 1 \rceil$, where T^* is the value of T from Theorem 1.

Proof: The proof for Proposition 4 in [13] shows that if the chosen poles $\lambda_1, \cdots, \lambda_n$ are distinct, positive and do not fall in some finite set F, then there is a choice of G such that the eigenvalues of $A\ (= A_0 + B)$ are exactly $\lambda_1, \cdots, \lambda_n$, and the corresponding eigenvectors v_i are such that $|\mathsf{supp}(Cv_i)| = p$. Thus, by Theorem 1, the number of correctable errors for (A, C) is $\lceil p/2 - 1 \rceil$. ■

In addition, recall that $E_{q,T}$ consists of the error vectors $e(0), \cdots, e(T-1)$ stacked vertically and our proofs for the existence of a unique solution to (10) are independent of how the individual error (nonzero) terms are distributed in the vector $E_{q,T}$. Thus, we can remove the assumption: $|\mathsf{supp}(e(k))| \leq q$ for all k, and allow $e(k)$ to appear in an arbitrary fashion, e.g. $|\mathsf{supp}(e(0))| = 2q$ and $|\mathsf{supp}(e(1))| = 0$, as long as $\sum_{k=0}^{T-1} |\mathsf{supp}(e(k))| \leq q \cdot T$, then our q-error-correcting decoder can still recover the true states. In other words, our proposed secure estimator can protect the system against more general attacks where the number of attacked sensors is not necessarily less than or equal to q at every time step.

IV. Combination of Secure Estimation and Kalman Filter

Consider the state estimation problem for the following LTI system under attack:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + e(k) + v(k), \end{aligned} \tag{11}$$

where x, y, u and e are as defined in (3); and v is a zero mean independent and identically distributed (i.i.d.) Gaussian measurement noise.

A KF can be used to estimate the states by modeling the attack signal as noise. More specifically, define a new measurement noise $\bar{v}(k) = e(k) + v(k)$ to give a new measurement equation $y(k) = Cx(k) + \bar{v}(k)$. A KF can then estimate the states from the inputs $u(k)$ and the corrupted measurements $y(k)$ [17]. One caveat with this method is that KFs assume zero mean and i.i.d. white Gaussian measurement noise, however, attack signals are usually erratic and may be poorly modeled by Gaussian processes [17], i.e., $e(k)$ and consequently, $\bar{v}(k)$ may not be Gaussian. Take Global Positioning System (GPS) spoofing attacks for example, attack signals are often structured to resemble normal GPS signals or can be genuine GPS signals captured elsewhere. When the system is subjected to attacks that are poorly modeled by Gaussian processes, it is reasonable to expect KFs to fail to recover the true states.

On the other hand, our proposed secure estimator does not assume the attack signal to follow any model, and therefore, it works for arbitrary and unbounded attacks. The only assumption is that the number of attacked sensors is sparse, i.e., less than $\lceil p/2 - 1 \rceil$. As the set of attacked sensors becomes less sparse, our secure estimator occasionally fails to recover the true states. Based on these observations, we propose to combine our secure estimator with a KF to improve its practical performance, as detailed in Algorithm 1.

The intuition is that the secure estimator acts as a pre-filter for the KF, so that $\tilde{v}(k)$ is close to a zero mean i.i.d. Gaussian process even when the true attack signal $e(k)$ is not. At most time steps k, the secure estimator perfectly recovers $e(k)$, i.e., $\hat{e}(k) = \hat{e}(k)$, hence $\tilde{v}(k) = v(k)$ and thus, is a zero mean Gaussian process. What happens when the secure estimator fails? (5) shows that the estimated state at time k, $\hat{x}(k)$, is independent from the estimated state at another time step $\hat{x}(l)$ ($l \neq k$). As a result, when the secure estimator fails, its estimation error, $e(k) - \hat{e}(k)$, appears to be quite random. Putting these together: $\tilde{v}(k) = e(k) - \hat{e}(k) + v(k)$ is closer to a zero mean i.i.d. white Gaussian process than $\bar{v}(k)$ (i.e., the corresponding measurement noise if a KF is applied directly to estimate the states), which improves the KF's performance. Finally, the *if* statement in Algorithm 1 ensures that the secure estimator always has access to T past measurements.

Algorithm 1 Combined secure estimator with KF

```
Initialize the KF
for each k do
    if k ≥ T then
        Estimate the attack signal at time k, ê(k), using secure estimator
    else
        Set ê(k) = 0
    end if
    Form a new measurement equation: ỹ(k) = Cx(k) + ṽ(k), where ỹ(k) = y(k) − ê(k) and ṽ(k) = e(k) − ê(k) + v(k)
    Apply standard KF using u and ỹ
end for
```

Next, we demonstrate the effectiveness of our proposed method through simulations of a UAV under two types of adversarial attacks, which also provides a realistic example illustrating the behaviors described in this section.

V. Numerical Examples

A. UAV Model

We consider a quadrotor with the following dynamics:

$$\begin{aligned} x(k+1) &= A_0 x(k) + Bu(k) + g \\ y(k) &= Cx(k) + e(k) + v(k), \end{aligned} \tag{12}$$

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, and v_i's are their corresponding

velocities. θ_x and θ_y are the pitch and roll angles respectively, and $\dot{\theta}_i$'s are their corresponding angular velocities. The input vector $u = [\theta_{r,x}, \theta_{r,y}, F]^T$, where $\theta_{r,i}$ 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 corrupted position measurements under attack e and measurement noise v. The constant vector g represents gravitational effects and can be dropped without loss of generality because we can always subtract it out in u. Further details about this model and its derivation can be found in [18]. Finally, the matrix C depends on the particular measurements taken in each example.

B. Decoder Design via Pole-Placement

We assume that the UAV uses the state feedback control law $u(k) = Gx(k)$[1], where G is the feedback matrix which can be designed. In this section, we show that we can design G to achieve our desired trade-off between the control performance and the secure estimation performance.

If the open loop pair (A_0, B) is controllable, then the closed loop poles can be placed anywhere in the complex plane by appropriate choice of G. First, we design a Linear Quadratic Regulator (LQR) and evaluate its secure estimation performance: we check the number of errors that the resulting secure estimator can correct by finding the maximum q for which $|\mathsf{supp}(Cv_i)| > q$ for all i. Figure 1 shows the results for a matrix $C \in \mathbb{R}^{5\times 10}$ (i.e., 5 measurements). Observe that $|\mathsf{supp}(Cv_i)| < p = 5$ for $i = 1, 2, 9, 10$ and furthermore, $|\mathsf{supp}(Cv_i)| = 1 > 0$ for $i = 9$ and 10, which means that the resulting secure decoder can correct zero errors! As shown in Figure 1, to improve the secure estimation performance, we perturb the closed-loop poles slightly until $|\mathsf{supp}(Cv_i)| = p$ for all i, i.e., we design a secure decoder that can achieve the maximum number of correctable errors within the limits of p (i.e., the number of measurements). By keeping the perturbations on the poles small, our final controller achieves both good control and estimation performances.

[1]In the GPS spoofing example, direct uncorrupted state measurements are not available. Therefore a KF is used to give estimated states which are then used for state feedback control.

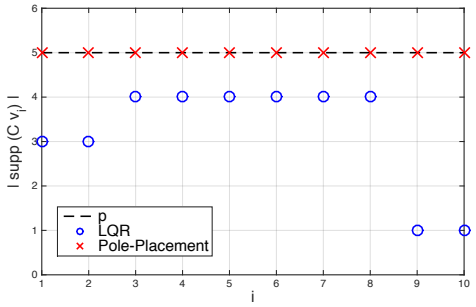

Fig. 1: $|\mathsf{supp}(Cv_i)|$ for all eigenvectors v_i of the closed-loop matrix A for 2 feedback controllers: a LQR and a controller designed by pole-placement. Black dashed line is at $p = 5$, i.e., the number of measurements.

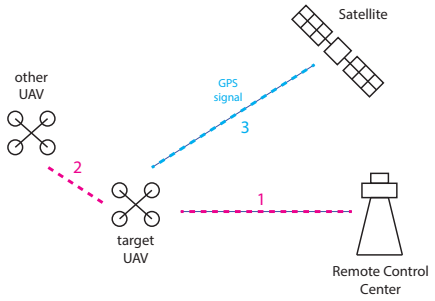

Fig. 2: Different communication channels that are subjected to cyber attacks.

C. UAV under Adversarial Cyber Attack

1) MITM Attack in Communication: In this section, we consider MITM attacks targeted at Channels 1 and 2 in Figure 2, where a malicious agent spoofs the information being sent and/or received over these channels. The goal of the remote control center or the other UAV is to accurately estimate the true flight path of the target UAV from compromised measurements. Note that the attack does not affect the actual path of the target UAV (as opposed to the GPS spoofing example later in this section).

Assume that the attacker aims to deceive the receiver that the target UAV is deviating in the x-direction, therefore she spoofs the x-position measurements by injecting a continuous and increasing attack signal in p_x. To make the estimation task even harder for the receiver, at each time step, the attacker injects a Gaussian noise to an additional randomly selected measurement, and the choice of this measurement changes over time.

In this example, we first demonstrate the effectiveness of our proposed decoder design using

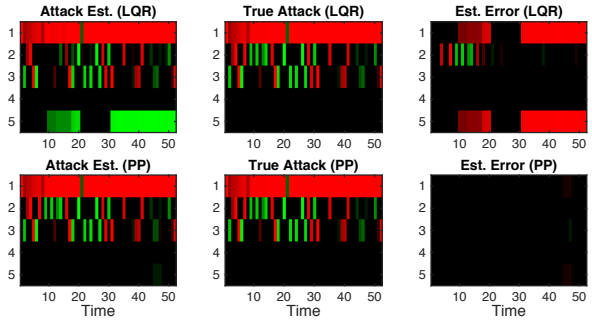

Fig. 3: Estimated attack signal, true attack signal and estimation error in the attack signal of the estimator (SE) with 2 different feedback controllers: LQR, controller designed via pole-placement (PP); with 5 measurements. Left column shows estimated attack signals. Middle column shows true attack signal. Right column shows estimation error. Each row corresponds to one type of measurement. Red pixels indicate positive values, green pixels are negative values and black indicates zero.

the pole-placement method by comparing the estimation performance of the decoder resulting from (1) a LQR controller and (2) a controller designed using pole-placement as described in the previous section. We then implement the latter controller, and compare the performance of three different estimation schemes: (1) KF only (KF), (2) secure estimator only (SE), and (3) secure estimator combined with KF (KF+SE).

Throughout this example, $y \in \mathbb{R}^5$ and the measurements include the x, y and z positions and two additional randomly selected states. Figure 3 compares the accuracy of the estimated attack signals by the LQ regulator (top) and the one designed via pole-placement (bottom). In each plot, one row corresponds to one sensor, and the first 3 rows are the x, y and z position measurements, respectively. The color of the pixel indicates the value of the signal or the estimation error. The middle plots show the true attack signal and they highlights three points: first, the attacked sensors change with time; second, the number of attacked sensors at each time step k is less or equal to 2; third, only position measurements are corrupted. The left plots show the estimated attacked signal by each decoder. It is easy to see that the decoder resulting from a feedback controller designed via pole-placement estimates the attack signal much more accurately. The right plots of this figure highlight this observation by explicitly showing the estimation error of the attack signal for each

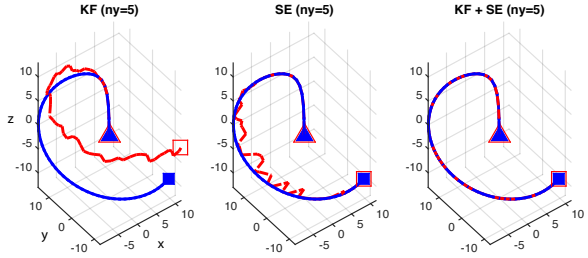

Fig. 4: Estimated UAV trajectory by three methods under MITM attack: KF only (KF), secure estimator only (SE), secure estimator with KF (KF+SE). Solid blue lines are the true UAV trajectories. They start from the blue triangle and end at the blue square. Red dashed lines represent estimated trajectories by each method, with 5 measurements.

measurement.

Figure 4 compares the estimated flight paths by all three methods: KF, SE and KF+SE. The UAV starts from the blue triangle and follows the solid blue line to land at the blue square. The estimated paths by each method are shown in red dashed lines. Observe that the KF fails to filter out the attack signal in the x-position measurements as the attack is highly non-Gaussian, and the estimated trajectory differs significantly from the true one. On the other hand, SE correctly estimates most portions of the trajectory and the final position of the vehicle, nevertheless it gives spontaneous errors. Finally the combined method KF+SE perfectly recovers the true path of the target UAV.

2) GPS Spoofing: In this section, we focus on adversarial attacks in the GPS navigation system (Channel 3 in Figure 2). Consider the scenario where a UAV uses a Linear Quadratic Gaussian (LQG) controller to follow a desired trajectory, $x_r(k)$. In other words, a KF uses corrupted and noisy measurements $y(k)$ to produce a state estimate $\hat{x}(k)$, which is then used for state feedback control: $u(k) = G(\hat{x}(k) - x_r(k))$, where G is the feedback matrix. Note that in the previous example (Section V-C .1), the feedback controller had access to uncompromised state measurements $x(k)$, therefore the true trajectory of the UAV is unaffected by attacks. In this example, however, the UAV uses estimated states $\hat{x}(k)$ for feedback control and path following. Therefore, if the measurements are corrupted and the state estimates are poor, then the UAV may deviate away from its desired path. Hence,

the goal of the UAV is to accurately follow its planned trajectory in the presence of cyber attacks.

Assume an attacker spoofs the GPS position measurements in order to deviate the UAV from its desired path. She injects a sinusoidal signal into the x position measurement, as well as a Gaussian noise to a randomly chosen position measurement at each time step.

In this example, we explore the effect of the number of sensor measurements on the secure estimation performance of two schemes: (a) KF only, (b) KF+SE. First, we assume that the UAV only uses GPS for navigation, i.e., 3 positional measurements. Figure 5 shows that KF completely fails to estimate the attack signal (KF, $n_y = 3$), consequently, the actual UAV trajectory (red dashed line) deviates significantly from its desired path (solid blue line). On the other hand, the combined method KF+SE's estimated attack signals are significantly more accurate, therefore the UAV can follow its planned path much more closely (Figures 5, KF + SE, $n_y = 3$). Recall from Proposition 1 that the maximum number of correctable errors for a system with p measurements is $\lceil p/2 - 1 \rceil$, which equals 1 in this case. However, at any time step k, there are at most 2 attacked sensors, which exceeds the above limit and explains the estimation error of the combined method KF+SE. Despite this small estimation error, KF+SE still outperforms the KF.

Next, we show the effect of increasing the number of measurements (n_y, or equivalently p) on the estimation performance and consequently, the UAV's path following performance. This can be achieved through sensor fusion. For example, autonomous UAVs often use IMUs in addition to GPS for navigation, the former provides additional measurements such as the UAV's velocities, pitch and roll angles. Figure 5 shows that increasing the number of measurements has no effect on the KF's estimation accuracy and hence, its path following ability. Even when 8 measurements are used the UAV equipped with a KF still fails to follow the desired trajectory. On the other hand, increasing the number of measurements improves the estimation performance of the secure estimator SE and consequently the performance of the combined scheme KF+SE.

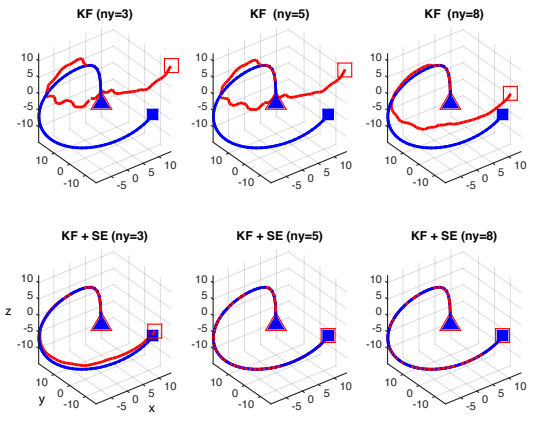

Fig. 5: Desired and actual UAV trajectory in different cases: KF and KF+SE, each using 3, 5 and 8 different measurements. Blue solid lines are the desired trajectory. Red dash lines are the actual UAV trajectory under adversarial attack.

Figure 5 shows that when 5 and 8 measurements are used, the UAV can follow its original planned path perfectly (KF + SE $n_y = 5$ and KF + SE $n_y = 8$).

Conclusion

In this paper, we consider the estimation problem for UAVs under adversarial cyber attack and propose a secure estimation based KF that is computationally efficient and makes no assumptions about the attack signal model. We demonstrate that our proposed secure estimator outperforms standard KF, using numerical examples of UAVs under adversarial cyber attacks. This is important not only for today's aviation system but also delivery systems with drones in the near future.

References

[1] "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.

[2] "Google project wing," http://www.theatlantic.com/technology/archive/2014/08/inside-googles-secret-drone-delivery-program/379306/?single_page=true, accessed: 2014-08-28.

[3] "Amazon Prime Air," http://www.amazon.com/b?node=8037720011.

[4] "Ascending Technologies," http://www.asctec.de/en/drone-uav/uav-uas-drone-powerline-infrastructure-inspection/.

[5] "UAS: the future of precision agriculture," http://www.croplife.com/equipment/precision-ag/uas-the-future-of-precision-agriculture/.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[15] D. Hayden, Y. H. Chang, J. Goncalves, and C. Tomlin, "Sparse network identifiability via compressed sensing," *Automatica*, vol. 52, 2016.

[16] Y. H. Chang, Q. Hu, and C. Tomlin, "Secure estimation based Kalman Filter for Cyber-Physical Systems against adversarial attacks," *arXiv:1512.03853v2*, 2015.

[17] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," *American Control Conference*, 2013.

[18] 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.

COPYRIGHT STATEMENT