On Analysis And Design Of Stealth-Resilient Control Systems

Shaunak D. Bopardikar

Alberto Speranzon

***Abstract*—The interest in cyber-physical system security has been growing exponentially in the last five years as the research community has realized that control loops embedded in complex systems can be compromised once attackers are capable to breach the cyber intrusion protection and detection systems. In this paper, we consider the class of attacks known as *stealth attacks* where the attacker can compromise information flow between all remotely located components of a closed-loop control system. We develop design strategies that can prevent or make stealth attacks difficult to be carried out. Our methods enhance a legacy system, so that stealth attacks can be detected and counteracted at the cost of an increased system complexity.**

I. Introduction

Recent years have witnessed a surge in research activity in the area of security for cyber-physical systems (CPSs). Typically, a CPS comprises of many devices, such as sensors, actuators, controllers, monitors, etc., that are both intra-connected to form local networks, and inter-connected to IT networks through ethernet and/or WiFi. CPSs are used in both, highly critical infrastructures such as power networks, water purification systems, airplanes, and in less critical infrastructure such as heating and lighting systems. Due to standardization, the need to lower costs and to have open system architectures, some of the core technologies have become common to both critical and non-critical systems, thereby increasing the risk of leveraging exploits and flaws for attacks. An attack on a CPS can have major physical consequences, ranging from significant energy waste in office buildings when HVAC system is tampered with, to catastrophic when instabilities are created in the control of nuclear power plants. Reports of attacks to CPSs are growing and they can be rather complex such as STUXNET [7] or less sophisticated [6], [14], but still leading to severe disruption/damages.

Although IT security has a relatively long history and a well established set of methods, tools and softwares to secure clients, servers and network devices, these typically do not apply in a straightforward manner to CPSs, as discussed in [4]. Although CPSs do comprise of devices that can have similar functionalities as standard PCs [3], patches, encryption and authentication methods might not be easily implementable. Indeed, these can introduce unwanted delays, require larger computation power than available or even require the system to be stopped to apply security fixes. Also, the possibility of attacks that do not have a direct counterpart in the digital world makes the cyber countermeasure rather ineffective, [4].

In the CPSs security research literature, authors have been considering various models of attacks. In [2] and [8], the authors have considered *false data injections* on static estimators. This attack is modeled as corruption of measurements that are used for state estimation. Conditions on the systems properties that prevent these attacks are derived. Mo and Sinopoli in [10] consider the false data injection attacks on a dynamical system for which they show that an attack exists if and only if the system dynamics has an unstable mode and the associated eigenvector satisfies a technical assumption. The same authors, in [9], analyze the effect of *replay attacks* on control systems. In a replay attack, the attacker records and plays back the same measurements while tampering with the control inputs. The authors assume that the measurements recorded are relative to the steady state of the system, and inspite of using classical failure detectors, it is still possible to carry out this attack. An ad-hoc method to increase the detectability of the attack is proposed. Smith, in [12], first introduced and demonstrated a *stealth/covert attack* on a closed-loop control system. In this scenario, the attacker has the capability to manipulate both, the input (actuators) and the output (measurements) of a closed loop control system so that the it is not possible to detect the attack from the output. Texeria et al. in [13] provide methods to change the system model so that a class of stealthy attacks on the actuators gets revealed. Dan and Sandberg in [5] consider stealth attacks on static linear systems and provide algorithms to secure measurements that require a maximum amount of attacker resources. Pasqualetti et al. [11] provide a more general framework to analyze several types of attacks on power systems and networks. In particular, general conditions for attack detection and identifiability for descriptor linear time-invariant systems are considered.

In this paper, we consider a legacy control system modeled as a discrete time linear time-invariant system, operating in closed loop. The measurements from the system are used by an observer/estimator to compute a state estimate that is then used by a controller to compute a control action. This is then applied back to the plant through actuators. We assume that an attacker has the capability to additively modify some/all signals in a stealthy fashion. In other words, an attacker seeks to inject additive signals to the control input that are coordinated with additive signals to the measurements and

Shaunak D. Bopardikar (BopardSD@utrc.utc.com) and Alberto Speranzon (SperanA@utrc.utc.com) are with United Technologies Research Center (UTRC). This work was supported by United Technologies Research Center, which is gratefully acknowledged. The authors would like to thank Nikola Trcka for critical comments on the manuscript.

possibly estimates, so that the measurements reaching the observer appear uncorrupted.

We first derive a necessary and sufficient condition based on the parameters of the legacy system, to ensure that there cannot exist any stealth attack. While similar conditions have been known to hold in the cases of power systems state identification [1], [2], [5], these were shown mainly for the estimation of static states. We extend the notion to address a coordinated attack on the actuators and the estimates as well. If this condition is not satisfied by the legacy system, i.e., if there exists a stealth attack, then we provide two techniques that enable us to modify the system so that the resulting system does not admit any stealth attacks. The first technique is based on optimal reconfiguration of the system to modify how any external input may affect the internal signals such as control inputs, measurements and estimates. As a particular case, this algorithm can provide a subset of signals that need to be protected to avoid stealth attacks. The second technique involves design of a secure augmented system, which could be implemented in software around the legacy system. This method essentially increases the total number of measurements of the system that need to be secured, so that the overall system admits no stealth attacks. This technique is similar in spirit to what proposed by Texeira et al. in [13] where the system matrices are modified to reveal a class of stealth attacks. In this paper, our technique is to design the system so that stealth attacks do not exist at all, and thus, extends their work.

This paper is organized as follows. Section II describes how the system and the attack are modeled. A necessary and sufficient condition for existence of stealth attacks is provided in Section III. Section IV describes two techniques that can be used to secure the system if it can be subjected to stealth attacks. These techniques are illustrated on a numerical example in Section V. Finally, Section VI summarizes our conclusions and directions for future research.

II. Problem Formulation

A. System Modeling

We consider the discrete-time version of a linear time invariant dynamical system given by

$$x_{k+1} = Ax_k + Bu_k\,, \tag{1}$$
$$y_k = Cx_k + Du_k\,, \tag{2}$$

where at every discrete time instant $k \in \mathbb{Z}_{\geq 0}$, the state $x_k \in \mathbb{R}^n$ and the measurement vector $y_k \in \mathbb{R}^m$. The control input $u_k \in \mathbb{R}^p$. For this system, the matrices, $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times p}, C \in \mathbb{R}^{m\times n}, D \in \mathbb{R}^{m\times p}$.

We will assume that the system is stabilizable and that the system operates in closed-loop, i.e., the control u_k is given by a state feedback law,

$$u_k = K\hat{x}_k,$$

where the matrix $K \in \mathbb{R}^{p\times n}$ is designed in a manner that the matrix $A+BK$ is Hurwitz and where $\hat{x}_k$ is the estimate of the state x_k. We thus assume that there is an observer that estimates the state from the measurements. The observer dynamics are given by

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k - L(y_k - C\hat{x}_k - Du_k)\,, \tag{3}$$

where the matrix $L \in \mathbb{R}^{n\times m}$ is designed in a manner that the matrix $A+LC$ is Hurwitz and with eigenvalues smaller (in absolute value) of those of $A+BK$.

B. Attack Modeling

Attacks can occur at different points of the closed loop system, depending on the various components.

In general, various cyber security mechanisms can be present in a cyber-physical system to prevent an attacker to tamper with the system. Such mechanisms can reduce the number of attack points and for each attack point, reduce the degree of freedom of an attack, namely the components and/or signals that can be attacked. For example, using authentication methods with different levels of security, certain sensors could be less or more difficult to attack. Using different encryption methods, signals to and from a controller can be less or more difficult to be corrupted, etc. Thus given a cyber-physical system, it is possible to rank the attacks points from likely to unlikely and for a given attack point, determine what could be the attack "progression", i.e., rank the number of devices/signals that could be corrupted as a function of time. This type of information could be determined from a cyber security audit performed by a red team.

In this paper, we will assume that the software modules that implement the control law and the observer are secure, i.e., it is unlikely for an attacker to modify system parameters and/or control/estimation algorithms. Additionally, we can identify various points of attack on the system, as shown in Figure 1. We assume that the attacker can independently inject additive signals into the control, estimation and the measurement vectors.

The closed loop system in Figure 1 is described by:

$$\begin{bmatrix} x_k^+ \\ \hat{x}_k^+ \end{bmatrix} = \begin{bmatrix} A & BK \\ -LC & A+BK+LC \end{bmatrix} \begin{bmatrix} x_k \\ \hat{x}_k \end{bmatrix} + \begin{bmatrix} BE \\ -LDE \end{bmatrix} v_k + \begin{bmatrix} 0 \\ -LF \end{bmatrix} w_k + \begin{bmatrix} BH \\ BH \end{bmatrix} s_k + \begin{bmatrix} 0 \\ LDG+BG \end{bmatrix} r_k \tag{4}$$

$$y_k^a = \begin{bmatrix} C & DK \end{bmatrix} \begin{bmatrix} x_k \\ \hat{x}_k \end{bmatrix} + DKHs_k + DEv_k + Fw_k\,, \tag{5}$$

where the attack signals are $v \in \mathbb{R}^e$, $w \in \mathbb{R}^f$, $r \in \mathbb{R}^g$ and $s \in \mathbb{R}^h$. The matrices E, F, G, H are of compatible dimensions and their column dimensions represent the degree of freedom of the attacker. For example, if the matrix $E \in \mathbb{R}^{p\times e}$ with $p \geq e$, then the attacker has e degrees of freedom to affect the p control signals.

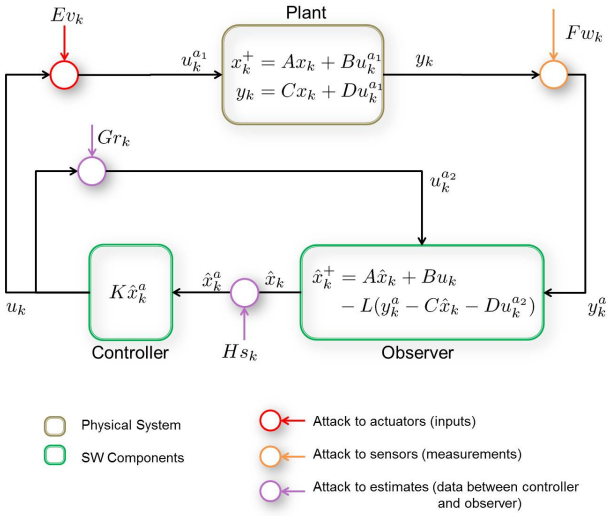

Fig. 1. A general closed loop controlled system with a plant a controller and observer. Attacks to the actuators and sensors are shown in red and orange, respectively, whereas an attack to the controller and observer signals is shown in purple.

The matrices E, F, G, H serve as a model of the cyber infrastructure, and how it introduces correlation between the attacker's inputs and their effect on the actual physical signals u, y and $\hat{x}$. These matrices could simply be the identity of appropriate dimensions, which would mean that the attacker can directly corrupt each individual component of u, y and $\hat{x}$. Without loss of generality, we assume that all four of these matrices are full column rank. Otherwise, they only provide redundant degrees of freedom to the attacker.

C. Problem Statement

We first formalize the notion of stealth attacks considered in this paper. Since the system is linear, we can assume $x_0 = \hat{x}_0 = 0$, without any loss of generality. Note that for the system (4)-(5) the output is in general a function of the attack vectors v_k, w_k, r_k, s_k, namely we have that $y_k(v_k, w_k, r_k, s_k)$.

Definition II.1 (Stealth Attack)**.** *Given a linear-time invariant system* (4)-(5) *a* stealth attack *exists if there exist attack vectors v_k, w_k, r_k, s_k, not all equal to zero, and a time instant $T > 0$ such that*

$$y_k(v_k, w_k, r_k, s_k) = y_k(0,0,0,0)\,, \quad \forall k \in \{0, 1, \ldots, T\}\,.$$

In other words, this means that there exists an action of the attacker on the internal signals of the system that is never seen by the measurement vector y.

Specifically, we will address the following problems:

1) Determine conditions on the system matrices so that there does not exist any stealth attack.
2) Suppose that a stealth attack exists, design techniques to modify the system so that the modified system does not admit stealth attacks.

III. Analysis

Let us define the following matrices:

$$\mathbf{A} = \begin{bmatrix} A & BK \\ -LC & A + BK + LC \end{bmatrix},$$
$$\mathbf{B} = \begin{bmatrix} BE & 0 & 0 & BH \\ -LDE & -LF & LDG + BG & BH \end{bmatrix},$$
$$\mathbf{C} = [C \quad DK]\,, \qquad \mathbf{D} = [DE \quad F \quad 0 \quad DKH]\,,$$

then the system can be rewritten as

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{a}_k \\ y_k^a &= \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{a}_k\,, \end{aligned} \tag{6}$$

with $\mathbf{x}_k = (x_k', \hat{x}_k')'$ is the state vector and $\mathbf{a}_k = (v_k', w_k', r_k', s_k')'$ is the attack vector.

We begin with the following necessary and sufficient condition for the existence of stealth attack.

Theorem III.1 (Stealth Attack Existence)**.** *There does not exist any stealth attack on the system* (4)-(5) *if and only if the matrix $\mathbf{D}$ has full column rank.*

Proof: From the definition of stealth attack, we have that $y_k^a = y_k$ for any $k \geq 0$. This holds true if and only if for any $T \geq 0$

$$\begin{aligned} &\mathbf{D}\mathbf{a}_0 = 0\,, \\ &\mathbf{C}\mathbf{B}\mathbf{a}_0 + \mathbf{D}\mathbf{a}_1 = 0\,, \\ &\quad\vdots \\ &\mathbf{C}\mathbf{A}^{T-1}\mathbf{B}\mathbf{a}_0 + \cdots + \mathbf{C}\mathbf{B}\mathbf{a}_{T-1} + \mathbf{D}\mathbf{a}_T = 0\,. \end{aligned}$$

Stacking these into a matrix equation, we obtain

$$\underbrace{\begin{bmatrix} \mathbf{D} & 0 & \cdots & 0 & 0 \\ \mathbf{C}\mathbf{B} & \mathbf{D} & \cdots & \cdots & 0 \\ \mathbf{C}\mathbf{A}\mathbf{B} & \mathbf{C}\mathbf{B} & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{C}\mathbf{A}^{T-1}\mathbf{B} & \mathbf{C}\mathbf{A}^{T-1}\mathbf{B} & \cdots & \mathbf{C}\mathbf{B} & \mathbf{D} \end{bmatrix}}_{\mathcal{M}_T} \underbrace{\begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_T \end{bmatrix}}_{\mathcal{A}_T} = 0\,.$$

From Definition II.1, for a stealth attack to exist, we must have the attack vectors $\mathcal{A}_T$ satisfying

$$\mathcal{M}_T \mathcal{A}_T = 0\,,$$

for $T \geq 0$. Therefore, a necessary and sufficient condition for a stealth attack to not exist is that the matrix $\mathcal{M}_T$ has full column rank, for all $T \geq 0$. Given the structure of $\mathcal{M}_T$, this holds if and only if the matrix $\mathbf{D}$ is full column rank. ■

It is important to note that from the structure of the matrix $\mathbf{D} = [DE \quad F \quad 0 \quad DKH]$ it is always possible for an attacker to generate a stealth attack, as the matrix $\mathbf{D}$ is not full column rank. This is because of the zero matrix, corresponding to the r signal. Thus, the attacker, through the signal r, can attack the system without changing the output. This means that we either need to assume that the cyber layer

enhances the protection of the control input computed by the controller which is communicated to the observer using an encrypted channel, or the data exchange occurs through memory, e.g., if the two software components (controller and observer) are co-located.

Remark III.1 ("Fat" $\mathbf{D}$ matrix)**.** *Since it is possible for an attacker to potentially have access to all measurements, control inputs and estimates, we could have $e+f+h > m$, in which case, the matrix $\mathbf{D}$ will be fat. For such cases, an attacker has a lot more freedom to inject a stealth attack, as shown in Section V.*

Remark III.2 (Relation with Geometric Control Literature)**.** *Akin to the analysis of stealth attacks considered in [13], one can characterize a class of stealthy attacks on the system* (6) *using geometric control theoretic techniques. Consider the following attack vector $\mathbf{a}_k \in \mathcal{N}(\mathbf{D}) \cap \mathcal{V}$, where $\mathcal{V}$ is the largest $(\mathbf{A}, Im(\mathbf{B}))$ controlled-invariant subspace in $\mathcal{N}(\mathbf{C})$, i.e., the null space of $\mathbf{C}$. Clearly, when $\mathbf{D}$ is full column rank, $\mathcal{N}(\mathbf{D})$ is empty, and so the above class of attacks does not exist.*

IV. Techniques to Prevent Stealth Attack

In this section, we describe how Theorem III.1 can be used as a design tool to secure a dynamical system against stealth attacks. We propose two approaches in which the system (6) may be secured against stealth attacks. The first method involves optimal allocation of countermeasures to secure selected signals, so that the resulting system does not admit any stealth attack. The second method involves augmenting the measurement vector with additional secure measurements so that the overall system does not admit any stealth attack.

A. Optimal Allocation of Countermeasures

In this approach, we assume that the matrices E, F, G, H can be reconfigured independently. In other words, each entry of the vector $\mathbf{u}$ can be forced to equal zero. However, this can be achieved with a certain cost. Let c_i denote the cost of securing against i-th attack parameter, for each $i \in \{1, \ldots, e+f+g+h\}$. Theorem III.1 tells us that if the rank of the matrix $X := \mathbf{D}$ is γ, then it suffices to secure $e+f+g+h-\gamma$ parameters so that the net cost is minimum. Equivalently, it suffices to find γ columns which are linearly independent and can be left *unsecured*, such that the sum of the costs of securing them is a maximum. More formally, let $\Pi \in \mathbb{R}^{(e+f+g+h)\times\gamma}$ denote a $0,1$ column selection matrix, i.e., each column of Π is a standard basis vector in $\mathbb{R}^{e+f+g+h}$. Then, the goal is to solve the following problem.

$$\begin{aligned} \max_{\Pi} \quad & c'\Pi\mathbf{1}\,, \\ \text{subject to} \quad & \text{rank}(X\Pi) = \gamma\,. \end{aligned} \tag{7}$$

As a first step, since $\mathbf{D}$ contains a block of zeros corresponding to the G term, it follows that it is imperative to secure all columns of G. Therefore, in what follows, we will only consider the E, F, H contributions. Further, while it appears that this problem is difficult to solve, especially due to the non-convex rank constraint, we will show there exists a greedy algorithm (Algorithm 1) that provides an optimal solution. This presentation assumes that without any loss of generality, the first γ columns of X are linearly independent. If not, then we can always re-order the columns accordingly. Further, for an index set $\mathcal{I} \subset \{1, \ldots, e+f+h\}$, let $X_{\mathcal{I}}$ denote the set of columns of X corresponding to the index set $\mathcal{I}$.

Algorithm 1 Greedy Securing

1: $\texttt{Opt} = \{1, 2, \ldots, \gamma\}$
2: $\Gamma = [c_1 \ c_2 \ \ldots \ c_\gamma]$
3: **for** $i = \gamma+1, \ldots, e+f$ **do**
4: $\quad \mathcal{E} = \text{RowEchelon}([X_{\texttt{Opt}} \ X_i])$
5: $\quad \Gamma = [\Gamma \ c_i]$
6: $\quad$ Let $\mathcal{B}$ denote the non-zero indices of $\mathcal{E}_{\gamma+1}$
7: $\quad$ Let $c_{\min} := \min\{\Gamma_{\mathcal{B}}\}, \mu := \arg\min\{\Gamma_{\mathcal{B}}\}$
8: $\quad$ **if** $\Gamma_{\gamma+1} > c_{\min}$ **then**
9: $\quad\quad$ Set $\Gamma_{\mathcal{B}(\mu)} = \Gamma_{\gamma+1}$
10: $\quad\quad$ Set $\texttt{Opt}(\mathcal{B}(\mu)) = i$
11: $\quad$ **end if**
12: $\quad$ Set $\Gamma = \Gamma_{1:\gamma}$
13: **end for**

Intuitively, this algorithm begins with an initial set of linearly independent columns of X. Then, it compares for each new column, whether it would be advantageous to include it within the existing set (in terms of increasing the total cost), or not. If the cost of the column exceeds the existing minimum, then we swap it with the column which has the minimum cost, include the new column in the set, and remove the existing minimum.

The following statement can be established.

Theorem IV.1 (Optimality of Algorithm 1)**.** *Given the optimization problem* (7)*, then following holds:*

1) *Algorithm 1 yields an optimal solution;*
2) *Algorithm 1 has a computational complexity of $O((e+f+h)m^3)$, where $e+f+h$ is the total number of independent attack parameters, and m is the number of measurements.*

Proof: Suppose that Algorithm 1 returns a solution $\texttt{Opt}$ which is not optimal. Let an optimal solution be given by S. Without any loss of generality, assume that $\texttt{Opt} \cap S = \emptyset$, because if it is not so, then we can remove the common elements and consider only the non-common ones from S in the rest of this proof. Since $X_{\texttt{Opt}}$ is full column rank by design, for each $s \in S$, there exists an $\bar{s} \in \texttt{Opt}$ such that $\bar{s} \in \mathcal{B}, c_{\bar{s}} = \min\{\Gamma_{\mathcal{B}}\}$ for $i = s$ in the for loop of Algorithm 1. Then, there are two possibilities:

1) If $c_s > c_{\bar{s}}$: Then, $\bar{s} \notin \texttt{Opt}$ due to steps 8 to 10 of Algorithm 1, which is a contradiction.

2) If $c_s \leq c_{\bar{s}}$: Since X_S must be full column rank due to Theorem III.1, for every $s \in S$, we conclude that c_s is less than or equal to that of a unique column $\bar{s}$ of $\texttt{Opt}$. This implies that the total cost of $\texttt{Opt}$ is at least equal to the total cost of S, which is a contradiction, since S is assumed to be optimal and $\texttt{Opt}$ is not.

This establishes the optimality of Algorithm 1. On the computational complexity, the row echelon computation is of $O(\gamma^3)$. The minimization step is of $O(\gamma \log(\gamma))$. This implies an overall complexity of $O((e+f+h)\gamma^3)$. Finally, the result follows since $\gamma \leq m$. ■

This approach is based on the fact that we can modify E, F, G, H, i.e., the manner in which any attacker input can affect the control, estimate and measurements. Systems for which this is infeasible or too expensive may be secured by an alternate approach described next.

B. Design of Augmented System

The second approach to securing a system is via the design an augmented system which expands the total number of measurements, by means of additional *secured* variables which measure directly the control input. We introduce the following secured measurements,

$$\begin{aligned} y^p_k &= D_p u^a_k = D_p K \hat{x}_k + D_p E v_k \,, \\ y^c_k &= C_c \hat{x}_k + C_c H s_k, \\ y^o_k &= D_o K \hat{x}_k + D_o G r_k + D_o K H s_k \,, \end{aligned}$$

where the set of measurements y^p are implemented at the plant side, y^c are implemented on the controller side, and y^o are implemented on the observer side. This set may be implemented as software modules within the three main components shown in Figure 1. We assume that this data is communicated between components over encrypted links so that it is difficult for the attacker to tamper with the system.

The above system along with (6) can be written in the standard matrix form with

$$\bar{\mathbf{C}} := \begin{bmatrix} C & DK \\ 0 & D_p K \\ 0 & C_c \\ 0 & D_o K \end{bmatrix}, \bar{\mathbf{D}} := \begin{bmatrix} DE & F & 0 & DKH \\ D_p E & 0 & 0 & 0 \\ 0 & 0 & 0 & C_c H \\ 0 & 0 & D_o G & D_o K H \end{bmatrix},$$

where the matrices to be designed are $D_p \in \mathbb{R}^{\tilde{m} \times p}, C_c \in \mathbb{R}^{\tilde{m}_c \times n}, D_o \in \mathbb{R}^{\tilde{m}_o \times p}$.

A simple design is described in the following result.

Theorem IV.2 (Augmented System)**.** *The augmented system designed with $D_p = E^\dagger$, $C_c = H^\dagger$, $D_o = G^\dagger$ where, the $\dagger$ represents the Moore-Penrose pseudo-inverse, is stealth-resilient.*

Proof: With this choice, the overall system has

$$\bar{\mathbf{D}} = \begin{bmatrix} DE & F & 0 & DKH \\ I_{e\times e} & 0_{e\times f} & 0_{e\times g} & 0_{e\times h} \\ 0_{h\times e} & 0_{h\times f} & 0_{h\times g} & I_{h\times h} \\ 0_{g\times e} & 0_{g\times f} & I_{g\times g} & G^\dagger K H \end{bmatrix},$$

which is full column rank, since F is assumed to be full column rank. Therefore, applying Theorem III.1 to the overall augmented system, the claim follows. ■

Remark IV.1 (Stealth Compensation)**.** *Both techniques proposed in this section lead to a modification of the legacy control system, which may be used to design a compensator for any (stealth) attack on the system in the following manner. Consider the residual $\xi_t := y^a_t$. Under no stealth attack, this residual is identically equal to 0. However, under attack, the residual will be non-zero. Since the modified system satisfies the conditions in Theorem III.1, we can write*

$$\hat{\mathbf{u}}_t = \mathbf{D}^\dagger_{mod,t} \xi_t.$$

which is the best (in the sense of least-squares) estimate for the vector of control attack $\mathbf{u}$, and $\mathbf{D}_{mod,t}$ is the modified system matrix. We can now compensate for the stealth attack as follows, since for every $k \geq 0$,

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k - \mathbf{B}\mathbf{D}^\dagger_{mod,k} \xi_k.$$

Note that the estimator for $\mathbf{u}_k$ is only the least squares solution, and will not guarantee that it cancels the effect of the stealth attack completely.

V. Numerical Example: Input Output Attack

In this section, we numerically demonstrate our results. For simplicity, we consider the case when $G = 0$ and $H = 0$. This is the case when only the actuators and the measurements are compromised. For this case, we considered the following system parameters:

$$A = \begin{bmatrix} 0.2080 & -0.0705 & 0.3765 & -0.3048 \\ -0.0504 & 0.6673 & -0.0770 & 0.0620 \\ -0.1818 & -0.0357 & 0.8382 & -0.3038 \\ 0.2994 & -0.0013 & -0.3647 & 0.8916 \end{bmatrix},$$

$$B = \begin{bmatrix} 2.2102 & -2.0235 & 0.4755 & -0.5273 \\ 0.9427 & 0.4696 & 0.0677 & -0.8228 \\ 2.1138 & -1.4192 & 2.3942 & -2.7286 \\ 3.9333 & -0.3943 & 0.8134 & -0.6077 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.7268 & -1.7720 & 0.3285 & 0.5364 \\ -2.2553 & 0.0266 & -0.9488 & 1.2590 \end{bmatrix},$$

$$D = \begin{bmatrix} 0.2588 & -1.9708 & -0.1163 & 1.4613 \\ 1.1028 & -0.3149 & 0.6507 & 1.0313 \end{bmatrix},$$

$$E = \begin{bmatrix} -0.2540 & 2.4125 & 1.0170 & -0.5178 \\ -1.3672 & -1.2000 & 0.2378 & 1.5906 \end{bmatrix}',$$

$$F = I_{2\times 2}\,.$$

The matrix $\mathbf{D} = [DE\ F]$ is clearly not full column rank, and therefore, there exists a stealth attack on this system. The effect of one such stealth attack can be visualized in the top two subfigures of Figure 2. The states x_i should have been equal to zero, but they show large deviations from zero, while the output remains at zero value. Note that the state evolution can be made arbitrarily large in magnitude by simply multiplying the value of the attack signal with a higher number.

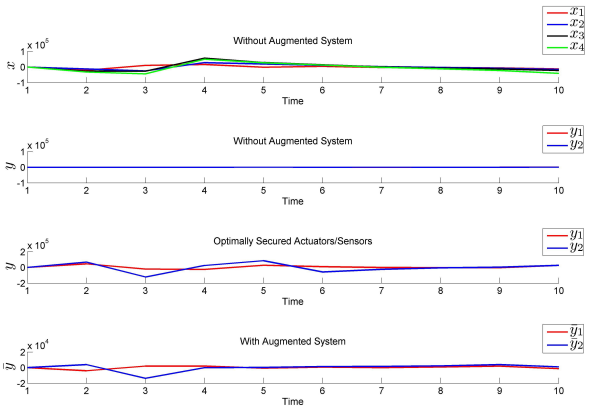

Fig. 2. Simulation results. The top two subfigures show the effect of a stealthy attack on the original system. The bottom subfigures shows how the augmented system can detect the attack.

A. Optimal Securing Approach

The rank of $\mathbf{D} = [DE\ F]$ is 2, which implies that we must secure at least $4 - 2 = 2$ degrees of freedom of the attacker to prevent stealth attacks. For this approach, we assumed the following costs for securing the actuators/sensors,

$$c = [0.3354 \quad 0.6797 \quad 0.1366 \quad 0.7212],$$

where the first two entries are the costs of securing the columns of E (actuator side), while the last two are the costs of securing the columns of F (sensor side). Applying Algorithm 1, we obtain that the second columns of E and F may be left unsecured. The optimal cost of securing the system is 0.4720 units. The third subfigure of Figure 2 shows the effect of the same stealth attack on the optimally secured system.

B. Augmented System Approach

The original system has only 2 measurements, while the attacker has 4 degrees of freedom. Therefore, the augmented system must introduce at least 2 secure measurements to ensure that there exists no stealth attack. The output of the two augmented measurements is shown in the bottom subfigure of Figure 2. The stealthy attack for the original system thus gets detected by the augmented system outputs.

In summary, the first technique requires securing the system against 2 degrees of freedom of the attacker, while the second technique requires us to secure 2 additional outputs. Either technique is clearly an advantage over the obvious solution for this problem, which is to secure all attacker degrees of freedom equal to 4.

VI. Conclusions

We addressed the problem of securing a legacy closed-loop cyber-physical control system against stealth attacks on the communication between system components. We first identified a condition on the system parameters that ensures that no stealth attacks exist. Then, we designed two techniques to secure the system based on either reconfiguring the system to modify the effect of an attack, or by deploying additional virtual but secure measurements. We demonstrated the techniques on a special case of input-output attacks numerically.

In future, we plan to extend our techniques to distributed networked control systems. We would also like to develop techniques which are based on a characterization of level of security against performance degradation of the system, instead of a worst-case approach.

References

[1] S. Bi and Y. J. Zhang. Defending mechanisms against false-data injection attacks in the power system state estimation. In *IEEE International Workshop on Smart Grid Communications and Networks, Houston, USA*, 2011.

[2] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye. Detecting false data injection attacks on DC state estimation. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK*, 2010.

[3] A. Cardenas, S. Amin, and S. S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd conference on hot topics in security (HOTSEC'08)*, 2008.

[4] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. S. Sastry. Challenges for securing cyber physical systems. In *Workshop on Future Directions in Cyber-physical Systems Security, DHS*, 2009.

[5] G. Dan and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *IEEE Int. Conf. on Smart Grid Communications*, 2010.

[6] F. Washkuch Jr. 'Dr. Chaos' gets seven more years in jail, 2005. Available online at: http://www.scmagazine.com/dr-chaos-gets-seven-more-years-in-jail/article/32757/.

[7] J. Leyden. Stuxnet 'a game changer for malware defence', 2010. Available online at: http://www.theregister.co.uk/2010/10/09/stuxnet_enisa_response/.

[8] Y. Liu, M. K. Reiter, and P. Ning. False data injection attacks against state estimation in electric power grids. In *ACM conference on Computer and communications security*, 2009.

[9] Y. Mo and B. Sinopoli. Secure control against replay attacks. In *Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing - Allerton House*, 2009.

[10] Y. Mo and B. Sinopoli. False data injection attacks in control systems. In *First Workshop on Secure Control Systems, CPS Week*, 2010.

[11] F. Pasqualetti, F. Dorfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transaction on Automatic and Control*, 2012. Conditionally accepted.

[12] R. Smith. A decoupled feedback structure for covertly appropriating networked control systems. In *18th International Federation of Automatic Control World Congress*, 2011.

[13] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson. Revealing stealthy attacks in control systems. In *50th Annual Allerton Conf. on Communication, Control and Comp.*, 2012.

[14] T. Wilson. Teenage hacker takes over Polish tram system, 2012. Available online at: http://www.darkreading.com/security/perimeter-security/208803765/teenage-hacker-takes-over-polish-tram-system.html.