Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Distributed Detection and Isolation of Covert Cyber Attacks for a Class of Interconnected Systems ⋆

Ahmad W. Al-Dabbagh * **Angelo Barboni** * **Thomas Parisini** **

* *Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK (e-mails: a.al-dabbagh@imperial.ac.uk and a.barboni16@imperial.ac.uk).*
** *Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK, Department of Engineering and Architecture, University of Trieste, 34127 Trieste, Italy, and KIOS Research and Innovation Center of Excellence, University of Cyprus, CY-1678 Nicosia, Cyprus (e-mail: t.parisini@imperial.ac.uk).*

Abstract: This paper deals with a topology for a class of interconnected systems, referred to as a highly interconnected system, consisting of interconnected plants and local controllers. We address the respective cyber attack surfaces as well as a design approach for detection and isolation of covert cyber attacks. For each pair of plant and controller, a cyber attack is implemented by a malicious agent, and its detection and isolation are achieved by associating the controller with two observers. These observers estimate the states of the plant, and compare the estimated states to determine if a neighbouring plant is under a covert cyber attack. The paper presents the modelling of the topology, the analysis of the covertness of cyber attacks, the design approach for the detection and isolation as well as a required existence condition. Simulation results are provided for the application of the design approach to interconnected pendula systems that are subject to a covert cyber attack.

Keywords: Highly interconnected system, covert cyber attack, distributed detection and isolation, large-scale system, networked control system.

1. INTRODUCTION

In engineering applications, interconnected control and monitoring systems are utilized to regulate processes that are physically coupled. This architecture of interconnected systems can become significantly large as distributed systems and processes are deployed and connected. It can also become more vulnerable to cyber attacks as more attack surfaces are introduced [1] . Thus, there is a growing need to account for cyber attack diagnosis in the context of large-scale interconnected systems.

In the literature, several research efforts attempt to capture control system topologies, or architectures that define the interconnection between plants and control and monitoring systems. For example, Pajic et al. (2011) present the control of a plant using a set of distributed and interconnected nodes, Al-Dabbagh and Chen (2016) and Al-Dabbagh (2019) show the control of a plant using a set of distributed and interconnected nodes as well as a centralized controller, Boem et al. (2019) address the distributed detection of faults in plants whose states are physically coupled using interconnected diagnosers, and Barboni et al. (2019) tackle the distributed detection of covert cyber attacks in plants whose states are physically coupled using interconnected local units consisting of controller and observer systems.

Also, several research efforts address cyber attacks in interconnected systems. For example, Dibaji et al. (2019) provide a survey of systems and control methods for cyber security in cyber-physical systems, Pasqualetti et al. (2015) discuss modelling and security analysis as well as monitoring design for cyber-physical security, Pasqualetti et al. (2013) design centralized and distributed detection and identification monitors for cyber attacks in cyber-physical systems, Teixeira et al. (2015) present a control framework for cyber security with modelling under different types of cyber attacks, Yang et al. (2019) study false data injection attacks in a wireless sensor network with distributed filtering, and Smith (2015) analyzes covert cyber attacks in networked control systems, where control commands of the controller and measurement signals of the plant are simultaneously manipulated to off-set the associated effect and render the attack undetectable.

⋆ The authors acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC), European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 739551 (KIOS CoE), Italian Ministry for Research in the Framework of the 2017 Program for Research Projects of National Interest (PRIN), Grant No. 2017YKXYXJ, and EPSRC Centre for Doctoral Training in High Performance Embedded and Distributed Systems (HiPEDS, Grant Reference EP/L016796/1).

[1] In this paper, we refer to communication channels subject to cyber attacks as attack surfaces.

In this paper, distributed detection and isolation of covert cyber attacks in large-scale interconnected systems are addressed, by significantly extending and generalizing the work presented by Barboni et al. (2019). The paper deals with a class of interconnected systems, referred to as a highly interconnected system, which consists of multiple interconnected plants and controllers. The controllers are associated with two observers to detect covert cyber attacks implemented by malicious agents. More specifically, the following contributions are provided:

- Modelling of the plant, controller, observers, and malicious agent is defined (Section 2).
- Implementation of covert cyber attacks is analyzed (Section 3).
- Design approach for detection and isolation of covert cyber attacks as well as a required existence condition are provided (Section 4).
- Simulation results for applying the design approach to interconnected pendula systems subject to a covert cyber attack are reported (Section 5).

The extension and generalization of the work presented by Barboni et al. (2019) are with regards to two directions: one is dealing with additional types of transmitted information in the interconnected system (hence, considering additional cyber attack surfaces), and the other is addressing the issue of false alarms in detection of cyber attacks, where a required existence condition for the detection as well as an isolation design approach are provided.

Subsequently, the following notations are used: $\mathbf{I}$ and $\mathbf{R}$ denote the identity matrix of appropriate dimensions and the set of real-valued numbers, respectively; $\mathcal{N}_i$ denotes the neighbourhood of a system i and $\mathcal{N}_i \setminus j$ denotes the set of all systems in $\mathcal{N}_i$ except system j; and $\|\cdot\|$ and $|\cdot|$ denote the Euclidean norm of vectors and the cardinality of a set, respectively.

2. HIGHLY INTERCONNECTED SYSTEMS

A highly interconnected system is defined to consist of multiple interconnected plants and local controllers, where the controllers can communicate with each other as well as with different plants for sensing and actuation. This captures the architecture of interconnected systems in some industrial settings, such as those with upstream and downstream processes in manufacturing facilities and water distribution networks. Consider a plant denoted by $\mathcal{P}_i$ and its local controller denoted by $\mathcal{K}_i$. The interconnection between the plant and controller as well as with other plants and controllers is depicted in Fig. 1, along with possible cyber attack surfaces introduced by the transmitted information between the plants and controllers.

The plant $\mathcal{P}_i$ under cyber attacks is modelled as a continuous-time linear time invariant (LTI) system as

$$\mathcal{P}_i \begin{cases} \dot{x}_i &= A_i^P x_i + B_i^P \tilde{u}_i + \sum_{j\in\mathcal{N}_i} A_{ij}^P x_j + \sum_{j\in\mathcal{N}_i} B_{ij}^P \tilde{u}_j^{P_i}, \\ y_i &= C_i^P x_i, \end{cases} \tag{1}$$

where $x_i \in \mathbf{R}^{n_i}$, $\tilde{u}_i \in \mathbf{R}^{m_i}$, $x_j \in \mathbf{R}^{n_j}$, $\tilde{u}_j^{P_i} \in \mathbf{R}^{m_j}$ and $y_i \in \mathbf{R}^{p_i}$ are its states and control commands, states of neighbouring plants and control commands of

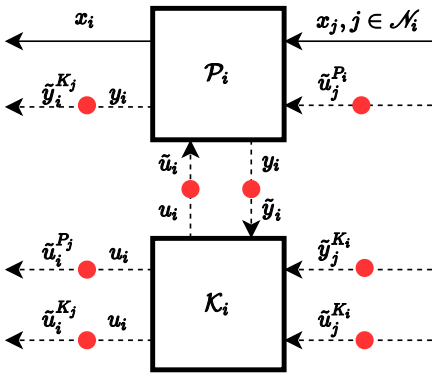

Fig. 1. Plant and controller, with physical coupling (solid arrows), transmitted information (dashed arrows), and cyber attack surfaces (red circles)

neighbouring controllers, and its measurement signals, respectively, and A_i^P, B_i^P, A_{ij}^P, B_{ij}^P, and C_i^P are system matrices of appropriate dimensions. Its local controller $\mathcal{K}_i$ under cyber attacks is modelled as a continuous-time LTI system as

$$\mathcal{K}_i \begin{cases} \dot{f}_i &= A_i^K f_i + B_i^K \tilde{y}_i + \sum_{j\in\mathcal{N}_i} B_{1_{ij}}^K \tilde{u}_j^{K_i} + \sum_{j\in\mathcal{N}_i} B_{2_{ij}}^K \tilde{y}_j^{K_i}, \\ u_i &= C_i^K f_i, \end{cases} \tag{2}$$

where $f_i \in \mathbf{R}^{q_i}$, $\tilde{u}_j^{K_i} \in \mathbf{R}^{m_j}$, and $\tilde{y}_j^{K_i} \in \mathbf{R}^{p_j}$ are its states, control commands of neighbouring controllers, and measurement signals of neighbouring plants, and A_i^K, B_i^K, $B_{1_{ij}}^K$, $B_{2_{ij}}^K$, and C_i^K are systems matrices of appropriate dimensions. As the transmitted information between the plants and controllers can be manipulated by cyber attacks, the transmitted information in (1) and (2) are modelled as

$$\begin{aligned} &\tilde{u}_i = u_i + \mu_i,\ \tilde{y}_i = y_i - \gamma_i,\ \tilde{u}_j^{P_i} = u_j + \mu_j^{P_i}, \\ &\tilde{u}_j^{K_i} = u_j + \mu_j^{K_i},\ \tilde{y}_j^{K_i} = y_j - \gamma_j^{K_i}, \end{aligned} \tag{3}$$

where μ_i and γ_i are bias signals injected into the transmitted information between $\mathcal{P}_i$ and $\mathcal{K}_i$, $\mu_j^{P_i}$ and $\mu_j^{K_i}$ are injected into the transmitted information from neighbouring controllers to $\mathcal{P}_i$ and $\mathcal{K}_i$, respectively, and $\gamma_j^{K_i}$ is injected into the transmitted information from neighbouring plants to $\mathcal{K}_i$. These cyber attacks are implemented by a malicious agent denoted by $\mathcal{A}_i$. Its aim is to perform covert cyber attacks by manipulating the control commands and offsetting the effect by simultaneously manipulating the measurement signals (namely, $\tilde{u}_i$, $\tilde{u}_j^{P_i}$, and $\tilde{y}_i$, respectively).

To detect the cyber attacks on plant $\mathcal{P}_i$, its local controller $\mathcal{K}_i$ is associated with two observers, similar to the approach in Barboni et al. (2019). The first observer is implemented based on an Unknown Input Observer (UIO) denoted by $\mathcal{O}_i^d$ and the second is implemented based on a Luenberger observer denoted by $\mathcal{O}_i^c$ (for further details on the design of UIO and Luenberger observer, refer to Chen et al. (1996) and Hwang et al. (2010)); together with $\mathcal{K}_i$, they form a local control and monitoring unit denoted by $\mathcal{CMU}_i$. The interconnection between the plant, controller, and observers and the malicious agent is depicted in Fig. 2.

By extending the modelling framework developed in Barboni et al. (2019), the malicious agent and observers are

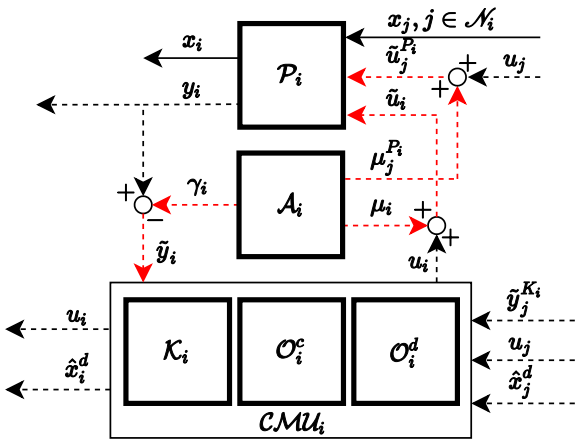

Fig. 2. Plant with the associated local control and monitoring unit and malicious agent, with physical coupling (solid arrows) and transmitted information with and without cyber attacks (red and black dashed arrows, respectively)

modelled. First, the malicious agent is modelled as a continuous-time LTI system as

$$\mathcal{A}_i \begin{cases} \dot{\tilde{x}}_i &= \tilde{A}_i^P \tilde{x}_i + \tilde{B}_i^P \mu_i + \sum_{j \in \mathcal{N}_i} \tilde{B}_{ij}^P \mu_j^{P_i}, \\ \gamma_i &= \tilde{C}_i^P \tilde{x}_i, \end{cases} \tag{4}$$

where $\tilde{x}_i \in \mathbf{R}^{n_i}$ is its states and $\tilde{A}_i^P$, $\tilde{B}_i^P$, $\tilde{B}_{ij}^P$, and $\tilde{C}_i^P$ are system matrices of appropriate dimensions, which capture the malicious agent's knowledge of A_i^P, B_i^P, B_{ij}^P, and C_i^P, respectively. The first observer $\mathcal{O}_i^d$ is modelled as a continuous-time LTI system as

$$\mathcal{O}_i^d \begin{cases} \dot{z}_i = & F_i z_i + T_i \left(B_i^P u_i + \sum_{j \in \mathcal{N}_i} B_{ij}^P u_j \right) \\ & + K_i \tilde{y}_i, \\ \hat{x}_i^d = & z_i + H_i \tilde{y}_i, \\ \hat{y}_i^d = & C_i^P \hat{x}_i^d, \end{cases} \tag{5}$$

where $z_i \in \mathbf{R}^{n_i}$ and $\hat{x}_i^d \in \mathbf{R}^{n_i}$ are its states and estimated states of $\mathcal{P}_i$, respectively, and F_i, T_i, K_i, and H_i are system matrices of appropriate dimensions (namely, they are the design variables, where F_i is selected to be Hurwitz). It should be noted that the physical coupling with neighbouring plants is not considered in estimating the states as they are decoupled by the observer (hence, the superscript d). The second observer $\mathcal{O}_i^c$ is modelled as a continuous-time LTI system as

$$\mathcal{O}_i^c \begin{cases} \dot{\hat{x}}_i^c = & (A_i^P - L_i C_i^P)\hat{x}_i^c + B_i^P u_i + \sum_{j \in \mathcal{N}_i} A_{ij}^P \hat{x}_j^d \\ & + \sum_{j \in \mathcal{N}_i} B_{ij}^P u_j + L_i \tilde{y}_i, \\ \hat{y}_i^c = & C_i^P \hat{x}_i^c, \end{cases} \tag{6}$$

where $\hat{x}_i^c \in \mathbf{R}^{n_i}$ and $\hat{y}_i^c \in \mathbf{R}^{p_i}$ are the estimated states and measurement signals of $\mathcal{P}_i$, respectively, and L_i is a system matrix of appropriate dimension (namely, it is the design variable, such that $A_i^P - L_i C_i^P$ is Hurwitz). It should be noted that the physical coupling with neighbouring plants is considered by using the estimated states $\hat{x}_j^d$ received from $\mathcal{O}_j^d$ and are coupled in the observer (hence, the superscript c). In the design of observers $\mathcal{O}_i^d$ and $\mathcal{O}_i^c$, the respective error and residuals under no cyber attacks are defined as

$$\begin{aligned} \epsilon_i^d &= x_i - \hat{x}_i^d, r_i^d = y_i - \hat{y}_i^d, \\ \epsilon_i^c &= x_i - \hat{x}_i^c, r_i^c = y_i - \hat{y}_i^c, \end{aligned} \tag{7}$$

and under covert cyber attacks are defined as

$$\begin{aligned} \tilde{\epsilon}_i^d &= x_i - \tilde{x}_i - \hat{x}_i^d, \tilde{r}_i^d = \tilde{y}_i - \hat{y}_i^d, \\ \tilde{\epsilon}_i^c &= x_i - \tilde{x}_i - \hat{x}_i^c, \tilde{r}_i^c = \tilde{y}_i - \hat{y}_i^c. \end{aligned} \tag{8}$$

Also, the following assumptions should be noted:

- Transmitted information between the controller and observers of control and monitoring unit $\mathcal{CMU}_i$ and those of other control and monitoring units is not subject to covert cyber attacks (namely, $\tilde{u}_j^{K_i} = u_j$ and $\hat{x}_j^d$ cannot be manipulated).
- Model of controller $\mathcal{K}_i$ in (2) is included for the sake of presentation and completeness, but is not designed in this paper. Therefore, covert cyber attacks on transmitted measurement signals from neighbouring plants (namely, $\tilde{y}_j^{K_i}$) are not considered in this paper.

3. ANALYSIS OF COVERT CYBER ATTACKS

In the highly interconnected system and with respect to plant $\mathcal{P}_i$, as depicted in Fig. 2, the objective of the malicious agent $\mathcal{A}_i$ in (4) is to simultaneously inject attack signals μ_i and $\mu_j^{P_i}$ into the control commands and γ_i into the measurement signals according to the definition of the transmitted information in (3). Its aim is to change the *real* measurement signals of $\mathcal{P}_i$ (namely, y_i) while remaining undetectable by the two observers $\mathcal{O}_i^d$ in (5) and $\mathcal{O}_i^c$ in (6) (hence, a covert cyber attack). With the above modelling framework and extending the results in Barboni et al. (2019), $\mathcal{A}_i$ indeed remains undetectable, as analyzed in the following three propositions.

Assumption 1. Malicious agent $\mathcal{A}_i$ has full knowledge of the system matrices of plant $\mathcal{P}_i$. That is, $\tilde{A}_i^P = A_i^P$, $\tilde{B}_i^P = B_i^P$, $\tilde{B}_{ij}^P = B_{ij}^P$, and $\tilde{C}_i^P = C_i^P$. In addition, it implements covert cyber attacks by injecting attack signals μ_i, $\mu_j^{P_i}$, and γ_i according to (3).

Proposition 2. Under Assumption 1, when agent $\mathcal{A}_i$ implements a covert cyber attack, the asymptotic response of $\mathcal{P}_i$ is the same as that with no cyber attack. Also, specifying $\tilde{x}_i(T_a) = 0$ at the initial time of the cyber attack results in the cyber attack being covert for all times.

Proof. First, prior to implementing a covert cyber attack for $t < T_a$, $\tilde{y}_i = y_i$. Then, at $t \geq T_a$, the response of $\mathcal{P}_i$ while specifying $\tilde{A}_i^P = A_i^P$, $\tilde{B}_i^P = B_i^P$, $\tilde{B}_{ij}^P = B_{ij}^P$, and $\tilde{C}_i^P = C_i^P$ can be derived as

$$\begin{aligned} \tilde{y}_i(t) =& C_i^P e^{A_i^P (t - T_a)} (x_i(T_a) - \tilde{x}_i(T_a)) + C_i^P \int_{T_a}^{t} e^{A_i^P (t-\tau)} \\ & \left[B_i^P u_i(\tau) + \sum_{j \in \mathcal{N}_i} A_{ij}^P x_j(\tau) + \sum_{j \in \mathcal{N}_i} B_{ij}^P u_j(\tau) \right] d\tau. \end{aligned} \tag{9}$$

The asymptotic responses of (9) is the same as that of $\mathcal{P}_i$ under no covert cyber attack. With specifying $\tilde{x}_i(T_a) = 0$, there is no discontinuity in the response, and therefore the cyber attack remains covert for all times. □

Proposition 3. Under Assumption 1 and suppose that $K_i = K_i^{(1)} + K_i^{(2)}$ in (5), when agent $\mathcal{A}_i$ implements a covert cyber attack, the *real* error ϵ_i^d in (7) is characterized as

$$\begin{aligned}\dot{\epsilon}_i^d =& F_i\epsilon_i^d + H_iC_i^PA_i^P\tilde{x}_i + B_i^P\mu_i + \sum_{j\in\mathcal{N}_i} B_{ij}^P\mu_j^{P_i} \\ &+ K_i^{(1)}\gamma_i\end{aligned} \tag{10}$$

and the cyber attack is undetectable by observer $\mathcal{O}_i^d$.

Proof. Consider $\dot{\epsilon}_i^d = \dot{x}_i - \dot{z}_i - H_i\dot{\tilde{y}}_i$. After algebraic operations, it is determined that

$$\begin{aligned}\dot{\epsilon}_i^d =& \left(\bar{A} - K_i^{(1)}C_i^P\right)\epsilon_i^d + \left[\left(\mathbf{I} - H_iC_i^P\right) - T_i\right]B_i^Pu_i \\ &+ B_i^P\mu_i + \left[\left(\mathbf{I} - H_iC_i^P\right) - T_i\right]\sum_{j\in\mathcal{N}_i} B_{ij}^Pu_j \\ &+ \sum_{j\in\mathcal{N}_i} B_{ij}^P\mu_j^{P_i} + H_iC_i^PA_i^P\tilde{x}_i \\ &+ \left[-\left(\bar{A} - K_i^{(1)}C_i^P\right)H_i + K_i\right]\gamma_i \\ &+ \left(\mathbf{I} - H_iC_i^P\right)\sum_{j\in\mathcal{N}_i} A_{ij}^Px_j + \left(\bar{A} - K_i^{(1)}C_i^P - F_i\right)z_i \\ &+ \left[\left(\bar{A} - K_i^{(1)}C_i^P\right)H_i - K_i^{(2)}\right]y_i,\end{aligned}$$

where $\bar{A} = A_i^P - H_iC_i^PA_i^P$. By specifying the conditions in Chen et al. (1996), the result in (10) is obtained. Then, consider $\tilde{\epsilon}_i^d$ in (8) that characterizes the error at $\mathcal{O}_i^d$, and using the result in (10), it can be found that

$$\dot{\tilde{\epsilon}}_i^d = F_i\tilde{\epsilon}_i^d.$$

This is the same as that under no cyber attack (namely, $\dot{\epsilon}_i^d = F_i\epsilon_i^d$), with system matrix F_i being Hurwitz by design. Thus, the cyber attack is undetectable by $\mathcal{O}_i^d$. □

Proposition 4. Under Assumption 1, when agent $\mathcal{A}_i$ implements a covert cyber attack, the *real* error ϵ_i^c in (7) is characterized as

$$\begin{aligned}\dot{\epsilon}_i^c =& (A_i^P - L_iC_i^P)\epsilon_i^c + \sum_{j\in\mathcal{N}_i} A_{ij}^P\epsilon_j^d + \sum_{j\in\mathcal{N}_i} B_{ij}^P\mu_j^{P_i} \\ &+ B_i^P\mu_i + L_i\gamma_i\end{aligned}$$

and the cyber attack is undetectable by observer $\mathcal{O}_i^c$.

Proof. The proof follows the same logic as that of Proposition 3. □

4. DETECTION AND ISOLATION OF COVERT CYBER ATTACKS

As demonstrated in Propositions 3 and 4, covert cyber attacks implemented by malicious agent $\mathcal{A}_i$ on plant $\mathcal{P}_i$ are undetected by both observers $\mathcal{O}_i^d$ and $\mathcal{O}_i^c$ of control and monitoring unit $\mathcal{CMU}_i$. However, following the same detection logic proposed in Barboni et al. (2019), neighbouring control and monitoring units $\mathcal{CMU}_j$ for $j \in \mathcal{N}_i$ can detect a covert cyber attack on $\mathcal{P}_i$. The detection logic is implemented by comparing the estimated states $\hat{x}_j^d$ of $\mathcal{O}_j^d$ and $\hat{x}_j^c$ of $\mathcal{O}_j^c$. In the absence of a covert cyber attack, the two estimated states are identical; and in its presence, they are different. Based on this comparison, neighbouring control and monitoring units $\mathcal{CMU}_j$ transmit alarm signals $\alpha_j = \{0, 1\}$ to $\mathcal{CMU}_i$ as well as their other neighbouring control and monitoring units. The alarm values $\alpha_j = 0$ and $\alpha_j = 1$ indicate the absence and presence of a covert cyber attack, respectively, on a plant associated with a neighbouring control and monitoring unit. Thus, for $\mathcal{CMU}_i$, when $\sum_{j\in\mathcal{N}_i}\alpha_j = |\mathcal{N}_i|$, it decides that $\mathcal{P}_i$ is under a covert cyber attack.

However, depending on the connectivity of the plants, the detection logic can lead to false alarms. Specifically, a control and monitoring unit can decide that its plant is under a covert cyber attack while it is not the case. The following proposition characterizes a required existence condition to prevent false alarms when implementing the detection logic.

Assumption 5. For any plant $\mathcal{P}_i$ under a covert cyber attack. Neighbouring plants $\mathcal{P}_j$ for $j \in \mathcal{N}_i$ are not under covert cyber attacks.

Proposition 6. Under Assumption 5 and for any plant $\mathcal{P}_i$ under a covert cyber attack, the detection logic provides no false alarms if one of the following conditions holds:

1. All plants $\mathcal{P}_j$ have no neighbouring plants except $\mathcal{P}_i$.
2. For any plant $\mathcal{P}_k$, where $k \in \mathcal{N}_j \setminus i$, there is at least one plant $\mathcal{P}_s$ such that $s \in \mathcal{N}_k$ and $s \notin \mathcal{N}_i$.

Proof. For plant $\mathcal{P}_i$ under a covert cyber attack, its $\mathcal{CMU}_i$ will correctly detect the cyber attack based on the design approach (namely, using the received alarm values). According to the proofs of Propositions 3 and 4, the error associated with its observers will converge to zero and the estimated states of both observers will be identical. For condition 1, $\mathcal{CMU}_i$ will transmit $\alpha_i = 0$, and $\mathcal{CMU}_j$ for $j \in \mathcal{N}_i$ will correctly decide that their plants are not under covert cyber attacks. For condition 2, any plant $\mathcal{P}_k$ having a neighbouring plant $\mathcal{P}_s$ such that $s \notin \mathcal{N}_i$ will receive $\alpha_s = 0$. Then, its $\mathcal{CMU}_k$ correctly decides that its plant is not under a covert cyber attack (namely, since $\sum_{s\in\mathcal{N}_k}\alpha_s \neq |\mathcal{N}_k|$). Thus, for the connectivity of the plants satisfying either conditions 1 and 2, the detection logic provides no false alarms. □

Remark 7. With condition 2 in Proposition 6, the assumption in Barboni et al. (2019) of only having a single plant in the overall interconnected system to be under a covert cyber attack is no longer necessary.

To improve the detection logic, such that there would be no false alarms regardless of the connectivity of the plants, while still satisfying Assumption 5, each control and monitoring unit can use an additional bank of observers. Each observer denoted by $\mathcal{O}_i^{d,j}$ is based on a UIO and is designed such that physical coupling from all neighbouring plants $\mathcal{P}_k$ for $k \in \mathcal{N}_i \setminus j$ is not considered, expect that from a *single* neighbouring plant $\mathcal{P}_j$. Consider re-defining the model of plant $\mathcal{P}_i$ in (1) as

$$\mathcal{P}_i \begin{cases} \dot{x}_i = & A_i^P x_i + B_i^P \tilde{u}_i + A_{ij}^P x_j + \sum_{k \in \mathcal{N}_i \setminus j} A_{ik}^P x_k \\ & + \sum_{j \in \mathcal{N}_i} B_{ij}^P \tilde{u}_j^{P_i}, \\ y_i = & C_i^P x_i. \end{cases}$$

Its observer denoted by $\mathcal{O}_i^{d,j}$ is modelled as

$$\mathcal{O}_i^{d,j} \begin{cases} \dot{z}_i^j = & F_i^j z_i^j + T_i^j \Big(B_i^P u_i + \sum_{j \in \mathcal{N}_i} B_{ij}^P u_j \\ & + A_{ij}^P \hat{x}_j^d \Big) + K_i^j \tilde{y}_i, \\ \hat{x}_i^{d,j} = & z_i^j + H_i^j \tilde{y}_i, \end{cases} \tag{11}$$

where the specification of the states and system matrices are similar to those in (5). The analysis of a covert cyber attack on $\mathcal{P}_i$ is provided in the following proposition.

Proposition 8. Under Assumptions 1 and 5, and suppose that $K_i^j = K_i^{j(1)} + K_i^{j(2)}$ in (11), when agent $\mathcal{A}_i$ implements a covert cyber attack, the *real* error $\epsilon_i^{d,j} = x_i - \hat{x}_i^{d,j}$ is characterized by

$$\dot{\epsilon}_i^{d,j} = F_i^j \epsilon_i^{d,j} + H_i^j C_i^P A_i^P \tilde{x}_i + B_i^P \mu_i + \sum_{j \in \mathcal{N}_i} B_{ij}^P \mu_j^{P_i} + K_i^{j(1)} \gamma_i$$

and the cyber attack is undetectable by $\mathcal{O}_i^{d,j}$, but is detectable by $\mathcal{O}_j^{d,i}$ for $j \in \mathcal{N}_i$.

Proof. The proof follows the same logic as that of Proposition 3. □

The detection logic is therefore modified by having neighbouring $\mathcal{CMU}_j$ for $j \in \mathcal{N}_i$ implement a comparison of the estimated states $\hat{x}_j^d$ and $\hat{x}_j^{d,i}$ for $i \in \mathcal{N}_j$. In the absence of a covert cyber attack on $\mathcal{P}_i$, the two estimated states are identical; and in its presence, they are different. Thus, using an additional bank of observers, each control and monitoring unit can precisely distinguish which neighbouring plant is under a covert cyber attack. The control and monitoring units therefore transmit alarm signals $\alpha_j^i = \{0, 1\}$ that are specific to $\mathcal{P}_i$. It should be noted that this design approach is only feasible when the conditions in Chen et al. (1996) are satisfied. Although this introduces additional restrictions, if it is possible to design such an additional bank of observers, both detection and isolation of covert cyber attacks in the highly interconnected system can be achieved.

5. SIMULATION

The design approach for detection of covert cyber attacks is applied to a system of six interconnected pendula, where each is denoted by $\mathcal{P}_i$ for $i = \{1, \ldots, 6\}$. The interconnection between the pendula is defined according to the following neighbourhood sets: $\mathcal{N}_1 = \{2, 3\}$, $\mathcal{N}_2 = \{1, 4, 6\}$, $\mathcal{N}_3 = \{1, 4, 5\}$, $\mathcal{N}_4 = \{2, 3, 6\}$, $\mathcal{N}_5 = \{3, 6\}$, and $\mathcal{N}_6 = \{2, 4, 5\}$. The dynamics of the i-th pendulum system is defined according to Šiljak (1978) as

$$\dot{x}_i = \begin{bmatrix} 0 & 1 \\ -\frac{g}{l_i} - \sum_{j \in \mathcal{N}_i} \frac{k_j a_j^2}{m_j l_j^2} & 0 \end{bmatrix} x_i + \begin{bmatrix} 0 \\ -\frac{1}{m_i l_i^2} \end{bmatrix} u_i + \sum_{j \in \mathcal{N}_i} \begin{bmatrix} 0 & 0 \\ \frac{k_i a_i^2}{m_i l_i^2} & 0 \end{bmatrix} x_j + \sum_{j \in \mathcal{N}_i} \begin{bmatrix} 0 \\ -\frac{1}{m_j l_j^2} \end{bmatrix} u_{ij},$$

where $x_i = [\theta_i \ \dot{\theta}_i]^\top$ is composed of its angle denoted by θ_i and angular velocity denoted by $\dot{\theta}_i$, and l_i, k_i, a_i, and m_i are its length, elastic constant for interconnecting springs, distance from the fulcrum of attaching the springs, and mass of the oscillating weight. In the simulation, it is assumed that measurements of both θ_i and $\dot{\theta}_i$ are available and the state-space representation is discretized with a sampling time of 1 millisecond. In addition, a local controller is implemented for each pendulum system using state feedback (namely, with using $\hat{x}_i^d$ to follow a specified sinusoidal reference signal for the position).

Further, a malicious agent is utilized to implement a covert cyber attack on pendulum system $\mathcal{P}_3$, which has $\mathcal{P}_1$, $\mathcal{P}_4$, and $\mathcal{P}_5$ in its neighbourhood. Its aim is to make the pendulum follow a malicious reference signal defined by a sine wave of different frequencies and amplitudes. It implements this by injecting attack signals μ_3, $\mu_4^{P_3}$, and γ_3 during time interval $[10, 20]$ seconds. The trajectories of the angle θ_i for $i \in \{1, 3\}$ are presented in Fig. 3. As can be observed, for pendulum system $\mathcal{P}_1$, the trajectories of the angle (namely, both θ_1 and $\tilde{\theta}_1$) follow that of the reference (namely, $\theta_{1,\text{ref}}$). A similar behaviour is obtained for pendula systems $\mathcal{P}_2$, $\mathcal{P}_4$, $\mathcal{P}_5$, and $\mathcal{P}_6$. However, for pendulum system $\mathcal{P}_3$ that is under a covert cyber attack, the trajectories of the actual angle θ_3 significantly deviates from that of the reference $\theta_{3,\text{ref}}$ and the received angle $\tilde{\theta}_3$ by the control and monitoring unit.

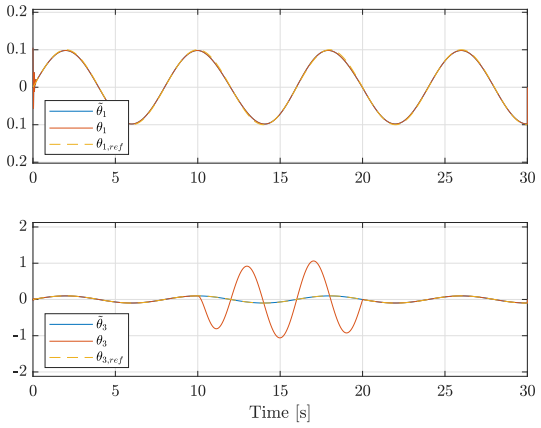

Fig. 3. Trajectories of the angle of pendula systems $\mathcal{P}_1$ (top) and $\mathcal{P}_3$ (bottom)

By implementing the proposed detection logic, each control and monitoring unit $\mathcal{CMU}_i$ of pendulum system $\mathcal{P}_i$ for $i \in \{1, \ldots, 6\}$ compares its estimated states $\hat{x}_i^d$ and $\hat{x}_i^c$. When the difference denoted by $\| \delta_i \|$ exceeds a specified threshold (namely, after large transients in the beginning of the simulation), it transmits an alarm value $\alpha_i = 1$ to its neighbours. The threshold is determined to be slightly larger than the steady-state peak value

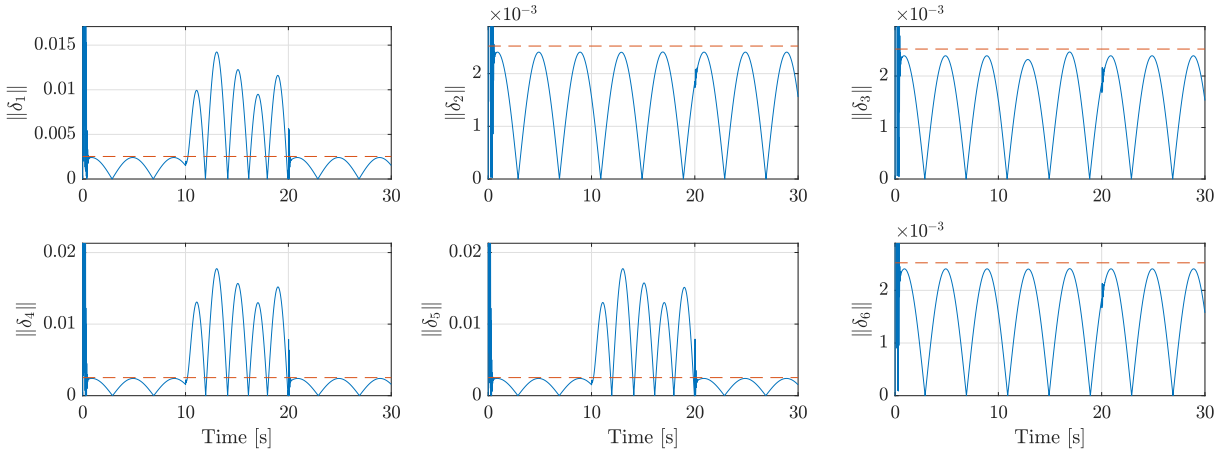

Fig. 4. Difference in the estimated states for each pendulum system $\mathcal{P}_i$ for $i \in \{1, \ldots, 6\}$, where $\delta_i = \hat{x}_i^d - \hat{x}_i^c$

under normal behaviour (namely, under no covert cyber attacks). It should be noted that the specification of a tight threshold is crucial in preventing false alarms, but is out of the scope of this paper. In addition, the connectivity of the pendula systems satisfies condition 2 of Proposition 6. Thus, no false alarms are introduced and the use of the additional bank of observers for isolation of cyber attacks is not required.

The trajectories of the computed difference in the estimated states are presented in Fig. 4. As can be observed, for pendulum system $\mathcal{P}_3$, the difference does not exceed the threshold, and the cyber attack is therefore covert. However, for all neighbouring pendula systems $\mathcal{P}_1$, $\mathcal{P}_4$, and $\mathcal{P}_5$, the difference exceeds the threshold during the time of the covert cyber attack. Consequently, alarm signals $\alpha_1 = \alpha_4 = \alpha_5 = 1$. Since $\mathcal{CMU}_3$ receives those alarm values and $\sum_{j\in\{1,4,5\}} \alpha_j = |\mathcal{N}_3| = 3$, it decides that $\mathcal{P}_3$ is under a covert cyber attack.

6. CONCLUSION

In this paper, a topology for a highly interconnected system is addressed, along with detection and isolation of covert cyber attacks. The paper provided the modelling of the topology, the analysis of covert cyber attacks, and the design approach for their distributed detection and isolation. Simulation results are also provided to demonstrate the effectiveness of the proposed detection approach. For future research, the following challenges are interesting: (i) considering uncertainties in knowledge of the plants by malicious agents as well as in transmitted information, (ii) determining further approaches for detection and isolation of covert cyber attacks, along with required existence conditions, (iii) designing controllers and observers in highly interconnect systems in a discrete-time framework, and (iv) addressing more complex applications and considering cyber attacks between control and monitoring units.

REFERENCES

Al-Dabbagh, A.W. (2019). Design of a wireless control system with unreliable nodes and communication links. *IEEE Transactions on Cybernetics*, 49(1), 315–327.

Al-Dabbagh, A.W. and Chen, T. (2016). Design considerations for wireless networked control systems. *IEEE Transactions on Industrial Electronics*, 63(9), 5547–5557.

Barboni, A., Rezaee, H., Boem, F., and Parisini, T. (2019). Distributed detection of covert attacks for interconnected systems. In *Proceedings of the 2019 18th European Control Conference*, 2240–2245. Napoli, Italy.

Boem, F., Carli, R., Farina, M., Ferrari-Trecate, G., and Parisini, T. (2019). Distributed fault detection for interconnected large-scale systems: A scalable plug & play approach. *IEEE Transactions on Control of Network Systems*, 6(2), 800–811.

Chen, J., Patton, R.J., and Zhang, H.Y. (1996). Design of unknown input observers and robust fault detection filters. *International Journal of control*, 63(1), 85–105.

Dibaji, S.M., Pirani, M., Flamholz, D.B., Annaswamy, A.M., Johansson, K.H., and Chakrabortty, A. (2019). A systems and control perspective of CPS security. *Annual Reviews in Control*, 47, 394–411.

Hwang, I., Kim, S., Kim, Y., and Seah, C.E. (2010). A survey of fault detection, isolation, and reconfiguration methods. *IEEE Transactions on Control Systems Technology*, 18(3), 636–653.

Pajic, M., Sundaram, S., Pappas, G.J., and Mangharam, R. (2011). The wireless control network: A new approach for control over networks. *IEEE Transactions on Automatic Control*, 56(10), 2305–2318.

Pasqualetti, F., Dörfler, F., and Bullo, F. (2013). Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11), 2715–2729.

Pasqualetti, F., Dörfler, F., and Bullo, F. (2015). Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine*, 35(1), 110–127.

Šiljak, D.D. (1978). *Large-scale dynamic systems: stability and structure*, volume 1–3. North Holland.

Smith, R.S. (2015). Covert misappropriation of networked control systems: Presenting a feedback structure. *IEEE Control Systems Magazine*, 35(1), 82–92.

Teixeira, A., Shames, I., Sandberg, H., and Johansson, K.H. (2015). A secure control framework for resource-limited adversaries. *Automatica*, 51, 135–148.

Yang, W., Zhang, Y., Chen, G., Yang, C., and Shi, L. (2019). Distributed filtering under false data injection attacks. *Automatica*, 102, 34–44.