2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

A Graph-Theoretic Security Index Based on Undetectability for Cyber-Physical Systems

Lijing Zhai[1], Kyriakos G. Vamvoudakis[1], Jérôme Hugues[2]

***Abstract*— In this paper, we investigate the conditions for the existence of dynamically undetectable attacks and perfectly undetectable attacks. Then we provide a quantitative measure on the security for discrete-time linear time-invariant (LTI) systems under both actuator and sensor attacks based on undetectability. Finally, the computation of proposed security index is reduced to a min-cut problem for the structured systems by graph theory. Numerical examples are provided to illustrate the theoretical results.**

***Index Terms*— CPS security, undetectability, graph theory.**

I. Introduction

Cyber-physical systems (CPS) are complex systems combining physical devices with computational and communication components. Actuators and sensors are vital components for CPS since the locations and numbers of actuators and sensors directly affect the control policies made by the system operators, who need to consider carefully where to put the available actuators and sensors to ensure systems operate in a desired and reliable way. Also, since actuators and sensors can be expensive it is important to figure out how many of them are needed in practice to be cost-efficient. CPS have gradually become large-scale and decentralized in recent years and rely more and more on communication networks. This high-dimensional and decentralized structure increases the exposure to malicious attacks that can cause faults, failures and even significant damage.

Research efforts have been made on the cost-efficient placement or allocation of actuators and sensors. However, most of these developed methods mainly consider controllability or observability properties and do not take into account the security aspect. Motivated by this gap, in this work, we consider the dependence of CPS security on the potentially compromised actuators and sensors, in particular, on deriving a security measure under both actuator and sensor attacks. The topic of CPS security has received increasing attention recently and different security indices are developed. The first kind of security measure is based on reachability analysis, i.e., quantifying the size of reachable sets, which are the sets of all states reachable by dynamical systems with admissible inputs [1], [2]. However, the direction on quantifying reachable sets under malicious attacks and using the developed security metrics to guide actuator and sensor selection from the perspective of security is not fully studied. The second kind of security index is defined as the minimal number of actuators/sensors that attackers need to compromise without being detected [3], [4]. Graph theory can be utilized to study on CPS security [5], [6]. The authors of [6] develop a generic actuator security and propose graph-theoretic conditions for computing the generic actuator security index with the help of maximum linking and the generic normal rank of the corresponding structured transfer function matrix. However, regarding calculating the proposed generic security index, they utilize a brute force search method to iterate through all attack sets. The main difficulties in this direction include how to decouple the computation of security index with the actuator and sensor selection problem. In particular, with the security index at hand, how to guide system operators to select the numbers or locations of actuators and sensors. Moreover, when both actuator and sensor attacks exist at the same time, how to distribute the security index between actuators and sensors.

Contributions: The contribution of this work is twofold. We provide conditions for the existence of dynamical and perfect undetectability. In term of the perfect undetectability, a security index for discrete-time LTI systems under actuator and sensor attacks is proposed. Then, a graph-theoretic approach for structured systems is used to compute the security index by solving a min-cut/max-flow problem.

II. Problem Formulation

Consider the following discrete-time LTI system,

$$x_{k+1} = Ax_k + Bu_k + B_a a_k,$$
$$y_k = Cx_k + D_a a_k,$$

where $k \in \mathbb{N}$ is the discrete time index, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$ and $y_k \in \mathbb{R}^l$ are the state vector, control input and potentially compromised output, respectively, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{l\times n}$ are the state matrix, input matrix and output matrix, respectively. The attack vector $a_k \in \mathbb{R}^{m+l}$ stands for the additive adversaries with the first m entries of a_k corresponding to actuator attacks while the remaining l entries corresponding to sensor attacks. The actuator attacks corrupt the controller command u_k by adding a value $Ba_k(1:m,:)$ that happens during the communication from controllers to actuators, where $a_k(1:m,:)$ stands for the first m entries of a_k. Similarly, the sensor attacks replace the true measurement

[1]L. Zhai and K. G. Vamvoudakis are with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA e-mail: lzhai3@gatech.edu, kyriakos@gatech.edu.

[2]J. Hugues is with the Carnegie Mellon University/Software Engineering Institute, Pittsburgh, PA 15213, USA e-mail: jhugues@andrew.cmu.edu.

This work was supported in part by the Department of Energy under grant No. DE-EE0008453, by ONR Minerva under grant No. N00014 − 18 − 1 − 2160, by NSF under grant Nos. CAREER CPS-1851588 and S&AS 1849198, and by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. DM21-0893.

signals Cx_k with a corrupted value that happens during the communication from sensors to controllers [7]. The matrices $B_a \in \mathbb{R}^{n\times(m+l)}$ and $D_a \in \mathbb{R}^{l\times(m+l)}$ represent attacker's capabilities to corrupt actuators and sensors respectively, given by $B_a = [B \quad \mathbf{0}_{n\times l}]$, $D_a = [\mathbf{0}_{l\times m} \quad \mathbf{I}_l]$, where $\mathbf{I}_l$ denotes an identity matrix with dimension l. The i-th column vector of B corresponds to the i-th actuator while the j-th row vector of C corresponds to the j-th sensor. Let U and S denote the set of actuators and sensors, respectively. Let $U_a \subseteq U$ be the set of attacked actuators with $|U_a| = m'$, and $S_a \subseteq S$ be the set of attacked sensors with $|S_a| = l'$. Assume there are no attack signals added to the safe actuators and sensors. The attacker has full information of the system dynamics, i.e., matrices A, B, and C, while the injected attack signals a_k are unknown to the system operator. Throughout the work, the attack signal a_k is assumed to be nonzero. The attacked actuators and sensors are assumed to be fixed but the values of attack signals may change over time. The matrix B is assumed to have a full column rank. The pairs (A, B) and (A, C) are assumed to be controllable and observable, respectively. Generally since control input u_k is given by the system operator, its contribution to output can be calculated accurately and does not affect the results in this work. Due to superposition properties of LTI systems, and without loss of generality, we neglect the control input term Bu_k throughout this work [8], [9]. Instead, we shall focus on the following system denoted as $\Sigma = (A, B, C, B_a, D_a)$ $\forall k \in \mathbb{N}$,

$$x_{k+1} = Ax_k + B_a a_k, \tag{1}$$

$$y_k = Cx_k + D_a a_k. \tag{2}$$

Assumption 1. The pairs (A, B) and (A, C) are controllable and observable, respectively. □

III. Conditions for Undetectable Attacks

The objective of this work is to investigate security measures of malicious attacks. The security level of CPS can be measured by their ability to detect attacks. So in this section we study undetectable attacks. During the time $0, 1, \ldots, N$ with $N \in \mathbb{Z}_{>0}$, for the system $\Sigma = (A, B, C, B_a, D_a)$, denote the corresponding output trajectory as $Y_N = [y_0^{\mathrm{T}} \quad y_1^{\mathrm{T}} \quad \ldots \quad y_N^{\mathrm{T}}]^{\mathrm{T}}$, and the corresponding unknown attack sequence as $E_N = [a_0^{\mathrm{T}} \quad a_1^{\mathrm{T}} \quad \ldots \quad a_N^{\mathrm{T}}]^{\mathrm{T}}$. The output trajectory Y_N during the time $0, \ldots, N$ is determined by the initial state x_0 and the unknown attack sequence E_N, formulated by $Y_N = \mathcal{O}_N x_0 + V_N E_N + (\mathbf{I}_{N+1} \otimes D_a)E_N$, where $\otimes$ stands for the Kronecker product, $\mathcal{O}_N = \begin{bmatrix} C^{\mathrm{T}} & (CA)^{\mathrm{T}} & (CA^2)^{\mathrm{T}} & \ldots & (CA^N)^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ is the extended observability matrix, and V_N is given as

$$\begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ CB_a & 0 & 0 & \ldots & 0 & 0 \\ CAB_a & CB_a & 0 & \ldots & 0 & 0 \\ CA^2B_a & CAB_a & CB_a & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{N-1}B_a & CA^{N-2}B_a & CA^{N-3}B_a & \ldots & CB_a & 0 \end{bmatrix}.$$

Likewise, the state vector at the time instant N is,

$$x_N = A^N x_0 + \mathcal{C}_{N-1} E_{N-1}, \tag{3}$$

where $\mathcal{C}_{N-1} = \begin{bmatrix} A^{N-1}B_a & A^{N-2}B_a & \ldots & B_a \end{bmatrix}$ is the extended controllability matrix. Now we introduce the following definitions [6], [10].

Definition 1. (*Dynamically Undetectable Attacks*) For the system $\Sigma = (A, B, C, B_a, D_a)$, there exist dynamically undetectable attacks if and only if the nonzero attack sequence E_N satisfies $\mathcal{O}_N x_0 + V_N E_N + (\mathbf{I}_{N+1} \otimes D_a)E_N = \mathcal{O}_N x_0'$, $\forall N \in \mathbb{Z}_{>0}$, with initial state x_0 and $x_0' \in \mathbb{R}^n \backslash \mathbf{0}$. □

Next, we shall provide the conditions for the existence of dynamically undetectable attacks. First, we show that it is sufficient to focus on the time period of $0, 1, \ldots, n-1$ to decide whether there exists a dynamically undetectable attack sequence for the time period of $0, 1, \ldots, N$, $\forall N \in \mathbb{Z}_{>0}$.

Lemma 1. For the system $\Sigma = (A, B, C, B_a, D_a)$, assume Assumption 1 holds and that there exists a dynamically undetectable attack sequence E_{n-1} in the time period of $0, 1, \ldots, n-1$. Then there exists a dynamically undetectable attack sequence in the time period of $0, 1, \ldots, N$, $\forall N \in \mathbb{Z}_{>0}$.

Proof: Given that there exists a dynamically undetectable attack sequence E_{n-1} during the time period of $0, 1, \ldots, n-1$, i.e., $\mathcal{O}_{n-1}x_0 + V_{n-1}E_{n-1} + (\mathbf{I}_n \otimes D_a)E_{n-1} = \mathcal{O}_{n-1}x_0'$. Then left multiply $\mathcal{O}_{n-1}^{\mathrm{T}}$ on both sides and rearrange to get $\mathcal{O}_{n-1}^{\mathrm{T}}[V_{n-1}E_{n-1} + (\mathbf{I}_n \otimes D_a)E_{n-1}] = \mathcal{O}_{n-1}^{\mathrm{T}}\mathcal{O}_{n-1}(x_0' - x_0)$. The observability matrix $\mathcal{O}_{n-1}$ has full column rank. Then it follows that the square matrix $\mathcal{O}_{n-1}^{\mathrm{T}}\mathcal{O}_{n-1}$ has full rank and thus is invertible. Then $\Delta_x = x_0' - x_0$ can be uniquely solved by $\Delta_x = (\mathcal{O}_{n-1}^{\mathrm{T}}\mathcal{O}_{n-1})^{-1}\mathcal{O}_{n-1}^{\mathrm{T}}[V_{n-1}E_{n-1} + (\mathbf{I}_n \otimes D_a)E_{n-1}]$. Now consider the time instant $N = n$. Assume that there exists an attack signal a_n such that $\mathcal{O}_n x_0 + V_n E_n + (\mathbf{I}_{n+1} \otimes D_a)E_n = \mathcal{O}_n x_0'$ holds with $E_n = [E_{n-1}^{\mathrm{T}} \; a_n]^{\mathrm{T}}$. Given now (2) and (3), it follows that $y_n = CA^n x_0 + CA^{n-1}B_a a_0 + CA^{n-2}B_a a_1 + \cdots + CB_a a_{n-1} + D_a a_n = CA^n x_0'$. Rearrange to get,

$$D_a a_n = CA^n \Delta_x - (CA^{n-1}B_a a_0 + \cdots + CB_a a_{n-1}). \tag{4}$$

Since Δ_x is uniquely solved, the right-hand-side (RHS) of (4) is uniquely determined. Considering $D_a = [\mathbf{0}_{l\times m} \; \mathbf{I}_l]$, the first m entries of a_n can be any values while the remaining l entries of a_n are uniquely determined from (4). Thus, the existence of a_n is guaranteed. Similarly, for $N = n+1, n+2, \ldots$, the newly added attack signal a_N always exists. ■

Definition 2. (*Perfectly Undetectable Attacks*) For the system $\Sigma = (A, B, C, B_a, D_a)$, there exist perfectly undetectable attacks if and only if for a nonzero attack sequence E_N, $V_N E_N + (\mathbf{I}_{N+1} \otimes D_a)E_N = \mathbf{0}$ holds, $\forall N \in \mathbb{Z}_{>0}$. □

Perfectly undetectable attacks leave zero trace in the sensory output. Therefore, Definition 2 is a stricter version of Definition 1. Consequently, Lemma 1 also applies to attacks of Definition 2. Now we have the following corollary.

Corollary 1. For the system $\Sigma = (A, B, C, B_a, D_a)$, suppose that there exists a perfectly undetectable attack sequence E_{n-1} during the time period of $0, 1, \ldots, n-1$, then there exists a perfectly undetectable attack sequence during the time period of $0, 1, \ldots, N$, $\forall N \in \mathbb{Z}_{>0}$.

Proof: The proof follows the same logic to that of Lemma 1 with $x_0 = \mathbf{0}$ and $x_0' = \mathbf{0}$. ■

Now we can only consider the time period of $0, 1, \ldots, n-1$. First we recall the following definitions [11], [12].

Definition 3. (*Input Unobservable Subspace*) For the system $\Sigma = (A, B, C, B_a, D_a)$, the input unobservable subspace over k steps is defined as $\mathcal{I}_k = \{x \in \mathbb{R}^n :$ there exists an attack sequence E_{k-1} such that $\mathcal{O}_{k-1}x + V_{k-1}E_{k-1} + (\mathbf{I}_k \otimes D_a)E_{k-1} = \mathbf{0}\}$. □

Definition 4. (*Weakly Unobservable Subspace*) For the system $\Sigma = (A, B, C, B_a, D_a)$, the weakly unobservable subspace, denoted as $\mathcal{W}(\Sigma)$, is defined as its input unobservable subspace over n steps, i.e., $\mathcal{W}(\Sigma) = \mathcal{I}_n$. □

Definition 5. (*Strongly Observable*) The system $\Sigma = (A, B, C, B_a, D_a)$ is strongly observable if and only if its corresponding weakly unobservable subspace is trivial. □

Theorem 1. For the system $\Sigma = (A, B, C, B_a, D_a)$, there exists a perfectly undetectable attack sequence E_{n-1} during the time period of $0, 1, \ldots, n-1$ if and only if the system Σ is strongly observable.

Proof: ($\Leftarrow$ If) Assume system Σ is strongly observable, which implies $V_{n-1}E_{n-1} + (\mathbf{I}_n \otimes D_a)E_{n-1} = \mathbf{0}$. By Definition 2, this shows that there exists a perfectly undetectable attack sequence E_{n-1} for the system $\Sigma = (A, B, C, B_a, D_a)$ with initial condition $x_0 = \mathbf{0}$. ($\Rightarrow$ Only if) Assume there exists a perfectly undetectable attack sequence E_{n-1} for the system Σ. By Definition 2, $V_{n-1}E_{n-1}+(\mathbf{I}_n \otimes D_a)E_{n-1} = \mathbf{0}$. If there exists $\delta \neq \mathbf{0}$ such that $\mathcal{O}_{n-1}\delta + V_{n-1}E_{n-1} + (\mathbf{I}_n \otimes D_a)E_{n-1} = \mathbf{0}$, $\mathcal{O}_{n-1}\delta = \mathbf{0}$ implies $\mathcal{O}_{n-1}$ does not have full column rank (contradicts Assumption 1). So, the system Σ is strongly observable. ■

Remark 1. The assumption that the system is strongly observable is a sufficient and necessary condition for the existence of perfectly undetectable attacks. □

Theorem 2. If the system $\Sigma = (A, B, C, B_a, D_a)$ is not strongly observable, there exists a dynamically undetectable attack sequence E_{n-1} in the time period of $0, 1, \ldots, n-1$.

Proof: Assume the system Σ is not strongly observable. By Definition 5, there exist a nonzero $\delta \in \mathcal{W}(\Sigma)$ and an attack sequence E_{n-1} such that $\mathcal{O}_{n-1}\delta + V_{n-1}E_{n-1} + (\mathbf{I}_n \otimes D_a)E_{n-1} = \mathbf{0}$. Define $x_0' = x_0 - \delta$ and substitute it into the above equation to get $\mathcal{O}_{n-1}x_0 + V_{n-1}E_{n-1} + (\mathbf{I}_n \otimes D_a)E_{n-1} = \mathcal{O}_{n-1}x_0'$. Therefore, according to Definition 1, there exists a dynamically undetectable attack sequence E_{n-1} during the time period of $0, 1, \ldots, n-1$. ■

Corollary 2. For the system $\Sigma = (A, B, C, B_a, D_a)$, non-existence of dynamically undetectable attacks implies the existence of perfectly undetectable attacks.

Proof: By Theorem 2, non existence of dynamically undetectable attacks implies the system is strongly observable. By Theorem 1, perfectly undetectable attacks exist. ■

Note that the non-existence of dynamically undetectable attacks implies the existence of perfectly undetectable attacks. However, the existence of dynamically undetectable attacks rules out the existence of perfectly undetectable attacks. Therefore, it is safe to consider perfectly undetectable attacks for the sake of system security.

IV. Security Index

Based on the previous discussions on undetectable attacks, we define security index in terms of perfect undetectability.

Definition 6. (*Security Index*) For the system $\Sigma = (A, B, C, B_a, D_a)$, the security index is defined as the minimal number of attacked sensors and actuators to conduct perfectly undetectable attacks, denoted as s_0 and given by,

$$s_0 = \min_{a_k} ||a_k||_0 \tag{5}$$

$$\text{s.t.} \quad x_{k+1} = Ax_k + B_a a_k, \tag{6}$$

$$\mathbf{0} = Cx_k + D_a a_k, \tag{7}$$

$$x_0 = \mathbf{0}, \tag{8}$$

where $||a_k||_0 = |\text{supp}(a_k)|$ and $\text{supp}(a_k) = \{i \in \mathcal{I} : a_k^{(i)} \neq 0\}$, with nonzero a_k, $a_k^{(i)}$ being the i-th element of a_k and $\mathcal{I}$ being a set of indices of elements of a_k. □

The constraints (6) and (7) make sure the system dynamics are obeyed. The constraints (7) and (8) imply that perfectly undetectable attacks are considered. Note that if (5) has no solution, which implies that the system Σ is not strongly observable based on Theorem 1, then the security index is denoted as $s_0 = \infty$. If $s_0 = m + l$, the system Σ is maximally secure, which implies that adversaries have to attack all the available actuators and sensors to remain perfect undetectable. The computation of security index is generally NP-hard due to l_0 norm in the objective function [13]. As a result, its computation is not efficient for high dimensional systems. Next we shall rely on structured model of systems to compute the security index.

A. Structured Model and Graph Representation

The structured matrices $[A], [C], [B_a], [D_a]$ have binary elements. The (i, j) entry of matrix $[A]$ equal to 0 means $A_{ij} = 0$ for every realization of matrix A while $[A]_{ij} = 1$ means A_{ij} is a free parameter and can be any value from $\mathbb{R}$ except 0. The same holds for matrices $[C]$, $[B_a]$ and $[D_a]$. Denote the set of all system realizations from structured matrices $[A], [C], [B_a], [D_a]$ as $\mathcal{R}$. Structured systems provide less knowledge of system dynamics, but the analysis based on structured models is robust to system dynamics variations and applied to any realizations. To simplify the analysis, we make the following assumption.

Assumption 2. Each state is influenced directly by only one actuator and measured directly by only one sensor. □

Assumption 3. It is assumed that each attack signal only corrupts one actuator. □

Remark 2. The formulation of the attack matrix D_a implies each attack signal only corrupts one sensor. □

We now associate a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the structured model $[A], [C], [B_a], [D_a]$ of the system (1)-(2). Denote $\mathcal{X} = \{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\}$ as the set of state vertices, $\mathcal{Y} = \{y^{(1)}, y^{(2)}, \ldots, y^{(l)}\}$ as the set of output vertices, $\mathcal{A} = \{a^{(1)}, a^{(2)}, \ldots, a^{(m+l)}\}$ as the set of attack vertices. The vertex set of $\mathcal{G}$ is formed by $\mathcal{V} = \mathcal{X} \bigcup \mathcal{Y} \bigcup \mathcal{A}$. The edge set is formed by $\mathcal{E} = \mathcal{E}_A \bigcup \mathcal{E}_C \bigcup \mathcal{E}_{B_a} \bigcup \mathcal{E}_{D_a}$, where $\mathcal{E}_A = \{(x^{(j)}, x^{(i)}) : [A]_{ij} = 1\}$ is the set of edges from vertex $x^{(j)}$ to vertex $x^{(i)}$, $\mathcal{E}_C = \{(x^{(j)}, y^{(i)}) : [C]_{ij} = 1\}$ is the set of edges from vertex $x^{(j)}$ to vertex $y^{(i)}$, $\mathcal{E}_{B_a} = \{(a^{(j)}, x^{(i)}) : [B_a]_{ij} = 1\}$ is the set of edges from vertex $a^{(j)}$ to vertex $x^{(i)}$, $\mathcal{E}_{D_a} = \{(a^{(j)}, y^{(i)}) : [D_a]_{ij} = 1\}$ is the set of edges from vertex $a^{(j)}$ to vertex $y^{(i)}$.

Malicious attacks against the system Σ can be considered as the attacker injecting signals into the system through attack vertices in $\mathcal{A}$. Perfect undetectability means that the injected signals do not flow to output vertices in $\mathcal{Y}$. Thus, we consider this problem as a flow network problem with source vertices in $\mathcal{A}$ and sink vertices in $\mathcal{Y}$. First, we add a dummy vertex t and add edges from all sink/output vertices in $\mathcal{Y}$ to t. Let $\mathcal{E}_{yt} = \{(y^{(i)}, t) : \forall y^{(i)} \in \mathcal{Y},\ i = \{1, 2, \ldots, l\}\}$ denote the set of edges from vertex $y^{(i)}$ to vertex t. The vertex t is considered as an operator who receives all the measurements in the process. For flow networks, in order to stay perfectly undetectable, the attacker needs to prevent the flow from reaching the operator vertex t, i.e., sensory output always equals to zero. If the maximum flow from attack vertices in $\mathcal{A}$ to t is zero, then attacks remain undetectable. Let $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$ be the extended graph of the system with the above modifications, i.e., $\mathcal{V}_t = \mathcal{V} \bigcup t$ and $\mathcal{E}_t = \mathcal{E} \bigcup \mathcal{E}_{yt}$.

B. Characterization of Security Index

Due to Assumption 2 and Remark 2), we denote the set of attack vertices corresponding to the set of attacked actuators, i.e., U_a, as $\mathcal{A}_a$, and the set of attack vertices corresponding to the set of attacked sensors, i.e., S_a, as $\mathcal{A}_s$. It follows that $\mathcal{A}_a \bigcup \mathcal{A}_s \subseteq \mathcal{A}$, $|\mathcal{A}_a| = m'$ and $|\mathcal{A}_s| = l'$.

Theorem 3. Consider the extended graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$. For each vertex $a_k^{(i)} \in \mathcal{A}_a$, $i \in \{1, 2, \ldots, |\mathcal{A}_a|\}$, define $\mathcal{X}_a^i = \{x_k^{(j)} \in \mathcal{X} : (a_k^{(q)}, x_k^{(j)}) \in \mathcal{E}_{B_a}, a_k^{(q)} \in \mathcal{A}_a \backslash a_k^{(i)}\}$. There exist perfectly undetectable attacks for the system Σ of any realization from $\mathcal{R}$ with the set of attacked actuators U_a and the set of attacked sensor S_a if and only if $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ is a vertex separator of $a_k^{(i)}$ and t in $\mathcal{G}_t$.

Proof: ($\Leftarrow$ If) Given $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ is a vertex separator of $a_k^{(i)}$ and t in $\mathcal{G}_t$, by Assumption 2, let the q-th actuator attack signal $a_k^{(q)}$ corrupting the j-th state $x_k^{(j)}$ be,

$$a_k^{(q)} = -A(j,:)x_k / B_a(j,q),\ k \in \mathbb{N} \tag{9}$$

where $A(j,:)$ denotes the j-th row of A, $B_a(j,q)$ is the (j,q)-th element of B_a. Note that $B_a(j,q) \neq 0$ due to Assumptions 2 and 3. For the p-th attacked sensor, let the corresponding sensor attack signal be,

$$a_k^{(m+p)} = -C(p,:)x_k,\ k \in \mathbb{N} \tag{10}$$

where $C(p,:)$ denotes the p-th row of C. Next we prove that attacks defined by (9) and (10) are perfectly undetectable, i.e., $y \equiv \mathbf{0}$ with $x_0 = \mathbf{0}$. For $x_k^{(j)} \in \mathcal{X}_a^i$ with $a_k^{(q)}$ influencing $x_k^{(j)}$, applying equation (9) we have $x_{k+1}^{(j)} = A(j,:)x_k + B_a(j,q)a_k^{(q)} = 0$, which implies that all states in $\mathcal{X}_a^i$ equal to 0. For the p-th attacked sensor, due to equation (10), we have $y_k^{(p)} = C(p,:)x_k + a_k^{(m+p)} = 0$. Next we define $\mathcal{X}_b^i = \{x_k^{(j)} \in \mathcal{X}$: there exists a directed path from $a_k^{(i)}$ to $x_k^{(j)}$ which does not include states in $\mathcal{X}_a^i\}$. We claim that states in $\mathcal{X}_b^i$ cannot be measured by attack-free sensors. If so, then there exists a directed path from $a_k^{(i)}$ to t, which contradicts that $\mathcal{X}_a^i \bigcup \mathcal{A}_s$ is a vertex separator of $a_k^{(i)}$ and t in $\mathcal{G}_t$. For the remaining states $\mathcal{X}_c^i = \mathcal{X} \backslash (\mathcal{X}_a^i \bigcup \mathcal{X}_b^i)$, we claim that the edge $(x_k^{(b)}, x_k^{(c)})$ with $x_k^{(b)} \in \mathcal{X}_b^i$ and $x_k^{(c)} \in \mathcal{X}_c^i$ does not exist. If so this would imply there exists a directed path from $a_k^{(i)}$ to $x_k^{(c)}$ which does not include states in $\mathcal{X}_a^i$. Then by the definition of $\mathcal{X}_b^i$, we get $x_k^{(c)} \in \mathcal{X}_b^i$, which is a contradiction since $x_k^{(c)} \in \mathcal{X}_c^i$. Thus, we conclude that states in $\mathcal{X}_c^i$ are not affected by states in $\mathcal{X}_b^i$. Since $x_0 = \mathbf{0}$ and we have proved that states in $\mathcal{X}_a^i$ equal to 0, then states in $\mathcal{X}_c^i$ always remain 0. We have showed that states in $\mathcal{X}_b^i$ cannot be measured by attack-free sensors. Thus, the attack-free sensor measurement equals to 0. We have proved that the attacked sensor measurement remains 0. Therefore, attacks with strategies of (9) and (10) are perfectly undetectable. ($\Rightarrow$ Only if) Suppose that $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ is not a vertex separator of $a_k^{(i)}$ and t. Then it follows that there exists a directed path from $a_k^{(i)}$ to t which does not include states in $\mathcal{X}_a^i \bigcup \mathcal{A}_s$. We denote this path as $p_i = \{a_k^{(i)} \rightarrow x_k^{(i_1)} \rightarrow x_k^{(i_2)} \rightarrow \cdots \rightarrow x_k^{(i_n)} \rightarrow y_k^{(q)} \rightarrow t\}$. Now we need to show that no perfectly undetectable attacks for at least one realization from $\mathcal{R}$ exist. For $x_k^{(i_1)}$ from path p_i, let $A(i_1,:) = 0$ so that other states cannot affect $x_k^{(i_1)}$. For other states $x_k^{(i_j)}$ from path p_i, where $2 \leqslant j \leqslant n$, set $A(i_j, h) \neq 0$ for $h = i_{j-1}$ and $A(i_j, h) = 0$ for $h \neq i_{j-1}$ so that only $x_k^{(i_{j-1})}$ can affect $x_k^{(i_j)}$. Let $C(q, i_n) \neq 0$ so that $y_k^{(q)} \neq 0$ as long as $x_k^{(i_n)} \neq 0$ due to Assumption 2. For $a_k^{(i)} \neq 0$, given $A(i_1,:) = 0$, we have $x_{k+1}^{(i_1)} = B(i_1, i)a_k^{(i)} \neq 0$. Now for other states from path p_i, we have $x_{k+2}^{(i_2)} = A(i_2, i_1)x_{k+1}^{(i_1)} \neq 0 \Rightarrow x_{k+3}^{(i_3)} = A(i_3, i_2)x_{k+2}^{(i_2)} \neq 0 \Rightarrow \ldots \Rightarrow x_{k+n}^{(i_n)} = A(i_n, i_{n-1})x_{k+n-1}^{(i_{n-1})} \neq 0$. Since $C(q, i_n) \neq 0$, then we have $y_{k+n+1}^{(q)} = C(q, i_n)x_{k+n}^{(i_n)} \neq 0$. Therefore, there does not exist perfectly undetectable attacks for this realization. ■

Note that in order for attacks to remain perfectly undetectable, at least one actuator needs to be attacked to hide sensor attack signals. Thus, we have the following corollary to formulate the condition for the existence of perfectly undetectable attacks from the view of sensor attack signals.

Corollary 3. Consider the extended graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$. For each vertex $a_k^{(j)} \in \mathcal{A}_s$ with $j \in \{1, 2, \ldots, |\mathcal{A}_s|\}$, let the corresponding j-th attacked sensor measures the p-th state $x_k^{(p)}$. Let $a_k^{(i)}$ denote the actuator attack signal which attacks the p-th state directly (i.e., there exists an edge from $a_k^{(i)}$ to

$x_k^{(p)}$) or indirectly (i.e, there exists a directed path from $a_k^{(i)}$ to $x_k^{(p)}$). Define $\mathcal{X}_a^i = \{x_k^{(j)} \in \mathcal{X} : (a_k^{(q)}, x_k^{(j)}) \in \mathcal{E}_{B_a}, a_k^{(q)} \in \mathcal{A}_a \backslash a_k^{(i)}\}$. Then there exist perfectly undetectable attacks for the system Σ of any realization from $\mathcal{R}$ with the set of attacked actuators U_a and the set of attacked sensor S_a if and only if $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ is a vertex separator of $a_k^{(i)}$ and t in $\mathcal{G}_t$.

Proof: Due to Assumption 4, Assumption 5 and Remark 7, for the j-th attacked sensor measuring the p-th state with attack signal $a_k^{(j)} \in \mathcal{A}_s$, for the existence of perfectly undetectable attacks, i.e., $y \equiv \mathbf{0}$ with $x_0 = \mathbf{0}$, there must exist an actuator attack signal denoted as $a_k^{(i)} \in \mathcal{A}_a$ directly or indirectly corrupting the p-th state. Thus, following the same logic to the proof of Theorem 3, $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ being a vertex separator of $a_k^{(i)}$ and t with $y_k^{(j)} \in \mathcal{S}_a$ is sufficient and necessary conditions for the existence of perfectly undetectable attacks. ■

Definition of security index (5)-(8) aims to find the minimum number of attacked actuators and sensors for adversaries to remain perfectly undetectable. Theorem 3 and Corollary 3 characterize the conditions for the existence of perfectly undetectable attacks. Therefore, we can compute the security index by solving a problem of finding the minimum size of $\mathcal{A}_a \bigcup \mathcal{A}_s$ such that $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ is a vertex separator of $a_k^{(i)}$ and t in $\mathcal{G}_t$, with $i \in \{1, 2, \ldots, |\mathcal{A}_a|\}$.

Problem 1. For the system $\Sigma = (A, B, C, B_a, D_a)$, the security index s is computed as $\forall i \in \{1, 2, \ldots, |\mathcal{A}_a|\}$,

$$s = \min_{\mathcal{A}_a, \mathcal{A}_s} (|\mathcal{A}_a \bigcup \mathcal{A}_s|)$$

s.t. $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ is a vertex separator of $a_k^{(i)}$ and t in $\mathcal{G}_t$ with $\mathcal{X}_a^i = \{x_k^{(j)} \in \mathcal{X} : (a_k^{(q)}, x_k^{(j)}) \in \mathcal{E}_{B_a}, a_k^{(q)} \in \mathcal{A}_a \backslash a_k^{(i)}\}$. □

C. Computation of Security Index

As we discussed above, the extended graph $\mathcal{G}_t$ can be considered as flows represented by attack signals from source vertices in $\mathcal{A}_a \bigcup \mathcal{A}_s$ to the system operator t. For attacks remaining perfectly undetectable, they need to prevent the flow from reaching t. Inspired by [14], now we convert the extended graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$ to a flow network $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ by adding a flow capacity for each edge. For each $i \in \{1, 2, \ldots, |\mathcal{A}_a|\}$, create $\mathcal{V}'$ and $\mathcal{E}'$ as follows:

Rule 1. For each vertex $x_k^{(j)} \in \mathcal{X}_a^i$, split $x_k^{(j)}$ into two vertices $x_{k,1}^{(j)}$ and $x_{k,2}^{(j)}$ with an edge from $x_{k,1}^{(j)}$ to $x_{k,2}^{(j)}$ and a flow capacity the same to the incoming flow to $x_{k,1}^{(j)}$.

Rule 2. For each vertex $x_k^{(q)} \in \mathcal{X} \backslash \mathcal{X}_a^i$, keep $x_k^{(q)}$ in $\mathcal{G}'$.

Rule 3. Keep $(a_k^{(j)}, x_k^{(q)}) \in \mathcal{E}_{B_a}$ in $\mathcal{G}'$ with ∞ flow capacity.

Rule 4. Consider $(x_k^{(j)}, x_k^{(q)}) \in \mathcal{E}_A$.

- For $x_k^{(j)} \in \mathcal{X}_a$ and $x_k^{(q)} \in \mathcal{X}_a$, include $(x_{k,2}^{(j)}, x_{k,1}^{(q)})$ in $\mathcal{E}'$ with a flow capacity ∞.
- For $x_k^{(j)} \in \mathcal{X}_a^i$ and $x_k^{(q)} \in \mathcal{X} \backslash \mathcal{X}_a^i$, include $(x_{k,2}^{(j)}, x_k^{(q)})$ in $\mathcal{E}'$ with a flow capacity ∞.
- For $x_k^{(j)} \in \mathcal{X} \backslash \mathcal{X}_a^i$ and $x_k^{(q)} \in \mathcal{X}_a^i$, include $(x_k^{(j)}, x_{k,1}^{(q)})$ in $\mathcal{E}'$ with a flow capacity ∞.
- For $x_k^{(j)} \in \mathcal{X} \backslash \mathcal{X}_a^i$ and $x_k^{(q)} \in \mathcal{X} \backslash \mathcal{X}_a^i$, include $(x_k^{(j)}, x_k^{(q)})$ in $\mathcal{E}'$ with a flow capacity ∞.

Rule 5. For vertex $x_k^{(j)} \in \mathcal{X}_a^i$, include $(x_{k,1}^{(j)}, x_{k,2}^{(j)})$ in $\mathcal{E}'$ with a flow capacity 1.

Rule 6. Consider $(x_k^{(j)}, y_k^{(q)}) \in \mathcal{E}_C$. For $x_k^{(j)} \in \mathcal{X}_a^i$, include $(x_{k,2}^{(j)}, y_k^{(q)})$ in $\mathcal{E}'$ with a flow capacity 1. For $x_k^{(j)} \in \mathcal{X} \backslash \mathcal{X}_a^i$, include $(x_k^{(j)}, y_k^{(q)})$ in $\mathcal{E}'$ with a flow capacity 1.

Rule 7. Consider $(y_k^{(j)}, t)$. If $y_k^{(j)}$ is not attacked, include $(y_k^{(j)}, t)$ in $\mathcal{E}'$ with a flow capacity ∞. If $y_k^{(j)}$ is attacked, include $(y_k^{(j)}, t)$ in $\mathcal{E}'$ with a flow capacity 1.

Given Assumptions 2 and 3, Remark 2, when assigning edge flow capacities, Rule 5 guarantees each attack signal only corrupts one actuator and each actuator directly influences on state. Rule 6 guarantees each sensor directly measures one state. Rule 7 guarantees each attack signal only corrupts one sensor. Next we derive a relationship between the minimum size of vertex separator and minimum cut.

Theorem 4. For the flow network $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, the minimum size of vertex separator $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ from $a_k^{(i)}$ to t is equivalent to the minimum $a_k^{(i)} - t$ cut.

Proof: For $x_k^{(j)} \in \mathcal{X}_a^i$, $x_k^{(j)}$ being a vertex separator implies that there is a cut of the edge from $x_{k,1}^{(j)}$ to $x_{k,2}^{(j)}$ in $\mathcal{G}'$ with a flow capacity 1. For $y_k^{(j)} \in \mathcal{S}_a$, $y_k^{(j)}$ being a vertex separator implies that there is a cut of the edge from $y_k^{(j)}$ to t with a flow capacity 1. Note that $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ being a vertex separator from $a_k^{(i)}$ to t means that the removal of all vertices in $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ eliminates all the directed paths from $a_k^{(i)}$ to t, and one vertex in $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ corresponds to a cut of edge with a flow capacity 1. Therefore, the minimum size of vertex separator $\mathcal{X}_a^i \bigcup \mathcal{S}_a$ from $a_k^{(i)}$ to t is equivalent to the minimum $a_k^{(i)} - t$ cut in $\mathcal{G}'$. ■

For the flow network $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$, let δ be the minimum capacity of $a_k^{(i)} - t$ on $\mathcal{G}'$, $\forall i \in \{1, 2, \ldots, |\mathcal{A}_a|\}$. Then by Theorem 4, the solution to Problem 1 is $s = \delta + 1$, with 1 added due to the consideration of the attacked actuator associated with the attack signal $a_k^{(i)}$. Based on the Max-flow Min-cut theorem, the minimum capacity of an $s-t$ cut equals to the size of maximum flow from s to t. Finding the maximum flow on a directed graph is a standard max-flow problem, which can be solved by Ford-Fulkerson algorithm or Edmonds-Karp algorithm in polynomial time.

V. Numerical Examples

In this section, we provide a numerical example to illustrate the computation of the proposed security index. Consider a structured system model with 5 states as follows,

$$[A] = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, [B] = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix},$$
$$[C] = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (11)$$

There are 2 actuators ($u_k^{(1)}$, $u_k^{(2)}$), and 3 sensors ($y_k^{(1)}$, $y_k^{(2)}$, $y_k^{(3)}$). Actuators $u_k^{(1)}$ and $u_k^{(2)}$ directly affect

states $x_k^{(2)}$ and $x_k^{(5)}$, respectively. Sensors $y_k^{(1)}$, $y_k^{(2)}$ and $y_k^{(3)}$ directly measure states $x_k^{(2)}$, $x_k^{(3)}$ and $x_k^{(5)}$, respectively. Assume that both actuators and only the second sensor are corrupted by malicious attacks, i.e., $a_k^{(1)} \neq 0$, $a_k^{(2)} \neq 0$, $a_k^{(4)} \neq 0$. The extended graph representation $\mathcal{G}_t$ for the structured model (11) is shown as Figure 1. For $i = 1$, $\mathcal{X}_a^1 =$

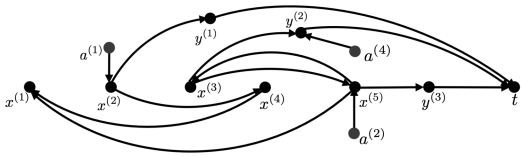

Fig. 1. Extended graph for the structured system model (11).

$\{x_k^{(5)}\}$ while for $i = 2$, $\mathcal{X}_a^2 = \{x_k^{(2)}\}$. The corresponding flow networks $\mathcal{G}'$ from the perspective of $a_k^{(1)}$ and $a_k^{(2)}$ are shown as Figure 2 and Figure 3, respectively. The security

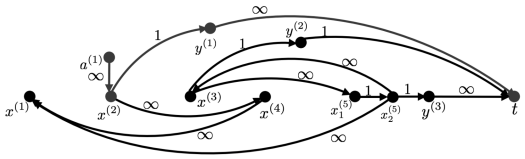

Fig. 2. Flow network from $a^{(1)}$ to t with highlighted max-flow path.

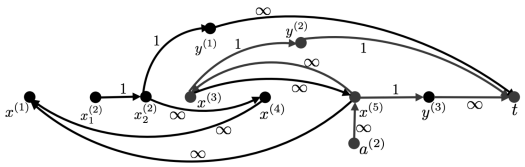

Fig. 3. Flow network from $a^{(2)}$ to t with highlighted max-flow path.

index is solved as $s = \min\{1, 2\} + 1 = 2$ with max-flow paths from a to t highlighted in red. Security index $s = 2$ implies that adversaries can only attack the first actuator $u_k^{(1)}$ and another one actuator or sensor to remain undetectable. Next, we aim to increase the security index by placing more actuators or sensors. Specifically, a secure sensor $y_k^{(4)}$ is added to directly measure $x_k^{(1)}$. Now the corresponding flow networks are shown as Figure 4 and Figure 5 with max-flow paths from a to t highlighted in red. The security index is

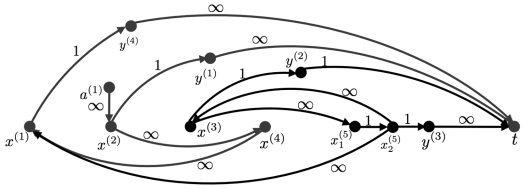

Fig. 4. Flow network from $a^{(1)}$ to t after adding $y^{(4)}$ to measure $x^{(1)}$.

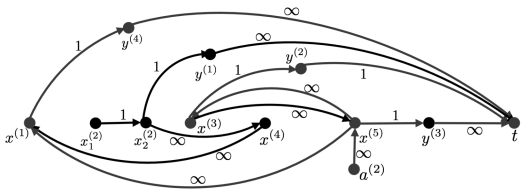

Fig. 5. Flow network from $a^{(2)}$ to t after adding $y^{(4)}$ to measure $x^{(1)}$.

now $s' = \min\{2, 3\} + 1 = 3$, which shows the placement of one more secure sensor makes the system less vulnerable.

VI. Conclusion and Future Work

In this work, we investigate the conditions for the existence of undetectable attacks and propose a security index for discrete-time LTI systems under both actuator and sensor attacks. Through structured models, the computation of the security index is considered as a min-cut/max-flow problem. Future work will focus on the actuator and sensor selection or placement strategies based on the proposed security index.

References

[1] C. Murguia, N. van de Wouw, and J. Ruths, "Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 2088–2094, 2017.

[2] C. Murguia, I. Shames, J. Ruths, and D. Nesic, "Security metrics of networked control systems under sensor attacks (extended preprint)," *arXiv preprint arXiv:1809.01808*, 2018.

[3] H. Sandberg and A. M. Teixeira, "From control system security indices to attack identifiability," in *2016 Science of Security for Cyber-Physical Systems Workshop (SOSCYPS)*. IEEE, 2016, pp. 1–6.

[4] J. Milošević, "Security metrics and allocation of security resources for control systems," Ph.D. dissertation, KTH Royal Institute of Technology, 2020.

[5] S. Weerakkody, X. Liu, S. H. Son, and B. Sinopoli, "A graph-theoretic characterization of perfect attackability for secure design of distributed control systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 60–70, 2016.

[6] S. Gracy, J. Milosevic, and H. Sandberg, "Actuator security index for structured systems," *arXiv preprint arXiv:2003.05752*, 2020.

[7] S. H. Kafash, J. Giraldo, C. Murguia, A. A. Cardenas, and J. Ruths, "Constraining attacker capabilities through actuator saturation," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 986–991.

[8] Y. Chen, S. Kar, and J. M. Moura, "Cyber-physical systems: Dynamic sensor attacks and strong observability," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1752–1756.

[9] ——, "Dynamic attack detection in cyber-physical systems with side initial state information," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4618–4624, 2016.

[10] J. Milošević, A. Teixeira, K. H. Johansson, and H. Sandberg, "Actuator security indices based on perfect undetectability: Computation, robustness, and sensor placement," *IEEE Transactions on Automatic Control*, 2020.

[11] B. Molinari, "Extended controllability and observability for linear systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 136–137, 1976.

[12] H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control theory for linear systems*. Springer Science & Business Media, 2012.

[13] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[14] J. Milošević, H. Sandberg, and K. H. Johansson, "A security index for actuators based on perfect undetectability: Properties and approximation," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2018, pp. 235–241.