Metrics to enhance the resilience of cyber-physical systems

Michel Barbeau, Frederic Cuppens, Nora Cuppens, Dagnas, Romain, Joaquin Garcia-alfaro

2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) |978-1-6654-0392-4/20/$31.00 ©2020 IEEE | DOI: 10.1109/TRUSTCOM50675.2020.00156

Metrics to Enhance the Resilience of Cyber-Physical Systems

Michel Barbeau[1], Frédéric Cuppens[2], Nora Cuppens[2], Romain Dagnas[3], and Joaquin Garcia-Alfaro[4]

[1]School of Computer Science, Carleton University, Ottawa, ON K1S 5B6, Canada

[2]École Polytechnique de Montréal, Montréal, QC H3T 1J4, Canada

[3]Institut de Recherche Technologique SystemX, Palaiseau, France

[4]Institut Polytechnique de Paris, SAMOVAR, Telecom SudParis, 91120, Palaiseau, France

***Abstract*—We focus on resilience towards covert attacks on Cyber-Physical Systems (CPS). We define the new k-steerability and ℓ-monitorability control-theoretic concepts. k-steerability reflects the ability to act on every individual plant state variable with at least k different groups of functionally diverse input signals. ℓ-monitorability indicates the ability to monitor every individual plant state variable with ℓ different groups of functionally diverse output signals. A CPS with k-steerability and ℓ-monitorability is said to be (k, ℓ)-resilient. k and ℓ, when both greater than one, provide the capability to mitigate the impact of covert attacks when some signals, but not all, are compromised. We analyze the influence of k and ℓ on the resilience of a system and the ability to recover its state when attacks are perpetrated. We argue that the values of k and ℓ can be augmented by combining redundancy and diversity in hardware and software techniques that apply the moving target paradigm.**

***Index Terms*—Cyber-Physical Systems, Control Theory, Cyber-Resilience, Covert Attacks, Security Metrics, Attack Remediation, Recoverability.**

I. Introduction

A Cyber-Physical System (CPS) is a plant controlled and monitored by embedded and network computers [8]. For the purposes of protection, CPS design aims at steerability and monitorability. Sending input signals to actuators, steerability refers to the ability to drive and maintain a CPS in a desired operating point. Interpreting output signals produced by sensors, monitorability indicates the capability to accurately deduce the internal state of a CPS. The control signals are chosen such that a CPS is asymptotically stable. This means that small variations in the input signals generate small variations in the output signals. The output stays bounded for any bounded input. There are no oscillations. Given the knowledge about the behavior of a CPS, i.e., steerability and monitorability, the goal is to increase the system recoverability (i.e., to adapt and bounce back from stability disruptions, as quick as possible). It has been acknowledged that CPS are vulnerable to integrity and availability attacks [13], [3], [1]. In this paper, we focus on covert attacks [18], [19], i.e., a family of cyber-physical attacks pointing out to physical aggression against the operation of a CPS, by manipulating output signals from sensors, and input signals to actuators.

We introduce the notion of k-steerability. The parameter k corresponds to the minimum number of input signals available to act on each individual plant state variable. We also define the concept of ℓ-monitorability. The parameter ℓ reflects the minimum number of output signals that can be used to monitor each individual plant state variable. We study values of k and ℓ with respect to system resilience and the ability to recover a plant state. If due to covert attacks, h input signals are compromised, steerability of each individual plant state is not entirely lost as long as h is lower than k. This partial steerability can be leveraged to run a covert attack mitigation plan. If due to covert attacks, g output signals are hacked, the ability to detect the condition is not entirely lost as long as g is lower than ℓ. We discuss how k and ℓ can be determined and augmented by adding redundant and diverse hardware. k-steerability and ℓ-monitorability combine together into the (k, ℓ)-resilient property. Building upon redundancy and diversity in actuators and sensors, our approach increases the level of difficulty for adversaries.

Our work assumes that both input and output signals can be correlated into functionally diverse groups, as a complement to the traditional use of redundancy in critical systems. The use of redundancy assume the inclusion of alternative copies of, e.g., sensors, actuators and controllers, in order to guarantee system availability. If the system finds itself under a situation of attack, and the values of a group of components are not behaving as expected, then the validation of such values can be contrasted with the values of redundant replicas, assuming that there was an attack affecting the system. This technique is complementary to fault tolerance techniques, also used to address situations in which some system components are victims of failures or faults. However, the use of redundancy for security purposes may have some drawbacks. Since the

978-1-6654-0392-4/20/$31.00 ©2020 IEEE
DOI 10.1109/TrustCom50675.2020.00156

replicas may be seen as identical, once an attacker has managed to compromise one of them, then the rest of the replicas can also be compromised very easily. Hence, we need to impose the use of diversity. For instance, if the replicas are geographically distributed, or the replicas compute their values using different physical phenomena, the approach can improve the way to handle those attacks exploiting the physical nature of vulnerable components. Hence, our approach assumes the existence of different replicas behaving in an independent manner and with non-overlapping patterns (e.g., physical patterns) to handle the attacks.

Paper Organization — Section II introduces CPS modeling, covert attacks and related work. The k-steerability, ℓ-monitorability and (k,ℓ)-resilient concepts are developed in Section III. Section IV concludes the paper.

II. Background and Related Work

A CPS consists of a plant and a controller. They are distributed and communicate through a network. Several mathematical models exist for representing them [5], [12]. In the sequel, we introduce the necessary modeling background, using a CPS with fluid dynamics as an example.

A. Differential Equation Representation

Let us consider as a plant an individual cylindrical tank, with a single inflow and a single outflow of liquid. The tank liquid level can be modeled by the following differential equation [20]:

$$\alpha \frac{dh\left(t\right)}{dt} = F(t) - a\sqrt{h\left(t\right)} \tag{1}$$

Eq. (1) models the relationship between instantaneous changes of liquid level and difference between the inflow rate and outflow rate. As a function of time t (second), the level of the liquid in the tank is $h(t)$ (cm). Variable α represents a cross-sectional area of the tank (cm^2). The term $F(t)$ represents the inflow rate (cm^3/second). The parameter a denotes the outlet valve coefficient. The outflow rate (cm^3/second) is proportional to the product a times square root of the pressure represented by the term $h\left(t\right)$. Note that because of the square root term, the system in nonlinear.

The model represented by Eq. (1) is linearized assuming a linear inflow rate and operation around a liquid level h_0, termed the *operating point*. It is assumed the level is maintained at point $h_0+\Delta$, with $|\Delta|$ small. Linearization is based on the observation that the expression $(1+\epsilon)^\beta$ is approximately equal to the expression $1+\beta\epsilon$, when $\epsilon \ll 1$. In the expression $a\sqrt{h\left(t\right)}$, substituting $h(t)$ by the sum $h_0+\Delta$, we get a linear model for the outflow rate:

$$a\sqrt{h\left(t\right)} = a\sqrt{h_0+\Delta} = a\sqrt{h_0}\sqrt{1+\Delta/h_0} \approx a\sqrt{h_0}\left(1+\frac{\Delta}{2h_0}\right)$$

Inflow rate $F(t)$ is modeled by the product $\gamma\kappa v$. The parameters γ, κ and v respectively denote the valve coefficient, pump coefficient (cm^3/V second) and voltage applied to the pump (V). The voltage v is the variable governed by the controller. The resulting linear differential equation modeling the liquid level is:

$$\alpha \frac{dh\left(t\right)}{dt} = \gamma\kappa v - a\sqrt{h_0}\left(1+\frac{\Delta}{2h_0}\right) \tag{2}$$

Eq. (2) is an approximation that remains valid as long as Δ is relatively small, that is, at the chosen operating point h_0 there are only small level fluctuations. To maintain that condition, the inflow rate $\gamma\kappa v$ must be equal to the outflow rate $a\sqrt{h_0}$, with small fluctuations $-a\sqrt{h_0}\left(\frac{\Delta}{2h_0}\right) = -\frac{a\Delta}{2\sqrt{h_0}}$ of inflow rate that translate to small fluctuations in liquid level $-\frac{a\Delta}{\alpha 2\sqrt{h_0}}$.

B. State Space Representation

The state space representation of a linear CPS is as follows:

$$x_{i+1} = Ax_i + Bu_i + w_i \tag{3}$$

$$y_i = Cx_i + Du_i + v_i \tag{4}$$

Eq. (3) models the evolution of the CPS. At time instant i, given input u_i, state x_i is transformed into state x_{i+1}, where the index i is in $\mathbb{Z}^+$, state column vectors x_i and x_{i+1} are in $\mathcal{X} \subseteq \mathbb{R}^m$, input column vector u_i is in $\mathcal{U} \subseteq \mathbb{R}^p$, output column vector y_i is in $\mathcal{Y} \subseteq \mathbb{R}^n$ and dimensions m, n and p are in $\mathbb{Z}^+$. The transition may also be affected by random noise w_i, in $\mathbb{R}^m$. Eq. (4) represents the CPS input, state and output relation. At time instant i and in state x_i, the sensor measurements are y_i. The sensor measurements may be also affected by random noise represented by v_i, in $\mathbb{R}^n$. Matrices A, B, C and D are respectively called the state (m by m), input (m by p), output (n by m), and direct transmission (n by p) matrices.

For example, let us map Eq. (2) to a state-space representation. The input u_i is the voltage, null at start. Let t be the continuous time corresponding to the discrete time instant i. The state variable x_i tracks the difference between the liquid level $h(t)$ and operating point h_0, i.e., $x_i = h(t) - h_0$, which is the symbol Δ in Eq. (2). The corresponding state, input, output and direct transmission matrices are:

$$A = \left(-\tfrac{a}{\alpha 2\sqrt{h_0}}\right), B = \left(\tfrac{\gamma\kappa}{\alpha}\right), C = \left(1\right) \text{ and } D = \left(0\right) \tag{5}$$

The state vector has one element $x_i[1]$, which is the current level difference Δ, w.r.t. the operating point h_0. The state matrix A contains one element, which is used to calculate in a transition, from time i to time $i+1$, the change in the amount of liquid leaving the tank, i.e., $\frac{a}{\alpha 2\sqrt{h_0}} \cdot x_i[1]$. Note that at the operating point, the total amount of liquid leaving the tank is the subtrahend in Eq. (2), divided by α. The input

vector has a single element $u_i[1]$. The input matrix B contains one element and calculates the amount of liquid coming into the tank in one transition, i.e., the product $\frac{\gamma\kappa}{\alpha} \cdot u_i[1]$. Note that this mapping has the linearity advantage, but fidelity if limited to small fluctuations around an operating point. As we move from the operating, the effect of gravity is distorted. This degree of fidelity is although more than sufficient for the type of analysis conducted in the sequel of this paper.

C. Adversary Model

We assume adversaries perpetrating covert attacks. Covert attacks are a family of cyber-physical attacks in which the adversary perturbs the state of a CPS while succeeding to evade detection, i.e., the adversary attempts to remain invisible [21], [24]. It is powerful attack because it is assumed that the adversary knows the plant dynamics (matrices A, B, C and D) and that input and output signals can be spoofed. While an attack is being carried out, the perpetrator manipulates the measurements to conceal the effect of the spoofed inputs. Hence, from the point of view of an observer, responsible for detecting attacks, the measurements look normal. Using Eqs. (3) and (4), attacks are represented as follows:

$$x_{i+1} = Ax_i + B(u_i + u_i^a) + w_i \tag{6}$$

$$y_i = Cx_i + D(u_i + u_i^a) + v_i + s_i^a \tag{7}$$

The variable u_i^a, in $\mathcal{U}$, denotes the addition of the adversary to the signals to the actuators. The term s_i^a, in $\mathbb{R}^n$, represents the manipulation done by the adversary on the sensor measurements.

D. Defense Methods

Methods have been devised to detect covert attacks. They all require the analysis of inputs and outputs of the plant. Rubio-Hernan et al. [14], [15], [16] have revisited challenge-response detectors via authentication techniques, initially proposed by Mo et al. [9], [10], [23]. Hoehn and Zhang [6] and Schellenberger and Zhang [17] developed the idea of external synthetic states that evolve in parallel and are coupled to the physical states of the CPS.

Adversaries can apply system identification [20] and machine learning [7], [20] to infer the dynamics of the plant. All detection methods acknowledge that the adversary has the ability to learn the dynamics of the CPS. However, they are all based on the important assumption that the knowledge of the adversary is not perfect. Due to this imperfect knowledge, the adversary makes errors that may be caught by the detection methods. Whether they are caught or not depends on the degree of knowledge of the adversary and the level of difficulty to avoid being detected. To make it challenging, detection methods comprise the integration of time-varying elements (inputs or states) concealed in the dynamics of the plant. Assuming the parameters of these elements are changed fast enough, the dynamics of the plant becomes a moving target for the adversary. In other words, the adversary does not have enough time to learn properly, makes errors and perpetrates attacks that are not covert. Next, we discuss in more details the concepts of challenge-response and auxiliary state.

1) Challenge-response Authentication: Challenge-response detectors, defined in [14], [15], [16], revisit the authentication signal in [9], [10] to extend error detectors into cyber-physical attack detectors. The resulting scheme provides a real-time protection of the linear time-invariant models of the plant. Built upon *Kalman filters* and *linear-quadratic regulators*, the scheme produces authentication signals to protect the integrity of physical measurements communicated over the cyber and physical control space of a networked control system. It is assumed that, without the protection of the networked messages, malicious actions can be conducted to mislead the system towards unauthorized or improper actions, i.e., by disrupting the plant services.

Assume u_i^* as the output of a controller and u_i the control input that is sent to the plant, cf. Eq. (3). The idea of challenge-response authentication is to superpose to the control law u_i^* an authentication signal $\Delta u_i \in \mathbb{R}^p$ that serves to detect integrity attacks. Thus, the control input u_i is given by:

$$u_i = u_i^* + \Delta u_i \tag{8}$$

The authentication signal is a Gaussian random signal with zero mean that is independent both from the state noise (w_i) and measurement noise (v_i). The authentication signal is used by the detector to identify the malicious signals originated by the adversary. Since the control law u_i^* carries the authentication signal Δu_i, the detector (physically co-located within the controller) triggers an alarm whenever a malicious signal is observed, i.e., whenever the challenge sent by the controller over the plant is not observed within the measurements returned by the plant. Towards this end, [9], [10] propose to employ a χ^2 detector, i.e., a well-known category of real-time anomaly detectors classically used for fault detection in control systems [2], for the purpose of signaling the anomalies identified in the behavior of the plant.

Further details about some more powerful challenge-response detectors, capable of identifying adversaries which are empowered by identification tools such as ARX (autoregressive with exogenous input) and ARMAX (autoregressive-moving average with exogenous input) [11], i.e., using identification tools to evade detection, are available in [15], [16].

2) Auxiliary States: The CPS can also be augmented with a synthetic auxiliary state, synthetic outputs and optionally new inputs [6], [17]. The auxiliary state has a linear time-varying dynamics that is evolved in parallel with the CPS. The dynamics is concealed to the adversary. Because it is

time-varying, it becomes a moving target that is challenging to identify by an adversary, a precondition to the covert attack. But, it is known to and used by the operator to detect the covert attack. The operator is in synchrony with the linear time-varying dynamics. It is therefore able to track it properly and compare the actual evolution of the auxiliary dynamics with the expected evolution. Significant discrepancies indicate the presence of anomalies, which can be used to identify the adversary.

The CPS model is extended with the auxiliary state $\tilde{x}_i$ and additional actuators and sensors ($\tilde{u}_i$ and $\tilde{y}_i$) related to the auxiliary state. The state x_i and auxiliary state $\tilde{x}_i$ are correlated. Together with the auxiliary state, the state transformation model is:

$$\begin{pmatrix}\tilde{x}_{i+1}\\ x_{i+1}\end{pmatrix} = \mathcal{A}_i \begin{pmatrix}\tilde{x}_i\\ x_i\end{pmatrix} + \mathcal{B}_i \begin{pmatrix}\tilde{u}_i\\ u_i\end{pmatrix} + \begin{pmatrix}\tilde{w}_i\\ w_i\end{pmatrix} \tag{9}$$

Together with the additional elements, the sensor measurements are:

$$\begin{pmatrix}\tilde{y}_i\\ y_i\end{pmatrix} = \mathcal{C}_i \begin{pmatrix}\tilde{x}_i\\ x_i\end{pmatrix} + \mathcal{D}_i \begin{pmatrix}\tilde{u}_i\\ u_i\end{pmatrix} + \begin{pmatrix}\tilde{v}_i\\ v_i\end{pmatrix} \tag{10}$$

with

$$\mathcal{A}_i = \begin{pmatrix}A_{1,i} & A_{2,i}\\ 0 & A\end{pmatrix}, \mathcal{B}_i = \begin{pmatrix}B_i\\ B\end{pmatrix}, \mathcal{C}_i = \begin{pmatrix}C_i & 0\\ 0 & C\end{pmatrix} \text{ and}$$

$$\mathcal{D}_i = \begin{pmatrix}D_i & 0\\ 0 & D\end{pmatrix}.$$

Hidden to the adversary, the state sub-matrices $A_{1,i}$ and $A_{2,i}$, the input matrix B_i, output matrix C_i and direct transmission matrix D_i are randomized variables. According to the approach proposed by Schellenberger and Zhang [17], the actual matrices are randomly switched from time-to-time. The operator and CPS are synchronized on the switching sequence, perhaps through a switching signal. This secret is not shared with the adversary. Sensor measurement $\tilde{y}_i$ is visible to the adversary, but changes over time in a random way. The adversary is challenged with learning the random auxiliary system state, input, output and direct transmission matrices.

We have introduced the system and adversary models and reviewed defense methods. In the next sections, we build upon that material and introduce new ideas to address resilience and state recovery.

III. The (k, ℓ)-Resilient Property

We define the k-steerability and ℓ-monitorability properties, which in conjunction define the (k, ℓ)-resilient property.

A. Inter-variable Dependencies

Firstly, to identify a dependency between two variables, we consider the use of the Pearson correlation coefficient.

Definition 1 (Pearson correlation coefficient): Given two random variables, A and B, and n observations for each of them, their correlation coefficient is defined by

$$\rho(A, B) = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{a_i - \mu_A}{\sigma_A}\right) \left(\frac{b_i - \mu_B}{\sigma_B}\right) \tag{11}$$

where $a_1, \ldots, a_n$ ($b_1, \ldots, b_n$), μ_A (μ_B) and σ_A (σ_B) are the observations, mean and standard deviation of random variable A (B).

A correlation coefficient is a unitless value between minus one and one. When $\rho(A, B)$ is equal to one, we have perfect positive correlation between A and B. When it is minus one, we have perfect negative correlation. Intuitively, when $|\rho(A, B)|$ is between zero and 0.2, the linear correlation is from null to weak. It is moderate between 0.2 and 0.6. Above 0.6, it is strong [22]. Note that null linear correlation does not mean necessarily that variables A and B are independent. In such a case, there is no linear dependency revealed by the observations, but a nonlinear dependency is possible. For example, Eq. (1) generates nonlinear output correlated with the input. Existence of correlation can be confirmed calculating the correlation coefficient using a linearized version of the output data. Furthermore, correlation is one way to establish dependencies between variables.

B. Dependency Graph

Let u, x and y be respectively p-element, m-element and n-element column vectors representing the input, state and output variables of a CPS. We define correlation coefficient matrices to capture in structures the relationships between state variables and input or output variables.

Definition 2 (Input and output correlation coefficient matrices): The $m \times p$ [$m \times n$] input [output] correlation coefficient matrix Q [R] is equal to $(q_{i,j})$ [$(r_{i,j})$] where $i = 1, \ldots, m$, $j = 1, \ldots, p$ [$j = 1, \ldots, n$]. An entry $q_{i,j}$ [$r_{i,j}$] is the correlation coefficient $\rho(x_i, u_j)$ [$\rho(x_i, y_j)$] between state variable x_i and input [output] variable u_j [y_j].

Definition 3 (Input and output dependency graphs): The input [output] dependency graph is a bipartite graph $G_U = (X, U, E)$ [$G_Y = (X, Y, E)$] where the two sets of vertices are $X = \{x_1, \ldots, x_m\}$ and $U = \{u_1, \ldots, u_p\}$ [$Y = \{y_1, \ldots, y_n\}$] the state and input [output] variables, respectively. When Pearson correlation is used to determine dependencies, there is an edge (x_i, u_i) [(x_i, y_i)] in E if-and-only-if the absolute value of the correlation between x_i and u_i [y_i], i.e., $|q_{i,j}|$ [$|r_{i,j}|$], is greater than or equal to a threshold T. The value of T is normally chosen to be close to one.

For the dependency graph G_U [G_Y] and a vertex x in X, let the expression $\deg(x)$ be its input [output] degree, i.e., the number of adjacent vertices in U [Y]. The ℓ-monitorability degree reflects the availability of at least ℓ sensor output signals for monitoring any state variable.

Definition 4 (ℓ-monitorability degree[1]): Let G_Y be the output dependency graph of a CPS. Let ℓ be equal to

$$\min_{x \in X} \deg(x).$$

Then, the CPS has ℓ-monitorability.

We introduce the notion of k-steerability. It indicates that there are at least k actuator input signals available for acting on every single plant state variable.

Definition 5 (k-steerability degree[2]): Let G_U be the input dependency graph of a CPS. Let k be equal to

$$\min_{x \in X} \deg(x).$$

Then, the CPS has k-steerability.

Definition 6 ((k, ℓ)-resilient): A CPS with k-steerability and ℓ-monitorability ($k \leq \ell$) is said to be (k, ℓ)-resilient.

There is a desirable relationship between k and ℓ. Increasing the steerability of all state variables (k), i.e., augmenting the number of actuators that can change each of them, offers more possibilities to act on the CPS. These additional possibilities can be leveraged to mitigate covert attacks. They, however, also increase the attack surface. In several instances, this increase can be mitigated by an equivalent number of observation points (ℓ), i.e., number of sensor outputs. Let us consider as an example the approach used for *in-home delivery* of parcels. In the absence of the customer, the driver is granted the ability to unlock the door and place a parcel inside the house. Unlocking the door to a stranger in your absence is risky. The attack surface is augmented. Nevertheless, for in-home delivery the risk is mitigated by the activation of a cloud camera filming all the delivery operation. In this case, the camera constitutes an additional point of observation. Hence, the relationship $k \leq \ell$ is desirable. That is to say, for any state variable when there are k different points for acting on it, there are also ℓ different observation points for monitoring the actions. In a design with $k > \ell$, there is higher steerability than monitorability. This configuration should be viewed as wrongly designed since the budget dedicated to the system resilience is not appropriately used. For such a configuration, the designer should redesign the system in order to increase monitorability by adding more sensors. The aim is the design of a CPS that has (k, ℓ)-resilient, where $\ell \geq k > 1$.

From the point of view of risk assessment, the (k, ℓ)-resilient property plays a role. Among other things, risk assessment evaluates the importance of threats taking into account the technical difficulties that must be overcome to perpetrate the attacks [4]. Risks are mitigated when the adversary has to traverse more technological barriers. For a given CPS, risk mitigation of covert attacks can be put into perspective comparing the different options for k and ℓ.

Definition 7 (Comparability): Let $c_1 = (k_1, \ell_1)$ and $c_2 = (k_2, \ell_2)$, then we have that $c_1 \leq c_2$ if-and-only-if $k_1 \leq k_2$ and $\ell_1 \leq \ell_2$. In such a case c_1 and c_2 are comparable.

Let us consider two CPS designs c_1 and c_2 that have the very same dynamics, with $c_1 \leq c_2$ while the c_2 design has everything that the c_1 design has, plus additional actuators and sensors. The design c_2 is more steerable and monitorable than the design c_1. The (k, ℓ)-resilient property can be used as a measure to compare the risk associated with two given configurations. We have established the principles of our approach. In the following section, we review a number of comparable designs and explain how the (k, ℓ)-resilient property translates into possibilities of recovering the state of a CPS when attacks are perpetrated.

IV. Conclusion

We have addressed covert attacks on CPS. We have defined the new k-steerability and ℓ-monitorability control-theoretic concepts. The k-steerability concept reflects the ability in a CPS to act on each of its individual plant state variables with at least k functionally diverse groups of input signals. In other words, it reflects the ability of the CPS to mitigate the impact of covert attacks when less than k groups of input signals are compromised, using static functional diversity. The ℓ-monitorability concept reflects the number of observations on each state variable of a CPS that can be used to identify covert attacks. Together, k-steerability and ℓ-monitorability determine the (k, ℓ)-resilient property of a CPS. If we assume that the detection process is conducted by combining strategies, such as redundancy and diversity in hardware and software techniques, the resulting (k, ℓ)-resilient concept applies the moving target paradigm, in which the CPS adapts itself to invalidate the acquired knowledge of the adversaries.

Acknowledgments — We acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the European Commission (H2020 SPARTA project, under grant agreement 830892).

References

[1] C. Barreto, A. A. Cárdenas, and N. Quijano. Controllability of dynamical systems: Threat models and reactive security. In *International Conference on Decision and Game Theory for Security*, pages 45–64. Springer, 2013.

[2] B. Brumback and M. Srinath. A chi-square test for fault-detection in Kalman filters. *IEEE Transactions on Automatic Control*, 32(6):552–554, Jun 1987.

[1] For simplicity, we refer to ℓ-monitorability hereafter.

[2] k-steerability for short.

[3] A. Chapman and M. Mesbahi. Security and infiltration of networks: A structural controllability and observability perspective. In *Control of Cyber-Physical Systems*, pages 143–160. Springer, 2013.
[4] ETSI. Telecommunications and internet protocol harmonization over networks (TIPHON) release 4; protocol framework definition; methods and protocols for security; part 1: Threat analysis. Technical Report Technical Specification ETSI TS 102 165-1 V4.1.1, European Telecommunications Standards Institute (ETSI), Januray 2003.
[5] G. Franklin, J. Da Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems*. Pearson Education, 2014.
[6] A. Hoehn and P. Zhang. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In *2016 American Control Conference (ACC)*, pages 302–307, July 2016.
[7] G. Horvath. Neural networks in system identification. *Nato Science Series Sub Series III Computer And Systems Sciences*, 185:43–78, 2003.
[8] E. Lee. Cyber-physical systems - are computing foundations adequate? In *NSF Workshop On Cyber-Physical Systems*, 01 2006.
[9] Y. Mo, R. Chabukswar, and B. Sinopoli. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, July 2014.
[10] Y. Mo and B. Sinopoli. Secure control against replay attacks. In *Communication, Control, and Computing. 47th Annual Allerton Conference on*, pages 911–918. IEEE, Sept 2009.
[11] H. Natke. System identification: Torsten Söderström and Petre Stoica. *Automatica*, 28(5):1069–1071, 1992.
[12] K. Ogata. *Modern Control Engineering*. Pearson Education, 2011.
[13] B. Ramasubramanian, M. Rajan, and M. G. Chandra. Structural resilience of cyberphysical systems under attack. In *2016 American Control Conference (ACC)*, pages 283–289. IEEE, 2016.
[14] J. Rubio-Hernan, L. De Cicco, and J. Garcia-Alfaro. Revisiting a watermark-based detection scheme to handle cyber-physical attacks. In *Availability, Reliability and Security (ARES), 2016 11th International Conference on*, pages 21–28. IEEE, August 2016.
[15] J. Rubio-Hernan, L. De Cicco, and J. Garcia-Alfaro. Adaptive control-theoretic detection of integrity attacks against cyber-physical industrial systems. *Trans. Emerging Telecommunications Technologies*, 32(09), 2017.
[16] J. Rubio-Hernan, L. De Cicco, and J. Garcia-Alfaro. On the use of watermark-based schemes to detect cyber-physical attacks. *EURASIP Journal on Information Security*, 2017(1):8, 2017.
[17] C. Schellenberger and P. Zhang. Detection of covert attacks on cyber-physical systems by extending the system dynamics with an auxiliary system. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1374–1379, December 2017.
[18] R. Smith. A decoupled feedback structure for covertly appropriating networked control systems. *IFAC Proceedings Volumes*, 44(1):90–95, 2011. 18th IFAC World Congress.
[19] R. Smith. Covert Misappropriation of Networked Control Systems: Presenting a Feedback Structure. *IEEE Control Systems*, 35(1):82–92, Feb 2015.
[20] A. K. Tangirala. *Principles of System Identification: Theory and Practice*. CRC Press, 2014.
[21] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135–148, 2015.
[22] S. Tutorials. Pearson Correlation. Retrieved on April 2020. https://libguides.library.kent.edu/SPSS/PearsonCorr, 2014.
[23] S. Weerakkody, Y. Mo, and B. Sinopoli. Detecting integrity attacks on control systems using robust physical watermarking. In *53rd IEEE Conference on Decision and Control*, pages 3757–3764, December 2014.
[24] S. Weerakkody, O. Ozel, Y. Mo, B. Sinopoli, et al. Resilient control in cyber-physical systems: Countering uncertainty, constraints, and adversarial behavior. *Foundations and Trends® in Systems and Control*, 7(1-2):1–252, 2020.