Preprints of the 22nd IFAC World Congress
Yokohama, Japan, July 9-14, 2023

Quantifying Impact on Safety from Cyber-Attacks on Cyber-Physical Systems ⋆

Eleftherios Vlahakis * **Gregory Provan** ** **Gordon Werner** ***
Shanchieh Yang *** **Nikolaos Athanasopoulos** *

* *Queen's University Belfast, UK (e-mail: {e.vlahakis, n.athanasopoulos}@qub.ac.uk).*
** *University College Cork, Ireland (e-mail: g.provan@cs.ucc.ie)*
*** *Rochester Institute of Technology, Rochester, USA, (e-mail: gxw9834@rit.edu, jay.yang@rit.edu)*

Abstract: We propose a novel framework for modeling attack scenarios in cyber-physical control systems: we represent a cyber-physical system as a constrained switching system, where a single model embeds the dynamics of the physical process, the attack patterns, and the attack detection schemes. We show that this is compatible with established results in hybrid automata, namely, constrained switching linear systems. The proposed attack modeling approach admits a large class of non-deterministic attack policies and enables the characterization of system safety as an asymptotic property. By calculating the maximal safe set, the resulting new impact metrics intuitively quantify the degradation of safety and the impact of cyber attacks on the safety properties of the system under attack. We showcase our results via an illustrative example.

Keywords: cyber-physical systems, cyber-security, attack modeling, regular language representation, constrained switching systems, safety.

1. INTRODUCTION

Cyber-physical systems (CPSs) represent a broad spectrum of safety-critical applications, ranging from power generation and distribution networks to autonomous mobility and industrial processes. Due to their extent and intrinsic link to society, the secure operation of such schemes is vital. Vulnerability to cyber attacks typically depends on the degree of integrating unsafe communication channels between computation, sensing, and actuation modules that control the underlying physical process.

CPS security (Sandberg et al., 2022) studies control problems under adversarial actions aiming to steer a control system into an unsafe region. Our modeling approach is motivated to an extent by the literature of networked control systems (Hespanha et al., 2007), where communication limitations and malfunctions (e.g., packet dropouts) can be embedded into a hybrid control system model with switching dynamics (Donkers et al., 2011; De Persis and Tesi, 2015). Due to their inherent complexity, CPSs are often modeled via hybrid systems. For example, hybrid linear automata, finite state machines, and Petri nets are important tools for modeling malicious and unpredictable behaviors, and threat propagations in CPSs (Beg et al., 2017; Meira-Góes et al., 2020).

Focusing on attack patterns that can be expressed via regular languages on directed labeled graphs (Cassandras and Lafortune, 2010), we propose a constrained switching systems framework for analyzing safety properties of CPSs. Although invariance and safety of constrained switching systems have received attention (De Santis et al., 2004), (Athanasopoulos and Lazar, 2014), they have not yet been studied in the context of CPS security. By modeling the overall attack scheme as a constrained switching system, our objective is to characterize the set of all initial states that cannot be driven to an unsafe state under any allowable attack. We call this the *safe set* of the attacked CPS. This is typically an infinite-reachability dynamic programming problem (Raković et al., 2006): the maximal safe set can be retrieved by computing recursively the fixed point of the sequence of sets $\{S_i\}_{i\in\{1,2,\dots\}}$ with $S_{i+1} = \mathrm{Pre}(S_i) \cap S_0$, where $S_0 = X_0$ denotes the state constraint set, and $\mathrm{Pre}(S_i)$ is the *preimage map* (Bertsekas, 1972), e.g., that is the set of states x for which, for all permissible attack patterns, the successor state $x^+ \in S_i$.

There has been significant research on modeling temporal aspects of attacks, with most of the works pertaining to automata-based approaches without considering timing behaviors or multiple simultaneous attack scenarios (Chen et al., 2003). Few papers have introduced general languages for describing attack patterns (Reda et al., 2022). One of the most comprehensive formal languages has been proposed in (Liu et al., 2017) building on probabilistic colored Petri nets combined with a game-theoretic mixed strategy. Also, most works derive receding horizon impact metrics of attacks or focus on a finite set of outcomes (Miao and Zhu, 2014). Our approach is aligned with the reachability-analysis-based works (Murguia et al., 2020; Mo and Sinopoli, 2016), providing asymptotic results on the closed-loop system safety. We treat stealthy attack perturbations as state- and/or input-dependent exogenous

⋆ E.V. and N.A. gratefully acknowledge support from EPSRC EP/T021942/1, and N.A. additionally from EU 2020-1-UK01-KA203-079283 and the UKRI Belfast Maritime Consortium 107138.

signals (Raković et al., 2006). To our knowledge, there is limited work dealing with state-dependent attacks directly in CPSs. Our contributions are summarised as follows:

We model the overall CPS under attack as a constrained switching system with the switching signal forming a regular language, generated by a nondeterministic directed graph. Each node corresponds to a set of states that evolve with time according to the modes assigned to its outgoing edges. Each labeled edge describes either an attack-free operation or a specific malicious action carried out over a subset of unsafe channels. The proposed approach permits the modeling of a large family of non-deterministic attacks, the impact of which can be quantified and analyzed asymptotically via the construction of maximal safe sets. To compute the maximal safe set of the system subject to all admissible attack sequences, we leverage reachability analysis techniques related to the notion of multi-set invariance (Athanasopoulos and Jungers, 2018). Based on the constructed sets, we assess vulnerability by two complementary security metrics, related to the Lebesgue measure and the Minkowski distance, providing scalar indices of system attack sensitivity.

The remainder of the paper is organized as follows. In Section 2, we present the family of systems we study and the type of attacks we are interested in. The main results, namely, the constrained switching system formulation, the safe set computations, and the introduction of scalar safety metrics, are in Section 3. A numerical example and concluding remarks are in Sections 4 and 5, respectively.

2. SYSTEM DESCRIPTION

In the following, we present the notation and the system under attack. See Fig. 1 for an illustration.

2.1 Notation

The transpose of ξ is $\xi^\top$. The $m \times m$ identity matrix is I_m and the vector with elements equal to one is $1 \in \mathbb{R}^n$. The jth row of matrix A and jth element of vector a are denoted by $(A)_j$ and $(a)_j$, respectively. The set of row indices of A is J_A. We write $\mathcal{G}(\mathcal{V}, \mathcal{E})$, or $\mathcal{G}$, a labeled directed graph with a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. We denote the p-norm of a vector x by $\|x\|_p$. $\mathbb{B}(\alpha)$, and $\mathbb{B}_\infty(\alpha)$ denote the balls of radius α of an arbitrary norm, and the infinity norm, respectively. The Minkowski sum of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ is denoted by $\mathcal{S}_1 \oplus \mathcal{S}_2$. The interior and the convex hull of a set $\mathcal{S}$ are denoted as $\text{int}(\mathcal{S})$ and $\text{conv}(\mathcal{S})$, respectively. A C-set $\mathcal{S} \subset \mathbb{R}^n$ is a convex compact polytopic set that contains the origin in its interior (Blanchini and Miani, 2015). By convention, for any C-set $\mathcal{S}$, we write its half-space representation by $\mathcal{S} = \{x : G_s x \leq g_s\}$ with the inequality applied elementwise. The cardinality of a set $\mathcal{V}$ is denoted by $|\mathcal{V}|$.

2.2 Dynamics, Controller, Estimator, Detector

We study discrete-time linear time-invariant (LTI) systems

$$P : \begin{cases} x(t+1) = A_p x(t) + B_p u(t) + v(t), \\ y(t) = C_p x(t) + w(t), \end{cases} \tag{1}$$

where $t \in \mathbb{N}$, $x(t) \in \mathcal{X} \subset \mathbb{R}^{n_x}$, $u(t) \in \mathcal{U} \subset \mathbb{R}^{n_u}$ and $y(t) \in \mathcal{Y} \subset \mathbb{R}^{n_y}$ are the state, input and output vectors,

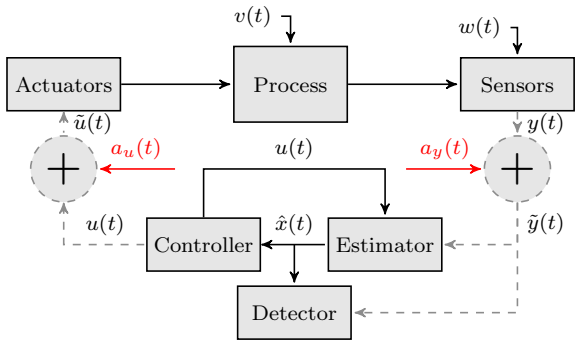

Fig. 1. Networked control loop with unsafe communication channels illustrated by dashed gray lines.

respectively, vectors $v(t) \in \mathcal{V} \subset \mathbb{R}^{n_x}$ and $w(t) \in \mathcal{W} \subset \mathbb{R}^{n_y}$ denote process and measurement uncertainties. We assume that $\mathcal{X}, \mathcal{U}, \mathcal{Y}, \mathcal{V}$, and $\mathcal{W}$ are C-sets. For meaningful control and estimation schemes, we assume the following.

Assumption 1. Pairs $(A_p,\ B_p)$, $(A_p^\top,\ C_p^\top)$ are stabilizable.

We are interested in *false data injection* (FDI) attacks poisoning sensors and input signals. Consistent with the notation in Fig. 1, we model the attacked output as

$$\tilde{y}(t) = y(t) + \Gamma_i^y a_y(t), \tag{2}$$

where $a_y(t) \in \mathbb{R}^{n_{\tilde{y}}}$ denotes *additive sensor poisoning attacks*, and $\Gamma_i^y \in \mathbb{R}^{n_y \times n_{\tilde{y}}}$, with $n_{\tilde{y}} \leq n_y$ denoting the number of vulnerable sensors. The subscript i denotes the ith attack strategy.[1] The jth row of Γ_i^y is $0 \in \mathbb{R}^{1 \times n_{\tilde{y}}}$ if the jth sensor is not corrupted under the ith attack action. Otherwise, it is the $\tilde{j}$th vector $\epsilon_{\tilde{j}}$ of the canonical basis of $\mathbb{R}^{n_{\tilde{y}} \times n_{\tilde{y}}}$, with $\tilde{j}$ denoting the index of a vulnerable sensor under attack by the ith attack action.

We consider a dynamic output feedback control law $u(t) = -K\hat{x}(t)$, where $K \in \mathbb{R}^{n_u \times n_x}$, and $\hat{x}(t)$ is obtained by

$$\hat{x}(t+1) = A_p\hat{x}(t) + B_p u(t) + L(\tilde{y}(t) - C_p\hat{x}(t)), \tag{3}$$

with $L \in \mathbb{R}^{n_x \times n_y}$. We call $r(t) = \tilde{y}(t) - C_p\hat{x}(t) \in \mathbb{R}^{n_y}$ the *residual* and $e(t) = x(t) - \hat{x}(t)$ the *estimation error*. From (1)-(3), we may write

$$\begin{cases} e(t+1) = (A_p - LC_p)e(t) - L\Gamma_i^y a_y(t) - Lw(t) \\ r(t) = C_p e(t) + \Gamma_i^y a_y(t) + w(t). \end{cases} \tag{4}$$

Remark 2. Under Assumption 1, a desirable robust performance for the attack-free system (1) can be achieved using robust control approaches, e.g., by constructing K and L via LMI-based algorithms (Gahinet and Apkarian, 1994).

The received control signal is corrupted as $\tilde{u}(t) = u(t) + \Gamma_i^u a_u(t)$, where $a_u(t) \in \mathbb{R}^{n_{\tilde{u}}}$ denotes *additive input poisoning attacks*, and $\Gamma_i^u \in \mathbb{R}^{n_u \times n_{\tilde{u}}}$, with $n_{\tilde{u}} \leq n_u$ denoting the number of unsafe channels over which actuation signals are transmitted. The structure of Γ_i^u is associated with the ith attack action and is defined similarly as Γ_i^y defined in (2).

An alarm is raised at $t \geq 0$ if $r(t) \notin \mathcal{R}$ with

$$\mathcal{R} = \{r \in \mathbb{R}^{n_y} : G_r r \leq h_r\}, \tag{5}$$

where G_r is a full row-rank matrix.

[1] See Section 2.3.

Remark 3. The polyhedral set $\mathcal{R}$ may be designed such that the number of false alarms is minimized.

Remark 4. Our anomaly detector is *stateless*. *Stateful* detectors with linear, or convex, dynamics can be accepted in our framework, (Milošević et al., 2018).

By defining augmented vectors $z(t) = [x(t)^\top \;\; e(t)^\top]^\top$, $a(t) = [a_u(t)^\top, \; a_y(t)^\top]^\top$, and $\eta(t) = [v(t)^\top, \; w(t)^\top]^\top$, the closed-loop dynamics under the ith attack action are

$$P_i : \begin{cases} z(t+1) = Az(t) + B_i a(t) + E\eta(t), \\ r(t) = Cz(t) + D_i a(t) + F\eta(t), \end{cases} \tag{6}$$

where

$$A = \begin{bmatrix} A_p - B_pK & B_pK \\ 0_{n_x \times n_x} & A_p - LC_p \end{bmatrix}, \; B_i = \begin{bmatrix} B_p\Gamma_i^u & 0_{n_x \times n_{\tilde{y}}} \\ 0_{n_x \times n_{\tilde{u}}} & -L\Gamma_i^y \end{bmatrix},$$
$$E = \begin{bmatrix} I_{n_x} & 0_{n_x \times n_y} \\ I_{n_x} & -L \end{bmatrix}, \; C = \begin{bmatrix} 0_{n_y \times n_x} & C_p \end{bmatrix},$$
$$D_i = \begin{bmatrix} 0_{n_y \times n_{\tilde{u}}} & \Gamma_i^y \end{bmatrix}, \; F = \begin{bmatrix} 0_{n_y \times n_x} & I_{n_y} \end{bmatrix}.$$

Switching between a set of attack actions is discussed next.

2.3 Attack patterns

We study attack actions made up of a set of logic rules (e.g., dwell-time) that can be expressed via a regular language (Cassandras and Lafortune, 2010, Chapter 2.4), leading to an overall attack policy described by a directed labeled graph. An edge indicates a set of attack operations and is associated with a specific dynamic mode (see (6)). Fig. 2 illustrates an example of how dwell-time restrictions and admissible sequences are modeled for attack tactics applied to two independent channels. An edge with a label 'N' denotes attack-free, nominal operation, whereas a label 'A' implies FDI attack on a channel. In channel I, we assume that FDI attacks cannot happen more than two consecutive time steps, whereas, in channel II, an FDI attack has to be followed by a nominal operation.

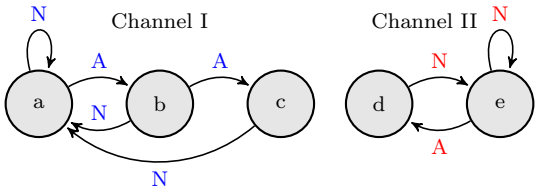

Fig. 2. Graph examples representing attack patterns.

Recalling Fig. 1, CPSs can have several vulnerable points and be subject to complex and varying attack actions. Fig. 3 shows the Kronecker product of the two graphs in Fig. 2 illustrating all possible combinations of two attacks acting simultaneously in different channels of the same CPS. Due to space limitations, see, e.g., (Saltik et al., 2015), for details on computing Kronecker products of multiple graphs.

3. MAIN RESULTS

3.1 Switching-system attack modeling

We consider a set of systems $\mathcal{P} = \{P_1, \ldots, P_N\}$, with P_i denoting the dynamics of mode i given in (6) and associated with an attack action. We describe an overall attack pattern, i.e., the switch between dynamic modes, by a directed labeled graph $\mathcal{G}(\mathcal{V}_\mathcal{G}, \; \mathcal{E}_\mathcal{G})$. Let the set of outgoing nodes of a node $s \in \mathcal{V}_\mathcal{G}$ be denoted by $\text{Out}(s, \; \mathcal{G}) := \{d \in \mathcal{V}_\mathcal{G} : (\exists \sigma \in \{1, \ldots, N\} : (s, d, \sigma) \in \mathcal{E}_\mathcal{G})\}$. We denote by n_z, n_a, and n_h, the augmented state, attack, and disturbance dimensions, respectively. We also consider the Cartesian product of the disturbance and uncertainty sets $\mathcal{H} = \mathcal{V} \times \mathcal{W}$. The dynamics of the overall attacked system is

$$z(t+1) = Az(t) + B_{\sigma(t)}a(t) + Eh(t), \tag{7}$$
$$\xi(t+1) \in \text{Out}(\xi(t), \; \mathcal{G}(\mathcal{V}_\mathcal{G}, \; \mathcal{E}_\mathcal{G})), \tag{8}$$
$$(z(0), \; \xi(0)) \in \mathcal{Z} \times \mathcal{V}_\mathcal{G}, \tag{9}$$

subject to the constraints

$$(\xi(t), \; \xi(t+1), \; \sigma(t)) \in \mathcal{E}_\mathcal{G}, \tag{10}$$
$$z(t) \in \mathcal{Z}, \tag{11}$$
$$h(t) \in \mathcal{H} \tag{12}$$
$$a(t) \in \mathcal{A}_{\sigma(t)}(z(t)), \tag{13}$$

for all $t \geq 0$, where $[z^\top \; \xi]^\top \in \mathbb{R}^{n_z} \times \mathcal{V}_\mathcal{G}$. We note that $\sigma = 1$ corresponds to the *nominal* attack-free dynamics, and the autonomous system $z(t+1) = Az(t)$ is stable by the stability of $A_p - B_pK$, $A_p - LC_p$.

Remark 5. Matrices A and E, respectively, remain identical for all modes of (7). Attack actions altering these matrices can also be considered in our framework.

Assumption 6. The constraint and disturbance sets $\mathcal{Z}$, $\mathcal{H}$, are C-sets.

Assumption 7. The sets $\text{Out}(i, \; \mathcal{G}(\mathcal{V}_\mathcal{G}, \; \mathcal{E}_\mathcal{G}))$, with $i \in \mathcal{V}_\mathcal{G}$, are nonempty.

Assumption 8. The attacker is aware of all system matrices, the control, estimation, and detection schemes, and the state, input, output and disturbance constraint sets.

Assumption 6 is standard, see e.g., (Blanchini and Miani, 2015). Assumption 7 guarantees the completeness of solutions. Assumption 8 is standard for the construction of stealthy data poisoning attacks.

The constraint (13) enforces attack stealthiness: First, we require that attacked inputs $\tilde{u}(t) = u(t) + \Gamma^u_{\sigma(t)}a_u(t) \in \mathcal{U}$, where $\mathcal{U} = \{u : G_u u \leq h_u\}$. Let $\mathcal{A}^u_\sigma(z) = \{a_u : (G_u)_j \Gamma^u_\sigma a_u \leq (h_u)_j + (G_u)_j K[I_{n_x} \; -I_{n_x}]z, \; j \in J_{G_u}\}$. We call an attack a *stealthy input attack* if $a_u(t) \in \mathcal{A}^u_{\sigma(t)}(z(t))$. The output poisoning attacks should respect two types of constraints, namely, the output constraints, i.e., $\tilde{y}(t) = y(t) + \Gamma^y_{\sigma(t)}a_y(t) \in \mathcal{Y}$, with $\mathcal{Y} = \{y : G_y y \leq$

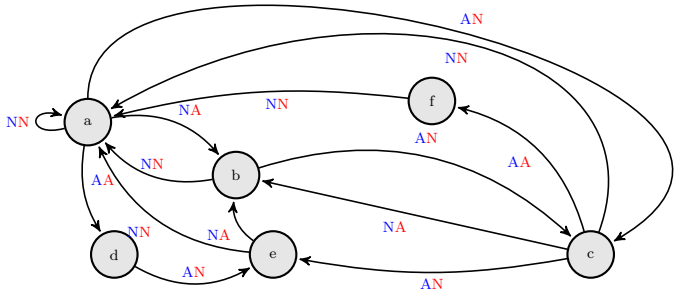

Fig. 3. All possible combinations of the attacks in Fig. 2.

$h_y\}$, and the residual constraints, i.e., $r(t) \in \mathcal{R}$, with $\mathcal{R}$ as in (5). Let $\mathcal{A}^y_\sigma(z) = \{a_y : (G_y)_j \Gamma^y_\sigma a_y \leq (h_y)_j - \max_{w\in\mathcal{W}}(G_y)_j w - (G_y)_j C_p [I_{n_x} \ 0_{n_x}] z, \ j \in J_{G_y}\}$ and $\mathcal{A}^r_\sigma(z) = \{a_y : (G_r)_j \Gamma^y_\sigma a_y \leq (h_r)_j - \max_{w\in\mathcal{W}}(G_r)_j w - (G_r)_j C_\sigma z, \ j \in J_{G_r}\}$. We call an attack a *stealthy output attack* if $a_y(t) \in \mathcal{A}^y_{\sigma(t)}(z(t)) \cap \mathcal{A}^r_{\sigma(t)}(z(t))$.

Definition 9. The attack signal $a(t) = \left[a_u(t)^\top \ a_y(t)^\top\right]^\top$ is called stealthy if $a(t) \in \mathcal{A}_{\sigma(t)}(z(t))$, where $\mathcal{A}_\sigma(z) = \mathcal{A}^u_\sigma(z) \times (\mathcal{A}^y_\sigma(z) \cap \mathcal{A}^r_\sigma(z))$, $\sigma = 1, \ldots, N$.

The H-representation of $\mathcal{A}_\sigma(z)$ is $\mathcal{A}_\sigma(z) = \{a : G_{a_\sigma} a \leq H_{a_\sigma}(z)\}$, with the inequality applied elementwise, where G_{a_σ} is a real matrix and $H_{a_\sigma}(z)$ is a convex piecewise affine function of $z \in \mathcal{Z}$, e.g., (Schaich and Cannon, 2015). The sets $\mathcal{A}_\sigma(z)$ is pointwise compact and polytopic for all $z \in \mathcal{Z}$.

3.2 Safe set computation

To define safety for system (7)-(13), we recall the notions of multi-sets and invariance.

Definition 10. (Multi-sets). We call multi-set a collection of sets $\{\mathcal{S}^i\}_{i\in\mathcal{V}_\mathcal{G}}$, with $\mathcal{S}^i \subset \mathbb{R}^{n_z}$, $i \in \mathcal{V}_\mathcal{G}$.

Definition 11. (Invariance). The multi-set $\{\mathcal{S}^i\}_{i\in\mathcal{V}_\mathcal{G}}$ is an invariant multi-set with respect to (7)-(13) if $z(0) \in \mathcal{S}^{\xi(0)}$ implies $z(t) \in \mathcal{S}^{\xi(t)}$ for all $t \geq 0$, $\xi(0) \in \mathcal{V}_\mathcal{G}$, and $\sigma(t)$ satisfying (10). If, additionally, $\mathcal{S}^i \subset \mathcal{Z}$, $i \in \mathcal{V}_\mathcal{G}$, then, $\{\mathcal{S}^i\}_{i\in\mathcal{V}_\mathcal{G}}$ is called an *admissible invariant multi-set* with respect to (7)-(13). The multi-set $\{\mathcal{S}^i_M\}_{i\in\mathcal{V}_\mathcal{G}}$ is the *maximal admissible invariant multi-set* if for any admissible invariant multi-set $\{\mathcal{S}^i\}_{i\in\mathcal{V}_\mathcal{G}}$, it holds that $\mathcal{S}^i \subseteq \mathcal{S}^i_M$, $i \in \mathcal{V}_\mathcal{G}$. The invariant multi-set $\{\mathcal{S}^i_m\}_{i\in\mathcal{V}_\mathcal{G}}$ is the *minimal invariant multi-set* if $\mathcal{S}^i_m \subseteq \mathcal{S}^i$, $i \in \mathcal{V}_\mathcal{G}$, for any invariant multi-set $\{\mathcal{S}^i\}_{i\in\mathcal{V}_\mathcal{G}}$.

Definition 12. (Safety). A set $\mathcal{S}_\mathcal{P} \subset \mathbb{R}^{n_z}$ is safe with respect to system (7)-(13) and the set of nodes $\mathcal{V}_\mathcal{G}$ if $(z(0), \ \xi(0)) \in \mathcal{S}_\mathcal{P} \times \mathcal{V}_\mathcal{G}$, implies $z(t) \in \mathcal{Z}$, $t \geq 0$.

Consider the system (7)-(9) and a switching signal $\sigma \in \{1, \ldots, N\}$. The *one-step forward reachability map* is $\Phi(\sigma, \mathcal{S}) = \{y : (\exists (z, a, h) \in \mathcal{S} \times \mathcal{A}_\sigma(z) \times \mathcal{H} : y = Az + B_\sigma a + Eh)\}$ and the *one-step backward reachability map* is $\Psi(\sigma, \ \mathcal{S}) = \{z : (A_\sigma z \oplus B_\sigma \mathcal{A}_\sigma(z) \oplus E_\sigma \mathcal{H}) \in \mathcal{S}\}$.The minimal-invariant multi-set is characterized next.

Proposition 13. Consider the forward reachability multi-set sequence $\{\mathcal{F}^i_l\}_{i\in\mathcal{V}_\mathcal{G}}$, $l \geq 0$, with

$$\mathcal{F}^i_0 = \{0\}, \ i \in \mathcal{V}_\mathcal{G}, \tag{14}$$

$$\mathcal{F}^i_{l+1} = \cup_{(s, \ i, \ \sigma)\in\mathcal{E}_\mathcal{G}} \Phi(\sigma, \ \mathcal{F}^s_l), \ i \in \mathcal{V}_\mathcal{G}. \tag{15}$$

The minimal invariant multi-set $\{\mathcal{S}^i_m\}_{i\in\mathcal{V}_\mathcal{G}}$ with respect to (7)-(13), if exists, is equal to $\mathcal{S}^i_m = \lim_{j\to\infty} \mathcal{F}^i_j$, $i \in \mathcal{V}_\mathcal{G}$.

The proof follows similar steps with (Athanasopoulos et al., 2017, Theorem 1). The difference concerns the involvement of the state-dependent set $\mathcal{A}_\sigma(z)$ in the multi-set sequence update (15), which is well defined as the sets $\mathcal{A}_\sigma(\mathcal{Z}) = \cup_{z\in\mathcal{Z}} \mathcal{A}_\sigma(z)$, $\sigma = 1, \ldots, N$ are compact by compactness of the set $\mathcal{A}_\sigma(z)$ and the compactness assumption of $\mathcal{Z}$.

Assumption 14. We assume that the minimal-invariant multi-set with respect to (7)-(13), denoted by $\{\mathcal{S}^i_m\}_{i\in\mathcal{V}_\mathcal{G}}$, exists, and that $\mathcal{S}^i_m \subset \mathcal{Z}$, $i \in \mathcal{V}_\mathcal{G}$.

We consider the backward reachability multi-set sequence $\{\mathcal{B}^i_l\}_{i\in\mathcal{V}_\mathcal{G}}$, where

$$\mathcal{B}^i_0 = \mathcal{Z}, \ \ i \in \mathcal{V}_\mathcal{G}, \tag{16}$$

$$\mathcal{B}^i_{l+1} = (\mathcal{B}^i_0 \cap_{(i, \ d, \ \sigma)\in\mathcal{E}_\mathcal{G}} \Psi(\sigma, \ \mathcal{B}^d_l)), \ i \in \mathcal{V}_\mathcal{G}. \tag{17}$$

The lth term of the multi-set sequence (16)-(17) contains the initial conditions $(z(0), \ \xi(0))$ which satisfy the state constraints for at least l consecutive instants. Intuitively, each set $\mathcal{B}^i_{l+1}$, $l \geq 0$, $i \in \mathcal{V}_\mathcal{G}$, contains the set of states in the state constraint set $\mathcal{Z}$ that can be stirred to $\mathcal{B}^d_l$ via the dynamics σ, where d is any outgoing node of i, (Athanasopoulos and Jungers, 2018).

Remark 15. Let $\mathcal{B}^d_l = \{z : (G^d_l)_j z \leq (g^d_l)_j, \ j \in J_{G^d_l}\}$. Then, the backward reachability map $\Psi(\sigma, \ \mathcal{B}^d_l))$ is computed by enforcing the constraint $(G^d_l)_j(Az + B_\sigma a + Eh) \leq (g^d_l)_j$, $\forall a \in \mathcal{A}_\sigma(z)$, $\forall h \in \mathcal{H}$, for all $j \in J_{G^d_l}$, or,

$$(G^d_l)_j Az \leq (g^d_l)_j - \max_{a\in\mathcal{A}_\sigma(z)} (G^d_l)_j B_\sigma a - (G^d_l)_j E h^*_j, \tag{18}$$

for all $j \in J_{G^d_l}$, where $h^*_j = \text{argmax}_{h\in\mathcal{H}}(G^d_l)_j Eh$. To compute the set induced by (18), we need to solve $\max_{a\in\mathcal{A}_\sigma(z)}(G^d_l)_j B_\sigma a$ which is a multi-parametric linear program (mpLP) with optimizers being affine functions of z. Solutions can be obtained, e.g., using off-the-shelf multi-parametric programming software. Typically, the set of *parameters* (here, the constraint set $\mathcal{Z}$) is divided into *critical regions*. Throughout a critical region, the optimality conditions derived from the KKT conditions are invariant (Borrelli et al., 2003). For each critical region, the solution of the problem in (18) is an affine function of z, forming the inequality constraints and the set corresponding to the backward reachable set in that critical region. Eventually, the set $\mathcal{B}^i_{l+1}$, with $(i, \ d, \ \sigma) \in \mathcal{E}_\mathcal{G}$, can be formed as the union of such sets. The multi-parametric solution is consistent with (Schaich and Cannon, 2015), with the setting therein expressing the state-dependent sets in vertex representation.

Proposition 16. Consider the backward reachability multi-set sequence (16)-(17) and let Assumption 14 hold. Then, the maximal admissible invariant multi-set $\{\mathcal{S}^i_M\}_{i\in\mathcal{V}_\mathcal{G}}$ is $\mathcal{S}^i_M = \lim_{j\to\infty} \mathcal{B}^i_j$, $i \in \mathcal{V}_\mathcal{G}$.

The proof follows similar steps with (Athanasopoulos et al., 2017, Theorem 3). The difference lies in the involvement of the state-dependent set $\mathcal{A}_\sigma(z)$ in the backward reachability map and consequently in the multi-set sequence (16)-(17), which is well defined as the sets $\mathcal{A}_\sigma(\mathcal{Z})$, $\sigma = 1, \ldots, N$, are compact.

From Definition 12 and Proposition 16, the maximal safe set of (7)-(13) is derived in the following corollary as in (Athanasopoulos et al., 2017).

Corollary 17. Let the maximal invariant multi-set with respect to (7)-(13) be $\{\mathcal{S}^i_M\}_{i\in\mathcal{V}_\mathcal{G}}$. The maximal safe set $\mathcal{S}_\mathcal{P}$ of (7)-(13) with node set $\mathcal{V}_\mathcal{G}$ is $\mathcal{S}_\mathcal{P} = \cap_{i\in\mathcal{V}_\mathcal{G}} \mathcal{S}^i_M$.

Remark 18. Assumption 14 can be lifted. In this case, if the inclusion $\mathcal{S}^i_m \subset \mathcal{Z}$, $i \in \mathcal{V}_\mathcal{G}$ does not hold, the multi-set sequence (16)-(17) converges to the empty set and, thus,

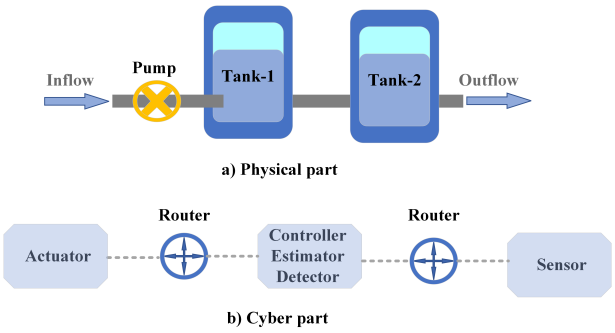

Fig. 4. Physical and cyber parts of a two-tank system.

the maximal safe set is empty indicating an attack with maximum impact.

3.3 Impact metrics

The maximal safe set $\mathcal{S}_\mathcal{P}$ in Corollary 17 provides a security measure of the system under attack. To construct scalar security indices, we propose two complementary impact metrics. First, we define the following.

Definition 19. The outer Lebesgue measure of $\mathcal{S} \subset \mathbb{R}^n$ is $\text{vol}(\mathcal{S}) = \inf \left\{ \sum_{j=1}^{\infty} \text{vol}(\mathcal{R}_j) : \mathcal{S} \subset \cup_{j=1}^{\infty} \mathcal{R}_j \right\}$, where the infimum is taken over all countable collections of rectangles $\mathcal{R}_j = [a_1^j,\ b_1^j] \times [a_2^j,\ b_2^j] \times \ldots \times [a_n^j,\ b_n^j] \in \mathbb{R}^n$, with $a_l^j \leq b_l^j \in \mathbb{R}$, whose union contains $\mathcal{S}$.

Definition 20. Let $\mathcal{S}_1 \subset \mathbb{R}^n$, $\mathcal{S}_2 \subset \mathbb{R}^n$ be two C-sets. The Minkowski distance between $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined as $\mu(\mathcal{S}_1,\ \mathcal{S}_2) = \max\{\lambda : \lambda\mathcal{S}_1 \subseteq \mathcal{S}_2\}$.

Denote the system (7)-(13) by $\mathcal{P}$, and let $\mathcal{Z}$, $\mathcal{H}$, and $\mathcal{A}$ be the constraint, disturbance, and attack sets, respectively. Let $\mathcal{S}_\mathcal{P}$ be the maximal safe set of (7)-(13) and $\mathcal{S}^0$ be the maximal safe set of the attack-free system. Then, $\mathcal{I}_1(\mathcal{P}, \mathcal{Z}, \mathcal{H}, \mathcal{A}) = \frac{\text{vol}(\mathcal{S}^0) - \text{vol}(\mathcal{S}_\mathcal{P})}{\text{vol}(\mathcal{S}^0)}$, and $\mathcal{I}_2(\mathcal{P}, \mathcal{Z}, \mathcal{H}, \mathcal{A}) = 1 - \mu(\mathcal{S}^0, \mathcal{S}_\mathcal{P})$ are two safety metrics of (7)-(13).

Since $\mathcal{S}_\mathcal{P} \subseteq \mathcal{S}^0$, it follows that $\text{vol}(\mathcal{S}_\mathcal{P}) \leq \text{vol}(\mathcal{S}^0)$ and $\mu(\mathcal{S}^0,\ \mathcal{S}_\mathcal{P}) \in [0, 1]$, thus, $0 \leq \mathcal{I}_i \leq 1$, $i = 1, 2$. A metric near zero indicates an attack with little impact whereas a metric almost equal to one translates an impactful attack inducing a small safe set. Metric $\mathcal{I}_1$ provides an index of the size of a safe set the shape of which is not critical to the metric calculation. Metric $\mathcal{I}_2$, however, is sensitive to the shape of the safe set (e.g., its skewness). These are exemplified in the following section.

4. NUMERICAL EXAMPLE

We consider a two-tank system as shown in Fig. 4, with state $x = [x_1\ x_2]^\top$ denoting liquid levels, input u the flow rate of the pump, and control objective maintaining the liquid levels at an operating point. The dynamics are $x(t+1) = Ax(t) + Bu(t) + v(t)$, where $A = \left[\begin{smallmatrix} 0.9 & 0.1 \\ 0.1 & 0.8 \end{smallmatrix}\right]$, $B = [0.1\ 0]^T$, and $\|v(t)\|_\infty \leq 0.01$. The plant has a sensor measuring the liquid level of a tank, an observer estimating the system state, and a detector monitoring attacks; an alarm is raised if the residual exceeds a value, in this case $|r(t)| > 0.01$. The output is $y(t) = Cx(t) + w(t)$, where $C \in \mathbb{R}^{1\times 2}$, $(C)_i = 1$ if the sensor is placed in Tank-i, $i = 1, 2$, or $(C)_i = 0$ otherwise, and $\|w(t)\|_\infty \leq 0.01$. We consider 1) stealthy attacks on the sensor's readings $y(t)$, 2) stealthy attacks on the actuation signal $u(t)$, and 3) stealthy attacks both on the sensor's readings $y(t)$ and the actuation signal $u(t)$. The controller and observer gains K, L, are designed such that the eigenvalues of $A - BK$ and $A - LC$ are $(0.7,\ 0.8)$ and $(0.86,\ 0.001)$, respectively. The operating point is $x^* = [2\ 1]^\top$ with $u^* = 1$. The state constraints are $1 \leq x_1(t) \leq 3$, $0 \leq x_2(t) \leq 2$, and the input constraint is $0 \leq u(t) \leq 2$. The attack signals $a_y(t)$, $a_u(t)$, are consistent with the stealthiness Definition 9, and are additionally bounded with lower and upper limits shown in Table 1. Attack patterns are listed in Table 1, where $N_{\max}$ is the maximum dwell time, i.e., the maximum length of consecutive attacks, and $N_{\min}$ is the minimum number of consecutive time steps that the system is attack-free.

Table 1. Attack actions

Vulnerable point	Attack bounds	Pattern
Sensor	$-0.05 \leq a_y \leq 0.05$	$N_{\max} = N_{\min} + 1$
Actuator	$-0.01 \leq a_y \leq 0.01$	$N_{\max} = N_{\min} - 1$

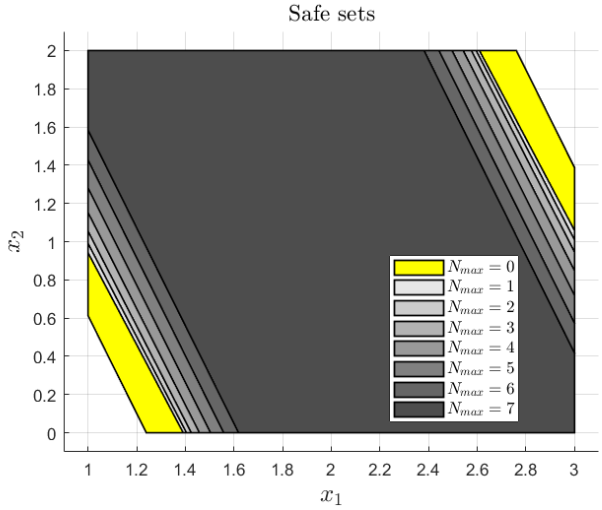

Fig. 5. Projections of safe sets onto $\mathbb{R}^2$ for $e = x - \hat{x} = 0$.

In Fig. 5, we compute the safe set of the system when the sensor placed in Tank-2 is under attack. The safe set of attack-free dynamics is illustrated in yellow, whereas safe state regions of the system under attack for the associated dwell-time specifications are in gray. Clearly, the safe region shrinks as $N_{\max}$ grows indicating safety degradation. In Fig. 6, we compute the safety metrics $\mathcal{I}_1, \mathcal{I}_2$, introduced in Section 3.3, for all attack scenarios considered. We show that attacks poisoning the actuation signal have a major effect on system safety in this particular example. From Fig. 6, we also conclude that a sensor placed at Tank-2 results in a less vulnerable plant preventing a safe set from collapsing to the empty set as $N_{\max}$ grows.

5. CONCLUSION

We have proposed a new approach to modeling attack scenarios in cyber-physical systems. We define a cyber-physical system under attack as a constrained switching system embedding the dynamics of the plant, the attack

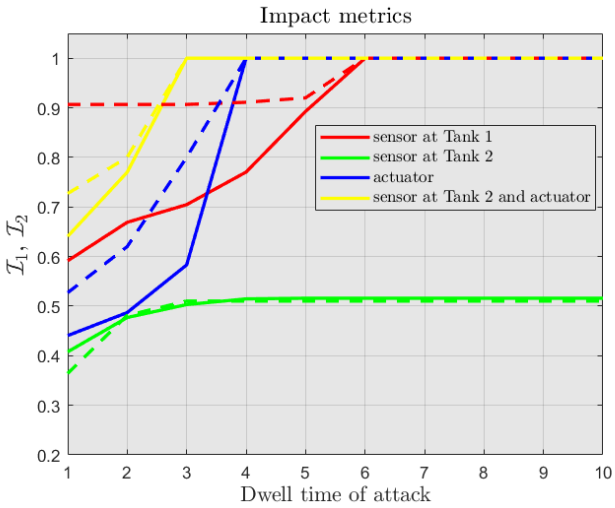

Fig. 6. Impact metrics: $\mathcal{I}_1$ (solid lines), $\mathcal{I}_2$ (dashed lines).

patterns, and the attack detection scheme. By calculating the maximal safe set of the underlying constrained switching system, we characterize system safety as an asymptotic property. Two complementary scalar security metrics are also introduced.

REFERENCES

Athanasopoulos, N., Smpoukis, K., and Jungers, R.M. (2017). Invariant sets analysis for constrained switching systems. *IEEE Control Systems Letters*, 1(2), 256–261.

Athanasopoulos, N. and Jungers, R.M. (2018). Combinatorial methods for invariance and safety of hybrid systems. *Automatica*, 98, 130–140.

Athanasopoulos, N. and Lazar, M. (2014). Stability analysis of switched linear systems defined by graphs. In *IEEE Conference on Decision and Control*, 5451–5456.

Beg, O.A., Johnson, T.T., and Davoudi, A. (2017). Detection of False-Data Injection Attacks in Cyber-Physical DC Microgrids. *IEEE Transactions on Industrial Informatics*, 13(5), 2693–2703.

Bertsekas, D.P. (1972). Infinite-Time Reachability of State-Space Regions by Using Feedback Control. *IEEE Transactions on Automatic Control*, 17(5), 604–613.

Blanchini, F. and Miani, S. (2015). *Set-Theoretic Methods in Control*. Birkhäuser Basel.

Borrelli, F., Bemporad, A., and Morari, M. (2003). Geometric algorithm for multiparametric linear programming. *Journal of Optimization Theory and Applications*, 118(3), 515–540.

Cassandras, C.G. and Lafortune, S. (2010). *Introduction to Discrete Event Systems*. Springer, second edition.

Chen, S., Kalbarczyk, Z., Xu, J., and Iyer, R. (2003). A data-driven finite state machine model for analyzing security vulnerabilities. In *2003 International Conference on Dependable Systems and Networks*, 605–614.

De Persis, C. and Tesi, P. (2015). Input-to-state stabilizing control under denial-of-service. *IEEE Transactions on Automatic Control*, 60(11), 2930–2944.

De Santis, E., Di Benedetto, M.D., and Berardi, L. (2004). Computation of Maximal Safe Sets for Switching Systems. *IEEE Transactions on Automatic Control*, 49(2), 184–195.

Donkers, M.C., Heemels, W.P., Van De Wouw, N., and Hetel, L. (2011). Stability analysis of networked control systems using a switched linear systems approach. *IEEE Transactions on Automatic Control*, 56(9), 2101–2115.

Gahinet, P. and Apkarian, P. (1994). A linear matrix inequality approach to H_∞ control. *International Journal of Robust and Nonlinear Control*, 4(4), 421–448.

Hespanha, J.P., Naghshtabrizi, P., and Xu, Y. (2007). A survey of recent results in networked control systems. *Proceedings of the IEEE*, 95(1), 138–172.

Liu, X., Zhang, J., and Zhu, P. (2017). Modeling cyber-physical attacks based on probabilistic colored Petri nets and mixed-strategy game theory. *International Journal of Critical Infrastructure Protection*, 16, 13–25.

Meira-Góes, R., Kang, E., Kwong, R.H., and Lafortune, S. (2020). Synthesis of sensor deception attacks at the supervisory layer of Cyber–Physical Systems. *Automatica*, 121.

Miao, F. and Zhu, Q. (2014). A moving-horizon hybrid stochastic game for secure control of cyber-physical systems. In *IEEE Conference on Decision and Control*, 517–522.

Milošević, J., Umsonst, D., Sandberg, H., and Johansson, K.H. (2018). Quantifying the Impact of Cyber-Attack Strategies for Control Systems Equipped with an Anomaly Detector. In *2018 European Control Conference*, 331–337.

Mo, Y. and Sinopoli, B. (2016). On the Performance Degradation of Cyber-Physical Systems under Stealthy Integrity Attacks. *IEEE Transactions on Automatic Control*, 61(9), 2618–2624.

Murguia, C., Shames, I., Ruths, J., and Nešić, D. (2020). Security metrics and synthesis of secure control systems. *Automatica*, 115, 108757.

Raković, S.V., Kerrigan, E.C., Mayne, D.Q., and Lygeros, J. (2006). Reachability analysis of discrete-time systems with disturbances. *IEEE Transactions on Automatic Control*, 51(4), 546–561.

Reda, H.T., Anwar, A., and Mahmood, A. (2022). Comprehensive survey and taxonomies of false data injection attacks in smart grids: attack models, targets, and impacts. *Renewable and Sustainable Energy Reviews*.

Saltik, M.B., Athanasopoulos, N., Ozkan, L., and Weiland, S. (2015). Safety analysis for a class of graph constrained scheduling problems. In *IEEE Conference on Decision and Control*, 1687–1692.

Sandberg, H., Gupta, V., and Johansson, K.H. (2022). Secure Networked Control Systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 5, 445–464.

Schaich, R.M. and Cannon, M. (2015). Robust positively invariant sets for state dependent and scaled disturbances. In *IEEE Conference on Decision and Control*, 7560–7565.