Design of stealthy attacks for cyber-physical systems

Luca A. L. Barbieri, Giancarlo Fortino and Giuseppe Franzè

***Abstract*— Stealthy attacks are formally designed to compromise the regulated state trajectories of networked control systems enjoying resilient capabilities. By taking advantage of the networked control system knowledge and the probability density function of the communication channel, a denial-of-service attack is conceived with the aim to deteriorate the control performance. This is achieved by properly exploiting Kullback-Leibler Divergence arguments in order to disguise as better as possible the attack occurrence. Specifically, the attack will be defined so that the anomaly detector is not capable of distinguishing between under attack and attack-free operating scenarios. A final example is provided to show the peculiarities of the proposed approach.**

I. Introduction

In recent years, a growing interest has been devoted to analyze cyber-physical systems and its applications. An important aspect in these contexts relies on the detection and possibly defense against cyber attacks, as testified by several contributions, see [1], [2]. Moreover, incidents, such as those in [3] and [4], have explicitly put in light the capability of these anomalies to jeopardize the control performance by injecting malicious data into the communication channels. As a consequence, the need of detection units to reveal attack occurrences has assumed a key role [5], [6] above all to define efficient countermeasures. Recently, resilient control, as a new concept in control systems, has been proposed as an important tool to address such issues, aiming at keeping an acceptable level of operational normalcy in the presence of these attacks while maintaining system robustness against disturbances. Recent results on resilient control of CPSs can be found in [7], [8], [9]. However, some attackers with enough resources can carefully design attacks so that they can avoid being detected. Hence, to investigate the vulnerability of CPSs, researchers study the design of stealthy attacks and evaluate how much destruction these attacks make. Two typical notions of stealthiness are usually introduced. The first, referred to as strict stealthiness, requires that the residual sequences under attack and those under nominal conditions share the same probability distribution. The other is ϵ-stealthiness, which is usually measured by the Kullback–Leibler Divergence between those two probability density functions [10]. From the perspective of attackers, the optimal attack on remote state estimation usually refers to the attack that yields the greatest estimation error under certain stealthiness constraints. This optimal attack can also be called a worst-case attack from the perspective of defenders. Hence, an interesting problem is how to obtain the worst-case attack for a certain system with specified stealthiness constraints. Along these lines, some relevant contributions have been recently presented. In [11], the authors investigate the concept of stealthiness for signal attacks in stochastic CPSs and provide a characterization of the performance degradation induced by ϵ-stealthy attacks. The authors of [12] presented the design of data-injection attacks based on the KLD for a class of cyber-physical systems subject to malicious actions on the controller-to-actuator channel. Conversely, in [13], the design of stealthy sensor attacks for CPSs modeled as linear quadratic Gaussian dynamics has been developed according to optimality criteria. Conversely, in [14] an on-line attack strategy is proposed where the malicious agent estimates the system state with the intercepted data at each instant and computes the optimal attack accordingly. Finally, along similar lines is the approach proposed in [15] where False Data Injection attack strategies are developed by exploiting the KLD properties.

The above literature analysis reveals an aspect not fully explored: designing stealthy denial-of-service (DoS) attacks by combining within a unique framework the KLD operator and the system knowledge. Essentially, the idea is to take advantage of the resilient control architecture to infer an optimal attack strategy capable of remaining indefinitely ϵ-stealthy. In particular, the proposed solution considers the constrained networked control architecture outlined in [16] for dealing with packet losses on both communication channels. Based on the available information and using the Kullback-Leibler Divergence, an optimization problem is defined in charge of generating stealthy DoS sequences and deteriorating as much as possible the regulated performance.

L. Barbieri, G. Fortino and G. Franzè are with DIMES, Università della Calabria, Via Pietro Bucci, Cubo 42-C, Rende (CS), 87036, Italy (e-mail: {g.fortino,giuseppe.franze}@unical.it, luca.barbieri@dimes.unical.it).

Notation

Let ξ be a continuous random variable. Then, the probability density function (pdf) is $f_\xi : \mathbb{R} \to \mathbb{R}_{\geq 0}$, while

$$F_\xi(\alpha) \triangleq \int_{-\infty}^{\alpha} f_\xi(z)\, dz \tag{1}$$

is the Cumulative Distribution Function (CDF) of ξ. Let $f_\xi^1(z)$ and $f_\xi^2(z)$ be two pdfs defined over the same random variable ξ, the Kullback-Leibler Divergence (KLD) from $f_\xi^1(z)$ to $f_\xi^2(z)$ is defined as follows:

$$D_{KL}(f^1 \| f^2) \triangleq \int_{-\infty}^{+\infty} f_\xi^1(z) \log\left(\frac{f_\xi^1(z)}{f_\xi^2(z)}\right) dz \tag{2}$$

The KLD is a non-negative quantity that measures the dissimilarity between two pdfs, with $D_{KL}(f^1\|f^2) = 0$ if and only if $f^1_\xi(z) = f^2_\xi(z)$ almost everywhere. In the sequel, for the sake of notational simplicity, f_ξ and F_ξ refer to the pdf and the CDF, respectively.
A sequence of L elements is denoted by $\tau^{[1:L]}$ and defined as:

$$\tau^{[1:L]} := \{\tau(1), \dots, \tau(L)\} \tag{3}$$

II. Problem statement

In the sequel, the class of networked control systems, depicted in Fig. 1, is of interest. There, the following operating hypotheses are made.

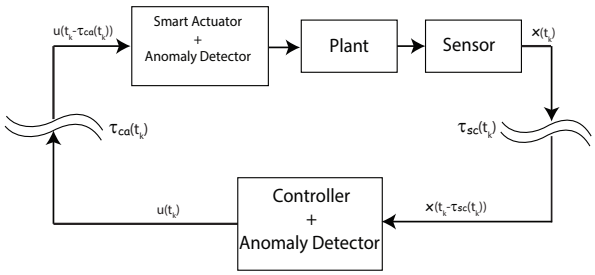

Fig. 1. Networked control system architecture

Plant - It is modeled in a generic form by the system

$$\Sigma: \ \dot{x}(t) = f(x(t), u(t), d(t)) \tag{4}$$

where $t \in \mathbb{R}_+$, $x(t) \in \mathbb{R}^{n_x}$ denotes the state and $u(t) \in \mathbb{R}^{n_u}$ the command input. Moreover, $d(t) \in \mathcal{D} \subset \mathbb{R}^{n_d}$ accounts for the process noise with $\mathcal{D}$ compact and convex and the origin being an interior point, i.e., $0_{n_d} \in \mathcal{D}$. Moreover, the system (4) is subject to the following set-membership constraints:

$$u(t) \in \mathcal{U}, \ x(t) \in \mathcal{X}, \ \forall t \geq 0 \tag{5}$$

with $\mathcal{U}$ and $\mathcal{X}$ compact subsets of $\mathbb{R}^{n_u}$ and $\mathbb{R}^{n_x}$, respectively, and $0_{n_u} \in \mathcal{U}$, $0_{n_x} \in \mathcal{X}$. Since the plant is remotely controlled via a communication network, it is assumed that it can export, via a sensor device, samples $x(t_k)$, with $k \in \mathbb{Z}_+$ and $t_k := T_s k$ for a given sampling time T_s. On the actuator side, it is assumed that a sample $u(t_k)$, conveyed through the network, is converted into an analog input.
Communication medium - The network is subject to induced time-delays occurring on both the communication channels. Specifically, $\tau_{sc}(t_k) : \mathbb{R}_+ \to \mathbb{R}_{\geq 0}$ and $\tau_{ca}(t_k) : \mathbb{R}_+ \to \mathbb{R}_{\geq 0}$ denote the *sensor-to-controller* (SC) and *controller-to-actuator* (CA) time-delay occurrences, respectively. Let $\bar{\tau}$ be the *Maximum Allowable Transmission Interval* (MATI), then a *Packet loss* occurs when there exists $t_{\bar{k}}$ such that $\tau_{sc}(t_{\bar{k}}) \geq \bar{\tau} - \bar{\tau}_c$ or $\tau_{ca}(t_{\bar{k}}) \geq T_s$. The statistical distribution of the induced time-delays $\tau_{sc}(t_k)$ and $\tau_{ca}(t_k)$ along the communication medium is characterized in terms of its pdf $f^N_\xi : \mathbb{R} \to \mathbb{R}_{\geq 0}$.
Controller - The controller is designed under the following requirements:

- $\forall k \in \mathbb{Z}_+$ the computed command $u(t_k)$ is admissible and can be consecutively applied for a finite number of steps, say τ_{max};
- the regulated state trajectory is Uniformly Ultimate Bounded (UUB) despite any noise realization $d(t_k) \in \mathcal{D}$.

Smart actuator - This unit is designed so that it is capable of recognizing that a feedback loss is underway on the *controller-to-actuator* channel and selecting the suitable sample $u(\cdot)$ for the plant among those stored in an internal memory.

For more details, the interested reader can refer to [16] where the properties of the resilient scheme of Fig. 1 are outlined. By collecting these premises, at each time instant t_k the NCS architecture is resilient to packet loss occurrences involving delays $\tau_{ca}(t_k)$ and/or $\tau_{sc}(t_k)$ if the following inequality holds:

$$\mu_{ca}(t_k) < \tau_{max} \ \wedge \ \mu_{sc}(t_k) < \tau_{max} \tag{6}$$

where $\mu_{ca}(t_k)$ and $\mu_{sc}(t_k)$ are integer counters aimed at evaluating the number of consecutive packet losses, i.e.,

$$\mu_{ca}(t_k) = \begin{cases} \mu_{ca}(t_{k-1}) + 1, & \text{if } \tau_{ca}(t_k) \geq T_s \\ 0, & otherwise \end{cases} \tag{7}$$

$$\mu_{sc}(t_k) = \begin{cases} \mu_{sc}(t_{k-1}) + 1, & \text{if } \tau_{sc}(t_k) \geq \bar{\tau} - \bar{\tau}_c \\ 0, & otherwise \end{cases} \tag{8}$$

with $\tau_{ca}(t_k) \leq \bar{\tau}_c, \forall t_k$. Then, let $\Delta T := [t_{k_{start}} \ t_{k_{end}}]$ be a time interval such that a sequence of packet losses occurs, and the following operating time-delay scenarios arise:

- **Normal mode -** each time delay occurrence is bounded: $\tau_{sc}(t_k) < \bar{\tau} - \bar{\tau}_c$ on the *sensor-to-controller* and $\tau_{ca}(t_k) < T_s$ on the *controller-to-actuator* links; the packet loss occurrences are such that $t_{k_{end}} - t_{k_{start}} < \tau_{max}$;
- **Anomalous mode -** there exists a time interval ΔT such that $t_{k_{end}} - t_{k_{start}} \geq \tau_{max}$.

It is evident that if the **Anomalous mode** takes place, the controller C is no longer admissible and some countermeasures must be adopted. Hence, a cyber-attack, aimed at producing an **Anomalous mode**, would result a *naive* attempt of inducing instability on Σ as it can be detected by simply checking condition (6). In view of this consideration and since the communication medium is unreliable, more sophisticated malicious actions can be carried out by external agents to deteriorate only the control performance. Specifically, it is assumed that a denial-of-service attack is performed according to the following guidelines:

- *System knowledge:* exploit the dynamical model (4)-(5), the controller structure and its properties, namely τ_{max};
- *Disclosure resources:* gather $\tau_{ca}(t_k)$ and/or $\tau_{sc}(t_k)$.

In order to identify this attack, the NCS is equipped with two **Anomaly detector** units straightforwardly conceived in terms of inequalities (6) and exploiting the counters (7)-(8), i.e., at each time instant t_k the shifting sequence of L consecutive time-delays $\tau^{[t_{k-L+1}:t_k]}$ ($\tau_{sc}^{[t_{k-L+1}:t_k]}$ or $\tau_{ca}^{[t_{k-L+1}:t_k]}$ respectively) is evaluated according to the following logic:

$$\mathbf{D}(\{\tau^{[t_{k-L+1}:t_k]}\}) = \begin{cases} \text{attack}, & \mu(t_k) \geq \tau_{max} \\ \text{no attack}, & otherwise \end{cases} \quad (9)$$

where, for the sake of simplicity from now onward, the notation $\tau(t_k)$ is used in place of both the realizations $\tau_{sc}(t_k)$ and $\tau_{ca}(t_k)$. Moreover, a positive scalar th will account for T_s and $\bar{\tau} - \bar{\tau}_c$ within (7) and (8), respectively.

Then, the problem to solve can be stated as follows.

Stealthy DoS Attack (S-DoS-A) - *Given the plant model (4)-(5), the control architecture of Fig. 1, resilient to τ_{max} consecutive DoS attacks, design sequences of consecutive packet losses $\tau^{[t_{k-L+1}:t_k]}$, compliant with the probability density function f_ξ^N, such that the detection logic (9) is not capable of revealing any attack occurrence (stealthiness).* □

III. Stealthy DoS attack design

In this section, a stealthy DoS attack is conceived under the following reasoning: find an artificial pdf $f_\xi^A \neq f_\xi^N$ such that any realization $\tau(t_k) \sim f_\xi^A$ is complying with f_ξ^N and any sequence $\tau^{[t_{k-L+1}:t_k]}$ remains stealthy to the logic (9).

According to the ideas presented in [11], the stealthiness of an attack sequence $\tau^{[t_{k-L+1}:t_k]}$ can be characterized in terms of the KLD as follows:

- the attack is said to be *strictly stealthy* if and only if $D_{KL}(f_\xi^A \| f_\xi^N) = 0$;
- the attack is said to be *ϵ-stealthy* if

$$D_{KL}(f_\xi^A \| f_\xi^N) \leq \epsilon \quad (10)$$

for a given scalar $\epsilon > 0$.

Notice that small divergences indicate that the attacker effects are statistically marginal and therefore difficult to be detected. Conversely, a large divergence highlights significant discrepancies between f_ξ^N and f_ξ^A so making the attack readily identifiable. Under these premises, the statistical hypotheses describing delay realizations can be defined as:

$$\mathcal{H}_0 : \tau(t) \sim f_\xi^N \quad \text{no attack} \quad (11)$$

$$\mathcal{H}_1 : \tau(t) \sim f_\xi^A \quad \text{under attack} \quad (12)$$

Therefore, the design of an ϵ-stealthy denial-of-service attack amounts to determining a distribution f_ξ^A satisfying the inequality (10), while simultaneously degrading the control performance. Specifically, the idea consists of a so-called redistribution of the probability of *positive* ($\tau(t_k) < th$) and *negative* ($\tau(t_k) \geq th$) events: in other words, the attacker moves a fraction of *positive* realizations to *negative* ones under the satisfaction of (10).

To formally deal with this reasoning, let

$$a := p \int_{-\infty}^{th} f_\xi^N(z) dz \quad (13)$$

be the shifted probability mass with $p \in [0, 1)$. Hence,

$$p := \frac{a}{F_\xi^N(th)}, \quad 0 \leq a < F_\xi^N(th) \quad (14)$$

accounting for the promotion probability applied to *positive* time-delay realizations. Then, the attacker pdf f_ξ^A can be selected as follows:

$$f_\xi^A := \begin{cases} f_{\xi|\xi<th}^A = \left(1 - \frac{a}{F_\xi^N(th)}\right) f_\xi^N \\ f_{\xi|\xi\geq th}^A = \left(1 + \frac{a}{q_N}\right) f_\xi^N \end{cases} \quad (15)$$

where

$$q_N := 1 - F_\xi^N(th) \quad (16)$$

is the nominal probability pertaining to the occurrence of *negative* time-delay realizations.

Since f_ξ^A is, by construction, normalized and differs from f_ξ^N only by the probability mass a which is strictly connected to the promotion probability p, the effective probability of observing a time-delay realization over the threshold th is given by:

$$q_{eff} := q_N + a \quad (17)$$

Remark 1 - It is worth underlining that a trade-off arises by looking at (15) and (17): increasing the number of *negative* realizations and keeping the statistical properties of f_ξ^A close to f_ξ^N both depend on the same suitable choice of the probability mass a. □

The next result provides an explicit evaluation of the ϵ-stealthiness in terms of the KLD operator.

Proposition 1: Given the pdf f_ξ^A (15), the KLD from f_ξ^A to f_ξ^N has the following expression:

$$D_{KL}(f_\xi^A \| f_\xi^N) = (1 - q_{eff}) \log\left(\tfrac{1-q_{eff}}{1-q_N}\right) + q_{eff} \log\left(\tfrac{q_{eff}}{q_N}\right) \quad (18)$$

Proof - According to definition (2) and formula (15), one has that

$$\begin{aligned} &D_{KL}(f^A || f^N) = \int_0^{+\infty} f_\xi^A(z) \log\left(\frac{f_\xi^A(z)}{f_\xi^N(z)}\right) dz \\ &= \int_0^{th} \left(1 - \tfrac{a}{F_\xi^N(th)}\right) f_\xi^N(z) \log\left(1 - \tfrac{a}{F_\xi^N(th)}\right) dz + \\ &\quad \int_{th}^{+\infty} \left(1 + \tfrac{a}{q_N}\right) f_\xi^N(z) \log\left(1 + \tfrac{a}{q_N}\right) dz \\ &= (F_\xi^N(th) - a) \log\left(1 - \tfrac{a}{F_\xi^N(th)}\right) + (q_N + a) \log\left(1 + \tfrac{a}{q_N}\right) \end{aligned}$$

from which the result straightforwardly follows by using the expression (17). □

Remark 2 - Notice that *Proposition 1* directly shows that the *strict stealthy* condition is in principle achievable. In fact, from (18) one has that, as $a \to 0$, $f_\xi^A \to f_\xi^N$ and, as a consequence, $D_{KL}(f^A || f^N) \to 0$. □

A. Optimization of the attack effect

Here, the probability mass a is optimized, in compliance with the structural constraint (14) and the stealthiness condition (10), to deteriorate as much as possible the overall performance of the control architecture of Fig. 1.

To this end, consider the following definition.

Definition 1: A *burst* of length $L_{burst} := L + 1$ is a sequence of $L \geq 0$ consecutive *negative* realizations followed by a *positive* occurrence. □

Let L be a discrete random variable, then the following equalities straightforwardly hold:

$$\frac{\mathbb{E}[L]}{\mathbb{E}[L_{burst}]} = q_{eff} \longleftrightarrow \mathbb{E}[L] = \frac{q_{eff}}{1 - q_{eff}} \quad (19)$$

with $\mathbb{E}[\cdot]$ the expected value operator. By imposing lower and upper bounds

$$L_{\min} \leq \mathbb{E}[L] \leq L_{\max},\ L_{\min} \geq 0,\ L_{\max} < \tau_{max}$$

formal constraints on a are promptly derived:

$$a \leq \frac{L_{\max}}{1+L_{\max}} - q_N \tag{20}$$

$$a \geq \frac{L_{\min}}{1+L_{\min}} - q_N \tag{21}$$

Then, the following claim holds true.

Statement 1: Given a tolerance level $\epsilon \geq 0$, the threshold th and the bounds L_{min}, L_{max}, the maximum level of probability mass a is obtained as the solution of the following optimization problem:

$$\begin{cases} a^* := \arg\max_{a} \ a \\ \quad \text{subject to} \\ (20)-(21) \\ 0 \leq a < F_{\xi}^{N}(th) \\ (1-q_{eff}) \log\left(\frac{1-q_{eff}}{1-q_N}\right) + q_{eff} \log\left(\frac{q_{eff}}{q_N}\right) \leq \epsilon \end{cases} \tag{22}$$

It is worth underlining that the optimal solution a^* represents the maximum statistically admissible f_{ξ}^{N}-deformation that preserves the prescribed ϵ-stealthiness level.

IV. Simulation

In this section, the effectiveness of the proposed approach is evaluated on a test-bed example. All the simulations have been carried out on a laptop equipped with an Intel Core i7-11800H (11th Gen, 2.30 GHz) processor.

A path-tracking problem for an autonomous vehicle described by a discrete-time linear time-invariant state space model has been considered. The state vector $x(t) = \begin{bmatrix} p_x(t) & v_x(t) & p_y(t) & v_y(t) \end{bmatrix}^T$ accounts for the position and velocity components in the xy plane, while the control input $u(t) = \begin{bmatrix} T_x(t) & T_y(t) \end{bmatrix}^T$ includes the force components along the two axes. Moreover, system matrices are:

$$A = \begin{bmatrix} 1 & T_s & 0 & 0 \\ 0 & 1-\zeta & 0 & 0 \\ 0 & 0 & 1 & T_s \\ 0 & 0 & 0 & 1-\zeta \end{bmatrix}, B = \begin{bmatrix} \frac{T_s^2}{2m} & 0 \\ \frac{T_s}{m} & 0 \\ 0 & \frac{T_s^2}{2m} \\ 0 & \frac{T_s}{m} \end{bmatrix}, B_d = B, \tag{23}$$

where the sampling time is set to $T_s = 0.03\,\text{s}$, the viscous damping factor to $\zeta = 0.01$, the vehicle mass to $m = 1\,\text{kg}$ and $d(t) \in \mathcal{D} := \{d \in \mathbb{R}^2 : ||d||_2 \leq 0.01\}$. Euclidean norm constraints are imposed on input and state vectors: $||u(t)||_2 \leq 3,\ ||x(t)||_2 \leq 10,\ \forall t \geq 0$. Data are exchanged by means of a non-acknowledged communication protocol (UDP) with the network MATI equal to $\bar{\tau} = 0.6\,\text{s}$. Moreover, in Fig. 2, the Independent-of-Delay $\{\mathcal{T}_i^{IOD}\}$ and Dependent-of-Delay $\{\mathcal{T}_i^{DD}\}$ families of controllable sets, computed according to the arguments developed in [16], are reported. As an outcome, the following design knobs are obtained: the round-trip delay upper bound $\bar{\tau}_c = 0.3\,\text{s}$ and the maximum admissible delay $\tau_{max} = 0.3\,\text{s}$.

As the network medium is concerned, the following lognormal probability density function is given by:

$$f_{\xi}^{N}(z) = \frac{1}{z\sigma\sqrt{2\pi}} \exp\left\{\frac{-(\log(z)-\mu)^2}{2\sigma^2}\right\}, z > 0 \tag{24}$$

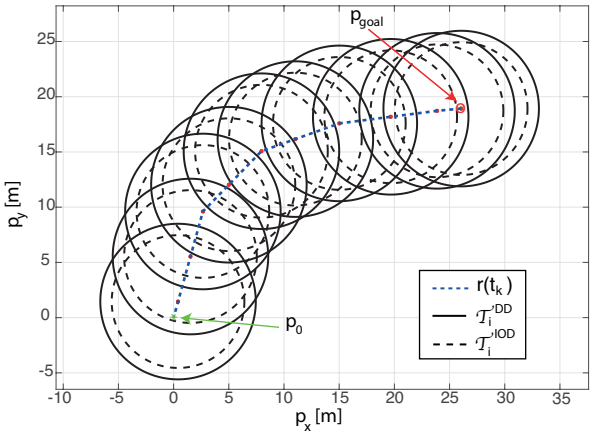

Fig. 2. Sequence of one-step ahead controllable sets

with $\mu = -4.8317$ and $\sigma = 0.9697$.

Operating scenario - Given the reference planar trajectory $r(t_k)$ (blue dashed line), the control architecture of Fig. 1 drives the autonomous vehicle from the initial position $p_0 = [0\ \ 0]^T$ to the target $p_{goal} = [26.05\ \ 18.93]^T$ under the occurrence of the induced time delays compatible with (24) and depicted in Fig. 3. □

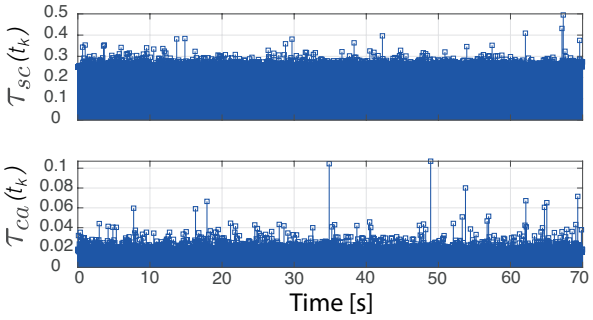

Fig. 3. Time-delay occurrences on the communication channels

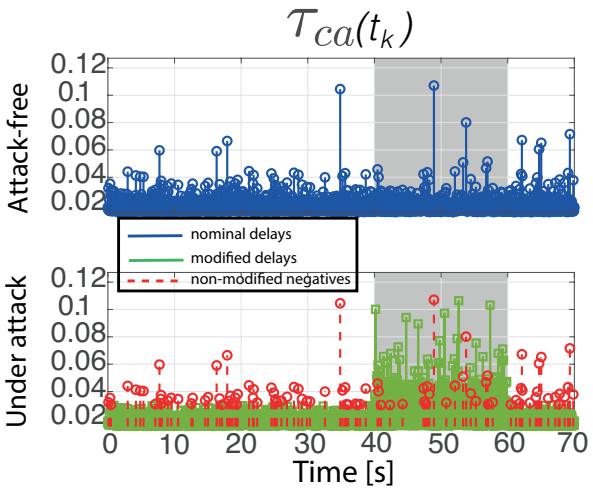

Fig. 4. Time-delay occurrences on the CA communication channel

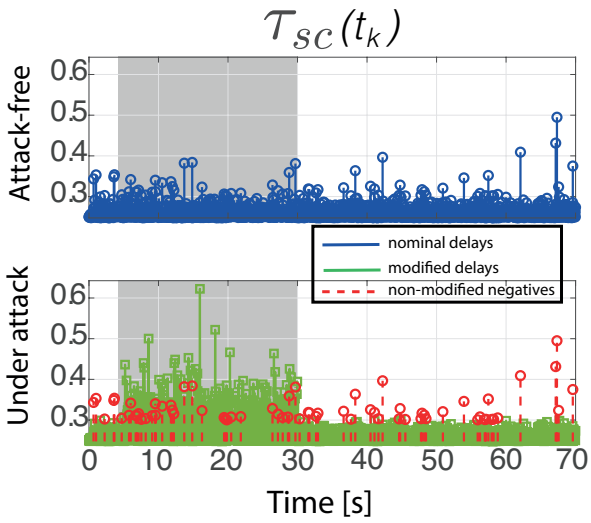

Fig. 5. Time-delay occurrences on the SC communication channel

In the sequel, the objective of this simulation can be summarized as follows.

Attack aim - Generate sequences of packet losses that are not detectable by the logic (9) such that the closed-loop control performance of the resilient architecture of Fig. 1 is deteriorated. □

To accomplish this task, a combined attack on the two channels is designed by exploiting the results presented in *Proposition 1* and *Statement 1*. In particular, the following artificial pdf has been selected:

$$f_{\xi}^{A} := \begin{cases} f_{\xi|\xi<0.3}^{A} = \left(1 - \dfrac{0.4813}{F_{\xi}^{N}(0.3)}\right) f_{\xi}^{N} \\ f_{\xi|\xi\geq 0.3}^{A} = \left(1 + \dfrac{0.4813}{q_N}\right) f_{\xi}^{N} \end{cases} \tag{25}$$

where $th = \bar{\tau} - \bar{\tau}_c = 0.3\,s$ and $a^* = 0.4813$ solution of the optimization problem (22). Moreover, the effective probability is $q_{eff} = 0.5219$. According to these premises, two attack windows have been defined: $S_1 = [5,\ 30]\ s$ on the *sensor-to-controller* and $S_2 = [40,\ 60]\ s$ on the *controller-to-actuator* channels, respectively (the shadow areas in Figs. 4-5).

The pdf (25) allows one to artificially alter the nominal latency behaviors as shown in Figs. 4-5. There, it can be noticed that, within the time interval S_1 a sequence of *bursts*, complying with the upper bound τ_{max} (consecutive number of negatives) and the threshold $th = 0.3\,s$ is properly introduced on the SC channel with respect to the nominal time-delay realizations; conversely, the artificial *bursts* modifies the *controller-to-actuator* link according to the sampling time $T_s = 0.03\,s$ on the second interval S_2.

These anomalies, though significant, remain stealthy to the detector units thanks to the satisfaction of the KLD condition (10). It is worth underlining that some nominal time-delay occurrences are not modified (red dashed spikes) because they normally overcome the data loss threshold. Moreover, to demonstrate the effectiveness of the proposed design attack with respect to the resilient control architecture, the evolution of the number of consecutive *negative* events over time is depicted in Fig. 6. There, it is shown that the length of each *burst* is always under the maximum admissible length, i.e., $\frac{\tau_{max}}{Ts} = 10$, so that providing a quantitative measure of the attack potential.

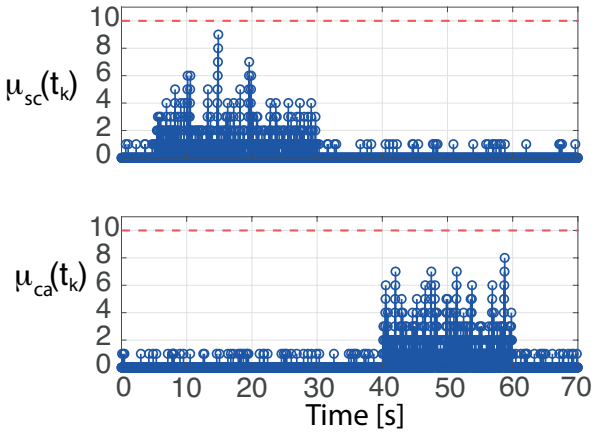

Fig. 6. Consecutive packet dropouts counter

The consequence of this *modus operandi* can be appreciated by looking at the regulated state trajectory shown in Fig. 7. First, it is important to underline that the resilient control architecture exhibits good performance when DoS attack-free scenarios are considered, see the latency effects depicted in the upper graphs of Figs. 4-5. Starting from $t_k = 5\,s$ when the DoS attack takes place on the SC channel a discrepancy arises: initially, this is more evident on the lateral velocity (v_y), then this undesired phenomenon propagates along the planar components with a significant loss of control performance. Such an effect compromises the effectiveness of the control action also because the second attack on the CA channel takes place deteriorating the overall performance. Moreover, the command input behaviors, pertaining to the scenarios, are shown in Fig. 8. As expected, a gradual growth of control effort can be observed in order to take care of such undesired and unexpected anomalies.

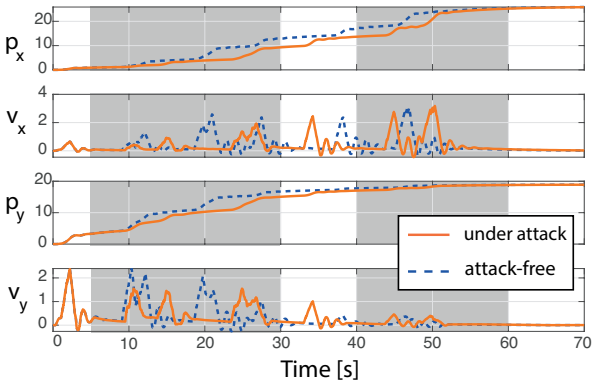

Fig. 7. State evolutions under attack and attack-free scenarios

Finally, it is important to analyze network congestion levels in terms of in-flight packets, to assess the capability of

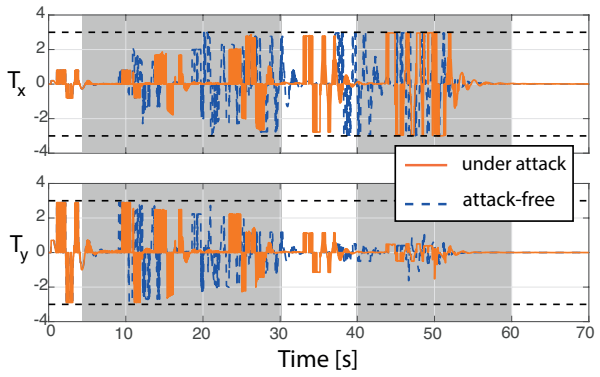

Fig. 8. Command input signals under attack and attack-free scenarios

the attacker to deteriorate the overall performance. As it is clearly shown in Figs. 9-10, the effect of DoS attacks falls back into a significant increase of not dispatched data.

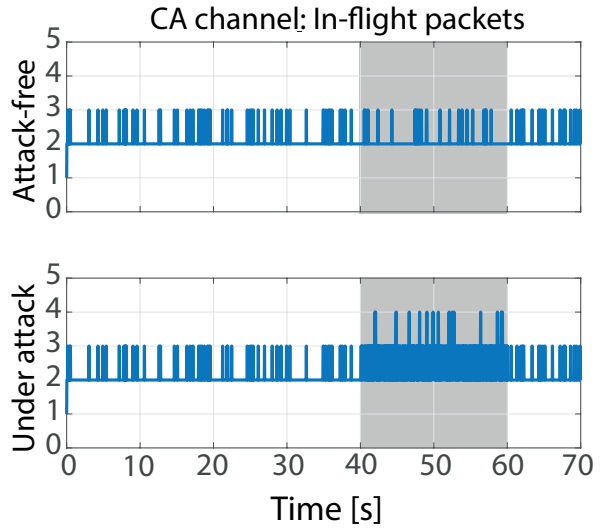

Fig. 9. CA channel congestion level under attack and attack-free scenarios

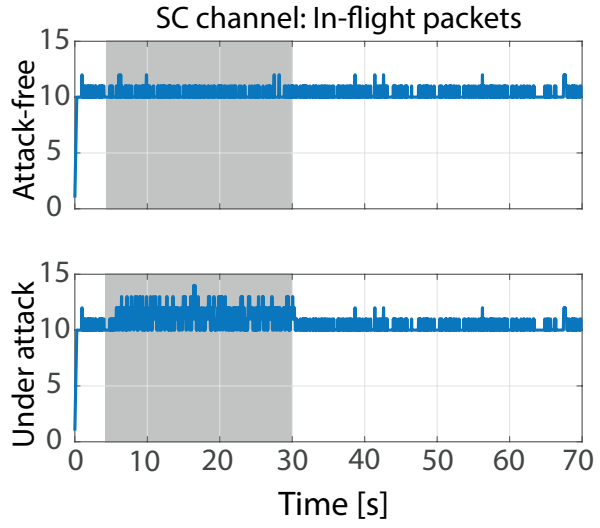

Fig. 10. SC channel congestion level under attack and attack-free scenarios

V. Conclusions

In this paper, the design of stealthy DoS attacks has been analyzed for cyber-physical systems enjoying resilient capabilities. By jointly taking advantage of systems knowledge, disclosure resources, the Kullback-Leibler Divergence and the communication medium statistics, a novel approach has been developed to artificially and properly define sequences of undetectable DoS attacks. The simulation results are promising, as they show the loss of performance due to the introduction of anomalies without being detected.

References

[1] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2013.

[2] C. Savaglio, V. Barbuto, F. Mangione, and G. Fortino, "Generative digital twins: A novel approach in the iot edge-cloud continuum," *IEEE Internet of Things Magazine*, vol. 8, no. 1, pp. 42–48, 2024.

[3] S. Karnouskos, "Stuxnet worm impact on industrial cyber-physical system security," in *IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2011, pp. 4490–4494.

[4] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *International conference on critical infrastructure protection*. Springer, 2007, pp. 73–82.

[5] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–36, 2018.

[6] P. Griffioen, S. Weerakkody, B. Sinopoli, O. Ozel, and Y. Mo, "A tutorial on detecting security attacks on cyber-physical systems," in *2019 18th European control conference (ECC)*. IEEE, 2019, pp. 979–984.

[7] G. Franzè, W. Lucia, and F. Tedesco, "Resilient model predictive control for constrained cyber-physical systems subject to severe attacks on the communication channels," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 1822–1836, 2021.

[8] G. Franze, D. Famularo, W. Lucia, and F. Tedesco, "Cyber–physical systems subject to false data injections: A model predictive control framework for resilience operations," *Automatica*, vol. 152, p. 110957, 2023.

[9] G. Liu, C. Wang, T. Tan, D. Shen, G. Fortino, G. Franze, and X. Zhong, "A fractional-order game-theoretic model with sparse attention multi-agent reinforcement learning for malware defense in iov," *IEEE Transactions on Automation Science and Engineering*, 2025.

[10] S. Kullback, *Information Theory and Statistics*. New York: Dover Publications, 1968.

[11] C. Fang, Y. Qi, J. Chen, R. Tan, and W. X. Zheng, "Stealthy actuator signal attacks in stochastic control systems: Performance and limitations," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3927–3934, 2019.

[12] H. Li, J. Zhang, and X. He, "Design of data-injection attacks for cyber-physical systems based on kullback–leibler divergence," *Neurocomputing*, vol. 361, pp. 77–84, 2019.

[13] X.-X. Ren and G.-H. Yang, "Kullback–leibler divergence-based optimal stealthy sensor attack against networked linear quadratic gaussian systems," *IEEE Transactions on Cybernetics*, vol. 52, no. 11, pp. 11 539–11 548, 2021.

[14] Q. Zhang, K. Liu, A. M. Teixeira, Y. Li, S. Chai, and Y. Xia, "An online kullback–leibler divergence-based stealthy attack against cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 68, no. 6, pp. 3672–3679, 2022.

[15] Z. Lian, P. Shi, C. P. Lim, M. Saif, and M. Chen, "Designing optimal stealthy false data injection attacks in cyber-physical systems: Leveraging historical data and kullback-leibler divergence constraints," *IEEE Transactions on Automatic Control*, 2025.

[16] G. Franzè, F. Tedesco, and D. Famularo, "Model predictive control for constrained networked systems subject to data losses," *Automatica*, vol. 54, pp. 272–278, 2015.