2025 IEEE 64th Conference on Decision and Control (CDC)
December 10-12, 2025. Rio de Janeiro, Brazil

Attack Detection Through Time Fingerprinting: A Stochastic Event-Triggered Control Approach

Ivo van Straalen, Alexander J. Gallo, Riccardo M. G. Ferrari and Manuel Mazo Jr.

2025 IEEE 64th Conference on Decision and Control (CDC) | 979-8-3315-2627-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CDC57313.2025.11312582

***Abstract*— We propose a novel cyber-attack detection scheme for control schemes regulated via Stochastic Event-Triggered Control, to detect packets that are maliciously injected by an adversary. The diagnosis scheme relies on assessing whether the arrival time of the information packets received from the controller are compatible with the nominal probability distribution of triggering, or whether they are anomalous. To contrast the threat of an eavesdropping adversary capable of estimating the nominal triggering distribution, we propose a switching scheme, whereby the probability of triggering is drawn among a set of stochastic triggering mechanisms, which is such that the reconstruction of the communication pattern by an eavesdropper becomes computationally infeasible. We design the set of stochastic triggering mechanisms via the solution of an optimization problem, which embeds an explicit trade-off between the properties of the nominal Stochastic Event-Triggered Controller and the detection scheme. The results are illustrated through a numerical example.**

I. Introduction

Cyber-Physical Systems (CPS) are a class of system which lie at the intersection between physical processes and digital components designed to interact with them, e.g. for control purposes. While integration of these systems over large networks has increased their utility, an unintended consequence of including communication networks in control systems has been their exposure to malicious agents capable of leveraging disruptive attacks. Indeed, a number of cyber-attacks against industrial systems have been made public over the last decades, e.g., Stuxnet, Industroyer, Blackenergy, among others [1].

Growing awareness of the problem has led to the development of secure control schemes, where cyber-attacks are countered in a number of different ways, such as detection schemes [2], [3], or encrypted control [4]. Passive methods, relying on system input-output data and model knowledge [5], are not robust to attacks that exhibit similar behavior to the plant dynamics, such as replay, zero dynamics, or covert attacks [6]. Hence, a multitude of active diagnosis methods have been proposed that deliberately alter the closed-loop behavior, such as multiplicative watermarking filters [7] or moving target defense strategies [8].

This work has been partially supported by the EU Horizon program through the project SUDOCO, grant id 101122256.

I. van Straalen, R. M. G. Ferrari, and M. Mazo Jr. are with the Delft Center for Systems and Control, Mechanical Engineering, Delft University of Technology, Delft, Netherlands (email:{i.vanstraalen,r.ferrari,m.mazo}@tudelft.nl);

A. J. Gallo is with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italia (email: alexanderjulian.gallo@polimi.it).

In this paper, we focus on CPSs that are regulated via an Event-Triggered Control (ETC) scheme. Such schemes have seen abundant usage, due to their characteristics of reducing the communication rate necessary for control, while limiting performance loss when compared to periodic sampling [9], [10]. In the context of secure control, ETC presents a source of information that can be exploited to detect malicious data injection, which remains largely untapped in the secure control community: the arrival time of information at the controller. A few works have been recently published investigating this direction: in [11] it has been shown that including ETC within the control system can result in conditions for undetectability of replay attacks that are harder to satisfy for an adversary; in [12], a probabilistic Self-Triggered Control (STC) scheme is proposed to detect replay attacks, relying on early triggers to induce discrepancies between the statistics of the outputs and their replayed values.

We propose an alternative approach, which leverages the information that is contained in the timing of ETC schemes directly. More precisely, we use the arrival time of new measurements to determine whether a packet is malicious. As such we treat the statistics of trigger instants as a *fingerprint*. For this purpose, we propose the use of Stochastic Event-Triggered Control (S-ETC) [13], for which the statistics of the triggering times can be determined explicitly [14]. Furthermore, by switching between several of these schemes, we prevent an eavesdropper from reconstructing the statistics of the communication patterns. By using time as another dimension for detection, we reveal attacks that would otherwise remain stealthy such as replay or covert attacks [5].

Our contributions are summarized as follows:

a. an attack detection method is proposed that relies solely on the arrival time of measurements for the detection of a wide range of attacks, including replay attacks and False Data Injection (FDI), by utilizing S-ETC;
b. theoretical conditions are given that guarantee confidentiality of the switching triggering policy against eavesdropping adversaries;
c. the switching triggering policy is determined via a multi-objective optimization problem is formulated, capable of trading off nominal system performance against false negative rate.

Notation

Let $\mathbb{R}, \mathbb{R}_+, \mathbb{N}, \mathbb{Z}_{\geq 0}$ denote the set of real, positive real, natural, and non-negative integer numbers. The notation $A \succ 0$ indicates that some symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive definite. For any matrix A, denote by A^+ its Moore-Penrose

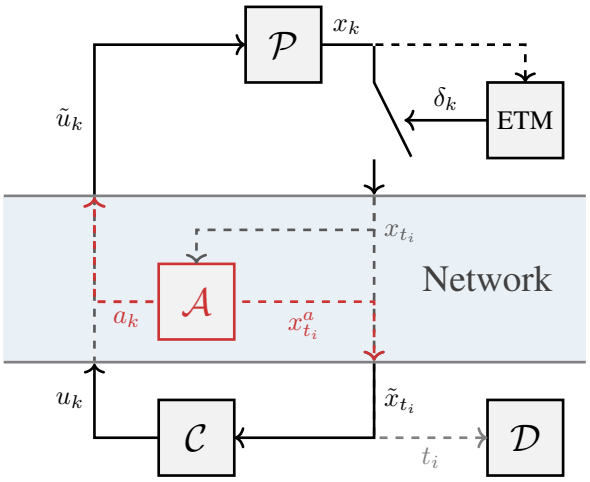


Fig. 1: Schematic of the considered CPS.

inverse, and define $\|x\|_A^2 = x^\top Ax$. The notation $[a < b]$ is used to denote $[a < b] = 1$ if $a < b$ and $[a < b] = 0$ otherwise. The symbols $\mathcal{U}(\mathcal{X})$, $\mathcal{N}(\mu, \Sigma)$, $\text{Exp}(\lambda)$ denote, respectively: the uniform distribution over the set $\mathcal{X}$; the normal distribution with mean μ and variance Σ; and the exponential distribution with rate parameter $\lambda > 0$.

II. Preliminaries and Problem Setup

A. System description

We consider the discrete-time CPS shown in Figure 1. The plant $\mathcal{P}$ is regulated by controller $\mathcal{C}$ over a network compromised by an attacker $\mathcal{A}$. The plant is equipped with an Event-Triggering Mechanism (ETM). The controller is co-located with a detector $\mathcal{D}$ which aims to detect the presence of $\mathcal{A}$. The plant is modeled by the following dynamics:

$$x_{k+1} = Ax_k + Bu_k + w_k, \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state, $u_k \in \mathbb{R}^m$ are control actions and $w_k \in \mathbb{R}^n$ are disturbances modeled by $w_k \sim \mathcal{N}(0, W)$, where $W \succ 0$. Furthermore, the pair (A, B) is controllable.

B. Stochastic Event-Triggered Control

The ETM is defined by the stochastic triggering policy δ_k:

$$\delta_k = \begin{cases} 1 & \text{if } \varphi_k(x_k, \bar{x}_k) > \zeta_k, \text{ or } \Delta t_i = S; \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\bar{x}_k$ is the last transmitted value of x available at time k; the threshold ζ_k is a random variable; $\varphi_k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is the so-called triggering function; Δt_i is the i-th Inter-Event Time (IET), determined by the triggering instants t_i:

$$\Delta t_i \triangleq t_i - t_{i-1}, \quad t_i = \min\{t' \in \mathbb{N} \mid t' > t_{i-1}, \delta_{t'} = 1\};$$

and $S \in \mathbb{Z}_{\geq 0}$ is the maximum allowed inter-event time. Unless specified otherwise, we consider that $k \in [t_i, t_{i+1})$. The triggering policy (2) and the dynamics (1) then characterize the time instants when information is transmitted from the plant to the controller: $\mathcal{K} = \{k \in \mathbb{Z}_{\geq 0} \mid \delta_k = 1\}$. These will be referred to as *triggering* or *arrival times*.

For general φ_k and ζ_k, the statistics of δ_k and x_k are complex or outright intractable to analyze. However, [13] proposes a φ_k for which the Gaussian nature of the disturbances w_k are preserved in the state. This allows to determine the probability of triggering in closed form $\mathbb{P}\left[t_{i+1} - t_i = \Delta \mid (\varphi_j, \zeta_j)_{j \leq t_{i+1}}\right] \triangleq p_{t_i}(\Delta)$ and furthermore design the parameters of the ETM (2) to prescribe these probabilities *a priori*. In Section III-B we exploit this property to detect cyber-attacks.

C. Cyber-Attack Modelling and Detection

We assume that a malicious agent with the following capabilities has compromised the communication network:

Attacker Knowledge: The agent has full knowledge of the plant and controller and is able to eavesdrop the *communication pattern* between $\mathcal{P}$ and $\mathcal{C}$. The information available to the agent at time k is given by:

$$\mathcal{I}_k^a \triangleq \{A, B, \kappa_k, (\delta_i)_{i \leq k}, (x_{i \in \mathcal{K}})_{i \leq k}, S\},$$

where $\kappa_k : \mathbb{R}^n \to \mathbb{R}^m$ is the feedback law: $u_k = \kappa_k(x_k)$. Furthermore, we assume that the adversary has knowledge of the structure of the ETM (2), as well as the detection scheme we propose in Section IV. However, we consider the exact details of the ETM as a secret, i.e. $p_{t_i}(\Delta), \varphi_k, \zeta_k \notin \mathcal{I}_k^a$.

Attacker model: The attack is modeled as a FDI on the communication between $\mathcal{P}$ and $\mathcal{C}$. We denote by $\mathcal{T}_a := [\bar{t}_0, \underline{t}_0] \times [\bar{t}_1, \underline{t}_1] \times \dots$ the set of time-intervals in which the adversary is active, where each $\bar{t}_j, \underline{t}_j \in \mathbb{Z}_{\geq 0}$. On the measurement channel, the adversary blocks communication when active, and injects malicious packets at time instants $\mathcal{K}_a \subset \mathcal{T}_a$. The attacked signals $\tilde{x}_k$ and $\tilde{u}_k$ are given by:

$$\tilde{x}_k = \begin{cases} x_{t_i}, & k \notin \mathcal{T}_a \\ x^a_{t^a_i} & k \in \mathcal{T}_a, \end{cases} \qquad \tilde{u}_k = \begin{cases} u_k, & k \notin \mathcal{T}_a \\ u^a_k & k \in \mathcal{T}_a, \end{cases}$$

where $t_i^a \triangleq \max\{t' \in \mathcal{K}_a \mid t' \leq k\}$. Generally, these injection times are described by an (unknown) distribution:

$$\mathbb{P}[t_i^a - t_{i-1}^a = \Delta \mid \mathcal{I}_{t_i}^a] = p_{t_i}^a(\Delta).$$

In order to attain maximum damage, an adversary tries to remain undetected for as long as possible. Hence, they try to define $p_{t_i}^a(\Delta)$ to match $p_{t_i}(\Delta)$ as closely as possible.

Remark 1. *In this work, we are interested in attack detection based solely on arrival times. Potential synergy with a residual-based detection scheme is left as future work.*

The discussed attack model covers a wide range of attack types, such as covert attacks, replay attacks and FDI attacks [6]. Here, we make a quick note on describing replay attacks as a means to motivate our approach.

Example (Replay Attacks)**.** *Starting from t_0^r, the adversary observes $N_r \in \mathbb{N}$ transmissions $T_r := \{t_0^r, \dots, t_{N_r-1}^r\}$, inter-event times $\Delta T_r := \{\Delta t_0^r, \dots, \Delta t_{N_r-1}^r\}$ and measurements $X_r := \{x_j^r \mid j \in T_r\}$. For simplicity, $\mathcal{T}_a = [T_s, \infty)$, where T_s is the start of the attack. The injection times are:*

$$\mathcal{K}_a = \{t_j^a \mid t_0^a = T_s,\ t_{j+1}^a = t_j^a + \Delta t^r_{j \bmod N_r},\ j \in \mathbb{Z}\},$$

with probability distribution $p^a_{t^a_j}(\Delta) = \delta(\Delta - \Delta t^r_{j \bmod N_r})$. *The injected measurements are given by* $x^a_{t^a_j} = x^r_{j \bmod N_r}$, *which follow the same statistics as nominal measurements.*

D. Problem formulation

The objective of this paper is to define a detection strategy capable of detecting the presence of maliciously injected data based solely on the arrival times.

Problem 1. *Given the system* (1)*, and the attack strategy in Section II-C, define an ETM* (2) *and detection scheme that allows to determine whether* $k \in \mathcal{K}_a$, *for all* $k \in \mathcal{K} \cup \mathcal{K}_a$, *based on* $p_{t_i}(\Delta)$, $k \in [t_i, t_{i+1})$.

For a measurement arriving at $k \in [t_i, t_{i+1})$, a detection scheme using $p_{t_i}(\Delta)$ can be designed as:

$$\mathcal{D}_k(\Delta) = \begin{cases} \mathcal{H}_0, & p_{t_i}(\Delta) \geq \eta, \\ \mathcal{H}_1. & \text{otherwise}, \end{cases} \tag{3}$$

where hypotheses $\mathcal{H}_0, \mathcal{H}_1$ denote that the received measurement is genuine or, respectively, maliciously injected. The threshold η is a design parameter that defines how unlikely a given IET must be, before triggering an alarm.

If the information $\mathcal{I}^a_{t_i}$ is sufficient for an adversary to predict for which $\Delta : p_{t_i}(\Delta) \geq \eta$, then our detection scheme will fail to detect their presence. As such we pose the following intermediary problem, that aims at counteracting an eavesdropping adversary.

Problem 2. *Design the ETM* (2) *such that the information* $\mathcal{I}^a_k$ *is not informative for predicting a genuine IET with a given significance* $N < 1$*:*

$$\mathbb{P}[\Delta^a \in \{\Delta \mid p_{t_i}(\Delta) \geq \eta\} \mid \mathcal{I}^a_{t_i}] \leq N, \, \forall \Delta^a \leq S, \, i \in \mathbb{Z}.$$

III. A Switching Triggering Policy

Problem 2 poses a strong constraint on the ETM (2). Hence, first we address these constraints and derive sufficient conditions that will be used in the solution of Problem 1.

A. Cryptographically Secure Triggering

In this work, the triggering probability is parametrized by a time-varying parameter θ_k, taken from a constant set of parameters Θ, $|\Theta| = N_\theta$:

$$p_{t_i}(\Delta) = \mathbb{P}[\Delta t_i = \Delta \mid \theta_{t_i}]. \tag{4}$$

The parameter θ_{t_i} is considered secret. Hence the triggering probability observed on the communication link is

$$\mathbb{P}[\Delta t_i = \Delta \mid \mathcal{I}^a_k] = \sum_{\vartheta \in \Theta} \mathbb{P}[\Delta t_i = \Delta \mid \theta_{t_i}] \mathbb{P}[\theta_{t_i} = \vartheta]. \tag{5}$$

To prevent an eavesdropper from reconstructing the conditional triggering probability, the distribution of Δt_i induced by the sequence $(\theta_{t_i})_{i \in \mathbb{N}}$ should be *hard* to distinguish from uniform randomness. Specifically, we use the notion of *computational indistinguishability* to quantify this objective.

Definition 1 (Negligible Function [15])**.** *A function* $\mu : \mathbb{N} \to \mathbb{R}_{\geq 0}$ *is* ***negligible*** *if for every positive polynomial* $p(\cdot)$, $\exists N > 0$ *such that* $\forall n > N$, $\mu(n) \leq \frac{1}{p(n)}$.

Definition 2 (Computational Indistinguishability [15])**.** *Two ensembles of random variables* $X \triangleq \{X_n\}_{n \in \mathbb{N}}$, $Y \triangleq \{Y_n\}_{n \in \mathbb{N}}$ *are* ***computationally indistinguishable****, denoted by* $X \approx Y$, *if for every Probabilistic Polynomial Time (PPT) algorithm* $\mathcal{A}$ *there is a negligible function* $\mu(n)$ *such that:*

$$|\mathbb{P}[\mathcal{A}(X_n) = 1] - \mathbb{P}[\mathcal{A}(Y_n) = 1]| \leq \mu(n).$$

Lemma 1. *Consider a switching probability mass function* (4)*, satisfying the following two conditions:*

(a) *The distribution of the parameter* θ_{t_i} *is computationally indistinguishable from the uniform distribution:*

$$\theta_{t_i} \approx \mathcal{U}(\Theta). \tag{6}$$

For example, θ_{t_i} *is generated by a Pseudo-Random Number Generator (PRNG).*

(b) *For every IET* Δ*, the average probability is uniform:*

$$\frac{1}{N_\theta} \sum_{\vartheta \in \Theta} \mathbb{P}[\Delta t_i = \Delta \mid \vartheta] = \frac{1}{S}, \, \forall \Delta \in \{1, \ldots, S\}. \tag{7}$$

Then the distribution of IETs (5) *is computationally indistinguishable from a uniform distribution:* $\tau_i \approx \mathcal{U}([1, S])$.

Proof. The proof follows from contradiction. Firstly, note that condition (*b*) allows the construction of a PPT $\mathcal{F}_i$ such that $\mathcal{F}_i(\theta_{t_i})$ is identically distributed to τ_i and $\mathcal{F}_i(\mathcal{U}(\Theta)) \equiv \mathcal{U}([1, S])$. Namely, $\mathcal{F}_i(\vartheta)$ outputs Δ with probability $\mathbb{P}[\Delta t_i = \Delta \mid \vartheta]$. Suppose that τ_i is *not* computationally indistinguishable from a uniform distribution: this implies that a PPT algorithm $\mathcal{A}$ can be constructed that is able to distinguish τ_i from $\tau^\star$, for $\tau^* \sim \mathcal{U}([1, S])$. Then, by combining $\mathcal{A}$ and $\mathcal{F}_i$, one can define a PPT $\mathcal{A}^* \triangleq \mathcal{A}(\mathcal{F}_i(\cdot))$ capable of distinguishing θ_{t_i} from θ^*, for $\theta^* \sim \mathcal{U}(\Theta)$. This contradicts condition (*a*). Hence, τ_i is *computationally indistinguishable* from the uniform distribution. □

B. Formulation of the Stochastic Event Triggering Mechanism

Lemma 1 poses constraints on the design of the ETM (2). To satisfy them, we design a switching ETM based on techniques from [14]. For the sake of simplifying notation, we define $\Theta \triangleq \{1, \ldots, N_\theta\}$, and abbreviate the probability mass function in (4) as $p^\vartheta_\Delta \triangleq \mathbb{P}[\Delta t_i = \Delta \mid \theta_{t_i} = \vartheta]$.

Event-Triggering Mechanism*:* The ETM is defined by:

$$\delta_k = \begin{cases} 1 & \text{if } \frac{1}{2} \|x_k - \hat{x}_{k|k-1}\|^2_{\Sigma^+_{k|k-1}} > \zeta_k \\ 0 & \text{otherwise}, \end{cases} \tag{8}$$

where the random variable $\zeta_k \sim \mathrm{Exp}(\lambda_k(\vartheta))$ is the stochastic threshold. After triggering, the values λ_k are reset: $\lambda_k(\vartheta) = \lambda_{k-t_i}(\vartheta)$, $k \in \{t_i + 1, \ldots, t_i + S\}$, $\vartheta \in \Theta$ and $\lambda_S(\vartheta) = \infty$. The state estimates are calculated recursively:

$$\begin{aligned} \hat{x}_{k|k-1} &= A\hat{x}_{k-1|k-1} + Bu_{k-1}, \\ \hat{x}_{k|k} &= \begin{cases} \tilde{x}_k & \text{if } \delta_k = 1, \\ \hat{x}_{k|k-1} & \text{otherwise}, \end{cases} \end{aligned} \tag{9}$$

The state covariances $\Sigma_{k|k-1}$, $\Sigma_{k|k}$ are given by:

$$\Sigma_{k|k-1} = \Omega_{k-t_i}(\theta_k),\ \Sigma_{k|k} = (1-\delta_k)\cdot\frac{\Sigma_{k|k-1}}{1+\lambda_k(\theta_k)}, \quad (10)$$

where $\Omega_\ell(\vartheta)$, $\ell \in \{1,\dots,S\}$ are matrices calculated *offline*:

$$\Omega_1(\vartheta) = W,\ \Omega_{\ell+1}(\vartheta) = \frac{1}{1+\lambda_\ell(\vartheta)}A\Omega_\ell A^\top + W. \quad (11)$$

Lastly, the probability of triggering at time $t_i + j$ can be written in closed form as [14]:

$$\mathbb{P}[\delta_{t_i+j}=1|\delta_{t_i+j-1},\dots,\delta_1] = 1-(1+\lambda_j(\vartheta))^{-\frac{n_j(\vartheta)}{2}} \quad (12)$$

where $n_j(\vartheta) = \operatorname{rank}(\Omega_j(\vartheta))$. The above can be inverted to compute $\lambda_j(\vartheta)$, $j \in \{1,\dots,S\}$:

$$\lambda_j(\vartheta) = \left(\frac{\bar{p}^\vartheta_j}{\bar{p}^\vartheta_{j-1}}\right)^{-\frac{2}{n_j(\vartheta)}} - 1, \quad (13)$$

where $\bar{p}^\vartheta_j = 1 - \sum_{\ell=1}^{j} p^\vartheta_\ell$.

Feedback law: The input is given by:

$$u_k = K\hat{x}_{k|k}, \quad (14)$$

such that $A_{cl} = (A+BK)$ is Schur and K is determined by minimizing the average quadratic cost with $Q \succ 0, R \succ 0$:

$$J = \limsup_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N}\mathbb{E}\left[x^\top Q x + u^\top R u\right], \quad (15)$$

which is given by $K = -(R+B^\top PB)^{-1}B^\top PA$, where P is the solution to the discrete-time algebraic Riccati equation with weighting matrices Q and R and we note that the optimal value for K is independent of p^ϑ_Δ. For fixed p^ϑ_Δ, the optimal performance cost (15) is equal to:

$$J_{lqr}(\vartheta) = \operatorname{tr}(PW) + \frac{1}{H}\sum_{\ell=1}^{S}\bar{p}^\vartheta_\ell \operatorname{tr}\left(\frac{\hat{R}\Omega_\ell(\vartheta)}{1+\lambda_\ell(\vartheta)}\right), \quad (16)$$

where H is the average inter-event time, and $\hat{R} = K^\top(R+B^\top PB)K$. Optimizing (16) with respect to p^ϑ_Δ provides the optimal performance in terms of (15).

Parameter Switching Law: To complete the definition of the ETM, a switching law for the parameters θ_k is designed that satisfies condition (6):

$$\theta_{k+1} \triangleq \begin{cases}\theta_k, & \delta_k = 0\\ \vartheta \sim \mathcal{U}(\{1,\dots,N_\theta\}) & \delta_k = 1.\end{cases} \quad (17)$$

In other words, after every trigger generated by (8), the next parameter is chosen uniformly at random. For our scheme to function correctly, we make the following assumption:

Assumption 1. *The parameter θ_k remains synchronized at the sensors $\mathcal{S}$ and detector $\mathcal{D}$, i.e. $\theta^{\mathcal{S}}_k \equiv \theta^{\mathcal{D}}_k$.*

Remark 2. *This assumption can be satisfied, e.g., by using a PRNG with identical seed at both $\mathcal{S}$ and $\mathcal{D}$.*

C. Nominal Stability

Lastly, we remark on the nominal closed-loop stability.

Proposition 1. *Consider dynamics* (1)*, control law* (14)*, ETM* (8)*, any choice of p^ϑ_Δ and switching of θ_k according to* (17)*. Then the closed-loop system is mean-square stable under nominal conditions. I.e. there exist $c_1, c_2 \in \mathbb{R}_+$ and $\mu \in (0,1)$, such that for any $x_0 \in \mathbb{R}^n$ and $k \in \mathbb{N}$:*

$$\mathbb{E}[x_k^\top x_k] \le c_2 + c_1\mu^k x_0^\top x_0. \quad (18)$$

Proof. The proof follows from [14, Lemma 2] by showing that there exists a $d \in \mathbb{R}$ such that $\operatorname{trace}(\Sigma_{k|k-1}) \le d$. Note that, for $k \in [t_i, t_{i+1})$, we have that $\theta_k = \vartheta$. Furthermore, $\Sigma_{t_i|t_i} = \mathbb{0}_{n\times n}$ and $t_{i+1} - t_i \le S$. Hence, the upper-bound is

$$d \triangleq \sup\left\{\operatorname{tr}(\Omega_j(\vartheta) \mid j \in \{1,\dots,S\},\ \vartheta \in \Theta\right\},$$

which exists and is finite, as $\lambda_j(\vartheta) > 0$ and the covariance can only grow to a finite size in finite time. Hence, $\operatorname{trace}(\Sigma_{k|k-1}) \le d$, and by [14, Lemma 2] the closed-loop system is *mean-square stable*. □

IV. Detection Scheme

A. Detector Logic

Using $p_{t_i}(\Delta) = p^{\theta_{t_i}}_\Delta$, the detector (3) is restated as:

$$\mathcal{D}(\Delta,\theta_{t_i}) = \begin{cases}\mathcal{H}_0, & p^{\theta_{t_i}}_\Delta \ge \eta,\\ \mathcal{H}_1. & \text{otherwise},\end{cases} \quad (19)$$

where Δ is the IET of the received packet, and η is a threshold to be designed. The threshold η establishes a trade-off between misidentifying valid packets and malicious packets, i.e., between false-positive and false-negative rates.

Detection of Replay Attacks: Replay attacks utilize realizations Δt_j of the IET with probability distribution $\mathbb{P}[\Delta t_i = \Delta \mid \theta_j]$. Detection of these attacks relies on the de-synchronization between the parameters θ_j used during an attack and used for detection. The attack is detected at time t^a_j if $\theta_{t^a_j} \neq \theta_{t^r_{j \bmod N_r}}$ and $\mathbb{P}[\Delta t_j = \Delta t^a_j \mid \theta_{t^r_{j \bmod N_r}}] < \eta$.

B. Detection Rates

Rate of False Positives: A false positive event is that in which a legitimate triggering instant is misclassified by the detection scheme as being a maliciously injected signal, i.e., $\mathcal{D}(t_i - t_{i-1}, \theta_{t_{i-1}}) = \mathcal{H}_1$, for $t_i \in \mathcal{K}$. The probability of such an event depends on the definition of p^ϑ_Δ, and as such can be computed *a priori* for a fixed $\theta_{t_i} = \vartheta$ as follows:

$$r_{FP}(\vartheta) \triangleq \mathbb{P}\left[\mathcal{D} = \mathcal{H}_1 \mid \mathcal{H}_0,\ \vartheta\right] = \sum_{\Delta=1}^{S}\left[p^\vartheta_\Delta < \eta\right]p^\vartheta_\Delta. \quad (20)$$

Rate of False Negatives: Conversely, the rate of false negatives can be understood as the rate of missed detections, when a malicious agent injects data in the communication network, i.e., $\mathcal{D}(t^a_{i-1} - t_{i-1}, \theta_{t_{i-1}}) = \mathcal{H}_0$, although $t^a_{i-1} \in \mathcal{K}^a$. The false negative rate can be computed as:

$$r_{FN}(\theta_{t_i}, t_i) \triangleq \mathbb{P}\left[\mathcal{D} = \mathcal{H}_0 \mid \mathcal{H}_1,\ \vartheta\right] = \sum_{\Delta=1}^{S}\left[p^{\theta_{t_i}}_\Delta \ge \eta\right]p^a_{t_i}(\Delta).$$

Note that the adversary probability $p^a_{t_i}(\Delta)$ is unknown. Nevertheless, we motivate an appropriate choice for $p^a_{t_i}(\Delta)$ and the threshold η in the next section.

C. Threshold Selection and Optimal Adversary Strategy

To maximize their effect on the control system, an adversary trying to remain hidden will design their probability $p^a_{t_i}(\Delta)$ to maximize the false negative rate. However, due to Lemma 1, they will not be able to directly use p^ϑ_Δ in their design. The best strategy they could employ, is to maximize the expected false negative rate, with unknown p^ϑ_Δ:

$$p^{a\,\star}_\Delta \triangleq \arg\max_{p^a_\Delta} \left\{ \mathbb{E}_{\mathbf{p}}\left[\hat{r}_{FN}(p, p^a_\Delta)\right] \,\middle|\, \sum_{\Delta=1}^{S} p^a_\Delta = 1 \right\}, \quad (21)$$

where $\mathbf{p} \in \mathcal{P} \triangleq \left\{ p \in \mathbb{R}^{N_\theta \times S}_{\geq 0} \mid \forall \vartheta : \sum_{\Delta=1}^{S} p^\vartheta_\Delta = 1 \wedge \forall \Delta : \sum_{\vartheta \in \Theta} p^\vartheta_\Delta = \frac{N_\theta}{S} \right\}$ is the set of any possible distribution satisfying (7). If the adversary concludes that all possible $\mathbf{p} \in \mathcal{P}$ are equally likely, the expected value of the false negative rate for unknown p^ϑ_Δ can be expressed as:

$$\begin{aligned}\mathbb{E}_{\mathbf{p}}\left[\hat{r}_{FN}(p, p^a_\Delta)\right] &= \frac{1}{A \cdot N_\theta} \sum_{\vartheta \in \Theta} \sum_{\Delta=1}^{S} p^a_\Delta \int_{\mathcal{P}} \left[p^\vartheta_\Delta \geq \eta\right] d\mathbf{p}, \\ &= \frac{c}{A} \sum_{\Delta=1}^{S} p^a_\Delta = \frac{c}{A},\end{aligned}$$

where $A \triangleq \int_{\mathcal{P}} d\mathbf{p}$. The second equality holds as each integral evaluates to the same value c. Thus, any p^a_Δ such that $\sum_{\Delta=1}^{S} p^a_\Delta = 1$ is a maximizer for (21). Hence, the adversary will expect to see the same false negative rate for any p^a_Δ.

This argument motivates the following assumption, which allows us to use the false negative rate in the design of p^ϑ_Δ:

Assumption 2. *The defender's belief on the malicious packet injection is that each IET occurs with the same probability:*

$$\hat{p}^a_{t_i}(\Delta) = \frac{1}{S}, \forall \Delta \in \{1, \ldots, S\}, \forall t_i \in \mathcal{K} \cup \mathcal{K}_a.$$

This suggests a natural choice for the detection threshold: $\eta = \frac{1}{S}$. Using this adversary probability, the defender's expected false negative rate for a fixed $\theta_{t_i} = \vartheta$ is:

$$\hat{r}_{FN}(\vartheta) = \frac{1}{S} \sum_{\Delta=1}^{S} \left[p^\vartheta_\Delta \geq \frac{1}{S} \right]. \quad (22)$$

V. Optimal Triggering Policy

In principle, the detection function (19) provides a solution to Problem 1 and 2 for any values for p^ϑ_Δ such that (7) holds. However, not all choices provide the same false-positive and false-negative rates. Hence, we pose the following optimization problem, as a trade-off between *nominal control performance* and *false-negative detection rate*:

$$\min_{p^\vartheta_\Delta} \frac{1}{N_\theta} \sum_{\vartheta=1}^{N_\theta} \left\{ J(\vartheta) + \gamma \hat{r}_{FN}(\vartheta) \right\} \quad (23)$$

$$\text{s.t.} \quad \sum_{\Delta=1}^{S} p^\vartheta_\Delta = 1, \forall \vartheta \in \{1, \ldots, N_\theta\} \quad (24)$$

$$\frac{1}{N_\theta} \sum_{\vartheta=1}^{N_\theta} p^\theta_\Delta = \frac{1}{S}, \forall \Delta \in \{1, \ldots, S\} \quad (25)$$

$$r_{FP}(\vartheta) \leq \beta, \forall \vartheta \in \{1, \ldots, N_\theta\} \quad (26)$$

where $\gamma, \beta, N_\theta > 0$ are design parameters and $J(\vartheta)$ is the nominal quadratic cost (15) for mode ϑ. Constraint (25) ensures condition (7) is met. The posed optimization problem is non-convex. However, in practice this is not an issue, because it only needs to be solved offline once. Moreover, sub-optimal solutions are acceptable, as satisfaction of constraint (25) provides cryptographic security.

Proposition 2. *For any given S, N_θ, the optimization problem* (23)-(26) *is feasible.*

Proof. For simplicity, consider the case where $p^\vartheta_\Delta = \frac{1}{S} \cdot \ell$, $\ell \in \{0, \ldots, S\}$. This automatically satisfies constraints (26) as $r_{FP}(\vartheta) = 0$. We can characterize the probabilities with a matrix $P \in \mathbb{Z}^{N_\theta \times S}_{\geq 0}$ such that $p^\vartheta_\Delta \equiv \frac{1}{S} P_{\vartheta\Delta}$. Constraints (24), (25) can then be written as

$$\begin{aligned} P \mathbb{1}_S &= S \mathbb{1}_{N_\theta} \\ \mathbb{1}^\top_{N_\theta} P &= N_\theta \mathbb{1}^\top_S, \end{aligned}$$

where $\mathbb{1}_n$ is a n-dimensional vector with every entry equal to 1. Such a matrix P is guaranteed to exist [16, Section 6.2]. Hence, the problem (23)-(26) is feasible. □

VI. Numerical Example

The effectiveness of the proposed approach is demonstrated by performing a replay attack on the classical example of the linearized batch plant [9], discretized with zero-order hold and sampling time $h = 0.05$, resulting in the matrices:

$$A = \begin{bmatrix} 1.080 & -0.005 & 0.290 & -0.237 \\ -0.027 & 0.810 & -0.003 & -0.032 \\ 0.045 & 0.189 & 0.731 & 0.235 \\ 0.001 & 0.189 & 0.055 & 0.911 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0.039 \\ 0.256 & 0 \\ 0.084 & 0.135 \\ 0.084 & 0.005 \end{bmatrix}, \quad \rho(A) = 1.1047,$$

where $u_k = K\hat{x}_{k|k}$, K is solved for by the algebraic Riccati equation with $Q = I_4$ and $R = I_2$ and $w_k \sim \mathcal{N}(0,\, 10^{-3} \cdot I_4)$. We have used Gurobi [17] with its Python interface to solve the optimization problem (23), resulting in the probability distributions shown in Figure 2, with $r^\star_{FN}(\theta^1) = 0.3333$, and $r^\star_{FN}(\theta^{\geq 2}) = 0.25$. The adversary records from $t^r_0 = 5\,[\text{s}]$ until $9\,[\text{s}]$ resulting in $N_r = 14$ samples observed. The replay attack is executed starting from $T_s = 10\,[\text{s}]$. In Figure 3, the results of this simulation are shown. The replay attack is effective at destabilizing the plant using the shown inputs,

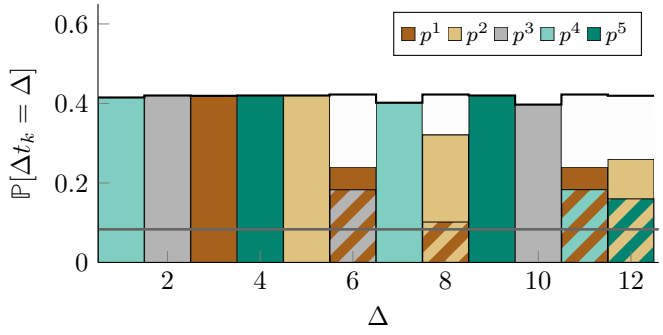
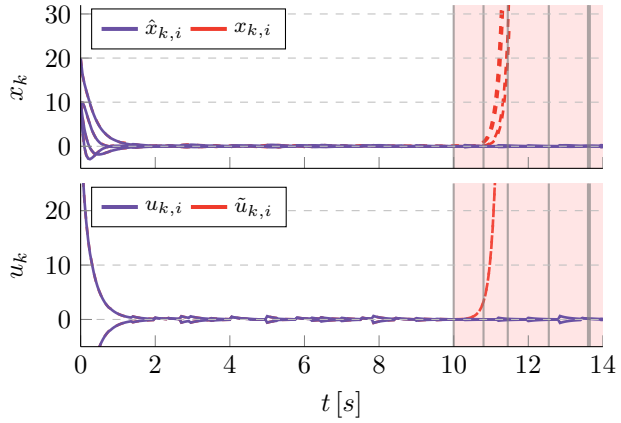


Fig. 2: Probability of triggering for each of the $N_\theta = 5$ different modes. The gray solid line indicates $1/S$; the black line indicates the sum $\sum_\vartheta p^\vartheta_\Delta$.

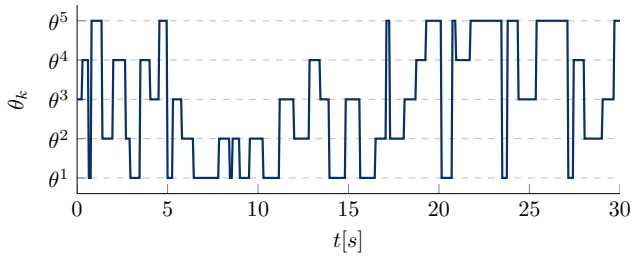
however the state estimate $\hat{x}_k$ calculated by the controller remains close to zero. Our proposed detector is able to detect the attack, with the first alarm raised at $10\,[\mathrm{s}]$. Running 100 Monte-Carlo simulations over the same horizon length, we find an empirical false negative rate $r_{FN} = 0.492$ and average first detection time $10.178\,[\mathrm{s}]$.

VII. Conclusions and Future Work

In this paper, we have provided a first approach to detection of injection attacks based on the arrival time of measurements in an ETC setting. We employ a switching S-ETC structure, which permits the explicit design of the probability of triggering and inhibits reconstruction by eavesdroppers. The former allows us to design the triggering conditions to achieve desirable detection performance, by limiting false positives and minimizing false negatives. Future work includes increasing robustness against network effects such as delays and packet drops, studying the effects of Denial-of-Service (DoS) attacks. Finally, we highlight the potential effectiveness of the combination of our proposed scheme with conventional observer-based detectors.



(a) System dynamics and estimated state (above). Controller input and attacked input (below). Attacks are detected at the time instant indicated by the gray lines.



(b) Value of θ_k over time, corresponding to (a).

Fig. 3: Simulation of the closed loop system, with triggering probabilities as in Figure 2 under replay attack starting from 10 [s] (indicated in the red region).

References

[1] K. E. Hemsley and D. R. E. Fisher, "History of Industrial Control System Cyber Incidents," tech. rep., United States, 2018.

[2] H. Sandberg, V. Gupta, and K. H. Johansson, "Secure networked control systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 445–464, 2022.

[3] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annual Reviews in Control*, vol. 47, Jan. 2019.

[4] M. S. Darup, A. B. Alexandru, D. E. Quevedo, and G. J. Pappas, "Encrypted Control for Networked Systems: An Illustrative Introduction and Current Challenges," *IEEE Control Systems Magazine*, vol. 41, no. 3, pp. 58–78, 2021.

[5] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, Jan. 2015.

[6] M. S. Chong, H. Sandberg, and A. M. H. Teixeira, "A Tutorial Introduction to Security and Privacy for Cyber-Physical Systems," in *18th European Control Conference (ECC)*, pp. 968–978, June 2019.

[7] R. M. G. Ferrari and A. M. H. Teixeira, "Detection and Isolation of Replay Attacks through Sensor Watermarking," *IFAC-PapersOnLine*, vol. 50, pp. 7363–7368, July 2017.

[8] P. Griffioen, S. Weerakkody, and B. Sinopoli, "A moving target defense for securing cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2016–2031, 2020.

[9] W. P. M. H. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *inproceedings*, pp. 3270–3285, 2012.

[10] D. P. Borgers and W. P. M. H. Heemels, "Event-separation properties of event-triggered control systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 10, pp. 2644–2656, 2014.

[11] A. Barboni, A. W. Al-Dabbagh, and T. Parisini, "An Event-Triggered Watermarking Strategy for Detection of Replay Attacks," *IFAC-PapersOnLine*, vol. 55, pp. 317–322, Jan. 2022.

[12] B. Wolleswinkel, R. Ferrari, and M. Mazo, "A Self-Triggered Control Watermarking Scheme for Detecting Replay Attacks," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*, Dec. 2024.

[13] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic Event-Triggered Sensor Schedule for Remote State Estimation," *IEEE Transactions on Automatic Control*, vol. 60, 2015.

[14] F. D. Brunner, D. Antunes, and F. Allgöwer, "Stochastic thresholds in event-triggered control: A consistent policy for quadratic control," *Automatica*, vol. 89, pp. 376–381, Mar. 2018.

[15] O. Goldreich, *Foundations of Cryptography: Volume 1: Basic Tools*, vol. 1. Cambridge: Cambridge University Press, 2001.

[16] R. A. Brualdi and H. J. Ryser, *Combinatorial Matrix Theory*. Encyclopedia of Mathematics and its Applications, Cambridge: Cambridge University Press, 1991.

[17] Gurobi Optimization, LLC, "Gurobi optimizer reference manual," 2024.