A masking protocol for private communication and attack detection in nonlinear observers

Andreu Cecilia, Daniele Astolfi, Giacomo Casadei, Ramon Costa-Castelló, Dragan Nešić

A masking protocol for private communication and attack detection in nonlinear observers

Andreu Cecilia[1,3], Daniele Astolfi[1], Giacomo Casadei[2], Ramon Costa-Castelló[3] and Dragan Nešić[4]

***Abstract*—This work presents a novel masking protocol to secure the communication between a nonlinear plant and a nonlinear observer. Communication is secured in two senses. First, the privacy of the plant is preserved during the communication. Second, the protocol can detect a false-data injection attack in the communication link. The masking protocol is based on the use of washout-filters in nonlinear observers and the internal model principle.**

***Index Terms*—Cyber-security, Eavesdropping, Privacy, false-data Injection Attack, Washout-filter, Nonlinear observers.**

I. Introduction

Cyber-physical systems are systems with both physical and cyber components, where, in many cases, networking is used to connect the various system components. Currently, with the widespread usage of cyber-physical systems, security has become a real concern as wireless communication links serve as new access points for malicious agents that want to disrupt the system [1], [2].

Although there is a rich literature on cyber-attack detection and mitigation, e.g. [2]–[5], cyber-physical systems should be resilient in the sense that their attack space is minimized [6]. Nonetheless, designing a completely secure system is infeasible in practice and the presence of potential vulnerabilities has to be assumed. With this in mind, it is imperative to prevent a malicious agent from gathering sensitive data that may uncover the potential sensitive points. This fact has motivated the design of multiple security strategies that preserve the privacy of the transmitter during the communication of system data [7]–[10]. In this sense, a delicate point is the communication link between a sensor and a remote observer. Indeed, observers are designed over measured signals that have to present (at least) some minimal observability condition [11]. Thus, any observer will be the receiver of highly-informative data that could be used by a malicious agent to study the system. This fact has motivated the design of a set of security strategies to preserve privacy during remote estimation [12]–[16]. Nonetheless, all these security strategies assume a linear plant and observer. To the best of the author's knowledge, no general solution has been proposed to the privacy problem of remote state estimation for the nonlinear case. In this work, we propose a novel masking framework to fill this gap.

The framework that we propose is based on adding a masking disturbance to the signal broadcasted from the plant (the transmitter) to the observer (the receiver). The masking signal is generated by a known autonomous system. Then, a filter is included at the observer side in order to eliminate the masking signal. Such a filter is based on a washout filters approach and the internal model principle [17]. We consider generic nonlinear plants and observers [11] satisfying an incremental passivity assumption [18].

The idea of perturbing the transmitted signal to preserve the privacy (i.e. resilience of eavesdropping attacks) is strongly related to the concept of chaotic masking [19], [20]. Although the idea is very similar, the architecture proposed in this work is completely different. In contrast, the proposed architecture presents a set of advantages over the one presented in [19]:

- No additional communication links between the plant and the observer need to be deployed.
- The proposed scheme can be used to detect false-data injections during the communication [3].

We remark that the second point is of significant interest in any masking protocol. Indeed, all masking protocols, in addition to hiding the transmitted data, have the unintended consequence of also masking any false-data injected by a malicious agent. As a consequence, a poor masking protocol can actually ease stealthy false-data injection attacks.

II. Problem Statement

A. The framework

We consider the problem of state estimation for nonlinear systems of the form

$$\dot{x} = f_p(x, u), \tag{1a}$$

$$y = h(x), \tag{1b}$$

where $x \in \mathbb{R}^{n_x}$ is the unknown state of the system, $u \in U$ is a known input signal taking values in a compact set $U \subset \mathbb{R}^{n_u}$ and $y \in \mathbb{R}^{n_y}$ is the output signal transmitted between the

[1] A. Cecilia and D. Astolfi are with Univ Lyon, Université Claude Bernard Lyon 1, CNRS, LAGEPP UMR 5007, 43 boulevard du 11 novembre 1918, F-69100, Villeurbanne, France (name.surname@univ-lyon1.fr).

[2] G. Casadei is with Laboratoire Ampere Dpt. EEA of the Ecole Centrale de Lyon, Universite de Lyon, 69134 Ecully, France (giacomo.casadei@ec-lyon.fr).

[3] R. Costa-Castelló and A. Cecilia are with Universitat Politécnica de Catalunya, Avinguda Diagonal, 647, 08028 Barcelona, Spain. (ramon.costa@upc.edu, andreu.cecilia@upc.edu).

[4] D. Nešić is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, 3010 Victoria Australia. (dnesic@unimelb.edu.au).

This research was partially supported by the French Grant ANR ALLIGATOR (ANR-22-CE48-0009-01), by the project ACROBA and the Spanish Ministry of Universities funded by the European Union - NextGenerationEU (2022UPC-MSC-93823), and is part of the projects MAFALDA (PID2021-126001OB-C31) and MASHED (TED2021-129927B-I00), funded by MCIN/ AEI/10.13039/501100011033 and by the European Union Next GenerationEU/PRTR.

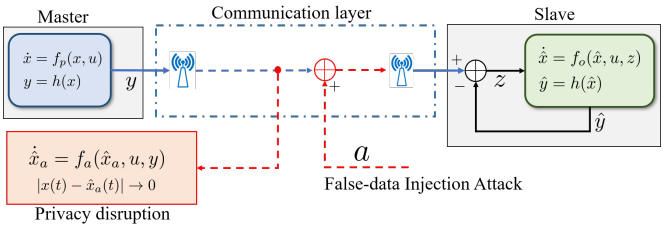

Fig. 1. Scheme of a privacy disruption and false-data injection attack attack.

plant and the observer. In the following, the plant is denoted as *master* while the observer as *slave*. Furthermore, we suppose that the slave has the form

$$\dot{\hat{x}} = f_o(\hat{x}, u, z), \qquad \hat{y} = h(\hat{x}), \tag{2a}$$

$$z = y - h(\hat{x}), \tag{2b}$$

where $\hat{x} \in \mathbb{R}^{n_x}$ is the state, f_o represents the slave dynamics satisfying $f_o(x,u,0) = f_p(x,u)$ for any $(x,u) \in \mathbb{R}^{n_x} \times U$, and $z \in \mathbb{R}^{n_y}$ represents the correction term which is selected, for the time being as in (2b). We suppose that the slave achieves a reliable estimation of the unknown master states, namely that the estimation error $\hat{x} - x$ exponentially converges to zero

$$|x(t) - \hat{x}(t)| \leqslant ke^{-\lambda t}|x(0) - \hat{x}(0)|, \qquad \forall\, t \geqslant 0,$$

for some $k, \lambda > 0$ and any initial condition $x(0), \hat{x}(0) \in \mathbb{R}^{n_x}$. Furthermore, we suppose that the master (1) and the slave (2b) satisfy the following technical assumptions that will be used in the sequel. In particular, we assume boundedness of the master trajectories and an incremental passivity assumption on the observer dynamics (2). Further comments on these assumptions are postponed to the end of Section III.

Assumption 1. *The states of the master* (1) *evolve in a compact set* X*, that is* $x(t) \in X \subset \mathbb{R}^{n_x}$ *for all* $t \geqslant 0$ *and* $u \in U$*.*

Assumption 2. *The observer dynamics* (2a) *is globally incrementally passive from* z *to* $\hat{y}$*, namely there exists a* C^1 *storage function* $V_o : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}_{\geqslant 0}$ *and class* $\mathcal{K}_\infty$ *functions* $\underline{\alpha}, \bar{\alpha}$ *satisfying*

$$\underline{\alpha}(|\hat{x} - \hat{x}'|) \leqslant V_o(\hat{x} - \hat{x}') \leqslant \bar{\alpha}(|\hat{x} - \hat{x}'|) \tag{3}$$

for any $\hat{x}, \hat{x}' \in \mathbb{R}^{n_x}$*, and*

$$\Big\langle \nabla V_o(\hat{x} - \hat{x}'), \big[f_o(\hat{x}, u, z), f_o(\hat{x}', u, z')\big] \Big\rangle \leqslant (h(\hat{x}) - h(\hat{x}'))^\top (z - z') \tag{4}$$

for any $\hat{x}, \hat{x}' \in \mathbb{R}^{n_x}$*,* $z, z' \in \mathbb{R}^{n_y}$ *and any* $u \in U$*.*

B. Objectives

The main objective of this article is securing the master-slave communication in order to protect the system in front of possible communication faults and/or malicious attack over the transmitted signal y. In particular, the goal is to secure the communication between agents to achieve simultaneously two system properties, as depicted in Fig. 1:

- *Privacy of the master:* Prevent an unwanted agent, denoted as *eavesdropper*, to estimate the master states by listening the transmitted signal, y. Specifically, prevent the possibility of designing an adversarial observer of the form
$$\dot{\hat{x}}_a = f_a(\hat{x}_a, u, y) \tag{5}$$
with exponential convergent properties
$$|x(t) - \hat{x}_a(t)| \leqslant ke^{-\lambda t}|x(0) - \hat{x}_a(0)|, \qquad \forall\, t \geqslant 0,$$
for some some constants $k, \lambda > 0$ and any initial condition $x(0), \hat{x}_a(0) \in \mathbb{R}^{n_x}$, by an eavesdropper that has the complete knowledge of the function f_p, the control input signal u, and can measure the signal y.
- *False-data injection attack detection*: Detect the injection of a signal a in the transmitted variable y,
$$y = h(x) + a \tag{6}$$
with a being a malicious signal.

A scheme of the proposed privacy and false-data injection attack problems is included in Fig. 1.

Finally, in order to widen the applicability of the proposed security mechanism, three additional restrictions have been imposed on the considered problem.

1) *Modularity constraint.* The structure of the master and the slave cannot be modified. This includes not modifying the control input u, nor the function f or the slave dynamics f_o, but we are only allowed to modify the input signal z of the observer dynamics (2a). This restriction imposes a *modular* design philosophy on our proposed solution, namely we assume that the slave algorithm has been already designed in the nominal case and we want to include an additional security layer
2) *Communication constraint.* The slave is only a receiver. This implies that the slave cannot send information back or communicate with the master. This additional communication would increase the cost of the overall architecture and allow the possibility of additional exogenous attacks. We also assume that no additional communication links can be implemented.
3) *Model-free constraint.* The proposed security mechanism design has to be independent of the functions f and f_o. This restriction ensures robustness to variations on the master and/or slave equations and, moreover, reduces the system knowledge required to implement the security mechanism.

III. Masking Protocol

A. Oscillator Masking Approach

In order to mask the signal y transmitted from the master (1) to the slave (2), we suppose to add to the output (1b) an extra signal which is rich enough so that to mask the nominal output trajectory $h(x)$. This is depicted in the left-hand side

of Fig. 2. In particular, we suppose that y in (1b) is now given by

$$y = h(x) + d \tag{7}$$

where d is a masking signal that can be thought of the form

$$d(t) = \sum_{i=1}^{N} d_i(t), \qquad d_i(t + T_i) = d_i(t), \quad \forall t \geqslant 0,$$

namely composed by a sum of N signals in which each d_i is T_i-periodic, and with periods T_i that are incommensurable reals, namely $\frac{T_i}{T_j}$ is an irrational number for any pair of i, j. The resulting signal d is quasi-periodic. In the reminder of the paper, we suppose then that the signal d can be thought of as generated by an autonomous system of the form

$$\dot{w} = \Phi w, \qquad d = \gamma(\Gamma w), \tag{8}$$

where $w \in \mathbb{R}^{n_w n_y}$ is the internal state of the masking generator and the matrices Φ, Γ are selected as follows

$$\Phi = \text{diag}(\underbrace{S, \ldots, S}_{n_y \text{ times}}), \quad \Gamma = \text{col}(\underbrace{G, \ldots, G}_{n_y \text{ times}}). \tag{9}$$

with the matrices S, G and the C^1 function $\gamma(\cdot) : \mathbb{R}^{n_y} \to \mathbb{R}^{n_y}$ satisfying the following assumption.

Assumption 3. *The matrix S is skew-symmetric, the pair (S, G) is observable and γ is monotonic, namely it satisfies*

$$(a - b)^\top [\gamma(a) - \gamma(b)] \geqslant (a - b)^\top (a - b)$$

for all $a, b \in \mathbb{R}^{n_y}$, $a \neq b$.

A possible choice of S, G is simply given by

$$\begin{aligned} S &= \text{blckdiag}(S_1, \ldots, S_m), & S_i &= \begin{pmatrix} 0 & \omega_i \\ -\omega_i & 0 \end{pmatrix}, \\ G &= \begin{bmatrix} G_1 & \ldots & G_m \end{bmatrix}, & G_i &= \begin{pmatrix} 1 & 0 \end{pmatrix}, \end{aligned} \tag{10}$$

where $\omega_i > 0$, $i = 1, \ldots, m$, correspond to the desired frequencies of the signal d. The function γ can be selected as a transcendental function. Consequently, the signal d may potentially contain an infinite number of harmonics.

The master dynamics (1a) with the masking generator (8) form now the next overall system

$$\begin{aligned} \dot{x}_p &= f_p(x, u), \qquad \dot{w} = \Phi w, \\ y &= h(x) + \gamma(\Gamma w). \end{aligned} \tag{11}$$

It is well known from standard system theory that if we want to preserve the ability of the slave to reconstruct the state x of the plant, an observability property on the overall dynamics (x, w) from the output y in (11) is needed. Such an assumption is stated in terms of detectability properties of the plant (11) as follows.

Assumption 4. *The extended system*

$$\begin{aligned} \dot{\eta} &= F(\eta, u) := \begin{pmatrix} f_o(x, u, z) \\ \Phi\xi + \Gamma^\top \xi \end{pmatrix}, \quad \eta = \begin{pmatrix} x \\ \xi \end{pmatrix} \\ \zeta &= H(\eta) := h(x) + \Gamma\xi \end{aligned} \tag{12}$$

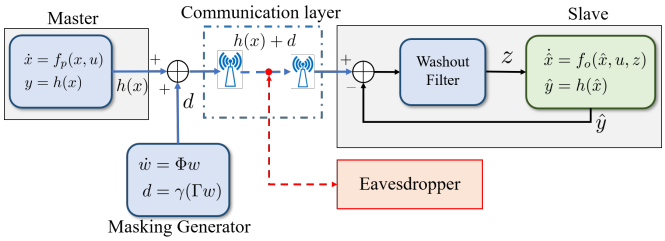

Fig. 2. Scheme of the proposed masking/de-masking protocol.

with state η and output ζ, is incrementally zero-state detectable for all $(\eta, u) \in \mathbb{R}^{n_x + n_\xi} \times \mathbb{R}^{n_u}$. That is, take any pair of solutions (η, η') of (12) and define $e := \eta - \eta'$. Then,

$$H(\eta) = H(\eta'), \quad \forall t \geqslant 0 \;\Rightarrow\; \lim_{t \to \infty} |e| = 0.$$

The oscillator masking model generator (namely the matrix S and the nonlinear function γ) is assumed to be known by the slave, but is unknown by the eavesdropper. By exploiting this knowledge, a washout filter based de-masking module can be included in the slave in order to remove the masking signal and allow the slave to reconstruct the master state. As the oscillator masking generator is unknown by the eavesdropper, reconstruction of the master state from the signal y from unwanted actors is not possible, so, the privacy problem is immediately solved, see Section IV. Furthermore, the states of the filter can be used to detect any additional signal a different from the oscillator masking one. This fact can be used to solve the false-data injection attack problem, as detailed later in Section V.

B. De-masking Module

The last part of the proposed architecture consists in a de-masking module based on the washout filters approach proposed in [17] to remove the masking signal in the slave. This washout fiter is depicted in the right-hand side of Fig. 2. Precisely, the washout filter takes the form

$$\dot{\xi} = \Phi\xi + \Gamma^\top z, \tag{13a}$$

$$z = y - h(\hat{x}) - \gamma(\Gamma\xi), \tag{13b}$$

where $\xi \in \mathbb{R}^{n_\xi}$ is the filter state, and the matrices $\Phi \in \mathbb{R}^{n_\xi \times n_\xi}$ and $\Gamma \in \mathbb{R}^{n_\xi \times n_y}$ are selected as in (9). The slave dynamics (2) is thus augmented by the dynamics (13a) and the signal z, redefined according (13b), drives the dynamics (2a). The overall scheme is therefore given by

$$\begin{aligned} \dot{\hat{x}} &= f_o(\hat{x}, u, z), \qquad \dot{\xi} = \Phi\xi + \Gamma^\top z \\ z &= y - h(\hat{x}) - \gamma(\Gamma\xi)\,. \end{aligned} \tag{14}$$

The zeros of the transfer function of the washout filter (13), considering $y - h(\hat{x})$ as the input and z as output, coincide with the eigenvalues of the masking generator (8). Therefore, the output disturbance, d, does not have any effect on the state estimation. Indeed, the masking is completely removed in the slave, if the overall filter-slave (14) possesses adequate convergence properties, as formalized in the next theorem.

Theorem 1. *Consider the master dynamics* (1a) *with masked output* (7) *where* d *is the masking signal generated by* (8)*. Moreover, consider the filter-slave system in* (14)*. If Assumptions 1-4 are satisfied, then, exponential estimation of the master state is achieved, namely*

$$|x(t) - \hat{x}(t)| \leqslant ke^{-\lambda t}|x(0) - \hat{x}(0)|, \qquad \forall\, t \geqslant 0 \tag{15}$$

for some $k, \lambda > 0$*, any* $x(0), \hat{x}(0) \in X$ *and* $u \in U$*.*

Proof. The overall system composed by the plant (1a) with output (7) with masking d generated by the dynamics (8), filter-slave dynamics (14) reads

$$\begin{aligned} \dot{x} &= f_p(x,u), & \dot{\hat{x}} &= f_o(\hat{x},u,z), \\ \dot{w} &= \Phi w, & \dot{\xi} &= \Phi\xi + \Gamma^\top z, \\ y &= h(x) + \gamma(\Gamma w), & z &= y - h(\hat{x}) - \gamma(\Gamma\xi). \end{aligned} \tag{16}$$

Let $\hat{x}_{ss}(t) = x(t)$ and $\xi_{ss}(t) = w(t)$ for all $t \geqslant 0$. By recalling that $f_p(x,u) = f_o(x,u,0)$ for any x, u, it can be verified that the pair $\hat{x}_{ss}, \xi_{ss}$ so defined is a steady-state solution of the overall system (16) because it can be verified that $z = 0$ for $\hat{x} = x$ and $\xi = w$.

Now, given any evolution of the plant dynamics (1a), (7), (8), consider any pair of solutions $\eta = (\hat{x}, \xi)$ and $\eta' = (\hat{x}', \xi')$ to (14) and define the error dynamics $e := \eta - \eta' = (\tilde{x}, \tilde{\xi})$. The e-dynamics evolves according to

$$\begin{aligned} \dot{\tilde{x}} &= f_o(\tilde{x} + \hat{x}', u, \tilde{z} + z') - f_o(\hat{x}', u, z'), \\ \dot{\tilde{\xi}} &= \Phi\tilde{\xi} + \Gamma^\top \tilde{z}, \qquad \tilde{z} = -\tilde{h}(\tilde{x}, \hat{x}') - \tilde{\gamma}(\tilde{\xi}, \xi'), \end{aligned}$$

where $\tilde{h}(\tilde{x}, \hat{x}') := h(\tilde{x} + \hat{x}') - h(\hat{x}')$ and $\tilde{\gamma}(\tilde{\xi}, \xi') := \gamma(\Gamma(\tilde{\xi} + \xi')) - \gamma(\Gamma\xi')$. Now, consider the Lyapunov function

$$V(e) = V_o(\hat{x} - \hat{x}') + \frac{1}{2}\tilde{\xi}^\top\tilde{\xi}$$

where the first factor, V_o, comes from Assumption 2. Using inequality (4) we compute the derivative of the term V_o as follows

$$\dot{V}_o \leqslant \tilde{h}(\tilde{x}, \hat{x}')^\top \tilde{z} \leqslant -\tilde{h}(\tilde{x}, \hat{x}')^\top \big[\tilde{h}(\tilde{x}, \hat{x}') + \tilde{\gamma}(\tilde{\xi}, \xi')\big].$$

Then, the derivative of the second term V_ξ satisfies

$$\dot{V}_\xi = \tilde{\xi}^\top\Phi\tilde{\xi} + \tilde{\xi}^\top\Gamma^\top\tilde{z} \leqslant -\tilde{\xi}^\top\Gamma^\top\big[\tilde{h}(\tilde{x}, \hat{x}') + \tilde{\gamma}(\tilde{\xi}, \xi')\big]$$

where the second inequality is deduced from the skew-symmetric property of S. Combining the previous inequalities we compute the derivative of V as follows

$$\begin{aligned} \dot{V} &\leqslant -\big[\tilde{h}(\tilde{x}, \hat{x}')^\top + \tilde{\xi}^\top\Gamma^\top\big]\big[\tilde{h}(\tilde{x}, \hat{x}') + \tilde{\gamma}(\tilde{\xi}, \xi')\big] \\ &\leqslant -|H(\eta) - H(\eta')|^2 \end{aligned}$$

where the second inequality is deduced from the monotonic property in Assumption 3 and H is defined in Assumption 4. By using (3) we deduce that V is an incrementally dissipative Lyapunov function for the $(\hat{x}, \xi)$-dynamics. Furthermore, according to Assumption 1, x, u evolve in compact sets. This and Assumption 2 imply that $\hat{x}, \xi$ too evolve in compact sets by the incremental passivity properties of the $(\hat{x}, \xi)$-dynamics. As a consequence, by the incremental LaSalle invariance theorem [21, Section VII], the system converges to the largest invariant set such that $H(\eta) = H(\eta')$. Thus, using the detectability property in Assumption 4, we conclude that the $(\hat{x}, \xi)$-dynamics is incrementally stable and satisfies

$$|\eta(t) - \eta'(t)| \leqslant ke^{-\lambda t}|\eta(0) - \eta'(0)|, \quad \forall\, t \geqslant 0.$$

Existence of a steady-state $(\hat{x}_{ss}, \xi_{ss})$ and the incremental stability property show that such a steady-state is also unique. As a consequence, recalling the definition of η, previous inequality leads to (15) when the initial condition of the filter (13) is selected as $\xi(0) = w(0)$. In case $\xi(0) \neq w(0)$ the bound (15) is modified as

$$|x(t) - \hat{x}(t)| \leqslant ke^{-\lambda t}|x(0) - \hat{x}(0)| + ke^{-\lambda t}|\xi(0) - w(0)|$$

for all $t \geqslant 0$. This concludes the proof. □

C. Discussion about the Assumptions

From a theoretical point of view, the whole architecture relies on the condition that the filter-observer interconnection in (14) presents an incremental stability property. To ensure this condition, this work exploits the fact that the washout filter (13) presents an incremental passivity property. Consequently, Assumption 2 is included to impose an incremental passivity condition on the observer, making the filter-observer interconnection passive as well. Finally, the compactness Assumption 1 and the detectability Assumption 4 are included to guarantee incremental stability of the interconnection.

From a practical viewpoint, the compactness condition in Assumption 1 is not a restrictive assumption, as physical systems are usually designed to evolve in bounded sets. Furthermore, Assumption 2 implies an incremental passivity property on the observer. Similarly to [18], following the so-called Demidovich conditions, in case of linear output maps, namely $h(x) = Cx$, the observer can be designed as $f_o(x,u,z) := f_p(x,u) + Lz$ provided the following condition hold

$$P\frac{\partial f_p}{\partial x}(x,u) + \frac{\partial f_p}{\partial x}(x,u)^\top P \leqslant 0, \qquad L = \mu P^{-1}C^\top, \tag{17}$$

for all $x \in \mathbb{R}^{n_x}$ and all $u \in \mathbb{R}^{n_u}$, for any constant $\mu > 0$.

Note that if the output is nonlinear and monotonic, a passive observer can still be designed. See, for instance, [22, Eq (9)] of [23]. Finally, Assumption 4 imposes a minimal observability condition on the system composed by the master and the masking generator. Without this assumption, it would not be possible to distinguish between the master states and the masking generator states. In many cases, this can be done for instance by selecting the frequencies of Φ much faster than the dynamics of the master. In the context of linear dynamics, namely $f_p(x,u) = Ax + Bu$, the detectability Assumption 4 reduces to ask A and Φ to have disjoint spectra.

IV. Privacy Protection

Theorem 1 shows that after adding the oscillator masking to the transmitted signal y via the signal d generated by (8), state estimation of the master x can be still obtained by using

the washout filter-observer architecture in (14). Since only the observer knows the oscillator masking generator model, the eavesdropper cannot implement a washout filter or similar and convergence of the state of the adversarial observer (5) to the one of the master (1a) (7) is impossible. At most, an adversarial observer can achieve an input-to-state stability property with respect to the oscillator masking, d, namely

$$|x(t) - \hat{x}_a(t)| \leqslant ke^{-\lambda t}|x(0) - \hat{x}_a(0)| + \sup_{s\in[0,t]} \rho(|d(s)|)$$

for some class $\mathcal{K}$ function ρ, some constants $k, \lambda > 0$ and $\forall t$ and any initial condition $\hat{x}_a(0) \in \mathbb{R}^{n_x}$. Therefore, oscillator masking signals with large values d will make any adversarial observer estimation practically useless.

As an illustration of this result consider as a master system the ball and beam system studied in [24] described as

$$\dot{x} = f_p(x,u) := \begin{pmatrix} x_2 \\ \frac{x_1 x_4^2 - g\sin x_3}{\mathcal{J}_b/(M\mathcal{R}^2)+1} \\ x_4 \\ -\frac{x_1 x_2 x_4 + g x_2 \cos x_3}{x_1^2 + \mathcal{J}/M + \mathcal{J}_b/M} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} u \qquad (18)$$

$$y = Cx := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} x$$

where $x = (x_1, \ldots, x_4) \in \mathbb{R}^4$ is the state vector, $y = (y_1, y_2) \in \mathbb{R}^2$, and the input is designed as

$$u = \frac{2x_1x_2x_4 + gx_1\cos x_3}{x_1^2 + \mathcal{J}/M + \mathcal{J}_b/M} + \frac{24(\mathcal{J}_b/\mathcal{R}^2 + M)x_1}{Mg} + \frac{50(\mathcal{J}_b/R^2 + M)x_2}{Mg} - 35x_3 - 10x_4 + \sin t.$$

To make the simulation more realistic, random white noise of variance 0.0005 and 0.00001, have been added to the outputs y_1, y_2, respectively. The parameters of the model are taken from [25]: $\mathcal{J} = 0.02$, $M = 0.05$, $\mathcal{J}_b = 2 \times 10^{-6}$, $\mathcal{R} = 0.01$, $g = 9.81$. According to [24], the considered model satisfies the dissipativity condition in (17) with

$$P = \begin{bmatrix} 3.6087 & 2.1779 & -4.2801 & -0.292 \\ 2.1779 & 3.1127 & -6.9958 & -0.4454 \\ -4.2801 & -6.9958 & 25.5017 & 1.4720 \\ -0.292 & -0.4454 & 1.4720 & 0.3972 \end{bmatrix}$$

for $|x_1| \leqslant 3$, $x_2 \in \mathbb{R}$, $|x_3| \leqslant 0.65$ and $x_4 \leqslant 0.19$. Thus, it is possible to design an observer for the system as $f_o(\hat{x}, u, z) = f_p(\hat{x}, u) + Lz$ by selecting the gain L as in (17). To secure the communication between the master and the slave the oscillator masking generator (8) is implemented as in (10) with $m = 3$ oscillators, $\omega_1 = 20$, $\omega_2 = 20\sqrt{2}$ and $\omega_3 = 20\sqrt{3}$, and with the function γ as $\gamma(s) = (\bar{\gamma}(s_1), \bar{\gamma}(s_2))$, for any $s = (s_1, s_2) \in \mathbb{R}^2$ and $\bar{\gamma}$ defined as $\bar{\gamma}(\bar{s}) = 3 + 5\,\mathrm{atan}(1/3\bar{s}^3 - 1/2\bar{s}^2 + \bar{s})$ for any $\bar{s} \in \mathbb{R}$. The function $\gamma(\cdot)$ so selected is monotonic, that is, satisfies Assumption 3. In simulations, the initial conditions of (8) are taken as $w(0) = (5, 0, 5, 0, 5, 0)$. The true output signal, y, and its masked version are depicted in Fig. 3.

We that an adversarial observer of the form (5) has intruded the system and is trying to break the privacy of the system with dynamics in (5) described as

$$f_a(\hat{x}_a, u, y) = f(\hat{x}_a, u) + P^{-1}C^\top(y + d - h(x_a)).$$

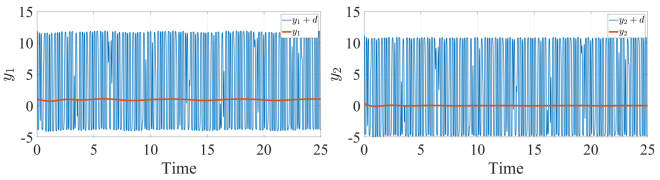

Fig. 3. Evolution of the system output y and the masked signal $y + d$.

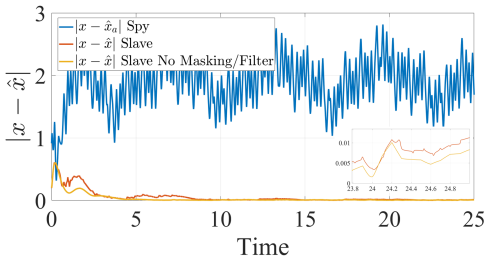

Fig. 4. Evolution of the adversarial observer (Spy) error norm, $|x - \hat{x}_a|$, and the error norm of the slave with the washout filter, $|x - \hat{x}|$ and the error norm of the same slave without the masking/filter architecture.

We also assume that the slave implements the washout filter-slave architecture (14) with the matrices Φ, N as in (9). The evolution of the norm of the state estimation error, $|x - \hat{x}|$, of the secured slave and the adversarial observer are depicted in Fig. 4. The evolution of the slave without the masking/filter architecture is also depicted in the same figure. It can be seen that the estimation error of the adversarial observer is order of magnitudes larger than the one from the slave and practically useless. Thus, privacy of the master is preserved. Moreover, it can be seen that the slave with the masking/filter security layer presents a similar transient and converges to a similar error than the same slave without the security layer.

V. Detection of False-Data Injection Attacks

Additionally, the washout filter (13) can also be used to detect any deviation of the system from the expected steady-state behaviour. Indeed, once the system converges to the steady-state trajectories $\xi_{ss} = w$, the filter reduces to an autonomous system of the form $\dot{\xi} = \Phi\xi$.

The principle we are going to follow is indeed to detect whether or not the washout filter has reached such a steady state. To this end, since the state of the washout filter ξ is part of our architecture, an observer can be trivially designed as

$$\dot{\hat{\xi}} = \Phi\hat{\xi} + \kappa(\xi - \hat{\xi}), \qquad (19)$$

where $\hat{\xi} \in \mathbb{R}^{n_\xi}$ is the state of the filter observer and $\kappa > 0$ is a positive gain. Note that the matrix $\Phi - \kappa I$ is Hurwitz for any $\kappa > 0$. As a consequence, it is easy to show that if ξ is at steady-state, i.e. $\xi_{ss} = w$, then the observer (19) satisfies

$$\lim_{t\to\infty} |\xi(t) - \hat{\xi}(t)| = 0. \qquad (20)$$

Hence, the value $|\xi - \hat{\xi}|$ can be used as a metric to detect any malfunction of the communication link.

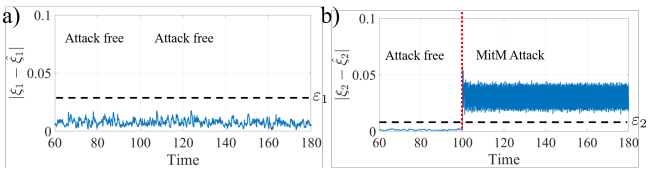

Fig. 5. a) Evolution of the norm $|\xi_1 - \hat{\xi}_1|$ of the first 6 states of ξ. b) Evolution of the norm $|\xi_1 - \hat{\xi}_1|$ of the last 6 states of ξ. At 100 seconds there is a man-in-the middle (MitM) attack in the transmitted signal, y_2.

Moreover, divide the filter state as $\xi = (\xi_1, \ldots, \xi_{n_y})$ with $\xi_i \in \mathbb{R}^{n_w}$ for all $i = 1, \ldots, n_y$, and divide the output as $y = (y_1, \ldots, y_{n_y})$. Then, a set of metrics can be designed, $|\xi_1 - \hat{\xi}_1|, \ldots, |\xi_{n_y} - \hat{\xi}_{n_y}|$, associated to each component of the output, $y_1, \ldots, y_{n_y}$, respectively. Finally, the presence of noise in the transmitted signal will prevent the metric $|\xi - \hat{\xi}|$ to be zero. This fact has to be taken into account in order to design the detection mechanism. In particular, it is possible to define a set of threshold variables ε_i for $i = 1, \ldots, n_y$ and define the following detection mechanism

$$\begin{cases} \text{Attack } a_i \text{ at } y_i, & \text{if } |\xi_i - \hat{\xi}_i| > \varepsilon_i \\ \text{No attack } a_i \text{ at } y_i, & \text{otherwise.} \end{cases} \tag{21}$$

As an illustration of this result consider exactly the previous example of Section IV. We select the same set of master, slave and washout filter dynamics, the same initial conditions and the same type of measurement noise. Moreover, we add the observer in (19) in order to compute the security metrics $|\xi_1 - \hat{\xi}_1|$ and $|\xi_2 - \hat{\xi}_2|$. Now, we simulate a false-data injection attack as in (6) by selecting $a = (a_1, a_2)$ with $a_1(t) = 0$ for all $t \geqslant 0$ and $a_2(t) = 0$ for $t \in [0, 100]$ and $a_2(t) = 0.3\sin(10t)$ for $t \geqslant 100$. In order to detect this attack, the detection mechanism in (19), (21) is implemented with $\varepsilon_1 = 0.025$ and $\varepsilon_2 = 0.01$. The evolution of the security metrics $|\xi_1 - \hat{\xi}_1|$ and $|\xi_2 - \hat{\xi}_2|$ are depicted in Fig. 5. It can be seen that in the absence of an attack, both metrics remain below the defined thresholds. Nonetheless, during the attack, the metric, $|\xi_2 - \hat{\xi}_2|$, related to the second output, y_2, increases and activates the security mechanism.

VI. Conclusions

This work has presented a masking protocol in a plant/observer system to preserve the privacy of the plant while allowing the detection of false-data injection attacks in the communication link. The proposed protocol can be implemented in any observer satisfying an incremental passivity property. The protocol doesn't require any knowledge of the plant or the observer dynamics and the protocol doesn't need any additional communication links between the plant and the observer. Future works will focus on extending the proposed framework to secure dynamic controllers and multi-agent systems. Moreover, we will explore the idea of using multiplicative noise instead of additive one.

References

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

[2] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[3] A. J. Gallo, M. S. Turan, F. Boem, T. Parisini, and G. Ferrari-Trecate, "A distributed cyber-attack detection scheme with application to dc microgrids," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3800–3815, 2020.

[4] T. Yang, C. Murguia, C. Lv, D. Nešić, and C. Huang, "On joint reconstruction of state and input-output injection attacks for nonlinear systems," *IEEE Control Systems Letters*, vol. 6, pp. 554–559, 2022.

[5] T. Yang, C. Murguia, M. Kuijper, and D. Nešić, "An unknown input multi-observer approach for estimation and control under adversarial attacks," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 475–486, 2021.

[6] C. Murguia, I. Shames, J. Ruths, and D. Nešić, "Security metrics and synthesis of secure control systems," *Automatica*, vol. 115, p. 108757, 2020.

[7] C. Murguia, I. Shames, F. Farokhi, D. Nešić, and H. V. Poor, "On privacy of dynamical systems: An optimal probabilistic mapping approach," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2608–2620, 2021.

[8] D. Umsonst and H. Sandberg, "On the confidentiality of controller states under sensor attacks," *Automatica*, vol. 123, p. 109329, 2021.

[9] Y. Kawano and M. Cao, "Design of privacy-preserving dynamic controllers," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3863–3878, 2020.

[10] J. Kim, D. Kim, Y. Song, H. Shim, H. Sandberg, and K. H. Johansson, "Comparison of encrypted control approaches and tutorial on dynamic systems using learning with errors-based homomorphic encryption," *Annual Reviews in Control*, vol. 54, pp. 200–218, 2022.

[11] P. Bernard, V. Andrieu, and D. Astolfi, "Observer design for continuous-time dynamical systems," *Annual Reviews in Control*, vol. 53, pp. 224–248, 2022.

[12] K. Ding, X. Ren, A. S. Leong, D. E. Quevedo, and L. Shi, "Remote state estimation in the presence of an active eavesdropper," *IEEE Transactions on Automatic Control*, vol. 66, no. 1, pp. 229–244, 2021.

[13] L. Wang, X. Cao, H. Zhang, C. Sun, and W. X. Zheng, "Transmission scheduling for privacy-optimal encryption against eavesdropping attacks on remote state estimation," *Automatica*, vol. 137, p. 110145, 2022.

[14] M. Wiese, T. J. Oechtering, K. H. Johansson, P. Papadimitratos, H. Sandberg, and M. Skoglund, "Secure estimation and zero-error secrecy capacity," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1047–1062, 2019.

[15] A. Tsiamis, K. Gatsis, and G. J. Pappas, "State-secrecy codes for stable systems," in *American Control Conference*, 2018, pp. 171–177.

[16] A. S. Leong, D. E. Quevedo, and S. Dey, "State estimation over markovian packet dropping links in the presence of an eavesdropper," in *IEEE 56th Conference on Decision and Control*, 2017, pp. 6616–6621.

[17] L. Wang, L. Marconi, C. Wen, and H. Su, "Pre-processing nonlinear output regulation with non-vanishing measurements," *Automatica*, vol. 111, p. 108616, 2020.

[18] A. Pavlov and L. Marconi, "Incremental passivity and output regulation," *Systems & Control Letters*, vol. 57, no. 5, pp. 400–409, 2008.

[19] Ö. Morgül and M. Feki, "A chaotic masking scheme by using synchronized chaotic systems," *Physics Letters A*, vol. 251, no. 3, pp. 169–176, 1999.

[20] C. Murguia, I. Shames, F. Farokhi, and D. Nešić, "Information-theoretic privacy through chaos synchronization and optimal additive noise," *Privacy in Dynamical Systems*, pp. 103–129, 2020.

[21] F. Forni and R. Sepulchre, "A differential lyapunov framework for contraction analysis," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 614–628, 2013.

[22] A. Pavlov, E. Steur, and N. van de Wouw, "Nonlinear integral coupling for synchronization in networks of nonlinear systems," *Automatica*, vol. 140, p. 110202, 2022.

[23] J. Lei and H. K. Khalil, "High-gain observers in the presence of sensor nonlinearities," in *American control conference*, 2017, pp. 3282–3287.

[24] C. Wu, A. van der Schaft, and J. Chen, "Robust trajectory tracking for incrementally passive nonlinear systems," *Automatica*, vol. 107, pp. 595–599, 2019.

[25] J. Hauser, S. Sastry, and P. Kokotovic, "Nonlinear control via approximate input-output linearization: the ball and beam example," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 392–398, 1992.