A robust protection scheme against cyber-physical attacks in power systems

A. Di Giorgio, A. Pietrabissa, F. Delli Priscoli, A. Isidori

Department of Computer, Control and Management Engineering 'A. Ruberti', University of Rome La Sapienza, Via Ariosto 25, 00185 Rome, Italy
Corresponding author email: digiorgio@diag.uniroma1.it

Abstract

This paper presents a robust defence strategy in reaction to destabilizing cyber-physical attacks launched against linear time invariant systems and its application to power systems. The proposed protection scheme aims at making the dynamics of a selected subsystem decoupled from the dynamics of the subsystem targeted by the attack. The standard decoupling methods are made robust, in spite of poor information about plant parameters and lack of state measurement, with the aid of an extended observer. In this way it is possible to keep the protected dynamics arbitrarily close to the one of a suitably chosen stable system, so long as the dynamics being targeted by the attack remain within prescribed bounds. The proposed defence strategy is presented in the context of modern power systems, wherein generators and transmission network are operated by different players, and shown to be effective using the Western System Coordinating Council 9-bus test power network.

1 Introduction

Over the last two decades the cyber physical systems have increasingly attracted the attention of academics and industry, posing new opportunities and challenges. Indeed, nowadays the integration of information and communication technology into physical processes is commonly recognized as fundamental to improve the operation of industrial plants and large scale infrastructures dedicated to the production of goods and the provisioning of primary importance services. Despite of this, the new paradigm increases the vulnerability of this class of systems, which are exposed to attacks leveraging the potential access to operational data in order to alter the behaviour of the underlying physical process. Well-known examples are the Maroochy Water breach [22] in 2000, the SQL Slammer worm attack on the Ohio nuclear plant network [11] in 2002, the coordinated attack on the Ukrainian power grid [23] in 2016.

The cyber-physical system security topic is receiving a lot of attention within the control theory community, as witnessed by the growing number of papers and special issues in relevant journals, e.g., [20][4] and references therein. Several classes of attack design and detection have been identified and studied, among the others: *deception* and *denial of service* attacks [2][6], *replay* attacks [16][28], *false data injection* attacks [13][14], *random* and *constant bias* attacks [12], and *zero dynamics* attacks [24][25][19].

In particular, it has been observed that systems having unstable zero dynamics are vulnerable to *stealthy attacks*. In fact, as shown e.g. in [17], in a system whose zero dynamics are unstable, with an (output feedback) control chosen so as to guarantee asymptotic stability in the absence of attacks, an attack generator may inject signals that make the internal state diverge, while the effects of such attack are not visible from the mere observation of the output (on the measure of which the stabilizing control is designed). An attack of this kind is commonly referred to as a *zero-dynamics attack*. Recent researches have focused on the *design* of a zero dynamics attack as well as on the *detection* of (or *defence* from) such an attack. In particular, [17] shows how a zero dynamics attack can be implemented that is robust in spite of model uncertainties, by means of a technique reposing on the design of a robust disturbance observer [21]. Researches on the detection of zero dynamics attacks are based on the design of centralized and decentralized observers [19], Kalman filtering [9], adaptive sliding mode observers [3], or by suitably altering the input behaviour of the process [7].

In this paper we focus on the design of *defence* strategies and we consider a slightly different scenario. Specifically we address the case in which the purpose of the attacker is to influence a portion of the dynamics of the plant (for instance in such a way that the zero dynamics associated with a selected output become unstable, so as to make a zero-dynamics attack possible) in a malicious way. The defence strategy is based on the (indeed elementary) idea of making the portion of the dynamics affected by the attack *decoupled* from the portion of the dynamics that needs to be defended. In this context, though, the standard decoupling methods are of no use because relying upon exact cancelation of coupling terms and availability of a measure of the entire state of the plant. Instead, we propose a design technique by means of which the result in question is achieved, robustly, in "practical terms", over a finite time horizon. The method in question basically reposes on some fundamental results of [5], in which a *high-gain extended observer* is employed, to the purpose of obtaining a robust "proxy" of a control law based on exact cancelation.

Among the application fields of interest, power systems are typical cyber physical systems, characterized by lack of information, asking for enhanced defence schemes. It is well known that power system dynamics results from the interconnection of synchronous machines, whose electromechanical behaviour is controlled by local prime mover governors. A fundamental role of the governor is the one of keeping the machine angular speed constant, against the oscillations of the electrical torque, by acting on the mechanical torque applied to the rotor. A malicious intervention on this control has the effect of inducing oscillations on the other machines in the network through the interconnections. In this context, a zero dynamics attack can be seen as the action of altering the mechanical torque of a properly selected set of machines, in order to induce instability in some machines without having an impact on some others. Conversely network decoupling can be achieved via feedback in order to allow a power plants operator to protect its machines.

Both the destabilization and the exact decoupling require a significant amount of information about the network model and full information about the rotor angle and the angular speed of machines. Despite the availability of measurements does not constitute a problem from the technological point of view in modern power systems, all the above information together are typically not available in practice. Indeed, following the unbundling of the electricity systems and the establishment of electricity market in most industrialized countries, the transmission network and the power plants have started to be operated by different operators, which typically share a limited amount of data about their infrastructures. Then a requirement for applying both the nominal attack and defence controls is the access to information owned by different players.

In the light of the above, this paper presents a robust defence strategy to attacks launched against linear time invariant systems and its application to power systems.

The paper is organized as follows. Section 2 recalls the model of a power system, describes the attack scenario and clarifies the requirements for a successful defence. Section 3 presents the attack model. Section 4 presents the robust decoupling control at the basis of the defence. In section 5 the proposed strategy is applied to a test power network in order to show the potential of the proposed defence strategy. Finally section 6 is dedicated to the concluding remarks.

2 Reference scenario

2.1 Power system model

In this section the power system model is recalled. The generic network here considered is constituted by $\bar{m}$ power plants and q load buses. It is well known that the electromechanical dynamics of a power system results from the composition of second order swing equations, through the nonlinear algebraic power flow equations [10][15]. As reported in [18], under the assumptions of lossless network, small angular differences and small deviations of bus voltages from rated values, the power system dynamics is described by the following time invariant linear descriptor model

$$\begin{pmatrix} I & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \dot{\delta} \\ \dot{\omega} \\ \dot{\theta} \end{pmatrix} = - \begin{pmatrix} 0 & -I & 0 \\ L_{\mathrm{gg}} & D & L_{\mathrm{g}\ell} \\ L_{\ell\mathrm{g}} & 0 & L_{\ell\ell} \end{pmatrix} \begin{pmatrix} \delta \\ \omega \\ \theta \end{pmatrix} + \begin{pmatrix} 0 \\ P_{\mathrm{g}} \\ P_{\ell} \end{pmatrix} \tag{1}$$

In (1) $\delta = \mathrm{col}(\delta_1, \delta_2, \dots, \delta_{\bar{m}})$ and $\omega = \mathrm{col}(\omega_1, \omega_2, \dots, \omega_{\bar{m}})$ denote the vectors of machines rotor angles and angular speeds, $\theta = \mathrm{col}(\theta_1, \theta_2, \dots, \theta_q)$ denotes the vector of load angles at load buses, $P_{\mathrm{g}} = \mathrm{col}(P_{\mathrm{g}1}, P_{\mathrm{g}2}, \dots, P_{\mathrm{g}\bar{m}})$ and $P_\ell = \mathrm{col}(P_{\ell 1}, P_{\ell 2}, \dots, P_{\ell q})$ are the vectors of mechanical input powers at generator buses and electrical powers at load buses, the matrices $M = \mathrm{diag}(M_1, \dots, M_{\bar{m}})$ and $D = \mathrm{diag}(D_1, \dots, \ , D_{\bar{m}})$ model the machines inertia and damping coefficients; finally L_{gg}, $L_{\mathrm{g}\ell}$, $L_{\ell\mathrm{g}}$ and $L_{\ell\ell}$ are properly sized submatrices of the network laplacian matrix

$$L_{\mathrm{N}} = \begin{pmatrix} L_{\mathrm{gg}} & L_{\mathrm{g}\ell} \\ L_{\ell\mathrm{g}} & L_{\ell\ell} \end{pmatrix} \tag{2}$$

where L_{gg} is diagonal and $L_{\ell\ell}$ is invertible. The model can be further simplified by explicity calculating θ from the third component of (1) as

$$\theta = -L_{\ell\ell}^{-1}[L_{\ell\mathrm{g}}\delta + P_\ell] \tag{3}$$

and substituting it into the angular speed dynamics to obtain

$$\begin{aligned} \begin{pmatrix} \dot{\delta} \\ \dot{\omega} \end{pmatrix} = &\begin{pmatrix} 0 & I \\ M^{-1}(-L_{\mathrm{gg}} + L_{\mathrm{g}\ell} L_{\ell\ell}^{-1} L_{\ell\mathrm{g}}) & -M^{-1}D \end{pmatrix} \begin{pmatrix} \delta \\ \omega \end{pmatrix} + \\ &+ \begin{pmatrix} 0 \\ M^{-1} \end{pmatrix} P_{\mathrm{g}} + \begin{pmatrix} 0 \\ M^{-1} L_{\mathrm{g}\ell} L_{\ell\ell}^{-1} \end{pmatrix} P_\ell \end{aligned} \tag{4}$$

Without loss of generality, it is possible to put $P_\ell = 0$, meaning that in what follows P_{g} will be intended as the deviation of the mechanical power from the value allowing to sustain a given loading condition during normal operation.

The resulting model has the standard linear time invariant form

$$\dot{x} \quad = \quad Ax + \bar{B}\bar{u} \tag{5}$$

where $x = \mathrm{col}(\delta, \omega) \in \mathbb{R}^{2\bar{m}}$, $\bar{u} = P_{\mathrm{g}} \in \mathbb{R}^{\bar{m}}$.

2.2 Attack scenario and defence requirements

In this section the attack scenario under investigation is described, raising the requirements for the defence design. In the reference scenario the $\bar{m}$ power plants are divided into three groups:

- a set of m_{a} power plants under the control of an *attacker*, able to alter the mechanical power input in order to induce instability in the rotor angle and angular speed dynamics of the other power plants in the network;
- a set of m_{p} power plants to be *protected* by a *defender* against the oscillations induced by the machines controlled by the attacker;
- a set of m_{u} *unprotected* power plants, which are exposed to the effect ot the attack.

As a result of this classification, the input to model (5) is partitioned as $\bar{u} = \mathrm{col}(u_\mathrm{a}, u_\mathrm{p}, u_\mathrm{u})$ where $u_\mathrm{a} \in \mathbb{R}^{m_\mathrm{a}}$ is the input available to the attacker, $u_\mathrm{p} \in \mathbb{R}^{m_\mathrm{p}}$ is the input of the protected machines and $u_\mathrm{u} \in \mathbb{R}^{m_\mathrm{u}}$ is the input of unprotected machines, the latter assumed not active in what follows.

This scenario is sufficiently general to cover some situations of practical interest; in this paper we take the perspective of a generation company operating m_p power plants and interested in protecting them from the spread of the instability occurring in other machines through network interconnections; the alteration of dynamics is supposed to be induced by the action of an *hacker* which, taking advantage of the vulnerabilty of the ICT infrastructure of a separate set of m_a power plants, uses their actuators to inject destabilizing signals out of the respective generation company's will.

The defender is supposed to have control on a limited number m_p of machines. Also, as a consequence of the power system industry unbundling, the defender is supposed to not have access to the state of the power plants which are not under its control (being operated by other generation companies) and to not know the laplacian matrix characterizing the connections in the network (which is an information owned by the transmission system operator); additionally it is assumed to have uncertain knowledge about the inertia and damping of its own machines. Notice that, in an attack scenario, even though the state of the attacked machines could be made available to the defender, such measurements should be considered unreliable, as coming from power plants under the influence of the attack.

In the light of the above, the fundamental requirements of the control strategy aimed at protecting the dynamics of interest are the following:

- the purpose of the defence control u_p is to *decouple* the dynamics (rotor and angular speed) of protected machines from the dynamics of the other machines operating in the network;

- the decoupling has to be *robust*, meaning that it has to be achieved without relying on the knowledge of machines state and network parameters.

To this purpose, a *protected* output $y_\mathrm{p} \in \mathbb{R}^{m_\mathrm{p}}$ is defined as the vector of protected machines' rotor angles; in this regard notice that the protection of rotor angles dynamics implies the one of angular speed dynamics, being the latter variable defined as the time derivative of the former.

In order to use standard notation, in what follows the subscript p will be omitted when referring to the decoupling control and the protected output, which will be denoted simply as $u \in \mathbb{R}^m$ and $y \in \mathbb{R}^m$.

3 An Attack Model

In what follows, we consider a system modeled by equations of the form

$$\begin{aligned} \dot{x} &= Ax + B_\mathrm{p}u + B_\mathrm{a}u_\mathrm{a} \\ y &= C_\mathrm{p}x \end{aligned} \tag{6}$$

with state $x \in \mathbb{R}^n$, control $u \in \mathbb{R}^m$, output $y \in \mathbb{R}^m$, in which the input u_a plays the role of an exogenous *attacker*. We focus our attention on the case in which the purpose of the attacker is to perturb the dynamics of the system and we describe how the input u can be designed so as to counter, in a sense that will be specified, the effects of such attack on the behavior of the *protected* output y.

To simplify matters, we consider hereafter the case in which

$$\begin{aligned} C_\mathrm{p} B_\mathrm{p} = C_\mathrm{p} A B_\mathrm{p} = \cdots = C_\mathrm{p} A^{r-2} B_\mathrm{p} = 0 \\ C_\mathrm{p} A^{r-1} B_\mathrm{p} = \mathrm{diag}(b_1, b_2, \ldots, b_m) \end{aligned} \tag{7}$$

where $b_i \neq 0$ for $i = 1, \ldots, m$, i.e. the case in which system (6), viewed by the *defender* as a system with input u and output y, has vector relative degree $\{r, r, \ldots, r\}$ and a purely diagonal "high-frequency gain matrix". This, in fact, is the case for the specific class of systems discussed in the previous section, consisting of the interconnection of a set of identical sub-systems, all of them independently actuated. However, we stress that without much complications one could as well address the more general case in which system (6) has vector relative degree $\{r_1, r_2, \ldots, r_m\}$ or even does not have a vector relative degree but has an invertible transfer function matrix $T(s) = C_\mathrm{p}(sI - A)^{-1} B_\mathrm{p}$.

It is also assumed

$$\begin{pmatrix} C_\mathrm{p} \\ C_\mathrm{p} A \\ \cdots \\ C_\mathrm{p} A^{r-1} \end{pmatrix} B_\mathrm{a} = 0\,. \tag{8}$$

which is yet another feature of the class of systems considered in the previous section.

As it is well known, under these assumptions system (6) can be – by means of a suitable change of coordinates – expressed in normal form as

$$\begin{aligned} \dot z_0 &= F_0 z_0 + G_0 \xi + G_{0,\mathrm{a}} u_\mathrm{a} \\ \dot \xi_i &= A_i \xi_i + B_i (H_{i,0} z_0 + K_i \xi + b_i u_i) \\ y_i &= C_i \xi_i \qquad\qquad i = 1, \ldots, m \end{aligned} \tag{9}$$

in which

$$A_i = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad B_i = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix},$$

$$C_i = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$$

and $\xi = \mathrm{col}(\xi_1, \ldots, \xi_m)$, $\dim(z_0) = n - mr$, $\dim(\xi_i) = r$.

Remark. In geometric terms, the previous setup can be characterized as follows (see [27, pp.87-90] and [27, pp.104-113] for definitions and basic properties related to the concepts of (A, B)-*invariant* subspace and of *controllability* subspace). Set $\bar{B} = \begin{pmatrix} B_{\rm p} & B_{\rm a} \end{pmatrix}$. If (7) holds, then $\mathcal{V}^*$, the largest $(A, \bar{B})$-*invariant* subspace contained in $\mathrm{Ker}(C_{\rm p})$, is given by

$$\mathcal{V}^* = \mathrm{Ker} \begin{pmatrix} C_{\rm p} \\ C_{\rm p} A \\ \cdots \\ C_{\rm p} A^{r-1} \end{pmatrix}.$$

In the coordinates of (9)

$$\mathcal{V}^* = \{(z, \xi) : \xi = 0\},$$

and (8) is equivalent to

$$\mathrm{Im}(B_{\rm a}) \subset \mathcal{V}^* .$$

Moreover $\mathcal{R}^*$, the largest *controllability subspace* of $(A, \bar{B})$ contained in $\mathrm{Ker}(C_{\rm p})$, can be identified with the reachable set of the pair $(F_0, G_{0,\rm a})$. If such pair is controllable, then $\mathcal{R}^* = \mathcal{V}^*$. ◃

The attacker can perturb the dynamics of (6) in various ways, depending on the information available. For instance, if the pair $(F_0, G_{0,\rm a})$ is controllable and z_0 is available for measurement, the attacker $u_{\rm a}$ can choose the strategy

$$u_{\rm a} = K_{\rm a} z_0 \tag{10}$$

so as to assign eigenvalues with positive real parts to the matrix $(F_0 + G_{0,\rm a} K_{\rm a})$. The effect of such attack is that of forcing, on the resulting system with input u and output y, an *antistable zero dynamics*.

This strategy presumes the availability of z_0 as well as an accurate knowledge of F_0 and $G_{0,\rm a}$. If this is not the case, an equivalent result could be obtained by means of a dynamic control law

$$\begin{array}{rcl} \dot{\zeta} & = & \tilde{A}_{\rm a} \zeta + \tilde{B}_{\rm a} y_{\rm a} \\ u_{\rm a} & = & \tilde{C}_{\rm a} \zeta + \tilde{D}_{\rm a} y_{\rm a} \end{array} \tag{11}$$

driven by a set of auxiliary measurements $y_{\rm a} = M_0 z_0 + N\xi$, so long as the system

$$\begin{array}{rcl} \dot{z}_0 & = & F_0 z_0 + G_{0,\rm a}(\tilde{C}_{\rm a} \zeta + \tilde{D}_{\rm a} M_0 z_0) \\ \dot{\zeta} & = & \tilde{A}_{\rm a} \zeta + \tilde{B}_{\rm a}(\tilde{C}_{\rm a} \zeta + \tilde{D}_{\rm a} M_0 z_0) \end{array}$$

can be rendered antistable.

Simple manipulations show that if the attack strategy is chosen as in (11), a system of the form

$$\begin{aligned} \dot{z} &= Fz + G\xi \\ \dot{\xi}_i &= A_i \xi_i + B_i (H_i z + K_i \xi + b_i u_i) \\ y_i &= C_i \xi_i \qquad\qquad i = 1, \ldots, m \end{aligned} \tag{12}$$

is obtained, where $z = \mathrm{col}(\zeta, z_0)$, and the matrix F is *anti-stable*.

While the specific target of the attack strategies (10) and (11) are the zero dynamics associated with the protected output y, it should be stressed that – because of the inherent coupling between the z's and the ξ's (which reflects, in the present context, the coupling between the protected, unprotected and attacked power plants of (4)) – any malicious signal u_a deliberately injected by the attacker might have a serious adverse effect on the behavior of the protected output y.

A simple strategy meant to counter the effects of an attack is indeed that of making the protected output y *decoupled* from u_a. As it is well-known, this is achieved if u is chosen as

$$u = -B^{-1}Hz + v\,, \tag{13}$$

in which, for convenience, we have set

$$B = \mathrm{diag}(b_1, b_2, \ldots, b_m), \qquad H = \begin{pmatrix} H_1 \\ \cdots \\ H_m \end{pmatrix}.$$

The control (13) renders the behavior of ξ, and consequently that of the protected output y, decoupled from z and hence unaffected by the attack. In fact, such control renders the state z unobservable through the output y. Moreover, one could pick the residual control v in (13) in such a way as to force a prescribed behavior of the ξ_i's. Setting, for convenience,

$$K = \begin{pmatrix} K_1 \\ K_2 \\ \cdots \\ K_m \end{pmatrix} \qquad K_0 = \begin{pmatrix} K_{01} & 0 & \cdots & 0 \\ 0 & K_{02} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & K_{0m} \end{pmatrix}$$

one could pick, to this end,

$$v = B^{-1}[-K\xi + K_0\xi] \tag{14}$$

so as to impose any prescribed (stable) dynamics on ξ. Note that the composition of (13) and (14) is a control of the form

$$u = B^{-1}[-Hz - K\xi + K_0\xi]\,. \tag{15}$$

The defence control (13), though, is not appropriate for various reasons. The main reason resides in its *lack of robustness*. In fact, the implementation of such control needs an *accurate knowledge* of B and H, as well as the availability of a *measurement* of z. Another reason is that, even if B and H were known and z were available for feedback, if the target of the attacker are the zero dynamics the state z will eventually diverge and so will the control u. Thus, the resulting control is sustainable only if implemented over a *finite interval of time*.

In what follows, we show that – from a practical viewpoint – the effect of decoupling y from z can be achieved by means of a *robust control law* that does not rely upon exact knowledge of B and H nor on availability of z, but rather appeals to techniques borrowed from the theory of the so-called *extended high-gain observer*, so long as the state z does not exceed a fixed (but that can be otherwise chosen arbitrarily large in the design stage) bound. In other words, we show that given any arbitrarily large number $\bar{M}$ and any arbitrarily small number $\bar{\varepsilon}$, it is possible to design a robust control law such that, so long as the norm of $z(t)$ does not exceed the bound $\bar{M}$, the behavior of $\xi(t)$ differs from the "ideal" behavior resulting from the implementation of the ("ideal" but not robust) control law (15) by a quantity that does not exceed $\bar{\varepsilon}$.

4 Robust Defence Against Destabilizing Attacks

4.1 The proposed defence strategy

We assume in what follows that all the b_i's are bounded from below and from above by fixed numbers, that is there exists numbers $0 < b_{\min} < b_{\max}$ such that

$$b_{\min} \leq |b_i| \leq b_{\max} \qquad \text{for all } i = 1, \dots, m.$$

If this is the case, one can find a number b_0 and a number $\delta_0 < 1$ such that

$$\left| \frac{b_i - b_0}{b_0} \right| \leq \delta_0 < 1 \qquad \text{for all } i = 1, \dots, m. \tag{16}$$

This number b_0 will be used in the design of the control.

With K_0 defined as above, let $\psi(\xi, \sigma)$ be the function defined as

$$\psi(\xi, \sigma) = B_0^{-1}[K_0\xi - \sigma]\,,$$

in which $\xi \in \mathbb{R}^{mr}$, $\sigma \in \mathbb{R}^m$ and $B_0 = \mathrm{diag}(b_0, b_0, \dots, b_0)$.

Let $g_L : \mathbb{R} \to \mathbb{R}$ be a smooth "saturation" function, that is a function characterized by the following properties:

- $g_L(s) = s$ if $|s| \leq L$,
- $g_L(s)$ is odd and monotonically increasing, with $0 < g_L'(s) \leq 1$,
- $\lim_{s\to\infty} g_L(s) = L(1 + c)$ with $0 < c \ll 1$.

With this in mind, define a function $G_L : \mathbb{R}^m \to \mathbb{R}^m$ as

$$G_L(s) = \mathrm{col}(g_L(s_1), \dots, g_L(s_m))$$

in which $g_L(\cdot)$ is a fixed saturation function.

System (12) will be controlled by a control law of the form

$$u = G_L(\psi(\hat{\xi}, \sigma)) = \begin{pmatrix} g_L(\psi_1(\hat{\xi}, \sigma)) \\ \cdots \\ g_L(\psi_m(\hat{\xi}, \sigma)) \end{pmatrix} \tag{17}$$

in which

$$\begin{aligned}\hat{\xi} &= \mathrm{col}(\hat{\xi}_1, \hat{\xi}_2, \ldots, \hat{\xi}_m)\\ \sigma &= \mathrm{col}(\sigma_1, \sigma_1, \ldots, \sigma_m)\end{aligned}$$

where, for $i = 1, 2, \ldots, m$, the vector

$$\hat{\xi}_i = \mathrm{col}(\hat{\xi}_{i,1}, \hat{\xi}_{i,2}, \ldots, \hat{\xi}_{i,r})$$

and the scalar σ_i are states of a dynamical system described by equations of the form

$$\begin{aligned}\dot{\hat{\xi}}_{i,1} &= \hat{\xi}_{i,2} + \kappa c_{i,r}(y_i - \hat{\xi}_{i,1})\\ \dot{\hat{\xi}}_{i,2} &= \hat{\xi}_{i,3} + \kappa^2 c_{i,r-1}(y_i - \hat{\xi}_{i,1})\\ &\cdots\\ \dot{\hat{\xi}}_{i,r-1} &= \hat{\xi}_{i,r} + \kappa^{r-1} c_{i,2}(y_i - \hat{\xi}_{i,1})\\ \dot{\hat{\xi}}_{i,r} &= \sigma_i + b_0 g_L(\psi_i(\hat{\xi}, \sigma)) + \kappa^r c_{i,1}(y_i - \hat{\xi}_{i,1})\\ \dot{\sigma}_i &= \kappa^{r+1} c_{i,0}(y_i - \hat{\xi}_{i,1}).\end{aligned} \tag{18}$$

In these equations, the coefficient κ and $c_{i,0}, c_{i,1}, \cdots, c_{i,r}$ are design parameters.

4.2 Main result

The controller proposed in this paper is defined by the couple of equations (17)–(18). This controller is completely specified by the set of parameters $B_0, K_0, L, c_{i,0}, c_{i,1}, \cdots, c_{i,r}$ and κ. In the previous subsection, structure and values of B_0 and K_0 have been specified. In what follows, we will show how the remaining design parameters can be chosen so as to obtain the desired goal, which – in a nutshell – is to (practically) decouple the "protected" output y from the "attacked" set z of state variables, so long as $z(t)$ remains bounded by a fixed – but otherwise arbitrary – number $\bar{M}$.

More specifically, we will prove that, if the design parameters are appropriately chosen, the response $\xi(t)$ can be made arbitrarily close (so long as $z(t)$ remains bounded by $\bar{M}$) to the response resulting from the implementation of the "ideal" control (15). In this respect, observe that, under the effect of the control law (15), one would obtain for $\xi(t)$ a response

$$\xi(t) = \mathrm{col}(\xi_1^{\mathrm{id}}(t), \ldots, \xi_m^{\mathrm{id}}(t))$$

in which $\xi_i^{\mathrm{id}}(t)$ is a solution of

$$\dot{\xi}_i^{\mathrm{id}} = (A_i + B_i K_{0,i})\xi_i^{\mathrm{id}}\,.$$

The Proposition that follows (in which we use B_R to denote the closed ball of radius R and $\hat{\xi}_{\mathrm{ext}} = \mathrm{col}(\hat{\xi}, \sigma))$ is main result of the paper .

Proposition 1 *Consider system (12) with control (17)–(18). Let R and $R^* >> R$ be fixed. Assume $x(0) \in B_R$ and $\hat{\xi}_{\mathrm{ext}}(0) \in B_R$. There is a choice of saturation* *level L and of the design parameters $c_{i,0},\cdots,c_{i,r}$ such that, for any choice of $\bar\varepsilon > 0$ there exists a number κ^* such that, if $\kappa \geq \kappa^*$, then for all t such that $x(t) \in B_{R^*}$ the components $\xi_1(t),\ldots,\xi_m(t)$ of the response $\xi(t)$ satisfy*

$$\|\xi_i(t) - \xi_i^{\rm id}(t)\| \leq \bar\varepsilon\,.$$

4.3 Proof of the main result: a change of coordinates

The arguments used in the proof of the main result are essentially the same as those used to show that a feedback law of the form (17)–(18) is able to induce – under the assumption that the zero-dynamics of the controlled system are asymptotically stable – an input-output behavior that asymptotically recovers the behavior obtained under the action of a control of the form (15) (see [5][26]). The novelty here is that we no longer assume that the zero dynamics are globally asymptotically stable and we show that those arguments can be used to prove "practical" decoupling of $\xi(t)$, and hence $y(t)$, from $z(t)$, so long as the latter remains bounded.

In order to analyze the response of the closed-loop system defined by (12)–(17)–(18), it is useful to make a change of the variables, introducing

$$\begin{array}{rcl} e_{i,1} &=& \kappa^r(\xi_{i,1} - \hat\xi_{i,1})\\ e_{i,2} &=& \kappa^{r-1}(\xi_{i,2} - \hat\xi_{i,2})\\ && \cdots\\ e_{i,r} &=& \kappa(\xi_{i,r} - \hat\xi_{i,r})\\ e_{i,r+1} &=& H_i z + K_i\xi + [b_i - b_0]g_L(\psi_i(\xi,\sigma)) - \sigma_i\,. \end{array} \tag{19}$$

Setting

$$\begin{array}{rcl} e &=& \mathrm{col}(e_1,\ldots,e_m)\\ e_i &=& \mathrm{col}(e_{i,1},e_{i,2},\ldots,e_{i,r_i+1})\,, \qquad i=1,\ldots,m \end{array}$$

equations (19) define a map

$$\begin{array}{rcccl} T &:& \mathbb{R}^{m(r+1)} &\to& \mathbb{R}^{m(r+1)}\\ && \hat\xi_{\rm ext} &\mapsto& e = T(z,\xi,\hat\xi_{\rm ext}) \end{array} \tag{20}$$

As shown in [8, pp.300-301]), the following property holds.

Lemma 1 *If assumption (16) is fulfilled, the map (20) is globally invertible.*

As consequence, (19) define a legitimate (partial) change of coordinates and we can express the closed-loop system in the coordinates $x = \mathrm{col}(z,\xi)$ and e. Note that e is a function of $(x,\hat\xi_{\rm ext})$ and, conversely, that $\hat\xi_{\rm ext}$ is as a function of (x,e).

We make now some manipulations in the equations that describe the closed-loop system, so as to put it in a form of two mutually "coupled" subsystems, one with state x and the other with state e. Later, we will discuss the effects of the couplings.

So long as the dynamics of x is concerned, adding and subtracting the function (15) to the control u defined in (17), one obtains

$$u = B^{-1}(-Hz - K\xi + K_0\xi) + \Delta_3(x,e)$$

in which (recall that $(\hat{\xi},\sigma)$ can be regarded as a function (x,e))

$$\Delta_3(x,e) = G_L(\psi(\hat{\xi},\sigma)) - B^{-1}(-Hz - K\xi + K_0\xi)\,.$$

As a consequence, the equations of system (12) controlled by (17) can be regarded as equations of the form

$$\begin{aligned} \dot{z} &= Fz + G\xi \\ \dot{\xi}_i &= (A_i + B_i K_{0,i})\xi_i + B_i \Delta_{3,i}(x,e) \end{aligned} \tag{21}$$

in which $\Delta_{3,i}(x,e)$ is the i-th row of $\Delta_3(x,e)$. These equations appear as a perturbed version of the equations resulting from the implementation of the "ideal" control (15). Recall also that $K_{0,i}$ is chosen in such a way as to make $(A_i + B_i K_{0,i})$ a Hurwitz matrix.

So long as the dynamics of the e_i's are concerned, appropriate calculations show that

$$\begin{aligned} \dot{e}_{i,1} &= \kappa(e_{i,2} - c_{i,r}e_{i,1}) \\ \dot{e}_{i,2} &= \kappa(e_{i,3} - c_{i,r-1}e_{i,1}) \\ &\cdots \\ \dot{e}_{i,r-1} &= \kappa(e_{i,r} - c_{i,2}e_{i,1}) \end{aligned} \tag{22}$$

$$\dot{e}_{i,r} = \kappa[e_{i,r+1} - c_{i,1}e_{i,1}] + \Delta_{1,i}(x,e)\,, \tag{23}$$

in which

$$\Delta_{1,i}(x,e) = \kappa[b_i - b_0][g_L(\psi_i(\hat{\xi},\sigma)) - g_L(\psi_i(\xi,\sigma))]\,,$$

and that

$$\dot{e}_{i,r+1} = -\kappa c_{i,0}e_{i,1} - \kappa\Delta_{0,i}(x,e)\begin{pmatrix} c_{1,0}e_{1,1} \\ c_{2,0}e_{2,1} \\ \cdots \\ c_{m,0}e_{m,1} \end{pmatrix} + \Delta_{2,i}(x,e)\,. \tag{24}$$

in which

$$\Delta_{0,i}(x,e) = [b_i - b_0]g'_L(\psi_i(\xi,\sigma))b_0^{-1}\,,$$

$$\Delta_{2,i}(x,e) = H_i\dot{z} + K_i\dot{\xi} + [b_i - b_0]g'_L(\psi_i(\xi,\sigma))b_0^{-1}K_{0i}\dot{\xi}_i\,.$$

Altogether, (22), (23) and (24) characterize a system of the form

$$\begin{aligned} \dot{e}_i = \kappa A_{e,i}e_i - \kappa B_{e2,i}\Delta_{0,i}(x,e)\begin{pmatrix} C_{e,1}e_1 \\ \cdots \\ C_{e,m}e_m \end{pmatrix} + \\ + B_{e1,i}\Delta_{1,i}(x,e) + B_{e2,i}\Delta_{2,i}(x,e) \end{aligned} \tag{25}$$

in which $A_{e,i} \in \mathbb{R}^{(r+1)\times(r+1)}$, $B_{e1,i} \in \mathbb{R}^{(r+1)}$, $B_{e2,i} \in \mathbb{R}^{(r+1)}$, $C_{e,i}^{\mathrm{T}} \in \mathbb{R}^{(r_i+1)}$ are matrices defined as

$$A_{e,i} = \begin{pmatrix} -c_{i,r} & 1 & 0 & \cdots & 0 & 0 \\ -c_{i,r-1} & 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ -c_{i,1} & 0 & 0 & \cdots & 0 & 1 \\ -c_{i,0} & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

$$B_{e1,i} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 1 \\ 0 \end{pmatrix}, \quad B_{e2,i} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix},$$

$$C_{e,i} = \begin{pmatrix} c_{i,0} & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Relevant, in the analysis that follows, is the possibility of showing that the functions $\Delta_{0,i}(x,e), \Delta_{1,i}(x,e), \Delta_{2,i}(x,e)$, have the following properties (see, in this respect, [8, 304-305]).

Lemma 2 *If (16) holds and $\kappa \geq 1$, there exist numbers $\delta_0 < 1$ and δ_1 such that*

$$\begin{array}{rcll} \|\Delta_{0,i}(x,e)\| & \leq & \delta_0 < 1 & \text{for all } (x,e) \text{ and all } \kappa \\ \|\Delta_{1,i}(x,e)\| & \leq & \delta_1 \|e\| & \text{for all } (x,e) \text{ and all } \kappa\,. \end{array} \tag{26}$$

Moreover, for each $R > 0$ there is a number M_R such that

$$\|x\| \leq R \quad \Rightarrow \quad \|\Delta_{2,i}(x,e)\| \leq M_R \qquad \text{for all } e \text{ and all } \kappa. \tag{27}$$

Finally, note that $(x,e) = (0,0)$ is an equilibrium point of the system defined by (21)–(25).

4.4 Proof of the main result: analysis of the response

We assume in what follows that the initial conditions $(x(0), \hat{\xi}_{\text{ext}}(0))$ of the controlled system are in a fixed bounded set, that is we assume that $x(0) \in B_R$ and $\hat{\xi}_{\text{ext}}(0) \in B_R$, for some $R > 0$. Pick any number $R^* >> R$ and let $[0, T_{\max}]$ be a time interval such that $x(t) \in B_{R^*}$ for all $t \in [0, T_{\max}]$. We will prove that, if the design parameters are appropriately chosen, on the entire time interval $[0, T_{\max}]$, the states $\xi_i(t)$ remain arbitrarily close to the trajectories of the stable systems $\dot{\xi}_i = (A_i + B_i K_{0,i})\xi_i$.

First of all, the threshold L of the saturation function is fixed, as

$$L = \max_{x \in B_{R^*}} \|B^{-1}(-Hz - K\xi + K_0\xi)\| + 1\,. \tag{28}$$

Then, observe that, since $G_L(\cdot)$ is bounded by $L(c+1)$, the quantity $\Delta_3(x,e)$ remains bounded so long as $x(t) \in B_{R^*}$, by a bound that does not depend on the design parameter κ (rather, this bound only depends on the choice of R^*). As a consequence, with $x(0) \in B_R$ and $\hat{\xi}_{\rm ext}(0) \in B_R$, given any arbitrarily small number $0 < \delta << (R^* - R)$ there is a time T_0, independent of the design parameter κ, such that, for all times $t \in [0, T_0]$, $x(t) \in B_{R+\delta}$. During the time interval $[0, T_0]$ also the state $e(t)$ remains bounded. This is seen from the bottom equation of (25), using the bounds determined for $\Delta_{0,i}(x, e)$, $\Delta_{1,i}(x, e)$, $\Delta_{2,i}(x, e)$ and the fact that $x(t) \in B_{R^*}$ for all $t \in [0, T_0]$. It is worth observing, in this respect, that the value of κ does affect the bound on $e(t)$. In fact, looking at the definitions of the various components of e, it is seen that $\|e(0)\|$ grows with κ (despite of the fact that, by assumption, $\|x(0)\| \leq \bar{M}$ and $\|\hat{\xi}_{\rm ext}(0)\| \leq \bar{M}$). This is not a problem, though, as it will be shown in the sequel.

We study now the behavior of $e(t)$ for times larger than T_0. To this end, we make use of the following results (see, in this respect, [8, pp.308-312]).

Lemma 3 *Consider the set of systems*

$$\dot{e}_i = A_{e,i} e_i - B_{e2,i} \Delta_{0,i}(x, e) \begin{pmatrix} C_{e,1} e_1 \\ \cdots \\ C_{e,m} e_m \end{pmatrix} \qquad i = 1, \ldots, m$$

where $A_{e,i}, B_{e2,i}, C_{e,i}$ and $\Delta_{0,i}(x, e)$ are defined as in (25). There is a choice of the coefficients $c_{i,0}, \cdots, c_{i,r}$ such with this system is asymptotically stable, with a quadratic, x-independent, Lyapunov function.

Lemma 4 *Let the $c_{i,j}$'s be chosen so as to make stability property indicated in Lemma 3 fulfilled. Suppose $x(t) \in B_{R^*}$ for all $t \in [0, T_{\max})$ and suppose that $\xi_{\rm ext}(0) \in B_R$. Then, for every $0 < T_0 \leq T_{\max}$ and every $\varepsilon > 0$, there is a κ^* such that, for all $\kappa \geq \kappa^*$,*

$$\|e(t)\| \leq 2\varepsilon \quad \text{for all } t \in [T_0, T_{\max}).$$

Lemma 5 *Suppose $\|e(t)\| \leq 2\varepsilon$ for all $t \in [T_0, T_{\max})$. If ε is small enough, then*

$$\|\psi(\xi, \sigma)\| \leq L - \frac{1}{2}\,.$$

As a consequence, $G_L(\psi(\xi, \sigma)) = \psi(\xi, \sigma)$.

From Lemma 3 and 4, we learn that there is a choice of the design parameters $c_{i,0}, \cdots, c_{i,r}$ such that, for any choice of ε, there is a number κ^* such that, for any $\kappa \geq \kappa^*$, so long as $x(t)$ remains in the set B_{R^*} on the time interval $[T_0, T_{\max})$, on the same time interval $\|e(t)\|$ is bounded by 2ε. From Lemma 5 we see that, if ε is chosen sufficiently small, on the same interval $G_L(\psi(\xi(t), \sigma(t))) = \psi(\xi(t), \sigma(t))$. We use this latter property to show that, on the same time interval,

$$G_L(\psi(\hat{\xi}(t), \sigma(t))) = \psi(\hat{\xi}(t), \sigma(t)), \tag{29}$$

i.e. that none of the components of the control (17) is "saturated".

To this end, observe that

$$\hat{\xi} = \xi - D(\kappa)e \tag{30}$$

in which $D(\kappa) = \text{diag}(D_1(\kappa), \cdots, D_m(\kappa))$ where $D_i(\kappa)$ is the $r_i \times (r_i+1)$ matrix

$$D_i(\kappa) = \begin{pmatrix} \kappa^{-r_i} & 0 & \dots & 0 & 0 \\ 0 & \kappa^{-r_i-1} & \dots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \dots & \kappa^{-1} & 0 \end{pmatrix}$$

and note that, if (without loss of generality) $\kappa \geq 1$, then $\|D(\kappa)\| \leq 1$. Since,

$$\psi(\hat{\xi}, \sigma) = \psi(\xi, \sigma) - B_0^{-1} K_0 D(\kappa) e\,,$$

if $\kappa > 1$ we have

$$\|\psi(\hat{\xi}, \sigma)\| \leq \|\psi(\xi, \sigma)\| + \|B_0^{-1}\| \|K_0\| \|e\|.$$

Thus, if $\|e\| \leq 2\varepsilon$ and ε is small enough we conclude from the previous Lemma that $\|\psi(\hat{\xi}, \sigma)\| < L$, and this proves that (29) holds on the time interval $[T_0, T_{\max})$.

We return now to equation (21) and observe that, for all $t \in [T_0, T_{\max})$,

$$\begin{aligned} \Delta_3(x, e) &= G_L(\psi(\hat{\xi}, \sigma)) - B^{-1}(-Hz - K\xi + K_0\xi) \\ &= \psi(\hat{\xi}, \sigma) - B^{-1}(-Hz - K\xi + K_0\xi) \\ &= \psi(\xi, \sigma) - B_0^{-1} K_0 D(\kappa) e - B^{-1}(-Hz - K\xi + K_0\xi)\,. \end{aligned}$$

In this expression, $\psi(\xi, \sigma)$ can be replaced by

$$\psi(\xi, \sigma) = B^{-1}(-Hz - K\xi + K_0\xi + \varsigma) \tag{31}$$

in which $\varsigma = \text{col}(e_{1,r+1}, e_{2,r+1}, \ldots, e_{m,r+1})$. In fact, from the last of (19) it is seen that

$$\varsigma = Hz + K\xi + [B - B_0] G_L(\psi(\xi, \sigma)) - \sigma\,.$$

Adding and subtracting $K_0\xi$, using the fact that $G_L(\psi(\xi, \sigma)) = \psi(\xi, \sigma) = B_0^{-1}[K_0\xi - \sigma]$, we obtain

$$\varsigma = Hz + K\xi - K_0\xi + B\psi(\xi, \sigma)$$

from which (31) follows.

As a consequence, it is seen that

$$\Delta_3(x, e) = -B_0^{-1} K_0 D(\kappa) e + B^{-1}\varsigma\,.$$

Recalling that ς is part of e and using again the property $\|D(\kappa)\| \leq 1$, we see that

$$\|\Delta_3(x, e)\| \leq (\|B_0^{-1}\| \, \|K_0\| + \|B^{-1}\|)\|e\|\,. \tag{32}$$

It is seen from this estimate that on the time interval $[T_0, T_{\max})$, on which $\|e(t)\| \leq 2\varepsilon$, the dynamics of the $\xi_i(t)$'s can be made arbitrarily close (by lowering the value of ε) to those of the "ideally decoupled" systems

$$\dot{\xi}_i^{\mathrm{id}} = (A_i + B_i K_{0,i})\xi_i^{\mathrm{id}} .$$

Specifically, observe that the difference $\delta\xi_i = \xi_i - \xi_i^{\mathrm{id}}$ obeys

$$\delta\dot{\xi}_i = (A_i + B_i K_{0,i})\delta\xi_i + B_i \Delta_{3,i}(x, e)$$

with initial condition $\delta\xi_i(0) = 0$. As a consequence

$$\begin{aligned} \delta\xi_i(t) = & \int_0^{T_0} e^{(A_i+B_iK_{0,i})(t-\tau)} B_i \Delta_{3,i}(x(\tau), e(\tau)) d\tau + \\ & + \int_{T_0}^{t} e^{(A_i+B_iK_{0,i})(t-\tau)} B_i \Delta_{3,i}(x(\tau), e(\tau)) d\tau \end{aligned}$$

in which $(A_i + B_i K_{0,i})$ is a Hurwitz matrix. The first term of this expression can be arbitrarily lowered by lowering the value of T_0 (recall that $\Delta_{3,i}(x, e)$ is bounded). Note, in this respect, that T_0 can be arbitrarily lowered by increasing κ (see Lemma 4). The second term, on the other hand, using the bound (16) can be bounded as

$$\int_{T_0}^{t} e^{(A_i+B_iK_{0,i})(t-\tau)} B_i \Delta_{3,i}(x(\tau), e(\tau)) d\tau \leq \bar{M} \sup_{\tau \in [T_0, t)} \|e(\tau)\|$$

Thanks to Lemma 4, this term can be arbitrarily lowered by lowering ε. Thus, in summary, we conclude that, given any choice of $\bar{\varepsilon}$, if we pick a sufficiently large value of κ, we have

$$\|\delta\xi_i(t)\| \leq \varepsilon \quad \textit{for all } t \in [T_0, T_{\max}).$$

and this proves Proposition 1.

4.5 Remarks

It is worth stressing that we have considered a scenario in which a fixed set of generators is to be protected by attacks affecting a different set of generators. The proposed control (17)–(18) does not rely on any information about the entry points of the attack in the system (in the context of power systems represented by the generators under the control of the hacker); it only uses data from the generators that have to be protected.

By means of the control law (17)–(18) we are able to practically decouple $\xi(t)$ from $z(t)$ on the finite time internal $[0, T_{\max})$, in which $T_{\max}$ is determined by the bound R^* chosen for $x(t)$, bound which in turn determines the values of the design parameters $L, c_{i,0}, c_{i,1}, \cdots, c_{i,r}$ and κ^*. In practice, as kindly pointed out by an anonymous reviewer, this should not be seen as a limitation of the method.

In fact, it is reasonable to conceive that in practice $z(t)$ will remain bounded for all t, in which case we may well take $T_{\max} = \infty$. As a matter of fact, the attack on $z(t)$ takes place through actuators that have saturations. This means that, although the goal of the attacker is to make the zero dynamics anti-stable, $z(t)$ will only diverge so long as the actuators are not saturated. In the end, if the un-attacked zero dynamics is stable, a consequence of a saturated attack is a bounded attacked $z(t)$. Another reason why one can consider $T_{\max} = \infty$ is that, since the attacked generators have automatic protections, if $z(t)$ gets excessively large then such generators at some time will be automatically disconnected from the network. In this case, our defence strategy is effective on the finite time interval during which the attacked generators are connected and will be no longer needed after the disconnection of such generators.

5 Simulation Results

5.1 Case study

In order to validate the proposed decoupling strategy, the WSCC 9-bus test power network reported in Fig. 1 has been considered [18]; it represents an approximation of the Western System Coordinating Council (WSCC) to an equivalent system with 3 generation buses and 6 load buses [1]. For this power system the inertia and damping matrices characterizing the power plants are respectively $M = \mathrm{diag}(0.125, 0.034, 0.016)$ and $D = \mathrm{diag}(0.125, 0.068, 0.048)$, while the interconnections within the network are characterized by the laplacian matrix indicated in eq. (33).

In the considered test case the power plant 1 is unprotected, the power plant 2 is the one used for launching the attack, while the power plant 3 is the one to be protected using the proposed decoupling approach. With reference to the nomenclature used in section 3 (see in particular eq. (9)), in this case $n = 6$, $m = 1$, $r = 2$; the controls used by the attacker and by the defender are $u_{\mathrm{a}} = P_{\mathrm{g}2}$ and, respectively, $u = P_{\mathrm{g}3}$ and the protected output is $y = \delta_3$ (recall that $\omega_3 = \dot{\delta}_3$). The simulations reported in the following have been performed

---

$$L_{\mathrm{N}} = \begin{pmatrix} 0.058 & 0 & 0 & -0.058 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.063 & 0 & 0 & -0.063 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.059 & 0 & 0 & -0.059 & 0 & 0 & 0 \\ -0.058 & 0 & 0 & 0.235 & 0 & 0 & -0.085 & -0.092 & 0 \\ 0 & -0.063 & 0 & 0 & 0.296 & 0 & -0.161 & 0 & -0.072 \\ 0 & 0 & -0.059 & 0 & 0 & 0.330 & 0 & -0.170 & -0.101 \\ 0 & 0 & 0 & -0.085 & -0.161 & 0 & 0.246 & 0 & 0 \\ 0 & 0 & 0 & -0.092 & 0 & -0.170 & 0 & 0.262 & 0 \\ 0 & 0 & 0 & 0 & -0.072 & -0.101 & 0 & 0 & 0.173 \end{pmatrix} \tag{33}$$

using Simulink.

The attack control has been chosen so as to force an anti-stable zero dynamics characterized by four identical positive real eigenvalues with time constant $\tau = 20$ s. The attack is triggered by a small initial condition on the angular speed of machine 2 ($\omega_2(0) = 1.3 \times 10^{-2}$rad/s), and all the machines are affected by the attack due to their mutual coupling.

During the time period considered for running the simulation, the attack control remains approximately in the range of 0.1 p.u.. Also the largest differences among angles occur at the interconnection of generators to load buses, and remain below 0.5 rad (see Fig. 2), value beyond which the linear model approximating the behaviour of the power system becomes questionable. Then the validation of the proposed approach is here performed considering a time horizon of approximately 6 s.

5.2 Protected case

In order to test the effectiveness of the proposed approach, both the cases of *ideal* and *robust* decoupling of the protected dynamics are analyzed and compared. For the purpose of tuning both controllers, the matrix $K_0 = [-9 \ \ -6]$ has been chosen, using which, under the assumption of exact decoupling, the protected dynamics is characterized by two identical negative real eigenvalues with time constant $\tau \approx 0.33$ s.

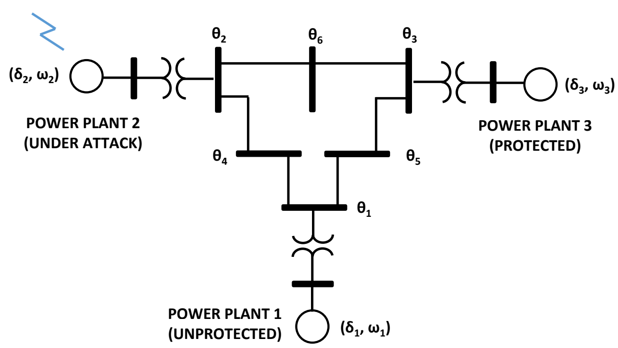

Figure 1: The WSCC 9-bus power network.

As far as the robust controller is concerned, at first the matrix B_0 (reduced here to a single scalar coefficient b_0) has been set to $0.8b = 0.8M_3^{-1}$. Then, according to Lemma 3, the choice $c_0 = 6, c_1 = 11, c_2 = 6$ has been performed, which guarantees asymptotic stability of the observer's error dynamics resulting from the robust control.

Fig. 3 and 4 report respectively the evolution of the error components and its norm for three different values of the gain ($\kappa \in \{50, 100, 200\}$): the higher is κ, the shorter are the transient and the residual magnitude of the error. As a matter of fact it is seen that if the value $\varepsilon = 0.01$ is chosen, the value $\kappa = 200$

guarantees that the error's norm reduces to and remains below the threshold 2ε (-34 dB) approximately after $T_0 = 32$ ms.

Fig. 5 reports the unsaturated defence control ψ, the actual control P_{g3} subject to saturation and the deviation Δ_3 between the robust and ideal control (for each of the different values of κ considered before). As observed in the general analysis, the higher is κ, the higher is the "peak" in $\psi(t)$, which explains why for larger values of κ a saturation in the actual control may occur (Fig. 5).

In the light of the above, the decoupling control becomes effective after approximately T_0 seconds, and in particular on a time scale that is one order of magnitude smaller than the time constant imposed by the choice of the matrix K_0.

Having completed the tuning phase, the effectiveness of the decoupling control is here analyzed on a larger time scale. Fig. 6 reports the rotor angles for all the machines in the network, showing the comparison between the evolutions obtained when the robust and ideal controls are applied to the plant. Differently from the ideal case, in which the exact decoupling is achieved along the whole considered time period, in the robust case the rotor angle δ_3 of the protected machine experiences a transient, due to the initial coupling with the infected zero dynamics, after which δ_3 becomes very small, meaning that the decoupling is occurring in practice over the entire period in which machines 1 and 2 loses stability. Compared to the ideal case, notice that the deviation of δ_3 from zero remains in the order of 10^{-2} rad. Similar considerations hold for the evolutions of machines' angular speeds reported in Fig. 7; again notice how ω_1 and ω_2 diverge, while the angular speed ω_3 remains substantially unaffected by the attack.

Finally Fig. 8 shows the attack and defence controls, again the ones resulting from the evolution of the power system state when the ideal and robust decoupling controls are applied to the plant. It can be seen here that, evaluated on the whole considered time period, the defence control has an opposite sign with respect to the attack, due to the need of balancing the excess energy introduced by the attack in the power system; also the defence effort is smaller with respect to the attack, considered that the attack energy is distributed among all the machines in the network. Again deviations among the controls in the ideal and robust cases appear, due to the different evolutions characterizing the state of the system.

6 Conclusions

In this paper a robust protection scheme in reaction to destabilizing attacks operated against linear cyber-physical systems has been presented. The proposed defence control is able to decouple the protected dynamics from the infected one, the latter seen by the defender as the zero dynamics of the system at study; the distinctive aspect of the proposed method lays in its robustness, meaning that the control objective is achieved in practice despite the lack of information about the plant model and state.

The application to power systems has been shown to be effective in relation to the protection of power plants electromechanical dynamics (rotor angles and angular speeds) against attacks operated using the governing system of vulnerable machines. In particular the robust control approaches the ideal control (allowing exact decoupling) on a time scale smaller than the one characterizing the attack.

Motivated by the need of extending the rotor angles and speeds operational range in which the decoupling is required to be effective, a future direction for this research stream considers the design of a robust decoupling control in the context of the nonlinear-descriptor representation of power systems.

7 Acknowledgments

This work has been carried out in the framework of the ATENA project (Grant Agreement no. 700581) partially funded by the EU. The authors gratefully acknowledge the members of the ATENA project and, in particular, the participants from the Consortium for Research in Automation and Telecommunications (CRAT), Rome, Italy.

References

[1] A. AL Hinai. Voltage collapse prediction for interconnected power systems. Master's thesis, Master degree thesis in Electrical Engineering, West Virginia University, 2000.

[2] Saurabh Amin, Alvaro A Cárdenas, and S Shankar Sastry. Safe and secure networked control systems under denial-of-service attacks. In *International Workshop on Hybrid Systems: Computation and Control*, pages 31–45. Springer), 2009.

[3] W. Ao, Y. Song, and C. Wen. Adaptive cyber-physical system attack detection and reconstruction with application to power systems. *IET Control Theory and Applications*, 10(12):1458–1468, 2016.

[4] P. Cheng, L. Shi, and B. Sinopoli. Guest editorial special issue on secure control of cyber-physical systems. *IEEE Transactions on Control of Network Systems*, 4(1):1–3, 2017.

[5] L. B. Freidovich and H. K. Khalil. Performance recovery of feedback-linearization-based designs. *IEEE Transactions on Automatic Control*, 53(10):2324–2334, Nov 2008.

[6] Ziyang Guo, Dawei Shi, Karl Henrik Johansson, and Ling Shi. Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*, 4(1):4–13, 2017.

[7] Andreas Hoehn and Ping Zhang. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In *American Control Conference (ACC), 2016*, pages 302–307. IEEE), 2016.

[8] A. Isidori. *Lectures in Feedback Design for Multivariable Systems*. Springer Verlag, 2016.

[9] Jean-Yves Keller and Dominique Sauter. Monitoring of stealthy attack in networked control systems. In *Control and Fault-Tolerant Systems (SysTol), 2013 Conference on*, pages 462–467. IEEE), 2013.

[10] P. Kundur, N. J. Balu, and M. G. Lauby. *Power system stability and control.*, volume 7. McGraw-Hill New York, 1994.

[11] Svetlana Kuvshinkova. Sql slammer worm lessons learned for consideration by the electricity sector. *North American Electric Reliability Council*, 1(2):5, 2003.

[12] Cheolhyeon Kwon and Inseok Hwang. Cyber attack mitigation for cyber–physical systems: hybrid system approach to controller design. *IET Control Theory & Applications*, 10(7):731–741, 2016.

[13] Yao Liu, Peng Ning, and Michael K Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, 14(1):13, 2011.

[14] Walter Lucia, Bruno Sinopoli, and Giuseppe Franze. A set-theoretic approach for secure and resilient control of cyber-physical systems subject to false data injection attacks. In *Cyber-Physical Systems Workshop (SOSCYPS), Science of Security for*, pages 1–5. IEEE), 2016.

[15] J. Machowski, J. Bialek, and J. Bumby. *Power system dynamics: stability and control.* Second edition. John Wiley & Sons, 2011.

[16] Y. Mo and B. Sinopoli. Secure control against replay attacks. In *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 911–918. IEEE), Sept 2009.

[17] Gyunghoon Park, Hyungbo Shim, Chanhwa Lee, Yongsoon Eun, and Karl H Johansson. When adversary encounters uncertain cyber-physical systems: Robust zero-dynamics attack with disclosure resources. In *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pages 5085–5090. IEEE), 2016.

[18] F. Pasqualetti, A. Bicchi, and F. Bullo. A graph-theoretical characterization of power network vulnerabilities. In *American Control Conference ACC*, pages 3918–3923. IEEE), 2011.

[19] F. Pasqualetti, F. Dorfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, Nov 2013.

[20] Henrik Sandberg, Saurabh Amin, and Karl Henrik Johansson. Cyberphysical security in networked control systems: An introduction to the issue. *IEEE Control Systems*, 35(1):20–23, 2015.

[21] Hyungbo Shim, Gyunghoon Park, Youngjun Joo, Juhoon Back, and Nam Hoon Jo. Yet another tutorial of disturbance observer: robust stabilization and recovery of nominal performance. *Control Theory and Technology*, 14(3):237–249, 2016.

[22] Jill Slay and Michael Miller. Lessons learned from the maroochy water breach. In *International Conference on Critical Infrastructure Protection*, pages 73–82. Springer), 2007.

[23] Chih-Che Sun, Chen-Ching Liu, and Jing Xie. Cyber-physical system security of a power grid: State-of-the-art. *Electronics*, 5(3):40, 2016.

[24] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson. Revealing stealthy attacks in control systems. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1806–1813. IEEE), 2012.

[25] André Teixeira, Iman Shames, Henrik Sandberg, and Karl Henrik Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135 – 148, 2015.

[26] Lei Wang, Alberto Isidori, and Hongye Su. Output feedback stabilization of nonlinear mimo systems having uncertain high-frequency gain matrix. *Systems & Control Letters*, 83:1 – 8, 2015.

[27] W. M. Wonham. *Linear Multivariable Control: a Geometric Approach (3rd ed.)*. Springer Verlag, 1979.

[28] Junfeng Zhao, Jing Wang, and Lei Yin. Detection and control against replay attacks in smart grid. In *Computational Intelligence and Security (CIS), 2016 12th International Conference on*, pages 624–627. IEEE), 2016.

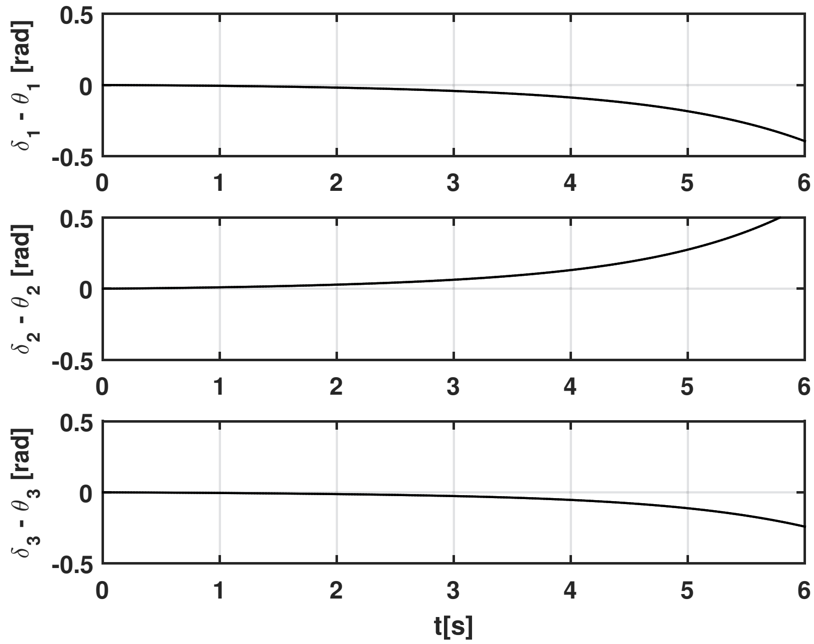

Figure 2: Differences among angles at the interconnection of generators to load buses, in absence of defence.

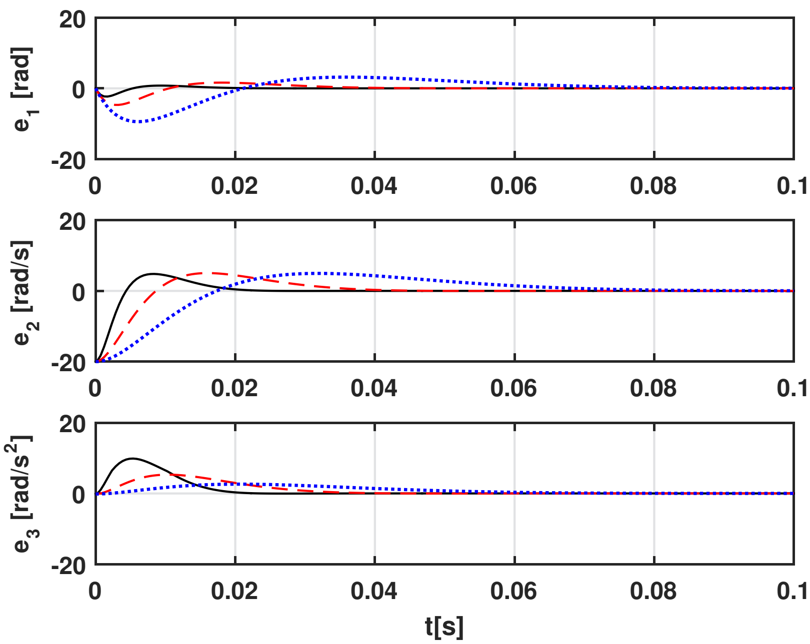

Figure 3: Components of the extended observer's error for $\kappa = 50$ (blue dotted line), 100 (red dashed line), 200 (black solid line).

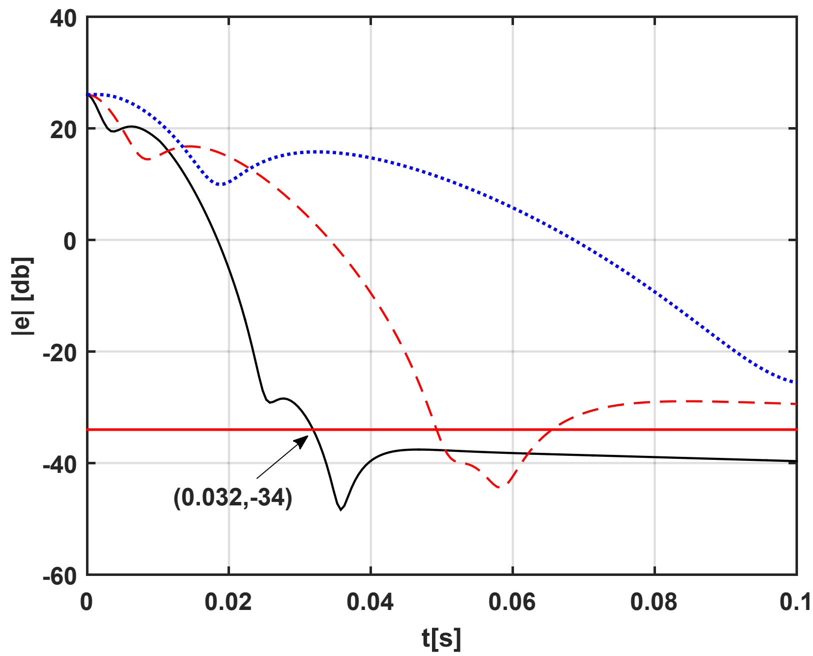

Figure 4: Absolute value of the extended observer's error for $\kappa = 50$ (blue dotted line), 100 (red dashed line), 200 (black solid line), expressed in dB to properly notice the residual magnitude of the error. The horizontal red solid line represents the threshold 2ϵ (-34 dB).

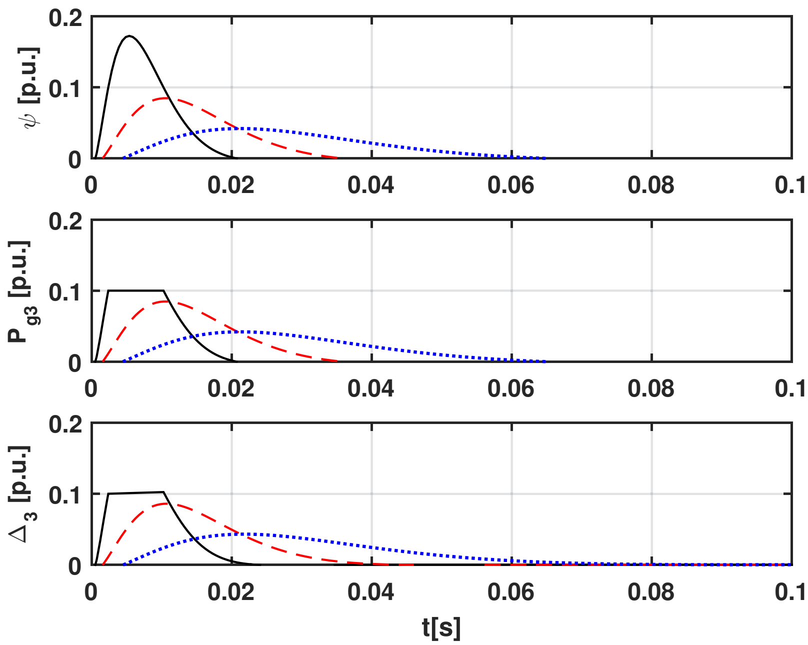

Figure 5: Unsaturated control, actual robust control and deviation from the ideal decoupling control for $\kappa = 50$ (blue dotted line), 100 (red dashed line), 200 (black solid line).

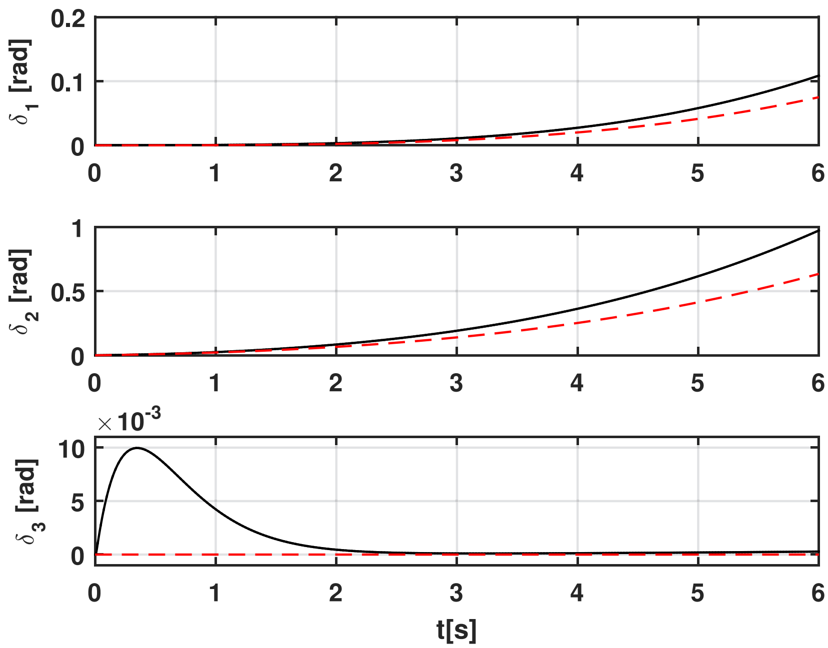

Figure 6: Rotor angles dynamics in case of robust control (black solid line) and ideal control (red dashed line).

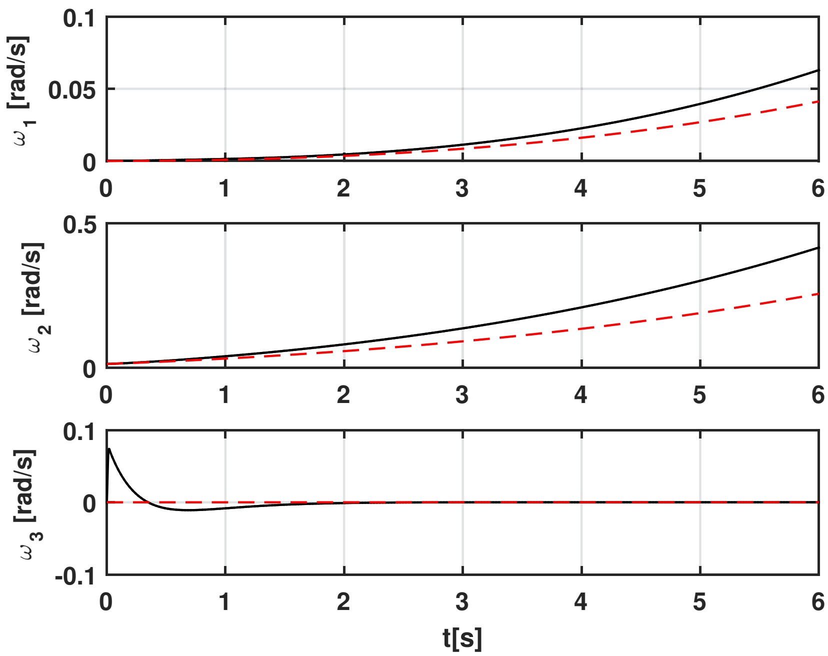

Figure 7: Angular speeds dynamics in case of robust control (black solid line) and ideal control (red dashed line).

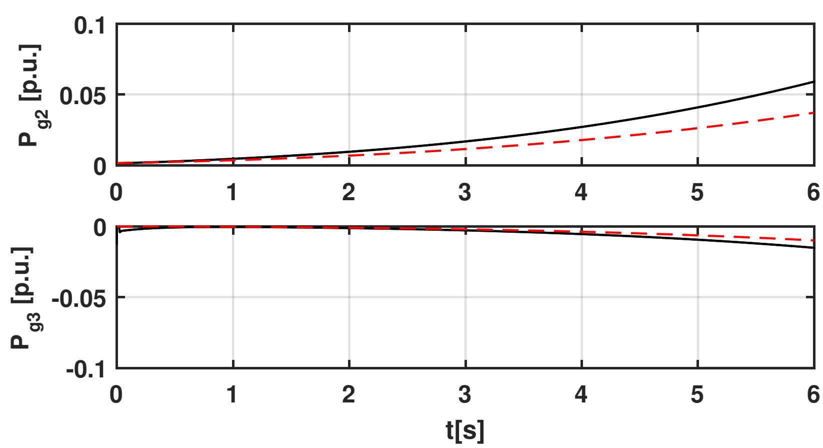

Figure 8: Attack and defence controls in case of robust control (black solid line) and ideal control (red dashed line).