


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

Alessandro Di Giorgio , Antonio Pietrabissa, Francesco Delli Priscoli, Alberto Isidori

¹ Department of Computer, Control and Management Engineering 'A. Ruberti', University of Rome La Sapienza, Via Ariosto 25, 00185 Rome, Italy
 E-mail: 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 [1] in 2000, the SQL Slammer worm attack on the Ohio nuclear plant network [2] in 2002, the coordinated attack on the Ukrainian power grid [3] 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., [4][5] and references therein. Several classes of attack design and detection have been identified and studied, among the others: *detection and denial of service* attacks [6][7], *replay* attacks [8][9], *false data injection* attacks [10][11], *random* and *constant bias* attacks [12], and *zero dynamics* attacks [13][14][15].

In particular, it has been observed that systems having unstable zero dynamics are vulnerable to *stealthy attacks*. In fact, as shown e.g. in [16], 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, [16] 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 [17]. Researches on the detection of zero dynamics attacks are based on the design of centralized and decentralized observers [15], Kalman filtering [18], adaptive sliding mode observers [19], or by suitably altering the input behaviour of the process [20].

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 [21], 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

87 applying both the nominal attack and defence controls is the access
88 to information owned by different players.

89 In the light of the above, this paper presents a robust defence strat-
90 egy to attacks launched against linear time invariant systems and its
91 application to power systems.

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

99 2 Reference scenario

100 2.1 Power system model

101 In this section the power system model is recalled. The generic
102 network here considered is constituted by \bar{m} power plants and q
103 load buses. It is well known that the electromechanical dynamics
104 of a power system results from the composition of second order
105 swing equations, through the nonlinear algebraic power flow equa-
106 tions [22][23]. As reported in [24], under the assumptions of lossless
107 network, small angular differences and small deviations of bus volt-
108 ages from rated values, the power system dynamics is described by
109 the following time invariant linear descriptor model

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

110 In (1) $\delta = \text{col}(\delta_1, \delta_2, \dots, \delta_{\bar{m}})$ and $\omega = \text{col}(\omega_1, \omega_2, \dots, \omega_{\bar{m}})$
111 denote the vectors of machines rotor angles and angular speeds, $\theta =$
112 $\text{col}(\theta_1, \theta_2, \dots, \theta_q)$ denotes the vector of load angles at load buses,
113 $P_g = \text{col}(P_{g1}, P_{g2}, \dots, P_{g\bar{m}})$ and $P_\ell = \text{col}(P_{\ell 1}, P_{\ell 2}, \dots, P_{\ell q})$
114 are the vectors of mechanical input powers at generator
115 buses and electrical powers at load buses, the matrices $M =$
116 $\text{diag}(M_1, \dots, M_{\bar{m}})$ and $D = \text{diag}(D_1, \dots, D_{\bar{m}})$ model the
117 machines inertia and damping coefficients; finally L_{gg} , $L_{g\ell}$, $L_{\ell g}$ and
118 $L_{\ell\ell}$ are properly sized submatrices of the network laplacian matrix

$$L_N = \begin{pmatrix} L_{gg} & L_{g\ell} \\ L_{\ell g} & L_{\ell\ell} \end{pmatrix} \quad (2)$$

119 where L_{gg} is diagonal and $L_{\ell\ell}$ is invertible. The model can be fur-
120 ther simplified by explicitly calculating θ from the third component
121 of (1) as

$$\theta = -L_{\ell\ell}^{-1}[L_{\ell g}\delta + P_\ell] \quad (3)$$

122 and substituting it into the angular speed dynamics to obtain

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

123 Without loss of generality, it is possible to put $P_\ell = 0$, mean-
124 ing that in what follows P_g will be intended as the deviation of the
125 mechanical power from the value allowing to sustain a given loading
126 condition during normal operation.

127 The resulting model has the standard linear time invariant form

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

128 where $x = \text{col}(\delta, \omega) \in \mathbb{R}^{2\bar{m}}$, $\bar{u} = P_g \in \mathbb{R}^{\bar{m}}$.

129 2.2 Attack scenario and defence requirements

130 In this section the attack scenario under investigation is described,
131 raising the requirements for the defence design. In the reference
132 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 of the attack.

As a result of this classification, the input to model (5) is
partitioned as $\bar{u} = \text{col}(u_a, u_p, u_u)$ where $u_a \in \mathbb{R}^{m_a}$ is the input
available to the attacker, $u_p \in \mathbb{R}^{m_p}$ is the input of the protected
machines and $u_u \in \mathbb{R}^{m_u}$ 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 genera-
tion 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 sup-
posed to be induced by the action of an *hacker* which, taking
advantage of the vulnerability 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 indus-
try unbundling, the defender is supposed to not have access to the
state of the power plants which are not under its control (being oper-
ated 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); addition-
ally 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 $y_p \in \mathbb{R}^{m_p}$ 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 angu-
lar 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
 $y \in \mathbb{R}^m$.

133 3 An Attack Model

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

$$\begin{aligned} \dot{x} &= Ax + B_p u + B_a u_a \\ y &= C_p x \end{aligned} \quad (6)$$

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 atten-
tion 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 .

194 To simplify matters, we consider hereafter the case in which

$$\begin{aligned} C_p B_p &= C_p A B_p = \dots = C_p A^{r-2} B_p = 0 \\ C_p A^{r-1} B_p &= \text{diag}(b_1, b_2, \dots, b_m) \end{aligned} \quad (7)$$

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

206 It is also assumed

$$\begin{pmatrix} C_p \\ C_p A \\ \dots \\ C_p A^{r-1} \end{pmatrix} B_a = 0. \quad (8)$$

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

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

$$\begin{aligned} \dot{z}_0 &= F_0 z_0 + G_0 \xi + G_{0,a} u_a \\ \dot{\xi}_i &= A_i \xi_i + B_i (H_{i,0} z_0 + K_i \xi + b_i u_i) \\ y_i &= C_i \xi_i \quad i = 1, \dots, m \end{aligned} \quad (9)$$

212 in which

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

$$C_i = (1 \quad 0 \quad \dots \quad 0)$$

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

215 *Remark.* In geometric terms, the previous setup can be characterized
216 as follows (see [25, pp.87-90] and [25, pp.104-113] for definitions
217 and basic properties related to the concepts of (A, B) -invariant sub-
218 space and of *controllability* subspace). Set $\bar{B} = (B_p \quad B_a)$. If (7)
219 holds, then \mathcal{V}^* , the largest (A, \bar{B}) -invariant subspace contained in
220 $\text{Ker}(C_p)$, is given by

$$\mathcal{V}^* = \text{Ker} \begin{pmatrix} C_p \\ C_p A \\ \dots \\ C_p A^{r-1} \end{pmatrix}.$$

221 In the coordinates of (9)

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

222 and (8) is equivalent to

$$\text{Im}(B_a) \subset \mathcal{V}^*.$$

223 Moreover \mathcal{R}^* , the largest *controllability subspace* of (A, \bar{B}) con-
224 tained in $\text{Ker}(C_p)$, can be identified with the reachable set of the
225 pair $(F_0, G_{0,a})$. If such pair is controllable, then $\mathcal{R}^* = \mathcal{V}^*$. \triangleleft

The attacker can perturb the dynamics of (6) in various ways, 226
depending on the information available. For instance, if the pair 227
 $(F_0, G_{0,a})$ is controllable and z_0 is available for measurement, the 228
attacker u_a can choose the strategy 229

$$u_a = K_a z_0 \quad (10)$$

so as to assign eigenvalues with positive real parts to the matrix 230
 $(F_0 + G_{0,a} K_a)$. The effect of such attack is that of forcing, on 231
the resulting system with input u and output y , an *antistable zero* 232
dynamics. 233

This strategy presumes the availability of z_0 as well as an accurate 234
knowledge of F_0 and $G_{0,a}$. If this is not the case, an equivalent result 235
could be obtained by means of a dynamic control law 236

$$\begin{aligned} \dot{\zeta} &= \tilde{A}_a \zeta + \tilde{B}_a y_a \\ u_a &= \tilde{C}_a \zeta + \tilde{D}_a y_a \end{aligned} \quad (11)$$

driven by a set of auxiliary measurements $y_a = M_0 z_0 + N \xi$, so 237
long as the system 238

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

can be rendered antistable. 239

Simple manipulations show that if the attack strategy is chosen as 240
in (11), a system of the form 241

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

is obtained, where $z = \text{col}(\zeta, z_0)$, and the matrix F is *anti-stable*. 242

While the specific target of the attack strategies (10) and (11) 243
are the zero dynamics associated with the protected output y , it 244
should be stressed that – because of the inherent coupling between 245
the z 's and the ξ 's (which reflects, in the present context, the cou- 246
pling between the protected, unprotected and attacked power plants 247
of (4)) – any malicious signal u_a deliberately injected by the attacker 248
might have a serious adverse effect on the behavior of the protected 249
output y . 250

A simple strategy meant to counter the effects of an attack is 251
indeed that of making the protected output y *decoupled* from u_a . 252
As it is well-known, this is achieved if u is chosen as 253

$$u = -B^{-1} H z + v, \quad (13)$$

in which, for convenience, we have set 254

$$B = \text{diag}(b_1, b_2, \dots, b_m), \quad H = \begin{pmatrix} H_1 \\ \dots \\ H_m \end{pmatrix}.$$

The control (13) renders the behavior of ξ , and consequently that 255
of the protected output y , decoupled from z and hence unaffected 256
by the attack. In fact, such control renders the state z unobservable 257
through the output y . Moreover, one could pick the residual control 258
 v in (13) in such a way as to force a prescribed behavior of the ξ_i 's. 259
Setting, for convenience, 260

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

one could pick, to this end, 261

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

so as to impose any prescribed (stable) dynamics on ξ . Note that the 262
composition of (13) and (14) is a control of the form 263

$$u = B^{-1} [-Hz - K\xi + K_0 \xi]. \quad (15)$$

264

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

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

287 4 Robust Defence Against Destabilizing Attacks

288 4.1 The proposed defence strategy

289 We assume in what follows that all the b_i 's are bounded from
 290 below and from above by fixed numbers, that is there exists numbers
 291 $0 < b_{\min} < b_{\max}$ such that

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

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

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

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

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

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

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

297 Let $g_L: \mathbb{R} \rightarrow \mathbb{R}$ be a smooth “saturation” function, that is a
 298 function characterized by the following properties:

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

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

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

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

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

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

305 in which

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

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

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

and the scalar σ_i are states of a dynamical system described by 307
 equations of the form 308

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

In these equations, the coefficient κ and $c_{i,0}, c_{i,1}, \dots, c_{i,r}$ are 309
 design parameters. 310

4.2 Main result 311

The controller proposed in this paper is defined by the couple 312
 of equations (17)–(18). This controller is completely specified by 313
 the set of parameters $B_0, K_0, L, c_{i,0}, c_{i,1}, \dots, c_{i,r}$ and κ . In the 314
 previous subsection, structure and values of B_0 and K_0 have been 315
 specified. In what follows, we will show how the remaining design 316
 parameters can be chosen so as to obtain the desired goal, which – 317
 in a nutshell – is to (practically) decouple the “protected” output y 318
 from the “attacked” set z of state variables, so long as $z(t)$ remains 319
 bounded by a fixed – but otherwise arbitrary – number \bar{M} . 320

More specifically, we will prove that, if the design parameters 321
 are appropriately chosen, the response $\xi(t)$ can be made arbitrarily 322
 close (so long as $z(t)$ remains bounded by \bar{M}) to the response 323
 resulting from the implementation of the “ideal” control (15). In this 324
 respect, observe that, under the effect of the control law (15), one 325
 would obtain for $\xi(t)$ a response 326

$$\xi(t) = \text{col}(\xi_1^{\text{id}}(t), \dots, \xi_m^{\text{id}}(t))$$

in which $\xi_i^{\text{id}}(t)$ is a solution of 327

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

The Proposition that follows (in which we use B_R to denote the 328
 closed ball of radius R and $\hat{\xi}_{\text{ext}} = \text{col}(\hat{\xi}, \sigma)$) is main result of the 329
 paper. 330

Proposition 1. Consider system (12) with control (17)–(18). Let 331
 R and $R^* \gg R$ be fixed. Assume $x(0) \in B_R$ and $\hat{\xi}_{\text{ext}}(0) \in B_R$. 332
 There is a choice of saturation level L and of the design parameters 333
 $c_{i,0}, \dots, c_{i,r}$ such that, for any choice of $\bar{\varepsilon} > 0$ there exists a number 334
 κ^* such that, if $\kappa \geq \kappa^*$, then for all t such that $x(t) \in B_{R^*}$ the 335
 components $\xi_1(t), \dots, \xi_m(t)$ of the response $\xi(t)$ satisfy 336

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

4.3 Proof of the main result: a change of coordinates 337

The arguments used in the proof of the main result are essentially 338
 the same as those used to show that a feedback law of the form (17)– 339
 (18) is able to induce – under the assumption that the zero-dynamics 340
 of the controlled system are asymptotically stable – an input-output 341
 behavior that asymptotically recovers the behavior obtained under 342
 the action of a control of the form (15) (see [21][26]). The novelty 343
 here is that we no longer assume that the zero dynamics are globally 344
 asymptotically stable and we show that those arguments can be used 345
 to prove “practical” decoupling of $\xi(t)$, and hence $y(t)$, from $z(t)$, 346
 so long as the latter remains bounded. 347

In order to analyze the response of the closed-loop system defined 348
 by (12)–(17)–(18), it is useful to make a change of the variables, 349

350 introducing

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

351 Setting

$$\begin{aligned} e &= \text{col}(e_1, \dots, e_m) \\ e_i &= \text{col}(e_{i,1}, e_{i,2}, \dots, e_{i,r_i+1}), \quad i = 1, \dots, m \end{aligned}$$

352 equations (19) define a map

$$\begin{aligned} T : \mathbb{R}^{m(r+1)} &\rightarrow \mathbb{R}^{m(r+1)} \\ \hat{\xi}_{\text{ext}} &\mapsto e = T(z, \xi, \hat{\xi}_{\text{ext}}) \end{aligned} \quad (20)$$

353 As shown in [27, pp.300-301]), the following property holds.

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

356 As consequence, (19) define a legitimate (partial) change of coordi-
357 nates and we can express the closed-loop system in the coordinates
358 $x = \text{col}(z, \xi)$ and e . Note that e is a function of $(x, \hat{\xi}_{\text{ext}})$ and,
359 conversely, that $\hat{\xi}_{\text{ext}}$ is as a function of (x, e) .

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

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

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

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

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

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

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

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

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

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

375

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

in which

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

376 and that

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

in which

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

377

$$\Delta_{2,i}(x, e) = H_i z + K_i \xi + [b_i - b_0]g'_L(\psi_i(\xi, \sigma))b_0^{-1}K_{0,i}\xi_i. \quad (25)$$

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

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

in which $A_{e,i} \in \mathbb{R}^{(r+1) \times (r+1)}$, $B_{e1,i} \in \mathbb{R}^{(r+1)}$, $B_{e2,i} \in \mathbb{R}^{(r+1)}$,
380 $C_{e,i}^T \in \mathbb{R}^{(r_i+1)}$ are matrices defined as

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

382

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

383

$$C_{e,i} = (c_{i,0} \ 0 \ 0 \ \dots \ 0 \ 0).$$

384 Relevant, in the analysis that follows, is the possibility of show-
385 ing that the functions $\Delta_{0,i}(x, e)$, $\Delta_{1,i}(x, e)$, $\Delta_{2,i}(x, e)$, have the
386 following properties (see, in this respect, [27, 304-305]).

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

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

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

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

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

4.4 Proof of the main result: analysis of the response

392

393 We assume in what follows that the initial conditions
394 $(x(0), \hat{\xi}_{\text{ext}}(0))$ of the controlled system are in a fixed bounded set,
395 that is we assume that $x(0) \in B_R$ and $\hat{\xi}_{\text{ext}}(0) \in B_R$, for some
396 $R > 0$. Pick any number $R^* \gg R$ and let $[0, T_{\text{max}}]$ be a time
397 interval such that $x(t) \in B_{R^*}$ for all $t \in [0, T_{\text{max}}]$. We will prove
398 that, if the design parameters are appropriately chosen, on the entire
399 time interval $[0, T_{\text{max}}]$, the states $\xi_i(t)$ remain arbitrarily close to
400 the trajectories of the stable systems $\xi_i = (A_i + B_i K_{0,i})\xi_i$.

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

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

402 Then, observe that, since $G_L(\cdot)$ is bounded by $L(c+1)$, the
403 quantity $\Delta_3(x, e)$ remains bounded so long as $x(t) \in B_{R^*}$, by a
404 bound that does not depend on the design parameter κ (rather, this
405 bound only depends on the choice of R^*). As a consequence, with
406 $x(0) \in B_R$ and $\hat{\xi}_{\text{ext}}(0) \in B_R$, given any arbitrarily small number
407 $0 < \delta \ll (R^* - R)$ 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 $\|\hat{\xi}_{\text{ext}}(0)\| \leq \bar{M}$). 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, [27, pp.308-312]).

Lemma 3. Consider the set of systems

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

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 $c_{i,0}, \dots, c_{i,r}$ 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 $t \in [0, T_{\text{max}}]$ and suppose that $\xi_{\text{ext}}(0) \in B_R$. Then, for every $0 < T_0 \leq T_{\text{max}}$ and every $\varepsilon > 0$, there is a κ^* such that, for all $\kappa \geq \kappa^*$,

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

Lemma 5. Suppose $\|e(t)\| \leq 2\varepsilon$ for all $t \in [T_0, T_{\text{max}}]$. 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 $c_{i,0}, \dots, c_{i,r}$ 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 $[T_0, T_{\text{max}}]$, 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,

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

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

To this end, observe that

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

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

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

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

$$\psi(\hat{\xi}, \sigma) = \psi(\xi, \sigma) - B_0^{-1}K_0D(\kappa)e,$$

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 $[T_0, T_{\text{max}}]$.

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

$$\begin{aligned} \Delta_3(x, e) &= G_L(\psi(\hat{\xi}, \sigma)) - B^{-1}(-Hz - K\xi + K_0\xi) \\ &= \psi(\hat{\xi}, \sigma) - B^{-1}(-Hz - K\xi + K_0\xi) \\ &= \psi(\xi, \sigma) - B_0^{-1}K_0D(\kappa)e - B^{-1}(-Hz - K\xi + K_0\xi). \end{aligned}$$

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

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

in which $\varsigma = \text{col}(e_{1,r+1}, e_{2,r+1}, \dots, e_{m,r+1})$. 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

$$\Delta_3(x, e) = -B_0^{-1}K_0D(\kappa)e + B^{-1}\varsigma.$$

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

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

It is seen from this estimate that on the time interval $[T_0, T_{\text{max}}]$, 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

$$\dot{\xi}_i^{\text{id}} = (A_i + B_iK_{0,i})\xi_i^{\text{id}}.$$

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

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

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 $(A_i + B_iK_{0,i})$ 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

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

and this proves Proposition 1.

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

478 By means of the control law (17)–(18) we are able to practically
479 decouple $\xi(t)$ from $z(t)$ on the finite time interval $[0, T_{\max})$,
480 in which T_{\max} is determined by the bound R^* chosen for $x(t)$,
481 bound which in turn determines the values of the design parameters
482 $L, c_{i,0}, c_{i,1}, \dots, c_{i,r}$ and κ^* . In practice, as kindly pointed out
483 by an anonymous reviewer, this should not be seen as a limitation
484 of the method. In fact, it is reasonable to conceive that in practice
485 $z(t)$ will remain bounded for all t , in which case we may well
486 take $T_{\max} = \infty$. As a matter of fact, the attack on $z(t)$ takes place
487 through actuators that have saturations. This means that, although
488 the goal of the attacker is to make the zero dynamics anti-stable,
489 $z(t)$ will only diverge so long as the actuators are not saturated. In
490 the end, if the un-attacked zero dynamics is stable, a consequence of
491 a saturated attack is a bounded attacked $z(t)$. Another reason why
492 one can consider $T_{\max} = \infty$ is that, since the attacked generators
493 have automatic protections, if $z(t)$ gets excessively large then such
494 generators at some time will be automatically disconnected from the
495 network. In this case, our defence strategy is effective on the finite
496 time interval during which the attacked generators are connected and
497 will be no longer needed after the disconnection of such generators.

498 5 Simulation Results

499 5.1 Case study

500 In order to validate the proposed decoupling strategy, the WSCC
501 9-bus test power network reported in Fig. 1 has been considered
502 [24]; it represents an approximation of the Western System
503 Coordinating Council (WSCC) to an equivalent system with
504 3 generation buses and 6 load buses [28]. For this power system
505 the inertia and damping matrices characterizing the power
506 plants are respectively $M = \text{diag}(0.125, 0.034, 0.016)$ and $D =$
507 $\text{diag}(0.125, 0.068, 0.048)$, while the interconnections within the
508 network are characterized by the laplacian matrix indicated in
509 eq. (33).

510 In the considered test case the power plant 1 is unprotected, the
511 power plant 2 is the one used for launching the attack, while the
512 power plant 3 is the one to be protected using the proposed decoupling
513 approach. With reference to the nomenclature used in section 3
514 (see in particular eq. (9)), in this case $n = 6, m = 1, r = 2$; the
515 controls used by the attacker and by the defender are $u_a = P_{g2}$
516 and, respectively, $u = P_{g3}$ and the protected output is $y = \delta_3$ (recall
517 that $\omega_3 = \dot{\delta}_3$). The simulations reported in the following have been
518 performed using Simulink.

519 The attack control has been chosen so as to force an anti-stable
520 zero dynamics characterized by four identical positive real eigen-
521 values with time constant $\tau = 20$ s. The attack is triggered by a
522 small initial condition on the angular speed of machine 2 ($\omega_2(0) =$
523 1.3×10^{-2} rad/s), and all the machines are affected by the attack
524 due to their mutual coupling.

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

533 5.2 Protected case

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

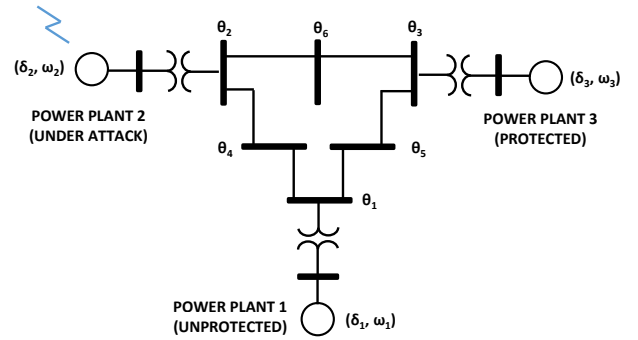


Fig. 1: The WSCC 9-bus power network.

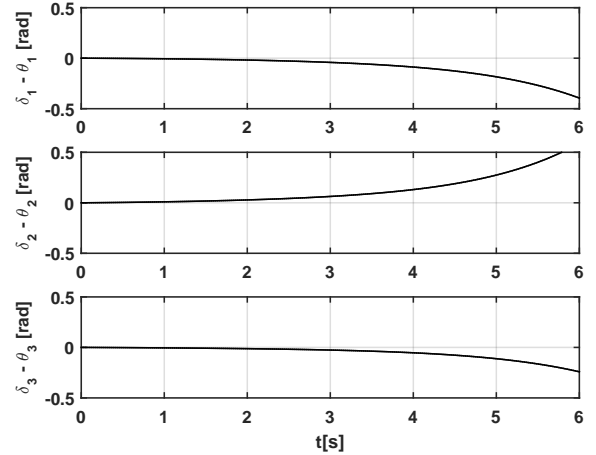


Fig. 2: Differences among angles at the interconnection of generators to load buses, in absence of defence.

$$L_N = \begin{pmatrix} 0.058 & 0 & 0 & -0.058 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.063 & 0 & 0 & -0.063 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.059 & 0 & 0 & -0.059 & 0 & 0 & 0 \\ -0.058 & 0 & 0 & 0.235 & 0 & 0 & -0.085 & -0.092 & 0 \\ 0 & -0.063 & 0 & 0 & 0.296 & 0 & -0.161 & 0 & -0.072 \\ 0 & 0 & -0.059 & 0 & 0 & 0.330 & 0 & -0.170 & -0.101 \\ 0 & 0 & 0 & -0.085 & -0.161 & 0 & 0.246 & 0 & 0 \\ 0 & 0 & 0 & -0.092 & 0 & -0.170 & 0 & 0.262 & 0 \\ 0 & 0 & 0 & 0 & -0.072 & -0.101 & 0 & 0 & 0.173 \end{pmatrix} \quad (33)$$

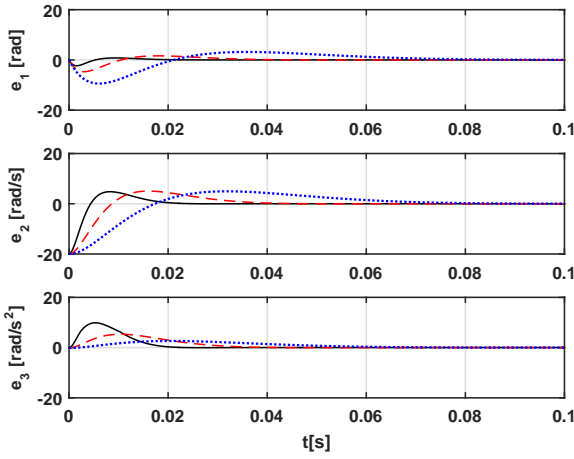


Fig. 3: Components of the extended observer's error for $\kappa = 50$ (blue dotted line), 100 (red dashed line), 200 (black solid line).

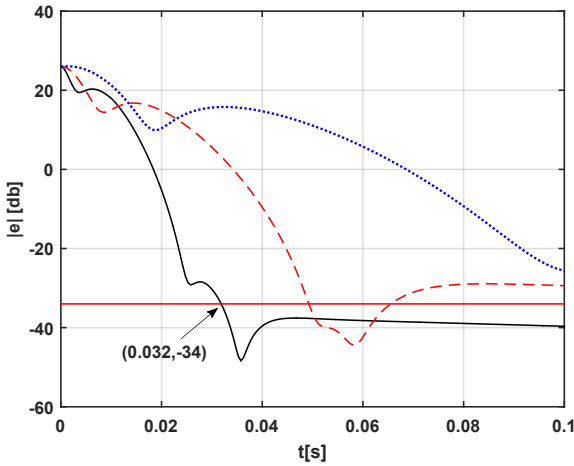


Fig. 4: Absolute value of the extended observer's error for $\kappa = 50$ (blue dotted line), 100 (red dashed line), 200 (black solid line), expressed in dB to properly notice the residual magnitude of the error. The horizontal red solid line represents the threshold 2ϵ (-34 dB).

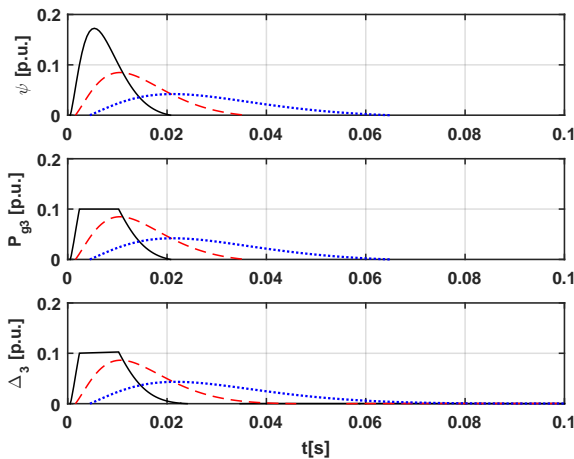


Fig. 5: Unsaturated control, actual robust control and deviation from the ideal decoupling control for $\kappa = 50$ (blue dotted line), 100 (red dashed line), 200 (black solid line).

541 As far as the robust controller is concerned, at first the matrix
 542 B_0 (reduced here to a single scalar coefficient b_0) has been set to
 543 $0.8b = 0.8M_3^{-1}$. Then, according to Lemma 3, the choice $c_0 =$
 544 $6, c_1 = 11, c_2 = 6$ has been performed, which guarantees asymptotic
 545 stability of the observer's error dynamics resulting from the
 546 robust control.

547 Fig. 3 and 4 report respectively the evolution of the error components
 548 and its norm for three different values of the gain ($\kappa \in$
 549 $\{50, 100, 200\}$): the higher is κ , the shorter are the transient and
 550 the residual magnitude of the error. As a matter of fact it is seen that
 551 if the value $\epsilon = 0.01$ is chosen, the value $\kappa = 200$ guarantees that
 552 the error's norm reduces to and remains below the threshold 2ϵ (-34
 553 dB) approximately after $T_0 = 32$ ms.

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

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

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

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

6 Conclusions

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

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

606 Motivated by the need of extending the rotor angles and speeds
 607 operational range in which the decoupling is required to be effective,
 608 a future direction for this research stream considers the design of a
 609 robust decoupling control in the context of the nonlinear-descriptor
 610 representation of power systems.
 611

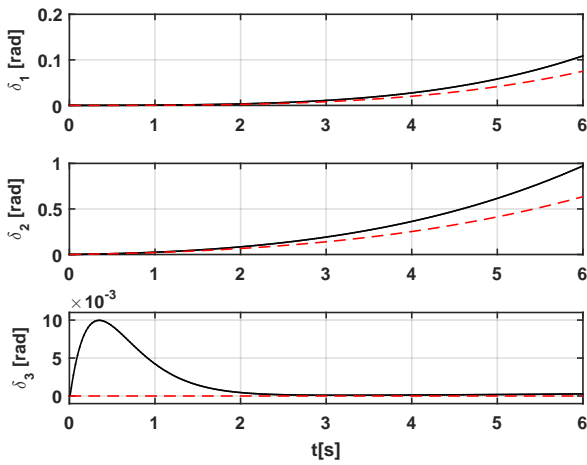


Fig. 6: Rotor angles dynamics in case of robust control (black solid line) and ideal control (red dashed line).

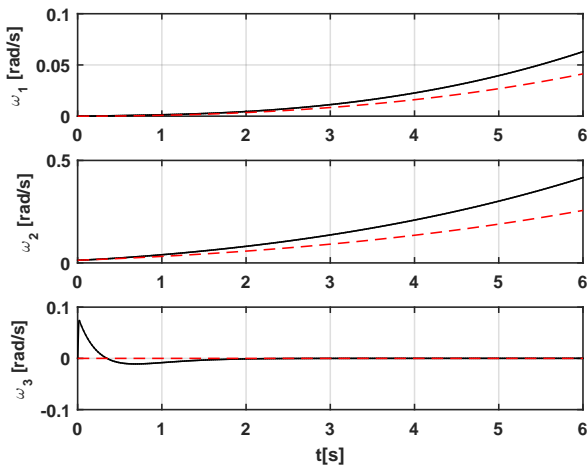


Fig. 7: Angular speeds dynamics in case of robust control (black solid line) and ideal control (red dashed line).

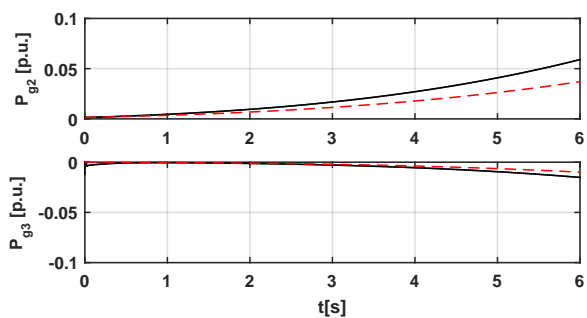


Fig. 8: Attack and defence controls in case of robust control (black solid line) and ideal control (red dashed line).

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.

8 References

- 1 Slay, J., Miller, M. ‘Lessons learned from the maroochy water breach’. In: International Conference on Critical Infrastructure Protection. (Springer), 2007. pp. 73–82
- 2 Kuvshinkova, S.: ‘Sql slammer worm lessons learned for consideration by the electricity sector’, *North American Electric Reliability Council*, 2003, **1**, (2), pp. 5
- 3 Sun, C.C., Liu, C.C., Xie, J.: ‘Cyber-physical system security of a power grid: State-of-the-art’, *Electronics*, 2016, **5**, (3), pp. 40
- 4 Sandberg, H., Amin, S., Johansson, K.H.: ‘Cyberphysical security in networked control systems: An introduction to the issue’, *IEEE Control Systems*, 2015, **35**, (1), pp. 20–23
- 5 Cheng, P., Shi, L., Sinopoli, B.: ‘Guest editorial special issue on secure control of cyber-physical systems’, *IEEE Transactions on Control of Network Systems*, 2017, **4**, (1), pp. 1–3
- 6 Amin, S., Cárdenas, A.A., Sastry, S.S. ‘Safe and secure networked control systems under denial-of-service attacks’. In: International Workshop on Hybrid Systems: Computation and Control. (Springer), 2009. pp. 31–45
- 7 Guo, Z., Shi, D., Johansson, K.H., Shi, L.: ‘Optimal linear cyber-attack on remote state estimation’, *IEEE Transactions on Control of Network Systems*, 2017, **4**, (1), pp. 4–13
- 8 Mo, Y., Sinopoli, B. ‘Secure control against replay attacks’. In: 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton). (IEEE), 2009. pp. 911–918
- 9 Zhao, J., Wang, J., Yin, L. ‘Detection and control against replay attacks in smart grid’. In: Computational Intelligence and Security (CIS), 2016 12th International Conference on. (IEEE), 2016. pp. 624–627
- 10 Liu, Y., Ning, P., Reiter, M.K.: ‘False data injection attacks against state estimation in electric power grids’, *ACM Transactions on Information and System Security (TISSEC)*, 2011, **14**, (1), pp. 13
- 11 Lucia, W., Sinopoli, B., Franze, G. ‘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. (IEEE), 2016. pp. 1–5
- 12 Kwon, C., Hwang, I.: ‘Cyber attack mitigation for cyber-physical systems: hybrid system approach to controller design’, *IET Control Theory & Applications*, 2016, **10**, (7), pp. 731–741
- 13 Teixeira, A., Shames, I., Sandberg, H., Johansson, K.H. ‘Revealing stealthy attacks in control systems’. In: 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton). (IEEE), 2012. pp. 1806–1813
- 14 Teixeira, A., Shames, I., Sandberg, H., Johansson, K.H.: ‘A secure control framework for resource-limited adversaries’, *Automatica*, 2015, **51**, pp. 135 – 148
- 15 Pasqualetti, F., Dorfler, F., Bullo, F.: ‘Attack detection and identification in cyber-physical systems’, *IEEE Transactions on Automatic Control*, 2013, **58**, (11), pp. 2700–2714

- 678 pp. 2715–2729
- 679 16 Park, G., Shim, H., Lee, C., Eun, Y., Johansson, K.H.
680 ‘When adversary encounters uncertain cyber-physical
681 systems: Robust zero-dynamics attack with disclosure
682 resources’. In: Decision and Control (CDC), 2016 IEEE
683 55th Conference on. (IEEE), 2016. pp. 5085–5090
- 684 17 Shim, H., Park, G., Joo, Y., Back, J., Jo, N.H.: ‘Yet
685 another tutorial of disturbance observer: robust stabi-
686 lization and recovery of nominal performance’, *Control
687 Theory and Technology*, 2016, **14**, (3), pp. 237–249
- 688 18 Keller, J.Y., Sauter, D. ‘Monitoring of stealthy attack
689 in networked control systems’. In: Control and Fault-
690 Tolerant Systems (SysTol), 2013 Conference on. (IEEE),
691 2013. pp. 462–467
- 692 19 Ao, W., Song, Y., Wen, C.: ‘Adaptive cyber-physical sys-
693 tem attack detection and reconstruction with application
694 to power systems’, *IET Control Theory and Applications*,
695 2016, **10**, (12), pp. 1458–1468
- 696 20 Hoehn, A., Zhang, P. ‘Detection of covert attacks and
697 zero dynamics attacks in cyber-physical systems’. In:
698 American Control Conference (ACC), 2016. (IEEE),
699 2016. pp. 302–307
- 700 21 Freidovich, L.B., Khalil, H.K.: ‘Performance recovery
701 of feedback-linearization-based designs’, *IEEE Transac-
702 tions on Automatic Control*, 2008, **53**, (10), pp. 2324–
703 2334
- 704 22 Kundur, P., Balu, N.J., Lauby, M.G.: ‘Power system sta-
705 bility and control.’. vol. 7. (McGraw-Hill New York,
706 1994)
- 707 23 Machowski, J., Bialek, J., Bumby, J.: ‘Power system
708 dynamics: stability and control.’. (Second edition. John
709 Wiley & Sons, 2011)
- 710 24 Pasqualetti, F., Bicchi, A., Bullo, F. ‘A graph-theoretical
711 characterization of power network vulnerabilities’. In:
712 American Control Conference ACC. (IEEE), 2011.
713 pp. 3918–3923
- 714 25 Wonham, W.M.: ‘Linear Multivariable Control: a Geo-
715 metric Approach (3rd ed.)’. (Springer Verlag, 1979)
- 716 26 Wang, L., Isidori, A., Su, H.: ‘Output feedback sta-
717 bilization of nonlinear mimo systems having uncertain
718 high-frequency gain matrix’, *Systems & Control Letters*,
719 2015, **83**, pp. 1 – 8
- 720 27 Isidori, A.: ‘Lectures in Feedback Design for Multivari-
721 able Systems’. (Springer Verlag, 2016)
- 722 28 AL Hinai, A. ‘Voltage collapse prediction for intercon-
723 nected power systems’. Master degree thesis in Electrical
724 Engineering, West Virginia University, 2000