IET Cyber-Physical Systems: Theory & Applications

Special Issue: Cyber Physical Power Systems: Advanced Intelligent Technologies and Applications

Developing correlation indices to identify coordinated cyber-attacks on power grids

ISSN 2398-3396
Received on 29th March 2018
Revised 15th September 2018
Accepted on 11th October 2018
E-First on 11th December 2018
doi: 10.1049/iet-cps.2018.5002
www.ietdl.org

Christian Moya[1] ✉, *Jiankang Wang*[1]

[1]*Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA*

✉ *E-mail: moyacalderon.1@osu.edu*

Abstract: Increasing reliance on Information and Communication Technology exposes the power grid to cyber-attacks. In particular, Coordinated Cyber-Attacks (CCAs) are considered highly threatening and difficult to defend against, because they (i) possess higher disruptiveness by integrating greater resources from multiple attack entities, and (ii) present heterogeneous traits in cyber-space and the physical grid by hitting multiple targets to achieve the attack goal. Thus, and as opposed to independent attacks, whose severity is limited by the power grid's redundancy, CCAs could inflict disastrous consequences, such as blackouts. In this study, the authors propose a method to develop Correlation Indices to defend against CCAs on static control applications. These proposed indices relate the targets of CCAs with attack goals on the power grid. Compared to related works, the proposed indices present the benefits of deployment simplicity and are capable of detecting more sophisticated attacks, such as measurement attacks. The method is demonstrated using measurement attacks against Security Constrained Economic Dispatch.

1 Introduction

The operation of today's power grid largely relies on automated control applications and Supervisory Control and Data Acquisition (SCADA) systems. While control applications compute the commands to operate the power grid, SCADA serves as the channel between control applications and field devices [1] by transmitting measurement and control signals. The desire to improve the efficiency and reliability of control applications and SCADA has led to the use of heterogeneous and non-proprietary information and communication technology (ICT) [2]. However, this heterogeneous and non-proprietary ICT increases the number of cyber-vulnerabilities, opening up a much wider scope of cyber-security concerns among utilities.

By exploiting cyber-vulnerabilities, malicious adversaries can launch cyber-attacks against control applications and SCADA, among which *Coordinated Cyber-Attacks* (*CCA*) are considered highly threatening and difficult to defend against. This is because CCAs (i) possess higher descriptiveness by integrating resources from multiple attack entities, and (ii) present heterogeneous traits in cyber-space and the physical power grid by hitting multiple targets to achieve the attack goal.

Thus, and as opposed to regular (or independent) cyber-attacks, whose severity is limited by the power grid's redundancy, CCAs could (i) inflict catastrophic consequences and (ii) be very challenging to detect in real-time. CCAs could inflict catastrophic consequences as exemplified by the cyber-attacks to the Ukrainian power grid (the 'BlackEnergy' malware attack in 2015 [3, 4], and the 'Crash Override' attack in 2016 [5]). These CCAs disconnected multiple substations that triggered power outages, leaving thousands of consumers and facilities without electricity. On the other hand, CCAs are challenging to detect in real-time due to the invisibility of attack goals, which are formed by and thus concealed by CCAs over space and time.

Intrusion Detection Systems (IDSs) are necessary tools to protect control applications and SCADA against cyber-attacks. IDSs record and analyse cyber-traces from adversaries that breach into the grid's cyber-system to exploit vulnerabilities. If, after analysing cyber-traces, the security of the grid appears to be compromised, then IDSs will generate alarms. In addition, some IDSs will also take action to mitigate attacks' effect. While IDSs can detect regular attacks or individual components of CCAs, they suffer from false alarms, fail to identify CCAs, and cannot estimate the attack consequences on the grid.

To identify CCAs and estimate attack consequences, recent works suggest integrating intrusion data from IDSs with attack templates –*attack templates* model cyber-attacks against control applications. This integration results in a set of *Correlation Indices* (CIs) describing the temporal and/or spatial correlation of coordinated attacks.

1.1 Related works

Many CIs have been proposed in the literature; however, they differ in their principles, which we summarise below.

1.1.1 CIs based on adversaries' cyber traces: Attack sequences of the same adversary have similar cyber-traces that can be identified as contributing to CCAs. IDSs use this detection principle to investigate the temporal correlation of intrusions in cyber-space. Anomaly matrices [6] and time failure propagation graphs [4] are proposed to relate intrusion time with intrusion actions. While capable of detecting CCAs at the cyber-space, CIs of this type fail to estimate the attack consequences on the grid.

1.1.2 CIs based on cyber-physical dependence: Logic graphs describing the conditions (in sequence in the cyber-space) for a physical consequence to take place can be used to derive CIs [7]. The logic graphs can take forms of attack trees [8], attack graphs [9], and PetriNets [10]. Temporal correlation of attacks is derived not only in the cyber-space but also in the physical power grid (see Fig. 2 in [11] for an example). However, constructing these logic graphs requires great computational effort due to a large number of cyber and physical components.

1.1.3 CIs based on attack goals on the physical grid: Adversaries' goals described with reliability metrics or in terms of the criticality of a certain target are used to derive CIs. For example, in [4], substations are attack targets and their criticality is first ranked. In [12], the attack goal is modelled as causing an insufficient power transfer. The work takes a numerical approach by disconnecting a set of substations at one time and running power flow. The substations in the set are identified as correlated if the power flow is divergent.

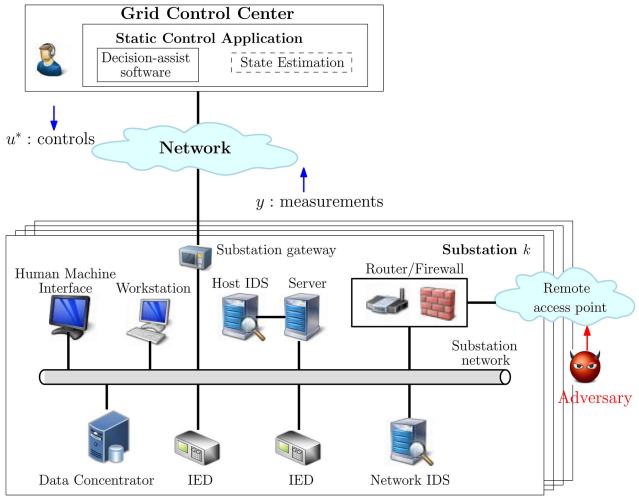

Fig. 1 *Static control application. At the substation s_k, we illustrate its ICT, including IDSs*

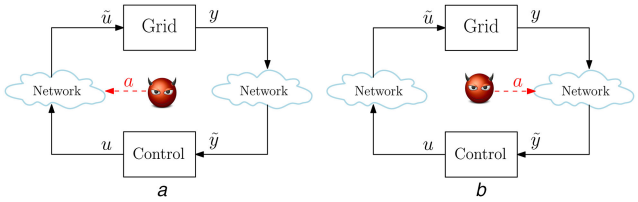

Fig. 2 *Control and measurement attacks. u: control command, y: measurements. $u \neq \tilde{u}$ ($y \neq \tilde{y}$) during the cyber-attack*
(a) Control attack, ***(b)*** Measurement attack

Given the great size of power grids, the combined deployment of CIs based on cyber-traces and attack goals promises better computation performance and higher accuracy than the CIs based on cyber-physical dependence. The existing CIs based on attack goals, however, are limited to a few goals achieved by corrupting control commands. Other cyber-attacks, such as measurement attacks, present much higher threats in coordination (as a rich body of the literature has shown their impact in electricity markets and security constrained power flows [13–15]). This is because measurement attacks are (i) difficult to detect by hiding in measurement signals and deceiving through control applications, and (ii) capable of inflicting disastrous consequences by coordinating attacks against multiple grid components.

1.2 Our work

This paper proposes a method to derive CIs based on attack goals for the following attack template: measurement attacks against Security Constrained Economic Dispatch (SCED). In particular, we make the following contributions:

i. An analytical method to derive CIs. We formulate the attack template as a bilevel mix-integer optimisation program. This problem is challenging due to its non-convex and combinatorial nature. To address these challenges, we propose an algorithm that computes the CIs based on attack goals.
ii. A collection of set-theoretic properties for the CIs. These properties relate attack goals to the targets of CCAs.
iii. Defence strategies against CCAs, a metric of defence effectiveness, and the application of CIs to identify CCAs.

Though we present our method to derive CIs for SCED, we emphasise that our method can be extended to other static control applications.

The remaining of the paper is organised as follows. Section 2 reviews the concepts of static applications and attack templates. The mathematical models of SCED and the attack template in bilevel form are presented in Sections 3 and 4, respectively. The CIs are derived in Section 5. Section 6 describes the CIs' properties, defence strategies, the metric of defence effectiveness, and the application of CIs to identify CCAs. In Section 7, the CIs are demonstrated with numerical experiments. Finally, Section 8 concludes the paper.

2 Background

In this section, we review the concepts of static applications and attack templates.

2.1 Static control applications

Static applications are control loops designed to monitor, supervise, and control the grid's operating point – i.e. they ignore the dynamics and work with the grid at a quasi-steady state. These applications can be automated or executed by a human operator. Examples include SCED, Optimal Reactive Power Support, and dispatch in Electricity Markets.

Fig. 1 illustrates a schematic diagram of a static application. These applications compute control commands by solving optimisation algorithms or by allowing direct manipulation via a human–machine interface. In any case, the control commands are computed based on measurements collected at remote substations. To verify the integrity of these measurements, the most well-known applications implement state estimation and bad data detection.

2.2 Attack templates

Attack templates describe models of cyber-attacks on control applications. We consider two CCAs: control and measurement attacks. In control attacks [16], adversaries *coordinately* corrupt or hijack multiple control devices and directly modify control commands (Fig. 2*a*). This class of attack is essentially the same as physical attacks, in the sense that the grid's configuration is altered by control signals in a way similar to mechanical operation. In measurement attacks [16], adversaries *coordinately* contaminate or falsify measurements at multiple substations to manipulate decision-making processes (i.e. control applications) (Fig. 2*b*). Since remote substations collect the measurements and operate physical control devices (e.g. circuit breakers or capacitors), we assume that coordinated control and/or measurement attacks are executed by hacking into multiple remote substations.

Attack templates have been used in the literature to determine the consequences of cyber-attacks, identify critical components of the grid, derive defence strategies and so on. For instance, by studying the attack template of measurement attacks on state estimation, several authors have proposed to stop the attacks by enhancing the screening methods of state estimation [17]. This defence strategy, however, fails if the static application does not have state estimation, which is often true for real-time and contingency dispatch.

In this paper, we consider the following attack template: measurement attacks against SCED. This attack template describes an adversary with the following characteristics:

i. The adversary knows the models of the power grid and SCED.
ii. The adversary can hack into the substations' ICT and inject falsified measurements to manipulate SCED.
iii. The adversary can coordinate the attack against multiple substations over a large geographic area – i.e. launch CCAs.

We use the attack template to derive CIs. These CIs describe a relation between the target substations and the attack goal (Fig. 3).

Remark 1: The adversary's characteristics might be restrictive. However, they were selected for convenience of CIs' development and can be relaxed at the expense of more involved computations. For example, to relax the first characteristic, existing studies [18, 19] developed stochastic methods to launch attacks with limited information. Other studies [20, 21] presented methods for estimating the power grid model with region-constrained information from multiple adversaries. The stochastic and

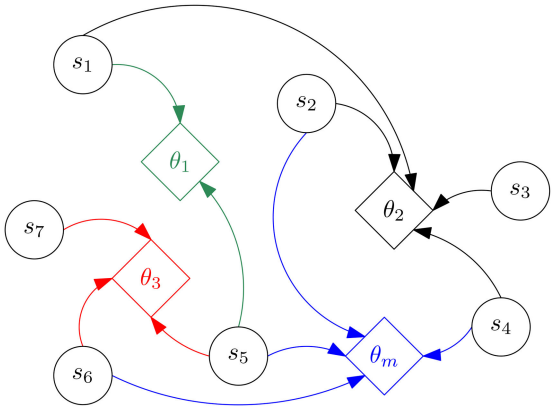

Fig. 3 *Relation graphs between n_s targets (s_k) and m attack goals (θ_i). For example, the targets associated to θ_3 are $\{s_5, s_6, s_{n_s}\}$, and to θ_m are $\{s_2, s_4, s_5, s_6\}$*

estimation methods can easily be applied to extend the CIs' development method in future studies.

3 Mathematical models

In this section, we describe the models of the power grid and SCED.

3.1 Mathematical notation

Throughout this paper, we use the following notation. Let $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ (resp. $\mathbb{R}_{>0}$) denote the set of real numbers and non-negative (resp. positive) real numbers. For $n > 1$, $\boldsymbol{I}_n$ denotes the n-dimensional identity matrix. $\mathbf{1}$ and $\mathbf{0}$ denote, respectively, the vectors (or matrices) with all components equal to one and zero. Given a finite set V, we let $|V|$ denote its cardinality, i.e. the number of elements of V, and 2^V the power set of V, i.e. the set of all subsets of V.

For a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$, $[A]_i$ and $[A]_{ij}$ denote its ith row and its (i, j)th elements. Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, x_i denotes the ith element, $(\boldsymbol{x})$ is the diagonal matrix of $\boldsymbol{x}$, and $\| x \|_0$ is the zero norm of $\boldsymbol{x}$, i.e. the number of non-zero elements of $\boldsymbol{x}$. We let $\| x \|_\infty$ denote the infinity norm defined as $\| x \|_\infty := \max\{|x_i|\}$. For two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $x \circ y = z \in R^n$ denotes the Hadamard or element-wise product, i.e. $z_i = x_i y_i$, and $x \preceq y$ denotes the element-wise inequality, i.e. $x_i \leq y_i$.

3.2 Power grid modelling

We model the power grid as the graph $G = (V, E)$, where V and $E \subset V \times V$ are the sets of $n := |V|$ buses and $m := |E|$ transmission lines. To each bus $i \in V$, we associate the generation $P_{g,i} \in \mathbb{R}_{\geq 0}$, and the demand $P_{d,i} \in \mathbb{R}_{\geq 0}$; to each transmission line $e := (i, j) \in E$, connecting buses $i, j \in V$, we associate the power flow $P_{f,e} \in \mathbb{R}$. In vector form, the generation, demand, and power flows are, respectively, $P_g = [P_{g,1}, \ldots, P_{g,n}]^\top$, $P_d = [P_{d,1}, \ldots, P_{d,n}]^\top$, and $P_f = [P_{f,1}, \ldots, P_{f,m}]^\top$.

In addition, we assume the grid has a set of n_s substations, i.e. $S = \{s_1, s_2, \ldots, s_{n_s}\}$. At the substation s_k, we represent the grid within its service area as the sub-graph $G_{s_k} = (V_{s_k}, E_{s_k})$ with the following properties:

i. Substation service areas compose the entire power grid, i.e. $\cup_{s_k \in S} G_{s_k} = G$.
ii. Substation service areas may overlap, i.e. for some $s_k, s_l \in S$, we may have $G_{s_k} \cap G_{s_l} \neq \emptyset$, but the overlapped areas do not have buses with generation.
iii. Each substation collects demand measurements, denoted as $\tilde{P}_d \in \mathbb{R}^n$, within its service area.

3.3 Security constrained economic dispatch

We consider a SCED problem that computes a new generation profile P_g^* based on demand measurements $\tilde{P}_d$.

The SCED problem is formulated based on the power flow equations. The power flow equations are the mathematical model to plan, operate, and analyse the power grid. They describe how generation and demand balance, and how active and reactive power flows through the grid.

For large-scale power grids, however, the coupled active and reactive power flow models might become computationally expensive and even unfeasible. Thus, a decoupled (DC) power flow might be the only viable alternative to solve large-scale problems. DC power flow is simpler and more robust due to sparsity and linearity, but it is only accurate close to the operating point [22]. We refer the interested reader to [23, 24] for more information on how utilities use DC power flow.

We formulate SCED (based on DC power flow) as a convex optimisation problem that minimises the total generation cost (1a) subject to the following security constraints: generation–demand balance (1b), operation limits of the generators (1c), and transmission limits on power flows (1d), i.e.

$$\min_{P_g} \quad \frac{1}{2} P_g^\top C_2 P_g + c_1^\top P_g + c_0, \tag{1a}$$

$$\text{s.t.} \quad \mathbf{1}^\top P_g - \mathbf{1}^\top \tilde{P}_d = 0, \tag{1b}$$

$$P_g \in [\mathbf{0}, \bar{P}_g], \tag{1c}$$

$$\underbrace{F(P_g - \tilde{P}_d)}_{=: P_f} \in [-\bar{P}_f, \bar{P}_f], \tag{1d}$$

where $c_2, c_1, c_0 \in \mathbb{R}^n_{\geq 0}$ are the cost coefficients for a generation, $C_2 = (c_2)$, $\bar{P}_g \in \mathbb{R}^n_{\geq 0}$ is the rated power from generators, $\bar{P}_f \in \mathbb{R}^m_{\geq 0}$ is the thermal capacity of transmission lines, and $\boldsymbol{F}$ is the generator shift matrix.

4 Attack template

In this section, we describe the attack template in bilevel form. The attack template models measurement attacks against SCED. We also describe the attack goal and constraints.

4.1 Measurement attacks

Let $a \in \mathbb{R}^n$ denote the attack signal. The adversary fabricates a to corrupt measurements of the demand as follows:

$$\tilde{P}_d(a) = P_d + a\,. \tag{2}$$

We assume the adversary injects a by hacking into substations and altering measurements at the data concentrator (or at a communication link via a man-in-the-middle attack). Thus, in the rest of the paper, we refer the target data concentrator and ICT within the substation as the target substation.

4.2 Attack goal

Using the corrupted measurements (2), the adversary has the following attack goal: to manipulate SCED and increase the power flow on a single target line $e \in E$, which occurs at

$$\left|P_{f,e}(a)\right| = \left|[F]_e(P_g^*(a) - P_d)\right| \geq (1 + \tau)\left|P_{f,e}(0)\right|, \tag{3}$$

where $P_{f,e}(a) \in \mathbb{R}$ (resp. $P_{f,e}(0) \in \mathbb{R}$) denotes the power flow on e after (resp. before) the attack, $P_g^*(a) \in \mathbb{R}^n$ denotes the new (after the attack) generation profile, and $\tau \in (0, \bar{\tau}] \subseteq \mathbb{R}_{>0}$ quantifies the *flow increase*.

We use the notation $(e, \bar{\tau}) \in E \times (0, \bar{\tau}]$ to describe attack goals satisfying (3). Since $\tau \in (0, \bar{\tau}]$, we can have (in theory) an infinite

number of attack goals. In practice, however, we study a finite number of attack goals τ. For example, the attack goal τ that will cause congestion (relating to economic loss), overloading (increasing long-term capital cost by accelerating asset depreciation, increasing losses), and loss of transmission lines (under very stressful operating condition). Thus, in the worst case scenario, we assume the adversary maximises the flow increase τ.

In the SCED example, based on the attack goals, the target lines are selected differently. For example, a line connected to a critical generator can be selected if the adversary aims to destabilise the system under heavily loaded condition. Similarly, a line/lines can also be selected to cause congestion (surrounding a load area) and induce market power. As a result, adversaries can deprive profit from generation assets outside or inside the load area. (In the latter case, the electricity market has a power mitigation procedure [25]).

4.3 Attack constraints

The attack might be constrained due to the following:

i. State estimation and bad data detection.
ii. Corruptible measurements and defence at substations.
iii. Attack resources.

Since SCED has state estimation, the adversary must design the attack signal a to bypass bad data detection. Other applications, however, might not have state estimation, and hence the attack signal a can take any (realistic) value. In any case, we write this constraint as $\| a \|_\infty \leq \bar{a}$ where $\bar{a} > 0$. We can use $\bar{a}$ as a design parameter to model different attack scenarios.

If the defender protects substation $s_k \in S$, then the adversary cannot corrupt measurements at s_k; otherwise, the adversary can corrupt all the measurements. We write this constraint as

$$a_i \in \delta_{s_k}[-\bar{a}, \bar{a}], \quad \forall i \in V_{s_k},\ \forall s_k \in S, \delta_{s_k} \in \{0, 1\}, \tag{4}$$

where $\delta_{s_k} = 1$ if the adversary attacks s_k, and $\delta_{s_k} = 0$ if not. The vector $\boldsymbol{\delta}(e, \tau) = [\delta_{s_1}, \delta_{s_2}, \ldots, \delta_{s_{n_s}}]^\top$ describes the safe and target substations during CCAs with an attack goal (e, τ).

If the adversary has limited resources, then (s)he must limit the number of target substations. We write this constraint as

$$\| \delta(e, \tau) \|_0 \leq \kappa, \tag{5}$$

where $\kappa \in \{1, 2, \ldots, n_s\}$ denotes the maximum number of target substations. In the worst-case scenario, the adversary minimises κ.

Remark 2: Note that in the worst-case scenario the adversary faces two conflicting objectives: maximise τ and minimise κ. The interaction $\tau - \kappa$ generates a Pareto-like behaviour between aimed flow increase (τ) and the number of target substations (κ).

4.4 Attack template in the bilevel form

We use bilevel optimisation to model the attack template, describing the worst-case scenario of measurement attacks against SCED. Since *bilevel optimisation* models decision making among agents [26] (e.g. adversary versus defender), researchers have used it to study cyber-attacks [27, 28]; or physical attacks [29] to power grids.

We write the attack template in bilevel form as follows:

$$\begin{aligned} \max_{\tau, \kappa, \delta, a} \quad & \tau - \kappa, \\ \text{s.t.} \quad & (3) - (5), \end{aligned} \tag{6}$$

where $P_g^*(a)$ denotes the optimal solution of the SCED optimisation algorithm, parametrised by the attack signal a, i.e.

$$\begin{aligned} P_{\mathrm{g}}^*(a) \in \arg\min_{P_{\mathrm{g}}} \quad & \frac{1}{2} P_{\mathrm{g}}^\top C_2 P_{\mathrm{g}} + c_1^\top P_{\mathrm{g}} + c_0, \\ \text{s.t.} \quad & \mathbf{1}^\top P_{\mathrm{g}} - \mathbf{1}^\top (P_{\mathrm{d}} + a) = 0, \\ A_0 \quad & P_{\mathrm{g}} + A_1 a - b \preceq 0, \end{aligned} \tag{7}$$

with

$$A_0 := \begin{bmatrix} -I_n \\ I_n \\ F \\ -F \end{bmatrix}, \quad A_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -F \\ F \end{bmatrix}, \quad b = \begin{bmatrix} \mathbf{0} \\ \bar{P}_{\mathrm{g}} \\ \bar{P}_{\mathrm{f}} \\ -\bar{P}_{\mathrm{f}} \end{bmatrix} - A_1 P_{\mathrm{d}}.$$

In the above, the upper level problem (6) models the attack goal and constraints, while the lower level problem (7) models the SCED manipulated through corrupted measurements (a).

The optimal solution of the bilevel form $(\tau^*, \kappa^*, \delta^*, a^*, P_{\mathrm{g}}^*)$, if it exists, describes an adversary that targets the least number of substations (κ^* and δ^*) and maximises the flow increase (τ^*) on the single line $e \in E$.

The bilevel form (6) and (7) depends on several parameters, including the power grid parameters, the SCED parameters, and the maximum value for the attack signal $\bar{a}$. Thus, a defender, using the attack template, can select the parameters to study different scenarios.

Remark 3: By defining the corresponding attack goal, constraints, and control algorithm, we can model measurement attacks against other static applications, using the attack template in bilevel form. In addition, we can model control attacks using the upper level problem (6).

5 Deriving the CIs

In this section, we derive the key concepts, CIs and security index. We obtain the indices by transforming the attack template in bilevel form into a Mathematical Program with Equilibrium Constraints (MPEC) and addressing its mathematical challenges.

5.1 Attack template in mathematical programming form

Since the lower level problem (7) is strictly convex on P_g for a fixed a, its Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality [30]. So, we can write the bilevel form (6) and (7) as an MPEC (i.e. a single-level optimisation problem [31]), by replacing (7) with the KKT conditions. This yields

$$\max_{\tau, \kappa, \delta, a, P_{\mathrm{g}}^*} \quad \tau - \kappa, \tag{8a}$$

$$\text{s.t.} \quad (3) - (5), \tag{8b}$$

$$\mathbf{1}^\top P_{\mathrm{g}}^* - \mathbf{1}^\top (a + P_{\mathrm{d}}) = 0, \tag{8c}$$

$$C_2 P_{\mathrm{g}}^* + c_1 - \mathbf{1}\nu^* + A_0^\top \lambda^* = 0, \tag{8d}$$

$$A_0 P_{\mathrm{g}}^* + A_1 a - b \preceq 0, \tag{8e}$$

$$\lambda^* \succeq 0, \tag{8f}$$

$$\lambda^* \circ (A_0 P_{\mathrm{g}}^* + A_1 a - b) = 0. \tag{8g}$$

In the above, (8c) and (8g) are the KKT conditions of (7) and ν^* (resp. λ^*) denotes the Lagrange multiplier of the equality (resp. inequality) constraint of (7).

5.2 Mathematical challenges

The MPEC (8) is a challenging problem. Its properties are far more complex than the properties of traditional mathematical

1: $(\kappa^*, \text{CIs}) \leftarrow \text{CorrelationIndices}(e, \tilde{\tau})$
2: **procedure** CORRELATIONINDICES$(e, \tilde{\tau})$
3: CIs $\leftarrow \{\emptyset\}$
4: Compute κ^* by solving (11)
5: **for** $j = 1$ to $\binom{\kappa^*}{n_s}$ **do**
6: **if** $\delta(j)$ is *feasible* (of (11) with $\kappa = \kappa^*$) **then**
7: CIs $= \{\text{CIs}, \delta(j)\}$
8: **return** (κ^*, CIs)

Fig. 4 *Deriving the security index and CIs*

programming problems, making the standard non-linear programming approach inapplicable [31]. These challenges arise because the MPEC (8) is non-convex, is non-differentiable, and has two conflicting objectives.

The complementary slackness constraint (8g) makes the MPEC (8) non-convex. To address this challenge, we linearise (8g) using the Big M method [31]. Let $M > 0$ be a sufficiently large constant, then (8g) is equivalent to

$$\lambda^* \preceq M(\mathbf{1} - \omega), \quad -(A_0 P_{\text{g}}^* + A_1 a - b) \preceq M\omega, \tag{9}$$

where $\omega \in \{0,1\}^{2(n+m)}$ is a binary decision variable.

The attack goal constraint (3) makes the MPEC (8) non-differentiable. To address this challenge, we proceed as follows. Since the flow on e before the attack $P_{f,e}(0)$ can be computed using (1), the attack goal constraint (3) can be written as

$$\begin{cases} [F]_e(P_{\text{g}}^*(a) - P_{\text{d}}) \geq (1+\tau)\left|P_{f,e}(0)\right| - M_\infty(1-\omega_e^+), \\ [F]_e(P_{\text{g}}^*(a) - P_{\text{d}}) \leq -(1+\tau)\left|P_{f,e}(0)\right| + M_\infty(1-\omega_e^-), \\ \omega_e^+ + \omega_3^- = 1, \end{cases} \tag{10}$$

where $M_\infty > 0$ is a sufficiently large constant and $\omega_e^+, \omega_e^- \in \{0,1\}$ are binary decision variables.

The MPEC (8) has two conflicting objectives, i.e. max $\tau - \kappa$. To address this challenge, we minimise κ (i.e. the number of target substations) and let $\tau \geq \tilde{\tau}$ where $\tilde{\tau}$ is a predefined flow increase. We can attach semantics to $\tilde{\tau}$, e.g. the $(\tilde{\tau})$ that triggers the line's protection.

The proposed solutions for the challenges transform the MPEC (8) into the following mixed-integer linear programming problem:

$$\begin{aligned} \min_{\tau,\kappa,\delta,a,P_{\text{g}}^*} \quad & \kappa, \\ \text{s.t.} \quad & \tau \geq \tilde{\tau}, \\ & (4),(5),(8\text{c})-(8\text{f}),(9) \text{ and } (10). \end{aligned} \tag{11}$$

5.3 Algorithm: deriving the CIs

The optimal solutions κ^* and $\delta^*(e,\tilde{\tau})$ of (11) denote, respectively, the security index and the CI for the attack goal $(e,\tilde{\tau})$. The *security index* κ^* determines the least number of target substations to increase the flow $(\tilde{\tau})$ on line e, while the *CI* $\delta^*(e,\tilde{\tau})$ describes which target substations. This CI represents a strongly correlated CCA since it relates the least number of target substations with the attack goal $(e,\tilde{\tau})$.

Though the security index κ^* is unique, the CI might not be. Other CCAs attacking κ^* substations might also increase the flow $(\tilde{\tau})$ on line e –i.e. a consequence of the combinatorial nature of (11). All the CIs, however, are feasible solutions of (11) with $\kappa = \kappa^*$, which we use to develop the following algorithm (see Fig. 4).

Given the attack goal $(e,\tilde{\tau})$, Algorithm 1 (depicted in Fig. 4) computes the security index first, and then the CIs by exploring which of the $\binom{\kappa^*}{n_s}$ combinations of target substations are feasible solutions of (11) with $\kappa = \kappa^*$.

The mathematical procedure, i.e. deriving the CIs from the embedded optimisation problem (11), is applicable to other static control applications, which are formulated as an optimisation problem. Examples are emergency voltage control, economic dispatches in electricity markets under various time frameworks and so on.

5.4 Limitations

Our method has a limitation, namely the computation performance of Algorithm 1 (Fig. 4), which we discuss next.

Algorithm 1 (Fig. 4) only promises local optimal solutions in finite time. This is because the mixed-integer linear problem (11) and the $\binom{\kappa^*}{n_s} - 1$ feasibility problems are in general NP-hard. Given that there are only a few substations in a power grid, the computation time of the proposed algorithm is unlikely to be a problem. However, in the case of abrupt changes occurring in the power grid, CIs will need to be updated at run-time and an algorithm providing theoretic bounds of convergence must be sought after. These tasks are out of the scope of this paper, but they will be part of our future work.

6 Applying the CIs to protect against CCAs

In this section, we describe the properties of CIs using a set-theoretic approach. These properties allow us to derive defence strategies against CCAs. In particular, CIs defend against CCAs in the following ways: (i) CIs imply defence strategies (in terms of physical and cyber assets criticality) under limited resources and (ii) CIs reveal the attack goals of CCAs, which can be used in IDS to allow the runtime detection of CCAs.

6.1 CIs' properties

Let $S_{\alpha,j} \in 2^S$ describe the set of target substations during a CCA. If the CCA is effective, i.e. if the CCA increases the flow $(\tilde{\tau})$ on line $e \in E$, we use the notation $S_{\alpha,j} \rightarrow (e,\tilde{\tau})$; otherwise, we use $S_{\alpha,j} \nrightarrow (e,\tilde{\tau})$. We collect all effective CCAs in the set

$$\mathcal{S}_\alpha(e,\tilde{\tau}) := \{S_{\alpha,j} \,\big|\, S_{\alpha,j} \rightarrow (e,\tilde{\tau})\} \subset 2^S.$$

The next proposition shows that if the CCA $S'_{\alpha,j}$ fails to increase the flow $(\tilde{\tau})$ on line e, then all subordinated attacks $S_{\alpha,j} \subset S'_{\alpha,j}$ also fail to increase the flow on e.

Proposition 1: (Subordinated CCAs): If $S'_{\alpha,j} \notin \mathcal{S}_\alpha(e,\tilde{\tau})$, then $S_{\alpha,j} \notin \mathcal{S}_\alpha(e,\tilde{\tau})$ for any $S_{\alpha,j} \subset S'_{\alpha,j}$.

Proof: See the Appendix. □

Definition 1: Let $\delta^*(e,\tilde{\tau})$ denote a feasible solution of Algorithm 1 (Fig. 4) (i.e. $\| \delta^*(e,\tilde{\tau}) \|_0 = \kappa^*$). A CI, denoted as $S^*_{\alpha,j}$, is a strongly correlated CCA that extracts target substations from $\delta^*(e,\tilde{\tau})$ as follows:

$$S^*_{\alpha,j} := \{s_k \in S \,\big|\, \delta^*_k(e,\tilde{\tau}) \neq 0\}, \tag{12}$$

and reaches the goal $(e,\tilde{\tau})$, i.e. $S^*_{\alpha,j} \in \mathcal{S}_\alpha(e,\tilde{\tau})$.

Proposition 2: (Minimal cardinality): Let $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$ be a CI. Then $S'_{\alpha,j} \nrightarrow (e,\tilde{\tau})$ for any $S'_{\alpha,j} \in 2^S$ satisfying $\left|S'_{\alpha,j}\right| < \kappa^*$.

Proof: See the Appendix. □

Note that Propositions 1 and 2 guarantee security against subordinated CCAs of the CI $S^*_{\alpha,j}$.

CIs are not unique since there might be another CCA $S'_{\alpha,j}$ satisfying $\left|S'_{\alpha,j}\right| = \kappa^*$ and $S'_{\alpha,j} \in \mathcal{S}_\alpha(e,\tilde{\tau})$. We collect all the CIs in the set

$$\mathcal{S}^*_\alpha(e,\tilde{\tau}) := \{S^*_{\alpha,j} \mid S^*_{\alpha,j} \text{ is a CI}\} \subseteq \mathcal{S}_\alpha(e,\tilde{\tau}).$$

The following lemma states that any CCA containing the CI $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$ can increase the flow on line $e \in E$.

Lemma 1: Let $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$ denote a CI. Suppose there exists another set of target substations $S'_{\alpha,j} \in 2^S$, then $(S^*_{\alpha,j} \cup S'_{\alpha,j}) \in \mathcal{S}_\alpha(e,\tilde{\tau})$.

Proof: See the Appendix. □

Lemma 1 implies that CCAs might be targeting multiple lines, which we generalise in the next theorem.

Theorem 1: (Multiple targets): Let J be an arbitrary index set, and let $S^*_{\alpha,j} \in \mathcal{S}^*(e_j,\tilde{\tau}_j)$ be CIs (for different lines) for all $j \in J$. Suppose there exist a CCA targeting substations $S^* \in 2^S$ such that $S^*_{\alpha,j} \subset S^*$ for all $j \in J$. Then the CCA S^* can increase the flow $(\tilde{\tau}_j)$ on any line from the set $\{e_j\}_{j\in J}$, i.e.$S^* \in \mathcal{S}(e_j,\tilde{\tau}_j)$ for all $j \in J$.

Proof: See the Appendix. □

The CIs' properties described allowing us to study what happens if we protect the measurements on a substation, which we state in the next theorem.

Theorem 2: (Defense at a single substation): Let J be an arbitrary index set, and let $S^*_{\alpha,j} \in \mathcal{S}^*(e_j,\tilde{\tau}_j)$ be CIs (targeting different lines) for all $j \in J$. Moreover, suppose the collection of target substations $\{S^*_{\alpha,j}\}_{j\in J}$ satisfies $\cap_{j\in J} S^*_{\alpha,j} \neq \emptyset$. If the grid's defender protects measurements at the substation $s^*_k \in \cap_{j\in J} S^*_{\alpha,j}$, then one of the following occurs (for all $j \in J$):

i. If $S^*_{\alpha,j}$ is the unique CI that increases the flow $(\tilde{\tau}_j)$ on line e_j, then after s^*_k is protected, the new CI $S^*_{\beta,j}$ (not necessarily unique) will require targeting more substations, i.e. $\kappa^*_\beta := |S^*_{\beta,j}| > |S^*_{\alpha,j}|$.
ii. If $S^*_{\alpha,j}$ is not the only CI, then, after s^*_k is protected, the following might occur: (a) if s^*_k is common to all CIs $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$, then the conclusion in (i) applies; or (b) if $\{s^*_k\} \cap S^*_{\alpha,j} = \emptyset$ for some $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$, then the new collection of CIs, denoted as $\mathcal{S}^*_\beta(e_j,\tilde{\tau}_j)$, satisfies $\mathcal{S}^*_\beta(e_j,\tilde{\tau}_j) \subset \mathcal{S}^*_\alpha(e_j,\tilde{\tau}_j)$.
iii. The attack is infeasible, i.e. $\mathcal{S}^*_\beta(e_j,\tilde{\tau}_j) \equiv \{\emptyset\}$.

Proof: See the Appendix. □

Theorem 2 suggests that protecting a substation will pivot the CI (from one set to a different set). It also suggests that protecting substation $s_k \in S$ becomes more critical if s_k is related to more target lines. We discuss more defence implications next.

6.2 Defence implications

In this subsection, we derive the best defence for a line $e \in E$, the best defence against CIs, and the best defence for substations based on the CIs' properties.

6.2.1 Best defence for a line $e \in E$: Suppose the CCA $S_{\alpha,j}$ increases the flow $(\tilde{\tau})$ on line $e \in E$, i.e. $S_{\alpha,j} \in \mathcal{S}_\alpha(e,\tilde{\tau})$. Then, the best defence for e is to protect the minimal set of substations $D_{\beta,e} \subseteq S_{\alpha,j}$ that renders the new CCA $S'_{\alpha,j} := S_{\alpha,j} \setminus D_{\beta,e}$ ineffective, i.e. $S'_{\alpha,j} \notin \mathcal{S}_\alpha(e,\tilde{\tau})$. If $S_{\alpha,j}$ is a CI (i.e. $S_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$), then $D_{\beta,e} = \{s_k\}$ with $s_k \in S_{\alpha,j}$, i.e. protecting any substation from $S_{\alpha,j}$ renders the new CCA ineffective.

On the other hand, CCAs might not remain static; that is, the adversary might switch between a CI $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$ and a CCA $S_{\alpha,j} \in \mathcal{S}_\alpha(e,\tilde{\tau})$ to identify vulnerabilities and hide from detection. However, if the attacks ($S^*_{\alpha,j}$ and $S_{\alpha,j}$) have common substations, i.e. if $S^*_{\alpha,j} \cap S_{\alpha,j} \neq \emptyset$, then the best defence for the line e is to protect substations satisfying $s^*_k \in S^*_{\alpha,j} \cap S_{\alpha,j}$.

6.2.2 Best defence against CIs: Suppose κ^* denotes the security index for the attack goal $(e,\tilde{\tau})$. Then, the best defence against CIs (i.e. $\mathcal{S}^*_\alpha(e,\tilde{\tau})$) is to protect the minimal set of substations $D_{\beta,e}$ that renders the new security index κ^*_β greater than κ^*. Thus, the adversary is required to attack more substations after $D_{\beta,e}$ is protected. Note that Theorem 2(i) and (ii-a) describe two special cases of this defence, i.e. when $D_{\beta,e} \equiv \{s^*_k\}$.

6.2.3 Metrics of defence effectiveness: We describe the metric used to compare defence at substations and to identify the best defence strategy. Suppose $\kappa^*(e,\tilde{\tau})$ denotes the security index for the attack goal $(e,\tilde{\tau})$. The security index measures the likelihood of a CCA since it is less likely to attack more substations than $\kappa^*(e,\tilde{\tau})$. We define the *average likelihood* to increase the flow $(\tilde{\tau})$ on all lines as

$$R(\tilde{\tau}) = \frac{1}{m}\sum_{e\in E} \kappa^*(e,\tilde{\tau}).$$

Using the average likelihood, we have the following definition.

Definition 2: For a target flow increase $\tilde{\tau}$, the defence effectiveness for substation s_k can be estimated by calculating

$$\Delta R_{\beta,s_k}(\tilde{\tau}) := R_{\beta,s_k}(\tilde{\tau}) - R(\tilde{\tau}),$$

where $R_{\beta,s_k}(\tilde{\tau}) := (1/m)\sum_{e\in E}\kappa^*_{\beta,s_k}(e,\tilde{\tau})$ denotes the average likelihood after protecting substation s_k.

Theorem 2 implies that $\Delta R_{\beta,s_k}(\tilde{\tau}) \geq 0$ for all $s_k \in S$, i.e. after protecting substation s_k, the number of target substations increases, while the average likelihood decreases. Thus, the best defence strategy is to protect substation s^*_k such that $\Delta R_{\beta,s^*_k} \in \text{argmax}_{s_k\in S}\{\Delta R_{\beta,s_k}\}$. Note that we derive the metric for a specific flow increase value $(\tilde{\tau})$, but we can always derive it for the case when $\tilde{\tau} \in (0,\bar{\tau}]$ is a free parameter for all target lines.

Remark 4: If we integrate κ^* (i.e. the likelihood of CCAs) and the likelihood of exploits at the cyber network, we can derive a risk metric for CCAs at both the cyber and physical networks. This metric will be studied in our future work.

6.3 Application: identifying CCAs

In this subsection, we briefly describe how CIs based on cyber-traces and CIs based on attack goals identify CCAs and estimate their possible consequences.

CIs based on cyber-traces identify in real time individual components of CCAs, i.e. the set of suspected target substations. These CIs interpret intrusion data from sensors of IDSs (or other security tools) installed at substations.

CIs based on attack goals estimate the possible consequences of CCAs. The grid's defender computes these CIs and stores them in a knowledge base. The grid's defender can update this knowledge base of CIs as needed.

Fig. 5 depicts a schematic diagram of how the CIs work together. The CIs based on cyber-traces output the set of suspected target substations $S_{\alpha,j}(t)$ (at some time t) to the knowledge base of CIs. The knowledge base of CIs compares this set of suspected target substations with the CIs (and their mathematical properties) to estimate possible consequences $(e,\tilde{\tau})$. This approach is analogous to signature-based (also known as a blacklist) detection techniques [32]. Nevertheless, the CIs (signatures) are derived based on the attack template instead of direct network knowledge. Thus, combining CIs with other direct-knowledge based approaches in IDS will significantly improve defence performance

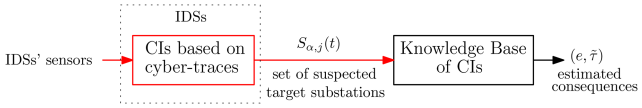

Fig. 5 *Schematic diagram: identifying CCAs*

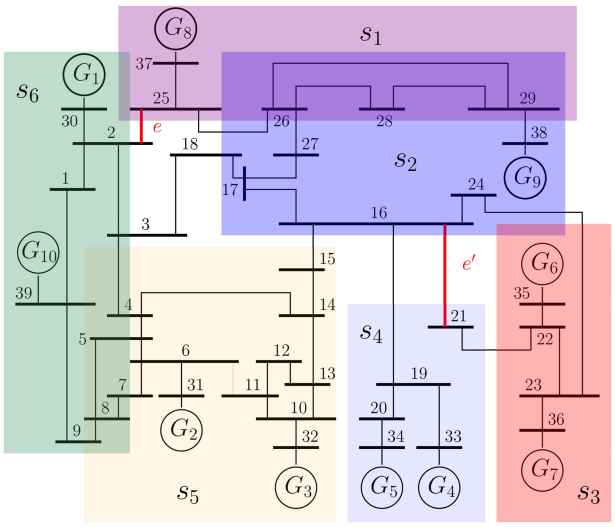

Fig. 6 *New England 39 bus system*

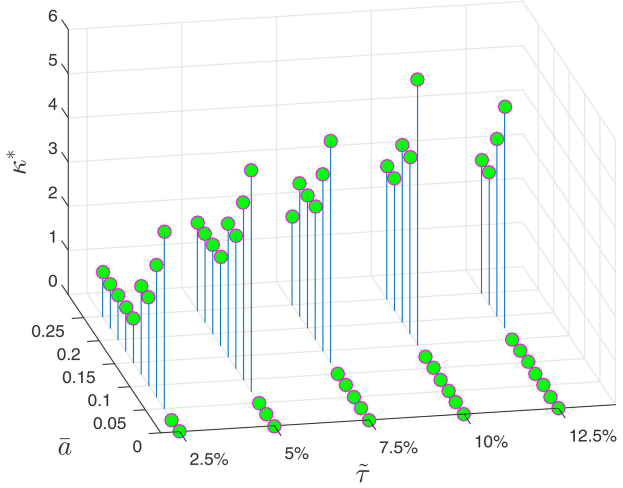

Fig. 7 *CIs' dependence on $\bar{a}$. Target line $e = (2, 25)$*

and allow taking immediate actions upon most harmful attacks. We will provide details of this approach in a different paper.

7 Numerical experiments

In this section, we provide numerical simulations, using the reduced model of the New England power grid, to demonstrate (i) how our method deduces CIs (described in Section 5); and (ii) the CIs properties, defence implications, and the metric of defence effectiveness (described in Section 6). We remark, however, that the deductive approach to construct CIs is not limited to the experimented power system but also applicable to any power system configuration.

Fig. 6 shows the New England 39 bus system used to model a power grid with $n_s = 6$ substations. We selected two target lines $e = (2, 25)$ and $e' = (16, 21)$; and the target flow increase $\tau \in \{2.5\%, 5\%, 7.5\%, 10\%\}$. The line $e \in E$ (resp. $e' \in E$) connects substations s_1 and s_6 (resp. s_2 and s_4), and allow us to mimic attacks aiming to cause overloading, trip the protective relays on the lines, and disconnect the substations from each other. The parameters used in our experiments were $M, M_\infty = 10^3$, $\bar{a} = 0.1$, and the SCED base case data for the New England system taken from MATPOWER software package [33].

Table 1 CIs for line $e = (2, 25)$

Attack goal	CIs	Security index
$\mathcal{S}^*_\alpha(e, 2.5\%)$	$\{2\}$	$\kappa^* = 1$
$\mathcal{S}^*_\beta(e, 2.5\%)$	$\{5, 6\}$	$\kappa^*_\beta = 2$
$\mathcal{S}^*_\alpha(e, 5\%)$	$\{2, 5, 6\}, \{4, 5, 6\}$	$\kappa^* = 3$
$\mathcal{S}^*_\beta(e, 5\%)$	$\{4, 5, 6\}$	$\kappa^*_\beta = 3$
$\mathcal{S}^*_\alpha(e, 7.5\%)$	$\{1, 2, 5, 6\}$	$\kappa^* = 4$
$\mathcal{S}^*_\beta(e, 7.5\%)$	$\{\varnothing\}$	$\kappa^*_\beta = 0$
$\mathcal{S}^*_\alpha(e, 10\%)$	$\{S\}$	$\kappa^* = 6$
$\mathcal{S}^*_\beta(e, 10\%)$	$\{\varnothing\}$	$\kappa^*_\beta = 0$

Note: $\mathcal{S}^*_\alpha(e, \tilde{\tau})$ (resp. $\mathcal{S}^*_\beta(e, \tilde{\tau})$) denotes the CIs before (resp. after) protecting $s^*_k = 2$. $\{\varnothing\}$ (or $\kappa^* = 0$) implies the attack is ineffective.

Table 2 CIs for line $e' = (21, 24)$

Attack goal	CIs	Security index
$\mathcal{S}^*_\alpha(e', 2.5\%)$	$\{2\}$	$\kappa^* = 1$
$\mathcal{S}^*_\beta(e', 2.5\%)$	$\{5, 6\}$	$\kappa^*_\beta = 2$
$\mathcal{S}^*_\alpha(e', 5\%)$	$\{4, 5, 6\}$	$\kappa^* = 3$
$\mathcal{S}^*_\beta(e', 5\%)$	$\{4, 5, 6\}$	$\kappa^*_\beta = 3$
$\mathcal{S}^*_\alpha(e', 7.5\%)$	$\{\varnothing\}$	$\kappa^* = 0$
$\mathcal{S}^*_\beta(e', 7.5\%)$	$\{\varnothing\}$	$\kappa^*_\beta = 0$

Note: $\mathcal{S}^*_\alpha(e', \tilde{\tau})$ (resp. $\mathcal{S}^*_\beta(e', \tilde{\tau})$) denotes the CIs before (resp. after) protecting $s^*_k = 2$. $\{\varnothing\}$ (or $\kappa^* = 0$) implies the attack is ineffective.

7.1 Experiment 1: Deducing the CIs

In this experiment, we derived the CIs for the attack goals $(e, \tilde{\tau})$ and $(e', \tilde{\tau})$ using Algorithm 1 (Fig. 4). We implemented Algorithm 1 (Fig. 4) using CVX (a package for solving convex and linear mixed-integer programs [34]). Tables 1 and 2 present the collection of CIs. We found that all attack goals have unique CIs but $(e, 5\%)$.

7.2 Experiment 2: CIs dependence on the parameter $\bar{a}$

In this experiment, we studied the CIs' dependence on $\bar{a}$. Fig. 7 shows how the security index κ^* changes as we increase the attack signal max value $\bar{a}$. We found that the security index decreases as $\bar{a}$ increases. This result implies that if the defender increases $\bar{a}$ in the attack template, the defence implications become more conservative.

7.3 Experiment 3: Defence implications of CIs

We studied the mathematical properties of CIs and defence implications from Theorem 2.Tables 1 and 2 show, respectively, the CIs for the attack goals $(e, \tilde{\tau})$ and $(e', \tilde{\tau})$, before and after protecting substation $s^*_k = s_2$. Before protecting substation $s^*_k = s_2$, the CIs have the following defence implications: for the attack goal $(e, 5\%)$, subordinated attacks of the CI $S^*_{\alpha,j} = \{2, 5, 6\}$ are ineffective (Propositions 1 and 2); and the CCA $S^* = \{2, 4, 5, 6\}$ can increase the flow $\tilde{\tau} = 5\%$ on both lines e and e' (Lemma 1 and Theorem 1). And, after protecting substation $s^*_k = 2$, the CIs have the following defence implications: for the attack goal $(e', 2.5\%)$, the new security index satisfies $\kappa^*_\beta = 2 > 1 = \kappa^*$ (Theorem 2 (i)), $\mathcal{S}^*_\beta(e, 5\%) \subset \mathcal{S}^*_\alpha(e, 5\%)$ (Theorem 2 (ii-b)), and $\mathcal{S}^*_\beta(e, 7.5\%) = \{\varnothing\}$ (Theorem 2 (iii)).

7.4 Experiment 4: The metric of defence effectiveness

In this final experiment, we computed the metric of defence effectiveness $\Delta R_{s_k}(\tilde{\tau})$ for all substations $s_k \in S$ and $\tilde{\tau} \in \{5\%, 7.5\%\}$. Table 3 presents the results. These results imply that the best defence is achieved by protecting substation $s^*_k = s_2$.

Table 3 Metric of defence effectiveness

Substation	s_1	s_2	s_3	s_4	s_5	s_6
$\Delta R_{s_k}(5\%)$	0.11	0.70	0.02	0.02	0.30	0.30
$\Delta R_{s_k}(7.5\%)$	0.13	0.54	0.02	0.07	0.2	0.2

8 Conclusion

In this paper, we provided a method to derive CIs based on attack goals, which can be used to estimate attack consequences and identify critical substations during coordinated attacks. Compared to existing approaches, our method does not rely on numerical simulation of a large number of attack events to conclude attack patterns for a specific victim power grid. In contrast, our method is deductive – by deriving CIs, we analytically reveal the cyber-physical causal chain of attack for any power system configuration and attack goals – and is able to detect more sophisticated attacks, such as measurement attacks. We modelled the attack template as a bilevel optimisation program and derived Algorithm 1 (Fig. 4) to solve it. Algorithm 1 (Fig. 4) computes the CIs for any given attack goal. These CIs describe strongly correlated attacks since the adversary reaches the goal by attacking the least number of target substations. We then used a set-theoretic approach to derive the CIs' properties. These properties suggest defence implications against coordinated attacks, including the best defence for a transmission line, the best defence against strongly correlated attacks, and the metric of defence effectiveness. Thus, our method to compute CIs and their properties present the benefit of deployment simplicity but face one limitation, namely the computational performance of Algorithm 1 (Fig. 4). However, given that there are only a few substations in the power grid, the computation performance is unlikely to be a problem. In our future work, we will use the CIs and their defence implications together with IDSs to protect the grid against coordinated attacks.

9 References

[1] Zhu, B., Joseph, A., Sastry, S.: 'A taxonomy of cyber attacks on SCADA systems'. Internet of Things (iThings/CPSCom), 2011 Int. Conf. on and 4th Int. Conf. on Cyber, Physical and Social Computing, Dalian, China, 2011, pp. 380–388

[2] Liu, C.C., Stefanov, A., Hong, J., *et al.*: 'Intruders in the grid', *IEEE Power Energy Mag.*, 2012, **10**, (1), pp. 58–66

[3] Lee, R.M., Assante, M.J., Conway, T.: 'Analysis of the cyber attack on the Ukrainian power grid', SANS Ind. Control Syst., 2016

[4] Sun, C.C., Hong, J., Liu, C.C.: 'A coordinated cyber attack detection system (CCADS) for multiple substations'. Power Systems Computation Conf. (PSCC), Genoa, Italy, 2016, pp. 1–7

[5] Dragos, Inc.: 'Crashoverride: Analysis of the threat to electric grid operations', 2017

[6] Ten, C.W., Hong, J., Liu, C.C.: 'Anomaly detection for cybersecurity of the substations', *IEEE Trans. Smart Grid*, 2011, **2**, (4), pp. 865–873

[7] Wang, P., Ashok, A., Govindarasu, M.: 'Cyber-physical risk assessment for smart grid system protection scheme'. 2015 IEEE Power & Energy Society General Meeting, Denver, CO, USA, 2015, pp. 1–5

[8] Ten, C.W., Liu, C.C., Govindarasu, M.: 'Vulnerability assessment of cybersecurity for SCADA systems using attack trees'. Power Engineering Society General Meeting, Tampa, FL, USA, 2007, pp. 1–8

[9] Liu, N., Zhang, J., Zhang, H., *et al.*: 'Security assessment for communication networks of power control systems using attack graph and mcdm', *IEEE Trans. Power Deliv.*, 2010, **25**, (3), pp. 1492–1500

[10] Chen, T.M., Sanchez Aarnoutse, J.C., Buford, J.: 'Petri net modeling of cyberphysical attacks on smart grid', *IEEE Trans. Smart Grid*, 2011, **2**, (4), pp. 741–749

[11] Jie, Y., Govindarasu, M., Chen Ching, L., *et al.*: 'Risk assessment framework for power control systems with pmu-based intrusion response system', *J. Modern Power Syst. Clean Energy*, 2015, **3**, (3), pp. 321–331

[12] Ten, C.W., Ginter, A., Bulbul, R.: 'Cyber-based contingency analysis', *IEEE Trans. Power Syst.*, 2016, **31**, (4), pp. 3040–3050

[13] Choi, D.H., Xie, L.: 'Ramp-induced data attacks on look-ahead dispatch in real-time power markets', *IEEE Trans. Smart Grid*, 2013, **4**, (3), pp. 1235–1243

[14] Xie, L., Mo, Y., Sinopoli, B.: 'Integrity data attacks in power market operations', *IEEE Trans. Smart Grid*, 2011, **2**, (4), pp. 659–666

[15] Ye, H., Ge, Y., Liu, X., *et al.*: 'Transmission line rating attack in two-settlement electricity markets', *IEEE Trans. Smart Grid*, 2016, **7**, (3), pp. 1346–1355

[16] Wang, J.: 'Resilient electric grid for smart cities'. in '*Smart cities: foundations, principles, and applications*' (John Wiley & Sons, Hoboken, NJ, USA, 2017), pp. 541–578

[17] Liang, G., Zhao, J., Luo, F., *et al.*: 'A review of false data injection attacks against modern power systems', *IEEE Trans. Smart Grid*, 2016, **PP**, (99), p. 1

[18] Tajer, A.: 'False data injection attacks in electricity markets by limited adversaries: stochastic robustness', *IEEE Trans. Smart Grid*, 2017, early access: DOI: 10.1109/TSG.2017.2733346

[19] Rahman, M.A., Mohsenian Rad, H.: 'False data injection attacks with incomplete information against smart power grids'. Global Communications Conf. (GLOBECOM), Anaheim, CA, USA, 2012, pp. 3153–3158

[20] Liu, X., Li, Z.: 'False data attacks against ac state estimation with incomplete network information', *IEEE Trans. Smart Grid*, 2017, **8**, (5), pp. 2239–2248

[21] Tajer, A., Kar, S., Poor, H.V., *et al.*: 'Distributed joint cyber attack detection and state recovery in smart grids'. 2011 IEEE Int. Conf. on Smart Grid Communications (SmartGridComm), Brussels, Belgium, 2011, pp. 202–207

[22] Dorfler, F., Bullo, F.: 'Novel insights into lossless ac and dc power flow'. Power and Energy Society General Meeting (PES), Vancouver, BC, Canada, 2013, pp. 1–5

[23] 'The pjm independent system operator'. Available at http://www.pjm.com/markets-and-operations.aspx

[24] 'The New England independent system operator'. Available at https://www.iso-ne.com/markets-operations

[25] 'California ISO, business practice manual for market instruments, chapter 2.1 the market power mitigation and reliability requirements, 2017'. Available at https://bpmcm.caiso.com/Pages/BPMDetails.aspx?BPM=Market%20Instruments

[26] Luo, Z.Q., Pang, J.S., Ralph, D.: '*Mathematical programs with equilibrium constraints*' (Cambridge University Press, Cambridge, UK, 1996)

[27] Yuan, Y., Li, Z., Ren, K.: 'Modeling load redistribution attacks in power systems', *IEEE Trans. Smart Grid*, 2011, **2**, (2), pp. 382–390

[28] Liang, J., Sankar, L., Kosut, O.: 'Vulnerability analysis and consequences of false data injection attack on power system state estimation', *IEEE Trans. Power Syst.*, 2016, **31**, (5), pp. 3864–3872

[29] Arroyo, J.M., Galiana, F.D.: 'On the solution of the bilevel programming formulation of the terrorist threat problem', *IEEE Trans. Power Syst.*, 2005, **20**, (2), pp. 789–797

[30] Boyd, S., Vandenberghe, L.: '*Convex optimization*' (Cambridge University Press, NY, USA, 2004)

[31] Dempe, S.: '*Foundations of bilevel programming*' (Springer Science & Business Media, Dordrecht, Netherlands, 2002)

[32] Zhu, B., Ghorbani, A.A.: 'Alert correlation for extracting attack strategies' , *Int. J. Netw. Secur.*, 2006, **3**, (3), pp. 244–258

[33] Zimmerman, R.D., Murillo Sánchez, C.E., Thomas, R.J.: 'Matpower: steady-state operations, planning, and analysis tools for power systems research and education', *IEEE Trans. Power Syst.*, 2011, **26**, (1), pp. 12–19

[34] Grant, M., Boyd, S..: 'CVX: Matlab software for disciplined convex programming, version 2.1'. 2014. Available at http://cvxr.com/cvx

10 Appendix

Proof: (Proposition 2): Assume $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$. We partition the CI in two arbitrary disjoint sets, i.e. $S^*_{\alpha,j} = S'_{\alpha,j} \cup D$ satisfying $|S'_{\alpha,j}| < \kappa^* := |S^*_{\alpha,j}|$. Since CIs are minimum cardinality CCAs, then $S'_{\alpha,j} \notin \mathcal{S}_\alpha(e,\tilde{\tau})$, which proves the proposition. □

Proof: (Lemma 1): Suppose, to get a contradiction, $S_{\alpha,j} \in \mathcal{S}_\alpha(e,\tilde{\tau})$; then any super-set S^* of $S_{\alpha,j}$, i.e. $S_{\alpha,j} \subseteq S^*$, is effective, i.e. $S^* \in \mathcal{S}_\alpha(e,\tilde{\tau})$. In particular, $S'_{\alpha,j} \equiv S^*$ reaches the goal $(e,\tilde{\tau})$, which contradicts $S'_{\alpha,j} \notin \mathcal{S}_\alpha(e,\tilde{\tau})$. □

Proof: (Lemma 1): We prove the lemma by cases. (i) Suppose $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$ is a super-set of $S'_{\alpha,j}$, i.e. $S^*_{\alpha,j} \supseteq S'_{\alpha,j}$. It follows that $S'_{\alpha,j} \cup S^*_{\alpha,j} \equiv S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau}) \subseteq \mathcal{S}_\alpha(e,\tilde{\tau})$. Similarly, (ii) suppose that $S^*_{\alpha,j} \subseteq S'_{\alpha,j}$. Then, it follows that $S'_{\alpha,j} \cup S^*_{\alpha,j} = S'_{\alpha,j} \in \mathcal{S}_\alpha(e,\tilde{\tau})$. Finally, (iii) suppose the CI $S^*_{\alpha,j}$ is neither a subset or super-set of $S'_{\alpha,j}$. Assume, to get a contradiction, that $(S'_{\alpha,j} \cup S^*_{\alpha,j}) \notin \mathcal{S}_\alpha(e,\tilde{\tau})$. Then, Proposition 1 implies that any CCA $S_{\alpha,j} \subseteq (S'_{\alpha,j} \cup S^*_{\alpha,j})$ does not reach the goal $(e,\tilde{\tau})$. In particular, $S_{\alpha,j} \equiv S^*_{\alpha,j} \subset (S'_{\alpha,j} \cup S^*_{\alpha,j})$ does not reach $(e,\tilde{\tau})$, i.e. $S^*_{\alpha,j} \notin \mathcal{S}_\alpha(e,\tilde{\tau})$. This yields the contradiction. □

Proof: (Theorem 1): We partition the CCA S^* in the union of two disjoint sets, i.e. $S^* = S^*_{\alpha,j} \cup (S^* \setminus S^*_{\alpha,j})$ for all $j \in J$. Then, by Lemma 1 we have $S^* \in \mathcal{S}_\alpha(e_j,\tilde{\tau}_j)$ for all $j \in J$, which proves the theorem. □

Proof: (Theorem 2): Assume the operator protects substation s^*_k. Moreover, assume that $S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$ is unique and the attack remains feasible after defence. Suppose, to get a contradiction, that

the new security index satisfy $\kappa^*_\beta = \kappa^*$. This implies that the new CI satisfies $S^*_{\beta,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau})$, which contradicts the fact that $\mathcal{S}^*_\alpha(e,\tilde{\tau}) = \{S^*_{\alpha,j}\}$. This proves (i). On the other hand, suppose that $S^*_{\alpha,j} \in S^*_{\alpha,j}$ is not unique. Part (ii-a) can be proven using the same arguments as in (i). We prove part (ii-b) as follows. Define $\mathcal{S} := \{S^*_{\alpha,j} \in \mathcal{S}^*_\alpha(e,\tilde{\tau}) \big| s^*_k \cap S^*_{\alpha,j} = \varnothing\}$. Then we partition the collection $\mathcal{S}^*_\alpha(e,\tilde{\tau})$ as follows $\mathcal{S}^*_\alpha(e,\tilde{\tau}) = \mathcal{S} \cup (\mathcal{S}^*_\alpha(e,\tilde{\tau}) \setminus \mathcal{S})$. Proposition 1 implies that after defence $S^*_{\alpha,j} \setminus \{s^*_k\} \notin \mathcal{S}^*_\beta(e,\tilde{\tau})$ for all $S^*_{\alpha,j} \in (\mathcal{S}^*_\alpha(e,\tilde{\tau}) \setminus \mathcal{S})$. Thus, $\mathcal{S}^*_\beta(e,\tilde{\tau}) \equiv \mathcal{S}$, and therefore $\mathcal{S}^*_\beta(e,\tilde{\tau}) \subset \mathcal{S}^*_\alpha(e,\tilde{\tau})$. This proves (ii-b). Finally, (c) follows trivially. □