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

Christian Moya†, Jiankang Wang†
†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.
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
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. $\mathbb{R}_{\leq 0}$) denote the set of real numbers and non-negative (resp. positive) real numbers. For $n > 1$, $I_n$ denotes the $n$-dimensional identity matrix, $1$ and $0$ 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 $A \in \mathbb{R}^{n \times m}$, $[A]_{i,j}$ denotes its $i$th row and its $j$th element. Given a vector $x \in \mathbb{R}^n$, $x_i$ denotes the $i$th element, $(x)$ is the diagonal matrix of $x$, and $\|x\|_\infty$ is the zero norm of $x$, i.e. the number of non-zero elements of $x$. We let $\|x\|_\infty := \max \{|x_i|\}$. For two vectors $x, y \in \mathbb{R}^n$, $x \circ y = z \in \mathbb{R}^n$ denotes the Hadamard element-wise product, i.e. $z = x \circ y$, 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 $E \subseteq V \times V$ are the sets of nodes $n := |V|$ 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 $P_{d,i} \in \mathbb{R}_{\leq 0}$ 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 $s_i$, we represent the grid within its service area as the sub-graph $G_{s_i} = (V_{s_i}, E_{s_i})$ with the following properties:

- Substation service areas compose the entire power grid, i.e. $\cup_{s \in S} G_{s} = G$.
- Substation service areas may overlap, i.e. for some $s_i, s_j \in S$, we may have $G_{s_i} \cap G_{s_j} \neq \emptyset$, but the overlapped areas do not have buses with generation.
- Each substation collects demand measurements, denoted as $P_d \in \mathbb{R}^s$, within its service area.

3.3 Security constrained economic dispatch

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

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 infeasible. 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.

$$\begin{align*}
\min_{P_g} & \quad \frac{1}{2} P_g^\top C P_g + c_1^\top P_g + c_0 \\
\text{s.t.} & \quad 1^\top P_g - 1^\top \hat{P}_d = 0, \\
& \quad P_g \in [0, \hat{P}_g], \\
& \quad F(P_g - \hat{P}_d) \preceq \theta_l
\end{align*}$$

where $c_0, c_1, c_2 \in \mathbb{R}^{n_n}$ are the cost coefficients for a generation, $C = (c_1)$, $\hat{P}_g \in \mathbb{R}^n_\geq 0$ is the rated power from generators, and $\hat{P}_d \in \mathbb{R}^n_\leq 0$ is the thermal capacity of transmission lines, and $F$ 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:

$$P_{d,a}(i) = P_{d,i} + a_i$$

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

$$P_{f_{a},i}^* = \begin{cases} 
|P_{f,i}^*(a)| & \text{if } P_{f,i}^*(a) > 0 \\
-|P_{f,i}^*(a)| & \text{if } P_{f,i}^*(a) < 0
\end{cases} \geq (1 + \tau)P_{f,i}^*(0)$$

where $P_{f,i}^*(a) \in \mathbb{R}$ (resp. $P_{f,i}^*(0) \in \mathbb{R}$) denotes the power flow on $e$ after (resp. before) the attack, $P_{f,i}^*(a) \in \mathbb{R}^n$ denotes the new (after the attack) generation profile, and $\tau \in (0,1]$ quantifies the flow increase.

We use the notation $(e, r) \in E \times (0, \tau]$ to describe attack goals satisfying (3). Since $\tau \in (0,1]$, we can have (in theory) an infinite number of attacks.

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)
number of attack goals. In practice, however, we study a finite number of attack goals $r$. For example, the attack goal $r$ 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 $r$.

In the SCED example, based on the attack goals, the target lines are selected differently. For example, a line connected to a critical substations during CCAs with an attack goal $(\kappa, \tau)$. Substation $s_k$'s measurements are selected differently. For example, a line connected to a critical substations during CCAs with an attack goal $(\kappa, \tau)$.

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 $\|a\|_2 \leq \bar{a}$ where $\bar{a} > 0$. We can use $\bar{a}$ as a design parameter to model different attack scenarios.

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

$$a_k \in \delta_k [-\bar{a}, \bar{a}], \quad \forall i \in V_{s_k}; \forall s_k \in S, \delta_k \in \{0, 1\},$$

where $\delta_k = 1$ if the adversary attacks $s_k$, and $\delta_k = 0$ if not. The vector $\delta(a, r) = [\delta_1, \delta_2, \ldots, \delta_n]^T$ describes the safe and target substations during CCAs with an attack goal $(e, r)$.

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

$$\|\delta(e, r)\|_2 \leq \kappa,$$  

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

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

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:

$$\max_{r, \kappa, \delta, a, P_g^\tau} \tau - \kappa,$$

s.t. \hspace{1cm} (3) - (5),

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

$$P_g^\tau(a) \in \arg \min_{P_g} \frac{1}{2} P_g^T C P_g + c^T P_g + c_a.$$  

s.t. \hspace{1cm} $1^T P_g - 1^T (P_g + a) = 0,$

$$A_0 P_g + A_1a - b \leq 0, \hspace{1cm} (7)$$

with

$$A_0 := [-I_n, I_n, 0, 0, -F], \quad A_1 = \begin{bmatrix} 0 \\ F \\ -F \end{bmatrix} - A_0 P_g.$$

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 ($\tau^*, \kappa^*, \delta^*, a^*, P_g^\tau$) if it exists, describes an adversary that targets the least number of substations ($\kappa^*$ and $\delta^*$) and maximises the flow increase ($\tau^*$) 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 $a$. 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

$$\max_{r, \kappa, \delta, a, P_g^\tau} \tau - \kappa,$$

s.t. \hspace{1cm} (3) - (5),

$$1^T P_g^\tau - 1^T (a + P_g) = 0,$$

$$C^T P_g^\tau + c_1 - 1^T \lambda^* + A_0^T \xi^* = 0,$$

$$A_0 P_g^\tau + A_1a - b \leq 0,$$

$$\xi^* \geq 0,$$

$$\lambda^* (A_0 P_g^\tau + A_1a - b) = 0.$$  

In the above, (8c) and (8g) are the KKT conditions of (7) and $\nu^*$ (resp. $\lambda^*$) 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
The security index and the CI for the attack goal target substations. This CI represents a strongly correlated CCA the Big M method [31]. Let since it relates the least number of target substations with the attack increase the flow (4). Though the security index (11) denote, respectively, the security index and the CI for the attack goal (e, τ). The security index index κ determines the least number of target substations to increase the flow (τ) on line e, while the CI δ(e, τ) 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, τ).

5.3 Algorithm: deriving the CIs

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

\[
\begin{align*}
\min & \quad \kappa, \\
\text{s.t.} & \quad \tau \geq \tau, \\
& \quad (4), (5), (8c) - (8f), (9) \quad \text{and} \quad (10).
\end{align*}
\]

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 program (11) and the \(\kappa^*\) - 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 \(S_{\alpha j} \in 2^\alpha\) describe the set of target substations during a CCA. If the CCA is effective, i.e. if the CCA increases the flow (τ) on line \(e\) \(\in E\), we use the notation \(S_{\alpha j} \rightarrow (e, \tau)\); otherwise, we use \(S_{\alpha j} \nrightarrow (e, \tau)\). We collect all effective CCA sets in the set

\[
\delta_a(e, \tau) := \{S_{\alpha j} \mid S_{\alpha j} \rightarrow (e, \tau)\} \subset 2^\alpha.
\]

The next proposition shows that if the CCA \(S_{\alpha j}\) fails to increase the flow (τ) on line \(e\), then all subordinated attacks \(S_{\alpha j} \in S_{\alpha j}\) also fail to increase the flow on \(e\).

**Proposition 1:** (Subordinated CCAs): If \(S_{\alpha j} \nrightarrow \delta_a(e, \tau)\), then \(S_{\alpha j} \nrightarrow \delta_a(e, \tau)\) for any \(S_{\alpha j} \in S_{\alpha j}\).

**Proof:** See the Appendix. □

**Definition 1:** Let \(\delta_a(e, \tau)\) denote a feasible solution of Algorithm 1 (Fig. 4) (i.e. \(\|\delta_a(e, \tau)\|_1 = \kappa^*\)). A CI, denoted as \(S_{\alpha j}\), is a strongly correlated CCA that extracts target substations from \(\delta_a(e, \tau)\) as follows:

\[
S_{\alpha j} := \{S_{\alpha j} \mid S_{\alpha j} \nbigvee (e, \tau) \neq 0\},
\]

and reaches the goal \(e, \tau\), i.e. \(S_{\alpha j} \in \delta_a(e, \tau)\).

**Proposition 2:** (Minimal cardinality): Let \(S_{\alpha j} \in \delta_a(e, \tau)\) be a CI. Then \(S_{\alpha j} \nrightarrow (e, \tau)\) for any \(S_{\alpha j} \in S_{\alpha j}\), satisfying \(|S_{\alpha j}| < \kappa^*\).

**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 \(|S_{\alpha j}| = \kappa^*\) and \(S_{\alpha j} \nbigvee \delta_a(e, \tau)\). We collect all the CIs in the set...
Theorem 1: (Multiple targets): Let \( J \) be an arbitrary index set, and let \( S_{e_j}^* \subseteq S_j(e, \tilde{r}) \) be CIs (for different lines) for all \( j \in J \). Suppose there exist a CCA targeting substations \( S_j^* \in 2^J \) such that \( S_{e_j}^* \subseteq S_j(e, \tilde{r}) \) for all \( j \in J \). Then the CCA \( S^* \) can increase the flow (\( \tilde{r} \)) on any line from the set \( \{ e_j \}_{j \in J}, \) i.e. \( S^* \subseteq S_j(e, \tilde{r}) \) for all \( j \in J \).

Proof: See the Appendix. □

Theorem 2: (Defense at a single substation): Let \( J \) be an arbitrary index set, and let \( S_{e_j}^* \subseteq S_j(e, \tilde{r}) \) be CIs (targeting different lines) for all \( j \in J \). Moreover, suppose the collection of target substations \( \{ S_{\alpha_j} \}_{j \in J} \) satisfies \( \cap_{k \in J} S_{\alpha_j} \neq \emptyset \). If the grid's defender protects measurements at the substation \( s_k^* \in \cap_{k \in J} S_{\alpha_j} \), then one of the following occurs (for all \( j \in J \)):

i. If \( s_{\alpha_j}^* \) is the unique CI that increases the flow (\( \tilde{r} \)) on line \( e_j \), then after \( s_{\alpha_j}^* \) is protected, the new CI \( S_{\alpha_j}^* \) (not necessarily unique) will require targeting more substations, i.e. \( \kappa_j^* := |S_{\alpha_j}^*| > |S_{\alpha_j}| \).

ii. If \( S_{\alpha_j}^* \) is not the only CI, then, after \( s_{\alpha_j}^* \) is protected, the following might occur: (a) if \( s_{\alpha_j}^* \) is common to all CIs \( S_{\alpha_j}^* \subseteq S_j(e, \tilde{r}) \), then the conclusion in (i) applies; or (b) if \( \{ s_{\alpha_j}^* \} \cap S_{\alpha_j} = \emptyset \), then the new collection of CIs, denoted as \( S_j^* \), satisfies \( S_j^* \subseteq S_j(e, \tilde{r}) \).

iii. The attack is infeasible, i.e. \( S_j^* \subseteq S_j(e, \tilde{r}) \) \( \equiv \{ \emptyset \} \).

Proof: See the Appendix. □

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 (\( \tilde{r} \)) on line \( e \), i.e. \( S_{\alpha_j} \subseteq S_{\alpha_j}(e, \tilde{r}) \). 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. \( S_{\alpha_j} \subseteq S_{\alpha_j}(e, \tilde{r}) \). If \( S_{\alpha_j} \) is a CI (i.e. \( S_{\alpha_j} \subseteq S_{\alpha_j}(e, \tilde{r}) \)), 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 \( s_j^* \) and a CCA \( S_{\alpha_j} \) 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 \( s_k^* \in S_{\alpha_j} \setminus S_{\alpha_j} \).

6.2.2 Best defence against CIs: Suppose \( \kappa^* \) denotes the security index for the attack goal (\( e, \tilde{r} \)). Then, the best defence against CIs (i.e. \( S_{\alpha_j}(e, \tilde{r}) \)) is to protect the minimal set of substations \( D_{\beta, e} \) that renders the new security index \( \kappa^* \) greater than \( \kappa^* \). Thus, the adversary is required to attack more substations after \( D_{\beta, e} \) is protected. Note that Theorem 2(ii) 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 \( \kappa^*(e, \tilde{r}) \) denotes the security index for the attack goal (\( e, \tilde{r} \)). The security index measures the likelihood of a CCA since it is less likely to attack more substations than \( \kappa^*(e, \tilde{r}) \). We define the average likelihood to increase the flow (\( \tilde{r} \)) on all lines as

\[
R(\tilde{r}) = \frac{1}{m} \sum_{k} \kappa^*(e, \tilde{r}),
\]

Using the average likelihood, we have the following definition.

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

\[
\Delta R_{s_k}(\tilde{r}) := R_{s_k}(\tilde{r}) - R(\tilde{r}),
\]

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

Theorem 2 implies that \( \Delta R_{s_k}(\tilde{r}) \geq 0 \) 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 \( \Delta R_{s_k}(\tilde{r}) \in \arg\max_{s_k \in S} \{ \Delta R_{s_k}(\tilde{r}) \} \). Note that we derive the metric for a specific flow increase value (\( \tilde{r} \)), but we can always derive it for the case when \( \tilde{r} \in (0, \hat{r}] \) is a free parameter for all target lines.

Remark 4: If we integrate \( \kappa^* \) (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 (\( e, \tilde{r} \)). 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.
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 \( n_e = 6 \) substations. We selected two target lines \( e = (2, 25) \) and \( e' = (16, 21) \); and the target flow increase \( r \in \{2.5\%, 5\%, 7.5\%, 10\%\} \). The line \( e \in E \) (resp. \( e' \in E \)) connects substations \( s_i \) and \( s_j \) (resp. \( s_{i'} \) and \( s_{j'} \)), 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 \( M, M_w = 10^3 \), \( \alpha = 0.1 \), 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 |
|-------------|-----|----------------|
| \( \delta'_{e, 2.5\%} \) | \{2\} | \( \kappa = 1 \) |
| \( \delta'_{e, 2.5\%} \) | \{5, 6\} | \( \kappa' = 2 \) |
| \( \delta'_{e, 5\%} \) | \{2.5, 6, 4.5, 6\} | \( \kappa = 3 \) |
| \( \delta'_{e, 5\%} \) | \{4.5, 6\} | \( \kappa' = 3 \) |
| \( \delta'_{e, 7.5\%} \) | \{1.2, 5, 6\} | \( \kappa = 4 \) |
| \( \delta'_{e, 7.5\%} \) | \{\} | \( \kappa' = 0 \) |
| \( \delta'_{e, 10\%} \) | \{\} | \( \kappa' = 0 \) |

Note: \( \delta'_{e, \epsilon} \) (resp. \( \delta'_{e, \epsilon} \)) denotes the CIs before (resp. after) protecting \( s_i = 2 \). \( \{\} \) (or \( \kappa' = 0 \)) implies the attack is ineffective.

### Table 2: CIs for line \( e' = (21, 24) \)

| Attack goal | CIs | Security index |
|-------------|-----|----------------|
| \( \delta'_{e', 2.5\%} \) | \{2\} | \( \kappa = 1 \) |
| \( \delta'_{e', 2.5\%} \) | \{5, 6\} | \( \kappa' = 2 \) |
| \( \delta'_{e', 5\%} \) | \{4.5, 6\} | \( \kappa = 3 \) |
| \( \delta'_{e', 5\%} \) | \{4.5, 6\} | \( \kappa' = 3 \) |
| \( \delta'_{e', 7.5\%} \) | \{\} | \( \kappa = 0 \) |
| \( \delta'_{e', 7.5\%} \) | \{\} | \( \kappa' = 0 \) |

Note: \( \delta'_{e', \epsilon} \) (resp. \( \delta'_{e', \epsilon} \)) denotes the CIs before (resp. after) protecting \( s_i = 2 \). \( \{\} \) (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 \( (e, \bar{e}) \) and \( (e', \bar{e}) \) 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 \( (e, \bar{e}) \).

#### 7.2 Experiment 2: CIs dependence on the parameter \( \bar{a} \)

In this experiment, we studied the CIs’ dependence on \( a \). Fig. 7 shows how the security index \( \kappa' \) 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 \( (e, \bar{e}) \) and \( (e', \bar{e}) \), before and after protecting substation \( s_i = 2 \). Before protecting substation \( s_i = 2 \), the CIs have the following defence implications: for the attack goal \( (e, 5\%) \), subordinated attacks of the CI \( S_j = \{2, 5\} \) are ineffective (Propositions 1 and 2), and the CCA \( S = \{2, 4, 5, 6\} \) can increase the flow \( \bar{f} = 5\% \) on both lines \( e \) and \( e' \) (Lemma 1 and Theorem 1). And, after protecting substation \( s_i = 2 \), the CIs have the following defence implications: for the attack goal \( (e', 2.5\%) \), the new security index satisfies \( \kappa' = 2 > 1 = \kappa \) (Theorem 2 (ii)), \( \delta'_{e', 5\%} \subseteq \delta'_{e', 5\%} \subseteq \delta'_{e', 5\%} \subseteq \delta'_{e', 5\%} \subseteq \delta'_{e', 5\%} \subseteq \delta'_{e', 5\%} \). Note: \( \delta'_{e', \epsilon} \) (resp. \( \delta'_{e', \epsilon} \)) denotes the CIs before (resp. after) protecting \( s_i = 2 \).

#### 7.4 Experiment 4: The metric of defence effectiveness

In this final experiment, we computed the metric of defence effectiveness \( \Delta R_{e, \epsilon}(\bar{f}) \) for all substations \( s_i \in S \) and \( \bar{f} \in \{5\%, 7.5\%\} \). Table 3 presents the results. These results imply that the best defence is achieved by protecting substation \( s_2 \).

| Table 3: The metric of defence effectiveness |
|-------------------------------------------|
| Attack goal | CIs | Security index |
|-------------|-----|----------------|
| \( \delta'_{e, 2.5\%} \) | \{2\} | \( \kappa = 1 \) |
| \( \delta'_{e, 2.5\%} \) | \{5, 6\} | \( \kappa' = 2 \) |
| \( \delta'_{e, 5\%} \) | \{2.5, 6, 4.5, 6\} | \( \kappa = 3 \) |
| \( \delta'_{e, 5\%} \) | \{4.5, 6\} | \( \kappa' = 3 \) |
| \( \delta'_{e, 7.5\%} \) | \{1.2, 5, 6\} | \( \kappa = 4 \) |
| \( \delta'_{e, 7.5\%} \) | \{\} | \( \kappa' = 0 \) |
| \( \delta'_{e, 10\%} \) | \{\} | \( \kappa' = 0 \) |

Note: \( \delta'_{e, \epsilon} \) (resp. \( \delta'_{e, \epsilon} \)) denotes the CIs before (resp. after) protecting \( s_i = 2 \). \( \{\} \) (or \( \kappa' = 0 \)) implies the attack is ineffective.
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/ICPSCom), 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. 58–68
[3] Lee, R.M., Assante, M.J., Conway, M.J.: ‘Cyber-physical threat assessment’, IEEE Trans. Smart Grid, 2011, 2 (4), pp. 741–749
[4] Wang, P., Ashok, A., Govindaraju, M.: ‘Cyber-physical risk assessment for smart grid system protection scheme’, IEEE Trans. Power & Energy Society General Meeting, Denver, CO USA, 2015, pp. 1–5
[5] Ten, C.W., Liu, C.C., Govindaraju, M.: ‘Vulnerability assessment of cyber-security for SCADA systems using attack trees’. Power Engineering Society General Meeting, Tampa, FL USA, 2007, pp. 1–8
[6] Liu, N., Zhang, J., Zhang, H., et al.: ‘Security assessment for communication networks of power control systems using attack graph and mcdc’, IEEE Trans. Power Syst., 2010, 25 (3), pp. 1492–1500
[7] 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. 1235–1243
[8] Xie, L., Mo, Y., Sinopoli, B.: ‘Integrity data attacks in power market operations’, IEEE Trans. Smart Grid, 2011, 2 (4), pp. 659–666
[9] 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. 1364–1375
[10] 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
[11] Liang, G., Zhao, J., Luo, F., et al.: ‘A review of false data injection attacks against modern power systems’, IEEE Trans. Smart Grid, 2016, 7 (3), pp. 991–1001

Table 3 Metric of defence effectiveness

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

Proof: (Proposition 2): Assume $S_{a_{j}} \subset S_{d}(e, \tau)$. We partition the CI in two arbitrary disjoint sets, i.e. $S_{a_{j}} = S_{a_{j}} \cup D$ satisfying $|S_{a_{j}}| \leq s_{n} \leftarrow |S_{a_{j}}|$. Since CIs are minimum cardinality CCAs, then $S_{a_{j}} \not\subset D_{d}(e, \tau)$, which proves the proposition.

Proof: (Lemma 1): Suppose, to get a contradiction, $S_{a_{j}} \subset D_{d}(e, \tau)$; then any super-set $S_{a_{j}}$ of $S_{a_{j}}$, i.e. $S_{a_{j}} \subset S_{a_{j}}$, is effective, i.e. $S_{a_{j}} \subset D_{d}(e, \tau)$. In particular, $S_{a_{j}} \subset S_{a_{j}}$ does not reach the goal $(e, \tau)$, which contradicts $S_{a_{j}} \not\subset D_{d}(e, \tau)$.

Proof: (Theorem 1): We partition the CCA $S_{a_{j}}$ in the union of two disjoint sets, i.e. $S_{a_{j}} = S_{a_{j}} \cup (S_{a_{j}} \not\subset D_{d}(e, \tau))$ for all $j$, i.e. $S_{a_{j}}$ do not reach $(e, \tau)$, which proves the theorem.

Proof: (Theorem 2): Assume the operator protects substation $s_{j}$, Moreover, assume that $S_{a_{j}} \subset D_{d}(e, \tau)$ 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 $S_{\beta, j} \in \Delta(d(e, \tilde{r}))$, which contradicts the fact that $\Delta(d(e, \tilde{r})) = \{S_{\alpha, j}\}$. This proves (i). On the other hand, suppose that $S_{\alpha, j} \in S_{\beta, j}$ is not unique. Part (ii-a) can be proven using the same arguments as in (i). We prove part (ii-b) as follows. Define $\delta := \{S_{\alpha, j} \in \Delta(d(e, \tilde{r})) \mid \delta \cap S_{\alpha, j} = \emptyset\}$. Then we partition the collection $\Delta(d(e, \tilde{r}))$ as follows $\Delta(d(e, \tilde{r})) = \delta \cup (\Delta(d(e, \tilde{r})) \setminus \delta)$. Proposition 1 implies that after defense $S_{\alpha, j} \setminus \{s_k\} \notin \Delta(d(e, \tilde{r}))$ for all $S_{\alpha, j} \in (\Delta(d(e, \tilde{r})) \setminus \delta)$. Thus, $\Delta(d(e, \tilde{r})) \equiv \delta$, and therefore $\Delta(d(e, \tilde{r})) \subseteq \Delta(d(e, \tilde{r}))$. This proves (ii-b). Finally, (c) follows trivially.