Power Grid State Estimation under General Cyber-Physical Attacks

Yudi Huang, Ting He, Nilanjan Ray Chaudhuri, and Thomas La Porta
Pennsylvania State University, University Park, PA 16802, USA
Email: {yxh5389, tzh58, nuc88, tfl12}@psu.edu

***Abstract*—Effective defense against cyber-physical attacks in power grid requires the capability of accurate damage assessment within the attacked area. While some solutions have been proposed to recover the phase angles and the breaker status of lines within the attacked area, existing solutions made the limiting assumption that the grid stays connected after the attack. To fill this gap, we study the problem of recovering the phase angles and the breaker status under a general cyber-physical attack that may partition the grid into islands. To this end, we (i) show that the existing solutions and recovery conditions still hold if the post-attack power injections in the attacked area are known, and (ii) propose a linear programming-based algorithm that can perfectly recover the breaker status under certain conditions even if the post-attack power injections are unknown. Our numerical evaluations based on the Polish power grid demonstrate that the proposed algorithm is highly accurate in localizing failed lines even though the conditions for perfect recovery can be hard to satisfy.**

***Index Terms*—Power grid state estimation, cyber-physical attack, failure localization.**

I. Introduction

Modern power grids are interdependent cyber-physical systems consisting of a power transmission system (power lines, substations, etc) and an associated control system (Supervisory Control and Data Acquisition - SCADA and Wide-Area Monitoring Protection and Control - WAMPAC) that monitors and controls the status of the power grid. This interdependency raises a legitimate concern: what happens if an attacker attacks both the physical grid and its control system simultaneously? The resulting attack, known as a *joint cyber-physical attack*, can cause devastating damage and large-scale blackouts, as the cyber attack can blindfold the control system and thus make the physical attack on the power grid more effective. For example, one such attack on Ukraine's power grid left 225,000 people without power for days [1].

The potential severity of cyber-physical attacks has attracted efforts in countering these attacks [2], [3]. One of the challenges in dealing with such attacks is that as the cyber attack blocks measurements (e.g., phase angles, breaker status, and so on) from the attacked area, the control center is unable to accurately identify the damage caused by the physical attack (e.g., which lines are disconnected) and hence unable to make accurate mitigation decisions. To address this challenge, solutions have been proposed to estimate the state of the power grid inside the attacked area using power flow models. Specifically, [2] developed methods to estimate the grid state under cyber-physical attacks using the *direct-current (DC) power flow model*, and [3] developed similar methods using the *alternating-current (AC) power flow model*. Both works made the *limiting assumption* that either (i) the grid remains connected after the attack, or (ii) the control center is aware of the supply/demand in each island formed after the attack, both leading to known post-attack active power injection at each bus.

In practice, however, disconnecting lines within the attacked area may cause partitioning of the grid and changes in the active power injections, and such changes within the attacked area will not be directly observable to the control center due to the cyber attack. Our goal is thus to estimate the power grid state, especially the breaker status of lines, under cyber-physical attacks without the above assumption.

A. Related Work

Power grid state estimation, as a key functionality for supervisory control, has been extensively studied in the literature [4]. Among them, secure state estimation under attack is of particular interest [5]. Specifically, the attackers can distort sensor data with noise [6] or inject stealthy data [7] so that the control center cannot correctly estimate the phase angles [8] or the topology [9] of the power grid. Recently, joint cyber-physical attack has gained attention, as the physical effect of such attack is harder to detect due to the cyber attack [2], [10], [11].

In particular, several approaches have been proposed for detecting failed links. In [12], [13], the problem was formulated as a mixed integer program, which becomes computationally inefficient when multiple links fail. Then, the problem was formulated as a sparse recovery problem over an overcomplete representation in [14], [15], where the combinatorial sparse recovery problem was relaxed to a linear programming (LP) problem. Based on this approach, the work in [2] further established graph-theoretic conditions for accurately recovering the failed links. All the algorithms in [2], [14], [15] aimed to find the sparsest solution among the feasible solutions under the assumption that the power grid remains connected after failure.

B. Summary of Contributions

We aim at estimating the power grid state within the attacked area, focusing on the phase angles and the breaker status of lines, with the following contributions:

1) We show that an existing rank-based condition for recovering the phase angles, previously established when

This work was supported by the National Science Foundation under award ECCS-1836827.

the grid remains connected after the attack, still holds without this limiting assumption.

2) We show that existing graph-theoretical conditions for localizing the failed lines, previously established under the same limiting assumption, still hold without this assumption if the post-attack power injections are known.
3) When the post-attack power injections are unknown, we develop an LP-based algorithm that can localize all the failed lines under certain conditions.
4) Our evaluations on a real grid topology show that while the conditions for perfect state estimation can be hard to satisfy, our proposed algorithm can localize the failed lines with a high accuracy.

Roadmap. Section II formulates our overall problem, which is divided into three subproblems addressed in Sections III–V. Then Section VI evaluates our solutions, and Section VII concludes the paper. All the proofs are in [16].

II. Problem Formulation

A. Power Grid Model

We adopt the DC power flow model, which is a relaxation of the AC power flow model that is commonly used in analyzing large power grids [14]. In this model, the power grid is modeled as a connected undirected graph $G = (V, E)$, where V is the set of nodes (buses) and E the set of links (transmission lines). Each link $e = (s, t)$ is associated with a *reactance* r_{st} ($r_{st} = r_{ts}$). Each node v is associated with a phase angle θ_v and an active power injection p_v. The phase angles $\boldsymbol{\theta} := (\theta_v)_{v\in V}$ and the active powers $\boldsymbol{p} := (p_v)_{v\in V}$ are related by

$$\boldsymbol{B\theta} = \boldsymbol{p}, \tag{1}$$

where $\boldsymbol{B} := (b_{uv})_{u,v\in V} \in \mathbb{R}^{|V|\times|V|}$ is the *admittance matrix*, defined as:

$$b_{uv} = \begin{cases} 0 & \text{if } u \neq v, (u,v) \notin E, \\ -1/r_{uv} & \text{if } u \neq v, (u,v) \in E, \\ -\sum_{w\in V\setminus\{u\}} b_{uw} & \text{if } u = v. \end{cases} \tag{2}$$

Given an arbitrary orientation of the links, the topology of G can also be represented by the *incidence matrix* $\boldsymbol{D} \in \{-1, 0, 1\}^{|V|\times|E|}$, whose (i, j)-th entry is defined as

$$d_{ij} = \begin{cases} 1 & \text{if link } e_j \text{ comes out of node } v_i, \\ -1 & \text{if link } e_j \text{ goes into node } v_i, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

We assume that each node is deployed with a phasor measurement unit (PMU) measuring the phase angle and remote terminal units (RTUs) measuring the breaker status along with the active power injection at this node and the power flows in its incident links. These reports are sent to the control center via a SCADA or WAMPAC system. The PMU data is assumed to be communicated over a relatively secure dedicated link and the RTU measurements over a more vulnerable SCADA network to the control center.

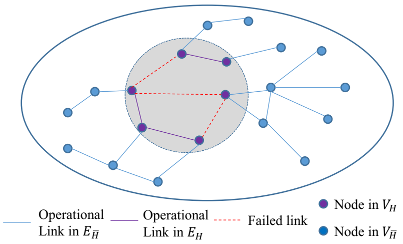


Figure 1. A cyber-physical attack that blocks information from the attacked area H while disconnecting certain lines within H.

Table I
Notations

Notation	Description
$G = (V, E)$	power grid
H, $\bar{H}$	attacked/unattacked area
F	set of failed links
$\boldsymbol{B}$	admittance matrix
$\boldsymbol{D}$	incidence matrix
$\boldsymbol{\theta}$	vector of phase angles
$\boldsymbol{p}$	vector of active power injections
$\boldsymbol{\Delta}$	vector of changes in active power injections

B. Attack Model

As illustrated in Fig. 1, an adversary attacks an area H of the power grid by: (i) blocking reports from the nodes within H (cyber attack on both SCADA and WAMPAC), and (ii) disconnecting a set F ($|F| > 0$) of links within H (physical attack). Formally, $H = (V_H, E_H)$ is a subgraph induced by a set of nodes V_H, where E_H is the set of links for which both endpoints are in V_H. Note that H does not have to be connected.

C. State Estimation Problem

Notation. The main notations are summarized in Table I. Moreover, given a subgraph X of G, V_X and E_X denote the subsets of nodes/links in X, and $\boldsymbol{x}_X$ denotes the subvector of a vector $\boldsymbol{x}$ containing elements corresponding to X. Similarly, given two subgraphs X and Y of G, $\boldsymbol{A}_{X|Y}$ denotes the submatrix of a matrix $\boldsymbol{A}$ containing rows corresponding to X and columns corresponding to Y. We use $D_H \in \{-1, 0, 1\}^{|V_H|\times|E_H|}$ to denote the incidence matrix of the attacked area H. For each quantity x, we use x' to denote its value after the attack.

Goal. Our goal is to recover the post-attack phase angles $\boldsymbol{\theta}'_H$ and localize the failed links F within the attacked area, based on the state variables before the attack and the measurements from the unattacked area $\bar{H}$ after the attack.

In contrast to the previous works, we consider cases where the attack may partition the grid into multiple islands, which can cause changes in active power injections to maintain the supply/demand balance in each island. Let $\boldsymbol{\Delta} = (\Delta_v)_{v\in V} := \boldsymbol{p} - \boldsymbol{p}'$ denote the change in active power injections, where $\Delta_v > 0$ if v is a generator bus and $\Delta_v < 0$ if v is a load bus.

III. Recovery of Phase Angles

Under the assumption that G remains connected after the attack and thus $\boldsymbol{\Delta} = \boldsymbol{0}$, [2] showed that the post-attack phase angles in the attacked area $\boldsymbol{\theta}'_H$ can be recovered if the submatrix $\boldsymbol{B}_{\bar{H}|H}$ of the admittance matrix has a full column rank. Below, we will show that the same condition actually holds without this limiting assumption.

Specifically, we have the following lemma (see proof in [16]) that extends [2, Lemma 1] to the case of arbitrary $\mathbf{\Delta}$. Here "supp" returns the indices of the non-zero entries.

Lemma III.1. *supp*$(\boldsymbol{B}(\boldsymbol{\theta}-\boldsymbol{\theta}')-\mathbf{\Delta})\subseteq V_H$.

Using Lemma III.1, we can prove that the recovery condition in [2, Theorem 1] remains sufficient even if the assumption of $\mathbf{\Delta}=0$ may not hold (see proof in [16]).

Theorem III.1. *The phase angles $\boldsymbol{\theta}'_H$ within the attacked area can be recovered correctly if $\boldsymbol{B}_{\bar{H}|H}$ has a full column rank.*

IV. Localizing Failed Links with Known Active Powers

Now assume that the post-attack phase angles $\boldsymbol{\theta}'$ have been recovered. They can be inferred when $\boldsymbol{B}_{\bar{H}|H}$ has a full column rank, or directly reported to the control center by PMUs if only SCADA has been compromised[1]. We will show that as long as the change in active powers $\mathbf{\Delta}$ is known, the failed links can be uniquely localized under the same conditions as specified in [2].

First, we note that under practical assumptions, the conditions presented in Section III for recovering the phase angles greatly simplify the recovery of the active powers. To this end, we assume that the adjustment of active power injections at generator/load buses follows the *proportional load shedding/generation reduction policy*, where (i) either the load or the generation (but not both) will be reduced upon formation of an island, and (ii) if nodes u and v are in the same island and of the same type (both load or generator), then $p'_u/p_u = p'_v/p_v$. This policy models the common practice in adjusting load/generation in the case of islanding [18], [19], and can help recovering the active powers in the following cases (see proof in [16]).

Lemma IV.1. *Let $N(v;\bar{H})$ denote the set of all the nodes in $\bar{H}$ that are connected to node v via links in $E\setminus E_H$. Then under the proportional load shedding policy, Δ_v for $v\in V_H$ can be recovered unless $N(v;\bar{H})=\emptyset$ or every $u\in N(v;\bar{H})$ is of a different type from v with $\Delta_u=0$.*

Remark: Under the condition of Theorem III.1, i.e., $\boldsymbol{B}_{\bar{H}|H}$ has a full column rank, each $v\in V_H$ must be the neighbor of at least one node in $\bar{H}$ (otherwise its corresponding column in $\boldsymbol{B}_{\bar{H}|H}$ will be $\mathbf{0}$), and thus $N(v;\bar{H})\neq\emptyset$. Moreover, majority of the nodes in practice are load buses, and thus each node in H is likely to be a load bus neighboring to another load bus in $\bar{H}$. Thus, we can usually recover $\mathbf{\Delta}_H$ under the proportional load shedding policy if the condition for recovering $\boldsymbol{\theta}'_H$ holds.

Next, we will establish the conditions for localizing the failed links F with known $\boldsymbol{\theta}'$ and $\mathbf{\Delta}$. The basic observation is the following property of the set F, proved in [16].

Lemma IV.2. *There exists a vector $\boldsymbol{x}\in\mathbb{R}^{|E_H|}$ that satisfies supp$(\boldsymbol{x})=F$, and*

$$\boldsymbol{D}_H\boldsymbol{x}=\boldsymbol{B}_{H|G}(\boldsymbol{\theta}-\boldsymbol{\theta}')-\mathbf{\Delta}_H. \tag{4}$$

This lemma, which replaces [2, Lemma 2], implies that if one can find the conditions under which the solution to (4) is unique, then the links corresponding to non-zero elements of this solution must be the failed links. To this end, [2] gave a set of graph-theoretic conditions. As these conditions are only about the solution space of $\boldsymbol{D}_H\boldsymbol{x}=\boldsymbol{y}$, they remain valid in our setting as long as the righthand side is known. We summarize these conditions below (see proof in [16]).

Theorem IV.1. *The failed links F within the attacked area can be localized correctly if:*

1) *H is acyclic (i.e., a tree or a set of trees), in which case (4) has a unique solution $\boldsymbol{x}$ for which supp$(\boldsymbol{x})=F$, or*
2) *H is a planar graph satisfying (i) for any cycle C in H, $|C\cap F|<|C\setminus F|$, and (ii) F^* is H^*-separable[2], in which case the optimization $\min\|\boldsymbol{x}\|_1$ s.t. (4) has a unique solution $\boldsymbol{x}$ for which supp$(\boldsymbol{x})=F$.*

Special cases satisfying the second condition in Theorem IV.1 include that (i) H is a cycle in which majority of the links have not failed, and (ii) H is a planar bipartite graph in which each cycle contains fewer failed links than non-failed links [2].

V. Localizing Failed Links with Unknown Active Powers

Although providing strong theoretical guarantees, the solutions for localizing failed links given in Section IV are only applicable to small attacked areas with simple topologies (e.g., trees or cycles in which every node is connected to another node outside the attacked area). To deal with larger attacked areas, we investigate alternative solutions and their accuracy in localizing the failed links. In this section, we tackle the joint estimation of the failed links F and the change in active power injections within the attacked area $\mathbf{\Delta}_H$. As in Section IV, we assume that the post-attack phase angles $\boldsymbol{\theta}'$ are known, which can be either inferred or directly measured.

A. Solution

Our approach is to formulate the joint estimation problem as an optimization as follows.

Constraints: Let $\boldsymbol{x}\in\{0,1\}^{|E|}$ be an indicator vector such that $x_e=1$ if and only if $e\in F$. Due to $\boldsymbol{B}=\boldsymbol{D}\mathbf{\Gamma}\boldsymbol{D}^T$ (where $\mathbf{\Gamma}:=\text{diag}\{\frac{1}{r_e}\}_{e\in E}$), we can write the post-attack admittance matrix as $\boldsymbol{B}'=\boldsymbol{B}-\boldsymbol{D}\mathbf{\Gamma}\text{diag}\{\boldsymbol{x}\}\boldsymbol{D}^T$, which implies

$$\mathbf{\Delta}_H=\boldsymbol{B}_{H|G}(\boldsymbol{\theta}-\boldsymbol{\theta}')+\boldsymbol{D}_H\mathbf{\Gamma}_H\text{diag}\{\boldsymbol{D}^T_{G|H}\boldsymbol{\theta}'\}\boldsymbol{x}_H, \tag{5}$$

where $\boldsymbol{D}_{G|H}\in\{-1,0,1\}^{|V|\times|E_H|}$ is the submatrix of the incidence matrix $\boldsymbol{D}$ only containing the columns corresponding to links in H. For simplicity, we define

$$\tilde{\boldsymbol{D}}:=\boldsymbol{D}\mathbf{\Gamma}\text{diag}\{\boldsymbol{D}^T\boldsymbol{\theta}'\}. \tag{6}$$

[1] This can occur in hybrid control systems where the PMU measurements are reported via a modern WAMPAC system with stronger defenses and the other sensor measurements are reported via a legacy SCADA system that is more vulnerable to cyber attacks [17].

[2] Here H^* is the dual graph of H, and F^* is the set of edges in H^* such that each edge in F^* connects a pair of vertices that correspond to adjacent faces in H separated by a failed link.

Let each link be oriented in the same direction as the post-attack power flow. Then, for link $e_k = (i,j)$ where post-attack power flows from i to j, $(\tilde{\boldsymbol{D}})_{i,k} = \frac{\theta'_i - \theta'_j}{r_{ij}}$ and $(\tilde{\boldsymbol{D}})_{j,k} = -\frac{\theta'_i - \theta'_j}{r_{ij}}$, where $\frac{\theta'_i - \theta'_j}{r_{ij}}$ is the post-attack power flow on link e_k if it has not failed.

Besides (5), $\boldsymbol{\Delta}_H$ is also constrained as

$$p_v \geq \Delta_v \geq 0, \ \ \forall v \in \{u \,|u \in V_H, p_u > 0\}, \tag{7a}$$

$$p_v \leq \Delta_v \leq 0, \ \ \forall v \in \{u \,|u \in V_H, p_u \leq 0\}, \tag{7b}$$

$$\boldsymbol{1}^T \boldsymbol{\Delta} = 0, \tag{7c}$$

which ensures that a generator/load bus will remain of the same type after the attack, and the total power is balanced. It is worth noting that (7c) is ensured by (5), which implies that $\boldsymbol{1}^T\boldsymbol{\Delta}_H - \boldsymbol{1}^T\boldsymbol{B}_{H|G}(\boldsymbol{\theta}-\boldsymbol{\theta}') = (\boldsymbol{1}^T\tilde{\boldsymbol{D}}_H)\boldsymbol{x}_H = 0$ since $\boldsymbol{1}^T\tilde{\boldsymbol{D}}_H = \boldsymbol{0}$ by definition (6). This implies that any $\boldsymbol{\Delta}_H$ satisfying (5) will satisfy $\boldsymbol{1}^T\boldsymbol{\Delta}_H = \boldsymbol{1}^T\boldsymbol{B}_{H|G}(\boldsymbol{\theta}-\boldsymbol{\theta}') = \boldsymbol{1}^T\boldsymbol{\Delta}^*_H$ ($\boldsymbol{\Delta}^*_H$: the ground-truth load shedding values in H), and thus satisfy (7c). Hence, we will omit (7c) in the sequel.

Objective: The problem of failure localization aims at finding a set $\hat{F}$ that is as close as possible to the set F of failed links, while satisfying all the constraints. The solution is generally not unique, e.g., if both endpoints of a link $l \in E_H$ are disconnected from $\bar{H}$ after the attack, then the status of l will have no impact on any observable variable, and hence cannot be determined. To resolve this ambiguity, we set our objective as using the fewest failed links to satisfy all the constraints, which is consistent with the previous approaches [2], [14], [15]. Mathematically, the problem is formulated as

$$\text{(P0)} \qquad \min_{\boldsymbol{x}_H, \boldsymbol{\Delta}_H} \ \boldsymbol{1}^T\boldsymbol{x}_H \tag{8a}$$

$$\text{s.t.} \quad (5),\ (7a)-(7b), \tag{8b}$$

$$x_e \in \{0,1\}, \quad \forall e \in E_H, \tag{8c}$$

where the decision variables are $\boldsymbol{x}_H$ and $\boldsymbol{\Delta}_H$.

Via a reduction from the *subset sum problem*, we characterize the complexity of (P0) (see proof in [16]).

Lemma V.1. *The optimization (P0) is NP-hard.*

To develop an efficient solution, we relax the integer constraint (8c), which turns (P0) into

$$\text{(P1)} \qquad \min_{\boldsymbol{x}_H, \boldsymbol{\Delta}_H} \ \boldsymbol{1}^T\boldsymbol{x}_H \tag{9a}$$

$$\text{s.t.} \quad (5),\ (7a)-(7b), \tag{9b}$$

$$\boldsymbol{0} \leq \boldsymbol{x}_H \leq \boldsymbol{1}. \tag{9c}$$

where $\boldsymbol{0} \leq \boldsymbol{x}_H \leq \boldsymbol{1}$ denotes element-wise inequality. The problem (P1) is a linear program (LP) which can be solved in polynomial time. Based on (P1), we propose an algorithm for localizing the failed links, given in Algorithm 1, where the input parameter $\eta \in (0,1)$ is a threshold for rounding the factional solution ($\eta = 0.5$ in our experiments).

B. Analysis

We now analyze when the proposed algorithm can correctly localize the failed links. In the sequel, $\boldsymbol{\Delta}^*$ denotes the ground-truth value of $\boldsymbol{\Delta}$ and $\boldsymbol{x}^*$ denotes the ground-truth value of $\boldsymbol{x}$ ($x^*_e = 1$ if $e \in F$ and $x^*_e = 0$ otherwise).

Algorithm 1: Failed Link Detection

```
input  : $\boldsymbol{B}, \boldsymbol{p}, \boldsymbol{\Delta}_{\bar{H}}, \boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{D}, \eta$
Output: $F$
Solve the problem (P1) to obtain $\boldsymbol{x}_H$;
Return $F = \{e : x_e \geq \eta\}$.
```

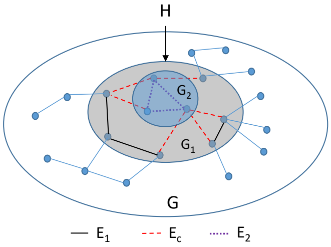


Figure 2. Decomposition of the attacked area H.

As illustrated in Fig. 2, we denote by G_1 the subgraph of H induced by nodes in V_H that stay connected to $\bar{H}$ after the attack, and the remaining part of H (if any) by G_2. Let V_i $(i = 1, 2)$ denote the set of nodes in G_i. Furthermore, we decompose each V_i into $V_{i,L}$ for nodes with $p_v \leq 0$ and $V_{i,G}$ for the rest. Define $E_i \subseteq E$ $(i = 1, 2)$ as the set of links with both endpoints in G_i, and $E_c \subseteq E$ as the remaining links that form a cut between G_1 and G_2, i.e., $\forall (s,t) \in E_c$ has $s \in G_1$ and $t \in G_2$ or vice versa.

Assumption 1. *We make the following assumptions:*

1) *$\forall e_c = (s,t) \in E_c$, where $t \in V_2$, $p'_t = 0$. This condition holds if node t is a load bus with zero load (i.e., $p_t = 0$), or lies in an island containing either no generator bus or no load bus after the attack.*
2) *As in [2], we assume that for each link $(s,t) \in E_H$, $\theta'_s \neq \theta'_t$, as otherwise the link will carry no power flow and hence its existence is not detectable.*

First, we simplify (P1) into an equivalent but simpler optimization problem. To this end, we combine the decision variables $\boldsymbol{\Delta}_H$ and $\boldsymbol{x}_H$ of (P1) into a single vector $\boldsymbol{y}_H = [\boldsymbol{\Delta}^T_H, \boldsymbol{x}^T_H]^T \in \mathbb{R}^{(|E_H|+|V_H|)}$ (where $[A, B]$ denotes horizontal concatenation), and explicitly represent the solution to $\boldsymbol{y}_H$ that satisfies (5). Notice that (5) can be written as $[\boldsymbol{I}, -\tilde{\boldsymbol{D}}_H]\boldsymbol{y}_H = \boldsymbol{B}_{H|G}(\boldsymbol{\theta}-\boldsymbol{\theta}')$ ($\boldsymbol{I}$: the $|V_H| \times |V_H|$ identity matrix). The ground-truth solution $\boldsymbol{y}^*_H = [(\boldsymbol{\Delta}^*_H)^T, (\boldsymbol{x}^*_H)^T]^T$ certainly satisfies (5). Next, consider the null space of $[\boldsymbol{I}, -\tilde{\boldsymbol{D}}_H]$, whose dimension is $|E_H|$. It is easy to verify that $[\tilde{\boldsymbol{d}}^T_e, \boldsymbol{u}^T_e]^T$ $(e \in E_H)$ are $|E_H|$ independent vectors spanning the null space of $[\boldsymbol{I}, -\tilde{\boldsymbol{D}}_H]$, where $\tilde{\boldsymbol{d}}_e$ is the column vector of $\tilde{\boldsymbol{D}}_H$ corresponding to link e, and $\boldsymbol{u}_e$ is a unit vector in $\mathbb{R}^{|E_H|}$ with the e-th element being 1 and the other elements being 0. Therefore, any $\boldsymbol{y}_H$ satisfying (5) can be expressed as

$$\boldsymbol{y}_H = \begin{bmatrix} \boldsymbol{\Delta}^*_H \\ \boldsymbol{x}^*_H \end{bmatrix} + \sum_{e \in E_H} c_e \begin{bmatrix} \tilde{\boldsymbol{d}}_e \\ \boldsymbol{u}_e \end{bmatrix}. \tag{10}$$

Let $\boldsymbol{c} := (c_e)_{e \in E_H} \in \mathbb{R}^{|E_H|}$. Let $\boldsymbol{\Lambda}_L \in \{0,1\}^{|V_H|}$ be a diagonal matrix such that $(\boldsymbol{\Lambda}_L)_{i,i} = 1$ if $p_i \leq 0$ and 0 otherwise, and $\boldsymbol{\Lambda}_G$ be defined similarly except that $(\boldsymbol{\Lambda}_G)_{i,i} = 1$ if $p_i \geq 0$. We can write (P1) into the following equivalent optimization of $\boldsymbol{c}$ to eliminate the equality constraint (5). *For notation simplicity, we omit the subscript $_H$ of $\tilde{\boldsymbol{D}}_H$, $\boldsymbol{\Delta}_H$ and $\boldsymbol{x}_H$ in the following* unless it causes confusion.

$$\min_{\boldsymbol{c}} \quad \boldsymbol{1}^T\boldsymbol{c} \tag{11a}$$

$$\text{s.t.} \quad -\boldsymbol{\Lambda}_L\tilde{\boldsymbol{D}}\boldsymbol{c} \geq \boldsymbol{\Lambda}_L\boldsymbol{\Delta}^*, \tag{11b}$$

$$\boldsymbol{\Lambda}_L\tilde{\boldsymbol{D}}\boldsymbol{c} \geq \boldsymbol{\Lambda}_L\boldsymbol{p} - \boldsymbol{\Lambda}_L\boldsymbol{\Delta}^*, \tag{11c}$$

$$\boldsymbol{\Lambda}_G\tilde{\boldsymbol{D}}\boldsymbol{c} \geq -\boldsymbol{\Lambda}_G\boldsymbol{\Delta}^*, \tag{11d}$$

$$-\boldsymbol{\Lambda}_G\tilde{\boldsymbol{D}}\boldsymbol{c} \geq -(\boldsymbol{\Lambda}_G\boldsymbol{p} - \boldsymbol{\Lambda}_G\boldsymbol{\Delta}^*), \tag{11e}$$

$$\boldsymbol{c} \geq -\boldsymbol{x}^*, \tag{11f}$$

$$-\boldsymbol{c} \geq \boldsymbol{x}^* - \boldsymbol{1}, \tag{11g}$$

For Algorithm 1 to correctly localize the failed links, it suffices to have $x_e^* + c_e \geq \eta$ for all $e \in F$ and $x_e^* + c_e < \eta$ for all $e \notin F$. Equivalently, it suffices to ensure that the optimal solution $\boldsymbol{c}^*$ to (11) satisfies $c_e^* \geq \eta - 1$ for all $e \in F$ and $c_e^* < \eta$ for all $e \notin F$.

Next, we use (11) to analyze the accuracy of Algorithm 1. Based on the decomposition of H, $\tilde{\boldsymbol{D}}$ can be written as

$$\tilde{\boldsymbol{D}} = \begin{array}{c} \\ V_{1,L} \\ V_{1,G} \\ V_{2,L} \\ V_{2,G} \end{array} \!\!\begin{array}{c} \begin{array}{ccc} E_1 & E_c & E_2 \end{array} \\ \begin{pmatrix} \tilde{\boldsymbol{D}}_{11,L} & \tilde{\boldsymbol{D}}_{1c,L} & \tilde{\boldsymbol{D}}_{12,L} \\ \tilde{\boldsymbol{D}}_{11,G} & \tilde{\boldsymbol{D}}_{1c,G} & \tilde{\boldsymbol{D}}_{12,G} \\ \tilde{\boldsymbol{D}}_{21,L} & \tilde{\boldsymbol{D}}_{2c,L} & \tilde{\boldsymbol{D}}_{22,L} \\ \tilde{\boldsymbol{D}}_{21,G} & \tilde{\boldsymbol{D}}_{2c,G} & \tilde{\boldsymbol{D}}_{22,G} \end{pmatrix} \end{array}. \tag{12}$$

Let $\tilde{\boldsymbol{D}}_{ij}$ denote $(\tilde{\boldsymbol{D}}_{ij,L}^T, \tilde{\boldsymbol{D}}_{ij,G}^T)^T$. It is easy to see that $\tilde{\boldsymbol{D}}_{12} = \boldsymbol{0}$ and $\tilde{\boldsymbol{D}}_{21} = \boldsymbol{0}$. Below, we will establish sufficient conditions under which Algorithm 1 can correctly identify the cut between G_1 and G_2, the proof of which can be found in [16].

Theorem V.1. *Under Assumption 1 and $F = E_c$, the ground-truth solution is the unique optimal solution to (P1), i.e., Algorithm 1 can correctly detect F, if the following two conditions hold:*

1) *$\tilde{\boldsymbol{D}}_{2c,L} \geq \boldsymbol{0}$ and $\tilde{\boldsymbol{D}}_{2c,G} \leq \boldsymbol{0}$, i.e., all the elements in $\tilde{\boldsymbol{D}}_{2c,L}$ are nonnegative and all the elements in $\tilde{\boldsymbol{D}}_{2c,G}$ are nonpositive, and*
2) *for each non-zero entry $(\tilde{D}_{2c})_{v,j}$ in row v of $\tilde{\boldsymbol{D}}_{2c}$, we have $|(\tilde{D}_{2c})_{v,j}| \geq (\tilde{D}_{22})_{v,k}$ for all k, i.e., the absolute value of each non-zero entry in $\tilde{\boldsymbol{D}}_{2c}$ is no less than the entries of $\tilde{\boldsymbol{D}}_{22}$ in the same row.*

Corollary V.1.1. *Assume Assumption 1 and $F = E_c$. If H contains no generator bus and $\tilde{\boldsymbol{D}}_{2c,L} \geq \boldsymbol{0}$, or H contains no load bus and $\tilde{\boldsymbol{D}}_{2c,G} \leq \boldsymbol{0}$, then the ground-truth solution is the unique optimal solution to (P1) and Algorithm 1 can correctly identify the cut between G_1 and G_2.*

VI. Performance Evaluation

We test our solutions on the Polish power grid ("Polish system - winter 1999-2000 peak") [20] with 2383 nodes and 2886 links, where parallel links are combined into one link. We generate the attacked area H by randomly choosing one node as a starting point and performing a breadth first search to obtain H with a predetermined $|V_H|$. We then randomly choose $|F|$ links within H to fail. The phase angles of each island without any generator or load are set to 0, and the rest are computed according to (1). We vary $|V_H|$ and $|F|$ to explore different settings, and for each setting, we generate 70 different H's and 300 different F's per H.

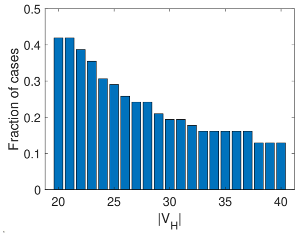


Figure 3. Prob. that conditions of Theorem IV.1.(1) hold.

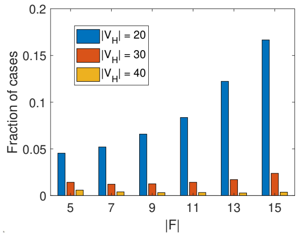


Figure 4. Prob. that $F = E_c$.

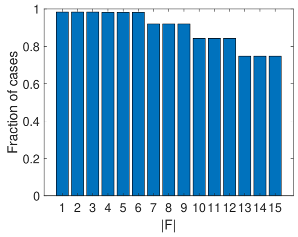


Figure 5. Prob. that conditions of Theorem V.1 hold if $F = E_c$ ($|V_2| \in [1, 4]$).

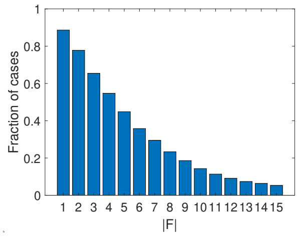


Figure 6. Prob. that setting $\boldsymbol{\Delta} = \boldsymbol{0}$ is feasible ($|V_H| = 40$).

We evaluate two types of metrics: (1) how often the theoretical recovery conditions are satisfied, and (2) how accurate Algorithm 1 is when its recovery conditions are not satisfied.

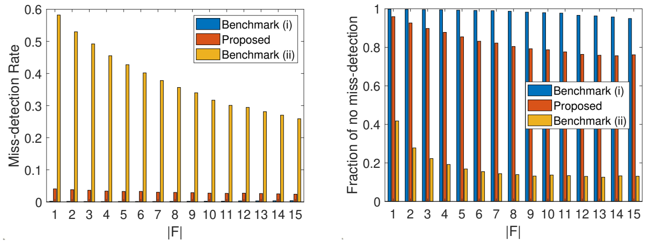
First, we evaluate the fraction of randomly generated cases satisfying the conditions in Theorem III.1 for recovering the phase angles, Theorem IV.1.(1)[3] for localizing the failed links with known phase angles and active powers, and Theorem V.1 for localizing the failed links with known phase angles and unknown active powers. We observe that (i) the condition in Theorem III.1 is almost never satisfied, (ii) the condition in Theorem IV.1.(1) is only satisfied with a limited probability as shown Fig. 3, which decreases with $|V_H|$ (note that Theorem IV.1.(1) does not depend on F), and (iii) the conditions in Theorem V.1 are only satisfied with a small probability due to the small probability of $F = E_c$, as shown in Fig. 4, which also decreases with $|V_H|$. However, Fig. 5 shows that the remaining conditions[4] of Theorem V.1 hold with high probability once $F = E_c$, indicating that Algorithm 1 can accurately detect the cut within H.

These results show that although it is possible to infer phase angles and failed links with perfect accuracy, guaranteeing perfect accuracy will require stringent conditions that are hard to satisfy, highlighting the need of protecting measurements.

For the second type of metrics, we focus on the accuracy of Algorithm 1 in comparison with benchmarks in localizing the failed links, assuming that PMU measurements are available due to their better protection against cyber attacks [17]. We

[3] We only tested condition (1) in Theorem IV.1, as the other condition relies on complicated graph properties that are difficult to test.

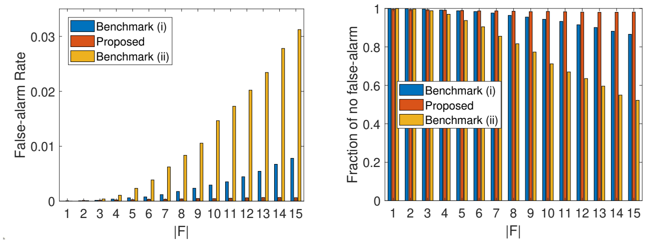
[4] Note that $|V_H|$ does not matter, as the conditions are only related to G_2.



(a) miss rate

(b) Probability of no miss

Figure 7. Performance comparison on miss rate ($|V_H| = 40$).



(a) false alarm rate

(b) Probability of no false alarm

Figure 8. Performance comparison on false alarm rate ($|V_H| = 40$).

consider two benchmarks: (i) the solution given in Theorem IV.1 (extended from [2]), i.e., estimating F by $\mathrm{supp}(\boldsymbol{x})$ for the solution to $\min \|\boldsymbol{x}\|_1$ s.t. (4), assuming the true $\boldsymbol{\Delta}_H$ to be known, and (ii) $\min \|\boldsymbol{x}\|_1$ s.t. $\|\boldsymbol{B}_{H|G}(\boldsymbol{\theta} - \boldsymbol{\theta}') - \boldsymbol{D}_H\boldsymbol{x}\|_2 \leq \|\boldsymbol{p}_H\|_2$, which is extended from the solution in [14], [15]. We note that the original solution in [2] (which assumes $\boldsymbol{\Delta} = \mathbf{0}$) is often infeasible for our problem, as shown in Fig. 6, and thus not used as a benchmark. Note that *benchmark (i) is meant to be a "performance upper bound"*, as it assumes more knowledge (i.e., $\boldsymbol{\Delta}_H$) than our proposed algorithm.

As shown in Fig. 7, benchmark (i) performs the best in miss rate, while Algorithm 1 performs much better than benchmark (ii). This confirms the importance of knowing or estimating load shedding values in failure localization. Fig. 8 shows that in terms of false alarm rate, Algorithm 1 performs even better than benchmark (i). This is because the decision variable $\boldsymbol{x}$ in benchmark (i) combines the information of both the failed links and the phase angles $\boldsymbol{\theta}'_H$, and thus does not fully exploit the knowledge of $\boldsymbol{\theta}'_H$.

VII. Conclusion

Observing that existing solutions for power grid state estimation under cyber-physical attacks relied on the limiting assumption that the grid stays connected after the attack, we revisited the problem without this limiting assumption, and showed that the existing solutions and conditions for recovering phase angles and breaker status remain valid as long as the post-attack power injections are known. We then focused on recovering the breaker status within the attacked area under unknown post-attack power injections, and proposed an LP-based algorithm that achieves perfect recovery under certain conditions. Our evaluations on Polish power grid showed that although the conditions for perfect recovery are hard to satisfy, our algorithm can always localize the failed lines with a high accuracy, which suggests that it admits a more general recovery condition. We leave the study of such conditions to future work.

References

[1] P. Fairley, "Cybersecurity at U.S. utilities due for an upgrade: Tech to detect intrusions into industrial control systems will be mandatory," *IEEE Spectrum*, vol. 53, no. 5, pp. 11–13, May 2016.

[2] S. Soltan, M. Yannakakis, and G. Zussman, "Power grid state estimation following a joint cyber and physical attack," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, p. 499–512, March 2018.

[3] S. Soltan and G. Zussman, "Power grid state estimation after a cyber-physical attack under the AC power flow model," in *IEEE PES-GM*, 2017.

[4] Y.-F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, "State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 33–43, 2012.

[5] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, pp. 1–33, 2011.

[6] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 4917–4932, 2017.

[7] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *2010 first IEEE international conference on smart grid communications*. IEEE, 2010, pp. 214–219.

[8] O. Vuković, K. C. Sou, G. Dán, and H. Sandberg, "Network-layer protection schemes against stealth attacks on state estimators in power systems," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2011, pp. 184–189.

[9] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1294–1305, 2013.

[10] R. Deng, P. Zhuang, and H. Liang, "Ccpa: Coordinated cyber-physical attacks and countermeasures in smart grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2420–2430, 2017.

[11] S. Soltan, M. Yannakakis, and G. Zussman, "React to cyber attacks on power grids," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 459–473, 2018.

[12] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1644–1652, 2008.

[13] ——, "Double line outage detection using phasor angle measurements," in *2009 IEEE Power & Energy Society General Meeting*. IEEE, 2009, pp. 1–5.

[14] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2215–2224, November 2012.

[15] J.-C. Chen, W.-T. Li, C.-K. Wen, J.-H. Teng, and P. Ting, "Efficient identification method for power line outages in the smart power grid," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1788–1800, 2014.

[16] Y. Huang, T. He, N. R. Chaudhuri, and T. L. Porta, "Power grid state estimation under general cyber-physical attacks," Technical Report, June 2020, https://sites.psu.edu/nsrg/files/2020/06/Huang20Report.pdf.

[17] "Wide area monitoring, protection, and control systems (WAMPAC) standards for cyber security requirements," National Electric Sector Cybersecurity Organization Resource (NESCOR), October 2012, https://smartgrid.epri.com/doc/ESRFSD.pdf.

[18] B. Pal and B. Chaudhuri, *Robust control in power systems*. Springer Science & Business Media, 2006.

[19] M. Lu, W. ZainalAbidin, T. Masri, D. Lee, and S. Chen, "Under-frequency load shedding (ufls) schemes-a survey," *International Journal of Applied Engineering Research*, vol. 11, no. 1, pp. 456–472, 2016.

[20] R. D. Zimmerman and C. E. Murillo-Sánchez, "Matpower 7.0 user's manual," *Power Systems Engineering Research Center*, vol. 9, 2019.