Link State Estimation under Cyber-Physical Attacks: Theory and Algorithms

Yudi Huang, *Student Member, IEEE*, Ting He, *Senior Member, IEEE*, Nilanjan Ray Chaudhuri, *Senior Member, IEEE*, and Thomas La Porta *Fellow, IEEE*

***Abstract*—Effective defense against cyber-physical attacks in power grids requires accurate damage assessment. While some solutions have been proposed to recover the line states within the attacked area, existing solutions are limited by the assumption of a connected post-attack grid and the lack of verifiable performance guarantees. To fill this gap, we study the recovery of line states under a cyber-physical attack that disconnects lines while blocking information from the attacked area. In contrast to existing solutions assuming a connected post-attack grid or known post-attack power injections, we consider a more challenging scenario where the attack may partition the grid into islands, which causes unknown changes in power injections. To address this problem, under the DC model we (i) propose a linear programming-based algorithm to recover the line states within the attacked area under unknown post-attack power injections, (ii) characterize the accuracy of the proposed recovery algorithm under certain conditions, and (iii) develop efficient algorithms to verify the recovery results using observable information. Then, we extend the DC-based algorithms to AC model. Our numerical evaluations demonstrate that the proposed recovery algorithm is highly accurate in localizing failed lines, most of which can be successfully verified by the proposed verification algorithms.**

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

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 states 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 large-scale blackouts, as the cyber attack can blindfold the control system and thus make the physical attack on the power grid more damaging. For example, one such attack on Ukraine's power grid left 225,000 people without power for days [2].

One of the challenges in dealing with such attacks is that in case the measurements (e.g., breaker status) within the attacked area are blocked by the cyber attacks, 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 efficiently schedule the repairing/restoration. To address this challenge, solutions [3]–[7] have been proposed to recover the state and topology of the power grid inside the attacked area under either direct-current (DC) or alternating-current (AC) power flow models. However, existing works are based on the limiting assumption that either the grid remains connected after attack, or the post-attack power injections at all the buses are known, which is often violated by large-scale attacks. In addition, most of the existing solutions lack the ability to verify the correctness of the recovered state in real time, which can result in a costly waste of time and resources during restoration due to false alarms.

As shown in Fig. 1, the defense system in the smart grid is usually composed of multiple subsystems [8], which can be categorized into pre-attack and post-attack subsystems. The pre-attack subsystem aims to prevent the attack from taking effect. For example, the prevention subsystem will reduce the information leakage, while the detection subsystem intends to detect the invasion. For the advanced attacks that can bypass all pre-attack modules, efficient recovery from the attacks, which is the focus of the work, is desired. A key step before repairing/restoration is to learn the current topology through line state estimation. In this work, we address this problem under a joint cyber-physical attack, where the cyber attack blocks information from an attacked area while the physical attack disconnects lines (i.e., transmission lines) within the area, with a focus on scenarios where the physical attack causes islanding and hence unknown changes in the power injections within the attacked area. Our first goal is to compute an estimate $\hat{F}$ of the failed lines within the attacked area (**Topology Recovery**). Then, in contrast to existing works [3], [4], [9], [10], we add a novel step called **Topology Verification** before repairing/restoration, to guide resource dispatch during repairing/restoration by avoiding the cost due to false alarms.

A. Related Work

Power grid state estimation, as a key functionality for supervisory control, has been extensively studied in the literature [11], [12]. Secure state estimation under attack is of particular interest [13], [14]. Specifically, the attackers can launch denial-of-service attacks [5], [8] or false data injection attacks [5], [6], [13]–[15] so that the control center cannot correctly estimate the phase angles [16] or/and the topology [17] of the power grid. Recently, joint cyber-physical attacks are gaining attention due to their stealthiness and severe consequences [5], [6].

The authors are with the School of Electrical Engineering and Computer Science, Pennsylvania State University, University Park, PA 16802, USA (e-mail: {yxh5389, tzh58, nuc88, tfl12}@psu.edu).

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

Aspects of this work were presented at SmartGridComm'20 [1].

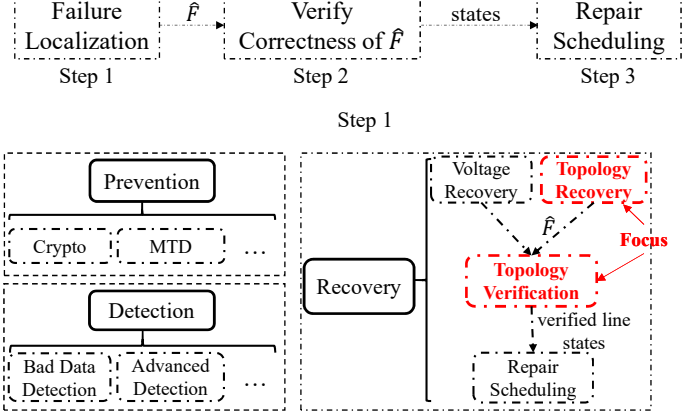


Figure 1. The defense system in smart grid. The proposed methods focus on **Topology Recovery** and **Topology Verification**.

The resilience of the power grid to attacks requires both pre-attack prevention [18] and post-attack restoration, the latter being the focus of this work. To facilitate the restoration, several approaches have been proposed for detecting failed lines. In [9], [10], the problem was formulated as a mixed-integer program, which becomes computationally inefficient when multiple lines fail. The problem was formulated as a sparse recovery problem over an overcomplete representation in [3], [4], where the combinatorial sparse recovery problem was relaxed to a linear programming (LP) problem. Machine learning-based recovery strategy was studied in [7]. All works discussed above assume DC power flow model. Recently, the detection of failed link caused by attack was studied under the AC power flow model [19]–[21]. Although the above works can successfully identify failures in some cases, *they all assume a connected post-attack grid and their recovery results cannot be verified in real time*. The works closest to ours are [5], [21], which established graph-theoretic conditions to guarantee the recovery accuracy. Our work differs from [5], [21] in the following aspects: (1) they still assume the grid to remain connected after attacks, while our solution is applicable to a partitioned post-attack grid where the power injections within the attacked area are unknown; (2) their conditions only characterize when all the line states can be identified correctly, while we provide recovery conditions at a finer granularity of individual lines. Such a finer granularity allows us to verify the states of a subset of lines (in Step 2 in Fig. 1) when the conditions in [5], [21] are not satisfied.

Besides blocking information as considered in this work, other types of cyber attacks are also possible, e.g., injecting false data into measurements. We refer readers to [15], [22], [23] and the references therein for details and leave failure localization under such attacks to future work.

The recovery from the joint cyber-physical attacks considered in this work is more challenging than traditional failure detection [24], since the jointly launched cyber attack will block the information from the attacked area and thus obfuscate the locations of the physical attack. Moreover, line failures due to naturally occurring faults typically do not occur at the same time and are mostly self-clearing, i.e., they rarely lead to line disconnection. On the contrary, a coordinated physical attack could take place simultaneously at multiple places, which may even lead to islanding.

B. Summary of Contributions

We aim at estimating the power grid state within an attacked area from which measurements have been blocked, with the following contributions:

1) Motivated by an observation that the existing method and condition for recovering the phase angles within the attacked area, previously developed for the case of connected post-attack grid, remain valid in the case of islanding, we focus on the recovery of the line states (i.e., breaker status of lines) within the attacked area using the phase angles, for which under the DC power flow model we develop an LP-based algorithm that allows for unknown changes in the power injections within the attacked area (due to islanding).
2) We establish conditions under which the accuracy of the proposed algorithm is guaranteed, which are further developed into verifiable conditions that can be tested using observable information.
3) Based on the above conditions, we develop a polynomial-time algorithm to verify the correctness of the estimated line states. We further provide an algorithm for verifying the states of potentially more lines based on the line states verified by the previous algorithm.
4) We extend the DC-based failed line detection and line state verification algorithms to their AC-based variants.
5) Our evaluations on real grid topologies show that the proposed recovery algorithm is highly accurate in localizing the failed lines with very few false alarms, and most of the failed lines can be successfully verified by the proposed verification algorithms.

Roadmap. Section II presents our models and problem formulation. Under the DC power flow model, Section III presents the proposed algorithm for localizing failed lines and its performance analysis. Section IV presents conditions and algorithms for verifying the correctness of the estimated line states. In Section V, we extend the DC-based line state detection and verification algorithms to the AC model. Section VI evaluates our solutions on real grid topologies, and Section VII concludes the paper. **All appendices can be found in the supplementary file** (proofs in Appendix D).

II. Problem Formulation

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/lines 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. For a set A, $\mathbb{I}_C = 1$ if condition C holds and $\mathbb{I}_C = 0$ otherwise. We use $\boldsymbol{\Lambda}_{(\cdot)} \in \{0,1\}^{m\times n}$ with one nonzero element in each row to select entries from a vector such that $\boldsymbol{\Lambda}_{(\cdot)}\boldsymbol{x}$ is a subvector of $\boldsymbol{x}$. For a vector $\boldsymbol{x}$, $[\boldsymbol{x}]$ denotes a diagonal matrix with $\boldsymbol{x}$ on the main diagonal. For a complex-valued number x, we use $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ to denote its real and imaginary part, respectively.

A. Power Grid Model

We model the power grid a
$G = (V, E)$, where V is th
the set of lines (transmission
associated with a *reactance*
{operational, failed} (assumed
Each node v is associated with
power injection p_v. The phas
active power injections $\boldsymbol{p} := ($

$$\boldsymbol{B\theta}$$

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

$$B_{uv} = \begin{cases} 0 \\ -1/r_{uv} \\ -\sum_{w\in V\setminus\{u\}} b_{u} \end{cases}$$

By arbitrarily assigning an
topology of G can also be repr
$\boldsymbol{D} \in \{-1, 0, 1\}^{|V|\times|E|}$, whose

$$D_{ij} = \begin{cases} 1 & \text{if line } e \\ -1 & \text{if line } e \\ 0 & \text{otherwis} \end{cases}$$

It is worth noting that the pr
are not restricted to any speci

As illustrated in Fig. 2, we a
as a composition of multiple
deployment of remote termina
surement units (PMUs), whic
measurements. The measuren
Supervisory Control and Data
Area Monitoring Protection and
power grid management. As envisioned by [25], we consider a heterogeneous smart grid with several operators where multiple communication networks and protocols coexist. More specifically, the measurements and control instructions can be communicated through traditional fiber optic cables, power lines [26] or wireless links [27]. Due to the wide range of communication media, heterogeneous protocols (such as DNP3 [28], IEEE 1901 FFT-OFDM [26], etc.) can coexist.

B. Attack Model

As illustrated in Fig. 2, we consider an adversary who launches joint cyber-physical attacks during the interval between two consecutive state estimations for a specific area $H = (V_H \subseteq V, E_H \subseteq E)$. We assume that the attacks have successfully bypassed both prevention and detection measures.

The physical part will disconnect a set F ($|F| > 0$) of lines within H by either manipulating the breaker status or cutting the power lines.

The cyber part will take the form of Denial-of-Service (DoS) attacks to block the measurements within H. In other words, although the control center can observe the information in $\bar{H} = (V_{\bar{H}} \subseteq V, E_{\bar{H}} \subseteq E)$, information within the attacked area H, especially the post-attack topology and power injections, is blocked. Such DoS attacks can be achieved by

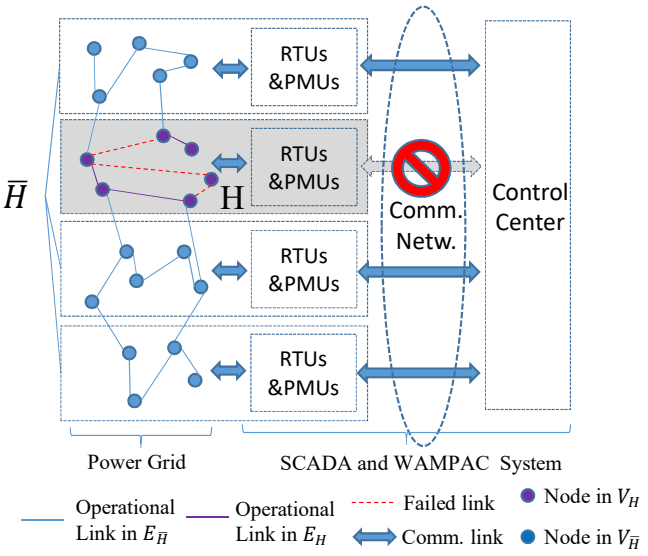


Figure 2. A cyber-physical attack that blocks information from the attacked area H while disconnecting certain lines within H.

destroying communication media (wireless or wired link), congesting the communication network (e.g., through telephonic floods), or remotely wiping the data in servers [2].

One of the motivations behind such joint attacks is to cause service disruption for consumers, especially critical infrastructures. Specifically, a large-scale line disconnection due to physical attacks can cause islanding and the associated load shedding, which can cause disruption of service to affected customers, and in the worst case, the collapse of the whole island in absence of generation. It is worth noting that the role of cyber attack considered in this work is not to hide the existence of physical attacks as considered in false data injection [15], [22], but to hinder the recovery process [2]. More specifically, repairing/restoration usually requires the knowledge of topology [29], which is normally available at the control center through binary breaker status measurements. However, the DoS attack will block such information and thus make the repairing/restoration scheduling challenging.

Formally, we denote x' as the post-attack counterpart of the pre-attack value x. Before attack, the control center has full access to measurements on power injections, line flows, and the information of breaker status (the topology $G = (E, V)$). The physical attack will change G to $G' = (V, E')$, where $E'_{\bar{H}} = E_{\bar{H}}$ while $E'_H \neq E_H$, as illustrated in Fig. 2. The cyber attack will block the information within $H = (V_H, E_H)$. To schedule the repairing/restoration to recover from the attack, we need to recover the topology information E'_H or equivalently detecting the failed lines F.

This work focuses on "Topology Recovery" (detecting $\hat{F}$) and "Topology verification" during the recovery from a general class of attacks that result in power line disconnection and information loss.

C. Voltage Recovery Problem

Our goal is to recover the post-attack state within H, based on the grid state before the attack (e.g., $\boldsymbol{B}$ and $\boldsymbol{\theta}$) and the

Table I
NOTATIONS

Notation	Description
$G=(V,E)$	power grid
H, $\bar{H}$	attacked/unattacked area
F , E_o	set of failed/operational lines after attack
$\boldsymbol{B}$, $\boldsymbol{D}$	admittance/incidence matrix
$\boldsymbol{\theta}$, $\boldsymbol{\theta}'$	phase angles before/after attack
$\boldsymbol{p}$, $\boldsymbol{p}'$	active power injections before/after attack
$\boldsymbol{\Gamma}$	$[\frac{1}{r_e}]_{e\in E}$ (r_e: reactance of line e)
$\boldsymbol{\Delta}$	change in active power injections
$\tilde{\boldsymbol{D}}$	hypothetical post-attack power flows (5)
η	rounding threshold in FLD
$S_U, f_{U,g}, f_{U,1(0)}$	definitions related to hyper-node (16)

information from the unattacked area $\bar{H}$ after the attack (e.g., $\boldsymbol{\theta}'_{\bar{H}}$). 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 due to such supply-demand balance, where $\Delta_v>0$ if v is a generator bus and $\Delta_v\le 0$ otherwise.

We observe in [30, Theorem A.1] that the existing condition in [5, Theorem 1] for recovering the phase angles $\boldsymbol{\theta}'_H$ in a connected post-attack grid remains valid under an attack that possibly partitions the grid into islands, which allows $\boldsymbol{\theta}'_H$ to be recovered through solving a linear system. Thus, we assume $\boldsymbol{\theta}'_H$ to be known in the sequel to focus on the recovery of the line states within H, i.e., the localization of F.

Remark: Alternatively, $\boldsymbol{\theta}'_H$ can be obtained from secured PMUs as assumed in [31]–[33]. We refer to Appendix A for details, which which also provides guidelines on placing secured PMUs for recovering $\boldsymbol{\theta}'_H$.

III. Localizing Failed lines with Unknown Active Power Injections

Although providing theoretical guarantees, the existing solutions for localizing failed lines in [5], [21] assume either a connected post-attack grid or a known $\boldsymbol{\Delta}_H$, which cannot be guaranteed in practice. To address this limitation, we will first present an algorithm extended from [5] that can jointly estimate the failed lines F and the changes in power injections $\boldsymbol{\Delta}_H$ (Section III-A), and then analyze the algorithm's accuracy in estimating F under unknown $\boldsymbol{\Delta}_H$ (Section III-B).

A. Algorithm

Our algorithm extends the failure localization algorithm in [5] (which assumes $\boldsymbol{\Delta}_H=\boldsymbol{0}$) by formulating the $(F,\boldsymbol{\Delta}_H)$ joint estimation problem as the following optimization.

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 $\boldsymbol{B}=\boldsymbol{D}\boldsymbol{\Gamma}\boldsymbol{D}^T$ (see Table I for the definitions), we can write the post-attack admittance matrix as $\boldsymbol{B}'=\boldsymbol{B}-\boldsymbol{D}\boldsymbol{\Gamma}[\boldsymbol{x}]\boldsymbol{D}^T$, which together with $\boldsymbol{\Delta}=\boldsymbol{p}-\boldsymbol{p}'$ implies

$$\boldsymbol{\Delta}_H=\boldsymbol{B}_{H|G}(\boldsymbol{\theta}-\boldsymbol{\theta}')+\boldsymbol{D}_H\boldsymbol{\Gamma}_H[\boldsymbol{D}_{G|H}^T\boldsymbol{\theta}']\boldsymbol{x}_H, \tag{4}$$

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 lines in H[1]. For simplicity, we define

$$\tilde{\boldsymbol{D}}:=\boldsymbol{D}\boldsymbol{\Gamma}[\boldsymbol{D}^T\boldsymbol{\theta}']. \tag{5}$$

For line $e_k=(i,j)$, $\tilde{D}_{i,k}=-\tilde{D}_{j,k}=\frac{\theta'_i-\theta'_j}{r_{ij}}$, which indicates the post-attack power flow on line e_k if it is operational.

In addition, $\boldsymbol{\Delta}_H$ is subject to the following constraints:

$$p_v\ge\Delta_v\ge 0,\ \ \forall v\in\{u\,|u\in V_H,p_u>0\}, \tag{6a}$$
$$p_v\le\Delta_v\le 0,\ \ \forall v\in\{u\,|u\in V_H,p_u\le 0\}, \tag{6b}$$
$$\boldsymbol{1}^T\boldsymbol{\Delta}=0, \tag{6c}$$

which ensure that (i) the bus type will remain the same after attack, (ii) the load will not increase after islanding, and (iii) the total power is balanced. It is worth noting that (6c) is ensured by (4), 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 (5). This implies that any $\boldsymbol{\Delta}_H$ satisfying (4) 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 power injection changes in H), and thus satisfy (6c). Hence, we will omit (6c) in the sequel.

Objective: The problem of failure localization aims at finding a failed line set $\hat{F}$ that is as close as possible to the ground-truth set F, while satisfying all the constraints. The solution is generally not unique, e.g., if both endpoints of a line $l\in E_H$ are disconnected from $\bar{H}$ after the attack, then the states 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 lines to satisfy all the constraints. This idea has been applied to failure localization in power grid in various forms [3]–[5]. Mathematically, the problem is formulated as

$$\text{(P0)}\quad \min_{\boldsymbol{x}_H,\boldsymbol{\Delta}_H}\ \boldsymbol{1}^T\boldsymbol{x}_H \tag{7a}$$
$$\text{s.t.}\quad (4),(6a)-(6b), \tag{7b}$$
$$x_e\in\{0,1\},\quad \forall e\in E_H, \tag{7c}$$

where the decision variables are $\boldsymbol{x}_H$ and $\boldsymbol{\Delta}_H$. Although binary linear programming is generally hard, (P0) is only a special case and hence needs to be analyzed separately.

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

By relaxing the integer constraint (7c), (P0) is relaxed into

$$\text{(P1)}\quad \min_{\boldsymbol{x}_H,\boldsymbol{\Delta}_H}\ \boldsymbol{1}^T\boldsymbol{x}_H \tag{8a}$$
$$\text{s.t.}\quad (4),(6a)-(6b), \tag{8b}$$
$$\boldsymbol{0}\le\boldsymbol{x}_H\le\boldsymbol{1}. \tag{8c}$$

where $\boldsymbol{0}\le\boldsymbol{x}_H\le\boldsymbol{1}$ denotes element-wise inequality. To take into account the error on recovered $\boldsymbol{\theta}'$, we can add a noise term in (6a)-(6b). For example, by denoting ϵ as a tunable noise term, (6a) becomes $p_v+\epsilon\ge\Delta_v\ge-\epsilon$. We assume $\epsilon=0$ in the rest of the paper to focus on the recovery from

[1]Because we focus on the post-attack recovery stage as shown in Fig. 1, we implicitly assume in (4) that the grid has reached the post-attack steady state. This means that each island is assumed to have reached a steady state in the case of islanding.

the information loss caused by the attack. 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 lines, called ***F**ailed **L**ine **D**etection (FLD)* as given in Alg. 1, where the input parameter $\eta \in (0,1)$ is a threshold for rounding the factional solution of $\boldsymbol{x}_H$ to an integral solution ($\eta = 0.5$ in our experiments). We will illustrate how to set η at the end of Section IV.

Algorithm 1: **F**ailed **L**ine **D**etection (FLD)

Input: $\boldsymbol{B}, \boldsymbol{p}, \boldsymbol{\Delta}_{\bar{H}}, \boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{D}, \eta$

Output: $\hat{F}$

Solve the problem (P1) to obtain $\boldsymbol{x}_H$;

Return $\hat{F} = \{e : x_e \geq \eta\}$.

B. Analysis

We now analyze when FLD can correctly localize the failed lines, the results of which will lay the groundwork for the verifiable conditions in the next section. In the sequel, $\boldsymbol{\Delta}_H^*$ denotes the ground-truth power injection changes in H and $\boldsymbol{x}_H^*$ denotes the ground-truth failure indicators.

According to (6), we decompose V_H into $V_{H,L}$ for nodes with $p_v \leq 0$ and $V_{H,S}$ for the rest. Define $E_o \subseteq E_H$ as the set of lines that operate normally after failure, and $F \subseteq E_H$ as the failed lines. We make the following assumption:

Assumption 1. *As in [5], we assume that for each line $(s,t) \in E_H$, $\theta'_s \neq \theta'_t$, as otherwise the line will carry no power flow and hence its states cannot be identified*[2].

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 $\boldsymbol{y}_H = [\boldsymbol{\Delta}_H^T, \boldsymbol{x}_H^T]^T \in \mathbb{R}^{(|E_H|+|V_H|)}$ (where $[A,B]$ denotes horizontal concatenation), and explicitly represent the solution to $\boldsymbol{y}_H$ that satisfies (4). Notice that (4) can be written as $[\boldsymbol{I}_{|V_H|}, -\tilde{\boldsymbol{D}}_H]\boldsymbol{y}_H = \boldsymbol{B}_{H|G}(\boldsymbol{\theta} - \boldsymbol{\theta}')$ ($\boldsymbol{I}_{|V_H|}$: 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 (4). Next, consider the null space of $[\boldsymbol{I}_{|V_H|}, -\tilde{\boldsymbol{D}}_H]$, whose dimension is $|E_H|$. It is easy to verify that $[\tilde{\boldsymbol{d}}_e^T, \boldsymbol{u}_e^T]^T$ ($e \in E_H$) are $|E_H|$ independent vectors spanning the null space of $[\boldsymbol{I}_{|V_H|}, -\tilde{\boldsymbol{D}}_H]$, where $\tilde{\boldsymbol{d}}_e$ is the column vector of $\tilde{\boldsymbol{D}}_H$ corresponding to line 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 pair of $(\boldsymbol{\Delta}_H, \boldsymbol{x}_H)$ satisfying (4) can be expressed as

$$\boldsymbol{y}_H = \begin{bmatrix} \boldsymbol{\Delta}_H \\ \boldsymbol{x}_H \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Delta}_H^* \\ \boldsymbol{x}_H^* \end{bmatrix} + \sum_{e \in E_H} c_e \begin{bmatrix} \tilde{\boldsymbol{d}}_e \\ \boldsymbol{u}_e \end{bmatrix}, \tag{9}$$

where c_e's are the coefficients. Based on the decomposition of V_H into $V_{H,L}$ and $V_{H,S}$, $\tilde{\boldsymbol{D}}_H$ and $\boldsymbol{\Delta}_H$ can be written as

$$\tilde{\boldsymbol{D}}_H = \begin{array}{c} \\ V_{H,L} \\ V_{H,S} \end{array} \begin{array}{c} \begin{array}{cc} E_o & F \end{array} \\ \begin{bmatrix} \tilde{\boldsymbol{D}}_{H,L,o} & \tilde{\boldsymbol{D}}_{H,L,F} \\ \tilde{\boldsymbol{D}}_{H,S,o} & \tilde{\boldsymbol{D}}_{H,S,F} \end{bmatrix} \end{array}, \tag{10a}$$

$$\boldsymbol{\Delta}_H = \begin{array}{c} \\ V_{H,L} \\ V_{H,S} \end{array} \begin{bmatrix} \boldsymbol{\Delta}_{H,L} \\ \boldsymbol{\Delta}_{H,S} \end{bmatrix}. \tag{10b}$$

Let $\tilde{\boldsymbol{D}}_{H,L} := [\tilde{\boldsymbol{D}}_{H,L,o}, \tilde{\boldsymbol{D}}_{H,L,F}]$, $\tilde{\boldsymbol{D}}_{H,S} := [\tilde{\boldsymbol{D}}_{H,S,o}, \tilde{\boldsymbol{D}}_{H,S,F}]$, and $\boldsymbol{c} := (c_e)_{e \in E_H} \in \mathbb{R}^{|E_H|}$. Since $\boldsymbol{\Delta}_{H,L}$ and $\boldsymbol{\Delta}_{H,S}$ are constrained differently in (6a) and (6b), we introduce $\boldsymbol{\Lambda}_L = [\boldsymbol{I}_{|V_{H,L}|}, \boldsymbol{0}]$ and $\boldsymbol{\Lambda}_S = [\boldsymbol{0}, \boldsymbol{I}_{|V_{H,S}|}]$ such that $\boldsymbol{\Delta}_{H,L} = \boldsymbol{\Lambda}_L \boldsymbol{\Delta}_H$, $\boldsymbol{\Delta}_{H,S} = \boldsymbol{\Lambda}_S \boldsymbol{\Delta}_H$, $\tilde{\boldsymbol{D}}_{H,L} = \boldsymbol{\Lambda}_L \tilde{\boldsymbol{D}}_H$ and $\tilde{\boldsymbol{D}}_{H,S} = \boldsymbol{\Lambda}_S \tilde{\boldsymbol{D}}_H$. According to (9), for FLD to correctly localize the failed lines, it suffices to have $x_e^* + c_e \geq \eta$ for all $e \in F$ and $x_e^* + c_e < \eta$ for all $e \in E_o$. Equivalently, it suffices to ensure that the optimal solution $\hat{\boldsymbol{c}}$ to the following optimization problem satisfies $\hat{c}_e \geq \eta - 1$ for all $e \in F$ and $\hat{c}_e < \eta$ for all $e \in E_o$:

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

$$\text{s.t.} \quad \tilde{\boldsymbol{D}}_{H,L} \boldsymbol{c} \leq -\boldsymbol{\Delta}_{H,L}^*, \tag{11b}$$

$$-\tilde{\boldsymbol{D}}_{H,L} \boldsymbol{c} \leq -(\boldsymbol{\Lambda}_L \boldsymbol{p}_H - \boldsymbol{\Delta}_{H,L}^*), \tag{11c}$$

$$-\tilde{\boldsymbol{D}}_{H,S} \boldsymbol{c} \leq \boldsymbol{\Delta}_{H,S}^*, \tag{11d}$$

$$\tilde{\boldsymbol{D}}_{H,S} \boldsymbol{c} \leq \boldsymbol{\Lambda}_S \boldsymbol{p}_H - \boldsymbol{\Delta}_{H,S}^*, \tag{11e}$$

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

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

where (11a) is equivalent to (8a), (11b)-(11c) correspond to (6b), (11d)-(11e) correspond to (6a), (11f)-(11g) correspond to (8c), and the change of variables $\boldsymbol{x}_H, \boldsymbol{\Delta}_H$ into $\boldsymbol{c}$ based on (9) ensures the satisfaction of (4). This equivalent formulation of (P1) will help to simplify our analysis by eliminating the equality constraint (4). *For notational simplicity, we will omit the subscript $_H$ in the sequel* unless it causes confusion.

Next, we use (11) to analyze the accuracy of FLD. Let $\hat{F}$ be the failed line set returned by FLD. We first define $Q_m = F \backslash \hat{F}$ as the set of missed failed lines, and $Q_f = \hat{F} \setminus F$ as the set of operational lines that are falsely detected as failed. Note that according to (11), a failed line $e \in F$ is missed if and only if $\hat{c}_e < \eta - 1$. Similarly, an operational line $e \in E_o$ is falsely detected as failed if and only if $\hat{c}_e \geq \eta$. To express this in a vector form, we define $\boldsymbol{W}_m \in \{0,1\}^{|Q_m| \times |E_H|}$ as a binary matrix, where for each $i = 1, \ldots, |Q_m|$, $(W_m)_{i,e} = 1$ if the i-th missed line is line e and thus we have $\boldsymbol{W}_m \hat{\boldsymbol{c}} \leq (\eta - 1)\boldsymbol{1}$. Similarly, $\boldsymbol{W}_f \in \{0,1\}^{|Q_f| \times |E_H|}$ is defined such that $(W_f)_{i,e} = 1$ if the i-th false-alarmed line is line e, which leads to $-\boldsymbol{W}_f \hat{\boldsymbol{c}} \leq -\eta \boldsymbol{1}$. For ease of presentation, we define

$$\boldsymbol{A}_D^T := [\tilde{\boldsymbol{D}}_L^T, -\tilde{\boldsymbol{D}}_L^T, -\tilde{\boldsymbol{D}}_S^T, \tilde{\boldsymbol{D}}_S^T] \in \mathbb{R}^{|E_H| \times 2|V_H|}, \tag{12a}$$

$$\boldsymbol{A}_x^T := [-\boldsymbol{I}_{|E_H|}, \boldsymbol{I}_{|E_H|}] \in \mathbb{R}^{|E_H| \times 2|E_H|}, \tag{12b}$$

$$\boldsymbol{W}^T := [\boldsymbol{W}_m^T, -\boldsymbol{W}_f^T] \in \mathbb{R}^{|E_H| \times (|Q_m| + |Q_f|)}, \tag{12c}$$

$$\boldsymbol{g}_D^T := [-(\boldsymbol{\Delta}_L^*)^T, (-\boldsymbol{p}'_L)^T, (\boldsymbol{\Delta}_S^*)^T, (\boldsymbol{p}'_S)^T], \tag{12d}$$

$$\boldsymbol{g}_x^T := [(\boldsymbol{x}^*)^T, \boldsymbol{1}^T - (\boldsymbol{x}^*)^T] \in \mathbb{R}^{1 \times 2|E_H|}, \tag{12e}$$

[2]This assumption essentially means that we will ignore the existence of such lines in failure localization. Specifically, in the case of islanding, if an island has a total blackout (because of not containing load/generation or frequency collapse), then lines in this island will be excluded from failure localization.

$$\boldsymbol{g}_w^T := [(\eta-1)\mathbf{1}^T, -\eta\mathbf{1}^T] \in \mathbb{R}^{1\times(|Q_m|+|Q_f|)}, \tag{12f}$$

where $\boldsymbol{p}'_L = \boldsymbol{p}_L - \boldsymbol{\Delta}^*_L$ and $\boldsymbol{p}'_S = \boldsymbol{p}_S - \boldsymbol{\Delta}^*_S$ denote the post-attack active power injections at $V_{H,L}$ and $V_{H,S}$. Then the constraints in (11) can be written as $[\boldsymbol{A}_D^T, \boldsymbol{A}_x^T]^T\boldsymbol{c} \le [\boldsymbol{g}_D^T, \boldsymbol{g}_x^T]^T$, and the optimal solution must satisfy $\boldsymbol{W}\boldsymbol{c} \le \boldsymbol{g}_w$. The following observation is the foundation of our analysis.

Lemma III.2. *A line $e \in F$ cannot be missed by FLD if for $Q_m = \{e\}$ and $Q_f = \emptyset$, there is a solution $\boldsymbol{z} \ge \mathbf{0}$ to*

$$[\boldsymbol{A}_D^T, \boldsymbol{A}_x^T, \boldsymbol{W}^T, \mathbf{1}]\boldsymbol{z} = \mathbf{0}, \tag{13a}$$

$$[\boldsymbol{g}_D^T, \boldsymbol{g}_x^T, \boldsymbol{g}_w^T, 0]\boldsymbol{z} < 0. \tag{13b}$$

Similarly, a line $e' \in E_o$ cannot be falsely detected as failed by FLD if there exists a solution $\boldsymbol{z} \ge \mathbf{0}$ to (13) *where $\boldsymbol{W}$ is constructed according to $Q_f = \{e'\}$ and $Q_m = \emptyset$.*

The proof is by contradiction: if $e \in F \setminus \hat{F}$, then for $\boldsymbol{W}$ corresponding to $Q_m = \{e\}$ and $Q_f = \emptyset$, there must be no $\boldsymbol{z} \ge \mathbf{0}$ satisfying (13); similar argument holds for $e' \in E_o$ by assuming $e' \in \hat{F} \setminus F$. See detailed proof in Appendix D.

For ease of presentation, we will introduce a few notations as follows. Denote $\tilde{\boldsymbol{D}}_u$ as the row in $\tilde{\boldsymbol{D}}$ corresponding to node u, and $\tilde{D}_{u,e}$ as the entry in $\tilde{\boldsymbol{D}}_u$ corresponding to line e. Recall that as defined in (5), if $e = (u, v)$, then $\tilde{D}_{u,e} = (\theta'_u - \theta'_v)r_{uv}^{-1}$. We decompose the l.h.s of (13a) into $\boldsymbol{A}_D^T\boldsymbol{z}_D + \boldsymbol{A}_x^T[\boldsymbol{z}_{x-}, \boldsymbol{z}_{x+}] + \boldsymbol{W}_m^T\boldsymbol{z}_{w,m} + \boldsymbol{W}_f^T\boldsymbol{z}_{w,f} + z_*\mathbf{1}$ such that its row corresponding to line e can be written as

$$\sum_{u\in V_H}\left(\tilde{D}_{u,e}z_{D,u} - \tilde{D}_{u,e}z_{D,-u}\right) + \left(z_{x+,e} - z_{x-,e}\right) + \mathbb{I}_{Q_m}(e)z_{w,m,e} - \mathbb{I}_{Q_f}(e)z_{w,f,e} + z_*. \tag{14}$$

Similarly, the l.h.s of (13b) can be expanded into

$$\sum_{u\in V_H}\left(g_{D,u}z_{D,u} + g_{D,-u}z_{D,-u}\right) + \sum_{e\in E_H}[z_{x+,e}(1-x_e^*) + z_{x-,e}x_e^*] + \boldsymbol{g}_w^T\boldsymbol{z}_w + z_*, \tag{15}$$

where $\boldsymbol{g}_w^T\boldsymbol{z}_w = \sum_{e\in E_H}[\mathbb{I}_{Q_m}(e)z_{w,m,e}(\eta-1) - \mathbb{I}_{Q_f}(e)z_{w,f,e}\eta]$, $g_{D,u} := -\Delta_u^*$ and $g_{D,-u} := -p'_u$ if $p_u \le 0$, whereas $g_{D,u} := p'_u$ and $g_{D,-u} := \Delta_u^*$ if $p_u > 0$. Then, a solution $\boldsymbol{z} \ge \mathbf{0}$ satisfies (13) if $\forall e \in E_H$, we have (14) equal to 0 and (15) less than 0.

Equipped with Lemma III.2, we are ready to explicitly characterize what types of lines are guaranteed to be correctly identified. Specifically, we will show that a line will satisfy the conditions in Lemma III.2 if its endpoints satisfy certain conditions. To make our conditions as general as possible, we introduce a generalization of node called *hyper-node* as follows (a single node is also a hyper-node):

Definition III.1. *A set of nodes $U \subseteq V_H$ is a hyper-node if they induce a connected subgraph before attack.*

We define a few properties of a hyper-node U. Define E_U as the set of lines with exactly one endpoint in U, i.e, $E_U := \{e | e = (s,t) \in E_H, s \in U, t \notin U\}$. If $E_U \cap F \ne \emptyset$, we define

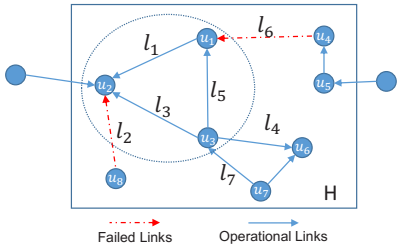


Figure 3. An example of hyper-node (arrow denotes the direction of a power flow or a hypothetical power flow).

$$\tilde{D}_{U,e} := \sum_{u\in U}\tilde{D}_{u,e}, \tag{16a}$$

$$S_U := \{e \in E_U \setminus F |\ \exists l \in E_U \cap F, \tilde{D}_{U,l}\tilde{D}_{U,e} > 0\}, \tag{16b}$$

$$f_{U,0} := \max_{e\in S_U}|\tilde{D}_{U,e}|, \text{ where } f_{U,0} := 0 \text{ if } S_U = \emptyset, \tag{16c}$$

$$f_{U,1} := \min_{e\in E_U\cap F}|\tilde{D}_{U,e}|, \tag{16d}$$

$$f_{U,g} := \begin{cases} \sum_{u\in U} g_{D,u} & \text{if } \exists l \in E_U \cap F, \tilde{D}_{U,l} < 0, \\ \sum_{u\in U} g_{D,-u} & \text{otherwise.} \end{cases} \tag{16e}$$

Example 1. *Consider an attacked area H as shown in Fig. 3, where blue circles denote nodes (buses) while the direction of each line indicates the direction of power flow*[3]*. Suppose that $F = \{l_2, l_6\}$ and all nodes are load buses. Nodes u_1, u_2 and u_3 form a hyper-node U, where $E_U = \{l_2, l_4, l_6, l_7\}$, $S_U = \{l_7\}$, $f_{U,0} = |\tilde{D}_{U,l_7}|$, $f_{U,1} = \min\{|\tilde{D}_{U,l_2}|, |\tilde{D}_{U,l_6}|\}$ and $f_{U,g} = -\sum_{v\in U}\Delta_v^*$. $\tilde{D}_{U,l_1} = \tilde{D}_{u_1,l_1} + \tilde{D}_{u_2,l_1} = 0$ since $l_1 \notin E_U$, while $\tilde{D}_{U,l_2} = \tilde{D}_{u_2,l_2} \ne 0$ since $l_2 \in E_U$.*

Based on these definitions and Lemma III.2, we are ready to present a condition under which a failed line $l \in F$ will not be missed by FLD (see proof in Appendix D).

Theorem III.1. *A failed line $l \in F$ will be detected by FLD, i.e., $l \in \hat{F}$, if there exists at least one hyper-node (say U) such that $l \in E_U$, for which the following conditions hold:*

1) $\forall e, l \in E_U \cap F$, $\tilde{D}_{U,e}\tilde{D}_{U,l} > 0$,
2) $S_U = \emptyset$, *and*
3) $f_{U,g} + (\eta-1)|\tilde{D}_{U,l}| < 0$.

Based on similar arguments, the following condition can guarantee that an operational line $l \in E_o$ will not be falsely detected by FLD ($l \notin \hat{F}$) (see proof in Appendix D). For notational simplicity, we extend the definition of $f_{U,g}$ to a hyper-node U with $E_U \cap F = \emptyset$:

$$f_{U,g} := \begin{cases} \sum_{u\in U} g_{D,u} & \text{if } \exists l \in E_U \setminus F, \tilde{D}_{U,l} > 0, \\ \sum_{u\in U} g_{D,-u} & \text{otherwise.} \end{cases} \tag{17}$$

Theorem III.2. *An operational line $l \in E_o$ will not be detected (as failed) by FLD, i.e., $l \notin \hat{F}$, if there exists at least one hyper-node (say U) such that $l \in E_U$, for which the following conditions hold:*

1) $\forall l, l' \in E_U \cap E_o$: $\tilde{D}_{U,l}\tilde{D}_{U,l'} > 0$,
2) $S_U = \emptyset$ *if* $E_U \cap F \ne \emptyset$, *and*
3) $f_{U,g} - \eta|\tilde{D}_{U,l}| < 0$.

[3]These may be hypothetical power flows, as a failed line carries no flow.

Remark: Theorems III.1 and III.2 provide sufficient conditions for FLD to correctly identify the states of a line l based on the direction and magnitude of "power flows" around a hyper-node U at the "endpoint" of l (i.e., $l \in E_U$): The (hypothetical) power flows on all the lines of the same state (failed or operational) around U should be in the same direction, i.e., all going into or out of U (condition 1); all lines of different states around U should have opposite (hypothetical) power flow directions (condition 2); the magnitude of the (hypothetical) power flow on the line of interest (i.e., l) should be sufficiently large (condition 3).

IV. Verifying Estimated Line States

Although the conditions in Section III-B can guarantee the accuracy of FLD, they are not testable in practice since the ground-truth (F and $\mathbf{\Delta}^*$) is required. In this section, we will develop conditions requiring only observable information such that they can be verified during operation.

To this end, we first notice that Δ_u for some $u \in V_H$ can be recovered directly. For example, if the adjustment of power injections after islanding follows *proportional load shedding/generation reduction* [34], [35][4], then we have the following observations (see proof in Appendix D):

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 lines in $E\backslash E_H$. Then under the assumption of proportional load shedding/generation reduction, 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 $\Delta_u = 0$.*

Define $U_B \subseteq V_H$ as the node set such that $\forall u \in U_B$, Δ_u has been recovered (e.g., through Lemma IV.1). Our key observation is that for any hyper-node U, $\tilde{D}_{U,l}$ for any $l \in E_U$ can be computed with the knowledge of $\boldsymbol{\theta}'$, and $f_{U,g}$ can be upper-bounded by

$$\hat{f}_{U,g} := \sum_{u\in U\cap U_B} f_{u,g} + \sum_{u\in U\setminus U_B} |p_u|, \tag{18}$$

where $f_{u,g}$ is defined in (16e) for $U = \{u\}$. Since $f_{u,g}$ is known for nodes in U_B and p_u (power injection at u before attack) is also known, $\hat{f}_{U,g}$ is computable. We now show how to use this information to verify the results of FLD based on Lemma III.2 and Theorems III.1–III.2. We note that our solution remains valid even if $U_B = \emptyset$, which only affects the tightness of the upper bound $\hat{f}_{U,g}$.

A. Verification without Knowledge of Ground Truth

We first tackle the lines whose states can be verified without any knowledge of the ground truth line states.

1) Verifiable Conditions: The basic idea is to rule out the other possibility by constructing *counterexamples* to the theorems in Section III-B if the estimated line state is incorrect.

lines in 1-edge cuts: If line $e = (u_1, u_2)$ forms a *cut* of H, i.e., $(V_H, E_H \setminus \{e\})$ contains more connected components than H, then by breadth-first search (BFS) starting from u_1 and u_2 respectively without traversing e, we can construct two hyper-nodes U_1 and U_2 such that $E_{U_1} = E_{U_2} = \{e\}$ and thus $S_{U_1} = S_{U_1} = \emptyset$. For example, in Fig. 3, line $e := l_6$ is a 1-edge cut, and thus $U_1 := \{u_4, u_5\}$ and $U_2 := V_H \setminus U_1$ satisfy this condition. Then the following verifiable conditions are directly implied by Theorems III.1–III.2:

Corollary IV.1. *If $e \in \hat{F}$ and $\min\{\hat{f}_{U_1,g}, \hat{f}_{U_2,g}\} - \eta|\tilde{D}_{U_1,e}| < 0$, then we can verify $e \in F$. If $e \in E_H \setminus \hat{F}$ and $\min\{\hat{f}_{U_1,g}, \hat{f}_{U_2,g}\} + (\eta-1)|\tilde{D}_{U_1,e}| < 0$, then we can verify $e \in E_H \setminus F$.*

Proof. If $e \in \hat{F}$ and $\min\{\hat{f}_{U_1,g}, \hat{f}_{U_2,g}\} - \eta|\tilde{D}_{U_1,e}| < 0$, then e must have failed, since otherwise e would have been estimated as operational according to Theorem III.2. Similarly, if $e \in E_H \setminus \hat{F}$ and $\min\{\hat{f}_{U_1,g}, \hat{f}_{U_2,g}\} + (\eta-1)|\tilde{D}_{U_1,e}| < 0$, then e must be operational, since otherwise e would have been estimated as failed according to Theorem III.1. Note that as our verification is based on contradiction, $\hat{f}_{U_i,g}$ should be computed as if $e \in E_H \setminus F$ to verify $e \in \hat{F}$ and vice-versa. □

lines in 2-edge cuts: If lines $e_1, e_2 \in E_H$ together form a cut of H but each individual line does not, then by BFS starting from the endpoints of e_1 (or e_2) without traversing e_1 or e_2, we can construct two hyper-nodes U_1, U_2 such that $E_{U_1} = E_{U_2} = \{e_1, e_2\}$. For example, as $e_1 := l_4$ and $e_2 := l_7$ form a 2-edge cut of H in Fig. 3, $U_1 := \{u_6, u_7\}$ and $U_2 := V_H \setminus U_1$ satisfy this condition. Moreover, any pair of lines in a cycle C form a 2-edge cut if they are not in any other cycle in H, e.g., any pair of lines in the cycle $\{l_1, l_3, l_5\}$ satisfy this condition. Based on this observation, we provide the following conditions for verifying the states of such lines

Theorem IV.1. *Consider a hyper-node U with $E_U = \{e_1, e_2\}$ and $e_1, e_2 \in E_H \setminus \hat{F}$. If $\tilde{D}_{U,e_1}\tilde{D}_{U,e_2} < 0$, then e_1, e_2 are guaranteed to both belong to $E_H \setminus F$ if*

1) $\hat{f}_{U,g} + (\eta-1)\min\{|\tilde{D}_{U,e_1}|, |\tilde{D}_{U,e_2}|\} < 0$, *and*
2) $\eta < 1 - \min\{\frac{\hat{f}_{U,g}+|\tilde{D}_{U,e_1}|}{|\tilde{D}_{U,e_2}|}, \frac{\hat{f}_{U,g}+|\tilde{D}_{U,e_2}|}{|\tilde{D}_{U,e_1}|}\}$.

If $\tilde{D}_{U,e_1}\tilde{D}_{U,e_2} > 0$, then we can verify:

1) $e_1 \in E_H \setminus F$ *if* $(1-\eta)|\tilde{D}_{U,e_1}| > \hat{f}_{U,g} + |\tilde{D}_{U,e_2}|$,
2) $e_2 \in E_H \setminus F$ *if* $(1-\eta)|\tilde{D}_{U,e_2}| > \hat{f}_{U,g} + |\tilde{D}_{U,e_1}|$.

Theorem IV.2. *Consider a hyper-node U with $E_U = \{e_1, e_2\}$ and $e_1 \in \hat{F}, e_2 \in E_H \setminus \hat{F}$. If $\tilde{D}_{U,e_1}\tilde{D}_{U,e_2} > 0$, then the states of e_1, e_2 are guaranteed to be correctly identified if*

1) $\hat{f}_{U,g} - \eta|\tilde{D}_{U,e_1}| < 0$, $\hat{f}_{U,g} + (\eta-1)|\tilde{D}_{U,e_2}| < 0$, *and*
2) *either* $\eta > \frac{\hat{f}_{U,g}+|\tilde{D}_{U,e_2}|}{|\tilde{D}_{U,e_1}|}$ *or* $\eta < 1 - \frac{\hat{f}_{U,g}+|\tilde{D}_{U,e_1}|}{|\tilde{D}_{U,e_2}|}$.

If $\tilde{D}_{U,e_1}\tilde{D}_{U,e_2} < 0$, then we can verify:

1) $e_1 \in F$ *if* $\eta|\tilde{D}_{U,e_1}| > \hat{f}_{U,g} + |\tilde{D}_{U,e_2}|$,
2) $e_2 \in E_H \setminus F$ *if* $(1-\eta)|\tilde{D}_{U,e_2}| > \hat{f}_{U,g} + |\tilde{D}_{U,e_1}|$.

Theorem IV.3. *Consider a hyper-node U with $E_U = \{e_1, e_2\}$ and $e_1, e_2 \in \hat{F}$. Then, we can verify:*

1) $e_1 \in F$ *if* $\eta|\tilde{D}_{U,e_1}| > \hat{f}_{U,g} + |\tilde{D}_{U,e_2}|$,
2) $e_2 \in F$ *if* $\eta|\tilde{D}_{U,e_2}| > \hat{f}_{U,g} + |\tilde{D}_{U,e_1}|$.

While in theory such verifiable conditions can also be derived for lines in larger cuts, the number of cases will grow

[4]Under this assumption, either the load or the generation (but not both) will be reduced upon the formation of an island. Moreover, if nodes u and v are in the same island and of the same type (both load or generator), then $p'_u/p_u = p'_v/p_v$. More details are given in Appendix G.

exponentially. We also find 1–2-edge cuts to cover th of lines in practice (see Fig. 11).

2) Verification Algorithm: Based on Theorems we develop an algorithm *Verification Of sTatEs (VOTE)* for verifying the line states estimated by FLD, whi applied to lines in 1–2-edge cuts. Here, E_a denotes all the lines in 1-edge cuts of H, while $\mathcal{E}_c$ denotes 2-edge cuts. In the algorithm, lines in E_a are tested b in $\mathcal{E}_c$ since it is easier to extend the knowledge of on the test results for E_a. As for the complexity, we ... that the time complexity of each iteration is $\mathcal{O}(|E_H| + |V_H|)$ due to BFS. Then, it takes $\mathcal{O}(|E_H|)$ iterations to verify E_a and $\mathcal{O}(|E_H|^2)$ iterations for $\mathcal{E}_c$, which results in a total complexity of $\mathcal{O}(|E_H|^2(|E_H| + |V_H|))$.

Algorithm 2: **V**erification **O**f s**T**at**E**s (VOTE)

Input: $\tilde{\boldsymbol{D}}, \boldsymbol{p}, \boldsymbol{\Delta}_{\bar{H}}, U_B, \eta, E_a, \mathcal{E}_c, \hat{F}$
Output: E_v

$E_v \leftarrow \emptyset$; /* verifiable lines */
foreach $e = (u_1, u_2) \in E_a$ **do**
  Construct hyper-nodes U_1 and U_2 such that $E_{U_1} = E_{U_2} = \{e\}$;
  if $e \in \hat{F}$ **then**
    Add e to E_v if $\min\{\hat{f}_{U_1,g}, \hat{f}_{U_2,g}\} - \eta|\tilde{D}_{U_1,e}| < 0$;
  else
    Add e to E_v if $\min\{\hat{f}_{U_1,g}, \hat{f}_{U_2,g}\} + (\eta - 1)|\tilde{D}_{U_1,e}| < 0$;
  if *e is verified to be in* $E_H \setminus F$ **then**
    Add u_i to U_B if Δ_{u_i} $(i = 1, 2)$ can be recovered through Lemma IV.1;
foreach $\{e_1, e_2\} \in \mathcal{E}_c$ **do**
  Construct hyper-nodes U_1 and U_2 such that $E_{U_1} = E_{U_2} = \{e_1, e_2\}$;
  Test the satisfaction of Lemma IV.1, IV.2, or IV.3 for U_1 and U_2, respectively;
  Add e_i $(i = 1, 2)$ to E_v if it is verified;

B. Verification with Partial Knowledge of Ground Truth

VOTE assumes no knowledge of the ground-truth line states, even if the states of some lines are already verified, e.g., line set E_v verified by VOTE. We observe that such partial knowledge of the ground truth can be exploited for better approximation of the unknown terms in (13) and thus verifying lines where VOTE fails. Following this idea, we propose a follow-up step designed to verify the states of additional lines in $E_H \setminus E_v$.

1) Verifiable Conditions: Recall that the idea for verifying the correctness of $e \in \hat{F}$ (or $e \in E_H \setminus \hat{F}$) is to construct a solution to (13) as if $e \in E_H \setminus F$ (or $e \in F$). Specifically, it can be shown that for a line $e \in \hat{F}$, if there exists $\boldsymbol{z} \geq \boldsymbol{0}$ for (13) where $\boldsymbol{W}$ is constructed for $Q_f = \{e\}$ and $Q_m = \emptyset$, then e is guaranteed to have failed since otherwise it must have been estimated to be operational. The challenge is the unknown $\boldsymbol{g}_D$, $\boldsymbol{g}_x$, and $\boldsymbol{g}_w$ due to unknown F and $\boldsymbol{\Delta}_H^*$. To tackle this challenge, we approximate these parameters by their worst possible values (in terms of satisfying (13)), which leads to the following result (see proof in Appendix D).

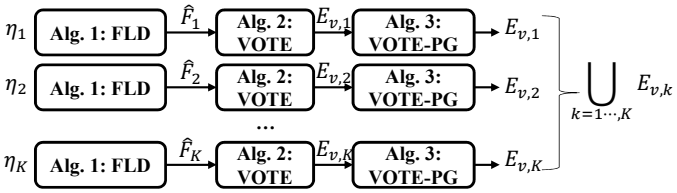


Figure 4. Guidelines for applying the proposed algorithms.

Theorem IV.4. *Given a set E_v of lines with known states, we define $\hat{\boldsymbol{g}}_D \in \mathbb{R}^{2|V_H|}$ and $\hat{\boldsymbol{g}}_x \in \mathbb{R}^{2|E_H|}$ as follows:*

$$\hat{g}_{D,u} = \begin{cases} g_{D,u} \text{ if } u \in U_B \\ |p_u| \text{ if } u \notin U_B \end{cases} \quad \hat{g}_{x,e} = \begin{cases} g_{x,e} \text{ if } e \in E_v \\ 1 \text{ if } e \notin E_v \end{cases} \tag{19}$$

and define $\hat{g}_{D,-u}$ and $\hat{g}_{x,-e}$ similarly. Then, a line $l \in \hat{F}$ is verified to have failed if there exists a solution $\boldsymbol{z} \geq \boldsymbol{0}$ to

$$[\boldsymbol{A}_D^T, \boldsymbol{A}_x^T, \boldsymbol{w}^T, \boldsymbol{1}]\boldsymbol{z} = \boldsymbol{0}, \tag{20a}$$

$$[\hat{\boldsymbol{g}}_D^T, \hat{\boldsymbol{g}}_x^T, g_w, \boldsymbol{0}]\boldsymbol{z} < 0, \tag{20b}$$

where $\boldsymbol{w} \in \{0,1\}^{|E_H|}$ is defined to be $\boldsymbol{W}_f$ with $Q_f = \{l\}$, and $g_w := -\eta$. Similarly, a line $e \in E_H \setminus \hat{F}$ is verified to be operational if $\exists \boldsymbol{z} \geq \boldsymbol{0}$ that satisfies (20)*, where $\boldsymbol{w} \in \{0,1\}^{|E_H|}$ is defined to be $\boldsymbol{W}_m$ with $Q_m = \{e\}$, and $g_w := \eta - 1$.*

2) Verification Algorithm: All the elements in (20) are known, and thus the existence of a solution can be checked by solving an LP. Based on this result, we propose an algorithm *VOTE with Partial Ground truth (VOTE-PG)* (Alg. 3) for verifying the estimated states of the remaining lines, which iteratively updates E_v. Each iteration of VOTE-PG involves solving $O(|E_H|)$ LPs, each of which has a time complexity that is polynomial[5] in the number of decision variables ($|E_H|$) and the number of constraints ($|V_H|+|E_H|$) [36]. Since VOTE-PG has at most $|E_H|$ iterations, the total time complexity of VOTE-PG is polynomial in $|E_H|$ and $|V_H|$.

Algorithm 3: VOTE with **P**artial **G**round truth (VOTE-PG)

Input: $\tilde{\boldsymbol{D}}, \boldsymbol{p}, \boldsymbol{\Delta}_{\bar{H}}, U_B, \eta, E_H, E_v, \hat{F}, \hat{\boldsymbol{g}}_D, \hat{\boldsymbol{g}}_x$

while $E_H \setminus E_v \neq \emptyset$ **do**
  $\bar{E}_v \leftarrow E_v$;
  foreach $e \in E_H \setminus E_v$ **do**
    if $\exists \boldsymbol{z} \geq \boldsymbol{0}$ *satisfying* (20) *for* e **then**
      $\bar{E}_v \leftarrow \bar{E}_v \cup \{e\}$;
      Update $\hat{\boldsymbol{g}}_x$;
  if $|\bar{E}_v| > |E_v|$ **then**
    $E_v \leftarrow \bar{E}_v$;
  else
    break;

Summary and Guidelines: In summary, Lemma III.2 is the foundation of our results. Based on Lemma III.2, we develop Theorems III.1-III.2 to understand the relationship between the feasibility of (13) in Lemma III.2 and the magnitudes/directions of power flows. Equipped with Lemma III.2 and Theorems III.1-

[5]The exact order of the polynomial depends on the specific algorithm used to solve the LP [36].

III.2, we develop Theorems IV.1-IV.3 based on verifiable conditions, which lead to the first verification algorithm (VOTE in Alg. 2). Then, we develop Theorem IV.4 to provide more verifiable conditions, which supports the second verification algorithm (VOTE-PG in Alg. 3).

The proposed algorithms form a three-step pipeline: FLD $\rightarrow$ VOTE $\rightarrow$ VOTE-PG, where FLD will estimate a set of failed lines ($\hat{F}$), based on which VOTE will identify a subset of lines (E_v) whose estimated states can be verified to be correct, and VOTE-PG will try to expand E_v.

All the proposed algorithms contain a parameter η. From Line 2 of FLD in Alg. 1, it is easy to see that a smaller η will make FLD less likely to miss failed lines. However, a smaller η will also make a failed line harder to be verified as analyzed in Theorem IV.1-IV.3. Similarly, an operational line is less likely to be detected as failed by FLD but harder to be verified with a larger η. Fortunately, there is no need to tune η. As shown in Fig. 4, the operator can run the proposed method with different values of η in parallel. For each value of η, the proposed method will return a set of lines with verified states (which are guaranteed to be consistent with the ground-truth states). Then, the operator can take the union of the sets obtained under different η values to recover more line states.

Remark: If control center knows that the post-attack grid remains connected, both Lemma III.2 and VOTE can be enhanced. The details are given in Appendix B.

V. Extension to AC Power Flow Model

So far we have assumed the DC power flow approximation as described in Section II. As the real grid behaves according to the AC power flow model, the natural questions are: (i) if we can directly apply the DC-based solution (FLD) under the AC model, and (ii) if we can adapt these algorithms to work better under the AC model.

For the first question, it is easy to see that FLD can be directly applied under the AC model. As for the verification algorithms, we have the following result (see proof in Appendix D).

Lemma V.1. *The DC-based line state verification algorithms VOTE can correctly verify the line states under the AC model.*

Despite the applicability of DC-based algorithms, the approximation error in the DC power flow model degrades the performance of failure detection and verification (see Section VI-A). Fortunately, we will show that FLD can be easily adapted to suit the AC model. Since only minor modification is needed, we will use "AC-X" to denote the AC-based modification of result "X".

A. Detection: Adaptation of FLD to AC-FLD

We first show how to adapt the failure detection algorithm FLD. Some notations necessary for presenting the results under the AC model are shown in Table II. Specifically, $D_{f,u,e} = 1$ and $D_{t,v,e} = 1$ if and only if $\exists e = (u, v) \in E$, i.e., $\boldsymbol{D} = \boldsymbol{D}_f - \boldsymbol{D}_t$. Based on a similar discussion as in Section II-C, we assume that the voltage after attack ($\vec{\boldsymbol{v}}'$) has been recovered through existing mechanisms [21, Lemma 1] or secured PMUs. We refer to Appendix A for details.

Table II
Notations for AC power flow

Notation	Description
$\boldsymbol{p}/\boldsymbol{q} \in \mathbb{C}^{\lvert V\rvert}$	Active/reactive power injection
$\boldsymbol{\Delta}_p/\boldsymbol{\Delta}_q \in \mathbb{C}^{\lvert V\rvert}$	Active/reactive power injection change
$\vec{v}_u = v_u e^{j\cdot\theta_u}$ / l_u	Nodal voltage/current
$\boldsymbol{Y} = \boldsymbol{G} + j\boldsymbol{B}$	Bus admittance matrix
$\boldsymbol{D}_f/\boldsymbol{D}_t \in \{0,1\}^{\lvert V\rvert\times\lvert E\rvert}$	*From/to* end incidence matrix
$\boldsymbol{Y}_f/\boldsymbol{Y}_t \in \mathbb{C}^{\lvert E\rvert\times\lvert V\rvert}$	*From/to* end line admittance matrix

The key is to extend (P1) in (7) to the AC model. To this end, we first derive the counterpart of (4). Recall that $\boldsymbol{x} \in \{0,1\}^{|E|}$ indicates which lines have failed, i.e., $x_e = 1$ indicates $e \in F$. Recalling that $[\boldsymbol{x}]$ denotes the diagonal matrix with $\boldsymbol{x}$ on the main diagonal and noticing that the post-attack bus admittance matrix is $\boldsymbol{Y}' = \boldsymbol{Y} - \boldsymbol{D}_f[\boldsymbol{x}]\boldsymbol{Y}_f + \boldsymbol{D}_t[\boldsymbol{x}]\boldsymbol{Y}_t$, we can transform the AC power flow equation $\boldsymbol{l}'_H = \boldsymbol{Y}'_{H|G}\vec{\boldsymbol{v}}'$ into

$$\boldsymbol{l}'_H = \boldsymbol{Y}_{H|G}\vec{\boldsymbol{v}}' - \boldsymbol{D}_{f,H|G}[\boldsymbol{x}_H]\boldsymbol{Y}_{f,H|G}\vec{\boldsymbol{v}}' - \boldsymbol{D}_{t,H|G}[\boldsymbol{x}_H]\boldsymbol{Y}_{t,H|G}\vec{\boldsymbol{v}}'. \quad (21)$$

Then, by left multiplying the conjugate of both sides of (21) by $[\vec{\boldsymbol{v}}'_H]$, we have

$$\boldsymbol{\Delta}_{p,H} = \boldsymbol{p}_H - \mathrm{Re}\left([\vec{\boldsymbol{v}}'_H]\overline{\boldsymbol{Y}_{H|G}\vec{\boldsymbol{v}}'}\right) + \tilde{\boldsymbol{D}}_{p,H}\boldsymbol{x}_H, \quad (22a)$$

$$\boldsymbol{\Delta}_{q,H} = \boldsymbol{q}_H - \mathrm{Im}\left([\vec{\boldsymbol{v}}'_H]\overline{\boldsymbol{Y}_{H|G}\vec{\boldsymbol{v}}'}\right) + \tilde{\boldsymbol{D}}_{q,H}\boldsymbol{x}_H, \quad (22b)$$

where $\tilde{\boldsymbol{D}}_{p,H} = \mathrm{Re}\left(\tilde{\boldsymbol{D}}_H\right)$, $\tilde{\boldsymbol{D}}_{p,H} = \mathrm{Im}\left(\tilde{\boldsymbol{D}}_H\right)$, and

$$\tilde{\boldsymbol{D}}_H = [\vec{\boldsymbol{v}}'_H]\left(\overline{\boldsymbol{D}_{f,H|G}[\boldsymbol{Y}_{f,H|G}\vec{\boldsymbol{v}}']} + \overline{\boldsymbol{D}_{t,H|G}[\boldsymbol{Y}_{t,H|G}\vec{\boldsymbol{v}}']}\right). \quad (23)$$

Here we slightly abuse the notation for $\tilde{\boldsymbol{D}}_H$ since it indicates the hypothetical power flow after attack in (4) for DC model and (23) for AC model, respectively. Let $V_{H,L,I} = \{u \in V_{H,L} : q_{H,u} \le 0\}$. Then, we introduce the row selection matrix $\boldsymbol{\Lambda}_I \in \{0,1\}^{|V_{H,L,I}|\times|V_H|}$ to select entries in $\boldsymbol{q}_H$ corresponding to nodes in $V_{H,L,I}$. Similarly, we introduce $V_{H,L,C} = \{u \in V_{H,L} : q_{H,u} > 0\}$ and the associated $\boldsymbol{\Lambda}_C \in \{0,1\}^{|V_{H,L,C}|\times|V_H|}$. As the counterpart of (6b) for reactive power, we have

$$\boldsymbol{0} \ge \boldsymbol{\Lambda}_I\boldsymbol{\Delta}_{q,H} \ge \boldsymbol{\Lambda}_I\boldsymbol{q}_H, \boldsymbol{0} < \boldsymbol{\Lambda}_C\boldsymbol{\Delta}_{q,H} \le \boldsymbol{\Lambda}_C\boldsymbol{q}_H. \quad (24)$$

Now, we are ready to give the counterpart of (P1) in (8) under the AC power flow model, referred to as AC-(P1), as follows

$$\min_{\boldsymbol{x}_H} \quad \boldsymbol{1}^T\boldsymbol{x}_H \quad (25a)$$

$$\text{s.t.} \quad (22), (24), (6a) - (6b), \quad (25b)$$

$$0 \le x_{H,e} \le 1, \forall e \in E_H, \quad (25c)$$

Thus, FLD can be adapted to the AC model by replacing (P1) in (8) by AC-(P1) in (25), which will be called AC-FLD in the sequel.

B. Verification: Adaptation of VOTE(-PG) to AC-VOTE(-PG)

We now show how to adapt the verification algorithms VOTE. Although Lemma V.1 guarantees that VOTE/VOTE-PG can still be used to verify the estimated line states under the AC model, they are developed under the DC model and thus have

degraded performance (see Section VI-A). To address this issue, we will develop AC-based counterparts of VOTE(-PG) by deriving the counterpart of Lemma III.2 for AC-FLD, which is the foundation of the verification algorithms.

To begin with, we will transform (25) into an equivalent LP without equality constraints, as in the transformation of (P1) into (11). To achieve this, we notice that any feasible $(\boldsymbol{p}'_H, \boldsymbol{q}'_H, \boldsymbol{x}_H)$ satisfying (22) can be represented as (26):

$$\begin{bmatrix} \boldsymbol{\Delta}_{p,H} \\ \boldsymbol{\Delta}_{q,H} \\ \boldsymbol{x}_H \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Delta}^*_{p,H} \\ \boldsymbol{\Delta}^*_{q,H} \\ \boldsymbol{x}^*_H \end{bmatrix} + \sum_{e\in E_H} c_e \begin{bmatrix} \tilde{\boldsymbol{d}}_{p,H,e} \\ \tilde{\boldsymbol{d}}_{q,H,e} \\ \boldsymbol{u}_e \end{bmatrix}. \quad (26)$$

Then, we modify the definitions in (12) as follows: we keep $\boldsymbol{A}_x, \boldsymbol{W}, \boldsymbol{g}_x, \boldsymbol{g}_w$ the same, and redefine $\boldsymbol{A}_D$ and $\boldsymbol{g}_D$ as

$$\begin{aligned} \boldsymbol{A}_D^T :=& \big[\boldsymbol{\Lambda}_L^T \tilde{\boldsymbol{D}}_{p,H}^T, -\boldsymbol{\Lambda}_L^T \tilde{\boldsymbol{D}}_{p,H}^T, -\boldsymbol{\Lambda}_S^T \tilde{\boldsymbol{D}}_{p,H}^T, \boldsymbol{\Lambda}_S^T \tilde{\boldsymbol{D}}_{p,H}^T \\ & \boldsymbol{\Lambda}_I^T \tilde{\boldsymbol{D}}_{q,H}^T, -\boldsymbol{\Lambda}_I^T \tilde{\boldsymbol{D}}_{q,H}^T, -\boldsymbol{\Lambda}_C^T \tilde{\boldsymbol{D}}_{q,H}^T, \boldsymbol{\Lambda}_C^T \tilde{\boldsymbol{D}}_{q,H}^T\big], \quad (27a) \\ \boldsymbol{g}_D^T :=& \big[-\boldsymbol{\Lambda}_L^T \boldsymbol{\Delta}_{p,H}^{*T}, -\boldsymbol{\Lambda}_L^T \boldsymbol{p}_H'^{*T}, \boldsymbol{\Lambda}_S^T \boldsymbol{\Delta}_{p,H}^{*T}, \boldsymbol{\Lambda}_S \boldsymbol{p}_H'^{*T} \\ & - \boldsymbol{\Lambda}_I^T \boldsymbol{\Delta}_{q,H}^{*T}, -\boldsymbol{\Lambda}_I^T \boldsymbol{q}_H'^{*T}, \boldsymbol{\Lambda}_C^T \boldsymbol{\Delta}_{q,H}^{*T}, \boldsymbol{\Lambda}_C^T \boldsymbol{q}_H'^{*T}\big]. \quad (27b) \end{aligned}$$

Equipped with (26) and (27), we can obtain the following LP that is equivalent to (25):

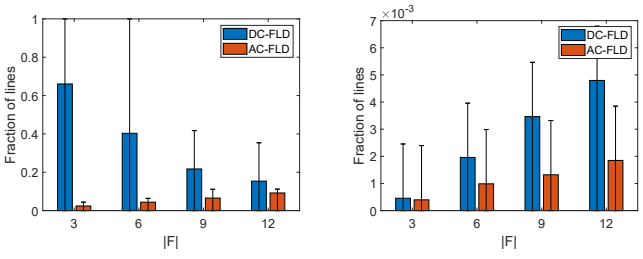
$$\begin{aligned} \min_{\boldsymbol{c}} \quad & \mathbf{1}^T \boldsymbol{c} && (28a) \\ \text{s.t.} \quad & \boldsymbol{A}_D \boldsymbol{c} \le \boldsymbol{g}_D, && (28b) \\ & \boldsymbol{A}_x \boldsymbol{c} \le \boldsymbol{g}_x, && (28c) \end{aligned}$$

where (28b) corresponds to (25b) while (28c) corresponds to (25c). Since the feasible region of (11) can also be written in the form of (28b)-(28c), AC-Lemma III.2 for AC-FLD has the same form as Lemma III.2 with $\boldsymbol{A}_D$ and $\boldsymbol{g}_D$ redefined as in (27). See Appendix D for the proof of AC-Lemma III.2.

Since Theorem IV.1-IV.4 are all proved by contradiction based on Lemma III.2, the corresponding algorithms (VOTE based on Theorem IV.1-IV.3 and VOTE-PG based on Theorem IV.4) can be easily adapted to AC-VOTE and AC-VOTE-PG with changed $\boldsymbol{A}_D$ and $\boldsymbol{g}_D$. In Appendix C, we provide the adapted theorems and discuss how they are used in AC-VOTE and AC-VOTE-PG.

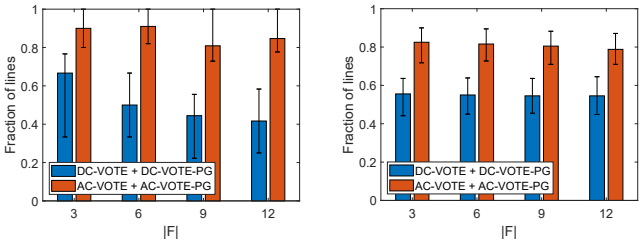
VI. Performance Evaluation

We will primarily evaluate our findings on the Polish power grid ("Polish system - winter 1999-2000 peak") [37] with 2383 nodes and 2886 lines (where parallel lines are combined). The ground-truth power grid states are generated according to AC power flow model. Key solutions (FLD, VOTE and VOTE-PG) will also be evaluated on the IEEE 300-bus system extracted from MATPOWER [37] to test their generality. We generate the attacked area H by randomly choosing one node as a starting point and performing a BFS to obtain H with a predetermined $|V_H|$. The generated H consists of buses topologically close to each other, which will intuitively share communication lines in connecting to the control center and can thus be blocked together once a cyber attack jams some of these lines. Note, however, that our solution does not depend on this specific way of forming H. We then randomly choose $|F|$ lines within H to fail. We vary $|V_H|$ and $|F|$ to explore different settings,



(a) Miss rate (b) False alarm rate

Figure 5. Performance of DC-FLD under the AC power flow model in Polish system ($|V_H| = 40$).



(a) Failed lines (b) Operational lines

Figure 6. Performance of DC-VOTE + DC-VOTE-PG under the AC power flow model in Polish system ($|V_H| = 40$).

and for each setting, we generate 70 different H's and 300 different F's per H. In contrast to previous works [3]–[7] where $|F| \le 3$, we focus on the scenarios where both $|V_H|$ and $|F|$ are large such that there are likely to be internal nodes in H whose post-attack active power injections cannot be recovered, and there are likely to be island formation in the post-attack grid. In our simulations, we assume that all viable islands have survived the attack (i.e., no frequency collapse), but this assumption is not necessary for our algorithms.

We evaluate two types of metrics: (1) how accurate FLD is, and (2) how often the line states estimated by FLD can be verified. Each evaluated metric is shown via the mean and the $25^{\text{th}}/75^{\text{th}}$ percentile (indicated by the error bars) when applicable. The threshold η is set to 0.5. A detailed case study is given in Appendix E.

A. Performance Loss of DC-based Algorithms

We start by evaluating the performance of the DC-based versions of FLD, VOTE, and VOTE-PG (denoted by DC-*) under the AC power flow model, since they are applicable under the AC model as discussed in Lemma V.1. In Fig. 5, we compare the performance of DC-FLD and AC-FLD in terms of miss rate and false alarm rate. In Fig. 6, we compare the performance of the combination of DC-VOTE and DC-VOTE-PG with their AC variants. We observe that although the DC-based algorithms are still applicable under the AC power flow model, the approximation error of the DC model leads to performance degradation in both detection and verification. Such observations validate the importance of deriving their AC variants, as shown in Section V. *In the rest of this section, all algorithms (including both proposed and benchmark algorithms) refer to their AC variants developed as in Section V.*

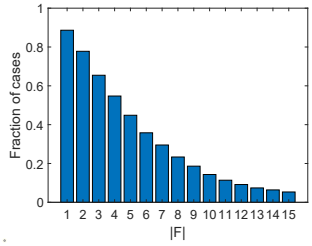
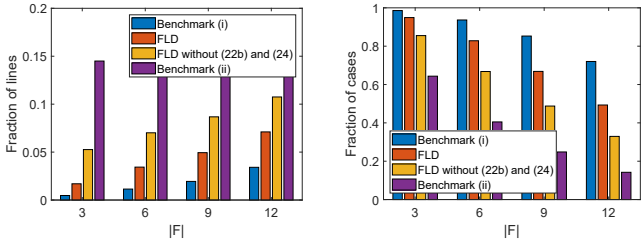


Figure 7. Prob. that assuming $\boldsymbol{\Delta} = \mathbf{0}$ leads to a feasible solution in Polish system ($|V_H| = 40$).



(a) Miss rate (b) Probability of no miss

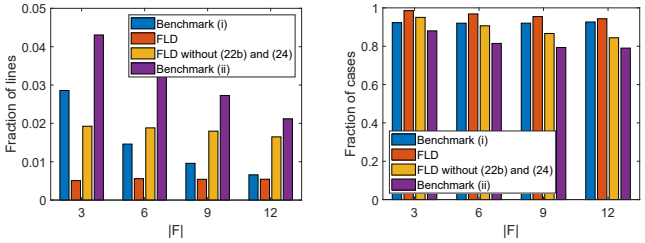
Figure 8. Performance comparison on miss rate in Polish system ($|V_H| = 40$).

B. Performance of Line State Recovery

First, we observe that for a nontrivial size of H ($|V_H| \geq 20$), there almost always exists $u \in V_H$ for which we cannot recover $\Delta_{H,u}$ by Lemma IV.1. We also observe that the solution in [5] (which assumes $\boldsymbol{\Delta} = \mathbf{0}$) is often infeasible, as shown in Fig. 7. These observations confirm the necessity of jointly estimating F and $\boldsymbol{\Delta}_H$ during failure localization.

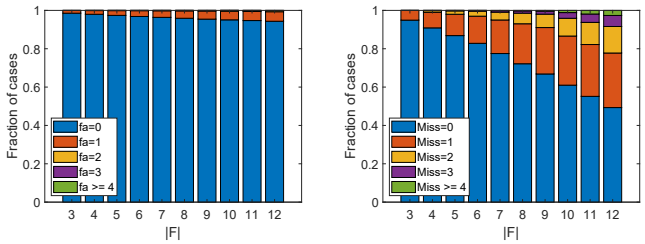
Next, we compare FLD with benchmarks in localizing the failed lines. We consider two benchmarks: (i) the solution extended from [21], i.e., i.e., estimating F by $\text{supp}(\boldsymbol{y})$ for the solution to $\min \|\boldsymbol{y}\|_1$ s.t. (22), assuming the knowledge of true $\boldsymbol{\Delta}_{p,H}$ and $\boldsymbol{\Delta}_{q,H}$, and (ii) $\min \|\boldsymbol{y}\|_1$ s.t. $\|\boldsymbol{p}_H - \text{Re}\left([\vec{\boldsymbol{v}}'_H]\overline{\boldsymbol{Y}_{H|G}\vec{\boldsymbol{v}}'}\right) + \tilde{\boldsymbol{D}}_{p,H}\boldsymbol{y}_H\| \leq \|\boldsymbol{p}_H\|$, which is adapted from the solutions in [3], [4]. In addition, we consider a variant of AC-FLD that removes the constraints (24) and (22b) to see the importance of adding constraints on reactive power. Note that benchmark (i) should be treated as a "performance upper bound" as it assumes more knowledge (of $\boldsymbol{\Delta}_H$) than our proposed algorithm.

As shown in Fig. 8, benchmark (i) demonstrates the best performance with regard to both the miss rate and the probability of having no miss, while FLD performs much better than benchmark (ii). This confirms the importance of knowing or estimating power injection changes in failure localization. Regarding the false alarm as shown in Fig. 9, FLD performs even better than benchmark (i). This is because the decision variable $\boldsymbol{x}$ in benchmark (i) combines the information about both the failed lines and the phase angles $\boldsymbol{\theta}'_H$, and thus does not fully exploit the knowledge of $\boldsymbol{\theta}'_H$. We also notice that adding the constraints (24) and (22b) can significantly improve detection accuracy by exploiting the knowledge on reactive power injections. Furthermore, from the specific number of false alarms/misses in Fig. 10, we see that besides having very few false alarms, FLD also correctly detects most of the failed lines with only a couple of misses for the majority of the time.



(a) False alarm rate (b) Probability of no false alarm

Figure 9. Performance comparison on false alarm rate in Polish system ($|V_H| = 40$).



(a) False alarm (b) Miss

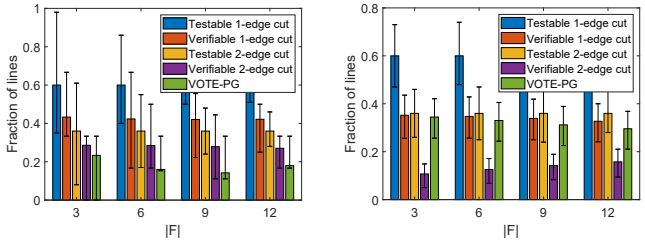
Figure 10. Number of false alarms/misses of FLD in Polish system ($|V_H| = 40$).

C. Performance of Line State Verification

In this subsection, we evaluate the performance of the proposed verification algorithms (VOTE and VOTE-PG) in terms of the fraction of verifiable lines.

We first evaluate the fraction of verifiable lines in E_a (lines in 1-edge cuts) and E_c (lines in 2-edge cuts, i.e., $E_c := \bigcup_{s\in\mathcal{E}_c} s$), as shown in Fig. 11. For each generated case (combination of H and F), denote $E_{a,v} := E_a \cap E_v$ and $E_{c,v} := E_c \cap E_v$. Then in Fig. 11(a), we evaluate the fractions of testable and verifiable lines in E_a (E_c) among failed lines, i.e., $\frac{|E_a \cap F|}{|F|}$ ($\frac{|E_c \cap F|}{|F|}$) and $\frac{|E_{a,v} \cap F|}{|F|}$ ($\frac{|E_{c,v} \cap F|}{|F|}$). The evaluation for operational lines is conducted similarly in Fig. 11(b). As can be seen, (i) the fractions of testable and verifiable lines both stay almost constant with varying $|F|$, which demonstrates the robustness of VOTE; (ii) among the testable lines ($E_a \cup E_c$), most of the failed lines are verifiable, but only half of the operational lines are verifiable.

Next, we use two metrics to evaluate the value of VOTE-PG. The first metric is the fraction of lines verified by VOTE-PG but not VOTE, as shown in Fig. 11 as 'VOTE-PG'. The second is the percentage of cases that VOTE-PG can verify additional lines, given in Table III. We observe that VOTE-PG

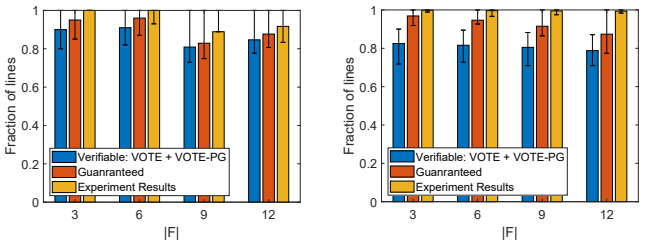


(a) Fraction of failed lines (b) Fraction of operational lines

Figure 11. Fraction of testable/verifiable lines in Polish system ($|V_H| = 40$).

Table III
PERCENTAGE OF CASES THAT VOTE-PG VERIFIES ADDITIONAL LINES IN POLISH SYSTEM

Type of lines	$\|F\| = 3$	$\|F\| = 6$	$\|F\| = 9$	$\|F\| = 12$
Failed lines	24.12%	38.94%	48.72%	57.43%
Operational lines	88.15%	91.45%	92.13%	92.45%
All lines	90.34%	93.61%	95.22%	95.71%



(a) Fraction of failed lines (b) Fraction of operational lines

Figure 12. Comparison between verifiable lines, theoretically guaranteed lines, and actually correctly identified lines in Polish system ($|V_H| = 40$).

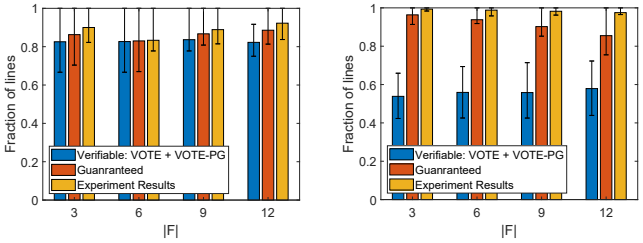
can usually verify more lines based on the results of VOTE.

Then, we compare the fraction of verifiable lines ('Verifiable – VOTE + VOTE-PG') to the fraction of lines whose states are guaranteed to be correctly estimated by FLD according to Lemma III.2 ('Guaranteed') and the actual fraction of lines whose states are correctly estimated by FLD ('Experiment Results'), as shown in Fig. 12. We see that over 80% of the lines are verifiable. Nevertheless, the fraction of lines whose states are correctly estimated by FLD is even higher: out of all the failed lines, over 80% will be estimated as failed and verified as so, while another 10% will be estimated as failed but not verified; out of all the operational lines, over 80% will be estimated and verified as operational, while the rest will also be estimated as operational but not verified. We have confirmed that the set of verifiable lines is a subset of the set of lines for which FLD is guaranteed to be correct (by Lemma III.2), which is in turn a subset of the set of correctly identified lines.

To validate our key observations, we repeat the experiments in Fig. 12 on the IEEE 300-bus system, as shown in Fig. 13. Compared with the results from the Polish system, most of the results from the IEEE 300-bus system are qualitatively similar. One notable difference is that although most of the failed lines remain verifiable in the 300-bus system, only half of the operational lines are verifiable. This indicates that most (80-90%) of the unverifiable lines are operational. To understand such a phenomenon, we observe that many operational lines carry small post-attack power flows, which makes the conditions in Theorem IV.1-IV.3 hard to satisfy. On the contrary, the values of hypothetical power flows on failed lines are usually large.

D. Summary of Observations

1) While the DC-based detection/verification algorithms can be applied in a grid that follows AC power flow equations, their AC-based variants provide better performance.
2) Under the possibility of islanding caused by the physical attack, existing failed line detection algorithms fail with high probability due to the change of power injections (Fig. 7), but the proposed FLD can handle the change of power injections and achieve high accuracy.
3) Despite its high accuracy, FLD can still miss failed lines and falsely report failures on operational lines, which will cause problems during recovery.
4) The proposed line state verification algorithms (VOTE and VOTE-PG) can substantially reduce the waste of resources during recovery by providing reliable information on the verifiable line states.



(a) Fraction of failed lines (b) Fraction of operational lines

Figure 13. Comparison between verifiable lines, theoretically guaranteed lines, and actually correctly identified lines in IEEE 300-bus system ($|V_H| = 40$).

VII. CONCLUSION

We investigated the problem of power grid state estimation under cyber-physical attacks that may decompose the grid into islands. Our focus was on recovering the line states within the attack area, due to an observation that existing solutions and their recovery conditions for recovering the phase angles (developed for the case of connected post-attack grid) remain applicable in the case of islanding. To handle the challenge of unknown changes in the power injections within the attack area caused by islanding, under the DC model we proposed an LP-based algorithm to jointly estimate the line states and the power injection changes within the attacked area. We established theoretical conditions under which the line states estimated by the proposed algorithm are guaranteed to be correct, including both more general conditions that depend on the ground-truth line states and less general conditions that only depend on observable information. The latter conditions are further used to develop two polynomial-time algorithms to verify the correctness of the estimated line states. In addition, we extend all results obtained under the DC model to their variants under the AC power flow model. Our evaluations based on the Polish power grid and the IEEE 300-bus system showed that the proposed algorithm is highly accurate in localizing the failed lines, and the correctness of its output can be verified in the majority of cases.

Compared to the previous solutions for line state estimation that label lines with binary states (failed/operational) without guaranteed correctness, our solution labels lines with ternary states (failed/operational/unverifiable), where the states of verifiable lines are identified with guaranteed correctness. This, together with the observation that most of the unverifiable lines are operational, provides valuable information for planning repair/restoration in the recovery process.

REFERENCES

[1] Y. Huang, T. He, N. R. Chaudhuri, and T. L. Porta, "Power grid state estimation under general cyber-physical attacks," in *IEEE SmartGridComm*. IEEE, 2020.

[2] 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.

[3] 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.

[4] 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.

[5] 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, pp. 499–512, 2018.

[6] ——, "React to cyber attacks on power grids," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 459–473, 2018.

[7] S. Soltan, P. Mittal, and H. V. Poor, "Line failure detection after a cyber-physical attack on the grid using bayesian regression," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3758–3768, 2019.

[8] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of cps security," *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.

[9] 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.

[10] ——, "Double line outage detection using phasor angle measurements," in *IEEE PES-GM*. IEEE, 2009, pp. 1–5.

[11] 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.

[12] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[13] 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*, vol. 14, no. 1, pp. 1–33, 2011.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] R. Kaviani and K. W. Hedman, "A detection mechanism against load-redistribution attacks in smart grids," *IEEE Transactions on Smart Grid*, 2020.

[19] M. Garcia, T. Catanach, S. Vander Wiel, R. Bent, and E. Lawrence, "Line outage localization using phasor measurement data in transient state," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3019–3027, 2016.

[20] 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.

[21] ——, "Expose the line failures following a cyber-physical attack on the power grid," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 1, pp. 451–461, 2018.

[22] L. Che, X. Liu, Z. Li, and Y. Wen, "False data injection attacks induced sequential outages in power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1513–1523, 2018.

[23] H.-M. Chung, W.-T. Li, C. Yuen, W.-H. Chung, Y. Zhang, and C.-K. Wen, "Local cyber-physical attack for masking line outage and topology attack in smart grid," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4577–4588, 2018.

[24] M. Jamei, R. Ramakrishna, T. Tesfay, R. Gentz, C. Roberts, A. Scaglione, and S. Peisert, "Phasor measurement units optimal placement and performance limits for fault localization," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 180–192, 2019.

[25] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on smart grid communication infrastructures: Motivations, requirements and challenges," *IEEE communications surveys & tutorials*, vol. 15, no. 1, pp. 5–20, 2012.

[26] S. Galli, A. Scaglione, and Z. Wang, "For the grid and through the grid: The role of power line communications in the smart grid," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 998–1027, 2011.

[27] H. Liang, B. J. Choi, A. Abdrabou, W. Zhuang, and X. S. Shen, "Decentralized economic dispatch in microgrids via heterogeneous wireless networks," *IEEE journal on Selected Areas in communications*, vol. 30, no. 6, pp. 1061–1074, 2012.

[28] I. S. Association *et al.*, "IEEE 1815-2012–ieee standard for electric power systems communications-distributed network protocol (DNP3)," 2012.

[29] C. Coffrin and P. Van Hentenryck, "Transmission system restoration with co-optimization of repairs, load pickups, and generation dispatch," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 144–154, 2015.

[30] Y. Huang, T. He, N. R. Chaudhuri, and T. L. Porta, "Link state estimation under cyber-physical attacks: Theory and algorithms," Technical Report, October 2021, https://sites.psu.edu/nsrg/files/2021/10/YudiStateRecoveryReport.pdf.

[31] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1216–1227, 2014.

[32] Q. Yang, L. Jiang, W. Hao, B. Zhou, P. Yang, and Z. Lv, "PMU placement in electric transmission networks for reliable state estimation against false data injection attacks," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1978–1986, 2017.

[33] Q. Yang, D. An, R. Min, W. Yu, X. Yang, and W. Zhao, "On optimal pmu placement-based defense against data integrity attacks in smart grid," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 7, pp. 1735–1750, 2017.

[34] P. Kundur, *Power System Stability and Control, ser. The EPRI power system engineering series*. New York: McGraw-Hill, 1994.

[35] 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.

[36] T. Terlaky, *Interior point methods of mathematical programming*. Springer Science & Business Media, 2013, vol. 5.

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

[38] "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.

[39] J. E. Dagle, "The North American SynchroPhasor Initiative (NASPI)," in *IEEE PES General Meeting*. IEEE, 2010, pp. 1–3.

[40] "SynchroPhasor technology fact sheet," North American SynchroPhasor Initiative, October 2014, https://www.naspi.org/sites/default/files/reference_documents/33.pdf.

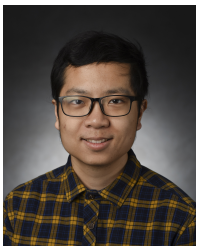
[41] "NASPI synchrophasor starter kit (draft)," North American SynchroPhasor Initiative (NASPI), October 2015, https://www.naspi.org/sites/default/files/reference_documents/4.pdf.

[42] K. D. Jones, J. S. Thorp, and R. M. Gardner, "Three-phase linear state estimation using phasor measurements," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

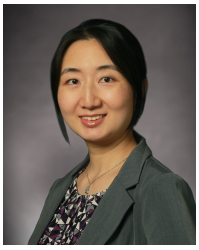
[43] K. D. Jones, A. Pal, and J. S. Thorp, "Methodology for performing synchrophasor data conditioning and validation," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1121–1130, 2014.

[44] S. Dasgupta, C. H. Papadimitriou, and U. V. Vazirani, *Algorithms*. McGraw-Hill Higher Education New York, 2008.

[45] O. L. Mangasarian, *Nonlinear programming*. SIAM, 1994.



Yudi Huang (S'21) received the B.Eng. and M.Eng. degrees in communication and information engineering from University of Electronic Science and Technology of China. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering in Pennsylvania State University. His main research interests are improving networked systems by applying statistical and optimization tools.



Ting He (SM'13) is an Associate Professor in the School of Electrical Engineering and Computer Science at Pennsylvania State University, University Park, PA. Her work is in the broad areas of computer networking, network modeling and optimization, and machine learning. Dr. He is a senior member of IEEE, an Associate Editor for IEEE Transactions on Communications (2017- 2020) and IEEE/ACM Transactions on Networking (2017-2021), a TPC Co-Chair of IEEE ICCCN (2022), and an Area TPC Chair of IEEE INFOCOM (2021). She received multiple Outstanding Contributor Awards from IBM, multiple awards for Military Impact, Commercial Prosperity, and Collaboratively Complete Publications from ITA, and multiple paper awards from ICDCS, SIGMETRICS, ICASSP, and IEEE Communications Society

Nilanjan Ray Chaudhuri (S'08-M'09-SM'16) received the Ph.D. degree in power systems from Imperial College London, London, UK in 2011. From 2005 to 2007, he worked in General Electric (GE) John F. Welch Technology Center. He came back to GE and worked in GE Global Research Center, NY, USA as a Lead Engineer during 2011 to 2014. Presently, he is an Associate Professor with the School of Electrical Engineering and Computer Science at Penn State, University Park, PA. He was an Assistant Professor with North Dakota State University, Fargo, ND, USA during 2014-2016. He is a member of the IEEE and IEEE PES. Dr. Ray Chaudhuri is the lead author of the book Multi-terminal Direct Current Grids: Modeling, Analysis, and Control (Wiley/IEEE Press, 2014). He served as an Associate Editor of the IEEE TRANSACTIONS ON POWER DELIVERY (2013 – 2019) and IEEE PES LETTERS (2016 - present). Dr. Ray Chaudhuri was the recipient of the National Science Foundation Early Faculty CAREER Award in 2016 and Joel and Ruth Spira Excellence in Teaching Award in 2019.

Thomas F. La Porta ((Fellow, IEEE) is the Director of the School of Electrical Engineering and Computer Science and Penn State University. He is an Evan Pugh Professor and the William E. Leonhard Chair Professor in the Computer Science and Engineering Department and the Electrical Engineering Department. He received his B.S.E.E. and M.S.E.E. degrees from The Cooper Union, New York, NY, and his Ph.D. degree in Electrical Engineering from Columbia University, New York, NY. He joined Penn State in 2002. He was the founding Director of the Institute of Networking and Security Research at Penn State. Prior to joining Penn State, Dr. La Porta was with Bell Laboratories for 17 years. He was the Director of the Mobile Networking Research Department in Bell Laboratories, Lucent Technologies where he led various projects in wireless and mobile networking. He is an IEEE Fellow, Bell Labs Fellow, received the Bell Labs Distinguished Technical Staff Award, and an Eta Kappa Nu Outstanding Young Electrical Engineer Award. He also won two Thomas Alva Edison Patent Awards.

Dr. La Porta was the founding Editor-in-Chief of the IEEE Transactions on Mobile Computing. He served as Editor-in-Chief of IEEE Personal Communications Magazine. He was the Director of Magazines for the IEEE Communications Society and was on its Board of Governors for three years. He has published numerous papers and holds 39 patents.