Joint Cyber and Physical Attacks on Power Grids: Graph Theoretical Approaches for Information Recovery

Saleh Soltan
Electrical Engineering
Columbia University
New York, NY
saleh@ee.columbia.edu

Mihalis Yannakakis
Computer Science
Columbia University
New York, NY
mihalis@cs.columbia.edu

Gil Zussman
Electrical Engineering
Columbia University
New York, NY
gil@ee.columbia.edu

ABSTRACT

Recent events demonstrated the vulnerability of power grids to cyber attacks and to physical attacks. Therefore, we focus on *joint cyber and physical attacks* and develop methods to retrieve the grid state information following such an attack. We consider a model in which an adversary attacks a zone by physically disconnecting some of its power lines and blocking the information flow from the zone to the grid's control center. We use tools from linear algebra and graph theory and leverage the properties of the power flow DC approximation to develop methods for information recovery. *Using information observed outside the attacked zone, these methods recover information about the disconnected lines and the phase angles at the buses.* We identify *sufficient conditions on the zone structure* and *constraints on the attack characteristics* such that these methods can recover the information. We also show that it is NP-hard to find an approximate solution to the problem of partitioning the power grid into the minimum number of attack-resilient zones. However, since power grids can often be represented by planar graphs, we develop a *constant approximation partitioning algorithm for these graphs*. Finally, we numerically study the relationships between the grid's resilience and its structural properties, and demonstrate the partitioning algorithm on real power grids. The results can provide insights into the design of a secure control network for the smart grid.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Reliability, availability, and serviceability; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms, Network problems*

Keywords

Power Grids; Cyber Attacks; Physical Attacks; Information Recovery; Graph Theory; Algorithms

SIGMETRICS'15, June 15–19, 2015, Portland, OR, USA.

ACM 978-1-4503-3486-0/15/06 ...$15.00.
http://dx.doi.org/10.1145/2745844.2745846.

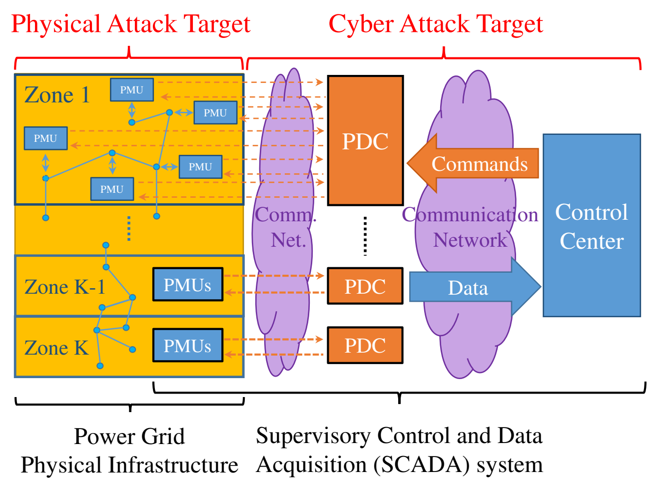


Figure 1: Components of the power grid and potential attacks: physical attacks target the physical infrastructure (lines, substations, etc.); Cyber attacks target the SCADA system – an adversary can disallow the information from the PMUs within the zone to reach the control center.

1. INTRODUCTION

Power grids are vulnerable to cyber attacks [7] and to physical attacks (e.g., the Apr. 2014 attack on a California substation [6]). These attacks may cause large-scale failures, initiate cascades (e.g., [3–5]), and have devastating effects on almost every aspect of modern life. As illustrated in Fig. 1, two main components of the power grid are (i) the physical infrastructure of the power transmission system (power lines, substations, power stations), and (ii) the Supervisory Control and Data Acquisition (SCADA) system responsible for monitoring and controlling the grid (we refer to it as the *control network*). Physical attacks target the former while cyber attacks target the latter.

The effects of a physical attack can be mitigated, if the control center has accurate understanding of its impacts and acts quickly to compensate for failures. However, if physical attacks are accompanied by cyber attacks that make information about the status of the attacked zones unavailable, the control center cannot take effective action. Hence, in this paper *we focus on developing methods for recovering the information about the status of the power grid following a joint cyber and physical attack as well as on studying the resilience of different topologies and the resilience to different attacks.*

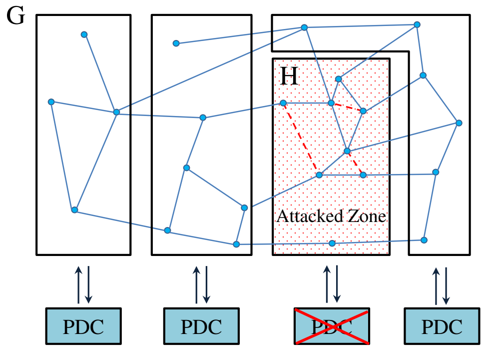


Figure 2: G is the power grid graph and H is an induced subgraph of G that represents the attacked zone. An adversary attacks a zone by disconnecting some of its power lines (red dashed lines) and disallowing the information from the PMUs within the zone to reach the control center.

We consider the linearized *direct-current (DC) power flow model*[1], a practical relaxation of the alternating-current (AC) model. We also use a simplified version of the *control network model* of [28] that includes Phasor Measurement Units (PMU), Phasor Data Concentrators (PDC), and a control center (see Fig. 1). We define a *zone* as a set of buses (nodes), power lines (edges), PMUs, and an associated PDC. We consider an attack on a zone that disconnects some edges within the zone (physical attack), and disallows the information from the PMUs within the zone to reach the control center (cyber attack). An adversary can perform the cyber attack by, for example, disabling the zone's associated PDC. Alternatively, the communication network between the PMUs and the PDC or between the PDC and the control center can be attacked.

As illustrated in Fig. 2, as a result of an attack, some lines get disconnected, and the phase angles and the status of the lines within the *attacked zone* $H = (V_H, E_H)$ become unavailable. Our objective is to recover the phase angles and detect the disconnected lines by using the information available outside of the attacked zone.

The idea underlying many of our results is that power flows are governed by the laws of physics, where a failure changes line flows and nodes' phase angles all over the power grid [40]. We show that based on this physical property, it is possible to estimate the state in the attacked zone using the information available outside of the zone. Specifically, we use the matrix representation of the DC equations and the notion of *admittance matrix*, and apply matrix analysis and graph theoretical tools to *develop methods for retrieving the information from the attacked zone.*

We present *necessary and sufficient conditions on the structure of a zone such that our methods are guaranteed to recover the information.* We prove that if there is a *matching* between the nodes inside and outside the attacked zone that covers the inside nodes (V_H), then the phase angles of the nodes in the attacked zone are recoverable by solving a set of linear equations of size $|V_H|$. We also prove that if H is *acyclic*, then the disconnected lines in H are detectable by solving a set of linear equations of size $|E_H|$. Moreover, we show that if H is *planar*, under some constraints, the disconnected lines are detectable by solving a Linear Programming (LP) problem.

We develop another method to simultaneously recover the phase angles and detect disconnected lines by solving a single LP problem. We show that if there is a *partial matching* between the nodes inside and outside of H, and H is *planar*, then under certain constraints on the attack (i.e., on the disconnected lines), this method is guaranteed to recover the information. Based on these results, we present the Post-Attack Recovery and Detection (PARD) Algorithm.

We study the problem of partitioning the power grid into the minimum number of attack-resilient zones (i.e., zones in which the information can be recovered by the methods mentioned above). We show that this problem is not approximable to within $n^{1-\epsilon}$ for all $\epsilon > 0$, unless P=NP. However, since power grids can often be represented by planar graphs, we introduce the Zone Selection (ZS) Algorithm and show that it provides a constant approximation ratio for these graphs. This algorithm can be used for designing a more secure control network for smart grids.

Finally, we present numerical results to assess the relationship between the structural properties of the power grid and its resilience to joint attacks and to demonstrate the operation of the ZS Algorithm on several power grids.

The main contributions of this paper are three fold. We use matrix analysis and graph theoretical tools to: (i) develop methods to recover the phase angles and detect the disconnected lines after a joint cyber and physical attack, (ii) find graph classes for which these methods are guaranteed to recover the information, (iii) develop an algorithm for partitioning the power grid into attack-resilient zones.

This paper is organized as follows. Section 2 reviews related work. Section 3 describes the models and reviews graph theoretical terms. In Section 4, we focus on information recovery and in Section 5, we study the grid partitioning problem. Section 6 provides numerical results and Section 7 provides concluding remarks and directions for future work. Most of the proofs appear in Appendix B.

2. RELATED WORK

The vulnerability of general networks to attacks was thoroughly studied in the past (e.g., [8, 31, 36] and references therein). Specifically, attacks and failures in power grids were studied using probabilistic failure propagation models (e.g., [18, 20, 47], and references therein) as well as using deterministic DC power flows [12, 14, 22, 32, 40]. Malicious data attacks on the power grid control network were also studied [21, 30, 33, 44]. However, to the best of our knowledge, none of the previous works considered power grid vulnerability to joint cyber and physical attacks.

In Section 4, we study the problem of recovering the phase angles and detecting disconnected edges after a joint cyber and physical attack. It is related to the problem of line outages identification from changes in phase angles that was studied in [42, 43]. However, these studies were limited to at most two line failures. The problem of line failure identification in an internal system using the information from an external system was studied in [49], which only provided a heuristic algorithm that was evaluated for only one or two line failures.

In Section 5, we study the problem of partitioning the power grid into the minimum number of attack-resilient zones. This problem is similar to the PMU placement problems

[1]The DC model is commonly used in large-scale contingency analysis of power grids [14, 15, 37, 49].

studied in [29,35,48]. Recently, this problem has attracted a lot of attention in India after the major blackouts of 2013 [29]. In [48] the problem of PMU placement for line outage detection was studied. However, none of the previous works considered the problem of PMU placement from the security point of view and when both the PDC/PMUs and the physical network can be attacked.

In Section 5, we reduce the attack-resilient zone partitioning problem to the problem of partitioning a graph into subgraphs such that each subgraph is (i) acyclic, and (ii) there is a matching between nodes inside and outside the subgraph that covers all the subgraph nodes. This problem is closely related to the problems of *vertex arboricity* (which is known to be NP-hard to be determined [26, p.193]) and *k-matching cover* of a graph (which can be found in $O(n^3)$ time [45]). However, to the best of our knowledge, the joint problem was not studied before.

Since the power grid and its robustness have drawn a lot of recent attention, several performance evaluation tools were used successfully to analyze power grids (e.g., [25, 27, 34], and references therein). This paper continues this line of work, using tools from linear algebra and graph theory.

3. MODEL AND DEFINITIONS

3.1 DC Power Flow Model

We adopt the linearized (or DC) power flow model, which is widely used as an approximation for the non-linear AC power flow model [11]. In particular, we follow [13–15] and represent the power grid by a connected undirected graph $G = (V, E)$ where $V = \{1, 2, \ldots, n\}$ and $E = \{e_1, \ldots, e_m\}$ are the set of nodes and edges corresponding to the buses and transmission lines, respectively. Each edge e_i is a set of two nodes $e_i = \{u, v\}$. p_v is the active power *supply* ($p_v > 0$) or *demand* ($p_v < 0$) at node $v \in V$ (for a *neutral node* $p_v = 0$). We assume *pure reactive* lines, implying that each edge $\{u, v\} \in E$ is characterized by its *reactance* $r_{uv} = r_{vu}$.

Given the power supply/demand vector $\vec{p} \in \mathbb{R}^{|V|\times 1}$ and the reactance values, a *power flow* is a solution $\mathbf{P} \in \mathbb{R}^{|V|\times|V|}$ and $\vec{\theta} \in \mathbb{R}^{|V|\times 1}$ of:

$$\sum_{v\in N(u)} p_{uv} = p_u,\ \forall\ u \in V \tag{1}$$

$$\theta_u - \theta_v - r_{uv}p_{uv} = 0,\ \forall\ \{u, v\} \in E \tag{2}$$

where $N(u)$ is the set of neighbors of node u, p_{uv} is the power flow from node u to node v, and θ_u is the phase angle of node u. Eq. (1) guarantees (classical) flow conservation and (2) captures the dependency of the flow on the reactance values and phase angles. Additionally, (2) implies that $p_{uv} = -p_{vu}$. When the total supply equals the total demand in each connected component of G, (1)-(2) has a unique solution [15, lemma 1.1].[2] Eq.(1)-(2) are equivalent to the following matrix equation:

$$\mathbf{A}\vec{\theta} = \vec{p} \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{|V|\times|V|}$ is the *admittance matrix* of G,[3] defined as follows:

$$a_{uv} = \begin{cases} 0 & \text{if } u \neq v \text{ and } \{u, v\} \notin E, \\ -1/r_{uv} & \text{if } u \neq v \text{ and } \{u, v\} \in E, \\ -\sum_{w\in N(u)} a_{uw} & \text{if } u = v. \end{cases}$$

Note that in power grids nodes can be connected by multiple edges, and therefore, if there are k multiple edges between nodes u and v, $a_{uv} = -\sum_{i=1}^{k} 1/r_{uv_i}$. Once $\vec{\theta}$ is computed, the flows, p_{uv}, can be obtained from (2).

Notation. Throughout this paper we use bold uppercase characters to denote matrices (e.g., $\mathbf{A}$), italic uppercase characters to denote sets (e.g., V), and italic lowercase characters and overline arrow to denote column vectors (e.g., $\vec{\theta}$). For a matrix $\mathbf{Q}$, q_{ij} denotes its $(i, j)^{\text{th}}$ entry. For a column vector $\vec{y}$, $\vec{y}^t$ denote its transpose, y_i denotes its i^{th} entry, $\|\vec{y}\|_1 := \sum_{i=1}^{n} |y_i|$ is its l_1-norm, and $\text{supp}(\vec{y}) := \{i|y_i \neq 0\}$ is its support.

3.2 Control Network

We use the simplified version of the model described in [28] to model the SCADA system to which we refer as the control network. Fig. 1 illusterates the components of the control network. We assume that there is a Phasor Measurement Unit (PMU) at each node of G. The PMU at node i reports the phase angle θ_i as well as the status of the edges (either *operational* or *failed*) adjacent to node i. Phasor Data Concentrators (PDC) gather the data collected by PMUs. The data gathered by PDCs is sent to a control center which monitors and controls the entire grid. A *zone* is a subgraph induced by a subset of nodes with a single associated PDC.

3.3 Attack Model

We study attacks on power grids that affect both the physical infrastructure and the control network. We assume that an adversary attacks a zone by: (i) disconnecting some edges within the attacked zone (physical attack), and (ii) disallowing the information from the PMUs within the zone to reach the control center (cyber attack). An adversary can perform the cyber attack by, for example, disabling the zone's associated PDC. Alternatively, the communication network between the PMUs and the PDC or between the PDC and the control center can be attacked. We assume that disconnecting edges within a zone does not make G disconnected.

Fig. 2 shows an example of an attack on the zone represented by H. Due to the attack, some edges are disconnected (we refer to these edges as *failed edges*) and the phase angles and the status of the edges within the *attacked zone* become unavailable. We denote the set of failed edges in zone H by $F \subseteq E_H$. Upon failure, the edges are removed from the graph and the flows are redistributed according to (1)-(2).

Notation. Throughout this paper, we denote an attacked zone by $H = (V_H, E_H)$. Without loss of generality we assume that the indices are such that $V_H = \{1, 2, \ldots, |V_H|\}$ and $E_H = \{e_1, e_2, \ldots, e_{|E_H|}\}$. We denote the complement of the zone H by $\bar{H} = G\backslash H$. If X, Y are two subgraphs of G, $\mathbf{A}_{X|Y}$ and $\mathbf{A}_{V_X|V_Y}$ both denote the submatrix of the admittance matrix of G with rows from V_X and columns from V_Y. For instance, $\mathbf{A}$ can be written in any of the following

[2]The uniqueness is in the values of p_{uv}s rather than θ_us (shifting all θ_us by equal amounts does not violate (2)).

[3]When $r_{uv} = 1\ \forall\{u, v\} \in E$, the admittance matrix $\mathbf{A}$ is the *Laplacian matrix* of the graph.

Table 1: Summary of notation.

Notation	Description
$G = (V, E)$	The graph representing the power grid
$\mathbf{A}$	Admittance matrix of G
$\vec{\theta}$	Vector of the phase angles of the nodes in G
H	A subgraph of G representing the attacked zone
F	Set of failed edges due to an attack
$\mathbf{D}$	Incidence matrix of G
$\bigcirc'$	The value of $\bigcirc$ after an attack
$\bar{\bigcirc}$	The complement of $\bigcirc$
$\bigcirc^*$	The dual of $\bigcirc$

forms,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{H|H} & \mathbf{A}_{H|\bar{H}} \\ \mathbf{A}_{\bar{H}|H} & \mathbf{A}_{\bar{H}|\bar{H}} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{A}_{G|H} & \mathbf{A}_{G|\bar{H}} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{A}_{H|G} \\ \mathbf{A}_{\bar{H}|G} \end{bmatrix}.$$

We use the very same notation for the vectors. For instance $\vec{\theta}_H$ and $\vec{\theta}_{\bar{H}}$ are the vectors of phase angle of the nodes in H and $\bar{H}$, respectively. We use the prime symbol ($'$) to denote the values after an attack. For instance, G', $\mathbf{A}'$, and $\vec{\theta}'$ are used to represent the graph, the admittance matrix of the graph, and the phase angles of the nodes after an attack.

3.4 Graph Theoretical Terms

In this paper, we use several graph theoretical terms and theorems, most of which are reviewed in Appendix A. We briefly review some of the important definitions in this subsection. Most of the definitions are borrowed from [16].

Subgraphs, Cuts, and Cycles: Let X and Y be subsets of the nodes of a graph G. $G[X]$ denotes the subgraph of G *induced* by X. We denote by $E[X, Y]$ the set of edges of G with one end in X and the other end in Y. We denote the complement of a set X by $\bar{X} = V \backslash X$. The *coboundary* of X is the set $E[X, \bar{X}]$ and is denoted by $\partial(X)$. $\partial(v)$ denotes the coboundary of $X = \{v\}$. $G[X, \bar{X}]$ denotes the subgraph of G induced by the edges from $E[X, \bar{X}]$. We say that $Q \subseteq E$ is *G-separable*, if there are pairwise edge-disjoint cycles $C_q (q \in Q)$, such that $\forall q \in Q, q \in C_q$ [39].

Planar Graphs: A graph G is *planar*, if it can be drawn in the plane so that its edges intersect only at their ends. A planar graph G partitions the rest of the plane into a number of edgewise-connected open sets called the *faces* of G.

Given a planar graph G, its dual graph G^* is defined as follows. Corresponding to each face c of G there is a node c^* of G^*, and corresponding to each edge e of G there is an edge e^* of G^*. Two nodes c_1^* and c_2^* are joined by the edge e^* in G^*, if and only if their corresponding faces c_1 and c_2 are separated by the edge e in G. It is easy to see that the dual G^* of a planar graph G is itself a planar graph [16].

Incidence Matrix: Suppose we assign an arbitrary orientation to the edges of G. We denote the set of oriented edges by $\mathcal{E} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_m\}$. The (node-edge) *incidence matrix* of G is denoted by $\mathbf{D} \in \{-1, 0, 1\}^{|V| \times |E|}$ and defined as follows,

$$d_{ij} = \begin{cases} 0 & \text{if } \epsilon_j \text{ is not incident to node } i, \\ 1 & \text{if } \epsilon_j \text{ is coming out of node } i, \\ -1 & \text{if } \epsilon_j \text{ is going into node } i. \end{cases}$$

When we use the incidence matrix, we assume an arbitrary orientation for the edges unless we mention an specific orientation. $\mathbf{D}_H \in \{-1, 0, 1\}^{|V_H| \times |E_H|}$ is the submatrix of $\mathbf{D}$ with rows from V_H and columns from E_H.

4. ATTACK ANALYSIS

In this section, we study the effects of an attack and provide analytical methods for recovering the phase angles and detecting failed edges in the attacked zone H. We find conditions on the structural properties of a zone and constraints on the failed edges for which these methods successfully recover the phase angles and detect the failed edges. These conditions depend on the connections between V_H and $\bar{V}_H$ as well as the inner connections of the nodes in H. Therefore, we refer to them as *external* and *internal* conditions on H, respectively. Finally, based on the methods, we present the Post-Attack Recovery and Detection (PARD) Algorithm. Table 2 summarizes the results regarding the resilience of a zone based on its internal and external conditions, and the constraints on the set of failed edges F.

In this section, when we describe our methods, we assume that there are no edges $\{i, j\} \in E_H$ for which $\theta'_i = \theta'_j$ (we refer to these edges as *null-edges*). Following (2), a null-edge does not carry any flow. Thus, we cannot detect the status of those edges since they cannot be distinguished from failed edges. However, we can detect the null-edges and treat them separately (we consider this in the PARD Algorithm).

4.1 Recovery of Phase Angles

In this subsection, we introduce a method to recover the phase angles of the nodes in an attacked zone H. We provide sufficient conditions on $G[V_H, \bar{V}_H]$ such that the method recovers the phase angles of the nodes in V_H successfully. As we mentioned, since these conditions depend only on the connections between V_H and $\bar{V}_H$, we refer to them as the *external conditions* on H.

The following lemma is the first step towards designing the method for recovering the phase angles and for detecting the failed edges (see Subsection 4.2).

Lemma 1. $\text{supp}(\mathbf{A}(\vec{\theta} - \vec{\theta}')) \subseteq V_H$.

Proof. Suppose $F = \{e_{i_1}, e_{i_2}, \ldots, e_{i_k}\} \subseteq E_H$ are the edges that are disconnected from the grid after the attack on the zone H. Define the column vectors $\vec{x_1}, \vec{x_2} \ldots \vec{x_k} \in \{-1, 0, 1\}^n$ associated with the failed edges as follows. If $e_{i_j} = \{s_j, t_j\}$ then $\vec{x_j}$ is 1 in its s_j^{th} entry, -1 in its t_j^{th} entry, and 0 everywhere else. It is easy to see that $\mathbf{A}'$ is related to $\mathbf{A}$ as $\mathbf{A}' = \mathbf{A} - \sum_{j=1}^{k} a_{s_j t_j} \vec{x_j} \vec{x_j}^t$. Since the graph G does not get disconnected after an attack, the flow equations in G' are $\mathbf{A}' \vec{\theta}' = \vec{p}$. On the other hand, $\mathbf{A}\vec{\theta} = \vec{p}$, therefore $\mathbf{A}\vec{\theta} - \mathbf{A}'\vec{\theta}' = 0$. Thus,

$$0 = \mathbf{A}\vec{\theta} - \mathbf{A}'\vec{\theta}' = \mathbf{A}\vec{\theta} - \mathbf{A}\vec{\theta}' + \sum_{j=1}^{k} a_{s_j t_j} \vec{x_j} \vec{x_j}^t \vec{\theta}'$$

$$\Rightarrow \mathbf{A}(\vec{\theta} - \vec{\theta}') = -\sum_{j=1}^{k} a_{s_j t_j} \vec{x_j} \vec{x_j}^t \vec{\theta}'$$

$$\Rightarrow \text{supp}(\mathbf{A}(\vec{\theta} - \vec{\theta}')) \subseteq \bigcup_{i=1}^{k} \text{supp}(\vec{x_j} \vec{x_j}^t \vec{\theta}') \subseteq \bigcup_{i=1}^{k} \{s_j, t_j\}$$

$$\Rightarrow \text{supp}(\mathbf{A}(\vec{\theta} - \vec{\theta}')) \subseteq \bigcup_{i=1}^{k} \{s_j, t_j\} \subseteq V_H. \quad \square$$

One of the immediate results of Lemma 1 is the following corollary. This corollary gives a true statement about $\vec{\theta}'$ (recall that $\vec{\theta}'$ is partly unknown). It states that $\vec{\theta}'$ is in the solution space of the matrix equation (4).

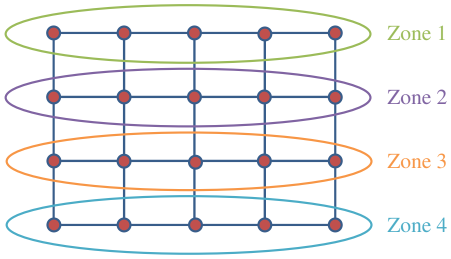


Figure 3: An example of a graph and set of zones such that each zone is both well-supported and acyclic.

Corollary 1. $\mathbf{A}_{\bar{H}|G}(\vec{\theta} - \vec{\theta}') = 0.$ (4)

We find sufficient conditions such that the solution $\vec{\theta}'_H$ to (4) is unique (given $\vec{\theta}$ and $\vec{\theta}'_{\bar{H}}$), and consequently $\vec{\theta}'_H$ can be recovered after any attack on H. We first define a *well-supported* zone.

Definition 1. A zone H is called *well-supported*, if $\vec{\theta}'_H$ can be recovered after any attack on H.

Using Corollary 1, the following theorem gives sufficient condition for a zone H to be *well-supported*.

Theorem 1. *A zone H is* well-supported*, if $\mathbf{A}_{\bar{H}|H}$ has linearly independent columns.*

Proof. From Corollary 1 we know that $\mathbf{A}_{\bar{H}|G}(\vec{\theta} - \vec{\theta}') = 0$, therefore $\mathbf{A}_{\bar{H}|H}\vec{\theta}'_H = \mathbf{A}_{\bar{H}|\bar{H}}(\vec{\theta}_{\bar{H}} - \vec{\theta}'_{\bar{H}}) + \mathbf{A}_{\bar{H}|H}\vec{\theta}_H$. The only unknown in this equation is $\vec{\theta}'_H$. Now since $\mathbf{A}_{\bar{H}|H}$ has linearly independent columns, this equation has a unique solution $\vec{\theta}'_H$ which can be computed in polynomial time. Thus, $\vec{\theta}'_H$ can be recovered in this case and zone H is well-supported. $\square$

It can be seen that the sufficient condition in Theorem 1 depends on the reactance values. However, the following corollary relaxes the condition in Theorem 1. It shows that if $G[V_H, \bar{V}_H]$ has a matching that covers V_H, then for almost any reactance values for the edges in $E[V_H, \bar{V}_H]$, H is *well-supported*. The idea is that the set of reactance values for the edges in $E[V_H, \bar{V}_H]$ for which $\mathbf{A}_{\bar{H}|H}$ does not have linearly independent columns is a measure zero set in the real space [38].

Corollary 2. *If there is a matching in $G[V_H, \bar{V}_H]$ that covers V_H, then H is* well-supported *almost surely.*[4]

In reality, since the reactance values are derived by the physical properties of the lines, we expect that these values are relatively random around a mean value. Thus, following Corollary 2, the existence of a matching that covers every node in V_H is enough for a zone to be well-supported (see Fig. 3 for an example of a graph in which every node in a zone is covered by a matching). Hence, in the following sections we consider the existence of a matching as a sufficient external condition on H to be well-supported.

4.2 Detecting Failed Edges

In this subsection, we assume that after an attack, the phase angles are recovered using the method in Subsection 4.1 (i.e., by solving (4)). We introduce methods to detect the failed edges using $\vec{\theta}'$. We provide sufficient conditions on H such that these methods detect the failed edges successfully. As we mentioned, since these conditions depend only on the connections between the nodes in H, we refer to them as *internal conditions* on H.

The following Lemma is the foundation for our approach to find the failed edges. It limits the set of failed edges to the solution space of the matrix equation (5). It can be considered as the complement of Corollary 1.

Lemma 2. *There exists a vector $\vec{x} \in \mathbb{R}^{|E_H|}$ such that* $\text{supp}(\vec{x}) = \{i | e_i \in F\}$ *and*

$$\mathbf{D}_H\vec{x} = \mathbf{A}_{H|G}(\vec{\theta} - \vec{\theta}'). \tag{5}$$

Proof. We use the notation that we used in proof of Lemma 1. Recall from the proof of Lemma 1 that $\mathbf{A}(\vec{\theta} - \vec{\theta}') = -\sum_{j=1}^{k} a_{s_j t_j}\vec{x_j}\vec{x_j}^t\vec{\theta}'$. It is easy to see that if $\vec{d_1}, \vec{d_2}, \ldots, \vec{d_m}$ are the columns of the incidence matrix $\mathbf{D}$, then $\forall j (1 \leqslant j \leqslant k)$, there exists $b_j \in \mathbb{R}$ such that $b_j\vec{d}_{i_j} = -a_{s_j t_j}\vec{x_j}\vec{x_j}^t\vec{\theta}'$. Therefore, $\mathbf{A}(\vec{\theta} - \vec{\theta}') = \sum_{j=1}^{k} b_j\vec{d}_{i_j}$. Thus, if we define $\vec{y} \in \mathbb{R}^m$ such that $\forall e_{i_j} \in F, y_{i_j} = b_j$ and 0 elsewhere, then $\mathbf{A}(\vec{\theta} - \vec{\theta}') = \mathbf{D}\vec{y}$ and $\text{supp}(\vec{y}) \subseteq \{i_1, i_2, \ldots, i_k\}$. However, from the Corollary 1 we know that $\mathbf{A}_{\bar{H}|G}(\vec{\theta} - \vec{\theta}') = 0$. Moreover, since $F \subseteq E_H$, $\vec{y}_{\bar{H}} = 0$. Thus, we can restrict the equation only to the components of the zone H, which means that $\mathbf{A}_{H|G}(\vec{\theta} - \vec{\theta}') = \mathbf{D}_H\vec{y}_H$. Now it is easy to see that since we assumed that no null-edges are in F, all the b_is are nonzero and $\text{supp}(\vec{y}_H) = \{i_1, i_2, \ldots, i_k\}$. Therefore, $\vec{x} = \vec{y}_H$ is a solution to (5) and $\text{supp}(\vec{x}) = \{i | e_i \in F\}$. $\square$

Lemma 2 provides important information regarding the failed edges. It states that there exists a solution $\vec{x}$ to (5) such that $\text{supp}(\vec{x})$ reveals the set of failed edges. However, the solution to (5) may not be unique. The Lemma below provides a necessary and sufficient condition on H such that the solution to (5) is unique.

Lemma 3. *The solution to (5) is unique and* $\text{supp}(\vec{x}) = \{i | e_i \in F\}$*, if and only if H is acyclic.*

According to Lemma 3 the set of failed edges for any attack can be detected, if and only if H is acyclic. Fig. 3 shows an example of a graph and set of zones such that each zone is both well-supported and acyclic (case I in Table 2).

Although Lemma 3 requires H to be an acyclic graph in order for the solution of (5) to be unique, by setting some constraints on the failed edges F, we provide a method to detect the failed edges in broader class of graphs. The underlying idea is that the set of failed edges is expected to be relatively sparse compared to the overall set of edges within a zone. Thus, we are interested in the solutions of (5) that are relatively sparse. The l_0-norm should be used to capture the sparseness of a vector. However, since minimizing l_0-norm is a combinatorial problem in general cases, we prefer to use l_1-norm which is known to be a good approximation of the l_0-norm. Thus, we consider the following minimization problem,

$$\min \|\vec{x}\|_1 \text{ s.t. } \mathbf{D}_H\vec{x} = \mathbf{A}_{H|G}(\vec{\theta} - \vec{\theta}'). \tag{6}$$

Notice that (6) is still linear and can be solved using Linear Programming. Moreover, when the solution to (6) also appears to be sparse, which is usually the case in the considered scenario, there are very fast algorithms to solve it [23].

The Lemma below states that by solving (6), the failed edges can be detected in more cases than by solving (5). The idea that we use in proof of Lemma 4 is the core idea in proofs of Theorems 2 and 3, as well. Namely, the null

[4] In probability theory, one says that an event happens almost surely, if it happens with probability one.

Table 2: Summary of the results in Section 4. The external/internal conditions on the structural properties of a zone H such that after an attack with certain constraints, the phase angles can be recovered and the failed edges can be detected by solving (8). Matching and partial matching refer to matchings in $G[V_H, \bar{V}_H]$ that cover V_H and $V_H \backslash (V_H^{\text{in}} \cup V_H^{\text{out}})$, respectively.

Case	External conditions	Internal conditions	Attack constraints	Resilience	Results
I	Matching	Acyclic	None	attack-resilient	Corollary 2/Lemma 3
II	Matching	Planar	$\forall$ cycle C, $\lvert C \cap F\rvert < \lvert C \backslash F\rvert$ F^* is H^*-separable	weakly-attack-resilient	Corollary 2/Theorem 2
III	Partial matching	Acyclic	$\forall v \in V_H^{\text{in}}$, $\lvert\partial(v) \cap F\rvert < \lvert\partial(v)\backslash F\rvert$	weakly-attack-resilient	Lemmas 3,6/Corollary 5
IV	Partial matching	Planar No cycle contains an inner-connected-node	$\forall$ cycle C, $\lvert C \cap F\rvert < \lvert C \backslash F\rvert$ $\forall v \in V_H^{\text{in}}$, $\lvert\partial(v) \cap F\rvert < \lvert\partial(v)\backslash F\rvert$ F^* is H^*-separable	weakly-attack-resilient	Theorem 3/Corollary 5

space of D_H is in one-to-one correspondence with the cycle space of the graph H. Therefore, there are graph theoretical interpretations to the solution space of (5). Hence, by using tools from graph theory and linear algebra, we find the solution to (5) with the minimum l_1-norm.

Lemma 4. *If H is a cycle and $|E_H \cap F| < |E_H \backslash F|$, the solution to (6) is unique and* $\text{supp}(\vec{x}) = \{i | e_i \in F\}$.

Proof. Here without loss of generality, we assume that $\mathbf{D}_H$ is the incidence matrix of H when edges of H has been oriented clockwise. Since H is connected, it is known that $\text{rank}(\mathbf{D}_H) = |V_H| - 1$ [9, Theorem 2.2]. Therefore, $\dim(\text{Null}(\mathbf{D}_H)) = 1$. Suppose $\vec{e} \in \mathbb{R}^{|E_H|}$ is the all one vector. It is easy to see that $\mathbf{D}_H \vec{e} = 0$. Since $\dim(\text{Null}(\mathbf{D}_H)) = 1$, $\vec{e}$ is the basis for the null space of $\mathbf{D}$. Suppose $\vec{x}$ is a solution to (5) such that $\text{supp}(\vec{x}) = \{i | e_i \in F\}$ (from Lemma 2 we know that such a solution exists). To prove that $\vec{x}$ is the unique solution for (6), we only need to prove that $\forall c \in \mathbb{R}\backslash\{0\}$, $\|\vec{x}\|_1 < \|\vec{x} - c\vec{e}\|_1$. Without loss of generality we can assume that $x_1, x_2, \ldots, x_k$ are the nonzero elements of $\vec{x}$, in which $k = |F|$. From the assumption we know that $|E_H \cap F| < |E_H \backslash F|$, therefore $k < |E_H|/2$. Hence, we have

$$\begin{aligned}
\|\vec{x} - c\vec{e}\|_1 &= \sum_{i=1}^{k} |x_i - c| + (|E_H| - k)|c| \\
&= \sum_{i=1}^{k} (|x_i - c| + |c|) + (|E_H| - 2k)|c| \\
&\geqslant \sum_{i=1}^{k} |x_i| + (|E_H| - 2k)|c| > \sum_{i=1}^{k} |x_i| = \|\vec{x}\|_1.
\end{aligned}$$

Thus, the solution to (6) is unique. □

Corollary 3. *If all the cycles in H are edge-disjoint and for any cycle C in H, $|C \cap F| < |C \backslash F|$, then the solution to (6) is unique and* $\text{supp}(\vec{x}) = \{i | e_i \in F\}$.

The following Theorem extends the idea in the proof of Lemma 4 and provides sufficient conditions for failed edges in a planar graph H to be detected by solving (6) (recall from subsection 3.4 that H^* is the dual of the planar graph H and F^* is the dual of the set of failed edges).

Theorem 2. *In a planar graph H, the solution to (6) is unique and* $\text{supp}(\vec{x}) = \{i | e_i \in F\}$, *if the following conditions hold:* (i) *for any cycle C in H, $|C \cap F| < |C \backslash F|$, and* (ii) *$F^*$ is H^*-separable.*

Fig. 4 shows an example of a zone H for which the set of failed edges can be detected by solving (6) based on Theorem 2 (case II in Table 2).

The Corollary below states that in planar bipartite graphs, condition (ii) in Theorem 2 immediately holds, if condition (i) holds.

Corollary 4. *In a planar bipartite graph H, the solution to (6) is unique and* $\text{supp}(\vec{x}) = \{i | e_i \in F\}$, *if for any cycle C in H, $|C \cap F| < |C \backslash F|$.*

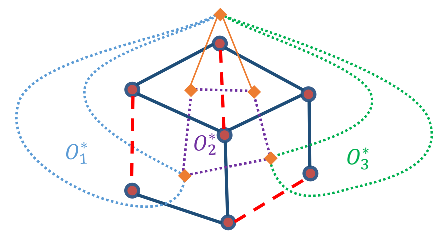


Figure 4: An example of a zone H and a set of failed edges (shown by red dashed lines) that can be detected by solving (6) based on Theorem 2. The diamond orange nodes are the nodes of the dual graph H^*. As can be seen, the dual of the failed edges can be covered by three edge disjoint cycles O_1^*, O_2^*, O_3^* (shown by dotted lines) in H^*. Thus, as Theorem 2 requires, F^* is H^*-separable.

Theorem 2 and Corollary 4 are important since power grids are usually considered to be planar. For instance, lattice graphs are planar bipartite.

4.3 Simultaneous Phase Angles Recovery and Failed Edges Detection

In Subsection 4.1 we showed that the phase angles of the zone H are recoverable, if there is a matching in $G[V_H, \bar{V}_H]$ that covers V_H. However, in reality, this condition might be very difficult and costly to maintain (i.e., it may require to increase the number of zones). Therefore, in this subsection, using similar ideas as in subsection 4.2, we relax the external conditions on H.

The key idea which is summarized in the following Lemma, is to combine Corollary 1 and Lemma 2.

Lemma 5. *There exist vectors $\vec{x} \in \mathbb{R}^{|E_H|}$ and $\vec{\delta}_H \in \mathbb{R}^{|V_H|}$ such that* $\text{supp}(\vec{x}) = \{i | e_i \in F\}$, $\vec{\delta}_H = \vec{\theta}_H - \vec{\theta}'_H$, *and*

$$\begin{aligned}
\mathbf{D}_H \vec{x} &= \mathbf{A}_{H|H} \vec{\delta}_H + \mathbf{A}_{H|\bar{H}} \vec{\delta}_{\bar{H}} \\
\mathbf{A}_{\bar{H}|H} \vec{\delta}_H &+ \mathbf{A}_{\bar{H}|\bar{H}} \vec{\delta}_{\bar{H}} = 0
\end{aligned} \tag{7}$$

where $\vec{\delta}_{\bar{H}} = \vec{\theta}_{\bar{H}} - \vec{\theta}'_{\bar{H}}$ and is known.

From Subsections 4.1 and 4.2 we know that the solution to (7) is unique, if and only if H is acyclic and $\mathbf{A}_{\bar{H}|H}$ has linearly independent columns. Therefore, to deal with cases for which $\mathbf{A}_{\bar{H}|H}$ does not have linearly independent columns, we consider a similar optimization problem as in (6) but with more constraints. For this reason, as we mentioned in Subsection 4.2, since the set of failed edges is expected to be relatively sparse compared to the overall set of edges, we consider the following optimization problem,

$$\begin{aligned}
&\min \|\vec{x}\|_1 \text{ s.t.} \\
&\mathbf{D}_H \vec{x} = \mathbf{A}_{H|H} \vec{\delta}_H + \mathbf{A}_{H|\bar{H}} \vec{\delta}_{\bar{H}} \\
&\mathbf{A}_{\bar{H}|H} \vec{\delta}_H + \mathbf{A}_{\bar{H}|\bar{H}} \vec{\delta}_{\bar{H}} = 0.
\end{aligned} \tag{8}$$

The following Lemma states that if there is an independent set of nodes in H with no neighbors in $\bar{H}$, then under some conditions on F, we can recover F and $\vec{\theta}'_H$ by solving (8) even when $\mathbf{A}_{\bar{H}|H}$ does not have linearly independent columns (case III in Table 2). First, we define *inner-connected* nodes.

Definition 2. A node $v \in V_H$ is called *H-inner-connected* if $N(v) \subseteq V_H$. It is called *H-outer-connected* if $N(v) \subseteq V_{\bar{H}}$. We denote the set of *$H$-inner-connected* and *$H$-outer-connected* nodes by V_H^{in} and V_H^{out}, respectively.

LEMMA 6. *Suppose H-inner-connected nodes form an independent set. If H is acyclic,* $\text{rank}(\boldsymbol{A}_{\bar{H}|H}) = |V_H| - |V_H^{in}|$, *and $\forall v \in V_H^{in}, |\partial(v) \cap F| < |\partial(v)\backslash F|$, then the solution $\vec{x}, \vec{\delta}$ to (8) is unique. Moreover,* $\text{supp}(\vec{x}) = \{i|e_i \in F\}$ *and $\vec{\delta}_H = \vec{\theta}_H - \vec{\theta}'_H$.*

PROOF. The idea of the proof is very similar to the proof of Lemma 4. Suppose $\vec{x}, \vec{\delta}_H$ is the solution to (7) such that $\text{supp}(\vec{x}) = \{i|e_i \in F\}$ and $\vec{\delta}_H = \vec{\theta}_H - \vec{\theta}'_H$. From Lemma 5 we know that such a solution exists. We show that this solution is the unique solution to (8) in this setting.

Without loss of generality in addition to assuming $V_H = \{1, 2, 3, \ldots, |V_H|\}$ and $E_H = \{e_1, e_2, \ldots, e_{|E_H|}\}$, we can assume the labeling of the nodes in G is such that $V_H^{\text{in}} = \{1, 2, \ldots, t\}$ is the set of H-inner-connected nodes. Suppose $\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_t \in \mathbb{R}^{|V_H|}$ are the coordinate vectors, in other words $\vec{\alpha}_i$ is 1 at its i^{th} entry and 0 everywhere else. It is easy to see that $\forall i \in V_H^{\text{in}} : A_{\bar{H}|H}\vec{\alpha}_i = 0$. On the other hand, since $\text{rank}(\mathbf{A}_{\bar{H}|H}) = |V_H| - t$ and $\vec{\alpha}_i$s are linearly independent, $\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_t$ form a basis for $\text{Null}(\mathbf{A}_{\bar{H}|H})$.

Assume $\mathbf{D}_H$ is the incidence matrix of H when its edges are oriented such that for each $i \in V_H^{\text{in}}$, the edges are coming out of i. Now suppose $\vec{z}$ is another solution to (8), it is easy to see that $\mathbf{D}_H(\vec{z} - \vec{x}) = \mathbf{A}_{H|H}\vec{\alpha}$ for a vector $\vec{\alpha} \in \text{Null}(\mathbf{A}_{\bar{H}|H})$. Since $\vec{\alpha} \in \text{Null}(\mathbf{A}_{\bar{H}|H})$, there are unique coefficients $c_1, c_2, \ldots, c_t \in \mathbb{R}$ such that $\vec{\alpha} = c_1\vec{\alpha}_1 + c_2\vec{\alpha}_2 + \cdots + c_t\vec{\alpha}_t$. Thus,

$$\begin{aligned} \mathbf{D}_H(\vec{z} - \vec{x}) &= \mathbf{A}_{H|H}\vec{\alpha} = \mathbf{A}_{H|H}(c_1\vec{\alpha}_1 + c_2\vec{\alpha}_2 + \cdots + c_t\vec{\alpha}_t) \\ &= c_1\mathbf{A}_{H|H}\vec{\alpha}_1 + c_2\mathbf{A}_{H|H}\vec{\alpha}_2 + \cdots + c_t\mathbf{A}_{H|H}\vec{\alpha}_t. \end{aligned}$$

Suppose $\vec{d}_j$ is the column associated with edge e_j in D_H. Notice that for each $i \in V_H^{\text{in}}$, $\partial(i) \subseteq E_H$. Therefore, $\forall i \in V_H^{\text{in}}$ and $\forall e_j \in \partial(i)$, $\vec{d}_j$ is a column of $\mathbf{D}_H$. It is easy to see that for any $i \in V_H^{\text{in}}$, $\sum_{j:e_j \in \partial(i)} \vec{d}_j = \mathbf{A}_{H|H}\vec{\alpha}_i$. If for any $i \in V_H^{\text{in}}$ we define vector $\vec{b}_i \in \{0, 1\}^{|E_H|}$ as follows,

$$b_{ij} := \begin{cases} 1 & \text{if } e_j \in \partial(i) \\ 0 & \text{otherwise,} \end{cases}$$

then $\mathbf{D}_H\vec{b}_i = \mathbf{A}_{H|H}\vec{\alpha}_i$ for any $i \in V_H^{\text{in}}$. Thus,

$$\begin{aligned} &\mathbf{D}_H(c_1\vec{b}_1 + \cdots + c_t\vec{b}_t) = c_1\mathbf{A}_{H|H}\vec{\alpha}_1 + \cdots + c_t\mathbf{A}_{H|H}\vec{\alpha}_t \\ &\quad \Rightarrow \mathbf{D}_H(\vec{z} - \vec{x}) = \mathbf{D}_H(c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_t\vec{b}_t). \end{aligned}$$

Now since H is acyclic, $\mathbf{D}_H$ has linearly independent columns. Thus, from the equation above we can conclude that,

$$\begin{aligned} &\vec{z} - \vec{x} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_t\vec{b}_t \\ \Rightarrow\ &\vec{z} = \vec{x} + c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_t\vec{b}_t. \end{aligned}$$

Using equation above, we show that $\|\vec{z}\|_1 > \|\vec{x}\|_1$ unless $c_1 = c_2 = \cdots = c_t = 0$. First, notice that since V_H^{in} is an independent set, $\forall i \neq j \in V_H^{\text{in}}$, $\partial(i) \cap \partial(j) = \varnothing$. Suppose $\forall i \in V_H^{\text{in}}, |\partial(i) \cap F| = k_i$, we have

$$\begin{aligned} \|\vec{z}\|_1 &= \|\vec{x} + c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_t\vec{b}_t\|_1 \\ &= \sum_{i \in V_H^{\text{in}}}\Big((|\partial(i)| - k_i)|c_i| + \sum_{j \in F \cap \partial(i)} |x_j + c_i|\Big) + \sum_{i \in F\backslash\partial(V_H^{\text{in}})} |x_i| \\ &= \sum_{i \in V_H^{\text{in}}}\Big((|\partial(i)| - 2k_i)|c_i| + \sum_{j \in F \cap \partial(i)}(|x_j + c_i| + |c_i|)\Big) + \sum_{i \in F\backslash\partial(V_H^{\text{in}})} |x_i| \\ &\leqslant \sum_{i \in V_H^{\text{in}}}\Big((|\partial(i)| - 2k_i)|c_i| + \sum_{j \in F \cap \partial(i)} |x_j|\Big) + \sum_{i \in F\backslash\partial(V_H^{\text{in}})} |x_i| \\ &\leqslant \sum_{i \in V_H^{\text{in}}}\big((|\partial(i)| - 2k_i)|c_i|\big) + \sum_{i \in V_H^{\text{in}}}\ \sum_{j \in F \cap \partial(i)} |x_j| + \sum_{i \in F\backslash\partial(V_H^{\text{in}})} |x_i| \\ &\leqslant \sum_{i \in V_H^{\text{in}}}\big((|\partial(i)| - 2k_i)|c_i|\big) + \|\vec{x}\|_1. \end{aligned}$$

Now, since from the assumptions $\forall i \in V_H^{\text{in}}, k_i < |\partial(i)|/2$, it is easy to see that $\sum_{i \in V_H^{\text{in}}}((\partial(i) - 2k_i)|c_i|) + \|\vec{x}\|_1 < \|\vec{x}\|_1$, unless $c_1 = c_2 = \cdots = c_t = 0$. Since $\vec{z}$ is a solution to (8), we should have $c_1 = c_2 = \cdots = c_t = 0$, and $\vec{z} = \vec{x}$. Thus, $\vec{x}$ is the unique solution to (8) and $\text{supp}(\vec{x}) = \{i|e_i \in F\}$. From the proof, it is easy to see that $\vec{\delta}_H$ is also unique. ☐

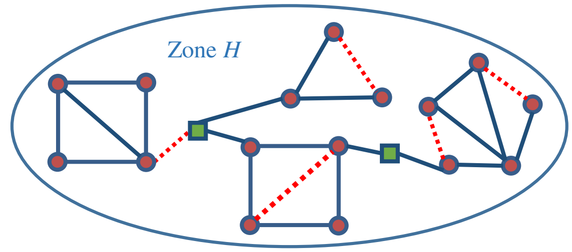


Figure 5: An example of a zone H and an attack such that the phase angles can be recovered and the failed edges can be detected by solving (8) based on Theorem 3. The squared green nodes are the H-inner-connected nodes. The failed edges are shown by red dashed edges.

To generalize Lemma 6, first let us consider cases in which H contains H-outer-connected nodes. The Lemma below shows that the value of δ for these nodes is unique.

LEMMA 7. *If v is H-outer-connected and $\vec{\delta}_H$ is a solution to (7), then δ_v is unique and equal to $\delta_v = 1/d(v) \sum_{u \in N(v)} \delta_u$, where $d(v)$ is the degree of node v.*

In the following theorem, we generalize Lemma 6. This theorem combines Lemma 6 and Theorem 2, and provides a broader class of graphs in which solving (8) recovers phase angles and detects the failed edges after an attack.

THEOREM 3. *In a planar graph H, the solution $\vec{x}, \vec{\delta}_H$ to (8) is unique with* $\text{supp}(\vec{x}) = \{i|e_i \in F\}$ *and $\vec{\delta}_H = \vec{\theta}_H - \vec{\theta}'_H$, if the following conditions hold:* (i) $\forall v \in V_H^{\text{in}}$, $|\partial(v) \cap F| < |\partial(v)\backslash F|$, (ii) *for any cycle C in H, $|C \cap F| < |C\backslash F|$,* (iii) *$F^*$ is H^*-separable,* (iv) *in $\boldsymbol{A}_{\bar{H}|H}$, columns associated with nodes that are neither H-inner-connected nor H-outer-connected are linearly independent,* (v) *no cycle in H contains a H-inner-connected node, and* (vi) *H-inner-connected nodes form an independent set.*

Note that when H is well-supported, there are no H-inner-connected or H-outer-connected nodes. Thus, conditions (i), (iv), (v), and (vi) immediately hold and Theorem 3 reduces to Theorem 2.

Algorithm 1 - Post-Attack Recovery & Detection (PARD)

Input: A connected graph G, phase angles before the attack $\vec{\theta}$, and partial phase angles after the attack $\vec{\theta}'_{\bar{H}}$.

1: Detect the attacked zone H by checking for missing data.
2: Compute $\vec{x}, \vec{\delta}_H$ the solution to (8) by Linear Programming.
3: Compute $\vec{\theta}'_H = \vec{\theta}_H - \vec{\delta}_H$.
4: Compute $F = \{e_i | i \in \mathrm{supp}(\vec{x})\}$.
5: Detect the set of null-edges that appear after the attack as $N = \{\{i,j\} \in E_H | \theta'_i = \theta'_j\}$.
6: **return** N, F, $\vec{\theta}'_H$.

Fig. 5 shows an example of a zone H and an attack such that the phase angles can be recovered and the failed edges can be detected by solving (8) using Theorem 3 (case IV in Table 2). As it can be seen, this theorem covers a broad set of graphs and attacks for which we can recover the phase angles and detect the failed edges. Notice that here, with similar argument as in Corollary 2 we can replace condition (iv) in Theorem 3 with a simpler matching condition as follows.

Corollary 5. *If there is a matching in* $G[V_H, \bar{V}_H]$ *that covers* $V_H \backslash (V_H^{\mathrm{in}} \cup V_H^{\mathrm{out}})$*, then condition* (iv) *in Theorem 3 holds almost surely.*

To conclude, we define the *attack-resilient* and *weakly-attack-resilient* notions to summarize the resilience of a zone to joint cyber and physical attacks.

Definition 3. A zone H is called *attack-resilient*, if it is both well-supported and acyclic.

Definition 4. A zone H is called *weakly-attack-resilient*, if $\vec{\theta}'_H$ and F can be uniquely found after a constrained attack on the zone H by solving (8).

It is easy to see that an attack-resilient zone is also weakly-attack-resilient.

4.4 Post-Attack Recovery and Detection Algorithm

In this subsection, we present the Post-Attack Recovery and Detection (PARD) Algorithm for recovering the phase angles and detecting the failed edges after an attack on a zone H. Based on the results provided in previous subsections, if zone H is weakly-attack-resilient, the PARD Algorithm will recover the phase angles and detect the failed edges after a constrained attack.

Notice that if there are some failed edges but no data is missing, then from the data that is gathered by the PDCs from the PMUs, all the information regarding the status of the lines and phase angles is available and there is no need for the algorithm. Thus, as the first step, the PARD Algorithm detects the attacked zone H by checking the missing data (line 1). Then, it solves (8) by Linear Programming to obtain $\vec{x}, \vec{\delta}_H$. If H is weakly-attack-resilient, from the results in previous subsections, we know that $\vec{x}, \vec{\delta}_H$ are unique, $\vec{\theta}'_H = \vec{\theta}_H - \vec{\delta}_H$ (line 3), and $F = \{e_i | i \in \mathrm{supp}(\vec{x})\}$ (line 4). Finally, using $\vec{\theta}'$ computed in previous line, the PARD Algorithm detects the set of null-edges N (line 5), and returns N, F, and $\vec{\theta}'_H$.

5. ZONE SELECTION ALGORITHM

In this section we use the results from Section 4 to provide an algorithm for partitioning the power grid into the minimum number of attack-resilient zones. From Lemma 3 and

Algorithm 2 - 3-Acyclic Partition of Planar (3APP)

Input: A non-empty planar graph G.

Find a node $v \in V$ such that $\deg(v) \leqslant 5$.
if $G \backslash v = \varnothing$ **then** set $Q_1 = Q_2 = Q_3 = \varnothing$.
else Find 3-partition of $G \backslash v$ using 3APP Algorithm as Q_1, Q_2, Q_3.
Add v to the partition that $|N(v) \cap Q_i|$ is minimum.
return Q_1, Q_2, Q_3.

Corollary 2, for a zone H to be attack-resilient, it is sufficient that H is acyclic and there is a matching in $G[V_H, \bar{V_H}]$ that covers every node in V_H. Fig. 3 shows an example of a partitioning such that each zone is attack-resilient. Thus, we define a *matched-forest* partition of a graph G as follows.

Definition 5. A *matched-forest* partition of a graph G into $H_1, H_2, \ldots, H_k$ is a partition such that for any i, H_i is acyclic and $G[V_{H_i}, \bar{V}_{H_i}]$ has a matching that covers V_{H_i}.

The problem of finding a matched-forest partition of G is closely related to two previously known problems of *vertex arboricity* and *k-matching cover* of a graph. The vertex arboricity $a(G)$ of a graph G is the minimum number of subsets into which the nodes of G can be partitioned so that each subset induces an acyclic graph. It is known that determining $a(G)$ is NP-hard [26, p.193].

A k-matching cover of a graph G is a union of k matchings of G which covers V. The matching cover number of G, denoted by $mc(G)$, is the minimum number k such that G has a k-matching cover. An optimal matching cover of a graph on n nodes can be found in $O(n^3)$ time [45].

Using these results, we study the time complexity of the *minimum matched-forest* partition problem.[5] The following Lemma shows that it is hard to find the minimum matched-forest partition of a graph.

Lemma 8. *The problem of finding the minimum matched-forest partition of a graph G is NP-hard.*

Moreover, we show that finding the minimum matched-forest partition is even hard to approximate. We use the well-known result by Zuckerman [51] that for all $\epsilon > 0$, it is NP-hard to approximate *chromatic number* to within $n^{1-\epsilon}$.

Lemma 9. *For all* $\epsilon > 0$*, it is NP-hard to approximate the minimum matched-forest partition of a graph G to within* $n^{1-\epsilon}$.

Despite these hardness results, we provide the polynomial-time Zone Selection (ZS) Algorithm to find a matched-forest partition of a graph. We prove that the ZS Algorithm provides a constant approximation for the minimum matched-forest partition of a graph G when G is planar.

Before describing the ZS Algorithm in detail, we first describe an algorithm that is used in the ZS Algorithm, when G is planar. It is known that for a planar graph G, $a(G) \leqslant 3$ [19]. Based on the proof provided in [19], we introduce a recursive 3-Acyclic Partition of Planar (3APP) Algorithm. The Lemma below shows the correctness of this Algorithm.

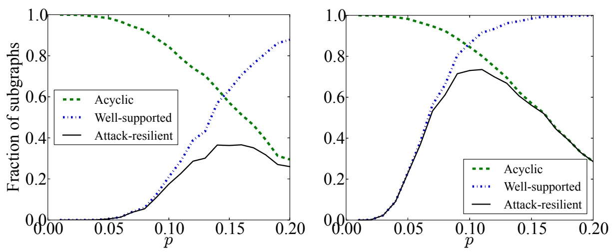
Lemma 10. *The 3APP Algorithm partitions the nodes of a planar graph G into 3 subsets such that each subset induces an acyclic graph.*

[5]To the best of our knowledge, this is the first time that the problem is studied.

Algorithm 3 - Zone Selection (ZS)

Input: A connected graph G.

1: Find an optimal matching cover $M_1, M_2, \ldots, M_t$ of G [45].
2: For each M_i, separate the matched nodes into two set of nodes V_{2i-1}, V_{2i} such that $\forall \{v, u\} \in M$, $v \in V_{2i-1}$ and $u \in V_{2i}$.
3: For any $1 \leqslant i \leqslant 2t$, $Q_i = V_i \backslash \bigcup_{j=1}^{i-1} Q_j$.
4: **for** each Q_i **do**
5:    **if** $G[Q_i]$ is acyclic **then continue**
6:    **if** $G[Q_i]$ is a planar graph **then**
7:       Use 3APP Algorithm to partition $G[Q_i]$.
8:    **else**
9:       Use any greedy algorithm to partition $G[Q_i]$ into acyclic subgraphs.
10: Name the resulted partitions $P_1, \ldots, P_k$.
11: **return** $P_1, \ldots, P_k$.



(a) $n = 30$ (b) $n = 50$

Figure 6: The relationships between the number of edges in random graphs (or equivalently p), and the fraction of induced subgraphs that form well-supported and acyclic zones of size $|V_H| = 10$. For each p, 100 random graphs were generated and in each graph, 100 subgraphs were chosen randomly.

We now present the ZS Algorithm. The ZS Algorithm first finds an optimal matching cover $M_1, M_2, \ldots, M_t$ of G using an $O(n^3)$ algorithm introduced in [45] (line 1). Then, in lines 2 and 3, it uses this matching cover to partition V into $Q_1, Q_2, \ldots, Q_{2t}$. It is easy to see that for each Q_i, $M_{\lceil i/2 \rceil} \cap E[Q_i, \bar{Q}_i]$ is the matching in $G[Q_i, \bar{Q}_i]$ that covers nodes in Q_i. Then, in order to satisfy the acyclicity condition on the partitions, it partitions Q_is that do not induce an acyclic graph, into subsets so that each subset induces an acyclic graph. When $G[Q_i]$ is a planar graph, it uses 3APP Algorithm to partition $G[Q_i]$. When it is not, it uses any *greedy algorithm* to do so. Thus, the resulted partition $P_1, P_2, \ldots, P_k$ satisfies the conditions of a matched-forest partition.

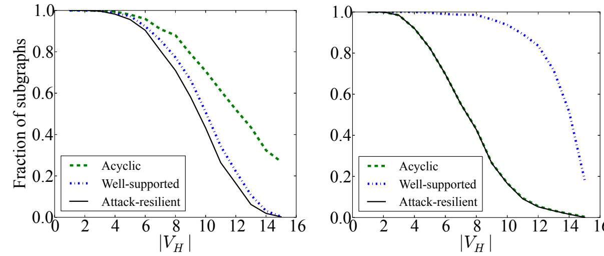
The lemma below states that when G is planar, the ZS Algorithm provides a constant approximation of the optimal matched-forest partition. We demonstrate the results obtained by the algorithm in the following section.

LEMMA 11. *If G is planar, the ZS Algorithm provides a 6-approximation of the minimum* matched-forest *partition of G in $O(n^3)$.*

6. NUMERICAL RESULTS

In this section, we (i) numerically study the relationship between the structural properties of a grid and its resilience to joint cyber and physical attacks, and (ii) demonstrate the results obtained by the ZS Algorithm in several known power grid networks.

To assess the relationship between the structural properties of a grid and its resilience to attacks, we quantify the resilience by the fraction of the node induced subgraphs of G that form attack-resilient zones. Recall from Section 4 that a zone H is attack-resilient, if it is well-supported and acyclic. We first study the relationships between the number of edges and the fraction of induced subgraphs that form well-supported, acyclic, and attack-resilient zones. These relationships can be best demonstrated in *random graphs* [24], since by increasing p (the probability that two nodes are connected), the total number of edges increases.



(a) $n = 30, m = 56$ (b) $n = 30, m = 104$

Figure 7: The relationship between the size of the zone $|V_H|$ in scale-free graphs, and the fraction of induced subgraphs that form well-supported and acyclic zones of size $|V_H|$. For each case, 1000 subgraphs were chosen randomly.

We generated random graphs with $n = 30$ and $n = 50$ nodes with various p values. For each p, we generated 100 random graphs and in each graph randomly selected 100 subgraphs of size 10. We then computed the fraction of subgraphs that form well-supported, acyclic, and attack-resilient zones. As can be seen in Fig. 6, as p increases, the fraction of subgraphs that form acyclic zones decreases and the fraction that form well-supported zones increases. In Fig. 6(a), when $p \approx 0.14$ these two fractions are equal and the fraction of attack-resilient subgraph is maximized. As can be seen in Fig. 6(b), the value of p for which these two fractions are equal (i.e., the fraction of attack-resilient subgraphs is maximized) decreases as n increases.

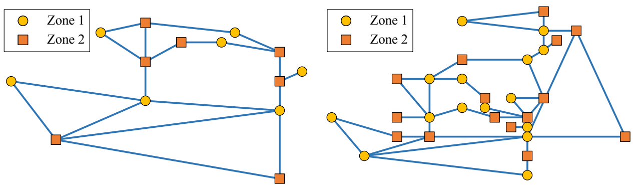
We also illustrate the relationships between the size of the zone $|V_H|$ and the fraction of induced subgraphs that form well-supported, acyclic, and attack-resilient zones. It is known that *scale-free graphs* are relatively good representatives of power grid networks [10], and therefore, we focus here on such graphs. Fig. 7 shows the relationships in scale-free graphs with $n = 30$ nodes and $m = 56$ or 104 edges. As can be seen in 7(a), as the size of the zone increases, the fraction of zones that are well-supported decreases faster than the fraction of zones that are acyclic. Thus, the well-supportedness is the restricting factor for attack-resilience. However, this trend is different in 7(b). Since the graph has more edges, as the size of the zone increases, the fraction of zones that are acyclic decreases faster than the fraction of zones that are well-supported. Here, the acyclicity is the restricting factor.

Overall, these results show that the structure of the grid plays an important role in its resilience and should be considered when designing a control network.[6]

[6]Clearly, random and scale free graphs do not model power grids perfectly. However, we used them to gain insight into the relations between structural properties and resilience. In future work, we will study more realistic network topologies [41].

Table 3: Number of partitions into which the ZS Algorithm divides different networks.

Network	Nodes	Edges	Partitions
IEEE 14-Bus	14	20	2
IEEE 30-Bus	30	41	2
IEEE 118-Bus	118	179	5
IEEE 300-Bus	300	409	14
Polish grid	3120	3684	10
Colorado state grid	662	864	6
Western interconnection	13135	16860	9



(a) IEEE 14 bus

(b) IEEE 30 bus

Figure 8: Partitioning of the IEEE 14 and IEEE 30 bus systems into 2 attack resilient zones (using the ZS Algorithm).

We now demonstrate results obtained by the ZS Algorithm in several known power networks. Table 3 lists the considered grids and number of resulting partitions. For example, Fig. 8 shows the partitions obtained by ZS Algorithm in the IEEE 14-Bus and 30-Bus benchmark systems [2]. As can be seen, in both cases the graphs can be partitioned into two attack-resilient zones. We also evaluated the ZS Algorithm on the IEEE 118 and 300-bus systems, the Polish grid (available with MATPOWER [50]), the Colorado state grid, and the U.S. Western Interconnection network.[7] For example, the 6 zones into which the Colorado grid is partitioned appear in Fig. 9. Recall form Section 5 that when G is planar, the ZS Algorithm is a 6-approximation algorithm for the minimum matched-forest problem. However, as can be seen from the examples above, in practice, it partitions the networks into few zones.

We note that the ZS Algorithm does not take the geographical constraints into account. Thus, when partitioning very large networks such as the Western Interconnection (see Fig. 10), the nodes in the same partition may be geographically distant from each other. This is impractical, since the PMUs from the same zone should send the data to a single PDC. However, it is easy to see that if a zone is attack-resilient, any of its subgraphs is also attack-resilient. Therefore, the partitions obtained by the ZS Algorithm can be further divided into smaller zones based on geographical constraints (e.g., into zones within different states in Fig. 10). This approach does not result in an optimal partitioning. Hence, obtaining an efficient partitioning with geographical constraints is a subject of future work.

7. CONCLUSION

We studied joint cyber and physical attacks on power grids. We developed methods to recover information about the phase angles of the nodes and the disconnected edges, using only the information available from outside of the attacked zone. We identified graph structures and constraints

[7]The data of the Western Interconnection (and of Colorado) was obtained from the Platts Geographic Information System (GIS) [1].

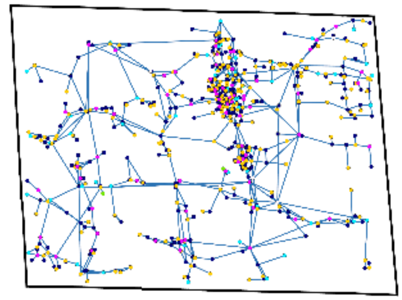


Figure 9: Partitioning of the Colorado state grid into 6 attack-resilient zones (using the ZS Algorithm). Nodes with the same color are in the same zone.

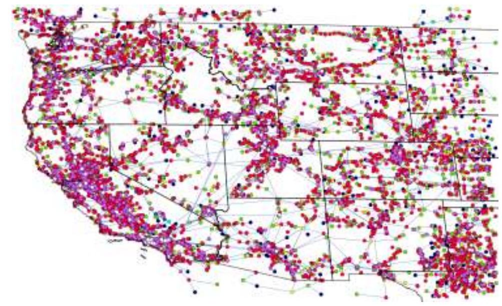


Figure 10: Partitioning of the U.S. Western Interconnection into 9 attack-resilient zones (using the ZS Algorithm). Nodes with the same color are in the same zone.

on the disconnected edges for which these methods are guaranteed to recover the state information. Moreover, we showed that the problem of partitioning the grid into the minimum number of attack-resilient zones is not approximable to within $n^{1-\epsilon}$ for all $\epsilon > 0$ unless P=NP. However, for planar graphs, we developed an approximation algorithm for the partitioning problem. Finally, we illustrated via numerical results the relationship between the structural properties of the power grid and its resilience to joint attacks as well as the operation of the partitioning algorithm.

This is one of the first steps towards understanding the vulnerabilities of power grids to joint cyber and physical attacks and developing methods to mitigate their effects. Hence, there are still many open problems. In particular, we plan to generalize Theorems 2 and 3 to a broader class of graphs. Moreover, we will develop algorithms to partition the grid into weakly-attack-resilient zones while taking into account geographical constraints and constraints on the number and positions of the PDCs. Finally, we will study the interdependencies between cascading failures in power grids and its control network when the control network faces sophisticated cyber attacks.

Acknowledgement

This work was supported in part by CIAN NSF ERC under grant EEC-0812072, the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. [PIIF-GA-2013-629740].11, NSF under grant CCF-1320654, and DTRA grant HDTRA1-13-1-0021. We thank Meric Uzunoglu and Andrey Bernstein for their help with processing the WI data.

8. REFERENCES

[1] Platts GIS Data. `http://www.platts.com/Products/gisdata`.

[2] Power systems test case archive. Available at: `http://www.ee.washington.edu/research/pstca/`.

[3] U.S.-Canada Power System Outage Task Force. Report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations. 2004.

[4] Report of the enquiry committee on grid disturbance in Northern region on 30th July 2012 and in Northern, Eastern and North-Eastern region on 31st July 2012, Aug. 2012.

[5] The Federal Energy Regulatory Comission (FERC) and the North American Electric Reliability Corporation (NERC). Arizona-Southern California Outages on September 8, 2011. 2012.

[6] Assault on California power station raises alarm on potential for terrorism. Source: `http://goo.gl/RiuhI1`, 2014.

[7] Cybersecurity and the North American electric grid: New policy approaches to address an evolving threat. Source: `http://goo.gl/WgUbPI`, 2014.

[8] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[9] R. Bapat. *Graphs and matrices*. Springer, 2010.

[10] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[11] A. R. Bergen and V. Vittal. *Power Systems Analysis*. Prentice-Hall, 1999.

[12] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman. Sensitivity analysis of the power grid vulnerability to large-scale cascading failures. *ACM SIGMETRICS Perform. Eval. Rev.*, 40(3):33–37, 2012.

[13] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman. Power grid vulnerability to geographically correlated failures - analysis and control implications. In *Proc. IEEE INFOCOM'14*, Apr. 2014.

[14] D. Bienstock. Optimal control of cascading power grid failures. *Proc. IEEE CDC-ECC*, Dec. 2011.

[15] D. Bienstock and A. Verma. The $N-k$ problem in power grids: New models, formulations, and numerical experiments. *SIAM J. Optimiz.*, 20(5):2352–2380, 2010.

[16] J. A. Bondy and U. Murty. Graph theory, volume 244 of graduate texts in mathematics, 2008.

[17] J. A. Bondy and U. S. R. Murty. *Graph theory with applications*. Macmillan London, 1976.

[18] S. Buldyrev, R. Parshani, G. Paul, H. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010.

[19] G. Chartrand, D. Geller, and S. Hedetniemi. Graphs with forbidden subgraphs. *J. Combin. Theory Ser. B*, 10(1):12–41, 1971.

[20] D. P. Chassin and C. Posse. Evaluating North American electric grid reliability using the Barabási–Albert network model. *Phys. A*, 355(2-4):667 – 677, 2005.

[21] G. Dán and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *Proc. IEEE SmartGridComm'10*, 2010.

[22] I. Dobson, B. Carreras, V. Lynch, and D. Newman. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. *Chaos*, 17(2):026103, 2007.

[23] D. Donoho and Y. Tsaig. Fast solution of l1-norm minimization problems when the solution may be sparse, 2006. *Preprint at `http://statweb.stanford.edu/~donoho/Reports/2006/kstep-20061005.pdf`*.

[24] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[25] L. Gan, A. Wierman, U. Topcu, N. Chen, and S. H. Low. Real-time deferrable load control: handling the uncertainties of renewable generation. In *Proc. ACM e-Energy'13*, May 2013.

[26] M. R. Garey and D. S. Johnson. Computers and intractability: a guide to the theory of np-completeness. 1979.

[27] Y. Ghiassi-Farrokhfal, S. Keshav, C. Rosenberg, and F. Ciucu. Firming solar power. In *Proc. ACM SIGMETRICS'13 (poster description)*, June 2013.

[28] 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 Process. Mag.*, 29(5):33–43, 2012.

[29] K. Khandeparkar, P. Patre, S. Jain, K. Ramamritham, and R. Gupta. Efficient PMU data dissemination in smart grid. In *Proc. ACM e-Energy'14 (poster description)*, June 2014.

[30] J. Kim and L. Tong. On topology attack of a smart grid: undetectable attacks and countermeasures. *IEEE J. Sel. Areas Commun*, 31(7):1294–1305, 2013.

[31] J. Kleinberg, M. Sandler, and A. Slivkins. Network failure detection and graph connectivity. In *Proc. ACM-SIAM SODA'04*, Jan. 2004.

[32] J. Liu, C. H. Xia, N. B. Shroff, and H. D. Sherali. Distributed optimal load shedding for disaster recovery in smart electric power grids: A second-order approach. In *Proc. ACM SIGMETRICS'14 (poster description)*, June 2014.

[33] Y. Liu, P. Ning, and M. K. Reiter. False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur.*, 14(1):13, 2011.

[34] L. Lu, J. Tu, C.-K. Chau, M. Chen, and X. Lin. Online energy generation scheduling for microgrids with intermittent energy sources and co-generation. In *Proc. ACM SIGMETRICS'13*, June 2013.

[35] N. M. Manousakis, G. N. Korres, and P. S. Georgilakis. Taxonomy of PMU placement methodologies. *IEEE Trans. Power Syst.*, 27(2):1070–1077, 2012.

[36] C. Phillips. The network inhibition problem. In *Proc. ACM STOC'93*, May 1993.

[37] A. Pinar, J. Meza, V. Donde, and B. Lesieutre. Optimization strategies for the vulnerability analysis of the electric power grid. *SIAM J. Optimiz.*, 20(4):1786–1810, 2010.

[38] W. Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 1987.

[39] P. D. Seymour. On odd cuts and plane multicommodity flows. *P. Lond. Math. Soc.*, 3(1):178–192, 1981.

[40] S. Soltan, D. Mazauric, and G. Zussman. Cascading failures in power grids – analysis and algorithms. In *Proc. ACM e-Energy'14*, June 2014.

[41] S. Soltan and G. Zussman. A statistical method for synthetic power grid generation based on the U.S. Western Interconnection. Abstract presented at *SIAM NS'15*, May 2015.

[42] J. E. Tate and T. J. Overbye. Line outage detection using phasor angle measurements. *IEEE Trans. Power Syst.*, 23(4):1644–1652, 2008.

[43] J. E. Tate and T. J. Overbye. Double line outage detection using phasor angle measurements. In *Proc. IEEE PES'09*, July 2009.

[44] O. Vukovic, K. C. Sou, G. Dán, and H. Sandberg. Network-layer protection schemes against stealth attacks on state estimators in power systems. In *Proc. IEEE SmartGridComm'11*, 2011.

[45] X. Wang, X. Song, and J. Yuan. On matching cover of graphs. *Math. Program.*, 147(1-2):499–518, 2014.

[46] D. B. West et al. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, 2001.

[47] H. Xiao and E. M. Yeh. Cascading link failure in the power grid: A percolation-based analysis. In *Proc. IEEE Int. Work. on Smart Grid Commun.*, June 2011.

[48] Y. Zhao, A. Goldsmith, and H. V. Poor. On PMU location selection for line outage detection in wide-area transmission networks. In *Proc. IEEE PES'12*, July 2012.

[49] H. Zhu and G. B. Giannakis. Sparse overcomplete representations for efficient identification of power line outages. *IEEE Trans. Power Syst.*, 27(4):2215–2224, 2012.

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

[51] D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proc. ACM STOC'06*, May 2006.

APPENDIX

A. PRELIMINARIES

In this appendix we provide some of the known theorems and definitions used in the proofs.

A *bond* of a graph is a minimal nonempty edge cut, that is, a nonempty edge cut none of whose nonempty proper subsets is an edge cut.

A graph in which each node has even degree is called an *even* graph. A *circuit* is the union of edge-disjoint *cycles*. It is easy to see that a *cycle* is also a *circuit*.

THEOREM A.1 (THEOREM 5.2 [39]). *Let $G = (V, E)$ be a planar Eulerian graph, and let $F \subseteq E$. Then the following are equivalent:* (i) *F is G-separable,* (ii) *for each bond D, $|D \cap F| \leqslant |D - F|$.*

THEOREM A.2 (THEOREM 10.16 [17]). *Let G be a connected planar graph, and let G^* be a planar dual of G.* (i) *If C is a cycle of G, then C^* is a bond of G^*.* (ii) *If B is a bond of G, then B^* is a cycle of G^*.*

THEOREM A.3 (THEOREM 6.1.16 [46]). *The followings are equivalent for a planar graph G,* (i) *G is bipartite,* (ii) *every face of G has even length, and* (iii) *the dual graph G^* is Eulerian.*

THEOREM A.4 (EULER'S FORMULA [16]). *For any planar graph G with n nodes, m edges, r faces, and c connected components, the following formula holds, $n + r - m = c$.*

LEMMA A.1 (COROLLARY 2.16 [16]). *The symmetric difference of two even subgraphs is an even subgraph.*

B. PROOFS

PROOF OF COROLLARY 2. Suppose $M = (U, V_H)$ is the matching for $G[V_H, \bar{V}_H]$ that covers V_H, and suppose $U \subseteq \bar{V}_H$ are the matched nodes which are in $\bar{V}_H$. Since M is the matching in $G[V_H, \bar{V}_H]$ that covers H, thus $|U| = |V_H|$. Regarding Theorem 1, to show that H is well-supported almost surely, we need to show that the columns of the matrix $\mathbf{A}_{\bar{H}|H}$ are linearly independent almost surely. For this reason, we show that $\det(\mathbf{A}_{U|V_H}) \neq 0$ almost surely. $\det(\mathbf{A}_{U|V_H})$ can be considered as a polynomial of the nonzero entries of the admittance matrix using Leibniz formula. Now assume $U = \{u_1, u_2, \ldots, u_{|V_H|}\}$ are matched to $V_H = \{v_1, v_2, \ldots, v_{|V_H|}\}$ in order. It can be seen that $\prod_{i=1}^{|V_H|} a_{u_i v_i}$ is a term with nonzero coefficient in $\det(\mathbf{A}_{U|V_H})$. Therefore, $\det(\mathbf{A}_{U|V_H})$ is not a zero polynomial in terms of its nonzero entries. Now since the set of reactance values for the edges in $E[V_H, \bar{V}_H]$ such that $\det(\mathbf{A}_{U|V_H}) = 0$ is a measure zero set in the real space, thus $\det(\mathbf{A}_{U|V_H}) \neq 0$ almost surely. $\square$

PROOF OF LEMMA 3. It is easy to see that the solution to (5) is unique if and only if $\mathbf{D}_H$ has linearly independent columns. It is known that $\mathrm{rank}(\mathbf{D}_H) = |V_H| - c$ in which c is the number of connected components of H [9, Theorem 2.3]. Therefore, $\mathbf{D}_H$ has linearly independent columns if and only if each connected component of $\mathbf{D}_H$ is a tree, which means that $\mathbf{D}_H$ should be acyclic. $\square$

PROOF OF THEOREM 2. Recall that we can assume $V_H = \{1, 2, \ldots, |V_H|\}$ and $E_H = \{e_1, e_2, \ldots, e_{|E_H|}\}$. We assign an arbitrary orientation to the edges of H and fix the embedding of H on the plane. We show the set of oriented edges by $\mathcal{E}_{\mathcal{H}} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_{|E_H|}\}$. Suppose H has r faces and $C_1, C_2, \ldots, C_r$ are the cycles that surrounded those faces. For each cycle C_i, define vector $\vec{c_i} \in \{-1, 0, 1\}^{|E_H|}$ as follows,

$$c_{ij} = \begin{cases} 0 & \text{if } \epsilon_j \notin C_i, \\ 1 & \text{if } \epsilon_j \in C_i \text{ and } \epsilon_j \text{ traverse } C_i \text{ clockwise}, \\ -1 & \text{if } \epsilon_j \in C_i \text{ and } \epsilon_j \text{ traverse } C_i \text{ counterclockwise}. \end{cases}$$

It is easy to see that $\forall i, \mathbf{D}_H \vec{c_i} = 0$. Therefore, $\vec{c_i} \in \mathrm{Null}(\mathbf{D})$. On the other hand it is easy to see that $\vec{c_i}$s are linearly independent. If c is the number of connected components of H, then $\dim(\mathrm{Null}(\mathbf{D})) = |E_H| - |V_H| + c$ which from *Euler's formula* is equal to r. Thus, $\vec{c_i}$s form a basis for the null space of the incidence matrix $\mathbf{D}$. Now suppose $\vec{x}$ is the solution to (5) such that $\mathrm{supp}(\vec{x}) = \{i | e_i \in F\}$ (from Lemma 2 we know that such a solution exists). We need to prove that for any $\vec{z} = \vec{x} + y_1\vec{c_1} + \cdots + y_r\vec{c_r}$, $\|\vec{z}\|_1 \leqslant \|\vec{x}\|_1$ iff

$y_1 = y_2 = \cdots = y_r = 0$. Please notice that since $\vec{c_i}$s are the cycles associated with the faces of the planar graph H, each edge e_j appears in at most two cycles C_t and C_s. Thus, the entries of $\vec{z}$, z_js are in one of the following forms, (i) $z_j = 0$, (ii) $z_j = x_j$, (iii) $z_j = x_j \pm y_t$, (iv) $z_j = x_j \pm (y_t - y_s)$, or (v) $z_j = \pm(y_t - y_s)$.

First, we show that $\|\vec{z}\|_1 \geqslant \|\vec{x}\|_1$. Suppose $F = \{e_{i_1}, e_{i_2}, \ldots, e_{i_k}\}$. From the assumption, we know that F^* is H^*-separable. Thus, there are pairwise edge-disjoint cycles $O_j^* (1 \leqslant j \leqslant k)$ in H^*, such that $\forall e_{i_j} \in F, e_{i_j}^* \in O_j^*$. For each $j(1 \leqslant j \leqslant k)$ define $N_O(j) = \{p | e_p^* \in O_j^*\}$. From the definitions it is easy to see that $\forall p, q (1 \leqslant p \neq q \leqslant k), N_O(p) \cap N_O(q) = \varnothing$. Thus, $\|\vec{z}\|_1 \geqslant \sum_{j=1}^{k} \sum_{p \in N_O(j)} |z_p|$. On the other hand, from the triangle inequality, it is easy to see that for all $1 \leqslant j \leqslant k$, $\sum_{p \in N_O(j)} |z_p| \geqslant |x_{i_j}|$. Thus, $\|\vec{z}\|_1 \geqslant \sum_{j=1}^{k} \sum_{p \in N_O(j)} |z_p| \geqslant \sum_{j=1}^{k} |x_{i_j}| = \|\vec{x}\|_1$.

Now we show that $\|\vec{z}\|_1 = \|\vec{x}\|_1$ is not possible unless $y_1 = y_2 = \cdots = y_r = 0$. If $\{j_1, j_2, \ldots, j_t\}$ are the indices for which $y_{j_1} = y_{j_2} = \cdots = y_{j_t} = w$, then $y_{j_1}\vec{c_{j_1}} + \cdots + y_{j_t}\vec{c_{j_t}} = w\vec{c''}$ for which from Lemma A.1, $\vec{c''}$ is the vector associated with the circuit $C'' = C_{j_1} \Delta C_{j_2} \Delta \ldots \Delta C_{j_t}$ (Δ is the symmetric difference). Therefore we can rewrite $\vec{z}$ as $\vec{x} + w_1\vec{c_1''} + w_2\vec{c_2''} + \cdots + w_q\vec{c_q''}$ for which $w_1 > w_2 > \cdots > w_q$. If $q = 0$, then there is nothing left to prove. If $q = 1$, then it can be easily concluded from Lemma 4 that it is not possible to have $\|\vec{z}\|_1 = \|\vec{x}\|_1$ unless $w_1 = 0 \Rightarrow y_1 = y_2 = \cdots = y_r = 0$. Now assume $q > 1$, from what we showed previously, we know that in order to have $\|\vec{z}\|_1 = \|\vec{x}\|_1$, we should have $\|\vec{z}\|_1 = \sum_{j=1}^{k} \sum_{p \in N_O(j)} |z_p| = \sum_{j=1}^{k} |x_{i_j}| = \|\vec{x}\|_1$. Equality $\|\vec{z}\|_1 = \sum_{j=1}^{k} \sum_{p \in N_O(j)} |z_p|$ shows that $\mathrm{supp}(\vec{z}) \subseteq \bigcup_{j=1}^{k} N_O(j)$. Equality $\sum_{j=1}^{k} \sum_{p \in N_O(j)} |z_p| = \sum_{j=1}^{k} |x_{i_j}|$ shows that w_1 should appear in $\|\vec{z}\|_1$, half of the time positive and half of the time negative. However, since w_1 has the largest value, it appears always as $w_1 - w_i$ in the instances like $|w_1 - w_i|$. Now from assumption we know that $|\mathrm{supp}(\vec{x}) \cap C_1''| < |C_1''|/2$. Thus, there are more than $|C_1''|/2$ instances like $|w_1 - w_i|$ or $|w_1|$ in $\|\vec{z}\|_1$. Therefore, it is not possible that w_1 appears half of the time positive and half of the time negative in $\|\vec{z}\|_1$, which shows that if $q > 1$ then $\|\vec{z}\|_1 \neq \|\vec{x}\|_1$. Thus, the proof is complete. ☐

Proof of Corollary 4. H is bipartite, therefore from Theorem A.3, H^* is Eulerian. For any C in H, $|C \cap F| < |C \backslash F|$, therefore from Theorem A.2 for any bond C^* in H^*, $|C^* \cap F^*| < |C^* \backslash F^*|$. Combining these two, from Theorem A.1 we can conclude that F^* is H^*-separable. Thus, we can simply apply Theorem 2. ☐

Proof of Lemma 5. From Corollary 1 we know that $\vec{\delta}_H = \vec{\theta}_H - \vec{\theta}'_H$ is a solution to $\mathbf{A}_{\bar{H}|H}\vec{\delta}_H + \mathbf{A}_{\bar{H}|\bar{H}}\vec{\delta}_{\bar{H}} = 0$. Now if $\vec{\delta} = \vec{\theta} - \vec{\theta}'$, from Lemma 2 we know that there exists a vector $\vec{x} \in \mathbb{R}^{|E_H|}$ such that $\mathrm{supp}(\vec{x}) = \{i | e_i \in F\}$ and $\mathbf{D}_H\vec{x} = \mathbf{A}_{H|H}\vec{\delta}_H + \mathbf{A}_{H|\bar{H}}\vec{\delta}_{\bar{H}}$. Thus, the proof is complete. ☐

Proof of Lemma 7. First, notice that since v is H-outer-connected, $N(v) \subseteq \bar{V}_H$. Thus, $\delta_v = 1/d(v) \sum_{u \in N(v)} \delta_u$ implies that δ_v is unique. Now, let us compute the v^{th} entry of the vectors on the both side of the equation $\mathbf{D}_H\vec{x} = \mathbf{A}_{H|H}\vec{\delta}_H + \mathbf{A}_{H|\bar{H}}\vec{\delta}_{\bar{H}}$. Since v is H-outer-connected, the v^{th} row of $\mathbf{D}_H$ is a zero vector. Thus, $(\mathbf{D}_H\vec{x})_v = 0$ for any $\vec{x}$. It is also easy to see that $(\mathbf{A}_{H|H}\vec{\delta}_H)_v = \delta_v d(v)$ and $(A_{H|\bar{H}}\vec{\delta}_{\bar{H}})_v = -\sum_{u \in N(v)} \delta_u$. Since $(\mathbf{D}_H\vec{x})_v = (\mathbf{A}_{H|H}\vec{\delta}_H)_v + (A_{H|\bar{H}}\vec{\delta}_{\bar{H}})_v$, we can conclude that $\delta_v = 1/d(v) \sum_{u \in N(v)} \delta_u$. Thus, the proof is complete. ☐

Proof of Theorem 3. First, using Lemma 7, since the phase angles for the H-outer-connected nodes can be computed uniquely, without loss of generality and for simplicity, we can assume that H does not contain any H-outer-connected nodes. Suppose $\vec{c}_1, \ldots \vec{c}_r$ are the vectors associated with the faces in H as we defined in the proof of Theorem 2. Suppose $\vec{b}_1, \ldots, \vec{b}_t$ are the vectors associated with the coboundry of the H-inner-connected nodes as we defined in the proof of Lemma 6. All we need to prove is that if $\vec{x}$ is the solution to (7) such that $\mathrm{supp}(\vec{x}) = \{i | e_i \in F\}$ then for any solution $\vec{z}$ for (7), $\|\vec{z}\|_1 > \|\vec{x} + y_1\vec{c}_1 + \cdots + y_r\vec{c}_1 + w_1\vec{b}_1 + \cdots + w_t\vec{b}_t\|_1$ unless $y_1 = \cdots = y_r = w_1 = \cdots = w_t = 0$. The rest of the proof is exactly similar to proofs of Lemma 6 and Theorem 2. The key is Condition (v) implies that the cycles ($\vec{c}_i$s) and coboundries ($\vec{b}_i$s) are edge disjoint. Thus, since all the conditions for Lemma 6 and Theorem 2 also hold here, with exactly the same approach as in the proofs of Lemma 6 and Theorem 2 we can conclude this Theorem. ☐

Proof of Lemma 8. We show that finding the minimum matched-forest partition is at least as hard as vertex arboricity. Suppose $G = (V, E)$ is given and we are interested in $a(G)$. We build a new graph $\hat{G} = (\hat{V}, \hat{E})$ by getting a copy of graph G, $G'' = (V'', E'')$ and connect every node in G to its counterpart in G''. Thus, $\hat{V} = V \cup V''$ and $\hat{E} = E \cup E'' \cup \{\{v_i, v_i''\} | v_i \in V\}$. We prove that $\hat{G}$ has a matched-forest partition of size k if and only if $a(G) \leqslant k$. First, let us assume $\hat{G}$ has a matched-forest partition of size k, namely $\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_k$. Since subgraphs induced by $\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_k$ in $\hat{G}$ are acyclic and partition $\hat{V}$, it is easy to see that subgraphs induced by $\hat{H}_1 \cap V, \hat{H}_2 \cap V, \ldots, \hat{H}_k \cap V$ in G are acyclic and partition V. Thus, $a(G) \leqslant k$. Now, let us assume $a(G) \leqslant k$. There should exists a partition of nodes of G into subsets $H_1, H_2, \ldots, H_k$ such that each subset induces an acyclic graph. Assume $H_1'', H_2'', \ldots, H_k''$ are the counterparts of these subsets in G''. For any $i, 1 \leqslant i \leqslant k-1$ define $\hat{H}_i = H_i \cup H_{i+1}''$, and $\hat{H}_k = H_k \cup H_1''$. It is easy to see that $\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_k$ is a matched-forest partition of size k for $\hat{G}$. Thus, we proved that $\hat{G}$ has a matched-forest partition of size k if and only if $a(G) \leqslant k$. It means that the minimum matched-forest partition of $\hat{G}$ is equal to $a(G)$, which shows that the minimum matched-forest partition is at least as hard as vertex arboricity and therefore it is NP-hard. ☐

Proof of Lemma 9. For a graph G, assume $\chi(G)$ is its chromatic number. Since each color gives an independent set of G, induced subgraph by the nodes with the same color is acyclic with no edges. Thus, it is easy to see that $a(G) \leqslant \chi(G)$. Suppose there is an α-approximation algorithm for the minimum matched-forest problem. Define $\hat{G}$ as in proof of Lemma 8. Assume this algorithm partitions $\hat{G}$ into k subsets. From the proof of Lemma 8, it is easy to see that $k \leqslant \alpha a(G)$. On the other hand, since each acyclic graph has the chromatic number of at most 2, this algorithm gives the $2k$ coloring of graph G. However, $2k \leqslant 2\alpha a(G) \leqslant 2\alpha\chi(G)$. Thus, this algorithm gives a 2α-approximation of the chromatic number of G. However, the result by Zuckerman [51] states that for all $\epsilon > 0$, it is NP-hard to approximate chromatic number to within $n^{1-\epsilon}$. Therefore, for all $\epsilon > 0$, it is NP-hard to approximate the minimum matched-forest problem to within $n^{1-\epsilon}$ as well. ☐

Proof of Lemma 10. It is known that every planar graph has a node of degree less than or equal to 5 [46]. Therefore, line 1 of the algorithm can always find v. At line 4, recursively we know that subgraphs induced by Q_1, Q_2, Q_3 in $G \backslash v$ are acyclic. Now since $\deg(v) \leqslant 5$, there exists a partition such that $|N(v) \cap Q_i| \leqslant 1$. Without loss of generality we can assume that $|N(v) \cap Q_1| \leqslant 1$. Hence, adding v to Q_1 does not produces any cycles. Thus, subgraphs induced by $Q_1 \cup \{v\}, Q_2, Q_3$ in G are acyclic. □

PROOF OF LEMMA 11. Suppose OPT is the minimum number of matched-forest partitions of G and OPT_m is the number of optimal matching cover of G. Since for any subset V_H in the partition of $G = (V, E)$ into matched-forest partitions there exists a matching that covers V_H, we can cover V with OPT matchings. Thus, $OPT_m \leqslant OPT$. Now since G is planar, ZS Algorithm uses $3APP$ to partition Q_is into atmost 3 subsets. Hence, if k is the number of subsets returned by ZS Algorithm then $k \leqslant 3 \times 2 \times OPT_m \leqslant 6 \times OPT$. Thus, ZS Algorithm provides a 6-approximation of the minimum *matched-forest* partition of G.

As for the running time, ZS algorithm takes $O(n^3)$ time to find the optimal matching cover of G and $O(|Q_i|)$ to partition Q_i. Now since $k \leqslant n$ and $|Q_i| \leqslant n$, line 4 does not take more than $O(n^2)$. Thus, the total running time is $O(n^3)$. □