EXPOSE the Line Failures following a Cyber-Physical Attack on the Power Grid

Saleh Soltan, *Member, IEEE,* and Gil Zussman, *Senior Member, IEEE*

***Abstract*—Recent attacks on power grids demonstrated their vulnerability to cyber and physical attacks. To analyze this vulnerability, we study cyber-physical attacks that affect both the power grid physical infrastructure and its underlying Supervisory Control And Data Acquisition (SCADA) system. We assume that an adversary attacks an area by: (i) disconnecting some lines within that area, and (ii) obstructing the information (e.g., status of the lines and voltage measurements) from within the area to reach the control center. We leverage the algebraic properties of the AC power flows to introduce the efficient EXPOSE Algorithm for detecting line failures and recovering voltages inside that attacked area after such an attack. The EXPOSE Algorithm outperforms the state-of-the-art algorithm for detecting line failures using partial information under *the AC power flow model* in terms of scalability and accuracy. The main advantages of the EXPOSE Algorithm are that its running time is independent of the size of the grid and number of line failures, and that it provides accurate information recovery under some conditions on the attacked area. Moreover, it approximately recovers the information and provides the confidence of the solution when these conditions do not hold.**

***Index Terms*—AC Power Flows, State Estimation, Line Failures Detection, Cyber Attack, Physical Attack.**

I. Introduction

Recent cyber attack on the Ukrainian grid in December 2015 [1] demonstrated the vulnerability of power grids to cyber attacks. As indicated in the aftermath report of the attack [1], once the attackers obtain access to the grid's Supervisory Control And Data Acquisition (SCADA) system, they can delete, modify, and spoof the data as well as remotely change the grid's topology by activating the circuit breakers.

The power grid infrastructure is also vulnerable to physical attacks. Such an attack occured in April 2014 in San Jose, California, when snipers tried to shut down a substation simply by shooting at its transformers [2]. Hence, a physical attack on the power lines and the measurement devices can have a similar effect to a cyber attack.

To analyze these vulnerabilities, in this paper, we study cyber-physical attacks that affect both the power grid physical infrastructure and its SCADA system. Fig. 1 shows the main components of the power grid. An adversary can attack the grid by damaging the power lines and measurement devices with a physical attack, by remotely disconnecting the lines and erasing the measurements data with a cyber attack, or by performing a combination of the both.

Independent of the attack strategy, we assume that an adversary attacks an area by: (i) disconnecting some lines within that area (*failed lines*), and (ii) obstructing the information (e.g., status of the lines and voltage measurements) from within the area to reach the control center. We call this area, the *attacked zone*. Our objective is to detect the failed lines and recover the voltages inside the attacked zone using the information available outside of the attacked zone as well as the information before the attack. An example of such an attack on the IEEE 300-bus system is depicted in Fig. 2.

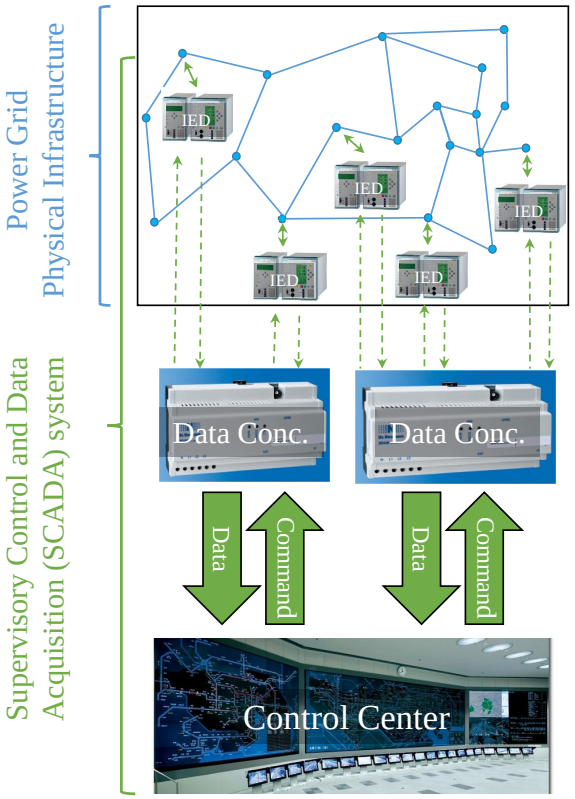


Fig. 1: The main components of the power grid. The Intelligent Electronic Devices (IEDs) measure voltage magnitudes and phase angles, and send these information via Data Concentrators to the Control Center.

We studied a similar attack scenario in [3] using the linearized DC power flows. In a recent extension [4], the methods in [3] were modified to statistically recover the information under the AC power flows. However, due to the inaccuracy of the DC power flow approximation, the methods in [4] could not guarantee the correct information recovery under the AC power flows.

In this paper, we *directly leverage the properties of the nonlinear AC power flows* to detect the line failures and recover the voltages after an attack with the guarantee of

S. Soltan is with the Department of Electrical Engineering, Princeton University, NJ, 08544, and G. Zussman is with the Department of Electrical Engineering, Columbia University, New York, NY, 10027.
E-mails: ssoltan@princeton.edu, gil@ee.columbia.edu

performance. In particular, we introduce the EXPress line failure detection using partially ObSErved information (EXPOSE) Algorithm and demonstrate that it outperforms the state-of-the-art algorithm for detecting line failures using partial information under the AC power flows in terms of scalability and accuracy. The main advantages of the EXPOSE Algorithm are that *its running time is independent of the size of the grid and number of line failures,* and that it provides accurate information recovery under some conditions on the attacked zone. Moreover, it approximately recovers the information and provides the confidence of the solution when these conditions do not hold.

Most of the related work rely on the DC power flows and deploy brute force search approaches. These approaches do not scale well, and therefore, are limited only to detecting single and double line failures using partial measurements [5], [6], [7], [8], [9]. To represent these approaches and for comparison purposes only, we also introduce a naive Brute Force Search (BFS) Algorithm for detecting line failures after the attack.

Finally, while we analytically prove that the EXPOSE Algorithm guarantees to accurately recover the voltages and detect line failures under some conditions, we also numerically evaluate its performance when those conditions do not hold. In particular, we evaluate the performance of the EXPOSE Algorithm as the attacked zone becomes topologically more complex and compare its running time to the BFS Algorithm by considering all single, double, and triple line failures in 5 nested attacked zones. Based on the simulation results, we conclude that despite its accuracy, the BFS Algorithm is not practical for line failures detection in large networks and that the EXPOSE Algorithm can provide relatively accurate results exponentially faster. For example, the EXPOSE algorithm recovers the voltages with less than 15% error and detects line failures with less than 1 false negative on average, after all single, double, and triple line failures in an attacked zone that satisfies none of the conditions for the accuracy of the EXPOSE Algorithm.

II. Related Work

Power grids vulnerability to failures and attacks has been widely studied [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. In particular, false data injection attacks on power grids and anomaly detection have been studied using the DC power flows in [20], [21], [22], [23], [24], [25]. These studies focused on the observability of the failures and attacks in the grid.

The problem studied in this paper is similar to the problem of line failures detection using phase angle measurements [5], [6], [7], [26]. Up to two line failures detection, under the DC power flow model, was studied in [5], [6]. Since the provided methods in [5], [6] are greedy-based methods that need to search the entire failure space, the running time of these methods grows exponentially as the number of failures increases. Hence, these methods cannot be generalized to detect higher order failures. Similar greedy approaches with likelihood detection functions were studied in [8], [9], [27], [28], [29] to address the PMU placement problem under the DC power flow model.

The problem of line failures detection in an area in the power grid using the information available outside of the area was first studied in [7] based on the DC power flow model. The proposed algorithm works for only one and two line failures, since it depends on the sparsity of line failures.

In a recent work [26], a linear multinomial regression model was proposed as a classifier for a single line failure detection using transient voltage phase angles data. Due to the time complexity of the learning process for multiple line failures, this method is impractical for detecting higher order failures. Moreover, the results provided in [26] are empirical with no performance guarantees.

Finally, in a recent series of papers [30], [31], [32], the vulnerability of power grids to undetectable cyber-physical attacks is studied using the DC power flows. These studies are mainly focused on designing attacks that affect the entire grid and therefore may remain undetected.

To the best of our knowledge, our methods presented in this paper and [4] are the only methods for line failures detection under the AC power flows that can be used to detect any number of line failures and scale well with size of the grid. However, the EXPOSE Algorithm provided in this paper is more accurate than the method provided in [4].

III. Model and Definitions

A. AC Power Flow Equations

A power grid with n nodes (buses) and m transmission lines can be represented by an undirected graph $\mathcal{G}(\mathcal{N}, \mathcal{L})$, where $\mathcal{N} = \{1, 2, \ldots, n\}$ denotes the set of nodes and $\mathcal{L} = \{l_1, l_2, \ldots, l_m\}$ denotes the set of lines or edges. In the steady-state, the status of each node i is represented by its voltage $V_i = |V_i|e^{\mathbf{i}\theta_i}$ in which $|V_i|$ is the voltage magnitude, θ_i is the voltage phase angle, and $\mathbf{i}$ denotes the imaginary unit.

The goal of the AC power flow analysis is the computation of the voltage magnitudes and phase angles at each bus in the steady-state conditions [33]. In the steady-state, the AC power flow equations can be written in matrix form as follows:

$$\mathbf{Y}V = I, \tag{1}$$

$$S = \mathrm{diag}(V)I^*, \tag{2}$$

where * denotes the complex conjugation, $V = [V_1, \ldots, V_n]^{\mathrm{T}}$ is the vector of node voltages, $I = [I_1, I_2, \ldots, I_n]^T$ is the vector of injected node currents, $S = [S_1, S_2, \ldots, S_n]^T$ is the vector of injected apparent powers, and $\mathbf{Y}$ is the admittance matrix of the graph.

The elements of the admittance matrix $\mathbf{Y}$ which depend on the topology of the grid as well as the admittance values of the lines, are defined as follows:

$$Y_{ik} = \begin{cases} y_{ii} + \sum_{i \neq k} y_{ik}, & \text{if } k = i \\ -y_{ik}, & \text{if } k \in N(i) \\ 0, & \text{if } k \notin N(i) \end{cases}$$

where $N(i)$ denotes the direct neighbors of node i, y_{ik} is the equivalent admittance of the lines from node i to k, and y_{ii} is sum of the *shunt admittances* at node i. In this paper,

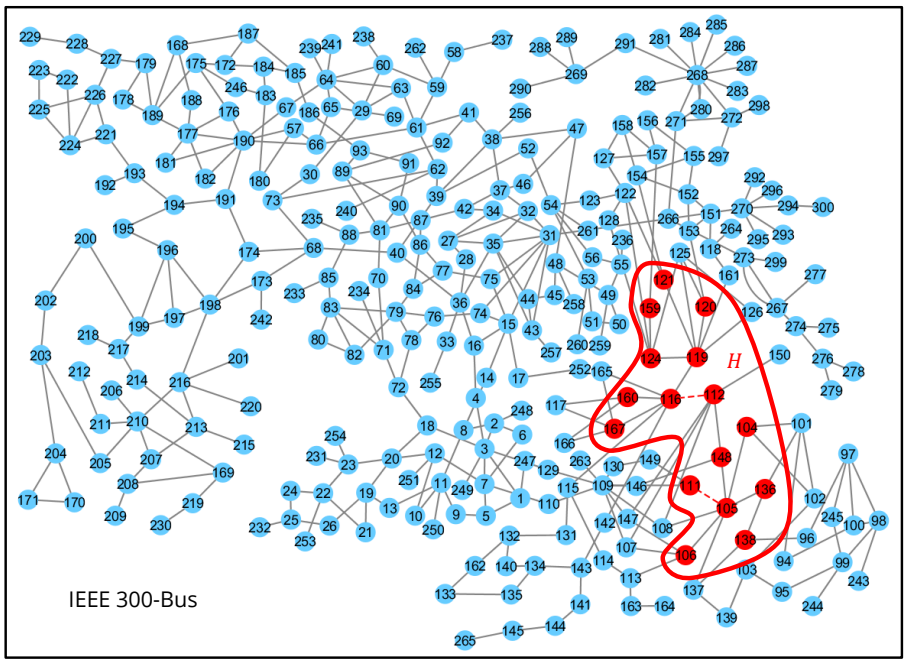


Fig. 2: The attack model. An adversary attacks a zone by disconnecting some of its lines (red dashed lines) and disallowing the information from within the zone to reach the control center. $\mathcal{G}$ is the power grid graph and $\mathcal{H}$ is a subgraph of $\mathcal{G}$ that represents the attacked zone.

we assume that the shunt admittances are negligible,[1] and therefore, $y_{ii} = 0$ for all $i \in \mathcal{N}$. The admittance matrix can also be written in term of its real and imaginary parts as $\mathbf{Y} = \mathbf{G} + \mathbf{i}\mathbf{B}$ where $\mathbf{G}$ and $\mathbf{B}$ are real matrices. Using this and the definition of the apparent power $S_i = P_i + \mathbf{i}Q_i$ in (1-2) results in the equations for the active power P_i and the reactive power Q_i at each node i as well.

B. Incidence Matrix

Under an arbitrary direction assignment to the edges of $\mathcal{G}$, the *incidence matrix* of $\mathcal{G}$ is denoted by $\mathbf{D} \in \{-1, 0, 1\}^{n \times m}$ and defined as,

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

For each line $l_j = (i, k)$, define $y_{l_j} := y_{ik}$. It can be verified that $\mathbf{Y} = \mathbf{D}\text{diag}([y_{l_1}, y_{l_2}, \ldots, y_{l_m}])\mathbf{D}^T$. As we demonstrate in Section IV, the incidence matrix is a very useful matrix for detecting line failures in power grids.

C. Basic Graph Theoretical Terms

Matching: A *matching* in a graph is a set of pairwise nonadjacent edges. If $\mathcal{M}$ is a matching, the two ends of each edge of $\mathcal{M}$ are said to be *matched* under $\mathcal{M}$, and each vertex incident with an edge of $\mathcal{M}$ is said to be *covered* by $\mathcal{M}$. In particular, a matching $\mathcal{M}$ between two subsets of the nodes like $\mathcal{A}$ and $\mathcal{B}$ that covers $\mathcal{A}$ is a set of pairwise nonadjacent edges of the graph such that: (i) each edge in $\mathcal{M}$ connects a node in $\mathcal{B}$ to a node in $\mathcal{A}$ and (ii) for each node in $\mathcal{A}$, there exists exactly one edge in $\mathcal{M}$ that is incident to that node. Hence, number of edges in $\mathcal{M}$ is exactly equal to the number of nodes in $\mathcal{A}$.

Cycle: A *cycle* in a graph is a sequence of its distinct nodes $u_1, u_2, \ldots, u_k$ such that for all $i < k$, $\{u_i, u_{i+1}\} \in \mathcal{L}$, and also $\{u_k, u_1\} \in \mathcal{L}$. A graph with no cycle is called *acyclic*.

D. Attack Model and Problem Formulation

We assume that an adversary attacks an area by: (i) disconnecting some lines within that area (*failed lines*), and (ii) obstructing all the information (e.g., status of the lines and voltage measurements) from within the area to reach the control center. We call this area, the *attacked zone*.

Fig. 2 shows an example of an attack on the area represented by $\mathcal{H} = (\mathcal{N}_\mathcal{H}, \mathcal{L}_\mathcal{H})$. We denote the set of failed lines in the attacked zone $\mathcal{H}$ by $\mathcal{F} \subseteq \mathcal{L}_\mathcal{H}$. Upon failure, the failed lines are removed from the graph and the flows are redistributed according to the AC power flows. *Our objective is to estimate the voltages and detect the failed lines inside the attacked zone using the changes in the voltages outside of the zone.*

We use the prime symbol ($'$) to denote the values after an attack (e.g., $\mathbf{Y}'$ denotes the admittance matrix of the grid and V' denotes node voltages after the attack) and $\bar{\mathcal{H}}$ to denote the set of nodes outside of the attacked zone. We also use subscripts $X_\mathcal{H}$ and $X_{\bar{\mathcal{H}}}$ to denote entries of vector X inside and outside of the attacked zone, respectively (e.g., $V_\mathcal{H}$ and $V_{\bar{\mathcal{H}}}$ denote voltage of the nodes inside and outside of the attacked zone, respectively).

Using this notation, given $V'_{\bar{\mathcal{H}}}$ and S', we want to recover $V'_\mathcal{H}$ and $\mathcal{F}$. We assume that $V'_{\bar{\mathcal{H}}}$ is measured directly by the PMUs outside of the attacked zone. It is reasonable to assume

[1]The shunt values are usually small (specially for the short lines). Here, for simplifying some of the theoretical constraints, we neglect these values. However, presented detection methods can be effectively used in practice and in the presence of shunt admittances.

that we know $S'_{\mathcal{H}}$ after the attack, since in a short time-scale from before to after the attack, we can assume that P' and Q' values remain unchanged and therefore $S' = S$.[2]

Detecting line failures after such an attack is crucial for maintaining the stability of the grid, since they may result in further line overloads and failures, if the proper load shedding mechanism is not applied. An effective load shedding requires the exact knowledge of the topology of the grid.

Notation. For any complex number $x = \mathrm{Re}(x) + \mathbf{i}\ \mathrm{Im}(x)$, real numbers $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ denote its real and imaginary values, respectively. For a vector X, $\mathrm{supp}(X)$ denotes the set of its nonzero entries. If $\mathcal{H}_1, \mathcal{H}_2$ are two subgraphs of $\mathcal{G}$, $\mathbf{Y}_{\mathcal{H}_1|\mathcal{H}_2}$ denotes the submatrix of $\mathbf{Y}$ with rows from $V_{\mathcal{H}_1}$ and columns from $V_{\mathcal{H}_2}$. Moreover, $\mathbf{Y}_{\mathcal{H}_1}$ denotes the submatrix of $\mathbf{Y}$ with all the rows associated with $V_{\mathcal{H}_1}$. For instance, $\mathbf{Y}$ can be written in any of the following forms,

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

Finally, $\mathbf{D}_{\mathcal{H}}$ is the $|\mathcal{N}_{\mathcal{H}}|$ by $|\mathcal{L}_{\mathcal{H}}|$ incidence matrix of $\mathcal{H}$. Notice that $\mathbf{D}_{\mathcal{H}}$ is known since it denotes the incidence matrix of $\mathcal{H}$ before the attack.

IV. State Estimation

In this section, we provide the analytical building blocks of the EXPOSE Algorithm which can be used to estimate the state of the grid following a cyber-physical attack. Notice that the state estimation problem considered here is different from the classical state estimation problem in power grids. Specifically, in addition to estimating the voltage magnitudes and phase angles in the attacked area, the algorithm needs to estimate the modified grid topology. However, here we mainly focus on the cases that the measurements are *noise-free.*

The idea of the approaches used in Subsections IV-A and IV-B for recovering the voltages and detecting line failures inside the attacked zone are similar to the approaches used in [3]. However, due to the nonlinear relationship between the voltages and the active/reactive powers in the AC power flows, extending those approaches to the AC power flows are more challenging. Moreover, due to this nonlinearity, those approaches are not as easily extendable for simultaneous voltage recovery and line failures detection as in [3]. We will discuss these difficulties and how to deal with them in details in Subsection IV-C.

In order to be coherent, in the final subsection of this section, we briefly describe the extension of the provided methods to noisy scenarios.

A. Voltage Recovery

Here, we provide a method to recover the voltages inside that attacked zone after the attack.

Observation 1: The admittance matrix of the grid does not change outside of the attacked zone (i.e., $\mathbf{Y}_{\bar{\mathcal{H}}} = \mathbf{Y}'_{\bar{\mathcal{H}}}$).

[2]Notice that since the grid operators do not have a direct access to power supply (controlled by the utilities) and the demand (controlled by the consumer) values, we can assume that the attacker cannot get access to these values either. Hence, the attacker cannot change these values directly.

Proof: Since the line failures only happen inside $\mathcal{H}$, following the definition of the admittance matrix (see Section III), after the attack only the entries of $\mathbf{Y}_{\mathcal{H}|\mathcal{H}}$ change. Hence, $\mathbf{Y}_{\bar{\mathcal{H}}}$ remains unchanged. ∎

From Observation 1 and using (1), we have:

$$\begin{aligned} &\mathbf{Y}'_{\bar{\mathcal{H}}} V' = I'_{\bar{\mathcal{H}}} \Rightarrow \mathbf{Y}_{\bar{\mathcal{H}}} V' = I'_{\bar{\mathcal{H}}} \Rightarrow \mathbf{Y}^*_{\bar{\mathcal{H}}} V'^* = I'^*_{\bar{\mathcal{H}}} \\ \Rightarrow &\mathrm{diag}(V'_{\bar{\mathcal{H}}}) \mathbf{Y}^*_{\bar{\mathcal{H}}} V'^* = \mathrm{diag}(V'_{\bar{\mathcal{H}}}) I'^*_{\bar{\mathcal{H}}} \\ \Rightarrow &\mathrm{diag}(V'_{\bar{\mathcal{H}}}) \mathbf{Y}^*_{\bar{\mathcal{H}}} V'^* = S'_{\bar{\mathcal{H}}} \\ \Rightarrow &\mathrm{diag}(V'_{\bar{\mathcal{H}}}) \mathbf{Y}^*_{\bar{\mathcal{H}}|\bar{\mathcal{H}}} V'^*_{\bar{\mathcal{H}}} + \mathrm{diag}(V'_{\bar{\mathcal{H}}}) \mathbf{Y}^*_{\bar{\mathcal{H}}|\mathcal{H}} V'^*_{\mathcal{H}} = S'_{\bar{\mathcal{H}}}. \end{aligned} \quad (3)$$

Notice that in (3) all the variables are known after the attack except $V'^*_{\mathcal{H}}$. Define $E_{\bar{\mathcal{H}}} := -\mathbf{Y}^*_{\bar{\mathcal{H}}|\bar{\mathcal{H}}} V'^*_{\bar{\mathcal{H}}} + \mathrm{diag}(V'^{-1}_{\bar{\mathcal{H}}}) S'_{\bar{\mathcal{H}}}$ which can be computed from the given variables after the attack. Then, we can separate the real and imaginary parts of (3) using block matrices as follows:

$$\begin{bmatrix} \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}} & -\mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}} \\ \mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}} & \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}} \end{bmatrix} \begin{bmatrix} \mathrm{Re}(V'_{\mathcal{H}}) \\ \mathrm{Im}(V'_{\mathcal{H}}) \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(E_{\bar{\mathcal{H}}}) \\ -\mathrm{Im}(E_{\bar{\mathcal{H}}}) \end{bmatrix}. \quad (4)$$

One can see that $\mathrm{Re}(V'_{\mathcal{H}})$ and $\mathrm{Im}(V'_{\mathcal{H}})$ can be uniquely recovered, if the matrix on the left hand side of (4) has full column rank. The following lemma provides the connection between the rank of that matrix and the topology of the grid.

Lemma 1: If there is a matching between the nodes in $\bar{\mathcal{H}}$ and $\mathcal{H}$ that covers the nodes in $\mathcal{H}$, then the following matrix has full column rank almost surely,[3]

$$\mathbf{M} := \begin{bmatrix} \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}} & -\mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}} \\ \mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}} & \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}} \end{bmatrix}.$$

Proof: Suppose $\mathcal{U} \subseteq \mathcal{N}_{\bar{\mathcal{H}}}$ are the matched nodes which are in $\bar{\mathcal{H}}$. Since the matching covers $\mathcal{H}$, thus $|\mathcal{U}| = |\mathcal{N}_{\mathcal{H}}|$. To show that $\mathbf{M}$ has full column rank, we show that

$$\det(\mathbf{M}_{U|H}) := \det \begin{bmatrix} \mathbf{G}_{\mathcal{U}|\mathcal{H}} & -\mathbf{B}_{\mathcal{U}|\mathcal{H}} \\ \mathbf{B}_{\mathcal{U}|\mathcal{H}} & \mathbf{G}_{\mathcal{U}|\mathcal{H}} \end{bmatrix} \neq 0,$$

almost surely. $\det(\mathbf{M}_{\mathcal{U}|\mathcal{H}})$ can be considered as a polynomial in terms of the entries of $\mathbf{M}_{\mathcal{U}|\mathcal{H}}$ using Leibniz formula. Now assume $\mathcal{U} = \{u_1, u_2, \ldots, u_{|\mathcal{N}_{\mathcal{H}}|}\}$ are matched to $\mathcal{N}_{\mathcal{H}} = \{v_1, v_2, \ldots, v_{|\mathcal{N}_{\mathcal{H}}|}\}$ in order. It can be seen that $\prod_{i=1}^{|\mathcal{N}_{\mathcal{H}}|} G^2_{u_i v_i}$ and $\prod_{i=1}^{|\mathcal{N}_{\mathcal{H}}|} B^2_{u_i v_i}$ are two terms with nonzero coefficient in $\det(\mathbf{M}_{\mathcal{U}|\mathcal{H}})$. Therefore, $\det(\mathbf{M}_{\mathcal{U}|\mathcal{H}})$ is a nonzero polynomial in terms of its entries. Now since the set of roots of a nonzero polynomial is a measure zero set in the real space, thus $\det(\mathbf{M}_{\mathcal{U}|\mathcal{H}}) \neq 0$ almost surely. ∎

Corollary 1: If there is a matching between the nodes in $\bar{\mathcal{H}}$ and $\mathcal{H}$ that covers the nodes in $\mathcal{H}$, then $V'_{\mathcal{H}}$ can be recovered almost surely.

B. Line Failures Detection

Assume $V'_{\mathcal{H}}$ is successfully recovered using (4). In this subsection, using $V'_{\mathcal{H}}$, we provide a method to detect the set of line failures $\mathcal{F}$. In the following lemma, recall that $\mathrm{supp}(X)$ denotes the index set of nonzero entries of vector X.

[3]Although the admittance values are fixed, since these values are determined by the physical properties of the lines, they are always accompanied with precision error. Hence, for a given grid, they can be considered as random around a mean value.

Lemma 2: There exists a complex vector $X \in \mathbb{C}^{|\mathcal{L}_\mathcal{H}|}$ such that

$$\mathbf{Y}_\mathcal{H} V' = I'_\mathcal{H} + \mathbf{D}_\mathcal{H} X, \tag{5}$$

and $\text{supp}(X) = \mathcal{F}$. Moreover, the vector X is unique if, and only if, $\mathbf{D}_\mathcal{H}$ has full column rank.

Proof: Without loss of generality assume $\mathcal{F} = \{l_1, l_2, \ldots, l_k\}$. Also assume that $\mathbf{0}_n$ denotes an all zero n by 1 vector. It can be seen that $\mathbf{Y}' = \mathbf{Y} - \mathbf{D}\text{diag}([y_{l_1}, y_{l_2}, \ldots, y_{l_k}, \mathbf{0}_{m-k}])\mathbf{D}^T$. Hence,

$$\begin{aligned}\mathbf{Y}'V' &= \mathbf{Y}V' - \mathbf{D}\text{diag}([y_{l_1}, y_{l_2}, \ldots, y_{l_k}, \mathbf{0}_{|\mathcal{L}|-k}])\mathbf{D}^T V' \\ I' &= \mathbf{Y}V' - \mathbf{D}\text{diag}([y_{l_1}, y_{l_2}, \ldots, y_{l_k}, \mathbf{0}_{|\mathcal{L}|-k}])\mathbf{D}^T V'.\end{aligned}$$

Now if we only focus on the rows associated with the nodes in $\mathcal{H}$ and ignore the multiplications by zero, it can been seen that

$$I'_\mathcal{H} = \mathbf{Y}_\mathcal{H} V' - \mathbf{D}_\mathcal{H}\text{diag}([y_{l_1}, y_{l_2}, \ldots, y_{l_k}, \mathbf{0}_{|\mathcal{L}_\mathcal{H}|-k}])\mathbf{D}_\mathcal{H}^T V'_\mathcal{H}.$$

Hence, vector $X := \text{diag}([y_{l_1}, y_{l_2}, \ldots, y_{l_k}, \mathbf{0}_{|\mathcal{L}_\mathcal{H}|-k}])\mathbf{D}_\mathcal{H}^T V'_\mathcal{H}$ satisfies (5). It can also be seen that only the entries of X that are associated with the failed lines are nonzero and therefore $\text{supp}(X) = \mathcal{F}$. In order for (5) to have a unique solution, $\mathbf{D}_\mathcal{H}$ should have full column rank. ∎

Corollary 2: There exist a real vector $X \in \mathbb{R}^{|\mathcal{L}_\mathcal{H}|}$ such that

$$\text{Re}\{\mathbf{Y}^*_\mathcal{H} V'^*\} = \text{Re}\{\text{diag}(V'_\mathcal{H})^{-1} S'_\mathcal{H}\} + \mathbf{D}_\mathcal{H} X, \tag{6}$$

and $\text{supp}(X) = \mathcal{F}$. Moreover, the vector X is unique if, and only if, $\mathbf{D}_\mathcal{H}$ has full column rank.

Proof: Using (2) and Lemma 2 gives the result. ∎

Corollary 2 indicates that the set of line failures can be detected by solving a matrix equation, if $\mathbf{D}_\mathcal{H}$ has full column rank. The following lemma provides the connection between the rank of that matrix and the topology of the attacked zone.

Lemma 3: $\mathbf{D}_\mathcal{H}$ has full column rank if, and only if, $\mathcal{H}$ is acyclic. Hence, the solution vector X to (6) is unique if, and only if, $\mathcal{H}$ is acyclic.

Proof: It is known in graph theory that $\text{rank}(\mathbf{D}_\mathcal{H}) = |\mathcal{N}_\mathcal{H}| - c$ in which c is the number of connected components of $\mathcal{H}$ [34, Theorem 2.3]. Therefore, $\mathbf{D}_\mathcal{H}$ has linearly independent columns if and only if $\mathcal{L}_\mathcal{H} = |\mathcal{N}_\mathcal{H}| - c$ which means that each connected component of $\mathbf{D}_\mathcal{H}$ is acyclic. The second part is the direct result of Lemma 2. ∎

Corollary 3: If $\mathcal{H}$ is acyclic, then the set of line failures $\mathcal{F}$ can be detected by solving (6) for X.

Corollary 3 states that the set of line failure can accurately be detected if $\mathcal{H}$ is acyclic. The importance of this result is in demonstrating that the set of line failures can be efficiently detected by solving a matrix equation, independently of the number of line failures.

We can use a similar idea as in [3] to extend this approach to the case in which $\mathcal{H}$ is not acyclic. If we assume that the set of line failures are sparse compared to the total number of lines in $\mathcal{H}$, we can detect line failures by finding the solution of the following optimization problem instead:

$$\begin{aligned}&\underset{X \in \mathbb{R}^{|\mathcal{L}_\mathcal{H}|}}{\text{minimize}} \|X\|_1 \text{ s.t.} \\ &\text{Re}\{\mathbf{Y}^*_\mathcal{H} V'^*\} = \text{Re}\{\text{diag}(V'_\mathcal{H})^{-1} S'_\mathcal{H}\} + \mathbf{D}_\mathcal{H} X.\end{aligned} \tag{7}$$

Notice that optimization problem (7) can be solved efficiently using Linear Programming (LP).

Lemma 4: If $\mathcal{H}$ is a cycle and less than half of its edges are failed, then the solution X to the optimization problem (7) is unique and $\text{supp}(X) = \mathcal{F}$.

Proof: The idea of the proof is similar to the idea used in the proof in [3, Lemma 4]. Without loss of generality, assume $\mathbf{D}_\mathcal{H}$ denotes the incidence matrix of $\mathcal{H}$ when edges of the cycle are directed clockwise. Since $\mathcal{H}$ is connected, $\text{rank}(\mathbf{D}_\mathcal{H}) = |\mathcal{N}_\mathcal{H}| - 1$ [34, Theorem 2.2]. Therefore, $\dim(\text{null}(\mathbf{D}_\mathcal{H})) = 1$. Suppose $\mathbb{1} \in \mathbb{R}^{|\mathcal{L}_\mathcal{H}|}$ is the all one vector. It can be seen that $\mathbf{D}_\mathcal{H}\mathbb{1} = 0$. Since $\dim(\text{null}(\mathbf{D}_\mathcal{H})) = 1$, span of $\mathbb{1}$ is the null space of $\mathbf{D}$. From Lemma 2, there exists a solution X to (7) such that $\text{supp}(X) = \mathcal{F}$. To prove that X is the unique solution for (7), we prove that $\forall c \in \mathbb{R}\backslash\{0\}$, $\|X\|_1 < \|X - c\mathbb{1}\|_1$. Without loss of generality assume that $x_1, x_2, \ldots, x_k$ are the nonzero elements of X, in which $k = |\mathcal{F}|$. From the assumption, we know that $k < |\mathcal{L}_\mathcal{H}|/2$. Hence,

$$\begin{aligned}\|X - c\mathbb{1}\|_1 &= \sum_{i=1}^{k} |x_i - c| + (|\mathcal{L}_\mathcal{H}| - k)|c| \\ &= \sum_{i=1}^{k} (|x_i - c| + |c|) + (|\mathcal{L}_\mathcal{H}| - 2k)|c| \\ &\geq \sum_{i=1}^{k} |x_i| + (|\mathcal{L}_\mathcal{H}| - 2k)|c| > \sum_{i=1}^{k} |x_i| = \|X\|_1.\end{aligned}$$

Thus, the solution to (7) is unique and $\text{supp}(X) = \mathcal{F}$. ∎

Lemma 4 can be extended to planar graphs similar to the result for the DC power flows presented in [3, Theorem 2]. The proof and the argument in [3] should apply with a very slight change here as well. To avoid repetition, we do not present a similar Lemma here.

C. Simultaneous Recovery and Detection

In order to extend our approach to the cases that (4) does not have a unique solution, we can solve (4) and (7) at the same time. Therefore, in order to recover the voltages and detect the line failures at the same time, one needs to solve the following optimization problem:

$$\begin{aligned}&\underset{X \in \mathbb{R}^{|\mathcal{L}_\mathcal{H}|}, V'_\mathcal{H} \in \mathbb{C}^{|\mathcal{N}_\mathcal{H}|}}{\text{minimize}} \|X\|_1 \text{ s.t.} \\ &\mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}}\text{Re}(V'_\mathcal{H}) - \mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}}\text{Im}(V'_\mathcal{H}) = \text{Re}(E_{\bar{\mathcal{H}}}) \\ &\mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}}\text{Re}(V'_\mathcal{H}) + \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}}\text{Im}(V'_\mathcal{H}) = -\text{Im}(E_{\bar{\mathcal{H}}}) \\ &\text{Re}\{\mathbf{Y}^*_\mathcal{H} V'^*\} = \text{Re}\{\text{diag}(V'_\mathcal{H})^{-1} S'_\mathcal{H}\} + \mathbf{D}_\mathcal{H} X.\end{aligned} \tag{8}$$

However, since $V'_\mathcal{H}$ is part of the variables, this optimization problem is not linear and convex anymore.[4] To resolve this issue, we need to approximate $\text{diag}(V'_\mathcal{H})^{-1}$ with a linear function in terms of $V'_\mathcal{H}$. For this, we have:

$$\text{diag}(V'_\mathcal{H})^{-1} = \text{diag}(|V'_\mathcal{H}|)^{-2}\big(\text{diag}(\text{Re}(V'_\mathcal{H})) - \mathbf{i}\ \text{diag}(\text{Im}(V'_\mathcal{H}))\big).$$

On the other hand, the voltage magnitudes are almost constant at each node before and after the failure ($|V'_\mathcal{H}| \approx |V_\mathcal{H}|$), hence:

$$\text{diag}(V'_\mathcal{H})^{-1} \approx \text{diag}(|V_\mathcal{H}|)^{-2}\big(\text{diag}(\text{Re}(V'_\mathcal{H})) - \mathbf{i}\ \text{diag}(\text{Im}(V'_\mathcal{H}))\big).$$

[4] This nonlinearity does not appear under the DC power flow model. For more details see [3].

We can use the approximation above in optimization (8) in order to relax its nonconvexity. Notice that since optimization (8) is for the cases in which the solution to (4) is not unique, and therefore the voltages cannot be recovered uniquely, some conditions should be placed on the voltages such that the recovered voltages are near operable conditions. To do so, we add a convex constraint on the voltage magnitudes of the nodes in $\mathcal{H}$ after the attack as $|V'_{\mathcal{H}}| \leq 1.1\mathbb{1}_{\mathcal{H}}$,[5] in which $\mathbb{1}_{\mathcal{H}}$ is an all ones vector of size $|\mathcal{N}_{\mathcal{H}}|$. Hence, the following convex optimization can be used to detect the set of line failures and recover the voltages when the solution to (4) is not unique:

$$\begin{aligned}
&\underset{X\in\mathbb{R}^{|\mathcal{L}_{\mathcal{H}}|},V'_{\mathcal{H}}\in\mathbb{C}^{|\mathcal{N}_{\mathcal{H}}|}}{\text{minimize}} \ \|X\|_1 \text{ s.t.}\\
&\mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}}\mathrm{Re}(V'_{\mathcal{H}}) - \mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}}\mathrm{Im}(V'_{\mathcal{H}}) = \mathrm{Re}(E_{\bar{\mathcal{H}}})\\
&\mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}}\mathrm{Re}(V'_{\mathcal{H}}) + \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}}\mathrm{Im}(V'_{\mathcal{H}}) = -\mathrm{Im}(E_{\bar{\mathcal{H}}})\\
&|V'_{\mathcal{H}}| \leq 1.1\mathbb{1}_{\mathcal{H}}\\
&\|\mathrm{Re}\{\mathbf{Y}^*_{\mathcal{H}}V'^*\} - \mathrm{diag}(|V_{\mathcal{H}}|)^{-2}\mathrm{diag}(\mathrm{Re}(V'_{\mathcal{H}}))P'_{\mathcal{H}}\\
&\qquad - \mathrm{diag}(|V_{\mathcal{H}}|)^{-2}\mathrm{diag}(\mathrm{Im}(V'_{\mathcal{H}}))Q'_{\mathcal{H}} - \mathbf{D}_{\mathcal{H}}X\|_2 \leq \epsilon. \quad (9)
\end{aligned}$$

In optimization (9), the ϵ value controls the approximation accuracy of $\mathrm{diag}(V'_{\mathcal{H}})^{-1}$ using $|V_{\mathcal{H}}|$. The ϵ value can be tuned using the confidence of the solution that we will define in the next subsection. In Section VII, we evaluate the accuracy of the results obtained by solving the convex optimization problem (9) as part of the EXPOSE Algorithm.

D. Confidence of the Solution

Once the set of line failures is detected and the voltages are recovered, one can compute the confidence of the solution using (1-2). Assume $\mathbf{Y}^\dagger$ and $V^\dagger$ denote the admittance matrix of the grid after removing the detected lines and the recovered voltages after the attack, respectively. If the detection and recovery are done correctly, then $\mathrm{Re}\{\mathrm{diag}(V^\dagger)^*\mathbf{Y}V^\dagger\} = P'$ and $\mathrm{Im}\{\mathrm{diag}(V^\dagger)^*\mathbf{Y}V^\dagger\} = Q'$. However, if the detection and recovery are not done correctly, these equalities do not hold. We can use the difference between the two sides of these equalities as a measure for the correctness of the solution.

We define c_P and c_Q to denote the confidence of the solution based on P and Q as follows:

$$c_P := (1 - \|\mathrm{Re}\{\mathrm{diag}(V^\dagger)^*\mathbf{Y}V^\dagger\} - P'\|_2/\|P'\|_2)^+ * 100, \quad (10)$$

$$c_Q := (1 - \|\mathrm{Im}\{\mathrm{diag}(V^\dagger)^*\mathbf{Y}V^\dagger\} - Q'\|_2/\|Q'\|_2)^+ * 100, \quad (11)$$

in which $(x)^+ := \max(0, x)$. If $c_P, c_Q \approx 100\%$, then the solution is reliable. If not, depending on the c_P or c_Q values, one can see how closely the solution fits the observed data.

E. Noisy Measurements

Although the main focus of this paper is on the information recovery after an attack when the measurements are noise-free (which is a reasonable assumption given the new generation of the PMUs), we briefly describe how some of the techniques provided in the previous subsections can be extended to noisy scenarios. The noise can also represent the small changes in P' and Q' that we assumed to remain unchanged after an attack (see Section III-D).

The extension to the cases in which the measurements are noisy is very straight forward and standard. The equality constraints (4) and (6) that appear also in optimization (9) when matrix M is not full-rank, should be replaced with the following relaxed versions,

$$\left\| \begin{bmatrix} \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}} & -\mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}} \\ \mathbf{B}_{\bar{\mathcal{H}}|\mathcal{H}} & \mathbf{G}_{\bar{\mathcal{H}}|\mathcal{H}} \end{bmatrix} \begin{bmatrix} \mathrm{Re}(V'_{\mathcal{H}}) \\ \mathrm{Im}(V'_{\mathcal{H}}) \end{bmatrix} - \begin{bmatrix} \mathrm{Re}(E_{\bar{\mathcal{H}}}) \\ -\mathrm{Im}(E_{\bar{\mathcal{H}}}) \end{bmatrix} \right\|_2 < \epsilon_1,$$

$$\|\mathrm{Re}\{\mathbf{Y}^*_{\mathcal{H}}V'^*\} - \mathrm{Re}\{\mathrm{diag}(V'_{\mathcal{H}})^{-1}S'_{\mathcal{H}}\} - \mathbf{D}_{\mathcal{H}}X\|_2 \leq \epsilon_2.$$

The ϵ_1 and ϵ_2 can then be tuned using confidence metrics c_P and c_Q, or they can be considered as part of the objective function along with $\|X\|_1$ in optimization (9). In the numerical results section, we only consider noise-free scenarios but exploring the noisy scenarios in detail is part of our future work.

Algorithm 1: EXPress line failure detection using partially ObSERved information (EXPOSE)

Input: A connected graph $\mathcal{G}$, attacked zone $\mathcal{H}$, V, $V'_{\mathcal{H}}$, and S'

1: **if** $\mathbf{M}$ is full rank **then**
2: Solve (4) to recover $V'_{\mathcal{H}}$
3: Use the recovered $V'_{\mathcal{H}}$ in (7) to find the solution vector X
4: **else**
5: Find the solution $V'_{\mathcal{H}}$ and X to the optimization (9)
6: Compute $\mathcal{F} = \mathrm{supp}(X)$
7: Compute the confidence of the solution c_P and c_Q
8: **return** $V'_{\mathcal{H}}, \mathcal{F}, c_P$, and c_Q

V. EXPOSE Algorithm

In this section, using the results provided in Section IV, we introduce the EXPress line failure detection using partially ObSERved information (EXPOSE) Algorithm. The EXPOSE Algorithm is summarized in Algorithm 1.

In the first step, the algorithm checks if matrix M is full rank (i.e., there is a matching between the nodes in $\bar{\mathcal{H}}$ and $\mathcal{H}$ that covers the nodes in H). If yes, it solves (4) to recover $V'_{\mathcal{H}}$ and then uses these recovered values in (7) to find the solution vector X. Else, it finds the solution $V'_{\mathcal{H}}$ and X to optimization (9) instead. Then, based on the obtained vector X, it detects the set of line failures by $\mathcal{F} = \mathrm{supp}(X)$. Finally, it computes the confidence of the solution metrics c_P and c_Q using (10-11) and returns $V'_{\mathcal{H}}, \mathcal{F}, c_P$, and c_Q.

Notice that for solving (4), (7), and (9) only the voltages of the nodes that are at most one hop away from the nodes in $\mathcal{H}$ are required. Hence, not only the running time of the EXPOSE algorithm is independent of the number of line failures, but also it is independent of the size of the entire grid. This makes the EXPOSE Algorithm suitable for detecting line failures in large networks.

VI. Brute Force Algorithm

In order to compare the performance of the EXPOSE Algorithm with the previous studies that were mostly conducted

[5]The 1.1 per unit upper bound on the voltages is to ensure that (9) is feasible. However, it can be replaced by more system specific upper bounds.

Algorithm 2: Brute Force Search (BFS)

Input: A connected graph $\mathcal{G}$, attacked zone $\mathcal{H}$, V, $V'_{\mathcal{H}}$, and S'

1: **for** Any $\mathcal{F}^\dagger \subseteq \mathcal{L}_{\mathcal{H}}$ **do**
2: Compute $V^\dagger$ after removing the lines in $\mathcal{F}^\dagger$ from $\mathcal{G}$
3: Compute $e_{\mathcal{F}^\dagger} = \|\mathrm{Re}(V^\dagger_{\mathcal{H}} - V'_{\mathcal{H}})\|_2 + \|\mathrm{Im}(V^\dagger_{\mathcal{H}} - V'_{\mathcal{H}})\|_2$
4: Select $\mathcal{F} = \arg\min_{\mathcal{F}^\dagger \subseteq \mathcal{L}_{\mathcal{H}}} e_{\mathcal{F}^\dagger}$
5: **return** $\mathcal{F}$

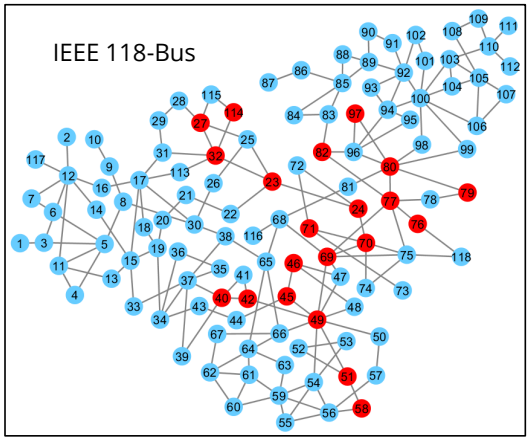


Fig. 3: An attacked zone used in [4], shown by red nodes, in the IEEE 118-bus system.

under the DC power flow model, we introduce the Brute Force Search (BFS) Algorithm for detecting line failures after the attack. The BFS Algorithm considers all possible line failure scenarios and returns the most likelihood solution. This method is the naive version of the approaches used in [5], [6], [7], [8], [9] in similar settings to detect line failures given a partial phase angle measurements under the DC power flow model.

The idea is to compute the voltages $V^\dagger_{\mathcal{H}}$ for any possible set of line failures $\mathcal{F}^\dagger \subseteq \mathcal{L}_{\mathcal{H}}$ and detect the one that is closest to $V'_{\mathcal{H}}$ as the most likely failure as follows:

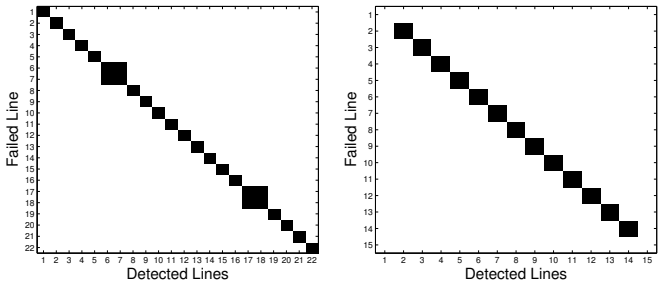
$$\mathcal{F} = \underset{\mathcal{F}^\dagger \subseteq \mathcal{L}_{\mathcal{H}}}{\arg\min} \|\mathrm{Re}(V^\dagger_{\mathcal{H}} - V'_{\mathcal{H}})\|_2 + \|\mathrm{Im}(V^\dagger_{\mathcal{H}} - V'_{\mathcal{H}})\|_2. \quad (12)$$

The BFS Algrithm is summarized in Algorithm 2. Notice that the BFS Algorithm is exponentially slower than the EXPOSE Algorithm, since it requires to solve the AC power flow solutions $2^{|\mathcal{L}_{\mathcal{H}}|}$ times. Moreover, since it requires to solve the power flow equations for the entire grid, unlike the EXPOSE Algorithm, its running time increases polynomially with the size of the grid.

The main shortcoming of the BFS Algorithm is its intractability for large networks. One way to speed up the BFS Algorithm is to stop whenever the $e_{\mathcal{F}^\dagger}$ (as defined in line 3 of the algorithm) is less than a threshold. This may speed up the process but does not solve the intractability issue.

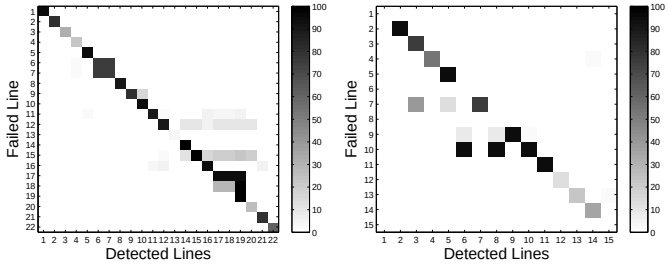
VII. Numerical Results

While in Section IV, we analytically proved that the EXPOSE Algorithm guarantees to accurately recover the voltages and detect line failures under some conditions (i.e., matched and acyclic attacked zones), in this section, we numerically confirm those results and also evaluate the EXPOSE Algorithm's performance when those conditions do not hold (i.e., general attacked zones).



(a) (b)

Fig. 4: Detected line failures after all possible single line failures using the EXPOSE Algorithm in (a) the zone shown in Fig. 3 in the IEEE 118-bus system, and (b) the zone shown in Fig. 2 in the IEEE 300-bus system.



(a) (b)

Fig. 5: Detected probability of line failures after all possible single line failures using the COPSSES Algorithm introduced in [4] in (a) the zone shown in Fig. 3 in the IEEE 118-bus system, and (b) the zone shown in Fig. 2 in the IEEE 300-bus system.

A. Matched and acyclic attacked zones

As we mentioned in Section II, to the best of our knowledge, our work in [4] is the only other method for information recovery under the AC power flows that can be used to detect any number of line failures and scales well with the size of the grid. In [4], we introduced the Convex OPtimization for Statistical State EStimation (COPSSES) Algorithm and demonstrated that when the attacked zone is matched and acyclic (i.e., matrices $\mathbf{M}$ and $\mathbf{D}_{\mathcal{H}}$ have full column rank), it can detect line failures with few errors. The COPSSES Algorithm uses a relaxation of the methods introduced in [3], which were based on DC power flow equations, for information recovery under the AC power flow equations. The advantage of the COPSSES Algorithm is that similar to the EXPOSE Algorithm, its running time is independent of the number of line failures.

In order to demonstrate the superiority of the EXPOSE Algorithm in this case, we compare its performance and running time to the COPSSES Algorithm in addition to the BFS Algorithm. For comparison purposes, we consider attacks on the same zones as considered in [4] within the IEEE 118- and 300-bus systems. The zones are depicted in Figs. 3 and 2, respectively.

Recall from subsections IV-A and IV-B that when matrices $\mathbf{M}$ and $\mathbf{D}_{\mathcal{H}}$ have full column rank, as it is the case here, the EXPOSE Algorithm can recover the voltages and detect the

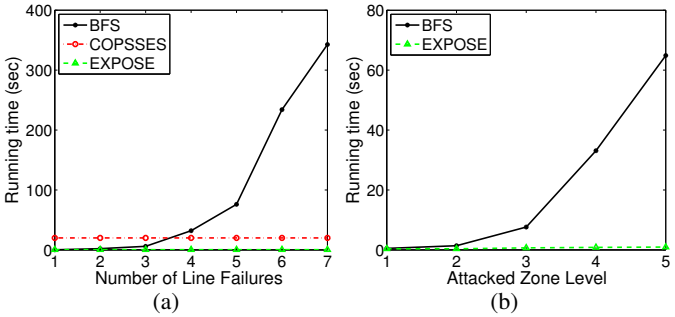


Fig. 6: Average running times of different line failures detection algorithms versus (a) the total number of line failures in the attacked zone shown in Fig. 2, and (b) the size of the attacked zone as shown in Fig. 7(a) in detecting triple line failures.

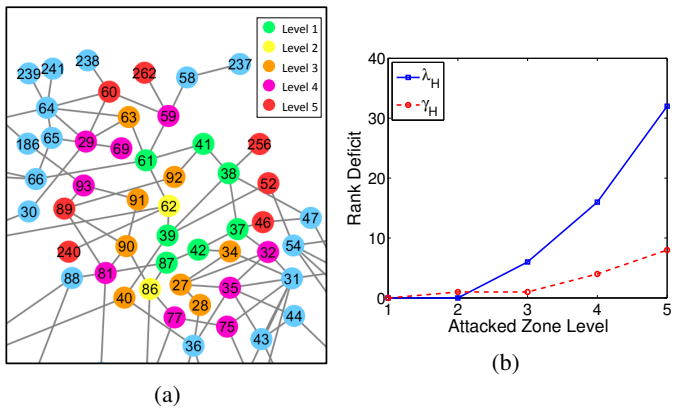


Fig. 7: Nested attacked zones in the IEEE 300-bus system. (a) Nodes corresponding to different levels are shown in different colors, and (b) the rank deficit in the attacked zones in different levels.

line failures accurately. Hence, as we expected and can be seen in Fig. 4, all the single line failures can be exactly detected using the EXPOSE Algorithm in the selected attacked zones within the IEEE 118- and 300-bus systems. Notice that the false positives in failures of lines 6 and 7 as well as 17 and 18 in the IEEE 118-bus system are due to the violation of the acyclicity of the attacked zone. Lines 6 and 7 (and also 17 and 18) are parallel lines that form a cycle with two nodes.

Moreover, lack of any detections after failures in lines 1 and 15 within that attacked zone in the 300-bus system is due to the fact that the AC power flows did not have a solution after those failures. Therefore those cases were not considered in evaluation of the EXPOSE Algorithm.

We considered up to 7 line failures in the zone depicted in Fig. 2. In all the cases, as we expected, the EXPOSE Algorithm could exactly detect the line failures. The BFS Algorithm could also detect the line failures exactly in those scenarios. However, as it was shown in [4] and also in Fig. 5 for detecting single line failures, the COPSSES Algorithm may result in few false positives and negatives in detecting single line failures (e.g., as can be seen in Fig. 5(b), lines 6, 8, and 10 are detected to be equally probable to be failed based on the COPSSES Algorithm after a failure in line 10), and more false positives and negatives as the number of line failures increases.

Fig. 6(a) compares the running times of the three algorithms in detecting line failures versus the number of line failures. As can be seen, since the running times of the EXPOSE and COPSSES Algorithms are independent of the number of line failures, they both provide a constant running time as the number of line failures increases. Moreover, since both rely on solving an LP for detecting line failures, they are both easy to implement. However, as can be seen, the running time of the BFS Algorithm increases exponentially as the total number of line failures increases.

Overall, when the attacked zone is matched and acyclic, the EXPOSE Algorithm detects line failures as accurately as the BFS Algorithm, but exponentially faster.

B. General attacked zones

In order to evaluate the performance of the EXPOSE Algorithm as the attacked zone becomes larger and topologically more complex, in this subsection, we consider 5 nested attacked zones as depicted in Fig. 7(a). We refer to the nodes that are added to the attacked zone at the i^{th} step by the *level* i nodes. The level i attacked zone is an attacked zone that consists of all the nodes in levels 1 to i.

As we proved in Section IV and briefly showed in Subsection VII-A, when matrices $\mathbf{M}$ and $\mathbf{D}_{\mathcal{H}}$ have full column rank, then the EXPOSE Algorithm can recover the voltages and detect the line failures accurately. In order to demonstrate the distance of the topological properties of an attacked zone to these conditions, we define $\lambda_{\mathcal{H}}$ and $\gamma_{\mathcal{H}}$ as follows:

$$\lambda_{\mathcal{H}} := 2n_{\mathcal{H}} - \text{rank}(\mathbf{M}),$$
$$\gamma_{\mathcal{H}} := m_{\mathcal{H}} - \text{rank}(\mathbf{D}_{\mathcal{H}}).$$

It can be verified that when matrices $\mathbf{M}$ and $\mathbf{D}_{\mathcal{H}}$ have full column rank, then $\lambda_{\mathcal{H}} = 0$ and $\gamma_{\mathcal{H}} = 0$, respectively. Hence, $\lambda_{\mathcal{H}}$ and $\gamma_{\mathcal{H}}$ indicate the *rank deficit* of matrices $\mathbf{M}$ and $\mathbf{D}_{\mathcal{H}}$.

Fig. 7(b) shows the $\lambda_{\mathcal{H}}$ and $\gamma_{\mathcal{H}}$ values for the different attacked zone levels. As can be seen, both values grow significantly in level 4 and level 5 attacked zones. This means that the data outside of the attacked zone is very insufficient to accurately detect the line failures based on the EXPOSE Algorithm in those levels.

First, in order to show the advantage of the EXPOSE Algorithm over brute force type algorithms, in Fig. 6(b), we compare the increase in the running times of the BFS and the EXPOSE Algorithms in detecting triple line failures as the number of nodes and lines increases in different levels. As can be seen in Fig. 6(b), the running time of the BFS Algorithm exponentially increases with the size of the attacked zone whereas that of the EXPOSE Algorithm only slightly increases. This along with Fig. 6(a) clearly indicates that the BFS Algorithm (and algorithms with similar approaches) do not scale well with the size of the attacked zone and the number of line failures.

In order to evaluate the performance of the EXPOSE Algorithm, we consider all single, double, and triple line failures in the nested zones. The results are presented in Fig. 8.

The average number of false negatives and positives in detecting line failures in different attacked zone levels for all single, double, and triple line failures are presented in Figs. 8(a) and 8(b). As expected, for the level 1 attacked

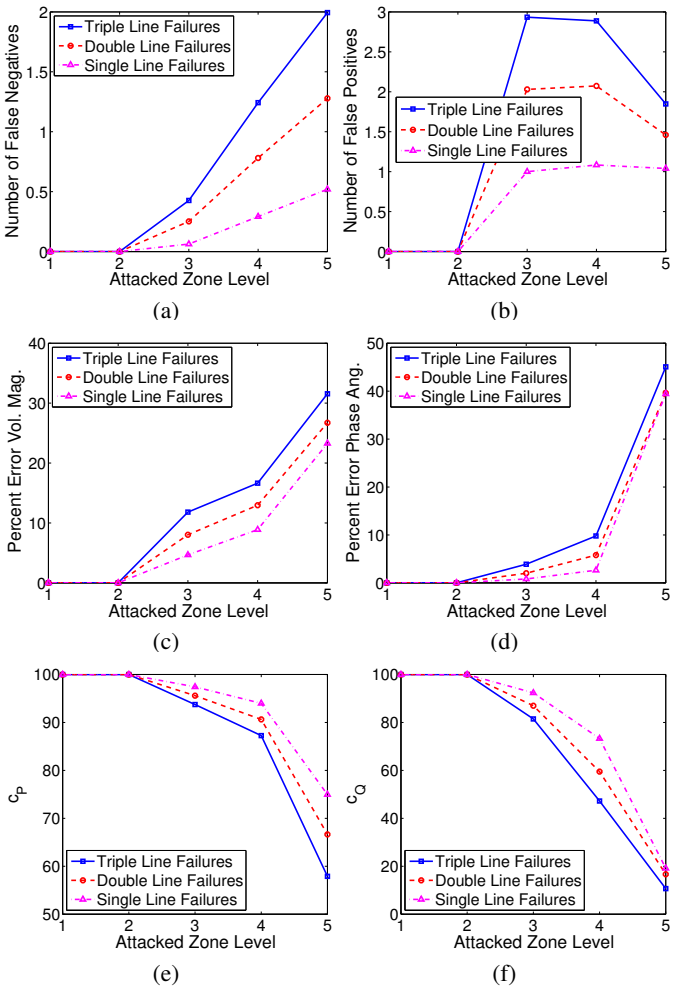


Fig. 8: The EXPOSE Algorithm's performance after all single, double, and triple line failures versus the size of the attacked zone as shown in Fig. 7(a). (a) Average number of false negatives, (b) average number of false positives, (c) average percentage error in recovered voltage magnitudes, (d) average percentage error in recovered voltage phase angles, (e) average confidence of the solutions (c_P), and (f) average confidence of the solutions (c_Q).

zone, there are no false negatives or positive. For the level 2 attacked zone also, although $\mathbf{D}_{\mathcal{H}}$ does not have full column rank (see Fig. 7(b)), the EXPOSE Algorithm can still detect the line failures accurately. However, as the attacked zone becomes larger in higher levels, and $\lambda_{\mathcal{H}}$ and $\gamma_{\mathcal{H}}$ increase, the EXPOSE Algorithm results in false positives and negatives. An important observation here is that the EXPOSE Algorithm results on average in more false positives than negatives. This is a good characteristic of the EXPOSE Algorithm, since it means that by having an extra brute force search step on the detected line failures set, one can reduce the number of false positives significantly.

The average error in recovered voltages in different attacked zone levels using the EXPOSE Algorithm for all single, double, and triple line failures are presented in Figs. 8(c) and 8(d). As can be seen, similar to the line failures detection, the EXPOSE Algorithm recovers the voltages accurately for the level 1 and level 2 attacked zones. Moreover, for the level 3 and level 4 attacked zones, the EXPOSE Algorithm recovers the voltage magnitudes and phase angles with less than 15% and 10% error, respectively. However, for the level 5 attacked zone, since $\lambda_{\mathcal{H}}$ is too high (see Fig. 7(b)), the EXPOSE Algorithm results in around 30% and 40% error in the recovered voltage magnitudes and phase angles, respectively.

Finally, the c_P and c_Q metrics introduced in subsection IV-D can be used to determine the confidence of the solutions obtained by the EXPOSE Algorithm. As can be seen in Figs. 8(e) and 8(f), the c_P and c_Q values are directly correlated with the errors in voltages and number of false negatives. Hence, these values can effectively be used to compute the confidence of the solution obtained by the EXPOSE Algorithm. Notice that c_Q is more sensitive than the c_P. Therefore, c_Q can be used as the upper bound for the error and c_P can be used as the lower bound.

We did not evaluate the performance of the BFS Algorithm here due to its very high running time (see Fig. 6(b)). However, we expect that the BFS Algorithm could detect the line failures and recover the voltages with almost no error. Despite its accuracy, the BFS Algorithm is impractical for line failures detection in large networks. As we showed in this section, the EXPOSE Algorithm can provide relatively accurate results exponentially faster than the BFS Algorithm.

VIII. Conclusion

We studied cyber-physical attacks on power grids under the AC power flows. We leveraged the algebraic properties of the AC power flows to develop the EXPOSE Algorithm for detecting line failures and recovering the voltages after the attack. We analytically proved that if the attacked zone has certain topological properties, the EXPOSE Algorithm can accurately recover the information. We also numerically demonstrated that in more complex attacked zones, it can still recover the information approximately well. The main advantages of the EXPOSE Algorithm are that its running time is independent of the size of the grid and number of line failures, and that it provides accurate information recovery under some conditions on the attacked zone. Moreover, it approximately recovers the information and provides the confidence of the solution when these conditions do not hold.

The results provided in this paper can be further used in different contexts as well. For example, the EXPOSE Algorithm can be used to detect line failures when measurement devices are scarce and not ubiquitous. Moreover, the conditions on the attacked zone such that the EXPOSE Algorithm can accurately detect the line failures and recover the voltages, can be used for optimal placement of measurement devices in the grid.

Despite its strengths, the EXPOSE Algorithm requires that the power system converges to a stable state after an attack. However, as the number of line failures increases, such an assumption may rarely hold. Therefore, the dynamics of the system after an attack should also be considered for an effective detection mechanism. Due to its complexity, studying the system's dynamics after an attack is a very challenging task. Hence, exploring this and other directions is part of our future work.

ACKNOWLEDGEMENTS

This work was supported in part by the U.S. DOE OE as part of the DOE Grid Modernization Initiative, U.S. DOE under Contract No. DE-AC36-08GO28308 with NREL, DARPA RADICS under contract #FA-8750-16-C-0054, and DTRA grant HDTRA1-13-1-0021. This work was done while Saleh Soltan was with Columbia University.

REFERENCES

[1] "Analysis of the cyber attack on the Ukrainian power grid," Mar. 2016, http://www.nerc.com/pa/CI/ESISAC/Documents/E-ISAC_SANS_Ukraine_DUC_18Mar2016.pdf.

[2] "Assault on California power station raises alarm on potential for terrorism," 2014, source: http://goo.gl/RiuhI1.

[3] S. Soltan, M. Yannakakis, and G. Zussman, "Power grid state estimation following a joint cyber and physical attack," *to appear in IEEE Trans. Control Netw. Syst.*, 2017.

[4] S. Soltan and G. Zussman, "Power grid state estimation after a cyber-physical attack under the AC power flow model," in *Proc. IEEE PES-GM'17*, July 2017.

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

[6] ——, "Double line outage detection using phasor angle measurements," in *Proc. IEEE PES-GM'09*, July 2009.

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

[8] 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-GM'12*, July 2012.

[9] H. Zhu and Y. Shi, "Phasor measurement unit placement for identifying power line outages in wide-area transmission system monitoring," in *HICSS'14*, 2014, pp. 2483–2492.

[10] A. Ashok, M. Govindarasu, and J. Wang, "Cyber-physical attack-resilient wide-area monitoring, protection, and control for the power grid," *Proc. IEEE*, vol. 105, no. 7, pp. 1389–1407, 2017.

[11] A. Pinar, J. Meza, V. Donde, and B. Lesieutre, "Optimization strategies for the vulnerability analysis of the electric power grid," *SIAM J. Optimiz.*, vol. 20, no. 4, pp. 1786–1810, 2010.

[12] T. Kim, S. J. Wright, D. Bienstock, and S. Harnett, "Analyzing vulnerability of power systems with continuous optimization formulations," *IEEE Trans. Net. Sci. Eng.*, vol. 3, no. 3, pp. 132–146, 2016.

[13] 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.

[14] 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.*, vol. 40, no. 3, pp. 33–37, 2012.

[15] I. Dobson, B. Carreras, V. Lynch, and D. Newman, "Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization," *Chaos*, vol. 17, no. 2, p. 026103, 2007.

[16] D. Bienstock, *Electrical Transmission System Cascades and Vulnerability: An Operations Research Viewpoint*. SIAM, 2016.

[17] S. Soltan, A. Loh, and G. Zussman, "Analyzing and quantifying the effect of k-line failures in power grids," *to appear in IEEE Trans. Control Netw. Syst.*, 2017.

[18] A. Srivastava, T. Morris, T. Ernster, C. Vellaithurai, S. Pan, and U. Adhikari, "Modeling cyber-physical vulnerability of the smart grid with incomplete information," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 235–244, 2013.

[19] F. Pasqualetti, F. Dörfler, and F. Bullo, "Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design," in *Proc. IEEE CDC-ECC'11*, 2011, pp. 2195–2201.

[20] J. Kim and L. Tong, "On topology attack of a smart grid: undetectable attacks and countermeasures," *IEEE J. Sel. Areas Commun*, vol. 31, no. 7, pp. 1294–1305, 2013.

[21] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, p. 13, 2011.

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

[23] 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.

[24] S. Li, Y. Yılmaz, and X. Wang, "Quickest detection of false data injection attack in wide-area smart grids," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2725–2735, 2015.

[25] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach." *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1102–1114, 2015.

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

[27] N. M. Manousakis, G. N. Korres, and P. S. Georgilakis, "Taxonomy of PMU placement methodologies," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 1070–1077, 2012.

[28] 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.

[29] Y. Zhao, J. Chen, A. Goldsmith, and H. V. Poor, "Identification of outages in power systems with uncertain states and optimal sensor locations," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 6, pp. 1140–1153, 2014.

[30] Z. Li, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Bilevel model for analyzing coordinated cyber-physical attacks on power systems," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2260–2272, 2016.

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

[32] J. Zhang and L. Sankar, "Physical system consequences of unobservable state-and-topology cyber-physical attacks," *IEEE Trans. Smart Grid*, vol. 7, no. 4, 2016.

[33] J. D. Glover, M. S. Sarma, and T. Overbye, *Power System Analysis & Design, SI Version*. Cengage Learning, 2012.

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

Saleh Soltan is a postdoctoral research associate in the department of Electrical Engineering at Princeton University. In 2017, he obtained the Ph.D. degree in Electrical Engineering from Columbia University. He received B.S. degrees in Electrical Engineering and Mathematics (double major) from Sharif University of Technology, Iran in 2011 and the M.S. degree in Electrical Engineering from Columbia University in 2012. He is the Gold Medalist of the 23rd National Mathematics Olympiad in Iran in 2005 and the recipient of Columbia University Electrical Engineering Armstrong Memorial Award in 2012.

Gil Zussman received the Ph.D. degree in electrical engineering from the Technion in 2004 and was a postdoctoral associate at MIT in 2004-2007. He is currently an Associate Professor of Electrical Engineering at Columbia University. He is a co-recipient of 7 paper awards including the ACM SIGMETRICS06 Best Paper Award, the 2011 IEEE Communications Society Award for Advances in Communication, and the ACM CoNEXT'16 Best Paper Award. He received the Fulbright Fellowship, the DTRA Young Investigator Award, and the NSF CAREER Award, and was a member of a team that won first place in the 2009 Vodafone Foundation Wireless Innovation Project competition.