Bayesian Regression for Robust Power Grid State Estimation Following a Cyber-Physical Attack

Saleh Soltan, Prateek Mittal, H. Vincent Poor
Department of Electrical Engineering
Princeton University, Princeton, NJ
Emails: {ssoltan,pmittal,poor}@princeton.edu

***Abstract*—Improving the power grid's security against cyber-physical attacks has been a major challenge for the power grid operators since the cyber attack on the Ukrainian grid in Dec. 2016. In order to partly address this challenge, we study the problem of power grid state estimation following such an attack. We assume that an adversary attacks an area by disconnecting some lines within the attacked area and blocking the measurements coming from inside the attacked area from reaching the control center in order to mask the failed lines. The objective is to use the phase angle measurements before and partial measurements after the attack to detect the failed lines. Despite recent efforts, there is no method that is both efficient and robust for estimating the state of the grid after such an attack in practical noisy settings. In this work, we provide such a method using Bayesian regression. Bayessian regression allows us to determine the probability that each line is failed, instead of a 0-1 hard decision on the status of the lines. These probabilities reflect the uncertainty in the detection, depending on the noise level. We show that these probabilities can further be used to limit the search space and significantly improve the running time of the existing brute force search methods for failed lines detection.**

I. Introduction

In this paper, we provide a novel application of Bayesian data analysis for power grid state estimation. By the *state* of the grid, we mean both its topological and operational states. We focus on the state estimation following a cyber-physical attack that has gained a lot of attention [1], [2], [3], [4], [5] since the Dec. 2016 cyber attack on the Ukrainian grid [6]. Although, here, we mainly focus on the scenarios caused by attacks, similar scenarios can result from natural disasters or the lack of enough measurement devices. So the provided methods have broad applicability for grid state estimation.

We follow [1] and assume that an adversary attacks an area by disconnecting some lines within the attacked area and blocking the measurements coming from inside the attacked area from reaching the control center in order to mask the failed lines. The objective is to use the phase angle measurements before and partial measurements after the attack to detect the failed lines. Detecting failed lines is necessary for obtaining a correct understanding of the grid topology and estimating the power flows on the lines after the attack.

This problem is thoroughly studied in [1] in the noise-free and low-noise settings. However, the previous work provides little insight on how to deal with the noisy measurements and inconsistency in power supply/demand. In this work, we adapt the convex relaxation of the failed lines detection problem provided in [1] and use Bayesian regression [7] to detect failed lines probabilistically based on noisy measurements. We follow the Bayesian Regression with Automatic Relevance Determination (ARD) model provided in [8]. The Bayesian approach provides the probability that each line is failed, instead of 0-1 hard decision on the status of the lines, which reflects the uncertainty in the detection caused by the noise.

These probabilities can further be used in two ways: (i) turn them into 0-1 hard decisions by considering a probability threshold p and detect all lines with probability of being failed greater than p as failed, and (ii) use them to limit the search space in the Brute Force Search (BFS) type methods by considering only lines with the high probability of being failed. We numerically compare these two approaches with the method provided in [1] as well as the BFS based methods provided in [9], [10], in the IEEE 300-bus system. We show that the second approach provides accuracy similar to (and sometimes better than) the BFS method but in an exponentially faster running time (e.g., in an attacked area with 15 lines, it runs 40 times faster). Hence, by combining the probabilities obtained by the Bayesian regression with the BFS, it reaches a sweet spot in accuracy and efficiency.

II. Model and Preliminaries

A. DC Power Flow Model

We use the linearized DC power flow model, which is widely used as an approximation for the non-linear AC power flow model in studying vulnerabilities of power grids. We represent the power grid by a connected undirected graph $G = (V, E)$ where $V = \{1, 2, \dots, n\}$ and $E = \{e_1, \dots, e_m\}$ are the set of nodes and edges corresponding to the buses and transmission lines, respectively. Each edge $e_i = \{u, v\} \in E$ is characterized by its *reactance* $x_{uv} = x_{vu} = x_{e_i}$.

Given the power supply/demand vector $\vec{p} \in \mathbb{R}^{|V|}$ and the reactance values, a *power flow* is a solution $\vec{f} \in \mathbb{R}^{|E|}$ and $\vec{\theta} \in \mathbb{R}^{|V|}$ of: $\mathbf{A}\vec{\theta} = \vec{p}$ and $\mathbf{B}\mathbf{D}^t\vec{\theta} = \vec{f}$, where $\mathbf{D} \in \{-1, 0, 1\}^{|V| \times |E|}$ is the *incidence matrix* of G defined as,

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

This work is supported by DARPA RADICS under contract #FA-8750-16-C-0054. We also thank Ghazal Fazelnia for her helpful insights and comments.

$\mathbf{B} := \mathrm{diag}([1/x_{e_1}, 1/x_{e_2}, \ldots, 1/x_{e_m}])$ is a diagonal matrix with diagonal entries equal to the inverse of the reactance values, and $\mathbf{A} \in \mathbb{R}^{|V|\times|V|}$ is the *admittance matrix* of G defined as $\mathbf{A} := \mathbf{DBD}^T$.

Notation. If X, Y are two subgraphs of G, $\mathbf{A}_{X|Y}$ denotes the submatrix of $\mathbf{A}$ with rows from X and columns from Y. For any matrix $\mathbf{C}$, $\mathbf{C}^T$ denotes its transpose and $\mathbf{C}^+$ denotes its *Moore-Penrose pseudo-inverse*. For a vector $\vec{y}$, $\|\vec{y}\|_1 := \sum_{i=1}^{n} |y_i|$ is its l_1-norm, $\|\vec{y}\|_2 := (\sum_{i=1}^{n} y_i^2)^{1/2}$ is its l_2-norm, and $\mathrm{supp}(\vec{y}) := \{i | y_i \neq 0\}$ is its support.

B. The Attack Model

We follow [1] and assume that an adversary attacks an area $H = (V_H, E_H)$ (representing a subgraph of G) by disconnecting some lines within the attacked area, referred to as *failed lines* and denoted by F, and blocking the measurements coming from inside the attacked area to mask the status of the lines in H. Hence, after the attack, the phase angles of the nodes and the status of the lines in H become unavailable to the control center. The objective is to use the phase angle measurements before and partial measurements after the attack to detect the failed lines. Detecting failed lines is necessary for obtaining a correct understanding of the grid topology and estimating the power flows on the lines after the attack, and for executing an effective control algorithm (e.g., power grid intentional islanding or load shedding) in order to stop the initial attack from affecting the entire grid through cascade of failures and causing a major blackout.

We use the prime symbol $(')$ to denote the values after an attack. For example $\mathbf{A}'$ and $\vec{\theta}'$ denote the admittance matrix and phase angle of the nodes after the attack. We denote the complement of the attacked area H by $\bar{H} = G \backslash H$. $\vec{\theta}_H$ and $\vec{\theta}_{\bar{H}}$ are the vectors of phase angles of the nodes in H and $\bar{H}$, respectively. Without loss of generality, we also assume that $E_H = \{e_1, e_2, \ldots, e_{|E_H|}\}$.

Using this notation, we assume that the control center receives T noisy measurements $\vec{\theta}^{(1)}, \vec{\theta}^{(2)}, \ldots, \vec{\theta}^{(T)}$ for the phase angles of the nodes in the grid before the attack, and T noisy measurements $\vec{\theta}'^{(1)}_{\bar{H}}, \vec{\theta}'^{(2)}_{\bar{H}}, \ldots, \vec{\theta}'^{(T)}_{\bar{H}}$ for the phase angles of the nodes outside of the attacked area after the attack. The T measurements are from different but very close time intervals (e.g., within microseconds). The measurements after the attack are assumed to be associated with the time that the system has been stabilized after the attack.

The noise corresponds to the measurement noise as well as fluctuations in the supply/demand vector $\vec{p}$. We assume that each measurement $\vec{\theta}^{(i)} = \vec{\theta} + \vec{e}$, in which $\vec{e} \sim \mathcal{N}(0, \beta\mathbf{I})$ has a multivariate Gaussian distribution with a zero mean and a diagonal covariance matrix $\beta\mathbf{I}$, and $\mathbf{A}\vec{\theta} = \vec{p}$. We assume the same for the measurements after the attack, $\vec{\theta}'^{(i)}_{\bar{H}} = \vec{\theta}'_{\bar{H}} + \vec{e}_{\bar{H}}$, in which $\vec{e}_{\bar{H}} \sim \mathcal{N}(0, \beta\mathbf{I}_{\bar{H}})$, and $\mathbf{A}'\vec{\theta}' = \vec{p}$. To measure the noise level, we define the Signal to Noise Ratio (SNR) based on the phase angles before the attack as $20\log_{10}(\|\vec{\theta}\|_2/\sqrt{|V|\beta})$.

One way of using the T measurements is to take average of these vectors. Hence, we define $\vec{\theta}^{(\mu)} := (\sum_i \vec{\theta}^{(i)})/T$ and $\vec{\theta}'^{(\mu)}_{\bar{H}} := (\sum_i \vec{\theta}'^{(i)}_{\bar{H}})/T$.

C. Bayesian Regression

Given a data set $\mathcal{D} = \{\mathbf{X}, \mathbf{Y}\}$, with $\mathbf{X} = [\vec{x}_1, \ldots, \vec{x}_n] \in \mathbb{R}^{d\times n}$ and $\mathbf{Y} = [y_1, \ldots, y_n] \in \mathbb{R}^{n\times 1}$, in linear regression, we are interested in finding a vector $\vec{w} \in \mathbb{R}^d$ that minimizes $\|\vec{w}^T\mathbf{X} - \mathbf{Y}\|_2$. Such a vector $\vec{w}$ determines the relation between in the input vectors $\mathbf{X}$ and outputs $\mathbf{Y}$.

The linear regression provides very little information on the uncertainties in the computed vector $\vec{w}$, specially when $\mathbf{Y}$ is noisy. One way of overcoming this issue is by adapting the Bayesian approach to regression [7]. In the Bayesian regression, instead of finding *the maximum likelihood* estimate for a vector $\vec{w}$ that describes the relationship between the inputs and the outputs, we are interested in computing a probability distribution on all possible vectors $\vec{w}$ that describe this relationship. This can be done by computing the *posterior* distribution on $\vec{w}$ using Bayes' rule, by assuming a *prior* distribution on $\vec{w}$ and an appropriate model for the way output data is related to the input. Due to the space constraints, for further details on Bayesian Regression see [7, Section 3.3].

Some sparsity constraints on the coefficients, as in the Lasso regression [7], can also be obtained in Bayesian regression by appropriate choice of the prior for the coefficients. In particular, in this paper, we follow the Bayesian Regression with Automatic Relevance Determination (ARD) model provided in [8]. This model assumes a linear relation between inputs $\vec{x}$ and outputs y, and constant-variance Gaussian noise, such that the likelihood is given by

$$\mathrm{P}(y|\vec{x}, \vec{w}, \tau) = \mathcal{N}(y|\vec{w}^T\vec{x}, \tau^{-1}) = \Big(\frac{\tau}{2\pi}\Big)^{1/2} \exp\Big(-\frac{\tau}{2}(y - \vec{w}^T\vec{x})^2\Big),$$

and the prior on $\vec{w}, \tau^{-1}$ is conjugate normal inverse-gamma

$$\mathrm{P}(\vec{w}, \tau|\vec{\alpha}) = \mathcal{N}\Big(\vec{w}|0, (\tau\mathrm{diag}(\vec{\alpha}))^{-1}\Big)\mathrm{Gam}(\tau|a_0, b_0),$$

where $\mathrm{diag}(\vec{\alpha})$ is a diagonal matrix with entries given by vector $\vec{\alpha} = [\alpha_1, \ldots, \alpha_d]$. The entries of $\vec{\alpha}$ are independent and the hyper-prior is given by $\mathrm{P}(\alpha) = \prod_i \mathrm{Gam}(\alpha_i|c_0, d_0)$.

Under this model, analytically computation of the posterior distribution $\mathrm{P}(\vec{w}, \tau, \vec{\alpha}|\mathcal{D})$ is intractable. Hence, in [8] the Variational Inference approach is used to approximate the posterior distribution (for further details on how to compute $\vec{w}_n, \mathbf{V}_n, a_n, b_n$ see [8])

$$\mathrm{Q}(\vec{w}, \tau) \quad = \quad \mathcal{N}(\vec{w}|\vec{w}_n, \tau^{-1}\mathbf{V}_n)\mathrm{Gam}(\tau|a_n, b_n). \qquad (1)$$

In Section IV, we use this model to detect the failed lines after an attack as described in the previous subsection.

III. Detection Methods

The problem of failed lines detection using partial phase angle measurements is NP-hard in general [11]. However, in special cases, it is possible to efficiently solve this problem. In this section, we provide an overview of the two main approaches to failed lines detection using phase angle measurements. In recent years, some efforts have been made to apply learning algorithms to this problem, such as [12]. However, since detection using these methods is limited to the state of the grid that they are trained on, they are not general enough to fit into the scope of this paper.

A. Brute Force Search

The classical approach to the failed lines detection problem is the Brute Force Search (BFS) [9], [10]. The BFS based methods consider all possible set of failed lines and return the set with the maximum likelihood, based on the observed measurements.

Assume $\mathcal{M}$ is the set of all admittance matrices associated with the graphs that can be obtained by removing any number of lines in E_H from the graph G. One way of finding the set of failed lines is by searching the entire space of matrices $\mathcal{M}$ to find a matrix $\mathbf{C}$ that minimizes the following

$$\min_{\mathbf{C}\in\mathcal{M}} \|(\mathbf{C}^{+}\mathbf{A}\vec{\theta}^{(\mu)})_{\bar{H}} - \vec{\theta}'^{(\mu)}_{\bar{H}}\|_2. \tag{2}$$

To see why (2) is an intuitive approach for detecting the set of failed lines, assume $F_{\mathbf{C}} \subseteq E_H$ is the set of lines that their removal results in the admittance matrix $\mathbf{C}$. Notice that $\mathbf{A}\vec{\theta}^{(\mu)} \approx \vec{p}$. Therefore, $\mathbf{C}^{+}\mathbf{A}\vec{\theta}^{(\mu)}$ is the approximate phase angles that we would have expected to observe if $F_{\mathbf{C}}$ was the actual set of failed lines. By comparing the expected phase angle of the nodes outside of the attacked area $(\mathbf{C}^{+}\mathbf{A}\vec{\theta}^{(\mu)})_{\bar{H}}$ with the average of the observed phase angles $\vec{\theta}'^{(\mu)}_{\bar{H}}$, one can check how much the guessed set of failed lines $F_{\mathbf{C}}$ is consistent with the observed data.

Despite the simplicity and effectiveness of the BFS based methods, however, their running time grow exponentially with the number of lines in the attacked area. This makes them inapplicable to the scenarios that require fast decision making in order to localize the attack and reduce its consequences. In Section V, we numerically investigate the limitations of the BFS based methods.

B. Convex Optimization

Another approach to the failed lines detection problem is to use topological properties of the grid in order to find the failed lines more efficiently for certain topologies. Such methods and their connection to the topology of the grid were first fully explored in [1]. However, similar approaches were first studied in [13] without making the connection to the topology.

It is proved in [1] that under some conditions on the topology of the attacked area, the optimal solution $\vec{t} \in \mathbb{R}^{|E_H|}$ and $\vec{z} \in \mathbb{R}^{|V_H|}$ of the following convex optimization problem is such that $\text{supp}(\vec{t}) = \{i|e_i \in F\}$ and $\vec{z} = \vec{\theta}'_H$:

$$\begin{aligned}
&\min_{\vec{t},\vec{z},\epsilon} \|\vec{t}\|_1 + \lambda\epsilon \text{ s.t.} \qquad\qquad (3)\\
&\|\mathbf{A}_{H|H}(\vec{\theta}^{(\mu)}_{H} - \vec{z}) + \mathbf{A}_{H|\bar{H}}(\vec{\theta}^{(\mu)}_{\bar{H}} - \vec{\theta}'^{(\mu)}_{\bar{H}}) - \mathbf{D}_H\vec{t}\|_2 \le \epsilon\\
&\|\mathbf{A}_{\bar{H}|H}(\vec{\theta}^{(\mu)}_{H} - \vec{z}) + \mathbf{A}_{\bar{H}|\bar{H}}(\vec{\theta}^{(\mu)}_{\bar{H}} - \vec{\theta}'^{(\mu)}_{\bar{H}})\|_2 \le \epsilon.
\end{aligned}$$

Therefore, by solving (3) the nonzero elements of $\vec{t}$ reveal the failed lines. It can be shown that in this case, if $e_i = \{j,k\} \in F$, then $t_i \approx -a_{jk}(\theta'_j - \theta'_k)$ which is the amount of the power flow that the line e_i would have carried if it was not failed.

When the noise is low, it is shown in [1] that (3) can detect the failed lines very well. However, as the noise level increases, the detection based on the solution to (3) produces false negatives and false positives. The main challenge here is to determine the weight λ that makes the solution space small enough to contain only the actual solution. In Section V, we numerically investigate the limitations of this method as well.

The main goal of this paper is to extend the idea of the optimization (3) and make it more robust to noisy measurements. In the next section, we show that why Bayesian regression is a suitable option for this purpose.

IV. Bayesian Regression for Failed Line Detection

In order to provide a robust method for failed lines detection, we use Bayesian regression, which is more suitable for dealing with uncertainties.

To see how Bayesian regression can be used here, recall that the key in (3) is that once it is solved, the failed lines can be detected by looking for the nonzero elements of vector $\vec{t}$. Now if the phase angle measurements are noisy, all we need to compute is the probability that each element of vector $\vec{t}$ is nonzero in the optimal solution. And this is where Bayesian regression can be used. Recall the regression notation from Subsection II-C and define

$$\mathbf{X}:=\begin{bmatrix}\mathbf{A}_{H|H} & D_H\\ \mathbf{A}_{\bar{H}|H} & 0\end{bmatrix}^T, \mathbf{Y}^{(i,j)}:=\begin{bmatrix}\mathbf{A}_{H|H}\vec{\theta}^{(i)}_{H} + \mathbf{A}_{H|\bar{H}}(\vec{\theta}^{(i)}_{\bar{H}} - \vec{\theta}'^{(j)}_{\bar{H}})\\ \mathbf{A}_{\bar{H}|H}\vec{\theta}^{(i)}_{H} + \mathbf{A}_{\bar{H}|\bar{H}}(\vec{\theta}^{(i)}_{\bar{H}} - \vec{\theta}'^{(j)}_{\bar{H}})\end{bmatrix}^T,$$

where $\mathbf{Y}^{(i,j)}$ is the output vector using the i^{th} phase angle measurement before that attack and j^{th} phase angle measurement after the attack. In regression, we want to find the vector $\vec{w}$ such that for any $1 \le k \le |V|$ and any $1 \le i,j \le T$: $y^{(i,j)}_k \approx \vec{w}^T x_k$. Notice that in this setting, the first $|V_H|$ elements of $\vec{w}$ are associated with vector $\vec{z}$ in (3) and the last $|E_H|$ elements are associate with vector $\vec{t}$.

To find a posterior distribution on $\vec{w}$, we use the Bayesian regression model with ARD introduced in [8] and summarized in Subsection II-C. As we mentioned in Subsection II-C, the prior distribution on $\vec{w}$ and the structure in this model is such that it promotes sparsity in $\vec{w}$ which can play the role of $\|\vec{t}\|_1$ minimizer in (3). Hence, this model can be considered as a Bayesian approach for solving (3).

Once the posterior distribution on $\vec{w}$ is computed as in (1), it can be used to estimate the phase angles of the nodes inside the attacked area and compute the probability that each line is failed. Define $\vec{z}$ to be the first $|V_H|$ elements of the posterior mean vector $\vec{w}_n$. It provides the maximum likelihood estimate for the phase angles of the nodes inside the attacked area.

Also define $\vec{t} = [t_1, \ldots, t_{|E_H|}]^T$ to be last $|E_H|$ entries of vector $\vec{w}$, $\vec{\mu}$ to be the last $|E_H|$ entries of the posterior mean vector $\vec{w}_n$, and $\mathbf{\Sigma}$ to be the submatirx of matrix $\mathbf{V}_n$ associated with its last $|E_H|$ rows and columns. Then according to the model, $\text{P}(\vec{t}|\tau) = \mathcal{N}(\vec{\mu}, \tau^{-1}\mathbf{\Sigma})$. Moreover, τ can be further approximated by its posterior mean a_n/b_n to obtain $\text{P}(\vec{t}) \approx \mathcal{N}(\vec{\mu}, b_n/a_n\mathbf{\Sigma})$. Using this, we can approximate the probability that each line is failed by computing the probability that $|t_i|$ is greater than $\gamma > 0$. Hence, since $\mathbf{\Sigma}$ is diagonal,

$$\text{P}(e_i \in F) \approx \text{P}(|t_i| \ge \gamma) \approx \frac{1}{2} - \frac{1}{2}\text{erf}\Big(\frac{\mu_i - \gamma}{\sqrt{2b_n/a_n\Sigma_{ii}}}\Big), \tag{4}$$

where $\mathrm{erf}(.)$ is the Gaussian error function. Here, we select $\gamma = 0.1$, which means that we are interested in detecting the entries of $\vec{t}$ that can get a substantial value.[1]

Once the probabilities are computed using (4), there are two ways to turn these probabilities into hard decisions:

(i) Consider a probability threshold p and detect all lines with probability of being failed greater than p as failed and use the probability values as the confidence in the detection. We refer to this method in Section V as *Bayes*.

(ii) Use the probabilities to sort the lines and only use the top k lines as the set of lines that are most probable to be failed in the search space of the BFS optimization (2). This significantly reduces the running time of the BFS method by keeping the size of the search space constant as the size of the attacked area increases. We refer to this method in Section V as *Bayes+BFS*.

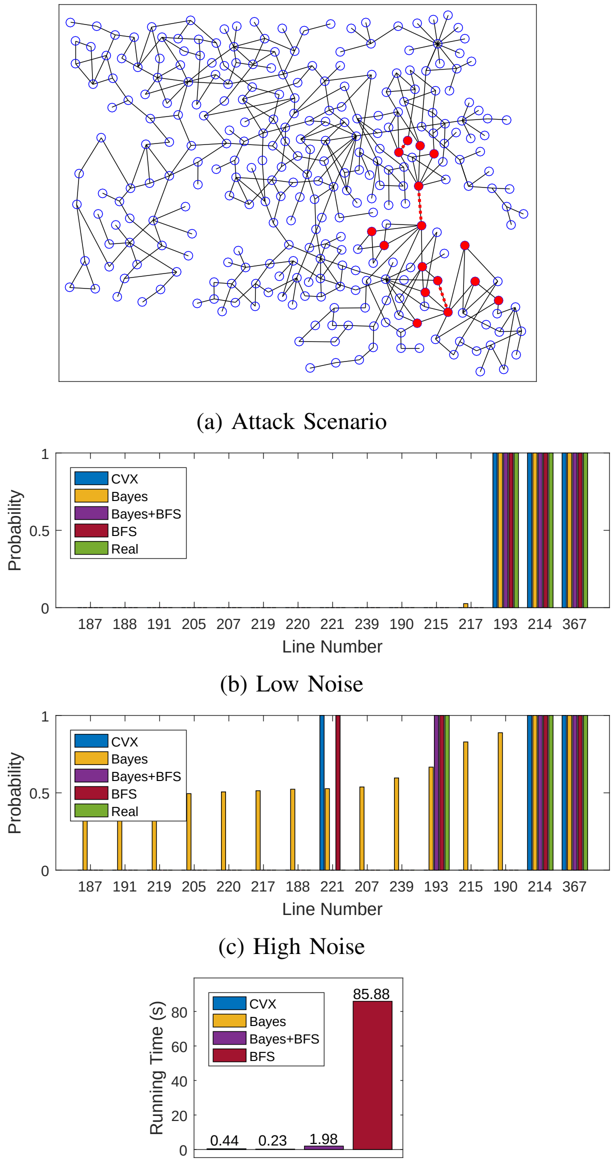
V. Numerical Results

In this section, we compare the two Bayesian regression based methods for failed lines detection, introduced in the previous section, with the methods described in Section III.

We assume $T = 10$ noisy measurements before and after the attack. In the BFS optimization (2), we limit the search to the set of failed lines of size at most 5. In the convex optimization (3), we set $\lambda = 1000$ and refer to the solution obtained by this method by *CVX*. Finally, we set $p = 0.9999$ as the threshold probability in Bayes, and $k = 7$ in Bayes+BFS.

Fig. 1 depicts an attack example and the detected failed lines using different methods. In this example, as depicted in Fig. 1(a), the phase angle measurements from the attacked area, shown by red filled nodes, are blocked and 3 lines, shown by red dashed lines, are failed. Figs. 1(b) and 1(c) compare the probability that each line is failed, computed by different methods, under low and high noise levels. Each color bar indicates the probability that a particular method assigns to a line. Since all the methods except Bayes make hard decisions, the probabilities they provide are either 0 or 1. Therefore, in these figures, the lines are ordered based on the probabilities obtained by the Bayes method.

When the noise level is low, as can be seen in Fig. 1(b), all the four methods can detect the failed lines accurately. When the noise level is high, however, all the methods except the Bayes+BFS result in a false positive or a false negatives, or both in their detections. For example, CVX misses failed line 193 and incorrectly detects line 221. Bayes with $p = 0.9999$ threshold also misses failed line 193. The interesting result, however, is that BFS method incorrectly detects additional line 193 as failed. On the other hand, Bayes+BFS by focusing only on the lines with high probability of being failed, avoids this false extra detection and detects all the failed lines correctly.



(a) Attack Scenario

(b) Low Noise

(c) High Noise

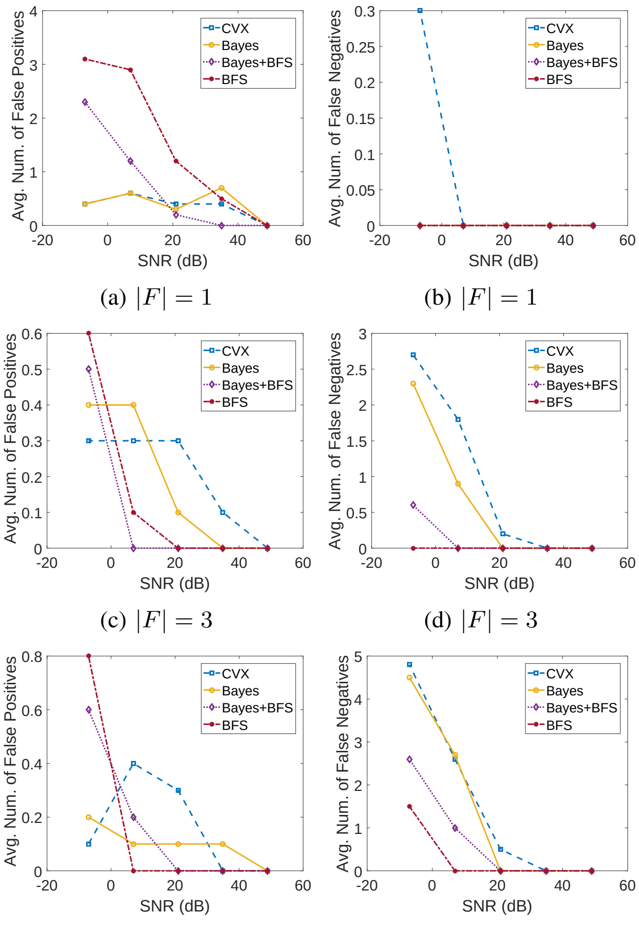
(d) Running Times

Fig. 1: An example of an attack and the failed lines detection using different methods. (a) The attacked area shown by red filled nodes and three failed lines shown by red dotted lines in the IEEE 300-bus system, (b) the probability that each line is failed as computed by different methods when the noise is low ($\mathrm{SNR} = 42.8dB$), (c) the probability that each line is failed as computed by different methods when the noise is high ($\mathrm{SNR} = 3.7dB$), (d) running times.

Fig. 1(d) shows the running times of these methods. As we mentioned in Section III and can be seen in this figure, the BFS method is significantly slower than other methods which make it unsuitable for practical purposes. While CVX method is very fast, it may provide less accuracy as the noise level increases. The Bayes method is the fastest among the four and by combining it with the BFS, as in Bayes+BFS, it reaches the sweet spot in accuracy and efficiency.

To further investigate these observations, we performed more simulations under different attack scenarios (e.g., different number of failed lines and different noise levels). The results are presented in Fig. 2. As can be seen, the results are consistent with the results that we presented in Fig. 1. While all the methods perform very well when the noise level is low,

[1] The threshold γ should be selected based on the active power nominal value. Here, based on the IEEE test system, the nominal value is $100MW$ which makes the 0.1 threshold in per unit, equal to $10MW$.



(a) $|F| = 1$

(b) $|F| = 1$

(c) $|F| = 3$

(d) $|F| = 3$

(e) $|F| = 5$

(f) $|F| = 5$

Fig. 2: The average number of false positives and negatives using different detection methods under different SNR values and different number of failed lines.

the BFS and Bayes+BFS perform better than the CVX and the Bayes, specially in the number of false negatives.

Finally, to show the main disadvantage of the BFS method more clearly, in Fig. 3, we compared the running times of the four methods as the number of lines in the attacked area increases. As can be seen, while the running times of the CVX, Bayes, and Bayes+BFS remain constant as the number of lines increases, the BFS method significantly slows down. Hence, Figs. 2 and 3 confirm the results we observed in Fig. 1 that the Bayes+BFS provides an accuracy level similar to BFS in significantly lower running time.

VI. Conclusion

In this work, we provided a new method based on Bayesian regression for power grid state estimation following a cyber-physical attack. We numerically showed that the method that uses Bayesian regression to find the set of lines with the highest probability of being failed and performs a limited BFS on that small set of lines, reaches a sweet spot in terms of accuracy and efficiency.

Our work is one of the first works in applying Bayesian data analysis tools for power grid state estimation. We believe that these tools have a vast potential and can be applied for power grid state estimation not only during an attack but also during the daily operation of the grid. Providing a general framework for power grid state estimation using Bayesian Regression is part of our future work.

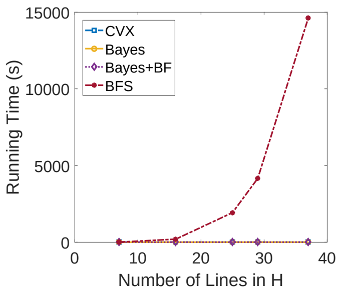


Fig. 3: The relationship between the running times of different methods and the number of lines in H.

Finally, in this work, we used the linearized DC power flows to describe the state of the grid. However, the provided methods can also be extended to the AC power flows using the extension of the optimization (3) to the AC power flows that is provided in a recent work [14]. Moreover, they can be extended to the case where there is false data injection instead of data blocking, using similar methods to ones in [11]. Exploring these directions are also part of our future work.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] "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.

[7] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[8] J. Drugowitsch, "Variational bayesian inference for linear and logistic regression," *arXiv preprint arXiv:1310.5438v3*, 2017.

[9] 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.

[10] 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.

[11] S. Soltan, M. Yannakakis, and G. Zussman, "REACT to cyber attacks on power grids," *arXiv preprint arXiv:1709.06934*, Sept. 2017.

[12] 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.

[13] 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.

[14] S. Soltan and G. Zussman, "EXPOSE the line failures following a cyber-physical attack on the power grid," *arXiv preprint arXiv:1709.07399*, Sept. 2017.