PMU Placement for Line Outage Identification via Multiclass Logistic Regression

Taedong Kim and Stephen J. Wright

Abstract—We consider the problem of identifying a single line outage in a power grid by using data from phasor measurement units (PMUs). When a line outage occurs, the voltage phasor of each bus node changes in response to the change in network topology. Each individual line outage has a consistent “signature,” and a multiclass logistic regression (MLR) classifier can be trained to distinguish between these signatures reliably. We consider first the ideal case in which PMUs are attached to every bus, but phasor data alone is used to detect outage signatures. We then describe techniques for placing PMUs selectively on a subset of buses, with the subset being chosen to allow discrimination between as many outage events as possible. We also discuss extensions of the MLR technique that incorporate explicit information about identification of outages by PMUs measuring line current flow in or out of a bus. Experimental results with synthetic 24-hour demand profile data generated for 14, 30, 57, and 118-bus systems are presented.

Index Terms—line outage identification, phasor measurement unit, optimal PMU placement, multiclass logistic regression

I. INTRODUCTION

In recent years, phasor measurement units (PMUs) have been introduced as a way to monitor power system networks. Unlike the more conventional Supervisory Control and Data Acquisitions (SCADA) system, whose measurements include active and reactive power and voltage magnitude, PMUs can provide accurate, high-sampling-rate, synchronized measurements of voltage phasor. There has been much ongoing research on how the real-time measurement information gathered from PMUs can be exploited in many areas of power system studies, including system control and state estimation.

In this paper, we study the use of PMU data in detecting topological network changes caused by single-line outages, and propose techniques for determining optimal placement of a limited number of PMU devices in a grid, so as to maximize the capability for detecting such outages. Our PMU placement approach can also be used as a tie-breaker for the other types of strategies that have multiple optimal solutions, for example, maximum observability problems.

Knowledge of topological changes as a result of line failure can be critical in deciding how to respond to a blackout. Rapid detection of such changes can enable actions to be taken that reduce risks of cascading failures that lead to large-scale blackouts. One of the main causes of the catastrophic Northeast blackout of 2003 was faulty topological knowledge of the grid following the initial failures (see [1]).

Numerous approaches have been proposed for identifying line outages using PMU device measurements. In [2], [3], phasor angle changes are measured and compared with expected phasor angle variations for all single- or double-line outage scenarios. Support vector machines (SVM) were proposed for identification of single-line outages in [4]. A compressed-sensing approach was applied to DC power balance equations to find sparse topological changes in [5], while a cross-entropy optimization technique was considered in [6]. Since the approaches in [5] or [6] use the linearized DC power flow models to represent a power system, their line outage identification strategies rely only on changes to phase angles. Voltage magnitude measurements from PMUs, which also provide useful information for monitoring a power system, are ignored. Our use of the AC power flow model may allow more accurate modeling of the system and more complete exploitation of the available data.

The key feature that makes line outage identification possible is that voltage phasor measurements reported by PMUs are different for different line-outage scenarios. Our approach aims to distinguish between these different “signatures” by using a multiclass logistic regression (MLR) model. The model can be trained by a convex optimization approach, using standard techniques. The coefficients learned during training can be applied during grid operation to detect outage scenarios. Our approach could in principle be applied to multilime outages too, but since the problem dimension is much larger in such cases — the number of possible outage scenarios is much greater — it is no longer practical. Second, even when trained only to recognize single-line outages, our classifier is useful in multilime outage situations on large grids, when the coupling between the lines is weak (as discussed in [7] Section 2.2). In other words, many multilime outage cases can be decomposed into single-line outage events on different parts of the grid.

Because of the expense of installation and maintenance, PMUs can reasonably be installed on just a subset of buses in a grid. We therefore need to formulate an optimal placement problem to determine the choice of PMU locations that gives the best information about system state. Several different criteria have been proposed to measure quality of a given choice of PMU locations. One of the most popular criteria is to place PMUs to maximize the number of nodes in the system that can be observed directly, either by a PMU located at that node or an adjacent connected node [8]. Another criterion is quality of state estimation results. In this approach, one can use PMU measurements alone, or combine them with traditional
SCADA measurement to decide the optimal PMU deployment (see for example [9]). Other criteria and techniques for locating PMUs optimally are discussed in the review papers [10], [11].

For the case in which line outage identification is used as a criterion for PMU placement, we mention [12], [5], [7]. In [12], the authors use pre-computed phase angles as outage signatures and attempt to find the optimal PMU locations by identifying a projection (by setting to zero the entries which are not selected as PMU locations) that maximizes the minimum distance in $p$-norm of the projected signatures. The problem is formulated as an integer program (IP) and a greedy heuristic and branch-and-bound approach are proposed. PMU placement for the line outage identification method discussed in [5] is studied in [7]. A non-convex mixed-integer nonlinear program (MINLP) is formulated, leading to a linear programming convex relaxation. Again, a greedy heuristic and a branch-and-bound algorithm is suggested as a solution methodology.

We build our PMU placement formulation on our MLR model for single-line outage detection, by adding nonsmooth “Group LASSO” regularizers to the MLR objective and applying several heuristics.

We assume that multiphase measurements are available at all buses. The reason is that the PMUs need to be deployed at all buses in order to have a complete view of the grid. In this way, we can detect any line outage, even if it is not directly connected to the PMU.

The rest of this paper is organized as follows: In Section II the line outage identification problem is described along with the multiclass logistic regression (MLR) formulation. The problem of PMU placement to identify a line outage is described in Section III and we describe the group-sparse problem of PMU placement to identify a line outage is described in Section IV. A conclusion appears in Section V.

In supplementary material, we describe an extension that makes use of explicit line outage information. This model uses the fact that when a PMU is attached to a line, it can detect the quasi-static equilibrium that is reached after the disruption. We use a quasi-steady state AC power flow model (see e.g. [13] Chapter 10) as a mapping from time varying load variation (and line outage events) to the polar coordinate “outputs” of voltage magnitude and angle.

PMUs report phasor measurements with high frequency, and changes in voltage due to topology changes of the power grid tend to be larger than the variation of voltage phasor during normal operation (for example, demand fluctuation that occurs during the sampling period). We construct signature vectors from these voltage changes under the various single-line outage situations, and use them to train a classifier.

Figure I shows an example of voltage changes for a 9-bus system (case9.m from MATPOWER [14]) on different line outage scenarios. The failure of lines connecting buses 4-5, 5-6, 6-7, 7-8, and 8-9 is considered as possible scenarios (columns in the figure) whose voltage phasors at buses 5, 6, and 7 (rows in the figure) observed. In each plot, $x$-axis shows voltage magnitude and $y$-axis shows voltage phase angle at the bus. The red dots in each plot indicate the voltage phasor when there is no line failure and the blue dots are the voltage values under the specified failure scenario. We observe that voltage values at these buses change in distinctive ways under different line outage scenarios. It therefore seems realistic to expect that by comparing voltage phasor data, gathered before and after a failure event, we can identify the failed line reliably. We now describe the multinomial logistic regression model for determining the outage scenario.

### A. Multinomial Logistic Regression Model

Multinomial logistic regression (MLR) is a machine-learning approach for multiclass classification. In our application, examples of voltage phasor changes under each outage scenario are used to train the classifier by determining parameter values in a set of parametrized functions. Once the parameters have been found, these functions determine the likelihood of a given set of phasor changes as being indicative of each possible failure scenario.

Suppose that there are $K$ possible outcomes (classes) labelled as $i \in \{1, 2, \ldots, K\}$ for a given vector of observations $X$. In the multinomial logistic regression model, the probability of a given observation $X$ has an outcome $Y$ (one of the $K$ possibilities $i \in \{1, 2, \ldots, K\}$) is given by the following formula:

$$
\Pr(Y = i|X) := \frac{e^{(\beta_i,X)}}{\sum_{k=1}^{K} e^{(\beta_k,X)}} \text{ for } i = 1, 2, \ldots, K, \quad (1)
$$

where $\beta_1, \beta_2, \ldots, \beta_K$ are regression coefficients, whose values are obtained during the training process. Note that there is one regression coefficient $\beta_i$ for each outcome $i \in \{1, 2, \ldots, K\}$. Once values of the coefficients $\beta_i$, $i \in \{1, 2, \ldots, K\}$ have been determined, we can predict the outcome associated with a given feature vector $X$ by evaluating

$$
k^* = \arg \max_{k \in \{1, 2, \ldots, K\}} \Pr(Y = k|X),
$$
or equivalently,

$$
k^* = \arg \max_{k \in \{1, 2, \ldots, K\}} \langle \beta_k, X \rangle. \quad (2)
$$

Training of the regression coefficients $\beta_1, \ldots, \beta_K$ can be performed by maximum likelihood estimation. The training data consists of $M$ pairs $(X_1, Y_1), (Y_2, Y_2), \ldots, (X_M, Y_M)$, each consisting of a feature vector and its corresponding outcome. Given formula (1), the a posteriori likelihood of observing $Y_1, Y_2, \ldots, Y_M$ given the events $X_1, X_2, \ldots, X_M$ is

$$
\prod_{i=1}^{M} P(Y = Y_i|X_i) = \prod_{i=1}^{M} \left( \frac{e^{(\beta_{Y_i},X_i)}}{\sum_{k=1}^{K} e^{(\beta_k,X_i)}} \right). \quad (3)
$$
By taking log of \( f(\beta) \), we have log-likelihood function
\[
f(\beta) := \sum_{i=1}^{M} \left( \langle \beta Y_i, X_i \rangle - \log \sum_{k=1}^{K} e^{(\beta_k, X_i)} \right),
\]
where the matrix \( \beta \) is obtained by arranging the coefficient vectors as \( [\beta_1 \beta_2 \ldots \beta_K] \). The maximum likelihood estimate \( \beta^* \) of regression coefficients is obtained by solving the following optimization problem:
\[
\beta^* = \arg \max_{\beta} f(\beta). \tag{4}
\]
This is a smooth convex problem that can be solved by fairly standard techniques for smooth nonlinear optimization, such as L-BFGS [15]. Note that \( f(\beta) \leq 0 \) for all \( \beta \).

If the training data is separable, the value of \( f(\beta) \) can be made to approach zero arbitrarily closely by multiplying \( \beta \) by an increasingly large positive value (see [16]). To recover meaningful values of \( \beta \) in this case, we can solve instead the following regularized form of (4):
\[
\beta^* = \arg \max_{\beta} f(\beta) - \tau w(\beta) \tag{5}
\]
where \( \tau > 0 \) is a penalty parameter and \( w(\beta) \) is a (convex) penalty function of the coefficient \( \beta \). The penalized form can also be used to promote some kind of structure in the solution \( \beta^* \), such as sparsity or group-sparsity. This property is key to our PMU placement formulation, and we discuss it further in Section III.

Training of the MLR model, via solution of (4) or (5), can be done offline, as described in the next subsection. Once the model is trained (that is, the coefficients \( \beta_i, i = 1, 2, \ldots, K \) have been calculated), classification can be done via (2), at the cost of multiplying the matrix \( \beta \) by the observed feature vector \( X \), an operation that can be done in real time.

B. Training Data: Observation Vectors and Outcomes

In our MLR model for line outage identification problems, the observation vector \( X_j \) is constructed from the change of voltage phasor at each bus, under a particular outage scenario. The corresponding outcome is the index of the failed line.

Suppose that a power system consists of \( N \) buses, all equipped with PMUs that report the voltage values periodically. Let \( (V_i, \theta_i) \) and \( (V_i', \theta_i') \), \( i = 1, 2, \ldots, N \), be two phasor measurements obtained from PMU devices, one taken before a possible failure scenario and one afterward. The observation vector \( X \) which describes the voltage phasor difference is defined to be
\[
X = \begin{bmatrix} \Delta V_1 & \cdots & \Delta V_N & \Delta \theta_1 & \cdots & \Delta \theta_N \end{bmatrix}^T \tag{6}
\]
where \( \Delta V_i = V_i' - V_i \) and \( \Delta \theta_i = \theta_i' - \theta_i \), for \( i = 1, 2, \ldots, N \). If we assume that the measurement interval is small enough that loads and demands on the grid do not change significantly between measurements, we would expect the entries of \( X \) to be small, unless an outage scenario (leading to a topological change to the grid) occurred. Some such outages would lead to failure of the grid. More often, feasible operation can continue, but with significant changes in the voltage phasors, indicated by large components of \( X \).

The training data \( (X_j,Y_j) \) can be assembled by considering a variety of realistic demand scenarios for the grid, solving the AC power flow equations for each possible outage scenario (setting the value of \( Y_j \) according to the index of that failure), then setting \( X_j \) to be the shift in voltage phasor that corresponds to that scenario. The phasor shifts for a particular scenario change somewhat as the pattern of loads and generations changes, so it is important to train the model using a sample of phasor changes under different realistic patterns of supply and demand.
The observation vector can be extended to include additional information beyond the voltage phasor information from the PMUs, if such information can be gathered easily and exploited to improve the performance of the MLR approach. For example, the system operator may be able to monitor the power generation level \( G \) (expressed as a fraction of the long-term average generation) that is injected to the system at the same time points at which the voltage phasor measurements are reported. If included in the observation vector, this quantity might need to scaled so that it does not dominate the phaser difference information. Also, a constant entry can be added to the observation vector to provide more flexibility for the regression. The extended observation vector thus has the form
\[
\mathbf{X} = [\Delta V_1 \ \cdots \ \Delta V_N \ \Delta \theta_1 \ \cdots \ \Delta \theta_N \ \rho G \ \rho]^T
\]
where \( \rho \) is a scaling factor that approximately balances the magnitudes of all entries in the vector. (Note that since \( G \) is not too far from 1, it is appropriate to use the same scaling factor for the last two terms.) The numerical experiments in Section IV make use of this extended observation vector.

### III. PMU Placement

As we mentioned in Section I, installing of PMUs at all buses is impractical. Indeed, if it were possible to do so, single-line outage detection would become a trivial problem, as each outage could be observed directly by PMU measurements of line current flows in or out of a bus; there would be no need to use the “indirect” evidence provided by voltage phasor changes. In this section we address the problem of placing a limited number of PMUs around the grid, with the locations chosen in a fashion that maximizes the system’s ability to detect single-line outages. This PMU placement problem selects a subset of buses for PMU placement, and assumes that PMUs are placed to monitor voltage phasors at the selected buses.

A naive approach is simply to declare a “budget” of the number of buses at which PMU placement can take place, and consider all possible choices that satisfy this budget. This approach is of course computationally intractable except for very small cases. Other possible approaches include a mixed-integer nonlinear programming formulation [12, 7], but this is very hard to solve in general. In this paper, we use a regularizer function \( w(\beta) \) in (5) to promote the a particular kind of sparsity structure in the coefficient matrix \( \beta \). Specifically, A group \( \ell_1 \)-regularizer is used to impose a common sparsity pattern on all columns in the coefficient matrix \( \beta \), with nonzeros occurring only in locations corresponding to the voltage magnitude and phase angle changes for a subset of buses. The numerical results show that approaches based on this regularizer give reasonable performance on the PMU placement problem.

#### A. Group-Sparse Heuristic (GroupLASSO)

Let \( \mathcal{P} \) be the set of indices in the vector \( X \), that is \( \mathcal{P} = \{1, 2, \cdots, |X|\} \). Consider \( S \) mutually disjoint subsets of \( \mathcal{P} \), denotes \( \mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_s \). For each \( s \in \mathcal{S} := \{1, 2, \cdots, S\} \), define \( q_s(\beta) \) as follows:
\[
q_s(\beta) = \|[\beta]_{\mathcal{P}_s}\|_F = \sqrt{\sum_{i \in \mathcal{P}_s} \sum_{k=1}^N (\beta_{ik})^2}
\]
where \([\beta]_{\mathcal{P}_s}\) is the submatrix of \( \beta \) constructed by choosing the rows whose indices are in \( \mathcal{P}_s \), \( \|\cdot\|_F \) is the Frobenius norm, and \( \beta_{ik} \) is the \((i, k)\) entry of matrix \( \beta \) (thus, \( \beta_{ik} \) is the \(i\)th entry of the coefficient vector \( \beta_k \) in (2) and (3)). The value of \( q_s(\beta) \) is the \( l_2 \)-norm over the entries of matrix \( \beta \) which are involved in group \( s \). For our observation vectors \( X \) (6) and \( \mathbf{X} \) (7), we can choose the number of groups \( |S| \) equal to the number of buses \( N \), and set
\[
\mathcal{P}_s = \{s, s + N\}, \quad s = 1, 2, \ldots, N.
\]

Thus, if bus \( s \) is “selected” in the placement problem, the coefficients associated with \( \Delta V_s \) and \( \Delta \theta_s \) are allowed to be nonzeros. Buses that are not selected need not of course be instrumented with PMUs, because the coefficients in \( \beta \) that correspond to these buses are all zero. Note that for the extended vector \( \mathbf{X} \), we do not place the last two entries (the constant and the total generation) into any group, as we assume that these are always “selected” for use in the classification process.

For any subset \( \mathcal{R} \) of \( \mathcal{S} \), we define a group-\( \ell_1 \)-regularizer \( w_{\mathcal{R}}(\beta) \) to be the sum of \( q_s(\beta) \) for \( s \in \mathcal{R} \), that is,
\[
w_{\mathcal{R}}(\beta) = \sum_{s \in \mathcal{R}} q_s(\beta).
\]

Setting \( \mathcal{R} = \mathcal{S} \), the penalized form (5) with \( w = w_{\mathcal{S}} \) can be used to identify a group-sparse solution:
\[
\max_{\beta} f(\beta) - \tau w_{\mathcal{S}}(\beta).
\]

With an appropriate choice of the parameter \( \tau \), the solution \( \beta^* \) of (9) will be group-row-sparse, that is, the set \( \{s \in \mathcal{S} \mid q_s(\beta^*) \neq 0\} \) will have significantly fewer than \( |S| \) elements. Given a solution \( \beta^* \) of (9) for some value of \( \tau \), we could define the \( r \)-sparse solution as follows (for a given value of \( r \), and assuming that the solution of (9) has at least \( r \) nonzero values of \( q_s(\beta^*) \)):
\[
\mathcal{R}^* := \arg \max_{\mathcal{R} \subseteq \mathcal{S}, \lvert \mathcal{R} \rvert = r} w_{\mathcal{R}}(\beta^*).
\]

Since the minimizer \( \beta^* \) of (9) is biased due to the presence of the penalty term, we should use the submatrix extracted from \( \beta^* \) according to the selected group \( \mathcal{R}^* \) as the regression coefficients for purposes of multiclass classification. Rather, we should solve a reduced, unpenalized version of the problem in which just the coefficients from sets \( \mathcal{P}_s \) that were not selected are fixed at zero. That is, we define a debiased solution \( \beta^* \) corresponding to \( \mathcal{R}^* \) as follows:
\[
\max_{\beta} f(\beta) \quad \text{subject to } \beta_{ik} = 0 \text{ for all } (i, k) \text{ with } k = 1, 2, \ldots, K \text{ and } i \in \mathcal{P}_s \text{ for some } s \notin \mathcal{R}^*.
\]
B. Greedy Heuristic

The regularization approach can be combined with a greedy strategy, in which groups are selected one at a time, with each selection made by solving a regularized problem. Suppose that $\mathcal{R}^{l-1}$ is set of selected groups after $l - 1$ iterations of the selection heuristic. The problem solved at iteration $t$ of the heuristic to choose the next group is

$$\hat{\beta}^t = \arg\max_{\beta} f(\beta) - \tau w_{\mathcal{S} \setminus \mathcal{R}^{l-1}}(\beta),$$

(12)

The next group $s^l$ is obtained from $\hat{\beta}^l$ as follows:

$$s^l = \arg\max_{s \in \mathcal{S} \setminus \mathcal{R}^{l-1}} q_s(\hat{\beta}^l),$$

and we set $\mathcal{R}^l = \mathcal{R}^{l-1} \cup \{s^l\}$. Note that we do not penalize groups in $\mathcal{R}^{l-1}$ that have been selected already, in deciding on the next group $s^l$. After choosing $r$ groups by this process, the debiasing step is performed to find the best maximum likelihood estimate for the sparse observation. Algorithm 1 describes this greedy approach. Note that the initial set of groups $\mathcal{R}^0$ might not be empty since we can use additional information that is independent from the PMU measurement, if available.

Algorithm 1 Greedy Heuristic

Require:
Choose an initial set of groups: $\mathcal{R}^0$.
Parameter $\tau$, $r$.
Ensure:
$\mathcal{R}^r$: Set of groups after selecting $r$ groups.
$\hat{\beta}^r$: Maximum likelihood estimate for $r$-group observation.
1: for $l = 1, 2, \ldots, r$ do
2:   Solve (12) with $\mathcal{R}^{l-1}$ for $\hat{\beta}^l$.
3:   $s^l \leftarrow \arg\max_{s \in \mathcal{S} \setminus \mathcal{R}^{l-1}} q_s(\hat{\beta}^l)$
4:   $\mathcal{R}^l \leftarrow \mathcal{R}^{l-1} \cup \{s^l\}$
5: end for
6: Solve (11) with $\mathcal{R}^r$ to get $\hat{\beta}^r$. 

(b) debiasing

The major advantage of this approach is that redundant observations are suppressed by already-selected, non-penalized observations at each iteration. We will give more details in discussing the experimental results in Section IV.

IV. NUMERICAL EXPERIMENTS

Here we present experimental results for the approaches proposed above. The test sets considered here are based on the power system test cases from MATPOWER (originally from [17]), with demands altered to generate training and test sets for the MLR approach.

A. Synthetic Data Generation

Since the data provided from IEEE test case archive [17] is a single snapshot of the states of power systems, we extend them to a synthetic 24-hour demand data cycle by using a stochastic process, as follows.

1) Take the demand values given by the IEEE test case archive as the average load demand over 24-hours.

2) Generate the demand variation profile by using an additive Ornstein-Uhlenbeck process as described in [18], separately and independently on each demand bus.

3) Combine the average demand and the variation ratio to obtain the 24-hour load demand profile for the system.

Figure 2 shows demand data generated by this procedure at three demand buses in the 9-bus system (case9.m) from MATPOWER. Figure 2(a) shows the data drawn from the MATPOWER file, now taken to be a 24-hour average. Figure 2(b) shows the ratio generated by the additive Ornstein-Uhlenbeck process, and Figure 2(c) shows the products of the average and ratio. Since the power injected to the system needs to increase proportionally to the total demands, all power generation is multiplied by the average of the demand ratios. This average of ratios is used as the generation level $G$ for the observation vector $\bar{X}$ defined by (7). The data assumes a 10-second interval between the measurements, so the total number of time points in the generated data is $24 \times 60 \times 6 = 8640$.

Once the 24-hour load demand profile is obtained, the AC power equations are solved using MATPOWER to calculate the voltage phasor values at each time point. These phasor values are taken to be the PMU measurements for a normal operation cycle over a 24-hour period. MATPOWER’s AC power flow equations solver is also used to evaluate voltage phasors for each single-line outage scenario that does not lead to an infeasible system. (During this process, if there exist duplicated lines that connect the same pair of buses, they are considered as a single line, that is, we do not allow only a fraction of multiple lines that connect the same set of buses to be failed.) Simulation of single-line failures to generate training data is necessary because there are typically few instances of actual outages available for study. The voltage variation for each line outage at time $t$ is calculated by subtracting these normal-operation voltages at timepoint $t - 1$ from line outage voltages at time point $t$. (The 10-second interval between measurements is usually sufficient time to allow transient fluctuations in phasor values to settle down; see (2).) This process leads to a number of labeled data pairs $(X,Y)$ (or $(\bar{X},Y)$) which we can use to train or tune the MLR classifier.

Table I provides the basic information on the power systems used for the experiments. The number of lines that are feasible is given in the column “Feas.”, while the number of lines that are duplicated or that lead to an infeasible power flow problem when removed from the system is shown in the column “Infeas./Dup.”. For each feasible line outage, five equally spaced samples are selected from the first half (that is, the first 12-hour period) of voltage variation data as training instances. Fifty samples are selected randomly from the second

| System | MATPOWER case | # of Lines | Train (5) | Test (50) |
|--------|---------------|------------|-----------|-----------|
| 14-Bus | case14        | 18         | 90        | 900       |
| 30-Bus | case_57       | 37         | 185       | 1850      |
| 57-Bus | case57        | 67         | 335       | 3350      |
| 118-Bus| case118       | 170        | 850       | 8500      |


(a) Base Case Demand (Considered as Average)  
(b) The Ratio Generated by Stochastic Process  
(c) Generated 24-Hour Demand Profile

Fig. 2. Generating Synthetic Demand Data by A Stochastic Process

### Table II

| System      | Using $X$ |          |          | Using $X$ |          |          |
|-------------|-----------|----------|----------|-----------|----------|----------|
|             | $\geq 0.9$ | $\geq 0.7$ | $\geq 0.5$ | $\geq 0.9$ | $\geq 0.7$ | $\geq 0.5$ |
| 14-Bus      | 100%      | 100%     | 100%     | 100%      | 100%      | 100%      |
| 30-Bus      | 99.7%     | 99.7%    | 99.7%    | 100%      | 100%      | 100%      |
| 57-Bus      | 99.5%     | 99.5%    | 99.5%    | 99.5%     | 99.5%     | 99.5%     |
| 118-Bus     | 99.5%     | 99.5%    | 99.5%    | 99.8%     | 99.8%     | 99.8%     |

(a) Based on Probability of Correct Answer

| System      | Using $X$ | $\leq 2$ | $\leq 3$ | Using $X$ | $\leq 2$ | $\leq 3$ |
|-------------|-----------|----------|----------|-----------|----------|----------|
| 14-Bus      | 100%      | 100%     | 100%     | 100%      | 100%     | 100%     |
| 30-Bus      | 99.7%     | 100%     | 100%     | 100%      | 100%     | 100%     |
| 57-Bus      | 99.8%     | 99.9%    | 99.9%    | 99.8%     | 99.9%    | 99.9%    |
| 118-Bus     | 99.5%     | 99.7%    | 99.7%    | 99.8%     | 99.9%    | 100%     |

(b) Based on Ranking of Correct Answer

- “Probability” indicates statistics for the probability assigned by the MLR classifier to the actual outage event.
- “Ranking” indicates whether the actual event was ranked in the top 1, 2, or 3 of probable outage events by the MLR classifier.

half of voltage variation data as test instances. The numbers of training and test instances are shown in the last two columns of the table.

### B. PMUs on All Buses

We present results for line outage detection when phasor measurement data from all buses is used. The maximum likelihood estimation problem with these observation vectors is solved by L-BFGS algorithm [14], coded in MATLAB. We measure performance of the identification procedure in two ways. The first measure is based on the probability assigned by the model to the actual line outage. Table III(a) shows the accuracy of the classifiers according to this measure, for both the original phasor difference vector $X$ (6) and the extended vector $\tilde{X}$ (7). Each column shows the percentage of testing samples for which the probability assigned to the correct outage exceeds 0.9, 0.7 and 0.5, respectively. The result shows that the performance of line outage identification is very good, even for the original observation vector $X$. For both of $X$ and $\tilde{X}$, the accuracy of line outage identification based on probability $\geq 0.5$ is at least 99%.

The second measure is obtained by ranking the probabilities assigned to each line outage on the test datum, and score a positive mark if the correct outage is one of the top one, two, or three cases in the ranking. We see in Table III(b) that the actual case appears in the top two in almost every case.

### C. PMU Placement

In this subsection we only consider the extended observation vector $\tilde{X}$ defined by (7). We assume too that a PMU is installed on the reference bus, for purposes of maintaining consistency in phase angle measurement. We describe in some detail the performance of the proposed algorithm on the IEEE 57 Bus system, showing that line-outage identification performance when PMUs are placed judiciously almost matches performance in the fully-instrumented case. We then summarize our computational experience on 14, 30, 57, and 118-bus systems.

For our regularization schemes, we used groups $P_s$, $s = 1, 2, \ldots, N$, defined as in (8). The final two entries in the extended observation vectors (the average-generation and constant terms) are not included in any group.

1) IEEE 57 Bus System: We describe here results obtained on the IEEE 57-bus system with two heuristics discussed in Section III: The GroupLASSO and Greedy Heuristics.

In Figure 3, results for the GroupLASSO heuristic are displayed for different values of $\tau$. The $x$-axis indicates the number of PMUs selected by this heuristic. The $y$-axis indicates the number of test cases for which the true outage was classified by the heuristic. Each bar is color-coded according to the probability assigned to the true outage by the MLR classifier. Blue colors indicate that a high probability is assigned (that is, the outage was identified correctly) while dark red colors indicate that the probability assigned to the true outage scenario is less than 0.5. For example, the second bar from the left in Figure 3(c) which corresponds to two PMUs, corresponds to the following distribution of probabilities assigned to the correct outage scenario, among the 3350 test instances.

| Probability | $[0.9, 1]$ | $[0.8, 0.9]$ | $[0.7, 0.8]$ | $[0.6, 0.7]$ | $[0.5, 0.6]$ | $[0.5, 0.5]$ |
|-------------|------------|--------------|--------------|--------------|--------------|--------------|
| # of Instances | 963        | 144          | 135          | 161          | 206          | 1411         |

Note that the dark blue color occupies a fraction 963/3350 of the bar, medium blue occupies 144/3350, and so on.

When only one PMU is installed, that bus naturally serves as the angle reference, so no phase angle difference information is
available, and identification cannot be performed. As expected, identification becomes more reliable as PMUs are installed on more buses. The value $\tau = .1$ (Figure 3(c)) appears to select locations better than the smaller choices of regularization parameter. For this value, about 10 buses are sufficient to assign a probability of greater than 90% to the correct outage event for more than 90% of the test cases, while near-perfect identification occurs when 30 PMUs are installed. Note that for $\tau = .1$, there is only slow marginal improvement after 10 buses; we see a similar pattern for the other values of $\tau$. The locations added after the initial selection are being chosen on the basis of information from the single regularized problem (9), so locations added later may be providing only redundant information over locations selected earlier.

Figure 4 shows performance of the Greedy Heuristic, plotted in the same fashion as in Figure 3. For each value of $\tau$, Algorithm I is performed with $R^0 = \emptyset$, with iterations continuing until there is no group $s \in S \setminus R^{l-1}$ such that $q_{\Delta}(\beta_l) > 0$. Termination occurs at 24, 16, and 11 PMU locations for the values $\tau = 10^{-5}$, $10^{-3}$, and $10^{-1}$, respectively. As the value of $\tau$ increases, the number of PMUs which are selected for line outage identification decreases. We can see by comparing Figures 3 and 4 that classification performance improves more rapidly as new locations are added for the Greedy Heuristic than for the GroupLASSO Heuristic. Larger values of $\tau$ give slightly better results. We note (Figure 4(c)) that almost perfect identification occurs with only 16 PMU locations, while only 6 locations suffice to identify 90% of outage events with high confidence.

Although we can manipulate the GroupLASSO technique to achieve sparsity equivalent to the Greedy Heuristic (by choosing a larger value of $\tau$), the PMUs selected by the latter give much better identification performance on this test set. In Table III the parameter $\tau$ in the GroupLASSO heuristics is chosen manually, to find the solutions with 10 PMUs and 15 PMUs for the 57 Bus system. Performance is compared to that obtained from the Greedy Heuristic, with a much smaller value of $\tau$. Results for the Greedy Heuristic are clearly superior.

2) Greedy Heuristic on 14, 30, 57 and 118 Bus System:
We applied the Greedy Heuristic to 14, 30, 57 and 118 Bus Systems with two values of $\tau = 5 \times 10^{-2}$ and $\tau = 5 \times 10^{-3}$, and found that the phasor measurements from the small set of buses are enough to have the similar line outage identification performance to the full measurement cases. Table IV shows the PMU locations selected for each case, and line outage identification performance. Identification performance is hardly degraded from the fully instrumented case, even when phasor measurements are available from only about 25% of buses.

V. CONCLUSIONS
A novel approach to identify single line outage using MLR model is proposed in this paper. The model employs historical load demand data to train a multiclass logistic regression classifier, then uses the classifier to identify outages in real
time from streaming PMU data. Numerical results obtained on IEEE 14, 30, 57 and 118 bus systems prove that the approach can identify outages reliably.

With this line outage identification framework, the optimal placement of PMU devices to identify the line outage is also discussed. Heuristics are proposed to decide which buses should be instrumented with PMUs. Experimental results show that detection is almost as good when just 25% of buses are instrumented with PMUs as when PMUs are attached to all buses.

ACKNOWLEDGMENT

The authors thank Professor Chris DeMarco and Mr. Jong-Min Lim for allowing the use of their synthetic 24-hour electric power demand data sets in our experiments, and for valuable discussions and guidance on this project.

REFERENCES

[1] U.S.–Canada Power System Outage Task Force, “Final Report on the August 14, 2003 Blackout in the United States and Canada,” 2004. [Online]. Available: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf

[2] J. E. Tate and T. J. Overbye, “Line Outage Detection Using Phasor Angle Measurements,” IEEE Transactions on Power Systems, vol. 23, no. 4, pp. 1644–1652, 2008.

[3] ——, “Double line outage detection using phasor angle measurements,” in 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Jul. 2009, pp. 1–5.

[4] A. Y. Abdelaziz, S. F. Mekhmar, M. Ezzat, and E. F. El-Saadany, “Line outage detection using Support Vector Machine (SVM) based on the Phasor Measurement Units (PMUs) technology,” in 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, Jul. 2012, pp. 1–8.

[5] H. Zhu and G. B. Giannakis, “Sparse Overcomplete Representations for Efficient Identification of Power Line Outages,” IEEE Transactions on Power Systems, vol. 27, no. 4, pp. 2215–2224, 2012.

[6] J. Chen, W. Li, and C. Wen, “Efficient Identification Method for Power Line Outages in Wide-Area Transmission System Monitoring,” IEEE Transactions on Power Systems, vol. 29, no. 4, pp. 1788–1800, 2014.

[7] H. Zhu and Y. Shi, “Phasor Measurement Unit Placement for Identifying Power Line Outages in Wide-Area Transmission System Monitoring,” in 2014 47th Hawaii International Conference on System Sciences, Waikoloa, HI, Jan. 2014, pp. 2483–2492.

[8] S. Chakrabarti, E. Kyriakides, and D. G. Eliades, “Placement of Synchronized Measurements for Power System Observability,” IEEE Transactions on Power Delivery, vol. 24, no. 1, pp. 12–19, Jan. 2009.

[9] V. Kekatos, G. B. Giannakis, and B. Wollenberg, “Optimal placement of phasor measurement units via convex relaxation,” IEEE Transactions on Power Systems, vol. 27, no. 3, pp. 1521–1530, 2012.

[10] N. M. Manousakis, G. N. Korres, and P. S. Georgilakis, “Optimal placement of phasor measurement units: A literature review,” in 2011 16th International Conference on Intelligent System Application to Power Systems (ISAP), Hersosnissos, 2011, pp. 1–6.

[11] W. Yuill, A. Edwards, S. Chowdhury, and S. P. Chowdhury, “Optimal PMU placement: A comprehensive literature review,” in 2011 IEEE Power and Energy Society General Meeting, San Diego, CA, Jul. 2011, pp. 1–8.

[12] Y. Zhao, A. Goldsmith, and H. V. Poor, “On PMU location selection for line outage detection in wide-area transmission networks,” in 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, Jul. 2012, pp. 1–8.

[13] A. R. Bergen and V. Vittal, Power systems analysis, 2nd ed. Prentice Hall, 1999.

[14] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, “MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education,” IEEE Transactions on Power Systems, vol. 26, no. 1, pp. 12–19, 2011.

[15] D. C. Liu and J. Nocedal, “On the limited-memory BFGS method for large scale optimization,” Mathematical Programming, Series A, vol. 45, pp. 503–528, 1989.

[16] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, “Sparse multinomial logistic regression: fast algorithms and generalization bounds.” IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 6, pp. 957–68, Jun. 2005.

[17] University of Washington, Electrical Engineering, “Power Systems Test Case Archive,” 2014. [Online]. Available: http://www.ee.washington.edu/research/pstca/pscas/

[18] M. Perninge, V. Knazkins, M. Amelin, and L. Söder, “Modeling the electric power consumption in a multiarea system,” European Transactions on Electrical Power, vol. 21, no. 2, pp. 413–423, 2011.

[19] B. Allauana and A. Laouifi, “Ant Colony Optimization Applied on Combinatorial Problem for Optimal Power Flow Solution,” Leonardo Journal of Sciences, no. 14, pp. 1–16, 2009.

[20] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” IEEE Transactions on Signal Processing, vol. 57, pp. 2479–2493, Aug. 2009.
Appendix A
Visualizing the Solution of the PMU Placement Problems

The location of PMUs for the IEEE 30-Bus and IEEE 57-Bus Systems (from Table IV) are displayed in Figure 5 with instrumented buses indicated by red circles.

![Diagram of PMU Locations for IEEE 30 Bus and IEEE 57 Bus Systems](image)

(a) IEEE 30 Bus System ($\tau = 5 \times 10^{-3}$, 5 PMUs)

(b) IEEE 57 Bus System ($\tau = 5 \times 10^{-3}$, 14 PMUs)

Fig. 5. PMU Locations for IEEE 30 Bus and IEEE 57 Bus Systems. (System diagrams are taken from [19, 20])

Appendix B
Extension: Use of Explicit Line Outage Information

We have assumed so far that only voltage angle and magnitude data from PMUs is used in detecting line outages. In fact, PMUs provide other information that is highly relevant for this purpose. For example, when the PMU measures current of a particular line (incident on a bus) it can detect immediately when an outage occurs on that line; we do not need to rely on the indirect evidence of voltage changes at the other PMUs. Another factor to consider is that when a decision is made to install a PMU at a particular bus, it is conventional to measure all lines that are incident on that bus, as the marginal cost of doing so is minimal. Although the phasor measurements are the same at all PMUs near a single bus, each of these PMUs provides direct information about the lines to which they are attached. Thus, if we choose to equip a particular bus with PMUs, we can immediately detect outages on all lines that touch that bus. In particular, if we install PMUs on all buses in the system, we have direct monitoring of all lines, and the outage detection problem becomes trivial.

We can extend the multiclass logistic regression technique to make effective use of these direct observations in choosing optimal buses for PMU placement. The key modification is to extend each feature vector $\beta_k$ to include additional entries that indicate the buses that touch line $k$. The observation vectors and the groups $P_s$, $s = 1, 2, \ldots, N$ are extended correspondingly.

For each line $k = 1, 2, \ldots, K$, let us define the following quantities:

$$\mathcal{T}_k := \{t^1_k, t^2_k\} \subset \{1, 2, \ldots, N\}$$

where $t^1_k$ and $t^2_k$ are indices of buses touched by line $k$.

We extend each observation vector $X$ by appending $2K$ additional elements to form $\overline{X}$, where each such vector has the form

$$\overline{X} := \begin{bmatrix} X \\ L_k \end{bmatrix}$$ (13)

for some $k = 1, 2, \ldots, K$, where $L_k$ is the $k$th column of the $2K \times K$ matrix $L$ defined as follows, for some $\eta > 0$:

$$L := \begin{bmatrix} \eta & 0 & \cdots & 0 \\ 0 & \eta & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \eta \\ 0 & 0 & \cdots & \eta \end{bmatrix}. \quad (14)$$

The two nonzero entries in each column $L_k$ indicate which two buses can detect outage of line $k$ directly. In other words, if line $k$ fails, we flag the PMUs on the buses that touch that line with a value $\eta$, since a fault on line $k$ is immediately detectable from the buses $t^1_k$ and $t^2_k$. We can say that the first part of the combined observation vector $\overline{X}$ contains indirect (voltage phasor, $\overline{X}$) observations while the second part contains direct (line outage, $L_k$) observations.

We need to extend too the definition (5) of the groups $P_s$, $s = 1, 2, \ldots, N$. We now distribute the additional $2K$ entries in the feature vector to these groups. The additional entries associated with bus $s$ are those in the index set $D_s$ defined as follows:

$$D_s = \{2(k-1) + i \mid t^i_k = s \text{ for } k = 1, 2, \ldots, K, \ i = 1, 2\}. $$

For the combined observation vector $\overline{X}$, we define groups $\mathcal{P}_s$, $s = 1, 2, \cdots, N$:

$$
\mathcal{P}_s = \mathcal{P}_s \cup \{2N + 2 + d \mid d \in \mathcal{D}_s\} \quad \text{for } s = 1, 2, \cdots, N,
$$

If PMUs are installed on every line, direct observations will identify each outage perfectly, so the solution of the maximum likelihood problem is rather trivial. In approximate solutions to the problem, the weight vector $\beta_k$ for line $k$ will have large positive entries in positions $2N + 2 + i$ for which $L_{ik} = \eta$, and large negative entries in positions $2N + 2 + i$ for which $L_{ik} = 0$. This would yield $\beta_k \overline{X}$ large and positive for observation vectors $\overline{X}$ that indicate a line-$k$ outage, with $\beta_k \overline{X}$ large and negative if there is no outage on line $k$, leading to assigned probabilities close to 1 and 0, respectively. (Entries in $\beta_k$ corresponding to the indirect observations may also have meaningfully large values, but these are less significant in the completely observed case.)

When PMUs are installed on a subset of buses, outages on some lines will be observed only indirectly, so the indirect observations in components $i = 1, 2, \ldots, 2N + 2$ of the vector $\overline{X}$ are critical to identification performance on those lines that are not directly observed.

We incorporate direct observations into our outage identification strategy in the following ways.

- **Indirect.** Direct observations are ignored. We use only the observation vector $\overline{X}$, as in Section IV.
- **Combined (Direct+Indirect).** Direct observations are incorporated into the observation vector, and we so MLR classification with the vectors $\overline{X}$.
- **Prescreening.** Instead of including the direct observation in the observation vector, line outages that can be identified by the direct observation are screened out before the MLR is applied. The number of outcomes in MLR is reduced since we do not need to consider the line outages identified already by the direct observation. Observation vectors $\overline{X}$ are used to train the MLR for those line outages that are not observed directly.
- **Postscreening.** First, we train an MLR classifier using only the indirect observations in vector $\overline{X}$. Then, during classification, we override the prediction result from the MLR when a direct observation is available for the line in question. Note that results from this strategy cannot be worse than results for the Indirect strategy.

We also compare solutions of the PMU placement problem using the Indirect strategy (as in Section IV) and the Combined strategy. We solve these two variants of the placement problem for the 57-bus case with the Greedy Heuristic of Algorithm I setting $\tau = 10^{-2}$ and the number of PMUs $r$ to the values 5 and 10. We note that the reference bus is always selected as one of the PMU locations, and it is used only to provide the phase angle reference for all the strategies above, as in the Indirect case. (In using the reference bus PMU in this restricted way, we allow a fairer comparison between the Indirect strategy and the strategies that use direct observations.)

Experimental results using PMU placements based on Indirect and Combined observations, and using each of the four classification strategies described above, are shown in Table V, using a similar format to Tables III and IV. When the PMU locations are selected using only indirect observations (Table VI(a)), the advantage of using direct line outage information during classification is not significant, especially when the larger number of 10 PMUs is installed. This observation is not too surprising. The biggest voltage phasor changes are produced by outages that are close to a bus, so even when a line outage is not detected by direct observation, it can usually be detected reliably by its “indirect” effect on nearby buses.

In Tables VI(b) and VI(c) the PMU locations are selected on the basis of the combined vectors. When only indirect data is used during classification, results are much worse, as the locations have been chosen under the assumption that direct observation data will be available. In fact, the results in Table VI(b) are generally slightly worse than those of Table VI(a). This is again because too much reliance is placed on direct observation in selecting PMU locations, and detection power is diminished slightly for those outages that are detectable only indirectly. Note that the PMU locations in Table VI(b) are essentially those with the greatest numbers of lines connected: a total of 35 in Table VI(b) (for $r = 10$), as compared with 18 in Table VI(a).

To reduce the weight placed on direct information in PMU placement, we scale down the values $\eta$ in (14). Reducing $\eta$ from 1 to $10^{-2}$ appears to strike a better balance between the use of direct and indirect information. Table VI(c) shows a marked improvement over Table VI(a) (which weights the direct observations more heavily), and slight improvements by most measures over Table VI(b) which uses only indirect information. The total number of lines that are directly connected to buses with PMUs (and which can thus be observed directly) in Table VI(b) is about halfway between the corresponding statistics in Tables VI(a) and VI(c).

**Appendix C
IMPLEMENTATION OF SPARSA**

We solve the regularized convex optimization problem (5) with the SpaRSA algorithm (21), a simple first-order approach that exploits the structure. We briefly describe the approach here, referring to (21) for further details.

The SpaRSA subproblem of (5) at iteration $n$ is defined as follows, for some scalar parameter $\alpha_n \in \mathbb{R}^+$:

$$
\beta^{n+1} := \arg \max_{\beta} \frac{1}{2} \|\beta - \gamma^n\|_F^2 + \tau \frac{1}{\alpha_n} w(\beta) \quad (15)
$$

where $\| \cdot \|_F$ is the Frobenius norm of a matrix, $\gamma^n := \begin{bmatrix} \gamma_1^n & \gamma_2^n & \cdots & \gamma_K^n \end{bmatrix}$ with

$$
\gamma_k^n := \frac{1}{\alpha_n} \nabla_{\beta_k} f(\beta),
$$

and

$$
\nabla_{\beta_k} f(\beta) = \sum_{p=1}^M x_p \sum_{j=1}^M \sum_{k=1}^K e^{(\beta_k x_j)}.
$$

When no regularization term is present ($w(\beta) = 0$), the solution for (15) is $\beta^{n+1} = \gamma^n$, so the approach reduces to the steepest descent algorithm on $f$ with step length $1/\alpha_n$. 

| $\gamma_k^n$ | $\beta_k^{n+1}$ | $f(\beta)$ |
|-------------|----------------|-------------|
| $\gamma_1^n$ | $\beta_1^{n+1}$ | $f(\beta)$ |
| $\gamma_2^n$ | $\beta_2^{n+1}$ | $f(\beta)$ |
| $\gamma_3^n$ | $\beta_3^{n+1}$ | $f(\beta)$ |
TABLE V
USE OF EXPLICIT LINE OUTAGE INFORMATION (τ = 10⁻², * indicates the reference bus.)

| Strategy       | 1⁺, 5 20 21 57 (5 PMUs) | 1⁺, 5 20 21 26 39 40 43 54 57 (10 PMUs) |
|----------------|--------------------------|------------------------------------------|
|                | Probability  | Ranking | Probability | Ranking |
|                | ≥ 0.9      | ≤ 0.7   | ≥ 0.5      | ≤ 2        | ≤ 3      | ≥ 0.9      | ≤ 0.7   | ≥ 0.5  | ≤ 2        | ≤ 3      |
| Indirect       | 83.1%      | 83.2%   | 88.0%      | 99.4%      | 99.8%    | 94.1%      | 94.1%    | 94.1%   | 99.1%      | 99.1%    |
| Combined       | 83.0%      | 83.8%   | 86.4%      | 97.5%      | 99.4%    | 95.4%      | 95.4%    | 95.4%   | 98.8%      | 98.8%    |
| Pre-screening  | 84.8%      | 85.1%   | 88.1%      | 98.4%      | 99.6%    | 95.2%      | 95.2%    | 95.3%   | 99.8%      | 99.9%    |
| Post-screening | 86.1%      | 86.2%   | 89.5%      | 99.9%      | 99.8%    | 94.2%      | 94.2%    | 94.2%   | 99.1%      | 99.1%    |

(a) PMU Placement Based on Indirect Observations

| Strategy       | 1⁺, 6 9 12 56 (5 PMUs) | 1⁺, 6 9 12 15 22 39 49 54 56 (10 PMUs) |
|----------------|--------------------------|------------------------------------------|
|                | Probability  | Ranking | Probability | Ranking |
|                | ≥ 0.9      | ≤ 0.7   | ≥ 0.5      | ≤ 2        | ≤ 3      | ≥ 0.9      | ≤ 0.7   | ≥ 0.5  | ≤ 2        | ≤ 3      |
| Indirect       | 77.2%      | 77.4%   | 83.1%      | 97.4%      | 99.0%    | 83.7%      | 83.7%    | 87.5%   | 93.5%      | 93.5%    |
| Combined       | 83.1%      | 83.1%   | 87.4%      | 92.7%      | 92.8%    | 93.1%      | 93.1%    | 94.7%   | 98.4%      | 98.5%    |
| Pre-screening  | 83.3%      | 83.8%   | 88.2%      | 93.4%      | 93.4%    | 93.8%      | 93.8%    | 95.2%   | 98.4%      | 98.4%    |
| Post-screening | 82.7%      | 82.8%   | 87.2%      | 98.5%      | 99.6%    | 93.7%      | 93.7%    | 95.1%   | 97.4%      | 97.4%    |

(b) PMU Placement Based on Combined (Direct + Indirect) Observations (η = 1)

| Strategy       | 1⁺, 5 9 49 56 (5 PMUs) | 1⁺, 5 9 21 26 39 45 49 56 (10 PMUs) |
|----------------|--------------------------|------------------------------------------|
|                | Probability  | Ranking | Probability | Ranking |
|                | ≥ 0.9      | ≤ 0.7   | ≥ 0.5      | ≤ 2        | ≤ 3      | ≥ 0.9      | ≤ 0.7   | ≥ 0.5  | ≤ 2        | ≤ 3      |
| Indirect       | 79.0%      | 79.3%   | 84.3%      | 98.9%      | 99.8%    | 93.3%      | 93.3%    | 94.7%   | 99.4%      | 99.9%    |
| Combined       | 85.4%      | 85.4%   | 90.0%      | 95.5%      | 95.6%    | 97.4%      | 97.4%    | 97.4%   | 99.9%      | 99.9%    |
| Pre-screening  | 84.5%      | 84.5%   | 88.5%      | 98.5%      | 99.8%    | 96.1%      | 96.1%    | 96.1%   | 97.9%      | 97.9%    |
| Post-screening | 85.1%      | 85.4%   | 89.5%      | 99.7%      | 100%     | 98.0%      | 98.0%    | 98.0%   | 100%       | 100%     |

(c) PMU Placement Based on Combined (Direct + Indirect) Observations (η = 10⁻²)

In the PMU placement problem, our regularizer $w_2(\beta)$ is group-separable. Thus the subproblem (13) can be divided into independent problems of the form

$$[\beta^{n+1}]_s := \arg\max_{\beta} \frac{1}{2} \|[\beta]_s - [\gamma^n]_s\|_2^2 + \tau \frac{1}{\alpha_n} q_s(\beta),$$

for all $s \in S$, where (as defined above), $[A]_s$ is the submatrix of $A$ consisting of the rows whose indices are in $P_s$. Since the penalty function $q_s(\beta)$ is the $\ell_2$-norm, this subproblem has a closed form solution (21), as follows:

$$[\beta^{n+1}]_s = [\gamma^n]_s \frac{\max\{\|[\gamma^n]_s\|_2 - \tau \alpha_n^{-1}, 0\}}{\max\{\|[\gamma^n]_s\|_2 - \tau \alpha_n^{-1}, 0\}} + \tau \alpha_n^{-1} \cdot [\gamma^n]_s.$$

For any row $i$ of $\beta$ that does not belong to any $P_s$, we have simply $[\beta^{n+1}]_i = [\gamma^n]_i$. Different strategies can be used to choose $\alpha_n$. We increase $\alpha_n$ at each iteration until sufficient decrease is obtained in the objective, terminating when $\alpha_n$ grows too large (indicating that a solution is nearby).