On-line Dynamic Security Assessment with Missing PMU Measurements: A Data Mining Approach

Miao He, *Student Member, IEEE,* Vijay Vittal, *Fellow, IEEE,* and Junshan Zhang, *Fellow, IEEE*

***Abstract*—A data mining approach using ensemble decision trees (DTs) learning is proposed for on-line dynamic security assessment (DSA), with the objective of mitigating the impact of possibly missing PMU data. Specifically, multiple small DTs are first trained off-line using a random subspace method. In particular, the developed random subspace method exploits the hierarchy of wide-area monitoring system (WAMS), the locational information of attributes, and the availability of PMU measurements, so as to improve the overall robustness of the ensemble to missing data. Then, the performance of the trained small DTs is re-checked by using new cases in near real-time. In on-line DSA, viable small DTs are identified in case of missing PMU data, and a boosting algorithm is employed to quantify the voting weights of viable small DTs. The security classification decision for on-line DSA is obtained via a weighted voting of viable small DTs. A case study using the IEEE 39-bus system demonstrates the effectiveness of the proposed approach.**

***Index Terms*—On-line dynamic security assessment, transient stability, missing PMU data, decision tree, data mining, ensemble learning, boosting, random subspace method.**

I. Introduction

DSA provides system operators with important information, e.g., transient security of a specific operating condition (OC) under various contingencies. Given a knowledge base, DTs can identify the attributes and the thresholds that are critical to assessing the transient performance of power systems [1], [2]. With the advent of synchrophasor technologies, a significant amount of effort has been directed towards *on-line* DSA, by using PMU measurements directly for decision making [3]–[6]. Upon a disturbance, by applying pre-determined decision rules to the PMU measurements of critical attributes, DTs can give security classification decisions in real-time.

Previous studies on PMU measurement-based on-line DSA implicitly assume that wide area monitoring systems (WAMS) provide reliable measurements. However, in on-line DSA, PMU measurements can become unavailable due to the unexpected failure of the PMUs or phasor data concentrators (PDCs), or due to loss of the communication links. Recently, it has been widely recognized that PMU failure can be an important factor that impacts the performance of WAMS. For example, AESO's newest rules on implementing PMUs [7] require that the loss or malfunction of PMUs, together with the cause and the expected repair time, has to be reported to the system operator in a timely manner. In the report [8], the deployment of redundancy is suggested by PMU manufacturers to reduce the impact of single PMU failure. Loss of PMUs has also been taken into account when designing WAMS and PMU placement [9]. Moreover, the delivery of PMU measurements from multiple remote locations of power grids to monitoring centers could experience high latency when communication networks are heavily congested, which could also result in the unavailability of PMU measurements. Therefore, it is urgent to design DT-based on-line DSA approaches that are robust to missing PMU measurements.

Intuitively, one possible approach to handle missing PMU measurements is to estimate the missing values by using other PMU measurements and the system model. However, with existing non-linear state estimators in supervisory control and data acquisition (SCADA) systems, this approach may compromise the performance of DTs. First, the scan rate of SCADA systems is far from commensurate with the data rate of PMU measurements, and thus using estimated values from SCADA data may result in a large delay for decision making. Second, SCADA systems collect data from remote terminal units (RTUs) utilizing a polling approach. Following a disturbance, it is possible that some post-contingency values are used due to the lack of synchronization, which can lead to inaccurate security classification decisions of DTs. It is worth noting that future fully PMU-based linear state estimators [10] can overcome the aforementioned limitations; but this is possible only when there is a sufficient number of PMUs placed in system. With this motivation, data-mining based approaches are investigated in this paper, aiming to use alternative viable measurements for decision making in case of missing data.

In DTs built by the classification and regression tree (CART) algorithm [11], missing data can be handled by using *surrogate*. However, a critical observation in this study is that when PMU measurements are used as attributes, most viable surrogate attributes have low associations with the primary attributes. Clearly, the accuracy of DSA would degrade if surrogate is used. This is because a DT is essentially a sequential processing method, and thus the wrong decisions made in earlier stages may have significant impact on the correctness of the final decisions. Thus motivated, this paper studies applying ensemble DT learning techniques, including random subspace methods and boosting, to improve the robustness to missing PMU measurements.

Manuscript received April 30, 2012; revised November 16, 2012 and January 6, 2013; accepted February 7, 2013. This work was supported in part by the US National Science Foundation under grant CPS-1035906, in part by the DTRA grant HDTRA1-09-1-0032, and in part by the Power System Engineering Research Center. Paper no. TPWRS-00446-2012.

The authors are with School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA (e-mail: Miao.He@asu.edu; Vijay.Vittal@asu.edu; Junshan.Zhang@asu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2013.xxxxxxx

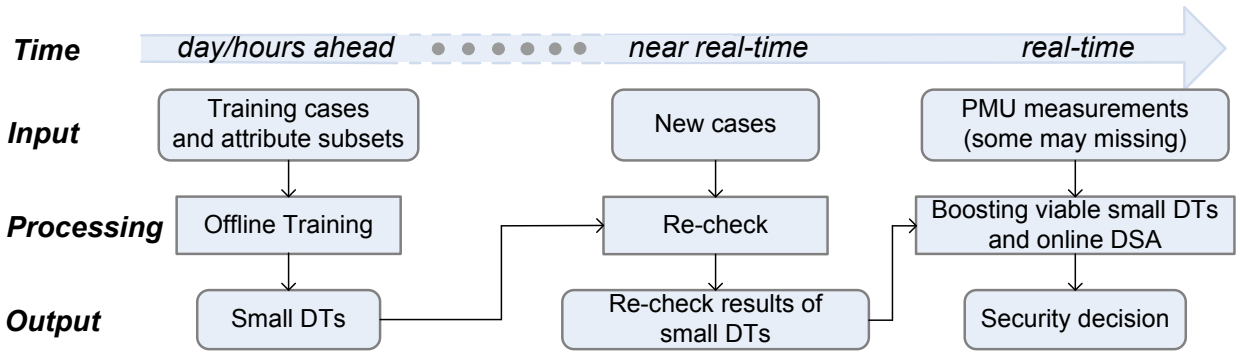

Fig. 1. Three-stage ensemble DT-based approach to on-line DSA with missing PMU measurements

Aiming to develop a *robust* and *accurate* on-line DSA scheme, the proposed approach consists of three processing stages, as illustrated in Fig. 1. Specifically, given a collection of training cases, multiple small DTs are trained off-line by using randomly selected attribute subsets. In near real-time, new cases are used to re-check the performance of small DTs. The re-check results are then utilized by a boosting algorithm to quantify the voting weights of a few *viable* small DTs (i.e., the DTs without missing data from their attribute subsets). Finally, security classification decisions of on-line DSA are obtained via a weighted voting of viable small DTs. More specifically, a random subspace method for selecting attribute subsets is developed by exploiting the locational information of attributes and the availability of PMU measurements. Conventionally, the availability of a WAMS is defined as the probability that the system is operating normally at a specified time instant [12]. In this study, the *availability* of PMU measurements is defined similarly, i.e., as the *probability* that PMU measurements are successfully collected and delivered to a monitoring center. The developed random subspace method guarantees that a significant portion of small DTs are viable for on-line DSA with high likelihood. Further, a boosting algorithm is employed to assign the viable small DTs with proper voting weights that are quantified by using the results from performance re-check, leading to the high robustness and accuracy of the proposed approach in case of missing PMU measurements. The proposed approach is applied to the IEEE 39-bus system with 9 PMUs. Compared to off-the-shelf DT-based techniques (including random forests (RFs) with and without using surrogate), the proposed ensemble DT-based approach can achieve better performance in case of missing PMU measurements.

The rest of the paper is organized as follows. An introduction to DTs with application to DSA is given in Section II. Section III focuses on the random subspace method for selecting attribute subsets. The proposed three-stage approach is presented in detail in Section IV. A case study is discussed in Section V. Finally, conclusions are given in Section VI.

II. Background on DTs

A *decision tree* is a tree-structured model that maps the measurements of the *attributes* $\mathbf{x}\in\mathcal{X}$ to a predicted value $\hat{y}\in\mathcal{Y}$ [11]. In a DT, a test on an attribute (thus called the primary attribute of the internal node) is installed at each internal node and decides which child node to drop the measurements into. Further, each leaf node of the DT is assigned a predicted value, and the measurements are thus labeled with the predicted value of the leaf node which it sinks into. The path from the root node to a leaf node specifies a decision region in the attribute space corresponding to that leaf node. Specifically, the length of the longest downward path from the root node to a leaf node is defined as the height of a DT.

A. Application of DTs in DSA

DTs with binary predicted values (i.e., $\mathcal{Y}=\{\pm1\}$) are used in DSA. Specifically, $\hat{y}=+1$ represents that an OC is classified as *insecure* under a given contingency. The numerical attributes used by DTs in DSA include voltage magnitudes, voltage phase angles and power/current flows. Moreover, the index of contingencies is used as a categorical attribute. In DSA, a collection of training cases are first created by applying DSA packages (e.g., DSATools$^{\text{TM}}$ [13]) to N_{OC} known OCs for a given list of N_C contingencies. Then, the training cases $\{\mathbf{x}_n, y_n\}_{n=1}^{N}$, where $N=N_{OC}\times N_C$, are used by the classification and regression tree (CART) algorithm [11] to build a DT that fits the training data. Intuitively, if the DT fits the training data well and the new OCs in on-line DSA are similar to the OCs corresponding to the training cases, the trained DT can give accurate security classification decisions for the new OCs in on-line DSA.

In this study, *small* DTs, which have a small height J are used. Generally, a small DT could have lower accuracy than a fully-grown DT, but is less prone to overfitting when the training data is noisy [14], and multiple DTs are usually combined together to improve the classification accuracy.

B. Handling Missing Data by using Surrogate in DTs

A surrogate split at an internal node is the one that "mimics" the primary split most closely, i.e., gives the most similar splitting results for the training cases. Usually, the similarity is quantified by the *association* between the surrogate split and the primary split [11]. The significance of a surrogate split that has a high association (i.e., over 0.9) with the primary split is that the DT could still use the surrogate split at this

TABLE I
CASE STUDY ON THE SURROGATES OF DTS

node	primary attribute[1]	by modified CART		by CART	
		surrogate	assoc.	surrogate	assoc.
1	$V_{\{217\}}$	$V_{\{207\}}$	0.76	$V_{\{207\}}$	0.76
2	$Q_{\{204,207\}}$	$Q_{\{212,216\}}$	0.33	$Q_{\{207,209\}}$	0.50
3	$Q_{\{204,207\}}$	$V_{\{209\}}$	0.28	$Q_{\{207,209\}}$	0.64
4	$I_{\{211,204\}}$	$P_{\{008,011\}}$	0.62	$P_{\{209,211\}}$	0.83
5	$P_{\{210,201\}}$	$P_{\{211,062\}}$	0.87	$P_{\{231,201\}}$	0.87
6	$Q_{\{005,033\}}$	$Q_{\{801,999\}}$	0.71	$Q_{\{801,999\}}$	0.71
7	$P_{\{213,222\}}$	$Q_{\{207,211\}}$	0.85	$P_{\{222,223\}}$	0.85
8	$Q_{\{041,060\}}$	$I_{\{011,051\}}$	0.50	$I_{\{011,051\}}$	0.50
9	$P_{\{211,062\}}$	$P_{\{213,216\}}$	0.50	$I_{\{062,211\}}$	0.75
10	$P_{\{236,219\}}$	$Q_{\{230,052\}}$	0.42	$P_{\{236,207\}}$	0.68

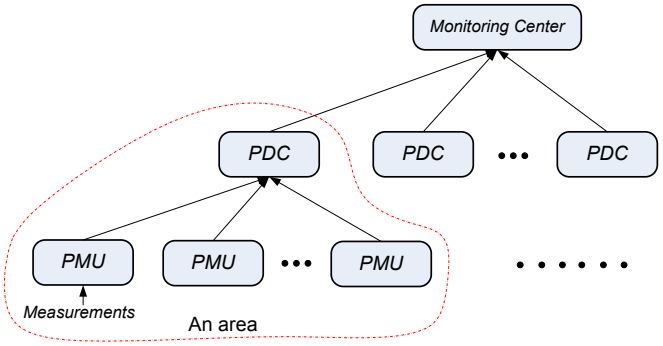

Fig. 2. Wide area monitoring system consisting of multiple areas

internal node to give almost the same decisions when the PMU measurement of the primary attribute is missing.

The performance of surrogate in DT-based DSA is evaluated via a case study, in which a single DT is built by using the same knowledge base for voltage magnitude violation analysis as in [6]. It is observed that *co-located* attributes (i.e., the attributes measured by the same PMU) would often be unavailable at the same time when the PMU fails, which implies that co-located attributes cannot be used as surrogate for each other in on-line DSA. Therefore, a modified CART algorithm in which co-located attributes are excluded from surrogate searching is used to build a single DT and identify the surrogate attributes. The results regarding the performance of the surrogates identified by both the modified CART algorithm and the CART algorithm are given in Table I. Two key observations are drawn. First, the results obtained by the modified CART algorithm suggest that all non-co-located surrogates have relatively low associations with the primary ones. The low association could be explained by the complex coupling structure of the attributes in power systems. According to the definition of surrogate, high association relies on the dependency between the surrogate and the primary attributes, i.e., the surrogate attribute gives similar decisions to the primary attribute on all the training cases regardless of any other attribute. However, in power systems, one attribute (i.e., voltage magnitude, voltage phase angle or power/current flow) is coupled with many other non-co-located attributes, as dictated by the AC power flow equations and the network interconnection structure. Second, it is observed in Table I that the surrogate attributes found by the CART algorithm are mostly co-located with the primary attributes. This observation signifies the redundancy between co-located attributes when used for splitting the training cases, and thus sheds lights on exploiting the locational information to create the attribute subsets, as described in Section III.

C. Ensemble DT Learning

Ensemble DT learning techniques (bagging, random subspace methods, boosting, RF [15]) combine multiple DTs to obtain better prediction performance. Studies (e.g., [16]) have shown that using random subspace methods can lead to improved accuracy and generalization capability, if the DTs are trained from a variety of compact and non-redundant attribute subsets. Usually, the attribute subsets used by DTs are selected in a randomized manner. For example, in the random decision forest algorithm [17], each DT is built by using an attribute subset that is randomly selected from all possible candidate attribute subsets with equal weights. For on-line DSA, it is observed that additional system information on the attributes could be utilized to create and select the attribute subsets. First, the candidate attribute subsets could be significantly refined by exploiting the locational information of attributes. Further, by putting more weights on the attribute subsets that have higher availability when randomly selecting attribute subsets, the resulting small DTs would be more likely to be robust to possibly missing PMU measurements.

[1]Bus numbers are given in the subscripts of attributes, but are different from the real ones of the practical system. For example, $Q_{\{204,207\}}$ and $Q_{\{207,209\}}$ are co-located; they are measured by the same PMU at bus 207.

III. RANDOM SUBSPACE METHOD FOR SELECTING ATTRIBUTE SUBSETS

A key step of the random subspace method is to identify a collection of candidate attribute subsets $\mathcal{S}$ and determine the weight $p_{\mathbf{s}}$ that dictates how likely a candidate attribute subset $\mathbf{s}\in\mathcal{S}$ is to be selected. In this study, by exploiting the locational information of attributes and the availability of PMU measurements, the random subspace method adheres to the following two guidelines:

- **G1**: Co-located attributes do not co-exist within an attribute subset.
- **G2**: The average availability of the selected attribute subsets should be sufficiently high.

Further, for a power system consisting of K areas, the corresponding WAMS is assumed to have a hierarchical architecture [18]. As illustrated in Fig. 2, each area of the power system has a PDC that concentrates the PMU measurements of this area and submits them to the monitoring center.

A. Candidate Attribute Subsets

The candidate attribute subsets are created based on the three following specific rules: **1**) Within a candidate attribute subset, all the attributes are from the same area. **2**) In area k (k=1,$\cdots$,K), three categories of pre-fault quantities measured by PMUs are used as the numerical attributes:

- Category 1: voltage magnitude V_i, for $i\in\mathcal{I}_k^{PMU}$;

- Category 2: active power flow P_{ij}, reactive power flow Q_{ij} and current magnitude I_{ij}, for $i \in \mathcal{I}_k^{PMU}$ and $j \in \mathcal{N}(i)$;
- Category 3: phase angle difference θ_{ij}, for $i,j \in \mathcal{I}_k^{PMU}$.

where $\mathcal{I}_k^{PMU}$ denotes the collection of the buses with PMU installation within area k, and $\mathcal{N}(i)$ denotes the collection of the neighbor buses of bus i. An attribute subset of area k is created by including one voltage or flow measurement from each bus $i \in \mathcal{I}_k^{PMU}$ and all phase angle difference measurements from this area. **3**) The index of contingencies is included as a categorical attribute in any attribute subset.

The criteria used in creating the attribute sets are elaborated below. By restricting the attributes of a subset to be the PMU measurements within the same area, the impact of some scenarios, i.e., when a PDC that concentrates PMU measurements within an area fails, is significantly reduced, since the small DTs using the PMU measurements from the other areas could still be viable. For a given bus, since Category 1 and Category 2 PMU measurements are co-located, it suffices to include only one of them in an attribute subset so that the redundancy within an attribute subset is minimal. Further, all measurable phase angle differences are included. This is because theoretical and empirical results (e.g., in [4]) suggest that angle differences contain important information regarding the level of stress in OCs, and thus are more likely to be the attributes critical to assessing transient instability. It is also worth noting that the Category 2 attributes from two different buses are unlikely to be redundant, in the sense that they are the measurements from different transmission lines, given the fact that PMUs could provide power flow measurements and it is usually unnecessary to place PMUs at both ends of a transmission line to achieve the full observability of power grids.

For convenience, let $\mathcal{S}_k$ denote the collection of candidate attribute subsets of area k. Then, the size of $\mathcal{S}_k$ is given by

$$M_k = \prod_{i \in \mathcal{I}_k^{PMU}} (3\deg(i) + 1), \tag{1}$$

where $\deg(i)$ denotes the degree of bus i, i.e., the number of buses that connect with bus i. Then, $\mathcal{S} = \bigcup_{k=1}^{K} \mathcal{S}_k$ is the collection of candidate attribute subsets.

B. Randomized Algorithm for Selecting Attribute Subsets

It is plausible to develop the randomized algorithm so as to achieve maximum randomness of the selected attribute subsets by maximizing the entropy of the weight distribution $\{p_{\mathbf{s}}, \mathbf{s} \in \mathcal{S}\}$. Without any other information of attribute, equal weights is usually used by existing random subspace methods (e.g., [19], [20]). Here, by adhering to guideline **G2**, an additional constraint is that the average availability of the randomly selected attribute subsets is above an acceptable level A_0. As a result, the weight distribution can be determined by solving the following problem:

$$\mathcal{P}_{\mathbf{s}}: \max_{\{p_{\mathbf{s}}, \mathbf{s} \in \mathcal{S}\}} \sum_{\mathbf{s} \in \mathcal{S}} p_{\mathbf{s}} \log_2 p_{\mathbf{s}}^{-1} \tag{2}$$

$$\text{s.t.} \quad \sum_{\mathbf{s} \in \mathcal{S}} p_{\mathbf{s}} A_{\mathbf{s}} \geq A_0, \tag{3}$$

$$\sum_{\mathbf{s} \in \mathcal{S}} p_{\mathbf{s}} = 1, \tag{4}$$

where $A_{\mathbf{s}}$ denotes the availability of an attribute subset $\mathbf{s}$. According to the rules for creating the candidate attribute subsets, it is easy to see that each of the attribute subsets of an area consists of exactly two measurements from each PMU within this area. Therefore, the availability of an attribute subset $\mathbf{s}$ of area k, which was formally defined in Section I as the probability that the measurements of $\mathbf{s}$ are successfully delivered to the monitoring center, equals that of the WAMS within area k, i.e.,

$$A_{\mathbf{s}} = A_k, \quad \forall \mathbf{s} \in \mathcal{S}_k. \tag{5}$$

In availability analysis of WAMS (e.g., in [12]), it is usually assumed that the availability of PMUs, PDCs and communication links are known (e.g., estimated from past operating data) and independent from each other. Under these assumptions, the availability of the WAMS within area k is given by:

$$A_k = \prod_{i \in \mathcal{I}_k^{PMU}} (A_i^{PMU} A_i^{link}) \tilde{A}_k^{PDC} \tilde{A}_k^{link}, \tag{6}$$

where A_i^{PMU}, A_i^{link}, $\tilde{A}_k^{PDC}$ and $\tilde{A}_k^{link}$ denote the availability of the PMU at bus i, the communication link from the PMU at bus i to the PDC, the PDC and the communication link from the PDC to the monitoring center, respectively. It is worth noting that (3) and (4) are derived for the case illustrated in Fig.2, and thus may not be directly applicable to the cases with measurement redundancy. For example, when multiple dual use PMU/line relays are utilized in substations, the availability of bus voltage phasor measurements can be enhanced. The procedure for analyzing the availability of WAMS in case of redundancy can be found in the literature (e.g., [21]).

By taking (5) into account, it follows that the solution to problem $\mathcal{P}_{\mathbf{s}}$ in (2) has the following property.

Proposition 3.1: The optimal solution to $\mathcal{P}_{\mathbf{s}}$ in (2) takes the following form:

$$p_{\mathbf{s}}^* = p_k^* / M_k, \quad \forall \mathbf{s} \in \mathcal{S}_k, \tag{7}$$

where M_k is the size of $\mathcal{S}_k$ as defined in (1), and $\{p_k^*, k=1,\cdots,K\}$ is the solution to the following problem:

$$\tilde{\mathcal{P}}_{\mathbf{s}}: \min_{p_1,\cdots p_K} \sum_{k=1}^{K} p_k \log_2(p_k / M_k), \tag{8}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} p_k A_k \geq A_0, \tag{9}$$

$$\sum_{k=1}^{K} p_k = 1. \tag{10}$$

Proof: Since $\mathcal{P}_{\mathbf{s}}$ maximizes a concave function with affine constraints, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for a solution to be optimal. Therefore,

$$(1 + \ln p_{\mathbf{s}}^*)/\ln 2 - \lambda^* A_{\mathbf{s}} + \mu^* = 0, \quad \forall \mathbf{s} \in \mathcal{S} \tag{11}$$

where λ^* and μ^* are the KKT multipliers for the two constraints of $\mathcal{P}_{\mathbf{s}}$. Then, by taking the equality in (5) into account,

it is easy to verify that $p_{\mathbf{s}}^*$ have the same value for all $\mathbf{s}\in\mathcal{S}_k$. Define $p_k=M_k p_{\mathbf{s}}$ for $\mathbf{s}\in\mathcal{S}_k$, then $\mathcal{P}_{\mathbf{s}}$ reduces to $\tilde{\mathcal{P}}_{\mathbf{s}}$. ■

The above result leads to the following implementation of the randomized algorithm, as summarized in Algorithm 1. Further, it is also observed from (11) that the attribute subsets which have higher availability are assigned higher weights.

Algorithm 1 Randomized algorithm for selecting an attribute subset

1: Calculate M_k and A_k according to (1) and (6), respectively, for $k = 1, \cdots, K$.
2: Find $\{p_k, k = 1, \cdots, K\}$ by solving $\tilde{\mathcal{P}}_{\mathbf{s}}$ in (8).
3: Select an area k among the K areas with weight p_k.
4: For the chosen area k, select an attribute subset s from $\mathcal{S}_k$ with weight $1/M_k$.

IV. Proposed Approach for On-line DSA with Missing PMU Measurements

First, L small DTs are trained off-line by using randomly selected attribute subsets. In case of missing PMU measurements in on-line DSA, $\tilde{L}$ ($\tilde{L}\leq L$) viable small DTs are identified, and are assigned different voting weights. Specifically, the results of performance re-check in near real-time are utilized to quantify these voting weights. Finally, the security classification decisions for the new OCs in on-line DSA are obtained via weight voting of the $\tilde{L}$ viable small DTs.

A. Off-line Training

Given a collection of training cases $\{\mathbf{x}_n, y_n\}_{n=1}^N$ and candidate attribute subsets $\mathcal{S}$, a primary objective of off-line training is to obtain small DTs $\{h_1,\cdots,h_L\}$ so that the majority voting of them, i.e., $F_L(\mathbf{x}) = \sum_{l=1}^L h_l(\mathbf{x})$ could fit the training data. The iterative process to obtain a F_L is summarized in Algorithm 2. In the l-th iteration, a small DT h_l is first obtained by solving the following problem:

$$\mathcal{P}_{DT}^{(l)} : \min_{h_l} \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{y_n \neq h_l(\mathbf{x}_n^l)\}}, \tag{12}$$

where $\mathbf{x}_n^l$ denotes the measurements of the attribute subset $\mathbf{s}_l$. It is well-known that the problem in (12) is NP-complete [22]. Here, the CART [11] algorithm is employed to find a sub-optimal DT, by using misclassification error rate as the splitting cost function. It is clear from (12) that equal weights, i.e., $\frac{1}{N}$, are assigned to all training data. When historical data that identifies potential weak spots of the system is available, these data can be integrated by assigning higher weights, and by replacing $\frac{1}{N}$ with unequal data weights.

B. Near Real-time Performance Re-check

In near real-time, a more accurate prediction of the imminent OC in on-line DSA can be made. Then, a collection of new cases $\{\tilde{\mathbf{x}}_n, \tilde{y}_n\}_{n=1}^{\tilde{N}}$ are created in a similar manner to that in off-line training and used to re-evaluate the accuracy of the L small DTs. The re-check results are then utilized by the boosting process in on-line DSA. In case of variations between the OCs used in off-line training and the new OCs in on-line DSA, near real-time re-check is also a critical step to make sure that the small DTs still work well.

Algorithm 2 Off-line training using the random subspace method

1: **Input:** Training cases $\{\mathbf{x}_n, y_n\}_{n=1}^N$, $\varepsilon_0 \in (0, 1)$.
2: **Initialization:** $F_0 = \mathbf{0}$.
3: **for** $l = 1 \rightarrow L$ **do**
4: Select an attribute subset $\mathbf{s}_l$ by using Algorithm 1.
5: Find a small DT h_l by solving $\mathcal{P}_{DT}^{(l)}$ in (12) using the CART algorithm.
6: $F_l \leftarrow F_{l-1} + h_l$.
7: **end for**

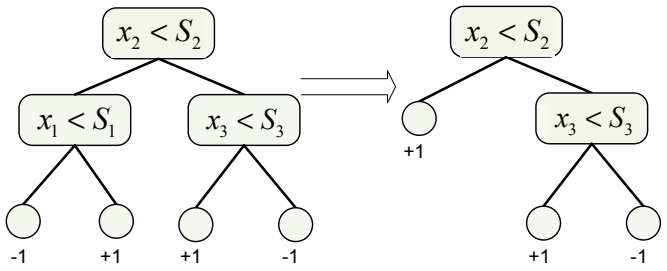

Fig. 3. Degeneration of a small DT as a result of missing PMU measurements of attribute x_1 when node ($x_1<S_1$) is originally assigned $+1$.

C. On-line DSA

The results of near real-time re-check $\{\tilde{h}_l(\tilde{\mathbf{x}}_n^l), \tilde{y}_n\}_{n=1}^{\tilde{N}}$, $\forall l=1,\cdots,\tilde{L}$ are utilized to choose a few viable small DTs to be used in on-line DSA and calculate the corresponding voting weights via a process of boosting small DTs. In order to make best use of existing DTs, the viable small DTs in on-line DSA include the small DTs without any missing PMU measurement and non-empty degenerate small DTs.

1) Degenerate Small DTs: A *degenerate* small DT is obtained by collapsing the subtree of an internal node with missing PMU measurement into a leaf node. Specifically, a small DT degenerates to a non-empty tree if the PMU measurements used by the internal nodes other than the root node are missing, an example of which is illustrated in Fig. 3. Further, since each internal node of the original small DT is also assigned a decision in building the DT, the new leaf node of the degenerate small DT is assigned the same decision as the original internal node. Therefore, for a non-empty degenerate small DT, the re-check results on the $\tilde{N}$ new cases could be easily obtained.

2) Weighted Voting of Viable Small DTs: Let $\tilde{\mathcal{H}}$ be the collection of viable small DTs. Then, weighted voting of the viable small DTs in $\tilde{\mathcal{H}}$ is utilized to obtain the security classification decisions of on-line DSA, due to the following two reasons. First, in case that some small DTs degenerate to empty trees and the accuracy of non-empty degenerate small DTs degrades, weighted voting could improve the overall accuracy compared to majority voting, provided that the voting

weights are carefully assigned based on the re-check results of the viable small DTs. Second, even though all the small DTs are viable, choosing the small DTs with proper voting weights based on their accuracy can still be a critical step to guarantee accurate decisions. This is because small DTs trained off-line fit the training cases that are created based on day-ahead prediction, while the re-check results on the $\tilde{N}$ new cases contain more relevant information on assessing the security of the imminent OCs in on-line DSA.

In the proposed approach, weighted voting of small DTs in $\tilde{\mathcal{H}}$ is implemented via a boosting process. Following the method in [6], initially with $\tilde{F}_0$ as a zero function, a small DT $\tilde{h}_l \in \tilde{\mathcal{H}}$ is first identified and added to $\tilde{F}_{l-1}$, i.e.,

$$\tilde{F}_l = \tilde{F}_{l-1} + a_l \tilde{h}_l, \tag{13}$$

iteratively for l=1,2,$\cdots$,$\tilde{L}$, so that the cost function, i.e.,

$$\hat{C}(\tilde{F}_{\tilde{L}}) = \frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} \log_2(1 + \mathrm{e}^{-\tilde{y}_n \tilde{F}_{\tilde{L}}(\tilde{\mathbf{x}}_\mathbf{n})}), \tag{14}$$

is minimized in a gradient descent manner. In the boosting process, $\tilde{h}_l$ is identified by solving the following problem:

$$\tilde{\mathcal{P}}_{DT}^{(l)} : \min_{h_l \in \tilde{\mathcal{H}}} \frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} w_n^{(l)} \mathbf{1}_{\{\tilde{y}_n \neq \tilde{h}_l(\tilde{\mathbf{x}}_n^l)\}}, \tag{15}$$

and the data weights and voting weight are given by

$$\begin{cases} w_n^{(l)} = \frac{1}{1+\mathrm{e}^{\tilde{y}_n \tilde{F}_{l-1}(\tilde{\mathbf{x}}_n)}} & n = 1, \cdots, \tilde{N} \\ a_l = \underset{a \in \mathcal{R}^+}{\operatorname{argmin}}\, g_l(a) \end{cases} \tag{16}$$

where $g_l(a) \triangleq \hat{C}(\tilde{F}_{l-1} + a\tilde{h}_l)$. Boosting viable small DTs in on-line DSA is summarized in Algorithm 3.

Algorithm 3 Boosting viable small DTs for on-line DSA

1: **Input:** Re-check results $\{\tilde{h}_l(\tilde{\mathbf{x}}_n^l), \tilde{y}_n\}_{n=1}^{\tilde{N}}, \forall l = 1, \cdots, \tilde{L}$,
2: **Initialization:** $\tilde{F}_0 = \mathbf{0}$.
3: **for** $l = 1 \rightarrow \tilde{L}$ **do**
4: Calculate the data weights according to (16).
5: Find a small DT $\tilde{h}_l$ from viable DTs by solving $\tilde{\mathcal{P}}_{DT}^{(l)}$ in (15).
6: Calculate the voting weight a_l according to (16).
7: $\tilde{F}_l \leftarrow \tilde{F}_{l-1} + a_l \tilde{h}_l$.
8: **end for**

D. Further Discussion

Through detailed complexity analysis, it is shown that the low computational complexity of the on-line processing renders that the time criticality of on-line DSA would not be compromised when the proposed approach is used. Specifically, the computationally intensive part of the on-line processing stage is the boosting process that consists of calculating the data weights $w_n^{(l)}$, solving $\tilde{\mathcal{P}}_{DT}^{(l)}$ and calculating the voting weights a_l of small DTs. According to (16), calculating the data weights requires evaluating $\tilde{F}_l$ for the new cases, which could be easily obtained from the re-check results of the small DTs. Therefore, it is easy to see that the complexity in calculating the data weights is $\mathcal{O}(\tilde{N})$. Solving $\tilde{\mathcal{P}}_{DT}^{(l)}$ boils down to searching for the small DT in $\tilde{\mathcal{H}}$ that has the least weighted misclassification error. Since the re-check results of the small DTs in $\tilde{\mathcal{H}}$ for the new cases are already known, the optimal small DT could be found by comparing the weighted misclassification errors of the small DTs in $\tilde{\mathcal{H}}$. Therefore, the complexity in solving $\tilde{\mathcal{P}}_{DT}^{(l)}$ is $\mathcal{O}(\tilde{L}\tilde{N})$. In the l-th iteration of the boosting process, the voting weight is obtained by minimizing $g_l(a)$. It is easy to verify that $g_l'(0)<0$ and $g_l''(a)>0$ holds for $a \in \mathcal{R}^+$. Therefore, $g_l(a)$ has a unique minimum in R^+ that could be found by using standard numerical methods (e.g., Newton's methods). Further, since $g_l(a)$ is convex, standard numerical methods could find the minimum in a few iterations. In each iteration, $\tilde{F}_{l-1}+a\tilde{h}_l$ needs to be evaluated for all the $\tilde{N}$ new cases. Therefore, the complexity in calculating the voting weight for a small DT is $\mathcal{O}(\tilde{N})$. Summarizing, the overall computational complexity of the boosting process is $\mathcal{O}(\tilde{L}^2\tilde{N})$.

The proposed approach above relates to that in [6] in the following sense: small DTs are utilized in both approaches; new cases are used in near real-time for accuracy guarantee by both approaches; the security classification decisions of on-line DSA are both obtained via a weighted voting of small DTs. However, the two approaches are tailored towards different application scenarios. The approach proposed here is more robust to missing PMU measurements, while the approach in [6] could give accurate decisions with less effort in off-line training when the availability of PMU measurements is sufficiently high. The major differences of the two approaches are outlined as follows. First, the small DTs in the proposed approach are trained by using attribute subsets for robustness, whereas the entire set of attributes is used in [6]. Second, the usage of new cases in near real-time is different. In [6], the new cases are used to update the small DTs, whereas in the proposed approach, the new cases are only used to re-check the performance of viable small DTs so as to quantify the voting weights.

V. Case Study

A. Test System

The IEEE 39-bus system [23] is used as the test system which contains 39 buses, 10 generators, 34 transmission lines and 12 transformers. Particularly, G1 represents the aggregated generation from the rest of eastern interconnection [23]. In this case study, the test system is assumed to consist of three areas. The three areas together with the PMU placement are illustrated in Fig. 4. It is worth noting that the PMU placement guarantee the full observability of the test system when zero-injection buses are taken into account.

B. Knowledge base

The knowledge base only consists of the OCs that are both pre-contingency secure and $N-1$ secure. The cases in the knowledge base are created from the combinations of the

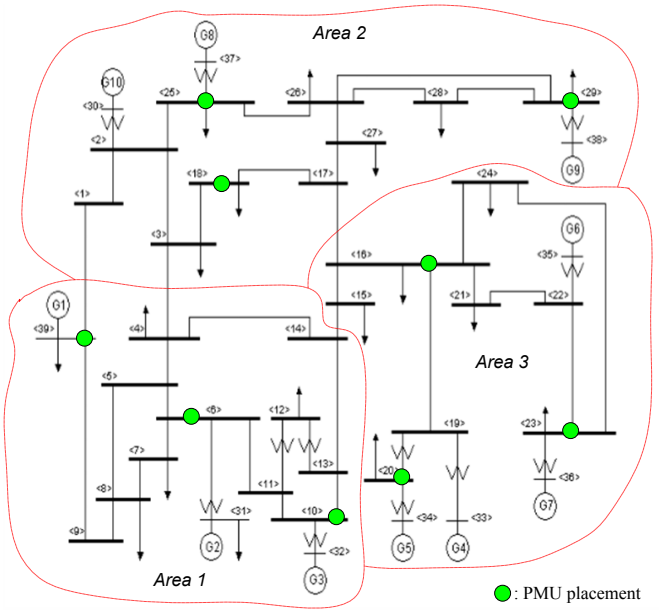

Fig. 4. The IEEE 39-bus system in three areas and PMU placement

"PMU measurements"[1] of the OCs and their transient security classification decisions for a few selected $N-2$ contingencies.

1) OC Generation: The OC given in [23] is used as the base OC. Following the method in [5], more OCs are generated for off-line training, by randomly changing the bus loads (both active and reactive) within 90% to 110% of their original values in the base OC; for the OCs generated for near real-time re-check and on-line DSA test, the bus loads varies from 97% to 103% of their original values in the base OC. The rationale for the above percentage values is that off-line training is usually carried out day/hours ahead, and thus the predicted OCs can have a larger prediction error than those in near real-time. The power flows of each generated OC are solved using the power flow and short circuit analysis tool (PSAT) [13], followed by a limit check such that the generated OCs with any pre-contingency overloading or voltage/angle limit violations are excluded from the knowledge base.

2) Contingencies: The loss of any of the 46 components (i.e., the 34 transmission lines, the 9 generator-transformer pairs and the 3 transformers at (11,12), (12,13) and (19,20)) is considered as an $N-1$ contingency. Due to the large number of possible $N-2$ contingencies, only a few of them are selected. Intuitively, a severe impact on the security of power systems is more likely if a second component gets overloaded after the loss of the first component. As such, the $N-2$ contingencies are selected in the following manner. First, each of the aforementioned 46 components is removed from the test system. Then, power flows are re-solved and limit check is re-run for the base OC using PSAT. The first removed component together with any overloaded component are regarded as the removed pair of an $N-2$ contingency. As a result, 15 pairs are identified, as listed in Table II.

[1]The power flow solutions of an OC are used as the "PMU measurements."

TABLE II
THE FIRST AND SECOND REMOVED COMPONENTS OF THE SELECTED $N-2$ CONTINGENCIES

line(4,14), line(6,11)	line(6,11), line(4,14)	line(6,11), line(13,14)
line(6,11), line(10,13)	line(10,11), line(10,13)	line(10,13), line(6,11)
line(10,13), line(10,11)	line(13,14), line(6,11)	line(13,14), line(10,11)
line(16,21), line(23,24)	line(21,22), line(23,24)	line(21,22), line(22,23)
line(21,22), line(16,24)	line(23,24), line(16,21)	line(23,24), line(21,22)

3) Transient Security Assessment: Transient security assessment tool (TSAT) [13] is used to assess the transient performance of the OCs that are pre-contingency secure. To create a contingency in TSAT, the "three-phase short circuit to ground" fault is applied at either of the two terminal buses of the first removed component with a primary clearing time of 4 cycles. Therefore, 92 $N-1$ contingencies and 30 $N-2$ contingencies are created. The "power angle-based stability margin" defined in TSAT [13] is used as the stability index.

C. Performance Evaluation

Three other approaches are used as benchmarks, including a single DT using surrogate, an RF using surrogate, and an RF without using surrogate. Following [20], unpruned DTs are used in RFs; in RFs, all training cases are used to build a single DT; in each split of DTs, a number of $\log_2 P+1$ attributes are randomly selected (where P=96 according to Column 3 of Table III); the optimal number of DTs in the forest is determined through out-of-bag validation [20]. Specifically, for the former two benchmark approaches, surrogate attributes are obtained from those which are not co-located with the primary attributes; for the third benchmark approach, degenerated DTs are used.

1) Attribute Subsets: The hypothetical WAMS for the test system has a hierarchical architecture similar to that in Fig. 2. Based on the evaluation results of the reference [24], it is assumed that all the PMUs have the same availability a ($a\in$ [0.979975,0.998920]), and all the communication links from PMUs to PDC have the same availability A^{link}=0.999. Further, the availability of the PDC and the communication link from the PDC to the monitoring center is assumed to be 1. Let $b=(0.999a)^3$, and thus $b\in$ [0.938299,0.993776]. Then, it follows that when $A_0\leq b$, the solution to $\tilde{P}_{\mathbf{s}}$ in (6) exists, as given in Table III. In what follows, the data in Table III is explained in detail. Specifically, column 2 provides the indices of PMU buses, which can also be seen from Fig.4. Column 3 contains the number of attributes for the three categories defined in Section III.A. Take area 1 for example, there are 3 voltage magnitude attributes, 24 transmission line (including power flow and current magnitude) attributes, and another 3 attributes from voltage phase angle difference. Given the system topology and availability information, M_k in column 4 and A_k in column 5 are calculated using (1) and (4), respectively. Then, p_k is obtained by solving (6).

2) Off-line Training: N_{OC}=200 generated OCs which are both pre-contingency and $N-1$ contingency secure are used for off-line training. Combining the generated OCs with their transient security classification decisions for the N_C=30 selected $N-2$ contingencies, are used to generate the N=6000

TABLE III
DATA USED BY ALGORITHM 1 FOR THE SYSTEM IN FIG. 4.

Area	Placement	Number of attributes			M_k	A_k	p_k
		Cat. 1	Cat. 2	Cat. 3			
1	8, 13, 39	3	24	3	700	b	0.28
2	18, 25, 29	3	24	3	700	b	0.28
3	16, 20, 23	3	30	3	1120	b	0.44

cases in the knowledge base. The size and the number of small DTs are determined by bias-variance analysis [25] and v-fold cross validation [4]. In this case study, L=40 and J=3 are used by the proposed approach; 45 DTs are used in the two RF-based approaches.

3) Near Real-time Re-check: By following the procedure described in Section V.B, 100 OCs are generated for performance re-check. The DTs trained off-line are applied to the new cases; the classification results are compared with the actual security classification decisions of the new cases. Then, these re-check results are used by Algorithm 3 to quantify the voting weights of DTs.

4) On-line DSA Test: Another 100 OCs are generated for testing, by following the procedure described in Section V.B. Recall that the availability of the PDCs and the communication links connecting PDCs is 1, and then it can be seen from Fig. 3 and Fig. 4 that the total number of failure scenarios can be reduced to 512 (2^9, since there are 9 pairs of PMUs and links). On-line DSA test is repeated for all failure scenarios, by identifying the missing PMU measurements and viable small DTs, calculating the voting weights of viable small DTs, and evaluating the misclassification error rate. The overall misclassification error of on-line DSA is calculated by:

$$\bar{\mathrm{e}}(\tilde{F}) = \sum_{k=1}^{512} \mathrm{Prob}(\Omega(k))\mathrm{e}(\tilde{F}|\Omega(k)), \tag{17}$$

where, $\Omega(k)$ denotes the k-th failure scenario, $\mathrm{Prob}(\Omega(k))$ denotes the probability for $\Omega(k)$ to happen, which can be easily calculated by using the assumed availability, and $\mathrm{e}(\tilde{F}|\Omega(k))$ denotes the misclassification error rate of $\tilde{F}$ in the failure scenario $\Omega(k)$ ($\mathrm{e}(\tilde{F}|\Omega)$ is set to be 1 when all PMUs fail). The test is performed for various values of b, and the test results are illustrated in Fig. 5. It is observed that the benchmark approaches are comparable to the proposed approach only around b=1. However, the gaps become more significant as b decreases.

5) Impact of Measurement Noise: In reality, PMU data can contain measurement noise. Following the approach in [26], numerical experiment is carried out to study the impact of measurement noise on the performance of the proposed approach.

For convenience, let $V\angle\theta_V$ and $I\angle\theta_I$ denote a voltage phasor and a current phasor, respectively; let $\tilde{V}\angle\tilde{\theta}_V$ and $\tilde{I}\angle\tilde{\theta}_I$ be the corresponding measurement. For PMUs complying with IEEE C37.118 standard [27], the PMU measurements should have a total vector error (TVE) less than 1%, i.e.,

$$|\frac{\tilde{V}\angle\tilde{\theta}_V - V\angle\theta_V}{V\angle\theta_V}| < 1\% \tag{18}$$

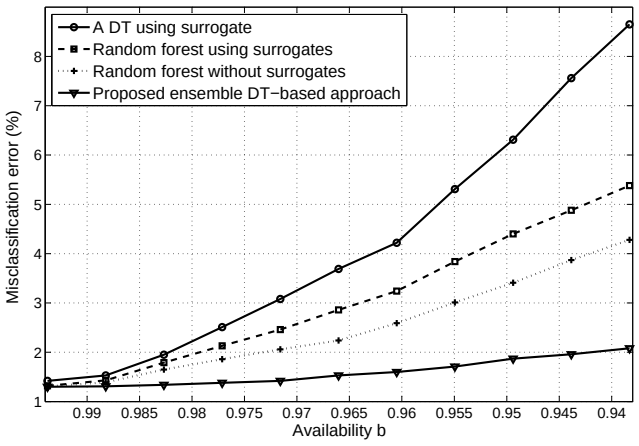

Fig. 5. Performance of on-line DSA in case of missing PMU measurements

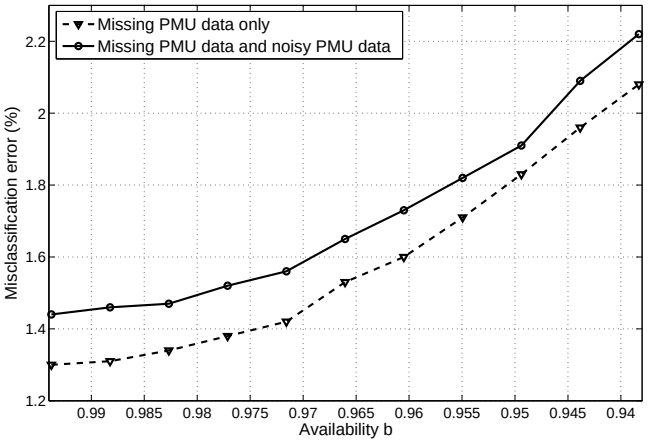

Fig. 6. Impact of measurement noise.

$$|\frac{\tilde{I}\angle\tilde{\theta}_I - I\angle\theta_I}{I\angle\theta_I}| < 1\% \tag{19}$$

Let n_V and n_I denote the measurement noise, respectively. In order to obtain PMU measurements that comply with the above specifications, n_V and n_I are randomly generated, by using the following density functions (note that other density functions can be also used) that are properly scaled and truncated from the standard complexity Gaussian distribution:

$$f(n_V) = \begin{cases} \frac{9}{\pi(1-e^{-9})10^{-4}V^2} e^{-\frac{9|n_V|^2}{10^{-4}V^2}} & \text{if } |n_V| \le 10^{-2}V \\ 0 & \text{o.w.} \end{cases} \tag{20}$$

$$f(n_I) = \begin{cases} \frac{9}{\pi(1-e^{-9})10^{-4}I^2} e^{-\frac{9|n_I|^2}{10^{-4}I^2}} & \text{if } |n_I| \le 10^{-2}I \\ 0 & \text{o.w.} \end{cases} \tag{21}$$

Then, it is clear that all noisy measurements have TVE not more than 1%, and are complex Gaussian distributed within their support. The generated random measurement noise is added to the both training and testing data. The test results are provided in Fig. 6.

VI. Conclusion

A data-mining approach has been proposed to mitigate the impact of missing PMU measurements in on-line DSA. In particular, the various possibilities of missing PMU measurements in on-line DSA can make off-the-shelf DT-based techniques (a single DT, RF, etc.) fail to deliver the same performance as expected. The proposed ensemble DT-based approach exploits the locational information and the availability information of PMU measurements in randomly selecting attribute subsets, and utilizes the re-check results to re-weight the DTs in the ensemble. These special treatments developed from a better characterization of power system dynamics guarantee that the proposed approach can achieve better performance than directly applying off-the-shelf DT-based techniques.

References

[1] L. Wehenkel, T. Van Cutsem, and M. Ribbens-Pavella, "An artificial intelligence framework for online transient stability assessment of power systems," *IEEE Trans. on Power Syst.*, vol. 4, no. 2, pp. 789–800, May 1989.

[2] S. Rovnyak, S. Kretsinger, J. Thorp, and D. Brown, "Decision trees for real-time transient stability prediction," *IEEE Trans. on Power Syst.*, vol. 9, no. 3, pp. 1417–1426, Aug. 1994.

[3] K. Sun, S. Likhate, V. Vittal, V. Kolluri, and S. Mandal, "An online dynamic security assessment scheme using phasor measurements and decision trees," *IEEE Trans. on Power Syst.*, vol. 22, no. 4, pp. 1935–1943, Nov. 2007.

[4] R. Diao, K. Sun, V. Vittal, R. O'Keefe, M. Richardson, N. Bhatt, D. Stradford, and S. Sarawgi, "Decision tree-based online voltage security assessment using PMU measurements," *IEEE Trans. on Power Syst.*, vol. 24, no. 2, pp. 832–839, May 2009.

[5] R. Diao, V. Vittal, and N. Logic, "Design of a real-time security assessment tool for situational awareness enhancement in modern power systems," *IEEE Trans. on Power Syst.*, vol. 25, no. 2, pp. 957–965, May 2010.

[6] M. He, J. Zhang, and V. Vittal, "A data mining framework for online dynamic security assessment: decision trees, boosting, and complexity analysis," in *Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES*, Washington D.C., USA, Jan. 2012, pp. 1–8.

[7] Alberta Electric System Operator Rules, "Section 502.9: synchrophasor measurement unit technical requirements," Aug. 2012, available online http://www.aeso.ca/downloads/2012-08-30_Section_502-9_phasor.pdf.

[8] Schweitzer Engineering Laboratories Technical Report, "Improving the availability of synchrophasor data," Aug. 2011, available online https://www.selinc.com/TheSynchrophasorReport.aspx?id=98004.

[9] R. Emami and A. Abur, "Robust measurement design by placing synchronized phasor measurements on network branches," *IEEE Trans. on Power Syst.*, vol. 25, no. 1, pp. 38–43, Feb. 2010.

[10] A. Gomez-Exposito, A. Abur, P. Rousseaux, A. de la Villa Jaen, and C. Gomez-Quiles, "On the use of PMUs in power system state estimation," in *Proc. 17th Power Systems Computation Conference*, Stockholm, Sweden, Aug. 2011.

[11] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Chapman and Hall/CRC, 1984.

[12] Y. Wang, W. Li, and J. Lu, "Reliability analysis of wide-area measurement system," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1483–1491, July 2010.

[13] Powertech Labs, "DSATools: Dynamic Security Assessment Softwares," available online http://www.dsatools.com.

[14] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[15] R. Banfield, L. Hall, K. Bowyer, and W. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 173–180, Jan. 2007.

[16] R. Bryll, R. Gutierrez-osuna, and F. K. Quek, "Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets," *Pattern Recognition*, vol. 36, no. 6, pp. 1291–1302, June 2003.

[17] T. K. Ho, "Random decision forests," in *Proc. Third Intl Conf. Document Analysis and Recognition*, Montreal, Canada, Aug. 1995, pp. 278–282.

[18] A. Phadke and J. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York: Springer, 2008.

[19] T. K. Ho, "Random decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[20] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[21] V. Khiabani, O. P. Yadav, and R. Kavesseri, "Reliability-based placement of phasor measurement units in power systems," *J. Risk and Reliability*, vol. 226, no. 1, pp. 109–117, Feb. 2012.

[22] L. Hyafil and R. L. Rivest, "Constructing optimal binary decision trees is NP-complete," *Information Processing Letters*, vol. 5, no. 1, pp. 15–17, May 1976.

[23] M. Pai, *Energy Function Analysis for Power System Stability*. Kluwer Academic Publishers, Boston, 1989.

[24] F. Aminifar, S. Bagheri-Shouraki, M. Fotuhi-Firuzabad, and M. Shahidehpour, "Reliability modeling of PMUs using fuzzy sets," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2384–2391, Oct. 2010.

[25] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2008.

[26] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," 2012, to appear, *IEEE Trans. on Power Syst.*

[27] "IEEE standard for synchrophasors for power systems," 2005, IEEE Std. C37. 118-2005.

Miao He (S'08) received the B.S. degree from Nanjing University of Posts and Telecommunications, in 2005, and the M.S. degree from Tsinghua University, in 2008. He is currently pursuing the Ph.D. degree at Arizona State University.

VijayVittal (S'78-F'97) received the B.E. degree in electrical engineering from the B.M.S. College of Engineering, Bangalore, India, in 1977, the M.Tech. degree from the India Institute of Technology, Kanpur, India, in 1979, and the Ph.D. degree from Iowa State University, Ames, in 1982.

He is the Ira A. Fulton Chair Professor in the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe. Currently, he is the director of the Power System Engineering Research Center (PSERC).

Dr. Vittal is a member of the National Academy of Engineering.

Junshan Zhang (F'12) received his Ph.D. degree from the School of ECE at Purdue University in 2000. He joined the EE Department at Arizona State University in August 2000, where he has been Professor since 2010. His research interests include communications networks, cyber-physical systems with applications to smart grid, stochastic modeling and analysis, and wireless communications.

Dr. Zhang is a recipient of the ONR Young Investigator Award in 2005 and the NSF CAREER award in 2003. He received the Outstanding Research Award from the IEEE Phoenix Section in 2003.