Decision Tree-Based Online Voltage Security Assessment Using PMU Measurements

Ruisheng Diao, *Student Member, IEEE*, Kai Sun, *Member, IEEE*, Vijay Vittal, *Fellow, IEEE*, Robert J. O'Keefe, *Member, IEEE*, Michael R. Richardson, Navin Bhatt, *Fellow, IEEE*, Dwayne Stradford, and Sanjoy K. Sarawgi, *Member, IEEE*

***Abstract*—Voltage collapse is a critical problem that impacts power system operational security. Timely and accurate assessment of voltage security is necessary to detect post-contingency voltage problems in order to prevent a large scale blackout. This paper presents an online voltage security assessment scheme using synchronized phasor measurements and periodically updated decision trees (DTs). The DTs are first trained offline using detailed voltage security analysis conducted using the past representative and forecasted 24-h ahead operating conditions. The DTs are also updated every hour by including newly predicted system conditions for robustness improvement. The associated synchronized critical attributes are obtained in real time from phasor measurement units (PMUs) and compared with the offline thresholds determined by the DTs to assess security. This approach is tested on the American Electric Power (AEP) system and properly trained DTs perform well in assessing voltage security. Several new ideas to improve DT performance are also introduced.**

***Index Terms*—Decision trees, online security assessment, phasor measurement units, voltage collapse.**

I. Nomenclature

A_x_y	Voltage phase angle of node x minus that of y.
CSR	Critical splitting rule in the decision tree.
DT	Decision tree.
FB	Faulted bus.
I_x_y	Current flow on branch between nodes x and y.
LS	Learning set.
OB	Other bus.
OC	Operating condition.
PMU	Phasor measurement unit.
Q_x_y	MVAr flow on branch between nodes x and y.
TS	Test set.
$\mathbf{V^2}$	Square of the measured voltage magnitude.
IZ_x_y	Current magnitude multiplied by branch impedance between nodes x and y.

Manuscript received September 02, 2008; revised November 14, 2008. First published April 10, 2009; current version published April 22, 2009. This work was supported by American Electric Power and the National Science Foundation under Grant NSF EEC-9908690 and the Power System Engineering Research Center. Paper no. TPWRS-00692-2008.

R. Diao and V. Vittal are with the Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: ruisheng.diao@asu.edu; vijay.vittal@asu.edu).

K. Sun is with EPRI, Palo Alto, CA 94304 USA (e-mail: ksun@epri.com).

R. J. O'Keefe, M. R. Richardson, N. Bhatt, D. Stradford, and S. K. Sarawgi are with American Electric Power Service, Columbus, OH 43215 USA (e-mail: rjo'keefe@aep.com; mrrichardson@aep.com; nbbhatt@aep.com; dstradford@aep.com; sksarawgi@aep.com).

Digital Object Identifier 10.1109/TPWRS.2009.2016528

II. Introduction

MODERN power systems are interconnected by transmission lines for both reliability and economic reasons. Countries with highly developed economies have seen the consumption of electric power rise at consistent annual rates, but more recently the load growth has not resulted in a concomitant increase in the transmission capability due to the difficulty of siting and approving new transmission lines [1]. Hence, power networks are likely to be operated under greater stress with transmission lines carrying electric power near their limits during peak load periods. In such situations, the loss of critical transmission lines or transformers may cause a continuous decline of bus voltages and result in a voltage collapse if the system lacks effective reactive power support [2].

Voltage collapse is one of the most critical problems that threaten system secure operations. It is usually initiated by: 1) a continuous load increase, and/or 2) a major change in network topology resulting from a critical contingency. During a voltage collapse event, bus voltages in a localized area decrease below an acceptable level. Without timely control actions, the low voltages may spread throughout adjacent areas of the power system and may eventually cause a large scale blackout instead of a localized outage. The time span for voltage instability ranges from 0.1 s to 1 h, representing transient and long-term voltage instability, respectively [2], [3]. For the fast transients, detailed voltage security analysis for online applications is a great challenge because of the high computational burden. As a result, an accurate assessment tool to determine the voltage security in real time becomes a necessity for system operators to determine remedial controls for preventing voltage collapse.

This paper focuses on the voltage collapse problems caused by severe disturbances in the system and presents a three-step decision tree-based scheme for online voltage security assessment using phasor measurements. The main procedures include: 1) conducting detailed post-contingency voltage security analysis for the past and predicted operating conditions (OCs) and training DTs offline from created databases containing these analysis results; 2) periodically updating the DTs by including anticipated OCs for the next hour to reduce the likelihood of misclassifications on whether the system is secure or insecure; and 3) continuously collecting the corresponding PMU measurements selected as critical attributes in the DTs.

By comparing these measurements with the thresholds in the DTs, an online voltage security assessment following severe disturbances in the system can be obtained. The proposed scheme takes full advantage of all the available PMU measurements across a system to assess post-contingency voltage security. The scheme is tested on an AEP operational system that consists of 2414 buses, 116 generators, and 2416 transmission lines. The result shows that with correctly selected parameters from current PMU locations, properly trained DTs perform well on the OCs during a specific load period. Two new ideas to improve DT performance and reliability including the applications of "Multiple DTs" and "Corrective DTs" are also introduced and analyzed.

This paper is organized as follows: Section III introduces the PMU and decision tree principles and explains the proposed scheme for online voltage security assessment in detail. Section IV tests the proposed approach on the AEP system and provides all the details regarding OC generation, parameter selection and DT training with practical considerations. The DT performance using different types of predictors is also compared. Section V discusses several methods to improve the DT performance for online application. Finally, conclusions are drawn in Section VI.

III. Proposed Method

A. PMU and Decision Tree Principles

State monitoring in power systems plays an important role in system operation and online decision making. Traditionally system states were monitored using current transformers and potential transformers and communicated to the energy management system (EMS) through the supervisory control and data acquisition (SCADA) system. This approach served the industry well but lacks the ability of observing measurements across the whole system because the data was not time synchronized. The advent of synchronized phasor measurements has revolutionized the field of power system state monitoring. PMUs have significant advantages over the traditional measurements in terms of both accuracy and speed of measurement by utilizing the global positioning system (GPS) receivers and microprocessors. The time-stamped digital phasors calculated in the PMUs are synchronized to a common time frame by satellites and then assembled into a series of data streams for communication to remote control centers [4]. The high data accuracy and data sampling frequency endow PMUs with the ability to observe the system state across the whole system in real time. Applications using phasor measurements in power systems have been developed for loss of synchronism detection, multi-area state estimation, oscillation mode identification, voltage stability protection and system dynamics monitoring [5]–[8].

The decision tree technique is an effective supervised data mining tool to solve the classification problems in high data dimensions. For a created database consisting of different cases that are represented by a vector of **predictors** (or variables) along with an **objective**, a DT is designed for successful classifications of this objective by using only a small number of these predictors. The decision tree structure is usually binary and there are two types of nodes in such a DT, the "internal node" with two successors and the "terminal node" without successors. For each internal node, a question or critical splitting rule (CSR) is asked to decide which successor the classification process should drop into. The splitting rule could be numerical by comparing the variable value with a threshold, or categorical by checking whether the current value belongs to a specific data set. For each terminal node, a classification result is assigned in terms of the majority class of the objective, e.g., "secure" or "insecure." The classification process is very simple and fast, which is to drop the associated predictors down the tree model by comparing the CSRs in different levels. One DT example that has four internal nodes and five terminal nodes is shown in Fig. 1. It is similar to the DTs trained in later sections, in which FB$ is a categorical predictor while the remaining A_195_4767, A_3901_5648, and A_35_2221 are all numerical ones.

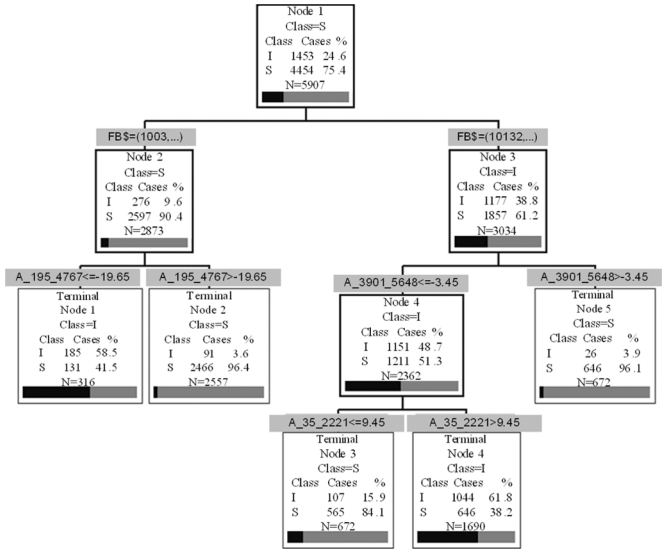

Fig. 1. DT sample.

Training a DT not only uncovers important system parameters that contribute to the final objective for the known cases, but also optimizes the prediction ability on the unknown cases. Therefore, a learning set (LS) and a test set (TS) with the same data format are required before a DT is trained. In the beginning, a maximum DT is first trained from the LS by recursively splitting a parent node into two purer child nodes in terms of the impurity reduction. The splitting process starts from the root of the tree and continues until further splitting of a node can not improve the overall DT performance or when a predefined threshold is reached. Details of identifying CSRs and split stopping rules are discussed in [9]. Each CSR in the DT has several "competitors," which could provide good alternate candidates if the CSR is missing. By definition, this maximal tree possesses the highest accuracy for the LS involved. It is then pruned using the TS to generate a series of smaller DTs in terms of the misclassification cost on the TS.

The **optimal** DT is then defined as the tree with the lowest misclassification cost. It always has a medium size, because a small tree does not utilize enough useful information and a large DT usually has the over-fitting problem caused by the data noise. Once properly trained, an optimal DT is quite suitable for identifying critical system attributes from various system states that are related to power system security problems. Several applications involving decision trees have been addressed in real-time transient stability prediction and assessment, voltage

security monitoring and estimation, loss of synchronism detection and timing of controlled separation in power systems [11]–[19]. A recent approach has combined DT with another data mining tool for prediction performance improvement in the field of dynamic security controls [20].

B. Main Procedures and Advantages

The proposed scheme for online voltage security assessment consists of three major steps, "Offline DT Training," "Periodic DT Update," and "Online Application," which is similar to the approach developed for online dynamic security assessment in Section III of [12]. In this method, two important assumptions are made: 1) The voltage profiles of the system at the base case should be maintained at normal levels; 2) Voltage collapse cases are initiated only by critical contingencies. The voltage instability caused by continuous load increase is not addressed here.

1) Offline DT Training: A number of operating conditions for the past representative data and the forecasted ones for the next 24 h are first collected a day ahead. For each of these N operating conditions (Noc), detailed voltage security analysis for N critical contingencies (Nc) are conducted. Each contingency case at different OCs is then assigned a voltage security label, secure (S) or insecure (I). A number of cases consisting of Noc*Nc cases are created in a database. By collecting different types of PMU-related system parameters for use as voltage security predictors from the pre-disturbance system information, DTs are then trained offline to obtain the security classifications for applications in the next day. The selection of voltage security predictors will be discussed further in Section IV-C.

2) Periodic DT Update: During the operation time horizon, system information is periodically checked and updated on an hourly basis in order to account for changing system states as accurately as possible so that the offline trained DTs may continue to perform well on the new system states. If significant changes exist in network topology, generator status or load level, voltage security simulations are conducted on these changed OCs to build new cases for DT test. Good performance of DTs indicates that the DTs do not need to be re-trained or modified. Otherwise, the newly created cases are combined with the original ones to build new DTs with better accuracy. The updated DTs are then used for online applications during the next hour.

3) Online Application: Measurements from PMUs are continuously collected in real time and compared with the CSRs in the final DTs. If certain CSRs that are measured by the PMUs violate the determined thresholds then preventive or corrective action is armed. No action is taken until the associated contingency which results in insecurity occurs. If the contingency occurs then the armed preventive or control action is initiated. This paper only addresses the topic of security assessment. The topic of corrective and preventive control is beyond the scope of the work done in this paper.

In the proposed scheme, the vulnerability of the current OC can be effectively characterized by only a few critical pre-disturbance system parameters instead of using thousands of system states. The simple structure and readability of the DT model make it very convenient to input the PMU measurements directly and compare with the thresholds on the CSRs to obtain a security assessment. This process is very fast since only a few comparisons are required. In addition, the pre-disturbance critical system attributes associated with the CSRs provide a "nomogram" in the space of these attributes in the system, which defines different operating regions regarding the probability of a contingency to cause voltage insecurity. The nomogram requires that these critical attributes be **measured simultaneously** to determine whether the current OC falls inside the nomogram or outside it and this requirement can be satisfied by the synchronized measurements obtained from the PMUs. By periodically updating the nomogram, operators can have a more accurate tool for online decisions. As a consequence, DTs with satisfactory performance are quite feasible for real time voltage security assessment in power systems.

The proposed method takes advantage of the PMU measurements across the whole system to obtain a voltage security assessment, instead of using only local measurements. Should a particular CSR measurement become missing or unavailable, the competitors for this CSR can be used in making the assessment, which can effectively reduce dependence on the local measurements in assessing voltage problems. In addition, the periodically updated DTs only pick up a small number of the system parameters as CSRs, which results in enhanced efficiency compared to traditional analysis methods to obtain a security result. The major computational burden in the process is due to the database generation including OC collection and voltage security simulation. However, it can be effectively reduced by parallel calculations since all the simulation cases are independent from each other. Furthermore, the DT training process from scratch only takes a few minutes on a PC with a Intel Core2 CPU 6700 (2.66 GHz) and 2.0 GB of RAM. Therefore, there is enough time to update these databases and DTs every hour. This 1-h time interval could be shortened if necessary.

IV. Case Study

The proposed method is tested on an AEP operations model consisting of more than 2400 buses, 100 generators, and 2400 transmission lines with the voltage level ranging from 4 to 765 kV. This system is also interconnected with the surrounding areas representing the eastern interconnected power system in North America. Therefore, a total number of 18 168 buses, 2753, generators, and 19 358 lines are included for voltage security studies. There are 27 either existing or proposed PMUs located across the eastern AEP system covering a wide area, which are shown in Fig. 2. One PMU is installed on a 138-kV bus, 11 PMUs are installed on 345-kV buses, and the remaining PMUs are all located on 765-kV buses. These PMUs are installed to monitor the states of the key buses and stations including bus voltage magnitudes, voltage phase angles, MW and MVAr flows, and current magnitudes of the associated branches. In this case study, a software platform involving a variety of simulation tools has been developed to test this scheme. Operating conditions are all generated using the Powerflow & Short-circuit Analysis Tool (PSAT) and voltage security studies are performed in the Voltage Security Assessment Tool (VSAT), both of which are components of the Dynamic Security Analysis Tool ($\mathrm{DSA}^{\mathrm{Tools}}$). $\mathrm{DSA}^{\mathrm{Tools}}$ is an advanced package for power system security evaluation and is developed by the Powertech Labs, Canada [21]. The decision trees are trained and tested using a commercial data mining package, Classification and Regression Trees (CART), which is developed by Salford Systems, CA [22]. The database generation, data conversion and analysis

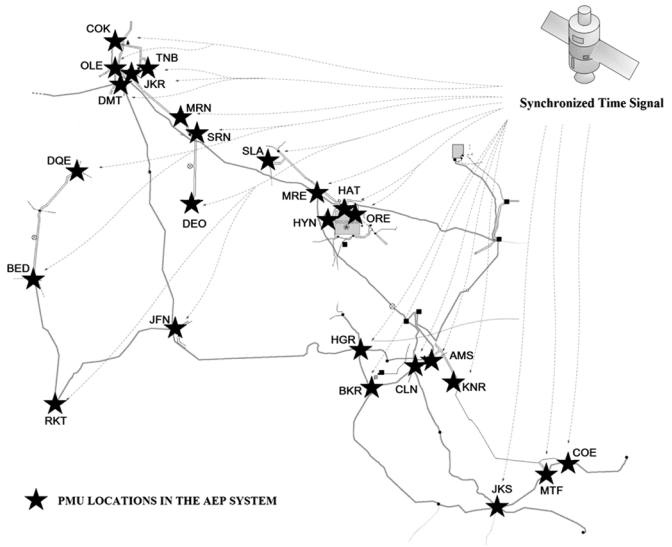

Fig. 2. PMU locations in the eastern AEP system.

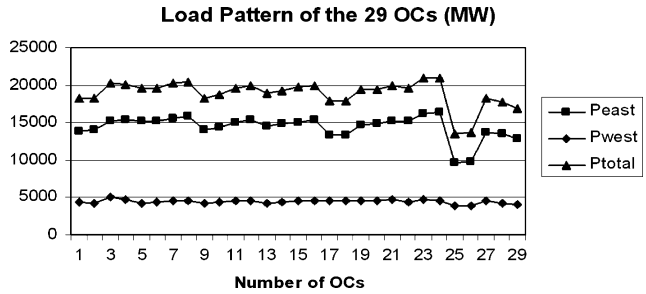

Fig. 3. Load pattern of the 29 stressed OCs in the AEP system.

work are conducted in MATLAB and Microsoft Visual Studio VC++ codes. Simulated PMU measurements across the system are collected from PSAT results.

A. OC Generation

In this step, 29 operating conditions (OCs) are generated in PSAT based on the generation and load patterns provided by the AEP operations staff. They represent stressed OCs that include all the details of load levels, generator outputs and branch power flows during a specific period of time. Some of these OCs stress the system well beyond the normal operating ranges. The load levels in these OCs are shown in Fig. 3. The voltages of all the buses in the AEP system are within reasonable levels at base case: 1) all the bus voltage magnitudes are adjusted to be between 0.90 and 1.10 in p.u. and 2) the voltage magnitudes of 138–765 kV buses are adjusted in the range of 0.95–1.10 p.u.

B. Voltage Security Evaluation

A list of critical contingencies that may cause severe voltage problems is selected according to previous operating experiences in the AEP system. This list contains 163 critical contingencies including different transmission line and transformer outages. This paper adopts a static analysis method to evaluate post-contingency voltage security and a voltage collapse judgment is given when a contingency results in the divergence of the power flow solution. Thus, the database consists of $29 * 163 = 4727$ voltage security simulations with either secure or insecure labels marked on each case. From the simulation results, 34.46% of the total cases are insecure and the remaining cases are labeled as "Secure."

TABLE I
DIFFERENT PMU-RELATED PREDICTOR GROUPS IN THE AEP SYSTEM

No.	Predictors
Group 1	Faulted bus (FB) and Other bus (OB) of the contingency branch
Group 2	Voltage phase angle differences (A_x_y)
Group 3	A_x_y, FB, OB
Group 4	Current magnitudes on branches (I_x_y), FB, OB
Group 5	MVAr flows on branches (Q_x_y), FB, OB
Group 6	Square of voltage magnitudes (V^2_x), FB and OB
Group 7	Absolute value of current magnitude multiplied by branch impedance (IZ_x_y), FB and OB
Group 8	A_x_y, Q_x_y, I_x_y, V^2_x, IZ_x_y, FB, OB

C. Predictor Selection and Database Generation

Since all the voltage insecure cases are caused by contingencies in the list, the voltage security predictors for DT training should include the contingency-dependent information together with pre-contingency system parameters. For the contingency-dependent parameters, an unordered bus pair x and y is adopted to denote the faulted transmission line or transformer, because the voltage security analysis method estimates voltage security by solving power flows without the faulted branch and it does not need any detail of the fault such as type, duration or location. The only useful information is the bus locations of the faulted branch. To eliminate the ordering of the two buses of a critical branch during DT training, each contingency case is doubled in the form of "Bus–1 = x and Bus–2 = y" and "Bus–1 = y and Bus–2 = x," which treats the two buses equally. As a result, a database containing $4727 * 2 = 9454$ cases is generated. This measure allows the DTs to identify buses or substations common to contingencies that are more prone to cause voltage collapse. The pre-contingency parameters reflect critical system states in real time; therefore they are chosen from the PMU monitored pre-disturbance system parameters in the AEP system. In order to compare the performance of different parameters in capturing post-contingency voltage behaviors, eight groups of predictors are used as defined in Table I.

Group 1 uses a categorical bus pair (FB and OB) to represent the contingency branch. Group 2 selects voltage phase angle differences among all of the 27 existing PMUs (A_x_y), where x and y are the PMU bus numbers. There are $27 * 26/2 = 351$ angle differences chosen as voltage security predictors. Although the information contained in the 351 angle differences are redundant since only 27 values are collected, the DT only picks up the most effective ones as CSRs. The predictors in Groups 1 and 2 help to compare the DT performance between contingency-dependent and pre-contingency parameters. Group 3 evaluates the DT performance using the combination of these two types of predictors. Groups 4 and 5 replace the phase angle differences with all the available current magnitudes and reactive power magnitudes on branches measurable by the PMUs. Groups 6 and 7 test the performance of square of voltage magnitudes on PMU buses and the absolute value of the current magnitude multiplied by the branch impedance. The selection of these

TABLE II
PERFORMANCE COMPARISON OF THE OPTIMAL DTS

Opt. DTs	Size	LS Accuracy (%)			TS Accuracy (%)		
		I	S	Overall	I	S	Overall
DT1	24	71.4	63.31	66.08	68.82	60.95	63.72
DT2	7	87.61	76.99	80.63	86.96	80.07	82.50
DT3	36	98.69	94.33	95.82	91.00	93.63	92.70
DT4	50	99.50	95.66	96.97	92.95	94.93	94.24
DT5	31	98.11	93.89	95.33	87.11	93.3	91.12
DT6	38	96.49	92.42	93.81	88.91	91.91	90.85
DT7	55	99.69	95.19	96.73	91.00	93.46	92.6
DT8	51	99.11	96.00	97.06	90.55	94.44	93.07

pre-contingency predictors is designed to collect critical system information that can indicate voltage problems more precisely and intuitively. A_x_y and I_x_y are good measures to indicate the degree of stress at an OC; Q_x_y plays a more important role than active power flow (P_x_y) in supporting voltage profiles; V^2_x is more sensitive than voltage magnitude itself in reflecting voltage security since it is coupled with reactive power more directly; and IZ_x_y covers more information than current magnitude in reflecting voltage problems.

Finally, Group 8 includes all the predictors indicated above in order to compare their performance in the DTs. With all the case labels and predictors collected, eight databases are correspondingly generated for the eight predictor groups.

D. DT Training and Performance

All the 9454 cases in the created database are treated equally and 1891 cases (20%) are randomly selected to form a test set. The remaining 7563 ones (80%) are used to form the learning set. In order that fewer insecure cases are misclassified as secure, the cost of misclassifying an insecure case to be a secure case is reasonably increased. Also, different algorithms including "Gini," "Symmetric Gini," "Entropy," "Class Probability," "Twoing," and "Ordered Twoing" are tested and compared during the DT training progress to obtain the optimal DT [10]. Therefore, eight optimal DTs from DT1 to DT8 are trained for the above databases respectively, the prediction accuracy for both learning set (LS) and test set (TS) of which are all shown in Table II.

From Table II, the comparison of DT1 and DT2 indicates that pre-contingency parameters are more important than contingency-dependent parameters in voltage security prediction at stressed OCs, because DT2 performs much better than DT1 for both secure and insecure cases. One explanation is that the system OC is more critical than fault location in voltage collapse prediction. Voltage phase angle differences can effectively indicate the degree of stress at an OC and large values of A_x_y usually indicate a more stressed OC. A severely stressed OC is more vulnerable to voltage collapse than a lightly loaded one following the same contingency. Similar tests were conducted by using the other pre-contingency parameters alone as predictors such as I_x_y, Q_x_y, V^2_x, and IZ_x_y, but improvement in prediction accuracy was not observed. Therefore, the remaining predictor groups in Table I (Groups 3 to 8) also include the contingency-dependent information for DT performance improvement and the results in DT3 to DT8 have proved this. The combination of FB, OB and various pre-contingency parameters are helpful for building much better DTs compared to DT2.

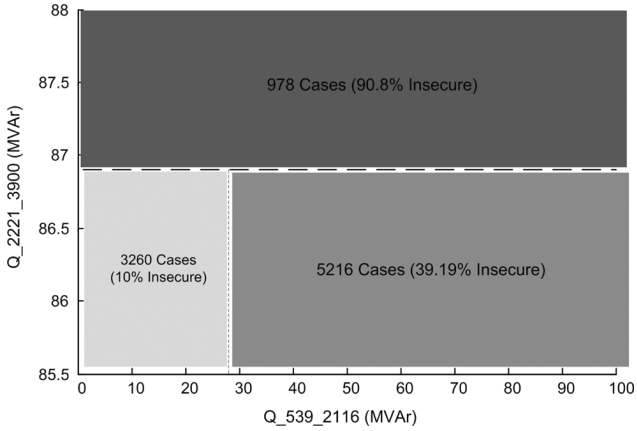

Fig. 4. Nomogram in terms of the top two predictors in DT5.

In DT3, the prediction accuracy for the test set is increased to 91.0% for insecure cases and 93.63% for secure cases. The overall accuracy is 92.7%. A better decision tree (DT4) with higher accuracy for the test set is obtained by using current magnitudes as predictors, which can accurately predict 92.95% insecure cases and 94.93% secure cases. Similarly, DT5, DT6 and DT7 are, respectively, trained by including values of Q_x_y, V^2_x, and IZ_x_y from the same PMU locations, and they perform similarly to DT3. DT8 attempts to cover all the above predictors for accuracy improvement in the test set because it includes the most system information. Unfortunately, the overall accuracy for the test set in DT8 does not see any significant improvement. Instead, the overall accuracy drops to 93.07%, which is even lower than DT4, although it has a better performance for the LS. This is mainly caused by the over-fitting problem in the DT training process, which uses too many parameters as predictors to build one DT. Another problem occurs because some of the important parameters are masked behind the CSRs in the DT since each splitting of the internal nodes guarantees the highest impurity reduction for the current node, instead of optimizing the prediction performance of the whole tree.

To illustrate the efficiency of the DTs trained, Table III lists the total number of CSRs selected in DT3 to DT8. Ns is the number of CSRs in the DT, while Nt represents the total number of predictors in the corresponding databases. It can be observed that only a small portion of the predictors are selected to be CSRs, indicating that the voltage security assessment process in real time is very fast. An example of the nomogram obtained based on the CSRs chosen by the DT5 and the thresholds calculated for the CSRs are depicted in Fig. 4. The operator can observe the PMU measurements of the CSRs with regards to the thresholds shown in the nomogram and arm for the appropriate control action as soon as a threshold is approached. The optimal DTs trained in Table II are not depicted in this paper due to space limitation.

V. DT PERFORMANCE IMPROVEMENT

The results in Section IV-D indicate that the best performance on the test set using a single DT is in DT4, the overall accuracy of which is 94.24%. In other words, there are 5.76% cases misclassified using this tree model. The misclassification problem in DTs could be caused by a variety of factors, among which the

TABLE III
NUMBER OF SYSTEM VARIABLES SELECTED IN THE OPTIMAL DTS

DTs	Ns/Nt	DTs	Ns/Nt
DT3	15/353	DT6	13/29
DT4	16/87	DT7	17/87
DT5	12/87	DT8	18/635

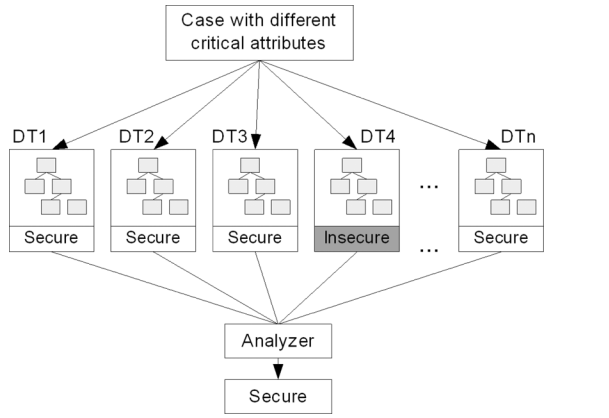

Fig. 5. Concept of multiple DT application.

formation of the LS and TS plays an important role. If the cases in the TS have great similarities with the ones in the LS, the DT has better performance on the TS. On the contrary, high accuracy of prediction on totally unknown cases may not be guaranteed. By periodically including new and unknown system states into the database, DTs are updated to learn more useful information for improving robustness and the classification accuracy can be effectively increased.

Another observation is that DT performance in predicting voltage security depends on the distribution of the DT input variable values and a preliminary test is conducted to illustrate this phenomenon. Focusing on one CSR that has a severe misclassification problem in the DT trained by using A_x_y, FB, and OB, it is found that most of the values of this parameter at the 29 OCs are concentrated in a small range. By creating more OCs to cover the values of this variable over a wider range but without changing the other parameters too much, a significant improvement in the DT trained for these new OCs is observed and this variable becomes a very important predictor in the new DT.

However, the proposed method for online voltage security assessment should work well for all kinds of unforeseen operating conditions no matter how the critical system parameters are distributed. Therefore, two methods to improve DT performance for online applications are discussed below when the future OCs for the next hour are relatively fixed.

A. Multiple Optimal DTs

As discussed above, the combination of all the available predictors may not improve the DT performance on the TS because of the over-fitting and variable masking problems. In order to improve reliability and prediction accuracy, an approach using multiple optimal DTs that are trained by combinations of different predictors is presented, the idea of which arises from the methods introduced in [23] and [24]. This method takes advantage of all the DTs that satisfy a desired threshold of performance instead of trusting the tree with the best performance only. The cases that are misclassified by the best DT may be correctly predicted by the other trees that use different PMU-related critical attributes. For online applications, all the optimal DTs are used to obtain a comprehensive classification result and the basic flow chart of this idea is shown in Fig. 5.

Based on the predictor groups introduced in Section IV-C, a heuristic search is conducted to identify different combinations of these predictors that contribute to good decision trees. The multiple DTs that are used for online application should be sufficiently different from each other because certain types of predictors can be totally masked when combined with other predictors to build a DT resulting in exactly the same tree as the one trained before predictor combination. As an example, when the predictor V^2_x was combined with any other type of predictors, they were never selected as CSRs. Thus, all different combinations of A_x_y, Q_x_y, I_x_y, and IZ_x_y are tested to create optimal DTs with sufficient difference. Nine DTs with good performance on the TS are obtained and shown in Table IV.

The decision trees from DT9 to DT17 are all different from each other although some of the branches share the CSRs. The result shows most of them offer better prediction performance on the TS than the DTs trained before predictor combination. As a result, all the 15 decision trees from DT3 to DT17 are used for online application since they are capable of correctly predicting over 91% of cases in the TS. For any contingency case, 15 voltage security assessments are obtained separately and a final assessment is given in terms of the majority classification results. A statistical analysis is conducted on the 1891 cases in the TS and 7563 cases in the LS using these 15 DT models. The results indicate that there are only 101 cases and 179 cases that are misclassified eight times or more in the TS and LS, respectively. Therefore the overall prediction accuracy on the TS is increased to 94.66%; while the accuracy on the LS is increased to 97.75%. Although the overall improvement is limited, the results obtained from a variety of DTs using different critical attributes are more convincing than that of only one tree. In addition, using multiple DTs in real time will not cause a significant increase in the total security assessment time because the DTs are all trained offline and the assessment process using different DTs can be conducted individually.

B. Corrective DTs

The above method is designed to build many decision trees by using different critical attributes for successful classifications on the misclassified cases in the TS. The main problem is that the absolute computation burden is increased although the DT training process can be conducted in parallel. According to the observation that most of the paths in a single DT trained by only one type of pre-contingency parameters (e.g., DT3) have excellent prediction behaviors on the TS and only a few paths have severe misclassification problems, another idea is developed to partially modify the DT by replacing the problematic paths with corrective DTs for accuracy improvement. For each of the paths with poor performance, a corrective DT is trained by including more system information for all the cases that fall into this path in the original database. During the voltage security assessment process, these problematic branches in the original DT are not abandoned; instead they are linked to these corrective DTs for further classification. For example, a path with poor performance in DT3 is taken to explain how this method works and this path is plotted in Fig. 6(a). The idea of using corrective

TABLE IV
DT PERFORMANCE FOR DIFFERENT PREDICTOR COMBINATIONS

Opt. DTs	Size	Predictor combinations	Test Set Accuracy (%)		
			I	S	Overall
DT9	59	A_x_y, I_x_y, FB, OB	90.4	95.59	93.76
DT10	49	A_x_y, Q_x_y, FB, OB	89.21	95.02	92.97
DT11	47	A_x_y, IZ_x_y, FB, OB	90.85	94.12	92.97
DT12	55	I_x_y, Q_x_y, FB, OB	92.95	95.26	94.45
DT13	45	Q_x_y, IZ_x_y, FB, OB	91.00	93.79	92.81
DT14	53	A_x_y, I_x_y, Q_x_y, FB, OB	92.95	95.18	94.39
DT15	47	A_x_y, I_x_y, IZ_x_y, FB, OB	90.85	94.12	92.97
DT16	53	I_x_y, Q_x_y, IZ_x_y, FB, OB	91.15	93.14	92.44
DT17	45	A_x_y, Q_x_y, IZ_x_y, FB, OB	91.00	93.79	92.81

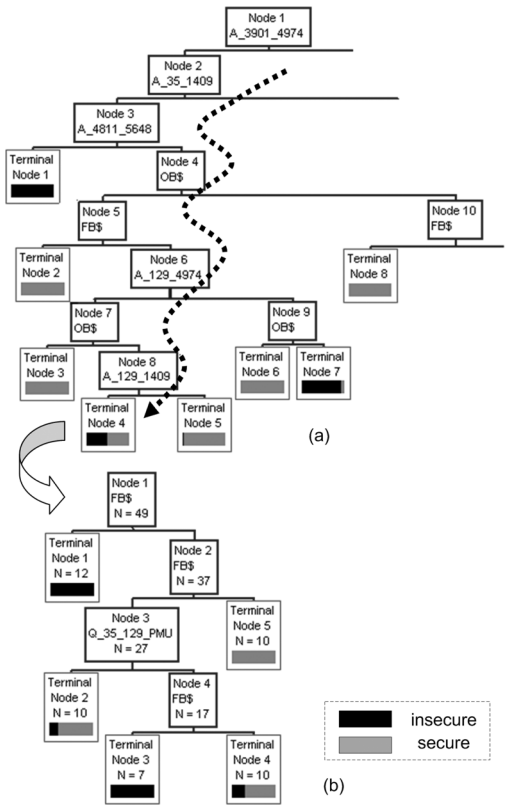

Fig. 6. Corrective DT for a problematic path in DT3.

DTs is similar to the work presented in [20], where additional classification techniques are applied to the nodes with poor performance.

DT3 is trained using the LS containing 7563 cases, among which there are 49 (24 insecure and 25 secure) cases falling into Terminal Node 4. 14 out of 1891 cases in the TS are tested using this path and six secure cases are misclassified as insecure cases. Although higher accuracy for these 49 cases in the LS can be achieved by further splitting Terminal node 4 it will jeopardize the prediction performance on the TS. The problem lies in using A_x_y alone, as this information is not enough to clearly classify these cases. Therefore, the 49 cases in the LS and 14 cases in the TS are picked to form a smaller learning set (LS') and a test set (TS'), respectively, but more predictors like I_x_y and Q_x_y are included. A corrective DT using LS' and TS' is now trained and it only misclassifies five cases in the LS' and 3 cases in the TS'.

TABLE V
PERFORMANCE OF CORRECTED DTS

Opt. DTs	Overall LS Accuracy (%)		Overall TS Accuracy (%)	
	Corrected	Original	Corrected	Original
DT3'	97.46	95.82	94.02	92.70
DT4'	98.21	96.97	**95.45**	94.24
DT5'	98.70	95.33	92.07	91.12
DT6'	97.38	93.81	93.39	90.85
DT7'	98.40	96.73	93.92	92.6

This small DT [depicted in Fig. 6(b)] performs much better than the original path that reaches Terminal Node 4 in DT3 for these $49 + 14 = 63$ cases. In real-time application, the original DT3 is first used to investigate voltage security. The cases that fall into Terminal Node 4 are further evaluated using this corrective DT to obtain the final security assessment. This approach preserves the branches with high accuracy and only modifies the paths with severe misclassification problems by using corrective DTs rather than re-training the whole tree. This method is tested on all the DTs from DT3 to DT7 that are trained using only one type of pre-contingency predictors and the improved DT performance are shown in Table V.

From Table V, all of the original DTs have better performance for both learning set and test set. The best performance is achieved in DT4' and it can correctly predict 95.45% cases in the TS, which has a 1.21% improvement. During the training of corrective DTs, not all of the problematic paths can be improved by adding more predictors for further classification. The main reasons are: 1) there are not enough cases falling into the problematic path to train a corrective DT; and 2) the cases in the LS that are classified by the problematic path all belong to one class, secure or insecure, which makes it impossible to build the corrective DT. Further investigation of this "multi-level DT" idea to improve prediction accuracy will be carried out in the future work.

VI. CONCLUSIONS AND FUTURE WORK

This paper presents a method for online voltage security assessment using PMUs and decision trees. The DTs are trained offline and periodically updated for robustness improvement. By comparing the PMU measurements with the CSRs in the DTs in real time, a fast voltage security assessment for severe contingencies can be obtained. Twenty-nine OCs for the AEP system are generated to represent the stressed system operating conditions during a peak load period. Voltage security is analyzed using VSAT to obtain a security label for each case following a severe contingency. Different PMU-related pre-disturbance system parameters are collected to create eight databases for DT training. The result shows that the combination of fault information and current magnitudes performs the best on the test set using a single DT. Two new ideas including using "Multiple optimal DTs" and "Corrective DTs" are also introduced and discussed to improve DT performance. In the future work, other voltage security evaluation tools with more criteria regarding voltage insecurity can be used to perform the offline studies. Time-domain analysis can also be included to obtain a more comprehensive assessment result. The DT enhancement methods will be further extended in more detail and other data mining tools like support vector machine and random forest will be tested to pursue better prediction accuracy on unknown cases.

REFERENCES

[1] Energy Information Administration, *World Net Conventional Thermal Electricity Generation (Billion Kilowatthours), 1980–2005.* [Online]. Available: http://www.eia.doe.gov/iea/elec.html.

[2] C. W. Taylor, *Power System Voltage Stability*. New York: McGraw-Hill, 1994.

[3] P. Kundur, *Power System Stability and Control*. New York: McGraw-Hill, 1994.

[4] A. G. Phadke, "Synchronized phasor measurements in power systems," *IEEE Comput. Appl. Power*, vol. 6, no. 2, pp. 10–15, Apr. 1993.

[5] N. Kakimoto, M. Sugumi, T. Makino, and K. Tomiyama, "Monitoring of interarea oscillation mode by synchronized phasor measurement," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 260–268, Feb. 2006.

[6] L. Zhao and A. Abur, "Multiarea state estimation using synchronized phasor measurements," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 611–617, May 2005.

[7] J. W. Ballance, B. Bhargava, and G. D. Rodriguez, "Monitoring power system dynamics using phasor measurement technology for power system dynamic security assessment," in *Proc. IEEE Power Tech Conf.*, Bologna, Italy, Jun. 2003.

[8] B. Milosevic and M. Begovic, "Voltage-stability protection and control using a wide-area network of phasor measurements," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 121–127, Feb. 2003.

[9] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees," *Wadsworth Int. Group*, 1984.

[10] CART for Windows User's Guide—An Implementation of the Original CART Methodology, 2002, Salford System software manual.

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

[12] 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. Power Syst.*, vol. 22, no. 4, pp. 1935–1943, Nov. 2007.

[13] A. R. Khatib, R. F. Nuqui, M. R. Ingram, and A. G. Phadke, "Real-time estimation of security from voltage collapse using synchronized phasor measurements," in *Proc. IEEE Power Eng. Soc. General Meeting*, 2004, vol. 1, pp. 582–588.

[14] L. Wehenkel *et al.*, "Decision tree based transient stability method a case study," *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp. 459–469, Feb. 1994.

[15] L. Wehenkel and M. Pavella, "Decision trees and transient stability of electric power systems," *Automatica*, vol. 27, no. 1, pp. 115–134, Jan. 1991.

[16] T. Cutsem, L. Wehenkel, M. Pavella, B. Heilbronn, and M. Goubin, "Decision tree approaches to voltage security assessment," *Proc. Inst. Elect. Eng.*, vol. 140, no. 3, pp. 189–198, May 1993.

[17] S. Rovnyak, C. Taylor, and Y. Sheng, "Decision trees using apparent resistance to detect impending loss of synchronism," *IEEE Trans. Power Del.*, vol. 15, no. 4, pp. 1157–1162, Oct. 2000.

[18] K. R. Padiyar and S. Krishna, "On-line detection of loss of synchronism using locally measurable quantities," in *Proc. Transmission and Distribution Conf. Exhib. 2001*, vol. 1, pp. 537–542.

[19] N. Senroy, G. T. Heydt, and V. Vittal, "Decision tree assisted controlled islanding," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1790–1797, Nov. 2006.

[20] E. M. Voumvoulakis and N. D. Hatziargyriou, "Decision trees-aided self-organized maps for corrective dynamic security," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 662–630, May 2008.

[21] DSATools Dynamic Security Assessment Software. [Online]. Available: http://www.dsatools.com/.

[22] Salford Systems, *CART*. [Online]. Available: http://www.salford-systems.com/cart.php.

[23] Y. Sheng, V. Phoha, and S. Rovnyak, "A parallel decision tree-based method for user authentication based on keystroke patterns," *IEEE Trans. Syst., Man, Cybern.*, vol. 35, no. 4, pt. B, pp. 826–833, Aug. 2005.

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

Ruisheng Diao (S'08) received the B.S. and M.S. degrees from the Department of Electrical Engineering of Zhejiang University, Hangzhou, China, in 2004 and 2006, respectively. He is currently pursuing the Ph.D. degree at Arizona State University, Tempe.

Kai Sun (M'06) received the B.S. degree in automation and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 1999 and 2004, respectively.

He is currently a project manager at ERPI, Palo Alto, CA.

Vijay Vittal (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 IIT, Kanpur, India, in 1979, and the Ph.D. degree from Iowa State University, Ames, in 1982.

Robert J. O'Keefe (S'79–M'80) received the B.S. and M.S. degrees in electric power engineering from Purdue University, West Lafayette, IN, in 1980 and 1983, respectively.

He has been with American Electric Power Service, Columbus, OH, since 1990 and is a Principal Engineer in the Advanced Transmission Studies and Technologies section.

Michael R. Richardson received the B.S.E.E. degree from Ohio State University.

He joined AEP Service Corporation, Columbus, OH, as a Distribution Engineer in December 1976. Currently, he is a Senior Engineer in the American Electric Power Transmission Operations Department.

Navin Bhatt (SM'82–F'09) received the M.S.E.E. and Ph.D. degrees from the West Virginia University, Morgantown.

Since joining American Electric Power Service, Columbus, OH, in 1977, he has been involved in conducting, managing, and coordinating advanced transmission studies.

Dwayne Stradford received the B.S.E.E. degree from Virginia Tech, Blacksburg, in December 1992.

He is the Director of Transmission Reliability in Transmission Operations for AEP He joined American Electric Power Service, Columbus, OH, as a Distribution Engineer in December 1976. Currently, he is a Senior Engineer in the American Electric Power Transmission Operations Department.

Sanjoy K. Sarawgi (M'03) received the B.Tech. (Hons.) in electrical engineering from IIT, Kharagpur, India, and the M.S.E.E. degree from Washington State University, St. Louis, MO, in 2002 and 2004, respectively.

He joined American Electric Power Service, Columbus, OH, in 2004.