PSERC

Decision Tree Based Online Voltage Security Assessment Using PMU Measurements

Final Project Report

Power Systems Engineering Research Center

A National Science Foundation Industry/University Cooperative Research Center since 1996

PSERC

Power Systems Engineering Research Center

Decision Tree Based Online Voltage Security Assessment Using PMU Measurements

Final Project Report

Vijay Vittal
Arizona State University

PSERC Publication 08-26

November 2008

Information about this project

For information about this project contact:

Vijay Vittal
Ira A. Fulton Chair Professor
Department of Electrical Engineering
Arizona State University
P.O. Box 875706
Tempe, AZ 85287-5706
Tel: 480-965-1879
Fax: 480-965-0745
Email: vijay.vittal@asu.edu

Power Systems Engineering Research Center

The Power Systems Engineering Research Center (PSERC) is a multi-university Center conducting research on challenges facing the electric power industry and educating the next generation of power engineers. More information about PSERC can be found at the Center's website: http://www.pserc.org.

For additional information, contact:

Power Systems Engineering Research Center
Arizona State University
577 Engineering Research Center
Box 878606
Tempe, AZ 85287-8606
Phone: 480-965-1643
FAX: 480-965-0745

Notice Concerning Copyright Material

Acknowledgements

This is the final report for the Power Systems Engineering Research Center (PSERC) research project, titled "Decision Tree Based Online Voltage Security Assessment Using PMU Measurements." We express our appreciation for the support provided by PSERC's industrial member, the American Electric Power (AEP) Service Corporation.

The author wishes to recognize his graduate student and postdoctoral researcher that contributed to the research and creation of the reports: Ruisheng Diao (Arizona State University, Ph.D. student) and Dr. Kai Sun (Arizona State University, now at EPRI).

The authors thank all PSERC members and industry advisors for their technical advice on this project, especially:

- Robert O'Keefe – AEP Service
- Michael Richardson – AEP Service
- Navin Bhatt – AEP Service
- Dwayne Stradford – AEP Service
- Sanjoy Sarawgi – AEP Service
- Sharma Kolluri – Entergy
- Sujit Mandal - Entergy
- Lei Wang – Powertech Labs

Executive Summary

This research project presents and examines an online voltage security assessment scheme to evaluate post-contingency voltage security in real time by using synchronized phasor measurement units (PMUs) and periodically updated decision trees (DTs). The main objective is to develop a fast and accurate tool to predict whether certain severe contingencies will initiate voltage collapse at the current operating condition for the purpose of preventing voltage collapse in a modern power system. The online assessment results are then used to guide system operators in real-time decision making. This project involved the participation of a PSERC member company – American Electric Power (AEP) Corporation – that provided supplemental funding for the project and the associated data, and participated as an industry advisor to the project.

With increasing difficulty in approving and installing new transmission lines, modern power systems are operated closer to their limits due to the continuous increase in load demand and generation capacity for the past few decades. Thus, severe contingencies, e.g., loss of critical tie lines or several generators simultaneously, may affect system stability leading to a large scale blackout such as the one that occurred on August 14th, 2003 in North America. Among different stability problems, voltage stability is one of the most critical issues that threaten system secure operations. Voltage stability is defined as the ability of power system to maintain acceptable bus voltage magnitudes at normal or contingent operating conditions. Once a disturbance like a fast load change or a severe contingency causes a continuous and uncontrollable decline in bus voltages, voltage instability is considered to occur, especially at stressed operating conditions with insufficient reactive power support. Without timely and correct control actions, the low voltages may spread throughout adjacent areas and eventually cause a wide area blackout instead of a localized outage.

Traditionally system states were monitored through the supervisory control and data acquisition (SCADA) system which did not have the capability of capturing fast transient behaviors due to the relatively low data sampling frequency. Besides, the measurements are not synchronized, which makes it impossible to obtain system snapshots in real time. The time span of losing voltage stability ranges from 0.1 second to several hours, representing transient and long-term voltage instability respectively. For the fast transients following severe disturbances, detailed voltage security analysis for online application becomes a great challenge because of the high computational burden in a large power system. Therefore, it is imperative to develop a fast and accurate voltage security assessment tool to detect the potential voltage problems, enable remedial control actions and effectively minimize the adverse impact of voltage insecurity on the whole system.

A literature survey was first conducted to examine the principles of voltage instability and review the approaches to evaluate voltage stability in the stage of real time operation. Taking advantage of the previous efforts, this research report presents a DT based voltage security assessment scheme by using PMUs for online application. The main idea is to first perform detailed voltage security simulations offline at various

operating conditions that represent all the system states forecasted for the next 24 hours. A supervised data mining tool DT is adopted to identify critical system attributes as voltage security indicators from the databases that are created by offline study. At the stage of online application, synchronized phasor measurements are collected to compare with these thresholds and obtain a final assessment result after a severe contingency is detected. The proposed scheme has been developed using the operational model that represents the North American Eastern Interconnection in which the eastern AEP system has approximately 2400 buses. The details of this scheme are developed in the following steps:

i. Offline DT Training (24 hour ahead). Based on the detailed power flow file of the AEP system and load changing pattern for a period of time in March 2007, 29 operating conditions are generated by adjusting the load levels and generator outputs. These operating conditions represent the system states under stressed conditions in order that voltage stability problems in the AEP system can be identified. The bus voltages are all at normal operating levels although some transmission lines are overloaded. Exhaustive voltage security simulations are conducted for a list of critical contingencies that have the possibility of losing voltage stability, provided by the AEP staff. DTs are then trained to obtain the desired prediction performance based on the databases that consist of offline simulation results and different PMU-related pre-disturbance system attributes.

ii. Periodic DT Update (1 hour). The offline trained DTs select a small number of system attributes as voltage security indicators and their thresholds are also determined in the tree models. In order to obtain good prediction performance in case of any significant change in the system, these DTs are updated on an hourly basis. For example, loss of a transmission path will change the power flow patterns and the post-disturbance operating conditions are not included in previous offline simulations. In this case, voltage security analysis for the new operating conditions needs to be carried out and included in the databases for DT updating. The modified DTs, if necessary, are finally used for online application during the next hour.

iii. Online Application. At this stage, real time measurements from PMUs are dropped into the DTs and compared with the corresponding thresholds to determine whether a severe disturbance will cause voltage instability. The process is very fast since only a few comparisons are performed. An insecure prediction result will then provide an alarm and guide system operators to take prompt remedial control actions to prevent the occurrence of losing voltage stability.

This scheme takes advantage of all the available and planned PMUs across the whole AEP system to evaluate global voltage stability problems. The synchronized signals and high sampling frequency endow PMUs with the capability of observing different states across the whole system in a common time frame with much higher accuracy. The DT technique serves as an effective data mining tool to solve the classification problems in high data dimensions. The periodically updated DTs not only uncover critical system attributes, but also build a nomogram in terms of the thresholds of these system parameters. This provides system operators with an effective alternative for system monitoring and online decision making. In addition, once a PMU measurement is missing or invalid, DTs will pick up effective "competitors" as candidates for online application. The test study shows that properly trained DTs perform quite well in assessing voltage security in real time. Several innovative ideas to improve DT performance in assessing voltage security including the applications of "Multiple DTs", "Corrective DTs" and "Maximum DTs" are also introduced. The main accomplishments resulted from this research project are listed below:

i. The feasibility of using periodically updated DTs and synchronized PMUs for online voltage security assessment is analyzed. The proposed scheme is tested on an operational snapshot of the AEP system. The test results have shown good prediction performance on the system model considered.

ii. The project work has incorporated (1) the Dynamic Security Analysis Tools (DSA^{Tools}) software developed by Powertech Labs for operating condition generation and voltage security analysis, and (2) 'Classification And Regression Trees' (CART) developed by the Salford system, for training and testing DTs. A software platform that implements the entire scheme, including data collection, data conversion and interface design, has been developed in MATLAB and VC++ codes.

Table of Contents

1 Introduction 1
1.1 Background 1
1.2 Phasor Measurement Unit 5
1.3 Decision Tree 8
1.4 Feasibility Analysis 11
2 The Proposed Online Voltage Security Assessment Scheme 13
2.1 Introduction 13
2.2 Offline DT Training 13
2.3 Periodic DT Update 13
2.4 Online Application 15
2.5 Advantages 15
3 Case Study 16
3.1 Introduction 16
3.2 Operating Condition Generation 17
3.3 Voltage Security Analysis 19
3.4 Predictor Selection & Database Generation 19
3.5 DT Training and Performance 20
3.6 DT Performance Improvement 23
3.6.1 Multiple Optimal DTs 25
3.6.2 Corrective DTs 27
3.6.3 Maximum DTs 29
4 Conclusions 30
5 References 31
5.1 Paper Resulted from this Project 31
5.2 Background Papers 31
Appendix A. Critical System Attributes Selected in the Opt. DTs 34
Appendix B. Structure of Optimal DTs in Section 3.5 37

Table of Figures

Figure 1-1: World Power Consumption Growth from 1980 to 2005 2
Figure 1-2: World Total Electricity Installed Capacity from 1980 to 2005 2
Figure 1-3: Progress of a Voltage Collapse Case 4
Figure 1-4: Basic Structure of a Phasor Measurement Unit 6
Figure 1-5: Block Diagram of the Macrodyne 1620 PMU 7
Figure 1-6: A DT Sample with 4 Internal Nodes and 5 Terminal Nodes 8
Figure 1-7: A Sample of DT Training Process in CART 10
Figure 2-1: Flowchart of the Proposed Scheme.. 14
Figure 3-1: Service Territories of the Eastern AEP System 16
Figure 3-2: PMU Locations in the Eastern AEP System (Installed and Planned)............ 17
Figure 3-3: Load Pattern of the 29 Stressed OCs 18
Figure 3-4: Distribution of A_35_539 for the original 29 OCs 24
Figure 3-5: Distribution of A_35_539 for the new 22 OCs 24
Figure 3-6: Concept of Multiple DT Application 25
Figure 3-7: Corrective DT for a Problematic Path in DT3 28
Figure B. 1 Structure of DT3 with critical splitting rules 37
Figure B. 2 Structure of DT4 with critical splitting rules 38
Figure B. 3 Structure of DT5 with critical splitting rules 39
Figure B. 4 Structure of DT6 with critical splitting rules 40
Figure B. 5 Structure of DT7 with critical splitting rules 41
Figure B. 6 Structure of DT8 with critical splitting rules 42

Table of Tables

Table 1-1: Performance Comparison of the 8 Selected DTs on the LS 10
Table 3-1: PMU Locations in the Eastern AEP System 18
Table 3-2: Different PMU-related Predictor Groups in the AEP System......................... 19
Table 3-3: Sample of the Created Databases 21
Table 3-4: Performance Comparison of the Optimal DTs 21
Table 3-5: Number of Critical System Attributes Selected in the Optimal DTs 22
Table 3-6: DT Performance for Different Predictor Combinations 26
Table 3-7: Performance of the Corrected DTs 27
Table 3-8: Maximum DT Performance on the Whole Database 29
Table A. 1 Contingency-independent CSRs in DT3 34
Table A. 2 Contingency-independent CSRs in DT4 34
Table A. 3 Contingency-independent CSRs in DT5 35
Table A. 4 Contingency-independent CSRs in DT6 35
Table A. 5 Contingency-independent CSRs in DT7 36
Table A. 6 Contingency-independent CSRs in DT8 36

1 Introduction

As modern power systems are operated closer to their operating limits, voltage stability has become a critical issue that threatens secure system operation. Without accurate security assessment and timely control action, voltage instability may lead to a large scale blackout with tremendous economic consequences. In real time operations setting, traditional analysis methods are not fast enough to evaluate post-disturbance voltage stability problems in a timely manner. This research effort explores the feasibility of an online voltage security assessment tool to rapidly assess voltage security by using synchronized phasor measurement units and periodically updated decision trees.

The main objective of this research project is the design of such a tool for the purpose of obtaining a fast and accurate voltage stability assessment. It is therefore important to review the mechanisms of voltage instability and the existing methods to evaluate voltage stability that are related to the objectives of this project. An online voltage security assessment scheme is proposed and tested on an operational snapshot of the American Electric Power system model. The properly trained DTs are able to select a small number of PMU related system attributes as voltage stability indicators and they have shown satisfactory prediction performance that will effectively guide system operators in real time decision making. This research report is organized as follows:

In Section 1, the authors first introduce the background of this research project and conduct a literature review to address the above issues. Feasibility analysis of the proposed online voltage security assessment scheme is performed while the principles of synchronized PMUs and DTs are also explored. In Section 2, the details of the proposed scheme that consist of 3 major steps, "Offline DT Building", "Periodic DT Update" and "Online Application" are discussed. Section 3 tests this scheme on the AEP operational power system. All the details regarding operating condition generation, database creation, predictor selection, voltage security analysis and DT training are included. In addition, 3 innovative methods to improve DT performance and robustness for online application are presented. Section 4 draws conclusions from this project; while Section 5 lists the publications resulting from this project and the references that are related to this research effort.

1.1 Background

Modern power systems are interconnected by transmission lines to transport large amounts of electric power over thousands of miles for both reliability and economic reasons. Uninterruptable electricity supply with high quality constitutes one of the most important requirements for the development of human societies. During the past few decades, countries with highly developed economies have seen the consumption of electric power in industrial, commercial and residential areas rise at consistent annual rates. The growth of electric power consumption all over the world from the year 1980 to 2005 is shown in Figure 1-1. In 2005, the world's total electricity power consumption increased to 15,746.54 million kWh, which is approximately 2.147 times of the consumption in 1980 [1]. In order to keep pace with the load growth, generation capacity has also been boosted by constructing large, concentrated power plants. The annual growth of world total generation capacity for the same time period is shown in Figure 1-2.

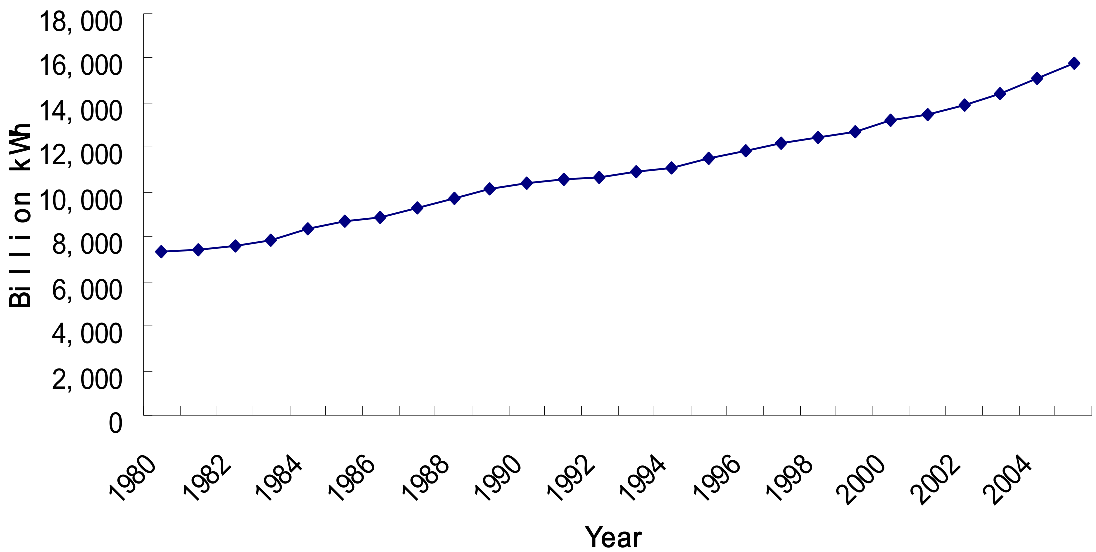

Figure 1-1: World Power Consumption Growth from 1980 to 2005

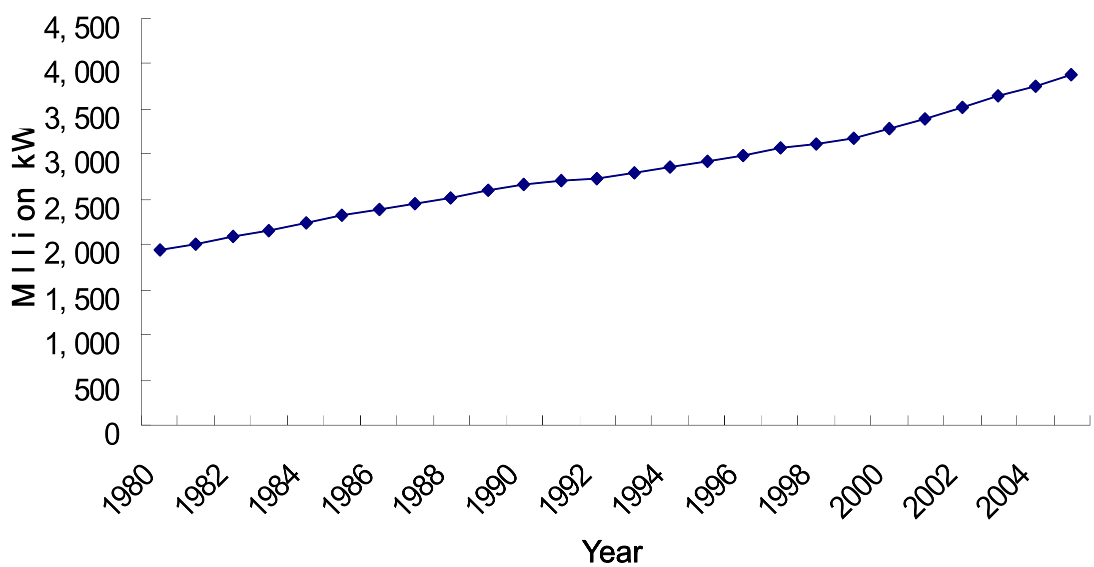

Figure 1-2: World Total Electricity Installed Capacity from 1980 to 2005

However the growth in load demand and generation capacity has not resulted in a concomitant increase in the transmission capability due to the difficulty of approving new transmission lines caused by the increasing environment and investment concerns. As a consequence, existing power networks are likely to be operated under great stress with transmission lines carrying power near their limits during peak load periods. At normal operating conditions, electric power transactions are well scheduled and planned in advance to minimize the overall economic cost, while during disturbances the interconnected systems can provide active and reactive power support to the area that is subjected to the disturbance, such as a fast load increase or loss of several critical transmission lines. If the disturbance is small, a new equilibrium point will be reached by re-scheduling power flows and switching equipments in the surrounding areas. Mature power systems in North America should have the capability of tolerating the loss of any one element in the system (N-1) according to NERC standards such as the outage of a large generator, a critical transmission line or a heavily loaded transformer. But in extreme cases, loss of one or more critical transmission path(s) may initiate system-wide stability problems such as transient stability and voltage stability. This situation could be worsened when any additional contingency occurs in an already disturbed system, which forms an N-2 or even N-k condition.

Once a severe disturbance causes a large imbalance between generator output and load demand in an area, transient stability is more likely to be affected with generator rotor angles swinging against each other; while voltage instability is likely to occur when such a contingency causes insufficient reactive power support to a load area with a large amount of reactive power demand. Many of previous research efforts were focused on addressing various aspects of the transient stability problem. Approaches for transient stability monitoring and detection, transient stability margin identification, transient stability prediction, online security assessment and the design of stability controls have been widely reported in [2]-[11]. On the contrary, effective measures to evaluate voltage stability for real time operation of a realistic interconnected power system are limited. Thus, this research project focuses on addressing the online voltage security assessment issue, takes advantage of previous efforts and presents a decision tree based scheme for this subject.

Voltage instability (or voltage collapse) is one of the most critical problems that threaten secure operations of an interconnected power system, which is typically associated with lack of sufficient reactive power support in a heavily loaded area. Once initiated by a serious disturbance, bus voltage magnitudes progressively decline to an unacceptable level and the decreasing voltages cannot be controlled back to normal levels. The low voltages not only affect the local load area, but also spread throughout adjacent areas of the power system. Cascading outages of transmission lines due to protective relay actions may be triggered that eventually cause a large scale blackout instead of a localized outage. Figure 1-3 shows the progress of a voltage collapse phenomenon. Several voltage collapse incidents in the world that resulted in tremendous losses were reported in [12].

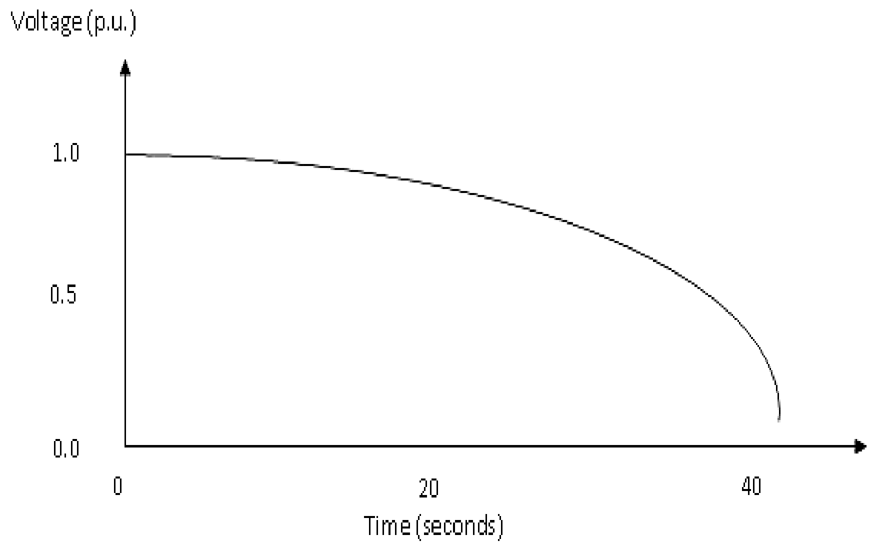

Figure 1-3: Progress of a Voltage Collapse Case

Voltage instability is usually caused by either a fast load increase or a major change in network topology resulting from a critical contingency, both of which will cause the lack of effective reactive power compensation. In addition, the decreasing bus voltage profiles may deteriorate during transients due to the excessive use of capacitor banks since the reactive power provided by the capacitors is proportional to the associated bus voltage magnitudes. Due to different mechanisms of losing voltage stability, the time scale for this phenomenon ranges from 0.1 second to several hours. For long-term stability that takes several minutes to several hours to occur, system operators may have ample time to take remedial control actions to ensure enough reactive power reserve by generation re-scheduling, gas-turbine starting up, fast load shedding and shunt element switching. But for the fast transients like loss of large induction motors, voltage collapse can start as soon as the disturbance occurs. In order to effectively prevent voltage collapse, both these types of voltage instability require an accurate and fast assessment result addressing whether a certain disturbance at the current OC will cause severe problems.

From the literature survey conducted, it is observed that different methods have been developed to evaluate voltage stability including sensitivity analysis, modal analysis, PV and QV curve calculation, L index and other voltage stability indices [12]-[18]. However as the size of a power system grows very large, detailed voltage stability analysis for online application becomes a great challenge because of the high computational burden. This leaves insufficient time for operators to examine the post-disturbance system behavior and take remedial control actions. As a result, an accurate and fast assessment tool to determine the voltage security in real time becomes a necessity for preventing the occurrence of voltage instability.

This report focuses on the voltage collapse problems that are caused by severe contingencies and adopts an effective data mining tool, decision tree (DT), to assess voltage stability in real time operation. The main idea is to first perform detailed voltage security simulations offline at various operating conditions that represent the system states forecasted for the next 24 hours. Decision trees are then trained to identify critical system attributes as voltage security indicators from the databases that are created from the offline study. The thresholds of these critical system attributes are also determined in the DT models. At the stage of online application, synchronized phasor measurements are collected to be compared with these thresholds and obtain a final prediction result. The details of this proposed scheme that consists of 3 major steps are discussed further in Section 2.

This scheme takes full advantage of all the available and planned PMU measurements across a power system to assess post-contingency voltage security and it is tested on an AEP operational snapshot 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. Several new ideas to improve DT performance and reliability including the applications of "Multiple DTs", "Corrective DTs" and "Maximum DTs" are also developed and tested. Before discussing the proposed method in more detail and its application on a realistic power system, principles and advantages of PMUs and DTs are introduced in the following two subsections.

1.2 Phasor Measurement Unit

State monitoring in power systems plays an important role in system operation and online decision making and is the basis for the real time power flow analysis used in voltage stability or transient stability analysis. Traditionally power system states like voltage magnitude, voltage phase angle, current magnitude and power flows on transmission lines were monitored using analog devices such as current transformers and potential transformers and communicated to the energy management system (EMS) through the supervisory control and data acquisition (SCADA) system. System information is periodically updated with the sampling frequency in the range of a few seconds. State estimation is then performed to obtain a converged solution for further application. This approach served the industry well but lacks the ability of observing measurements across the whole system because the data was not time synchronized. Thus, there is no way of obtaining a real time snapshot of the system.

The advent of phasor measurement units (PMUs) has revolutionized the field of power system state monitoring. PMUs have significant advantages over the traditional measurements in terms of both accuracy and frequency of sampling. A basic PMU configuration consists of a global positioning system (GPS) receiver, a filter, an analog/digital converter and a microprocessor as shown in Figure 1-4.

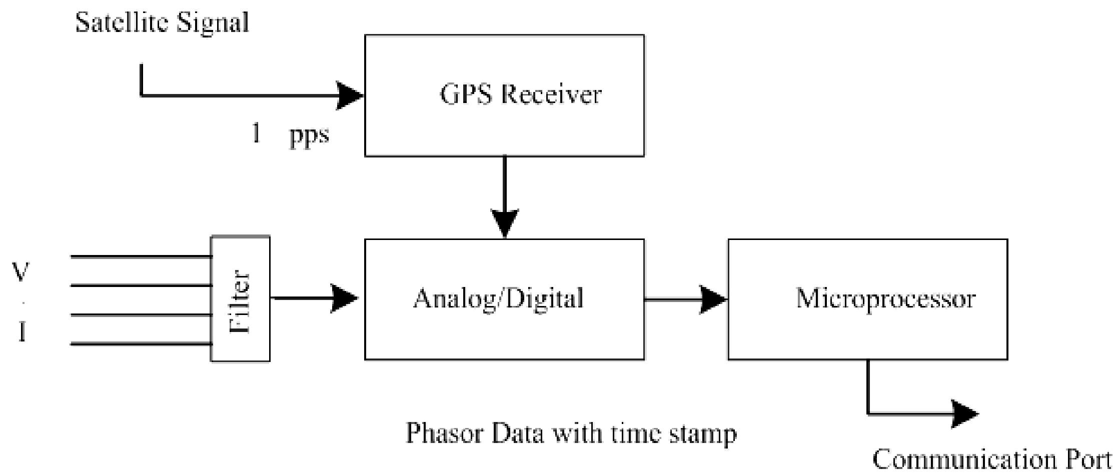

Figure 1-4: Basic Structure of a Phasor Measurement Unit

From the satellites, the GPS receiver receives a one pulse per second (PPS) signal with a time stamp containing the year, day, hour, minute and second, which is divided into a number of pulses for the sampling of the analog signals derived from conventional measurement devices. The number of pulses can be defined by the users, which is usually 12 times per cycle in a 60 Hz system. With these synchronized time stamps, analog signals are converted to digital signals and sent to the microprocessor for further calculation. Thus, generator angles, bus voltages, bus angles, power flows, frequencies, currents, power factors and other system information can be calculated in the microprocessor. The time-stamped 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 [19]. Therefore, a PMU has the ability to observe different system states across the whole system in real time even if they are thousands of miles away from each other.

The development of PMUs dates back to the middle 1980s when the PMU was first introduced at Virginia Tech. PMUs have been commercialized and widely used in f power systems all over the world, an example of which is shown in Figure 1-5 [20]. As more advanced control techniques and devices are used, the applications of PMUs in power system are being widely reported in the areas of wide area monitoring, loss of synchronism detection, multi-area state estimation, oscillation mode identification, voltage stability protection and system dynamics monitoring [21]-[25].

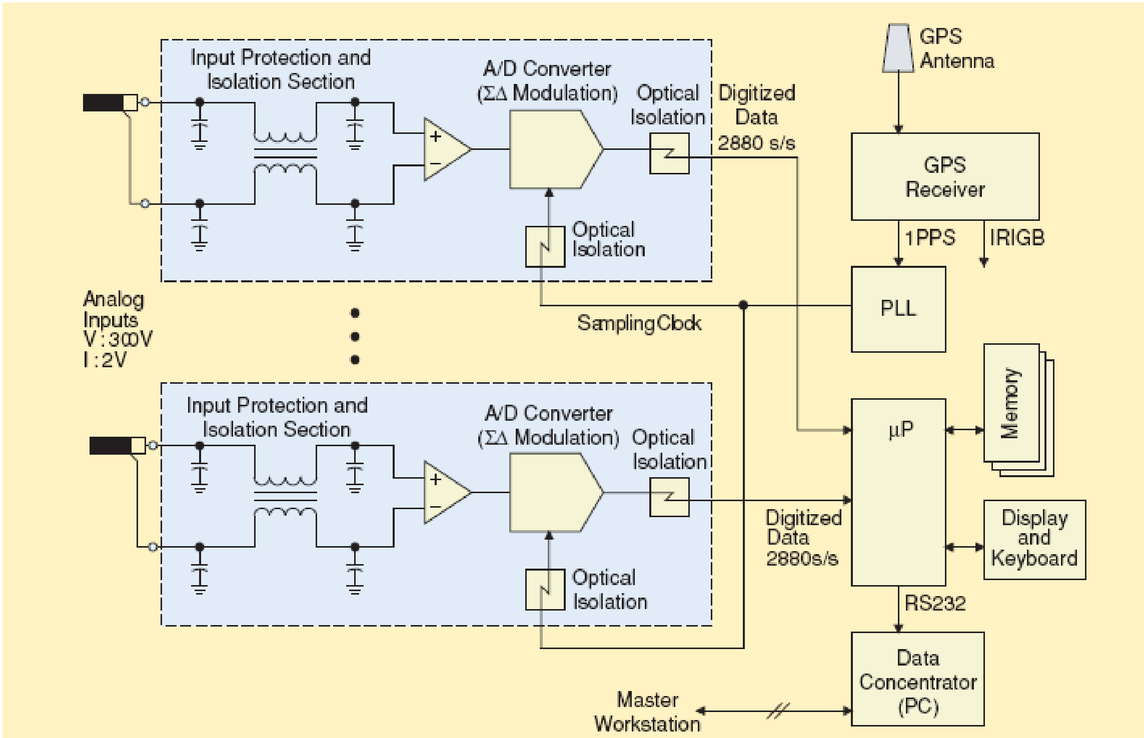

Figure 1-5: Block Diagram of the Macrodyne 1620 PMU (January 1992, taken from [20])

1.3 Decision Tree

The decision tree technique is an effective supervised data mining tool to solve the classification problems in high data dimensions. It was first developed by Leo Breiman in 1980's, and is now widely used in many fields including medical diagnosis, finance analysis, statistical analysis and decision makings in power system [26]. Different applications of DTs depend on the ability to describe the problem by creating a **database** that consists of a sufficient large number of **cases.** Each case is represented by a vector of **predictors** (or variables) along with an **objective**. The DT is designed to represent a classification or predictive model for this objective by identifying critical attributes that affect this objective most effectively and directly.

The DT structure is usually dimidiate and there are two types of nodes in a DT, the "internal node" with two successors (or children) 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 or categorical, by comparing the variable value with a threshold or checking whether the current value belongs to a specific data set, respectively. For each terminal node, a classification will be assigned in terms of the majority classes 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. The whole process starts from the root node until a terminal node along with a final classification is reached. One sample of decision tree that has 4 internal nodes (in blue boxes) and 5 terminal nodes (in red boxes) is depicted in Figure 1-6.

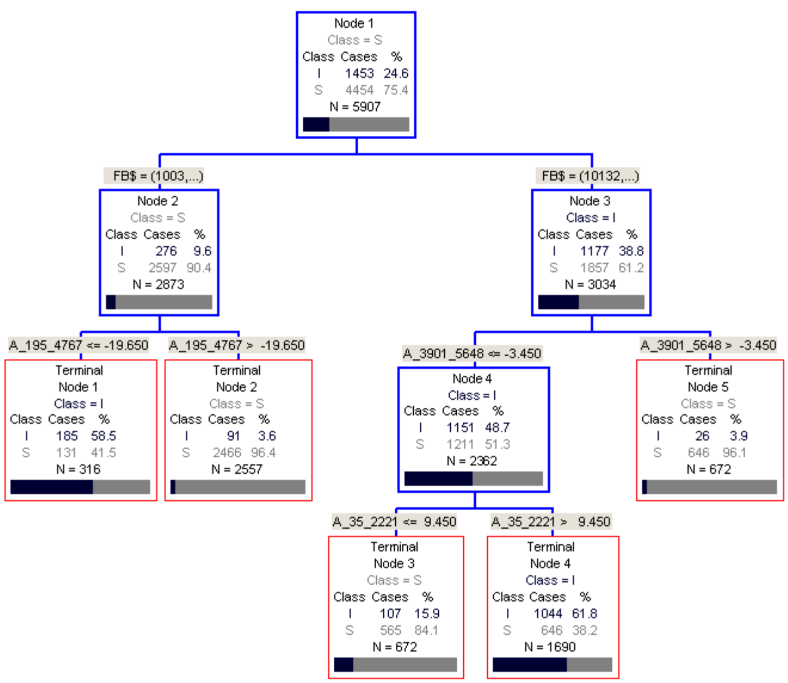

Figure 1-6: A DT Sample with 4 Internal Nodes and 5 Terminal Nodes

This DT is trained from a database with 5907 cases and is similar to the DTs trained in later sections, in which FB$ is a categorical predictor representing the faulted bus location of a contingency; while the remaining predictors that represent voltage phase angle difference between two PMU buses, A_195_4767, A_3901_5648 and A_35_2221 are all numerical predictors. The classifications in this tree are "Secure", represented by the gray bars at the terminal nodes and "Insecure", represented by the black bars. For each internal node, the node description includes the node name, the critical splitting rule (CSR) with its threshold and the number of cases that fall into this node. For each of the terminal nodes, the node name and the number of cases that fall into this classification are shown. The length of the bars at the bottom demonstrates the percentage of different classifications. An example of the classification progress is as follows. Starting from the root, if the faulted bus location is in the set {1003,…}, the whole classification goes to the left child node, otherwise it goes to the right. Let's assume the contingency occurs just at bus 1003. The second level comparison is conducted by checking whether the measured value of A_195_4767 is smaller or equal to -19.650. If so, an insecure assessment result will be provided, which is the "decision" of the DT.

Training a DT not only uncovers critical 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) in 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. 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. In order to obtain the CSRs in the DT, all the available predictors in the LS are scored in terms of the impurity reduction performance by using the equation below.

$$\Delta I(\alpha, t) = I(t) - I(t_L) - I(t_R)$$

Where t represents the current node in the tree, t_L is the left child node, t_R is the right child node, α is the splitting rule and $I(t)$ is the tree impurity function. Each predictor has its own splitting rule, α, and the one with the highest score is selected to be the critical splitting rule for this node. The remaining splitting rules with equal or less scores are called "competitors", which could provide good alternate candidates. Details of algorithms for identifying CSRs and split stopping rules are discussed in [26], [27]. 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, which is defined as

$$MC^{TS} = \frac{1}{N^{TS}} \sum_{i,j} c(i,j) N_{ij}^{TS}$$

where MC^{TS} is the misclassification cost of the whole tree, N^{TS} is the total number of cases in the TS, $c(i, j)$ represents the cost of misclassifying the i class as a j class and N_{ij}^{TS} is the number of j-class cases misclassified as i-class cases in the TS. The **optimal tree** is then defined as the one with the lowest misclassification cost. It always has a medium size, because a small tree does not contain enough useful information and a large DT usually has the over-fitting problem.

In order to illustrate how an optimal DT is obtained, a sample of the DT training process is shown in Figure 1-7. This optimal DT is trained to assess voltage security problems in the AEP system using commercially available software, Classification and Regression Trees (CART). In this case, there are altogether 37 DTs trained in a series. The first DT is the smallest with only 2 nodes, while the largest DT has 94 nodes. From the 37 DTs, eight trees are selected for comparison with their prediction accuracy shown in Table 1-1. In the column indicating the overall accuracy for the learning set, the accuracy keeps increasing as the DT grows larger, because the splitting algorithm in CART keeps using new predictors to find out more splits to obtain a higher accuracy. Then the test set is used to prune the large trees by calculating the misclassification costs for every built tree. Thus, DT-20 is selected to be the optimal one, since its misclassification cost value on the test set is the lowest, 0.194.

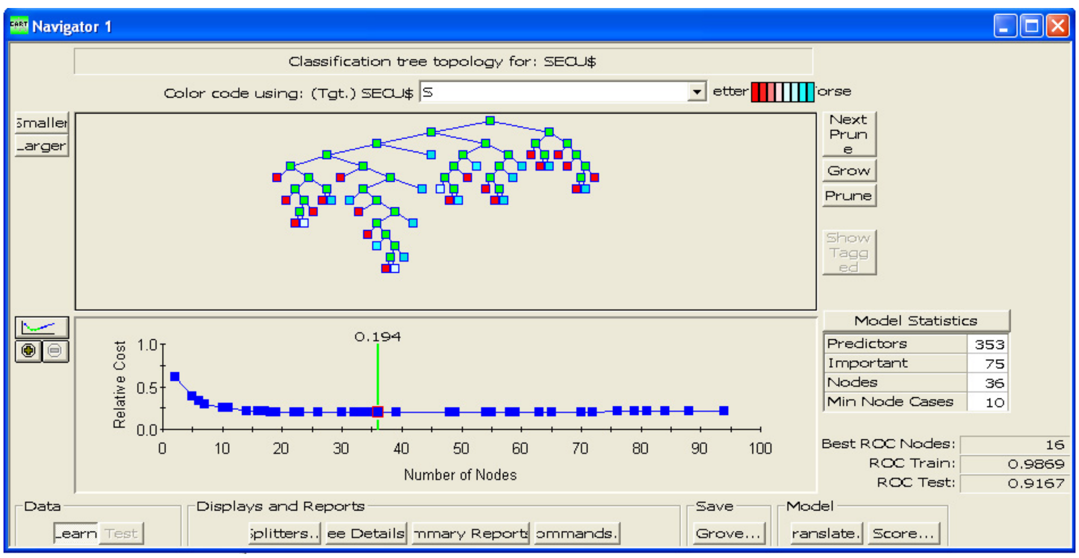

Figure 1-7: A Sample of DT Training Process in CART

Table 1-1: Performance Comparison of the 8 Selected DTs on the LS

No. of DTs	Size	Classification Accuracy (%)		
		Insecure	Secure	Overall
1	2	89.2%	52.19%	64.95%
5	10	90.73%	89.74%	90.08%
10	18	96.22%	91.28%	92.99%
15	30	95.49%	95.95%	95.79%
20	36	97.24%	95.16%	95.87%
25	55	98.77%	95.98%	96.94%
30	70	99.02%	97.01%	97.70%
37	94	99.29%	97.61%	98.19%

Training a good DT usually considers several constraints:

a) Minimum misclassification cost:

This ensures that the overall cost for both of the misclassified insecure cases and misclassified secure cases is the minimum.

b) Overall prediction accuracy:

This requirement considers the total number of misclassified cases to be minimized, which could be different from the first constraint since a higher cost is usually assigned to the misclassified insecure case.

c) Size of the DT:

It is usually desired that the obtained DT has a compact size in terms of depth and total number of terminal nodes, because a large tree has a much higher complexity than a smaller tree.

d) Number of cases in the terminal nodes:

Such a constraint helps to improve the efficiency of a DT.

e) Robustness:

If the system operation condition is disturbed, the performance of a properly trained DT on this OC is an important rule to measure this tree model.

Usually, a compromise needs to be made from the above requirements. In this report, the minimum misclassification cost on the test set is considered as the most important constraint before the final DT is trained. The size constraint of a DT is ignored in this research work, because all the DTs trained have a reasonable size and the voltage security predictors are collected from the existing PMUs across the AEP system.

1.4 Feasibility Analysis

The main advantages of the decision tree tool over the other data mining tools like neural network and support vector machine are the simpler structure and better readability of the model for prediction, which 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. Once a final DT is trained with satisfactory performance for online applications, the CSRs in the tree provide a "nomogram" in the space of critical attributes, which defines a secure operating region. When the system state is changing rapidly during and following disturbances, it is imperative that these critical attributes be measured simultaneously to determine whether the current OC falls inside the nomogram or outside it. This requirement can be satisfied by the synchronized measurements obtained from the PMUs across the system. From the perspective of speed, the DT training process from start to finish usually takes one or two minutes on a PC with a Intel Core2 CPU 6700 (2.66 GHz) and 2.0 GB of RAM once the database is created. The major computational burden rests with the need to conduct a large number of offline voltage security simulations to obtain a sufficient large database. However, this problem can be easily solved by parallel computations because all the cases in the database are independent of each other.

Therefore, properly trained DTs are quite suitable to identify critical system attributes from various system states that are related to power system security problems and feasible for real time transient stability assessment. Several applications involving decision trees have been addressed in real-time transient stability prediction and assessment, voltage security monitoring and estimation, and loss of synchronism detection and timing of controlled separation in power systems [28]-[33]. A tool has been recently developed to combine DTs with the other data mining tool for prediction performance improvement in the field of dynamic security controls [34]. Robustness and a high level of prediction accuracy are two aspects of great importance to be addressed before a final DT is obtained for online application. The following sections will consider a realistic power system to illustrate the process of this proposed scheme for online voltage security assessment in more detail.

2 The Proposed Online Voltage Security Assessment Scheme

2.1 Introduction

The proposed scheme consists of 3 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 [6]. The flow chart of the developed procedure is depicted in Figure 2-1, in which each of the 3 major procedures includes several aspects. In this scheme, two important assumptions are made:

i. The voltage profiles of the whole system at the base cases are maintained at normal levels, no matter how stressful an operating condition is. This assumption ensures that no low voltage problems exist at the base cases even if some of the transmission lines are overloaded.

ii. Voltage collapse cases are initiated only by critical contingencies. The voltage instability caused by continuous load increase is not directly addressed here. In fact, the load increase effect on system voltage security can be effectively reflected by including operating conditions at different load levels.

2.2 Offline DT Training

A number of OCs (Noc) representing all the details of system states including load levels, generation patterns, power flows and equipment status are first collected for the past 24 hours and the next 24 hours from load forecasting techniques. A list of contingency (Nc) is required for the purpose of identifying the critical contingencies that may cause voltage collapse and they can be obtained from historical events and experiences of system operators. For each of these Noc operating conditions, detailed voltage security analysis for all of the provided Nc contingencies are conducted. From simulation results, each contingency case is assigned a voltage security label, secure (S) or insecure (I). By collecting different types of PMU-related system parameters for use as voltage security predictors from the pre-disturbance system parameters, a database consisting of Noc*Nc contingencies is then created. Therefore, DTs can be trained offline to obtain the security classifications for the next day. The selection of voltage security predictors will be discussed in Section 3.

2.3 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 such as the loss of big generators or a sudden load increase, 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 database to build new DTs with higher prediction accuracy. The updated DTs are then used for online applications during the next hour.

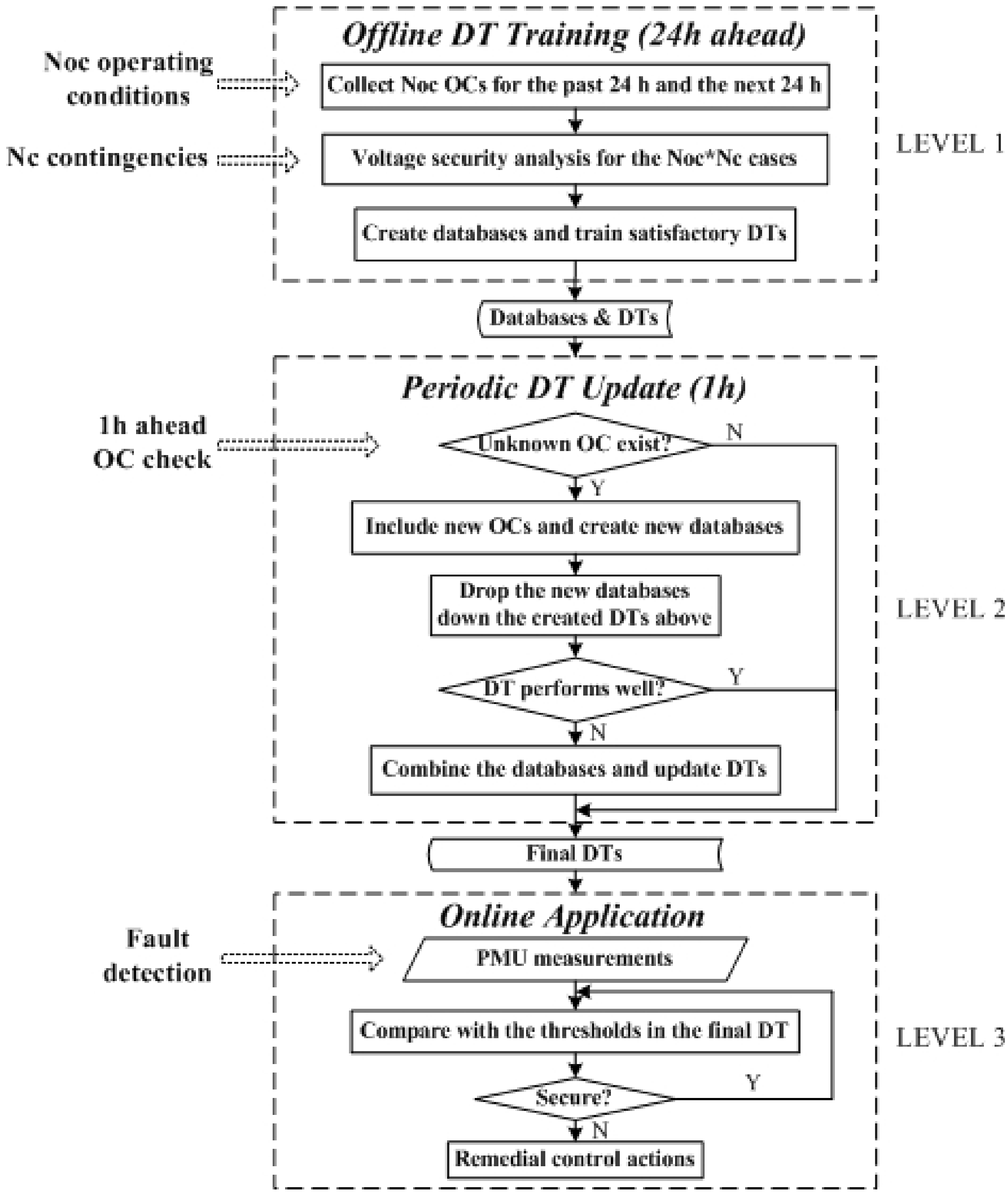

Figure 2-1: Flowchart of the Proposed Scheme

2.4 Online Application

During real time operation, a fault detection system is required to identify whether a critical contingency in the list occurs. If so, the obtained DT is used for online security assessment by reading the measurements from PMUs to check whether the thresholds in the DT are violated. Once an insecure assessment result is obtained, it will provide system operators an alarm and trigger remedial control actions. No action is taken until the contingency occurs. This research project only addresses the topic of security assessment. The topic of corrective and preventive control is beyond the scope of the work done in this research effort.

2.5 Advantages

This scheme 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. This approach 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 system parameters as voltage stability indicators, which results in enhanced efficiency compared to traditional analysis methods to obtain a security result. The CSRs can be recommended PMU locations. After severe disturbances occur, system operators have more time to take remedial control actions to adjust the current OC away from the insecure boundaries if an insecure assessment is obtained. Furthermore, since the offline voltage security simulations can be conducted by parallel calculation and the DT training process only takes a few minutes, it ensures enough time to update these databases and DTs every hour for real time application. This 1-hour time interval could be shortened if necessary.

3 Case Study

3.1 Introduction

The proposed scheme is tested on an American Electric Power operations snapshot. This power system model contains 18168 buses, 2753 generators, and 19358 lines of the North American Eastern Interconnection of which a subset numbering approximately 2400 buses represents the eastern AEP system. The service area of the eastern AEP system is shown in Figure 3-1.

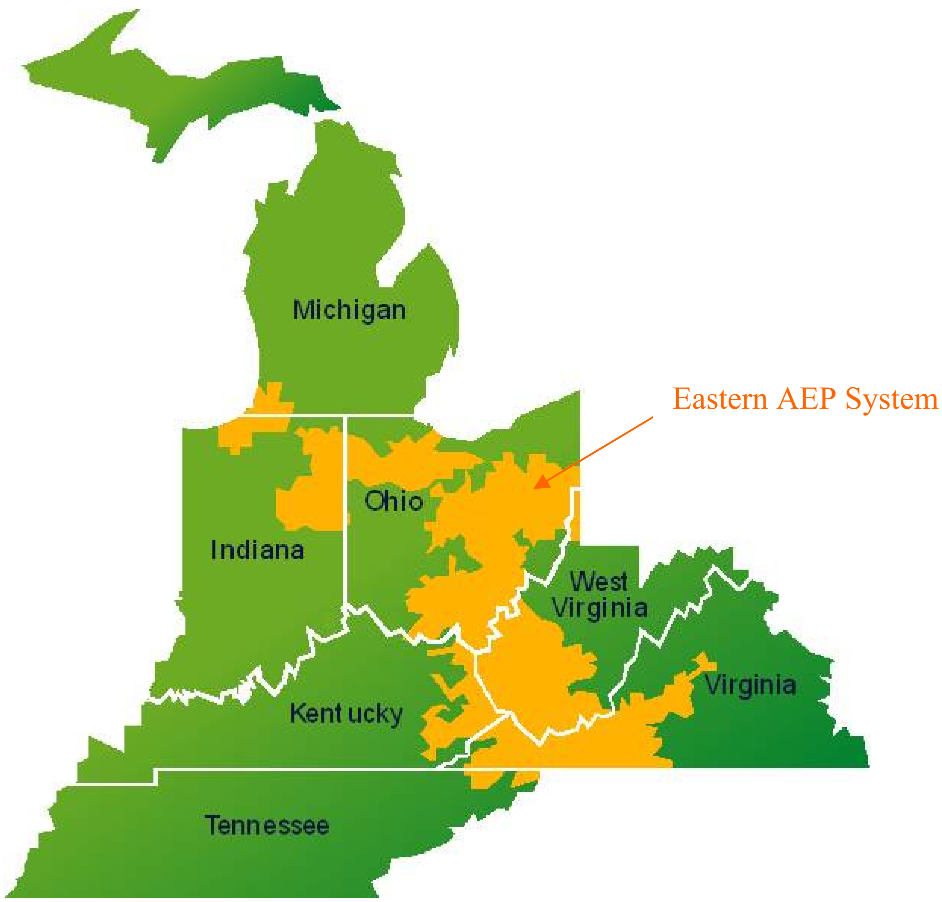

Figure 3-1: Service Territories of the Eastern AEP System

Currently, there are 12 PMUs located across the eastern AEP system and 15 additional units planned that cover a wide area and their geographical locations are shown in Figure 3-2. One PMU is on a 138 kV bus, 11 PMUs are on 345 kV buses while the remaining PMUs are on 765 kV buses, which are listed in Table 3-1. 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.

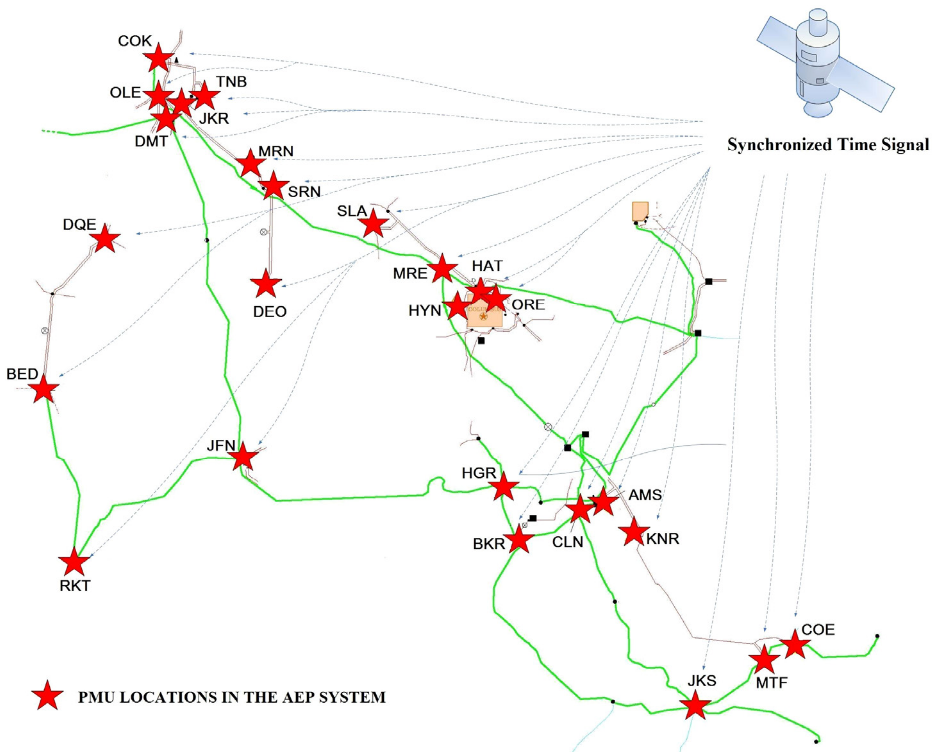

Figure 3-2: PMU Locations in the Eastern AEP System (Installed and Planned)

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 using the Voltage Security Assessment Tool (VSAT), both of which are components of the Dynamic Security Analysis Tool (DSA$^{\text{Tools}}$). DSA$^{\text{Tools}}$ is an advanced package for power system security evaluation and is developed by the Powertech Labs, Canada [35]. The decision trees are trained and tested using a commercial data mining package, Classification and Regression Trees (CART), which is developed by Salford System, CA [36]. The database generation, data conversion and analysis work are conducted using MATLAB and Microsoft Visual Studio VC++ codes.

3.2 Operating Condition 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 of these OCs are depicted in Figure 3-3. The voltages of all the buses in the AEP system are within reasonable levels at the base case:

(a) All the bus voltage magnitudes are adjusted to be between 0.90 and 1.10 in p.u.

(b) The voltages of 138~765kV buses are adjusted in the range of 0.95~1.10 p.u.

Table 3-1: PMU Locations in the Eastern AEP System

No.	Bus Name	Bus No.
1	Rockport 765 kV	4974
2	Kanawha River 345 kV	195
3	Jacksons Ferry 765 kV	1409
4	Matt Funk	1440
5	Maliszewski 765 kV	2221
6	Cook 345 kV	5599
7	Olive 345 kV	5872
8	Dumont 345 kV	5648
9	Dumont 765 kV	5649
10	Dequine 345 kV	4808
11	Twin Branch 345 kV	5983
12	Jackson Road 345 kV	5746
13	Meridian 345 kV	4913
14	Sorenson 345 kV	5407
15	Desoto 345 kV	4811
16	Breed 345 kV	4767
17	Jefferson 765 kV	4872
18	Sw Lima 345 kV	4186
19	Marysville 345	3900
20	Marysville 765 kV	3901
21	Hayden 345 kV	4491
22	Hyatt 345 kV	4513
23	Hanging Rock 765 kV	2116
24	Baker 765 kV	539
25	Culloden 765 kV	129
26	Amos 765 kV	35
27	Cloverdale 765 kV	1277

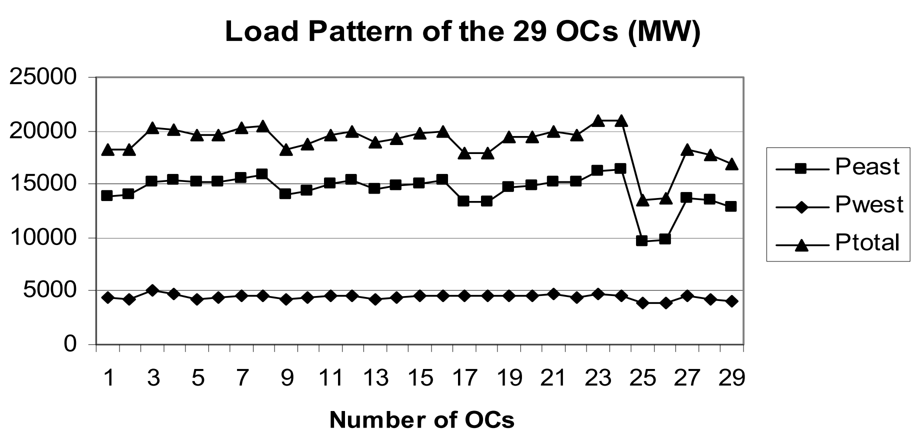

Figure 3-3: Load Pattern of the 29 Stressed OCs

3.3 Voltage Security Analysis

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 research project 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 [9], [13]. Thus, the database consists of 29*163=4727 voltage security simulations with either security or insecurity labels marked on each case.

3.4 Predictor Selection & 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 contingency-independent parameters in the system. 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 contingency-independent 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, 8 groups of predictors are used as defined in Table 3-2.

Table 3-2: 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

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 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 Group 1 and Group 2 help to compare the DT performance between contingency-dependent and contingency-independent parameters. Group 3 evaluates the DT performance using the combination of these two types of predictors. Group 4 and Group 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 contingency-independent 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 especially after contingencies; 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, 8 databases are generated in MATLAB codes for the 8 groups of predictors respectively. One sample of the created databases is shown in Table 3-3. The first column is the voltage security label obtained from VSAT results (S or I), the second and third columns represent the faulted bus (FB) and the other bus (OB) respectively, and the remaining columns are different types of measured values from PMUs. The '$' sign in the first three columns is used to differentiate the categorical variables from the numerical ones. Each row in the database represents one voltage security simulation case following one critical contingency. The pre-disturbance system parameters for each OC are the same from case to case, and the only difference rests with the contingency bus pairs, FB and OB. From the simulation results in VSAT, the database contains 3258 cases out of 9454 cases (34.46%) that fail to obtain a convergent power flow solution following a single branch outage. They are regarded as "I" cases leading to voltage collapses. The remaining cases are labeled as "S" cases.

3.5 DT Training and Performance

All the 9454 cases in the created database are given equal weight 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 [27]. Therefore, 8 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 listed in Table 3-4.

Table 3-3: Sample of the Created Databases

SECU$	FB$	OB$	Q_35_129	I_4513_4515	A_35_129	…
S	1998	2266	-398.5	13.4	0.9	
S	2266	1998	-398.5	13.4	0.9	
I	5983	9319	-398.5	13.4	0.9	
I	9319	5983	-398.5	13.4	0.9	**for OC1**
S	1205	1207	-398.5	13.4	0.9	
S	1207	1205	-398.5	13.4	0.9	
…	…	…	…	…	…	
I	1998	2266	-410	15	1.2	…
I	2266	1998	-410	15	1.2	
S	5983	9319	-410	15	1.2	
I	9319	5983	-410	15	1.2	**for OC2**
S	1205	1207	-410	15	1.2	
S	1207	1205	-410	15	1.2	
…	…	…	…	…	…	…

Table 3-4: Performance Comparison of the Optimal DTs

Opt. DTs	Size	Leaning Set Accuracy (%)			Test Set 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

From Table 3-4, the comparison of DT1 and DT2 indicates that contingency-independent 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 contingency-independent parameters alone as predictors such as I_x_y, Q_x_y, V2_x and IZ_x_y, but improvement in prediction accuracy was not observed. Therefore, the remaining predictor groups in Table 3-2 (Group 3 to Group 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 contingency-independent parameters are helpful for building much better DTs compared to DT2.

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, V2_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 on 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 performance of the whole tree.

To illustrate the efficiency of the DTs trained, Table 3-5 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. The important system parameters can be used to build an on line nomogram in identifying voltage security boundary and also provide an effective measure for the operators to steer the system conditions away from the security limits after severe disturbances occur. In addition, the critical system attributes that are selected in these DTs are listed in Appendix A and the structures of the optimal DTs trained in Table 3-4 are depicted in Appendix B.

Table 3-5: Number of Critical System Attributes 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

3.6 DT Performance Improvement

The results in Section 3.5 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 reasons, among which the 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 CSRs and a preliminary test is used to illustrate this phenomenon. Focusing on one CSR that has a severe misclassification problem in a DT trained by using A_35_539, FB and OB, it is found that most of the values of this parameter in the 29 OCs are concentrated in a small range, as shown in Figure 3-4. The horizontal axis represents the number of different operating conditions, while the vertical axis stands for the phase angle difference between bus 35 and bus 539 (in degrees). The values of A_35_539 range from 0.2 to 1.3 and most of them are concentrated in the range from 1.0 to 1.3. Only five OCs cover the range from 0.2 to 0.4. The values from 0.4 to 1.0 are totally missing. The distribution of these values shows a severe imbalance, which could be an important factor causing the misclassifications. As a result, more OCs were created for the values of A_35_539 to be distributed more evenly in this range. To accomplish this, generator active power outputs around bus 35 and bus 539 and the load levels in the AEP system are adjusted. This can effectively change the values of A_35_539 in the desired pattern. 22 new OCs are then created with the new distribution of A_35_539 shown in Figure 3-5. Another database including 22*163*2=7172 cases is created for the DT training, in which there are 288 insecure cases (4%). We can also observe that A_35_539 becomes a very important CSR in the new DT, which is located in the second level. This test indicates that by adjusting the previously problematic variable to cover more information near the splitting boundary, it can yields a better DT.

However, the proposed scheme 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, several methods to improve DT performance for online applications are discussed below when the future OCs for the next hour are relatively fixed.

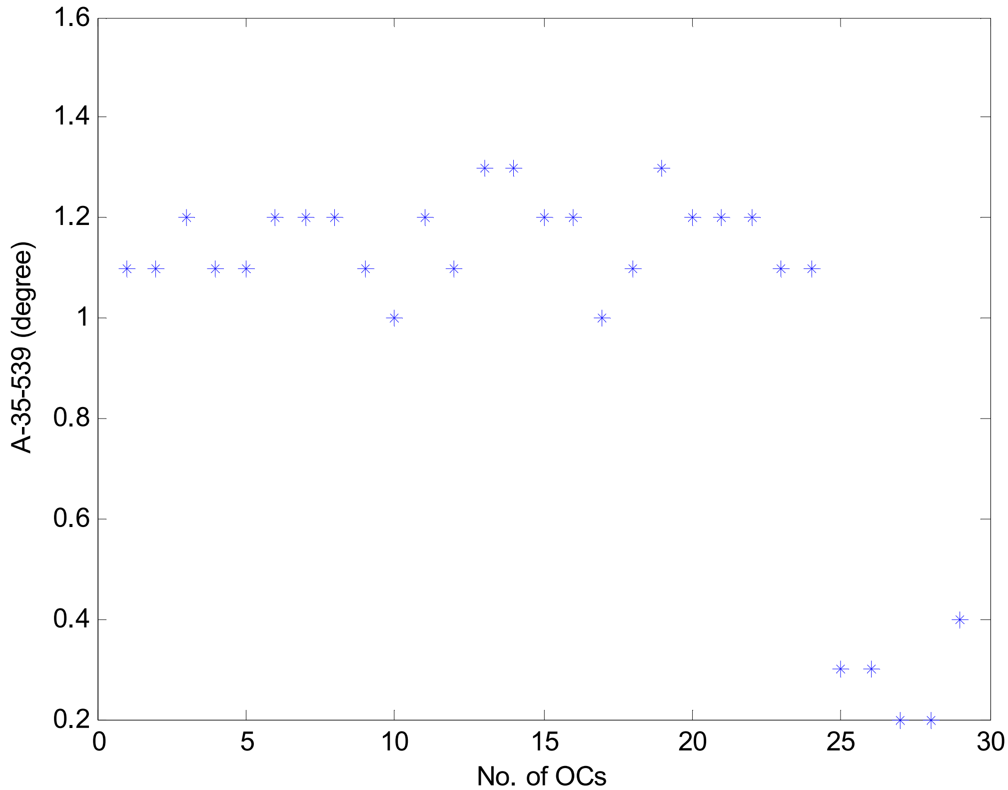

Figure 3-4: Distribution of A_35_539 for the original 29 OCs

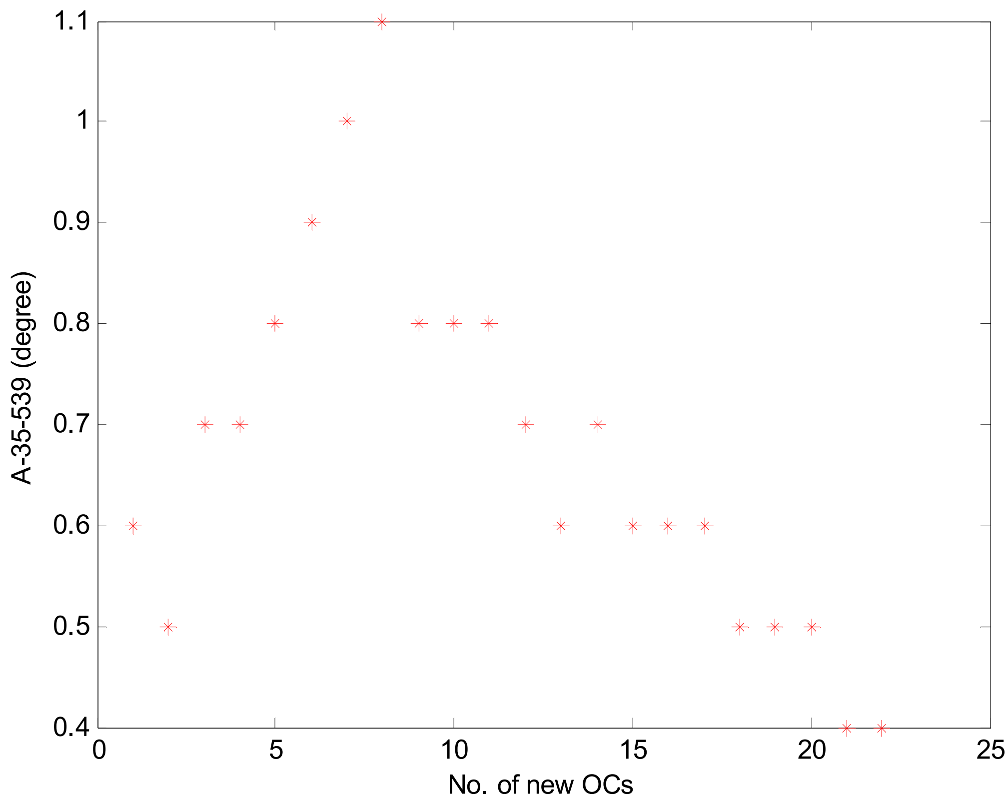

Figure 3-5: Distribution of A_35_539 for the new 22 OCs

3.6.1 Multiple Optimal DTs

As discussed in Section 3.5, 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. 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 Figure 3-6. For any incoming case, using multiple DTs can obtain different prediction results. As an example, although DT4 provides an insecure result, all the other DTs provide secure results. Therefore, a comprehensive result, Secure, is provided, which is considered as the final prediction result.

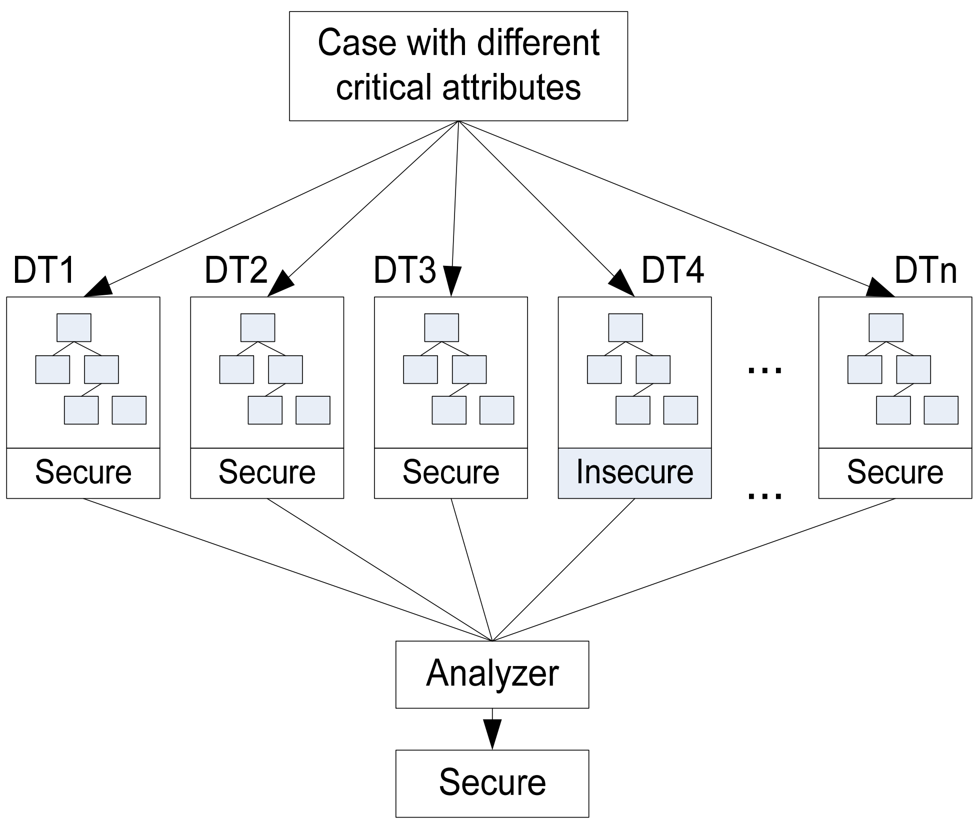

Figure 3-6: Concept of Multiple DT Application

Based on the predictor groups introduced in Section 3.4, 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. For 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. 9 DTs with good performance on the TS are obtained and shown in Table 3-6.

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 correctively 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 8 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.

Table 3-6: 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

3.6.2 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 contingency-independent 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 Figure 3-7(a).

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 6 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 5 cases in the LS' and 3 cases in the TS'. This small DT (depicted in Fig. 5(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 contingency-independent predictors and the improved DT performance are shown in Table 3-7.

Table 3-7: Performance of the 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

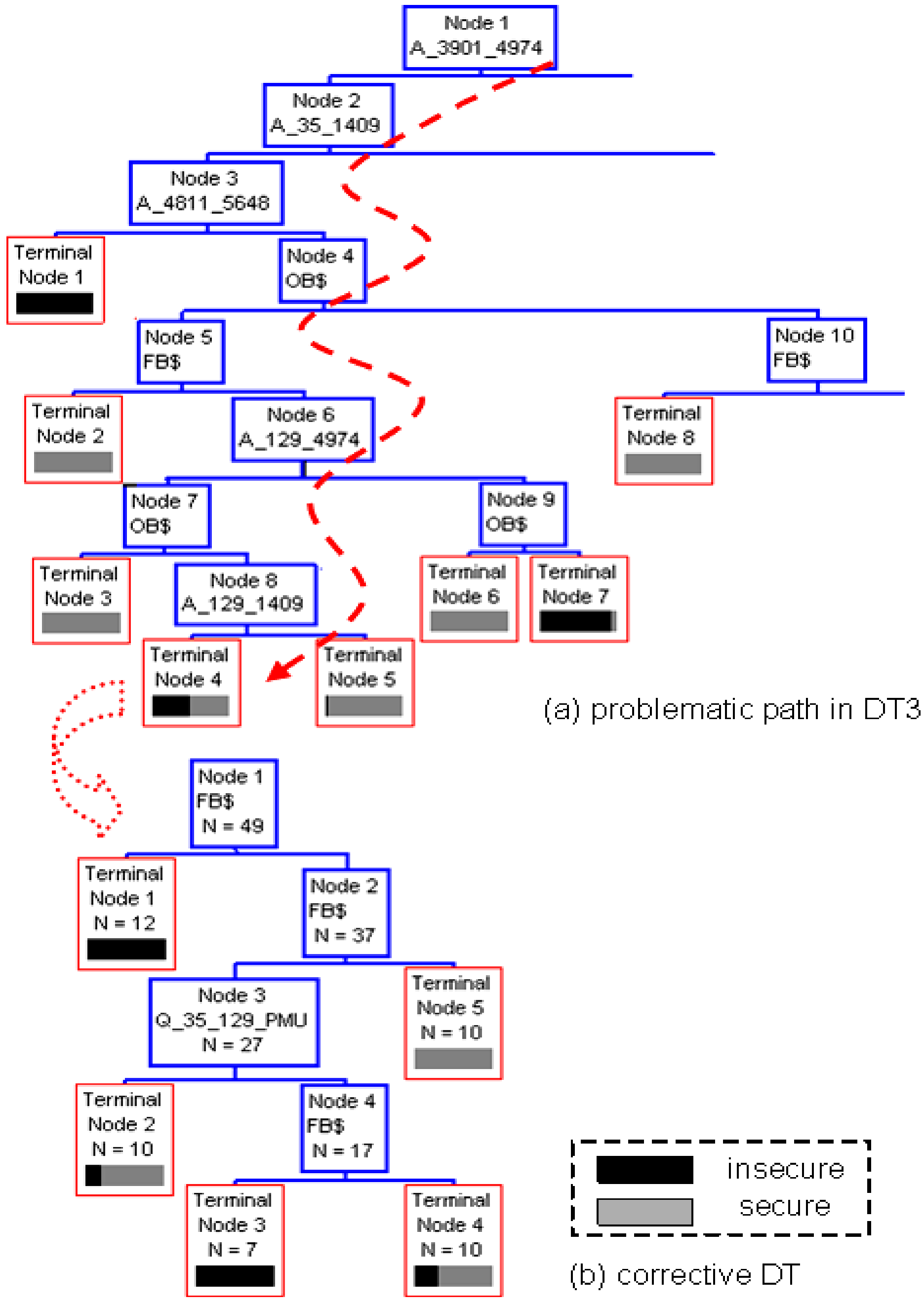

Figure 3-7: Corrective DT for a Problematic Path in DT3

From Table 3-7, all of the original DTs are improved with better performance for both learning set and test set. The best performance is achieved in DT4' and it can correctively 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: (a) there are not enough cases falling into the problematic path to train a corrective DT; and (b) 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.

3.6.3 Maximum DTs

The third measure for increasing DT accuracy is by using the maximum trees with high accuracy for the whole database. Critical system states for the next hour can be obtained an hour ahead with the help of load forecasting techniques and timely updated system information. In this situation, almost all the incoming OCs for the next hour with sufficient accuracy can be included in the database and the DT performance on the whole database becomes more important to system operators. Therefore, the trained maximum DTs with high accuracy on the whole database are a good choice for online applications in the industry.

To obtain a more reliable DT model, the "V-fold Cross Validation" method described in [27] is used to form the LS and TS. This method divides the whole database into n groups with equal number of cases, uses each group once as the test set while the remaining n-1 groups form the LS, and finally obtains a comprehensive DT based on the n generated DTs. Several tests were conducted to train the maximum DTs for the databases created in Section IV_C by setting n=10. The results in Table 3-8 indicate that any single DT trained from predictor Group 3 to Group 8 has an overall accuracy higher than 99.08%. More importantly, the overall prediction behaviors of these maximum DTs on the TS are around the level of 90%, which indicates that using the maximum trees for online application does not lose the prediction ability on the unknown cases. Higher accuracy on the whole database can be further achieved by setting the whole database as one learning set without any test set, but the corresponding maximum DT may have poor performance when system condition changes. The main problem of this maximum tree method is the high degree of complexity; however there is no need to install additional PMU devices because all the predictors are collected from the existing and planned PMUs.

Table 3-8: Maximum DT Performance on the Whole Database

Max. DTs	Size	Accuracy for the whole database (%)		
		I	S	Overall
DT3_M	117	99.91	98.72	99.13
DT4_M	125	99.85	98.68	99.08
DT5_M	112	99.85	99.11	99.37
DT6_M	142	99.97	98.66	99.11
DT7_M	126	99.85	99.05	99.32
DT8_M	125	99.91	99.11	99.39

4 Conclusions

This report presents a scheme for online voltage security assessment using synchronized phasor measurements 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. 29 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 8 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. Several new ideas including using “Multiple optimal DTs”, “Corrective DTs” and “Maximum DTs” are also introduced and tested to improve performance and reliability. In the future work, these methods will be further extended in more detail and other DT-combined data mining tools will also be tested without losing the readability during the voltage security assessment process.

5 References

5.1 Paper Resulted from this Project

Ruisheng Diao, Kai Sun, Vijay Vittal, Robert O'Keefe, Michael Richardson, Navin Bhatt, Dwayne Stradford and Sanjoy Sarawgi, “Decision Tree Based Online Voltage Security Assessment Using PMU Measurements,” Paper submitted to *IEEE Transactions on Power Systems*, August 2008.

5.2 Background Papers

[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] Stewart E. Stanton, “Transient stability monitoring for electric power systems using a partial energy function,” *IEEE Trans. Power Syst.*, vol. 4, no. 4, pp. 1389-1396, Nov. 1989.

[3] D. R. Ostojic and G. T. Heydt, “Transient stability assessment by pattern recognition in the frequency domain,” *IEEE Trans. Power Syst.,* vol. 6, no. 1, pp. 231-237, Feb. 1991.

[4] M. A. Pai, P. W. Sauer and F. Dobraca, “A new approach to transient stability evaluation in power systems,**”** *Proceedings of the 27th IEEE Conference on Decision and Control*, vol. 1, pp. 676-680, Dec. 1988.

[5] L. S. Moulin, A. P. Alves da Silva, M. A. El-Sharkawi and R. J. Marks II, “Support vector machines for transient stability analysis of large-scale power systems,” *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 818-825, May 2004.

[6] 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.

[7] S. Rovnyak, S. Kretsinger, J. Thorp, D. Brown, “Decision Trees for Real-time Transient Stability Prediction,” *IEEE Trans. Power Syst.*, vol. 9, no. 3, pp.1417-1426, Aug. 1994.

[8] Nima Amjady and Seyed Farough Majedi, “Transient Stability Prediction by a Hybrid Intelligent System,” *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp.1275-1283, Aug. 2007.

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

[10] Haibo You, Vijay Vittal and Xiaoming Wang, “Slow Coherency-Based Islanding,” *IEEE Trans. Power Syst.,* vol. 19, no. 1, pp. 483-491, Feb. 2004.

[11] Haibo You and Vijay Vittal, “Self-Healing in Power Systems: An Approach Using Islanding and Rate of Frequency Decline-Based Load Shedding,” *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 174-181, Feb. 2003.

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

[13] Thierry Van Cutsem and Costas Vournas, Voltage Stability of Electric Power Systems, Norwell, Massachusetts: Kluwer Academic Publishers, 1998.

[14] Muhammad Randhawa, Voltage Stability Assessment of the Entergy Electrical System, M.S. Thesis, Arizona State University, Tempe, AZ, 2007.
[15] P-A. Lof, G. Andersson and D. J. Hill, "Voltage stability indices for stressed power systems," *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 326-335, Feb. 1993.
[16] P. Kessel and H.Glavitsch "Estimating the Voltage Stability of a Power System," *IEEE Trans. Power Delivery*, vol.PWRD-1, no. 3, pp. 346-354, July 1986.
[17] A. R. Phadke, S. K. Bansal and K. R. Niazi, "A Comparison of Voltage Stability Indices for Placing Shunt FACTS Controllers," *Emerging Trends in Engineering and Technology, 2008. ICETET '08*, pp. 939-944, July 2008.
[18] Y. Gong, N. Schulz and A. Guzman, "Synchrophasor-Based Real-Time Voltage Stability Index," *Power Systems Conference and Exposition, 2006*, pp. 1029-1036, Oct. 2006.
[19] A. G. Phadke, "Synchronized Phasor Measurements in Power Systems," *IEEE Computer Applications in Power*, vol. 6, no. 2, pp. 10-15, April 1993.
[20] A. P. Sakis Meliopoulos, George J. Cokkinides, Floyd Galvan, Bruce Fardanesh and Paul Myrda, "Delivering Accurate and Timely Data to All – Model-Based Substation Automation Applications for Advanced Data Availability," IEEE Power & Energy Magazine, pp. 74-86, May/June 2007.
[21] 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.
[22] 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.
[23] J. W. Ballance, B. Bhargava, G. D. Rodriguez, "Monitoring Power System Dynamics using Phasor Measurement Technology for Power System Dynamic Security Assessment," *in Proc. IEEE Power Tech Conference*, Bologna, Italy, Jun. 2003.
[24] B. Milosevic, 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.
[25] A. R. Khatib, R. F. Nuqui, M. R. Ingram, 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.
[26] L. Breiman, J. Friedman, R. A. Olshen and C. J. Stone, Classification and Regression Trees, Wadsworth International Group, 1984
[27] Salford System, "CART for Windows User's Guide-An Implementation of the Original CART Methodology," software manual, 2002.
[28] L. Wehenkel, M. Pavella, E. Euxibie, et al, "Decision Tree Based Transient Stability Method A Case Study," *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp.459-469, Feb. 1994.
[29] L. Wehenkel, M. Pavella, "Decision Trees and Transient Stability of Electric Power Systems," *Automatica*, vol. 27, no. 1, pp.115-134, Jan., 1991.
[30] T. Cutsem, L. Wehenkel, M. Pavella, B. Heilbronn, M. Goubin, "Decision Tree Approaches to Voltage Security Assessment," *IEE Proceedings*, vol. 140, no. 3, pp.189-198, May 1993.

[31] S. Rovnyak, C. Taylor, Y. Sheng, "Decision Trees Using Apparent Resistance to Detect Impending Loss of Synchronism," *IEEE Trans. Power Syst.*, vol. 15, no. 4, pp 1157-11162, Oct. 2000.
[32] K. R. Padiyar and S. Krishna, "On-Line Detection of Loss of Synchronism Using Locally Measurable Quantities," *Transmission and Distribution Conference and Exhibition 2001*, Vol. 1, pp. 537-542.
[33] 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.
[34] 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.
[35] DSATools Dynamic Security Assessment Software. [Online]. Available: http://www.dsatools.com/
[36] Salford System, "CART," [Online]. Available: http://www.salford-systems.com/cart.php

Appendix A. Critical System Attributes Selected in the Opt. DTs

This Appendix lists all the contingency-independent critical system attributes that are selected in the optimal DTs in Section 3.5.

Table A. 1 Contingency-independent CSRs in DT3

No.	Bus1	Bus2	Bus1 Name	Bus2 Name	Predictor
1	3901	4974	MARYSVIL 345.	ROCKPORT 765.	A_x_y
2	4811	5648	DESOTO 345.	DUMONT 765.	A_x_y
3	35	1409	AMOS 765.	JACKSONS 765.	A_x_y
4	4186	4513	SWLIMA 345.	HYATTCS 345.	A_x_y
5	1409	4186	JACKSONS 765.	SWLIMA 345.	A_x_y
6	195	2116	KANAWHAR 345.	HANGINGR 765.	A_x_y
7	2221	4767	ORANGE 765.	BREED 345.	A_x_y
8	35	1277	AMOS 765.	CLOVERDA 765.	A_x_y
9	129	4974	CULLODEN 765.	ROCKPORT 765.	A_x_y
10	1440	4767	MATTFUNK 138.	BREED 345.	A_x_y
11	129	1409	CULLODEN 765.	JACKSONS 765.	A_x_y
12	35	2221	AMOS 765.	ORANGE 765.	A_x_y
13	4974	5983	ROCKPORT 765.	TWINBRAN 345.	A_x_y

Table A. 2 Contingency-independent CSRs in DT4

No.	Bus1	Bus2	Bus1 Name	Bus2 Name	Predictor
1	5648	8112	DUMONT 765.	WILTONCE 765.	I_x_y
2	1440	1586	MATTFUNK 138.	SHAWSVIL 138.	I_x_y
3	1097	1409	WYOMING 765.	JACKSONS 765.	I_x_y
4	35	129	AMOS 765.	CULLODEN 765.	I_x_y
5	4872	4974	JEFFERSO 765.	ROCKPORT 765.	I_x_y
6	2116	2154	HANGINGR 765.	MARQUIS 765.	I_x_y
7	5599	9473	COOK 345.	PALISADE 345.	I_x_y
8	129	1097	CULLODEN 765.	WYOMING 765.	I_x_y
9	2221	2672	ORANGE 765.	KAMMER 765.	I_x_y
10	1989	2116	NPROCTOR 765.	HANGINGR 765.	I_x_y
11	35	605	AMOS 765.	MOUNTAIN 765.	I_x_y
12	195	622	KANAWHAR 345.	SPORN 345.	I_x_y
13	129	2108	CULLODEN 765.	GAVIN 765.	I_x_y
14	4974	5012	ROCKPORT 765.	SULLIVAN 765.	I_x_y

Appendix A. Critical System Attributes Selected in the Opt. DTs

Table A. 3 Contingency-independent CSRs in DT5

No.	Bus1	Bus2	Bus1 Name	Bus2 Name	Predictor
1	2221	3900	ORANGE 765.	MARYSVIL 765.	Q_x_y
2	539	2116	BAKER 765.	HANGINGR 765.	Q_x_y
3	4767	7791	BREED 345.	WCASEY 345.	Q_x_y
4	35	605	AMOS 765.	MOUNTAIN 765.	Q_x_y
5	5746	5983	JACKSONR 345.	TWINBRAN 345.	Q_x_y
6	4290	4491	BEATTY 345.	HAYDEN 345.	Q_x_y
7	1277	1414	CLOVERDA 765.	JOSHUAFA 765.	Q_x_y
8	4767	12718	BREED 345.	WHEATLNA 345.	Q_x_y
9	35	129	AMOS 765.	CULLODEN 765.	Q_x_y
10	35	1989	AMOS 765.	NPROCTOR 765.	Q_x_y

Table A. 4 Contingency-independent CSRs in DT6

No.	Bus	Bus Name	Predictor
1	195	KANAWHAR 345.	V^2_x
2	129	CULLODEN 765.	V^2_x
3	1277	CLOVERDA 765.	V^2_x
4	4767	BREED 345.	V^2_x
5	5983	TWINBRAN 345.	V^2_x
6	35	AMOS 765.	V^2_x
7	1440	MATTFUNK 138.	V^2_x
8	1409	JACKSONS 765.	V^2_x
9	4808	DEQUINE 345.	V^2_x
10	4872	JEFFERSO 765.	V^2_x
11	2221	ORANGE 765.	V^2_x

Appendix A. Critical System Attributes Selected in the Opt. DTs

Table A. 5 Contingency-independent CSRs in DT7

No.	Bus1	Bus2	Bus1 Name	Bus2 Name	Predictor
1	5872	14830	OLIVE 345.	GREENACR 345.	IZ_x_y
2	2116	4872	HANGINGR 765.	JEFFERSO 765.	IZ_x_y
3	1989	2116	NPROCTOR 765.	HANGINGR 765.	IZ_x_y
4	1440	1676	MATTFUNK 138.	WSALEM 138.	IZ_x_y
5	35	129	AMOS 765.	CULLODEN 765.	IZ_x_y
6	2116	7132	HANGINGR 765.	CORNU 765.	IZ_x_y
7	4872	4974	JEFFERSO 765.	ROCKPORT 765.	IZ_x_y
8	4513	4515	HYATTCS 345.	HYATTCS 345.	IZ_x_y
9	36	195	AMOS 345.	KANAWHAR 345.	IZ_x_y
10	4767	12718	BREED 345.	WHEATLNA 345.	IZ_x_y
11	5599	9473	COOK 345.	PALISADE 345.	IZ_x_y
12	35	1989	AMOS 765.	NPROCTOR 765.	IZ_x_y
13	3901	4518	MARYSVIL 345.	HYATTOP 345.	IZ_x_y
14	195	622	KANAWHAR 345.	SPORN 345.	IZ_x_y
15	651	1409	BROADFOR 765.	JACKSONS 765.	IZ_x_y

Table A. 6 Contingency-independent CSRs in DT8

No.	Bus1	Bus2	Bus1 Name	Bus2 Name	Predictor
1	5872	14830	OLIVE 345.	GREENACR 345.	IZ_x_y
2	2116	4767	HANGINGR 765.	BREED 345.	A_x_y
3	539	2116	BAKER 765.	HANGINGR 765.	Q_x_y
4	35	605	AMOS 765.	MOUNTAIN 765.	Q_x_y
5	4872	4974	JEFFERSO 765.	ROCKPORT 765.	I_x_y
6	1376	1440	HANCOCK 138.	MATTFUNK 138.	Q_x_y
7	5648	8561	DUMONT 765.	GREENTOW 765.	I_x_y
8	35	129	AMOS 765.	CULLODEN 765.	Q_x_y
9	195	4974	KANAWHAR 345.	ROCKPORT 765.	A_x_y
10	5599	9473	COOK 345.	PALISADE 345.	I_x_y
11	35	129	AMOS 765.	CULLODEN 765.	I_x_y
12	129	1097	CULLODEN 765.	WYOMING 765.	Q_x_y
13	2116	2154	HANGINGR 765.	MARQUIS 765.	Q_x_y
14	4290	4491	BEATTY 345.	HAYDEN 345.	Q_x_y
15	4513	4514	HYATTCS 345.	HYATTCS 345.	Q_x_y
16	3901	4518	MARYSVIL 345.	HYATTOP 345.	I_x_y

Appendix B. Structure of Optimal DTs in Section 3.5

This appendix depicts all the CSRs in the structure of the optimal DTs that are trained in Section 3.5. Due to space limitation, the thresholds of the CSRs and details of each node are not shown. The blue nodes denote stable terminal nodes and the red ones represent the insecure nodes.

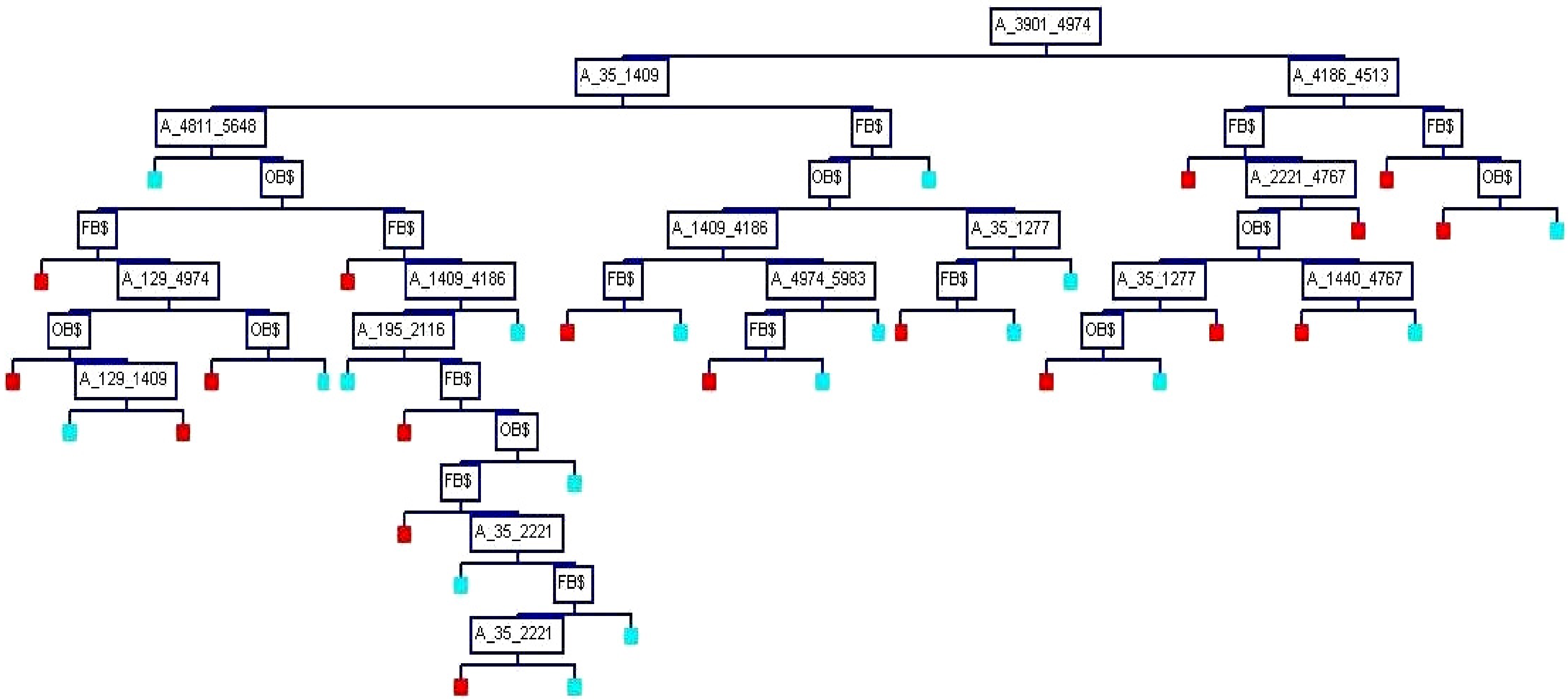

Figure B. 1 Structure of DT3 with critical splitting rules

Appendix B. Structure of Optimal DTs in Section 3.5

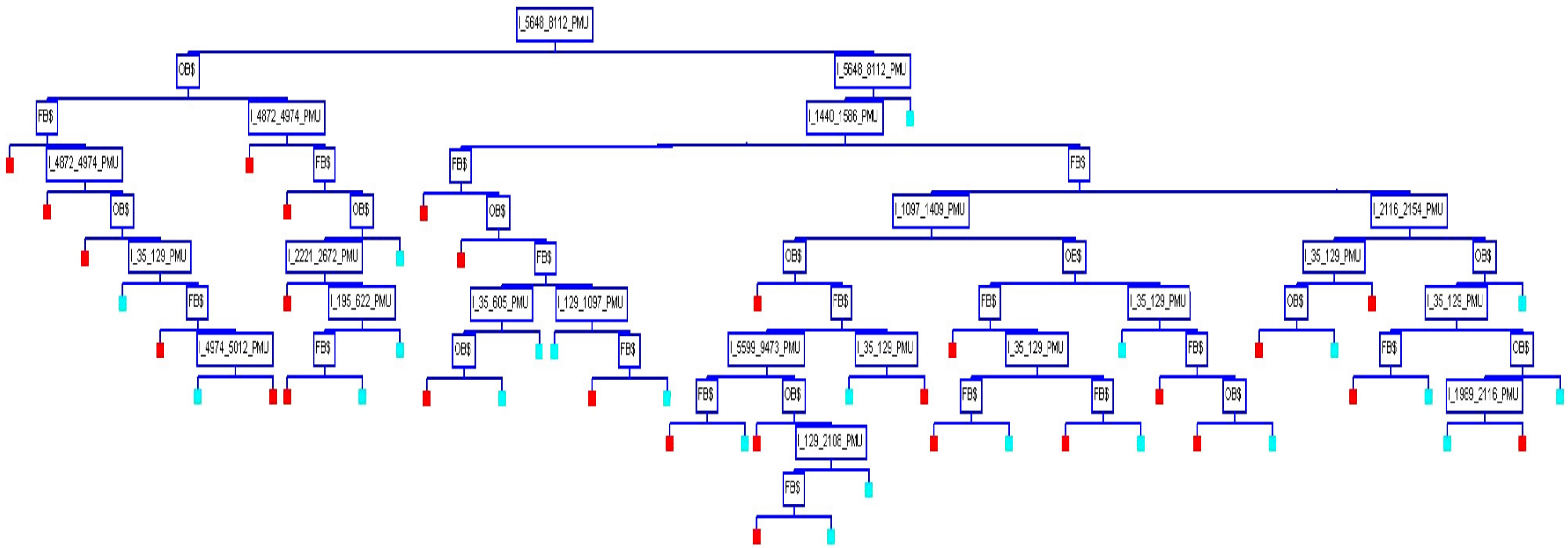

Figure B. 2 Structure of DT4 with critical splitting rules

Appendix B. Structure of Optimal DTs in Section 3.5

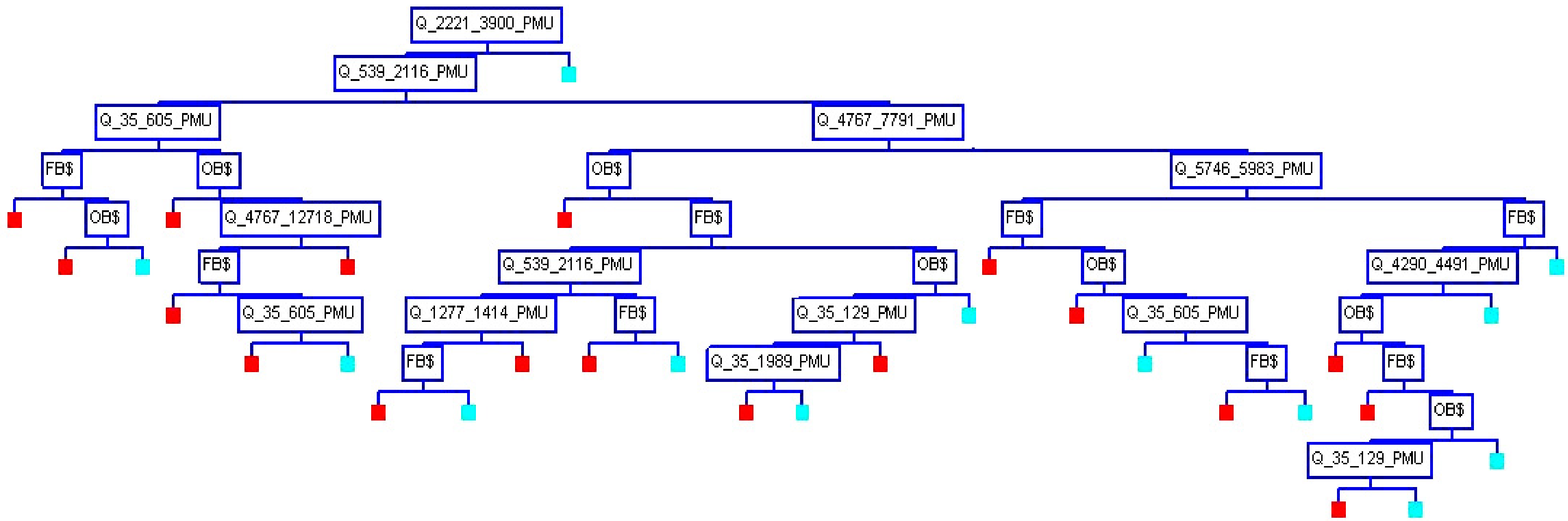

Figure B. 3 Structure of DT5 with critical splitting rules

Appendix B. Structure of Optimal DTs in Section 3.5

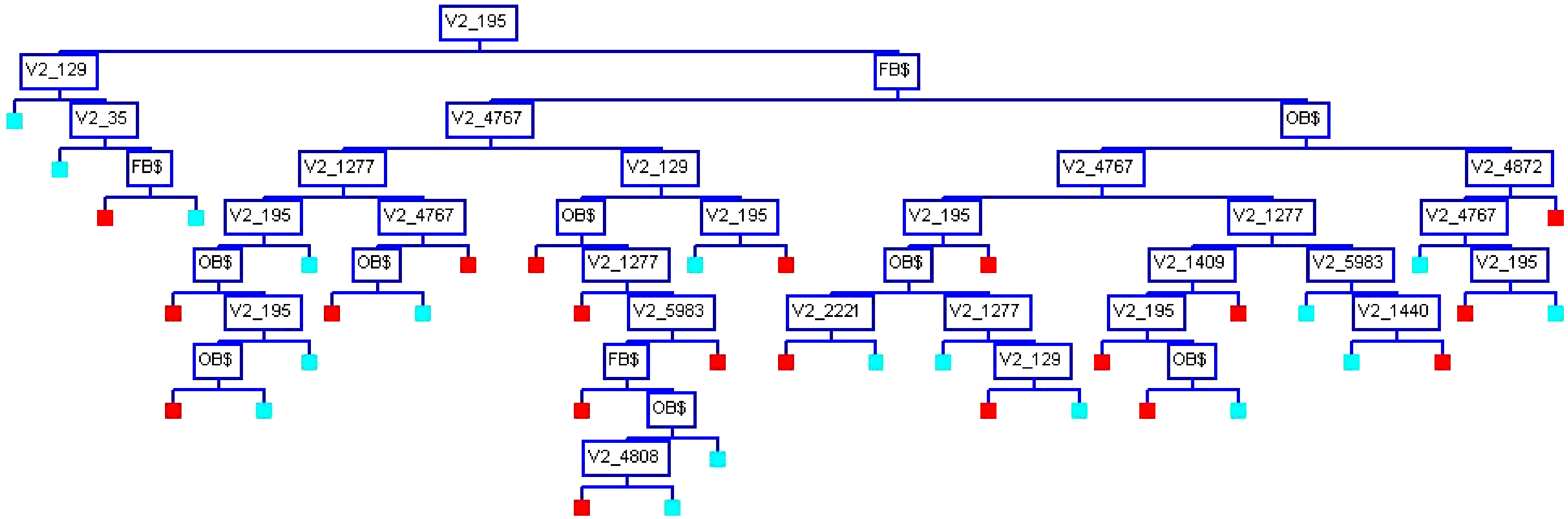

Figure B. 4 Structure of DT6 with critical splitting rules

Appendix B. Structure of Optimal DTs in Section 3.5

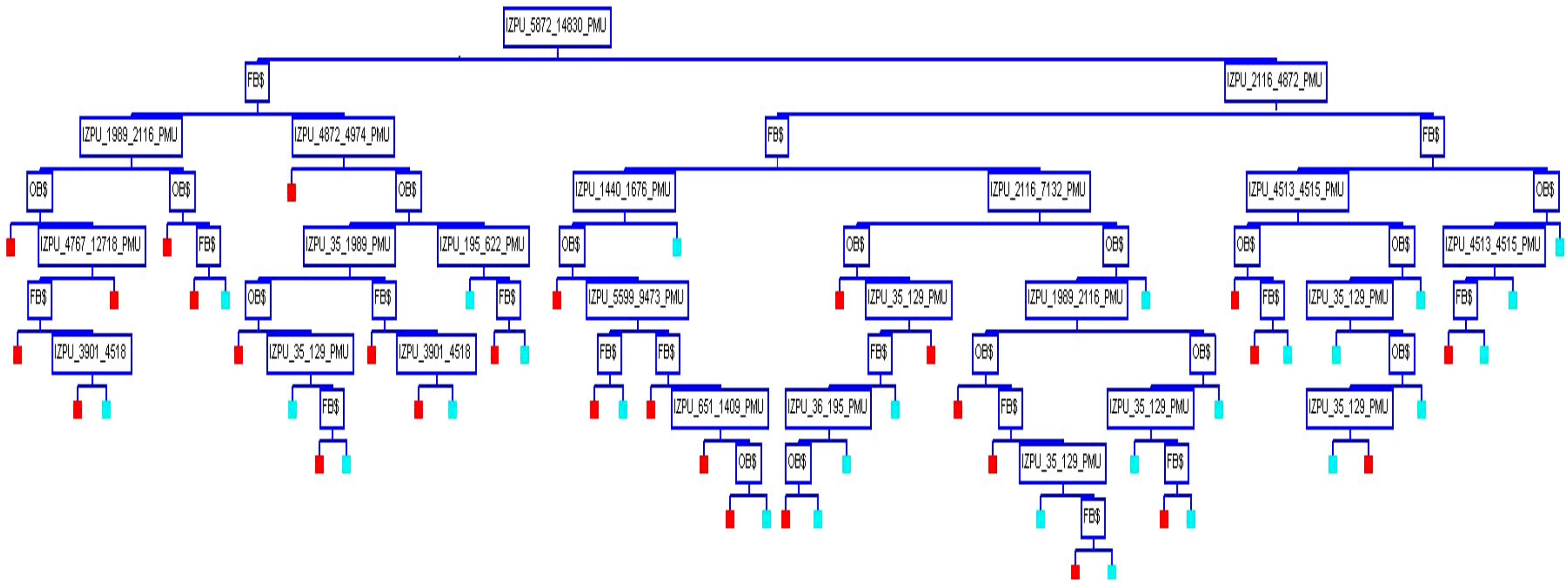

Figure B. 5 Structure of DT7 with critical splitting rules

Appendix B. Structure of Optimal DTs in Section 3.5

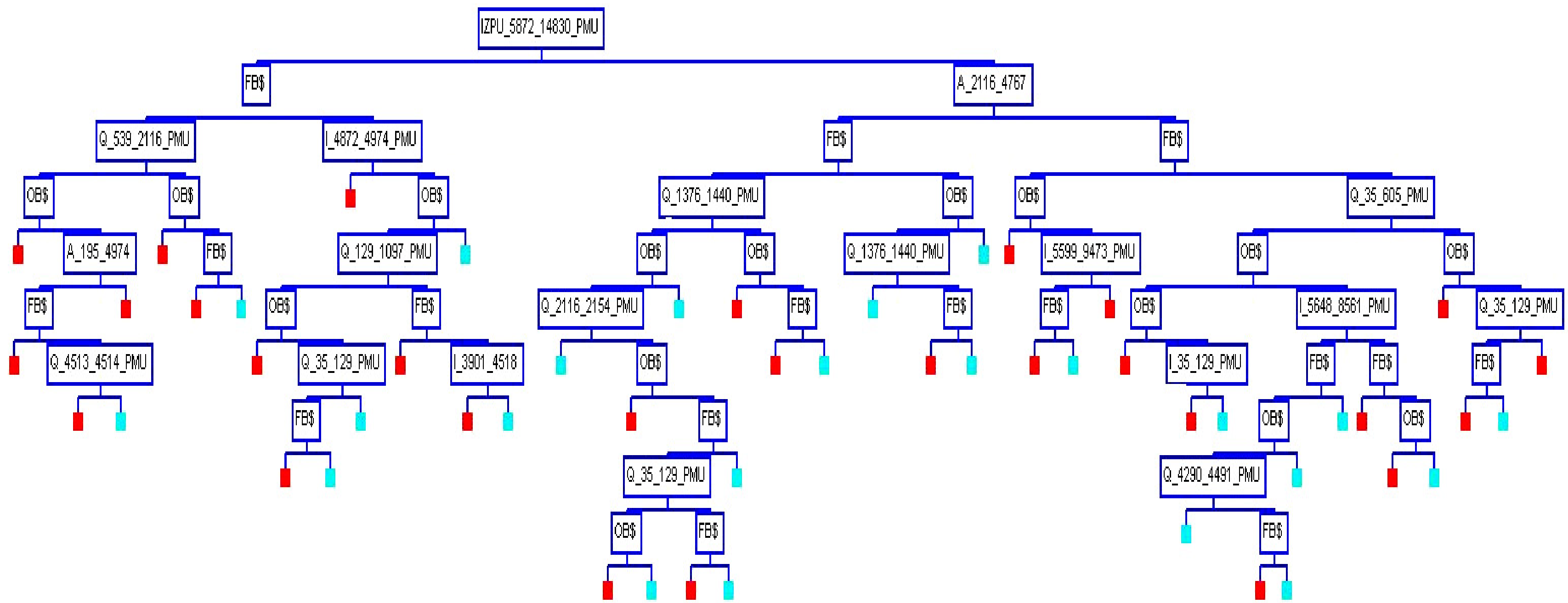

Figure B. 6 Structure of DT8 with critical splitting rules