Decision Tree Based Filter for Control Area External Contingencies Screening
Khaled Alcheikh-Hamoud, Nouredine Hadjsaid, Yvon Besanger, Jean-Pierre Rognon

To cite this version:

HAL Id: hal-00403410
https://hal.archives-ouvertes.fr/hal-00403410
Submitted on 10 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Decision Tree Based Filter for Control Area External Contingencies Screening

K. Alcheikh-Hamoud, N. Hadjsaid, Senior Member, IEEE, Y. Bésanger, Senior Member, IEEE and J.P. Rognon,

Abstract— The security of modern large interconnected power systems suffers from the absence of a unique security coordinator for the entire system. The presence of a security coordinator for such systems allows the detection of both internal and external harmful contingencies of each control area under his responsibility. The high economic costs and the technical and administrative complexities among others, makes it a very difficult task to create a security coordinator for such very large interconnected power systems. In the absence of a full information exchange, the operators’ alternative solution to the existence of a security coordinator is the on-line exchange of their post-contingency boundary buses power injections resulting from the analysis of their internal contingencies. In fact, this exchange enables the operators to assess the impact of the external contingencies on their own control areas. In this context, the operators’ major problem is how they can assess on-line, in relatively short time, the impact of a large number of external contingencies on their systems. This paper presents an innovative approach for a rapid screening of control area external contingencies allowing only the dangerous ones to be assessed in detail by the operators.

Index Terms— Security coordinator, decision tree based filter, control area external contingencies screening, control areas data exchange, multi-area electric power system security assessment.

I. INTRODUCTION

The security of electric power systems has been negatively affected by electric utilities deregulation and restructuring. Indeed, the competition imposed by the rules of liberalized market has pushed both the independent power producers (IPPs) and the independent system operators (ISOs) to exploit their systems in stressed operational conditions namely closer and closer to their stability limits. That is why the deregulation and the restructuring have become the main challenge for the security of today electric power systems [1],[2].

Another challenge for power system security is the horizontal expansion of this system by the interconnection of its high voltage transmission network with other networks of adjacent control areas. This integrated system constitutes in fact a multi-area electric power system. Actually, on the one hand, such interconnection, in addition to traditional benefits such as mutual help in case of emergency, is very useful for consolidating the competition and the rules of liberalized markets by the creation of larger marketplaces for the sake of market fluidity. On the other hand, the security of the individual control areas as well as the whole interconnected system can be seriously threatened [13],[14] especially when the energy exchanges (trading) between the systems are significant. The security is particularly affected in the case of synchronous interconnection by the propagation of the bad effects of some contingencies occurring in some control areas to the other control areas.

As a result, electric power systems have become more vulnerable to contingencies than before, and today it is possible that any contingency can lead the power system from a secure state to an insecure or an emergency state.

Usually, in the absence of a security coordinator of a multi-area electric power system, on-line security assessment of a multi-area system is done in a distributed and relatively independent way. The operator of each control area analyzes the security of his own electric system with some technical considerations for the borders (interconnections and some added information on the cluster and vicinity of the interconnection). He declares that his system is in a secure state if the standard security rule (N-1) of his area is satisfied regardless the possibility of occurrence of any external contingency that can be dangerous for his area. This situation is especially the case of the UCTE (large interconnected power system in Western Europe), where the TSOs of this system have no unique security coordinator who can analyze the security of each TSO for both internal and external contingencies [7],[15] although the situation is in progress. Also, in North America Interconnected Power System, there is no unique security coordinator for the whole system [7],[16].

The latest wave of blackouts, especially in North America and Europe [11], [12] has confirmed the inadequacy of operators’ present practices for security assessment of their control areas in the absence of a security coordinator for the whole interconnected system. In fact, the real weak point in these practices resides essentially in the low level of cooperation and coordination between these areas [3], [4]. Consequently, operators do not have adequate tools to assess the impact of external contingencies on their internal control areas, thus the electric systems continue to be vulnerable to those contingencies.

An important achievement in the field of distributed on-line security assessment of a multi-area electric system is
presented in paper [10]. This paper presents a cooperative
framework between operators which is based on the on-line
exchange of their post-contingency boundary buses power
injections resulting from the on-line assessment of their
internal contingencies. This exchange allows the operator of
each control area to assess the impact on his area of the
contingencies that may occur on other (neighboring) control
areas. Thus, the vulnerability of control areas with respect to
the external contingencies will be eliminated even in the
absence of security coordinator.

However, the proposed cooperative framework fails to
include an effective and rapid way for control areas external
contingencies screening.

Recently, the deregulated electric utilities have started to
exchange detailed data, such as modeling data and market
data, of their systems for certain time horizons [8], [9]. Based
on the possibility of data exchange between the operators of a
multi-area electric power system, this paper develops an
efficient tool, decision tree based tool, for control area external
contingencies screening.

II. CLASSIFICATION DECISION TREES (CDT)

The classification decision tree (CDT) is one of supervised
learning techniques which group in addition to (CDT),
regression decision tree (RDT), artificial neural networks
(ANN), statistical pattern recognition and others. These
techniques constitute with the unsupervised learning
techniques the two main divisions of automatic learning, data
mining, methods [20]. In fact, a CDT is usually used for
extracting meaningful classification rules, representing high
level synthetic information or knowledge, from a data base
(DB), containing large amounts of low level data, of a given
problem. Then, these rules are exploited to deduce rapidly the
class of new unseen cases of the considered problem. CDT has
the hierarchical form of an upside-down tree structure and is
build on the basis of a learning set (LS). Then, it is validated
against an independent test set (TS).

A. Learning and test sets

Each of the sets, learning and test, comprises a number of
pre-classified cases of the considered data base. Normally, a
small number of cases will be randomly selected from the data
base for the test set while the remaining cases, the largest part
of data base, form the learning set. In fact, each case (C) is
represented by a vector (υ) composed of the values of a group
of parameters called candidate attributes or predictors, and by
a value representing a class of the classification parameter also
called categorical attribute (g). Mathematically, the learning
and test sets can be defined by the following equations:

\[ C_T = \{a_1 = \eta_1, a_2 = \eta_2, \ldots, a_k = \eta_k \} \]
\[ u_T = \{\eta_1, \eta_2, \ldots, \eta_k, u_k \} \}
\[ C^R = \{u_1, u_2, \ldots, u_k, g \} \}

\[ LS=\{\eta_1, \eta_2, \ldots, \eta_k, g_1, g_2, \ldots, g_k\} \]
\[ TS=\{u_1, u_2, \ldots, u_k, g_1, g_2, \ldots, g_k\} \}

where:
\[ C_T : \text{is the (T)-th case of the considered data base.} \]
\[ a_S : \text{is the (S)-th attribute of any case of the considered data} \]
\[ base. \]
\[ \eta_{ST} : \text{is the numerical value of the (S)-th attribute of the (T)-th} \]
\[ case. \]
\[ g_T : \text{is the class of the (K)-th attribute, categorical attribute, of} \]
\[ the (T)-th case. \]
\[ K : \text{is the total number of attributes of any case in the} \]
\[ considered data base. \]
\[ (\eta_E, g_E) : \text{is the (E)-th case of the learning set.} \]
\[ (\eta_M, g_M) : \text{is the (M)-th case of the test set.} \]
\[ R : \text{is the number of cases in the learning set.} \]
\[ F : \text{is the number of cases in the test set.} \]
\[ P : \text{is the total number of cases in the considered data base.} \]

B. Building and validation of CDT

CDT is built from the entire cases of a learning set
according to the following steps:

Firstly, starting at the top-node of the tree, the classification
algorithm (CA) analyses all the cases of the learning set
looking for the most perfect test (MPT) which realizes the best
score for the splitting of learning set cases (LSCs), according
to their class, into subsets. In other words, the CA tries to
group a maximum number of cases belonging to a similar
class in the same subset. Actually, the CA acts in two steps to
determine the MPT:

- For each attribute, say \( a_1 \), the CA looks for its optimal
  threshold value \( \eta_{op} \) which splits the LSCs in the best way
  according to their class.

- The CA chooses, among the different attributes, the best
  attribute, \( a_{dop} \), along with its optimal threshold value, \( \eta_{dop} \),
  to split the LSCs according to the test:

\[ T : a_{dop} < \eta_{dop} \] (2)

Secondly, considering two classes of the LSCs, say
(secure, insecure), the MPT splits the LSCs of the top-node into two subsets, the two subsets constitute the successors nodes of the top-node, according to the following equations:

\[
LS_1 = \{ \nu_E \in LS \text{ with } a_{dop} < \nu_{dop} \} \quad (3) \\
LS_2 = \{ \nu_E \in LS \text{ with } a_{dop} \geq \nu_{dop} \}
\]

Thirdly, the CA repeats, in the same way as for the top-node LSCs, the splitting procedure on the resulting subsets \((LS_1, LS_2)\) of the successors nodes. In fact, the CA continues the splitting of a subset of a successor node into new subsets, new successors nodes, until it gets a sufficiently class pure subset, leaf terminal node, or it can stop the splitting process at a successor node if there is no possibility of enhancing the CDT accuracy in a statistically significant way by splitting the node further. Such a node is called a dead-end terminal node. At last, the CA assigns a class for each terminal node according to the prevailed class of cases contained in this node. A simple 5-node CDT building procedure is illustrated in Fig. 1.

The CDT validation phase (VP) comes after the CDT building phase (BP). The VP uses the test set pre-classified cases to measure the ability of CDT to properly classify unseen cases. The VP consists of dropping the numerical vector of each case of the test set down the CDT starting at the root node until a terminal node is reached. Thus, a class is assigned to each test set case according to the class of its terminal node. Actually, at each inner (test node) the test related to this node, specified during the BP, is applied to the numerical value of the concerned attribute of dropped vector to determine which successor node the vector should drop into. Finally by comparing the assigned class of each test case with its original class a misclassification rate of each class can be calculated. Thus, the performance of CDT can be easily evaluated [19].

III. DISTRIBUTED SECURITY ASSESSMENT OF A MULTI-AREA ELECTRIC POWER SYSTEM IN THE ABSENCE OF A SECURITY COORDINATOR

Actually, in the absence of a security coordinator of a multi-area electric power system, the on-line security analysis of this system is carried out in a distributed way at the level of each control area operator [6]. In fact, the operator of each control area receives in real-time only the raw measurements of his own SCADA (Supervisory Control And Data Acquisition) system.

Fig. 1. The building procedure of a simple 5-node CDT using LSCs.

As shown in Fig. 2 the control area topology processor treats the Boolean values (breaker status) of the SCADA measurements to determine the real-time network structure. Then, the observability analysis is used to determine whether the system is observable based on the results of the topology processor and the analog electrical values (Volts, Amps, Watts and VARs) of the SCADA measurements. If the control area electric system is not observable the observability analysis will further determine the pseudo-measurements (PM) that need to be added to make the system observable. Then, the state estimator (SE) is used to obtain the control area real-time model [5]. However, the external (neighboring) control areas are usually modeled by reduced equivalent models. In fact, for on-line security assessment, the reduced equivalent models of the external control areas will be updated periodically. Actually, these models plus the real-time detailed model of the internal control area, updated at each state estimation cycle, form the overall updated model that reflect the latest operational conditions of the whole interconnected power system. At each cycle of on-line security assessment, the operator of the internal control area will use the overall updated model to analyze the security of his area with respect to his own internal contingencies and to determine the external effects of each one of these contingencies [6], [10].

Fig. 2. Control area real-time model building.
Also, in the absence of a security coordinator, it is necessary that the operator of each control area analyzes online the security of the area under his jurisdiction with respect to all potential external contingencies [10]. In order to achieve this goal, first of all, each operator of a multi-area electric power system should compute the external effects, active/reactive power injections at all boundary buses subsequent to contingency occurrence, of each one of his internal contingencies. Then, the operators must exchange online the external effects of their internal contingencies. Thus, in order to analyze the security of a control area with respect to an external contingency, the operator of this area will use the real-time detailed model of his area, plus the post-contingency boundary buses power injections computed by area of origin of this external contingency, plus the equivalent models of the other external areas [10], [6].

**IV. THE IMPORTANCE OF CONTROL AREA EXTERNAL CONTINGENCIES SCREENING**

As it has been mentioned in the previous sections, the post-contingency power injections of the boundary buses will be exchanged between operators of the interconnected control areas. In fact, in the area control center, the on-line security assessment function is carried out in a cyclic manner usually every 10-15 minutes. Thus, in order to fulfill the requirements of the on-line security assessment function, it is important that the overall time, required to assess all the contingencies of some control area, must not exceed the specified time of one cycle. In this context, the real problem that the operators have to deal with is related to the relatively short time available to them for assessing the impact, one by one, of large number of external contingencies on their control areas.

Actually, in order to resolve this problem the control areas operators must have at their disposal an efficient tool for screening external contingencies. Therefore, a classification decision tree-based filter is proposed for this purpose. The proposed filter will allow each control area operator to reduce significantly the number of external contingencies that he should assess in real-time. Actually, the filter will enable the operators to screen rapidly their external contingencies. Thus, only potentially harmful contingencies will be selected for detailed on-line security assessment.

For the sake of clarity, the Fig. 3 shows the representation of a multi-area electric power system.

![Fig. 3. Representation of a multi-area electric power system.](image)

In this system, we designate the control area (A) as the internal control area. The other control areas (B, C and D) are considered as the external ones compared to (A). The set of the whole external contingencies $C(a)$ of the control area (A) can be divided into three sets $C(b)$, $C(c)$ and $C(d)$ depending on the external control areas (B,C and D). At the same time, these sets represent respectively the internal contingencies of the control areas (B,C and D). From the principal sets $C(b)$, $C(c)$ and $C(b)$, we can define the subsets $PC(b)$, $PC(c)$ and $PC(d)$ as the potentially dangerous external contingencies to the control area (A) from the external control areas (B,C and D) where:

\[
PC(b) \subseteq C(b) \\
PC(c) \subseteq C(c) \\
PC(d) \subseteq C(d)
\]  

(4)

In order to ensure the security of the internal control area (A) with respect to all potential external contingencies, it is sufficient that the operator of this area assesses only the impact of the potentially harmful external contingencies of all external control areas interconnected to his own control area. In our example the set $PC(tot)$ represents these contingencies where:

\[
PC(tot) = PC(b) \cup PC(c) \cup PC(d)
\]  

(5)

To this end, the operator of the control area (A) uses in real-time a classification decision tree-based filters that allow him to extract $PC(b)$ from $C(b)$ , $PC(c)$ from $C(c)$ and $PC(d)$ from $C(d)$. Hence, he can only assess the impact of $PC(tot)$ represented by the associated set of the post-contingency power injections at his boundary buses, received from the external control areas (B,C and D).
V. DATABASE GENERATING ALGORITHM FOR DECISION TREE BASED FILTER TRAINING AND TESTING

The algorithm described in this section has been developed to enable each control area operator to acquire decision tree-based filters. Actually, each one of these filters will be designed for a specific external and adjacent area. Consequently, each filter is used to screen on-line the contingencies of its own external area. It is important to indicate that the entering data of each filter will be the post-contingency boundary buses parameters, representing the external contingencies, received on-line from its external area. Its output will indicate to the operator which one of the external contingencies is potentially harmful and which one is not with respect to his own control area. In fact, the modeling and market, day-ahead market, data exchange between operators [8], [9] will be used on day to day basis by each operator to build up a model for the whole interconnected system for each hour of the next day. We designate these models as multi-area hourly models.

The developed algorithm uses these hourly models as base cases for modeling the next day real-time operational conditions, operating points, of the multi-area electric power system. Subsequently, the new obtained operating points, each one represents a possible real time operational scenario, will be used to generate the required database for training and testing the decision tree-based filters. The following three functions form the core of the developed algorithm:

A. Hourly model load curve random sampling

In order to consider the load fluctuations which occur in real time, this function samples randomly the scheduled load curve of each hourly model of the multi area electric power system. The sampling process will generate random values for all load buses of the interconnected system. In fact, the generated values, which are percent values of the original ones, will be included between the most probable maximum and minimum values of real-time load variations at each load bus. These values will be also randomly added to (or subtracted from) the original values of the hourly model load curve. Thus, a new load curve representing a possible real-time operational condition of the multi-area hourly model will be obtained.

The sampling process of all next day hourly models will be repeated as many times as necessary to get a sufficiently representing database. The final sufficiently representing database will be obtained once the classification error of the decision tree testing set of this database becomes smaller than a predefined value of an accepted error.

B. Readjustment of control areas generating units outputs

For each new load curve generated by the first function, this function will readjust the outputs of generating units of the hourly model from which the new load curve is derived. The readjustment procedure will be done at the level of each control area so that the overall change, increase or decrease, in load value of each control area will be compensated only by the generating units of this area. In our algorithm, we suppose, for simplicity reasons, that in the case of increase or decrease in control area load level the amount of control area load variation will be distributed between control area generators so that each generator in this area will be loaded or unloaded proportionally to its nominal capacity.

The previous two functions generate the necessary operating points that will be used by the security analysis and database generation function.

C. Security analysis and database generation

The core of this function is a power flow program. This program uses the data of each generated operating point as entering data to assess the impact of all external contingencies on the internal area for which we design the decision tree based filters. The following steps summarize how this function acts to generate the required database:

- For each operating point generated by the first two functions, each contingency of each external control area will be assessed using the power flow program.
- The impact on the internal control area of each assessed external contingency is determined by evaluating the two security criteria of this area as follows:

\[
\begin{align*}
U_{(\min)} & \leq U_{(x)} \leq U_{(\max)} \\
S_{(y)} & \leq S_{(y-Max)} \\
X & = 1, 2, \ldots, a \\
Y & = 1, 2, \ldots, b
\end{align*}
\]

where:
- \(U_{(x)}\) is the voltage magnitude at the load bus (x).
- \(S_{(y)}\) is the apparent power of the line (Y).
- \(a\) is the number of load buses of internal control area.
- \(b\) is the number of lines of the internal control area.

- The internal control area post-contingency boundary buses parameters obtained after each external contingency analysis will be classified as database secured scenarios if there are no violations in internal control area security criteria. Otherwise, they will be classified as non-secured database scenarios.
- The set of pre-classified internal control area post-contingency boundary buses parameters resulting from the security analysis of all contingencies of some external control area form the sub-database of that external area.
- The internal control area operator will use each external area sub-database for training and testing the decision tree-based filter used on-line to screen the contingencies of that external area.

Fig. 4 shows the off-line decision tree-based filter building of each one of the external control areas and, the on-line exploitation of the resulting classification rules for control area external contingencies screening.
Market and modelling data exchange between operators of adjacent control areas

Base cases (hourly models) building at the level of each control area operator

Specifying an initial number of iterations for random sampling process

At each iteration, Random sampling of all hourly models' load curves

New hourly models' load curves

For each operating point, assess each contingency of the last external and adjacent control area and evaluate the impact on the internal control area

First external area sub-database

Data mining using CDT supervised learning technique (building decision tree-based filter for each external area from its sub-database)

VI. STUDY CASE

The IEEE 39-bus test transmission network [18] shown in Fig. 5 has been used for demonstrating the effectiveness of the developed approach. For this purpose, this test network has been divided into two control areas (A) and (B).

For this test system, the operator of the control area (A) will apply the algorithm described in the section (V) to generate the necessary database for designing a decision tree-based filter allowing the contingencies of the control area (B) to be screened on-line.

Each case of the database represents the corresponding numerical values of thirteen attributes which are the control area (A) boundary buses parameters (V-9,I-9,P-9,Q-9,V-4,I-4,P-4,Q-4,V-15,I-15,P-15,Q-15) and the classification parameter (F), where: (V-9,V-4,V-15), (I-9,I-4,I-15), (P-9,P-4,P-15), (Q-9,Q-4,Q-15) represent respectively the voltage magnitudes, the injected current magnitudes, the active power injections, and the reactive power injections at the boundary buses (9,4,15).

For the secured cases, the parameter (F) takes the numerical value (0), while for the non-secured cases it takes the numerical value (1).

In our test system, the operator of the control area (A) will consider the designed filter as an appropriate tool for on-line external contingencies screening if its overall classification error for both secured and non-secured cases of the database test set is smaller than 3%, and its classification error of the non-secured cases of the database test set is lower than 1%. Otherwise, the operator should generate more scenarios and his filter should be retrained until he gets the filter with the desired accuracy.

For our test system, it was necessary to generate three databases for designing the decision tree-based filter of the control area (A) with the desired performance. The three generated databases correspond respectively to (15, 25 and 35) randomly generated load curves for each hourly model load curve. Each one of the three databases has been divided into training set and test set according to the table I.
Starting from the training sets, we used CART software [17], CART is a robust decision-tree tool for data mining, predictive modeling, and data preprocessing, to build the corresponding decision tree-based filter of each database. Then, using the same software, the performance of each filter is tested via its database test set. The overall results of the performance tests are shown in the table II.

### TABLE II
RESULTS OF PERFORMANCE TESTS OF THE THREE DECISION TREE-BASED FILTERS

<table>
<thead>
<tr>
<th>Test sets</th>
<th>Actual class</th>
<th>Total cases</th>
<th>Percent correct</th>
<th>Performance tests results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1322</td>
<td>95.23</td>
<td>1259/1322 63</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>86</td>
<td>96.5</td>
<td>3/86 83</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>1408</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td>Overall% correct</td>
<td></td>
<td>95.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2315</td>
<td>97.23</td>
<td>2251/2315 64</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>149</td>
<td>97.31</td>
<td>4/149 145</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>2464</td>
<td></td>
<td>149</td>
</tr>
<tr>
<td>Overall% correct</td>
<td></td>
<td>97.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2981</td>
<td>97.85</td>
<td>2917/2981 64</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>187</td>
<td><strong>99.46</strong></td>
<td>1/186 186</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>3168</td>
<td></td>
<td>187</td>
</tr>
<tr>
<td>Overall% correct</td>
<td></td>
<td>97.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From these results, we can conclude that only the decision tree-based filter obtained from the training set of the third database fulfills the required accuracy of the performance test. Thus, the operator of the control area (A) will select the decision tree-based filter of the third database shown in Fig. 6 to be his on-line filtering tool of the contingencies of the control area (B).

VII. DISCUSSION AND CONCLUSION

In this paper, we designed a CDT-based filter for rapid screening of control area external contingencies.

The decision tree intelligence technique has been chosen because in contrast with the other intelligence techniques, the decision tree has the following advantages:

- **The decision tree is an efficient natural attribute selection tool.** Accordingly, the association of decision tree approach with other techniques for the pre-selection of the most important attributes is not needed. For instance, in fig. 5 the decision tree has selected only three (I-9, U-15, Q-4) out of the 12 attributes to classify the cases.

- **Decision tree training process can be done in relatively short time compared to other techniques.** For example, the training based on neural networks is a very long and time-consuming process especially, for the large databases as we have in our study.

- **The decision tree provides an effective data analysis tool, thanks to its quantitative information such as attribute scores and information quantities, together with the tree interpretability features.**

Although quite simple, the decision tree shows to be a good classifier. In our example, the test performance of the third filter shows that only 64 cases out of the 5280 unseen test cases are misclassified. Actually, the decision tree has achieved low classification error (2.06%) on the test set despite the high imbalance between secured (2981/3168) and non-secured (187/3168) cases of the test set.
VIII. REFERENCES


