Classification of Load Change Transients and Incipient Abnormalities in Underground Cable Using Pattern Analysis Techniques

Mirrasoul J. Mousavi, *IEEE Student Member*, Karen L. Butler-Purry, *IEEE Senior Member*
Ricardo Gutierrez-Osuna, *IEEE Member*, Massieh Najafi

***Abstract*— This paper presents a feasibility study on the application of pattern analysis techniques to classify load change transients and incipient abnormalities in an underground distribution cable lateral. The data were collected using an on-line monitoring system installed in a residential area in Dallas. A set of features obtained from wavelet packet analysis was evaluated. Methods of dimensionality reduction were employed to overcome the curse of dimensionality while preserving a good classification rate. The classification results using k-nearest-neighbor (KNN) classifiers show that the proposed methodology can be used to classify load change transients and incipient abnormalities.**

***Index Terms*—Underground distribution cable, incipient abnormalities, wavelet packet analysis, pattern analysis.**

I. INTRODUCTION

Events in an underground distribution system can be broadly divided into two main categories, normal and abnormal. When normal events take place, no action needs to be taken. However, when an abnormality occurs, a series of corrective actions need to be used to ascertain the safe and reliable operation of the system. Abnormal events and transients can be categorized in terms of their severity level. For instance, transients due to load changes may occur frequently yet they are considered low level transients. In fact, it is the customer demand that dictates the amount of power to be delivered at any particular time. On the contrary, incipient fault-based events introduce high-level abnormalities that need to be identified so that the necessary corrective action can be taken. Hence, it is crucial to be able to distinguish among these events and issue an appropriate alarm signal upon completion of the detection of such abnormalities. The unique and dissimilar characteristics of these events suggest utilizing pattern analysis techniques for automated identification. Successful classification of abnormalities and transient events would be a great benefit to the utilities, enabling them to detect severe faults at an early stage of their development, and consequently preventing unscheduled outages due to failures in underground cable.

Ongoing research at the Power System Automation Laboratory (PSAL) at Texas A&M University aims to develop an incipient fault detection method having the capability of predicting the remaining life of underground cables. This paper presents preliminary work on the evaluation of a set of features obtained by a time-frequency multi resolution technique to classify load change transients and incipient abnormalities in an underground distribution cable lateral. This work was conducted to determine the most informative features that could distinguish between incipient abnormalities and load change transients.

In this paper, a load change transient is defined as an appreciable increase or decrease in the current signals, and an incipient abnormality is any activity in the low frequency or high frequency signal pertaining to an incipient behavior. Section II provides a concise description of the data collection system and explains the formulation of the classification problem in terms of pattern analysis terminologies. In section III a thorough discussion of the classification results is given. Section IV provides conclusions.

II. PROBLEM FORMULATION

As in any pattern analysis problem, there are four distinct steps to translate a classification problem into a pattern analysis formulation, data collection, preprocessing of data, feature extraction, and model selection / classifier design [1]-[2]-[3].

A. Data Collection

An underground distribution cable lateral installed in a residential area in Dallas was chosen to collect on-line data. This site was selected as the most appropriate location to capture possible incipient abnormalities. Fig. 1 shows the experiment site including the distribution transformer and the

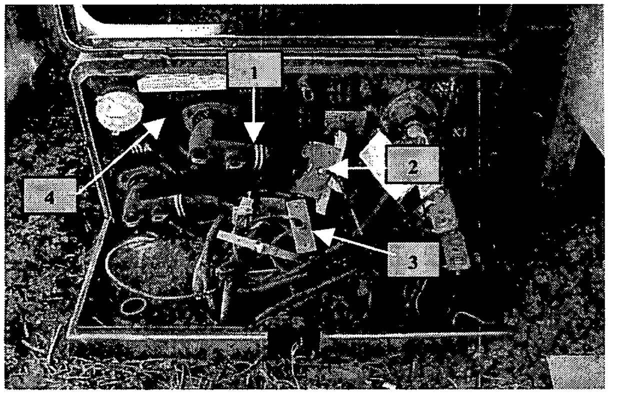
This work was supported in part by Oncor Service Delivery and the Texas Advanced Technology Program through Grant No. 000512-0311-1999.

M. J. Mousavi and K.L Butler-Purry are with Texas A&M University, Department of Electrical Engineering, Power System Automation Laboratory; College Station, TX 77843-3128 (email: klbutler@ee.tamu.edu, mousavi@ee.tamu.edu).

R. Gutierrez-Osuna is with Texas A&M University, Department of Computer Science, College Station, TX 77843-3112 (email: rgutier@cs.tamu.edu).

M. Najafi is with Texas A&M University, Department of Mechanical Engineering, College Station, TX 77843-3123, (email: mnajafi@tamu.edu)

underground cable. The underground distribution cable lateral is fed from a standard 7200 v distribution feeder and supplies power to the 7200v/120v/240v, 100 KVA, 60 Hz distribution transformer via a normally open distribution loop.



1 Cable Lateral Being Monitored 2 Phase Current CT
3 Neutral Current CT 4 Pad Mounted Distribution Transformer

Fig. 1. Monitoring Site in Dallas, Transformer and Cable

Data collection was performed using an on-line monitoring system installed in the site. The monitoring system, whose block diagram is shown in Fig. 2, comprises three basic components: signal transducers, analog signal conditioning unit, and a computer-based data acquisition system. The signal transducers transform the voltage signals to levels acceptable by the signal-conditioning unit. They also transform the current signals into equivalent voltage signals of acceptable range. The transformed signals are then fed to the analog signal-conditioning unit whose functions are to act as an isolation unit, and to filter the signals into various frequency ranges. Signals from the analog signal-conditioning unit are finally fed to the digital data-acquisition system embedded in the computer.

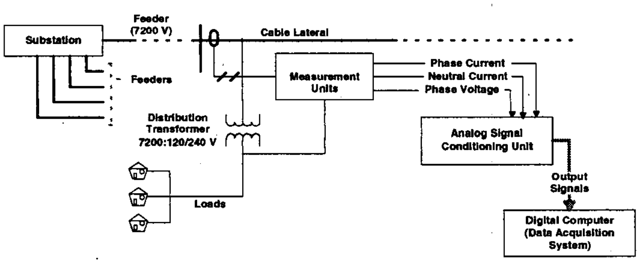


Fig. 2. Block Diagram of the Monitoring System

Using the monitoring system, three basic electrical signals, namely voltage, phase and neutral currents are observed. The system records the signals for one-second duration every 15 minutes. Moreover, various statistical and frequency parameters of these signals namely average, maximum, minimum, standard deviation and magnitude of the harmonics are calculated and recorded.

In the signal-conditioning unit, the phase and neutral current signals create four outputs, as shown in Table I, unique in frequency range and scale, to increase the magnitude resolution during data-acquisition. The notch in three of the output signals is to remove the dominant fundamental frequency (60Hz), thereby improving the magnitude resolution in the given frequency range [4].

TABLE I
CATEGORIES OF CURRENT SIGNALS

Category	Frequency Range of Output Signals
Low Frequency Signal	0 - 780 Hz
Notch Low Frequency Signal	0 -780 Hz, Notch at 60 Hz
Notch High Frequency Signal Scale 1 (NHF x 1)	2 - 7.5 KHz, Notch at 60 Hz
Notch High Frequency Signal Scale 10 (NHF x 10)	2 - 7.5 KHz, Notch at 60 Hz

B. Preprocessing Mechanism

Recorded current signals may contain normal or abnormal activities, however, most of the recorded data represent normal operation of the system. Furthermore, transient events can be initiated by load changes, incipient faults, or other events. A preprocessing scheme was employed in order to filter out normal events and categorize the remaining signals in terms of their corresponding predefined classes.

A second motivation for the use of preprocessing is to reduce redundancy. The sampling rate for the notch low frequency signals was set at 15360 samples/sec for recording, which gave a frequency range of 0 – 7680 Hz. However, these signals were limited to 0 – 780 Hz by a low pass anti-aliasing filter in the analog signal-conditioning unit. Thus, to reduce redundancy, these signals were decimated by a factor of eight yielding an effective sampling rate of 1920 samples/sec.

C. Feature Extraction / Selection

The goal of feature extraction or selection is to obtain a few features that discriminate classes with a high degree of accuracy. This important procedure is a key step for the success of any classifier. The feature extraction/selection process involves two steps. First, a number of raw features are obtained. This was accomplished in the present work by means of Discrete Wavelet Packet Analysis (DWPA) method [5]. Second, the raw features are projected onto a lower dimensional space by means of multivariate statistical techniques in order to reduce the dimensionality while preserving a good classification rate. These techniques are briefly explained in the following sections.

1) Wavelet Packet Analysis Design

Analyzing data using DWPA involves three steps, selection of the type of mother wavelet, the order of mother wavelet, and the level of decomposition. A number of wavelet families with unique properties have been proposed in the signal processing literature, but the most appropriate family is generally application-dependant. After literature review and from earlier wavelet analysis results, it was found that the fourth order Daubechies wavelet yields the best performance for studying power system transients and incipient behavior [4]-[6]-[7]. Thus, the fourth order Daubechies wavelet was chosen as the mother wavelet for the analysis. Roughly speaking, selection of the level of decomposition depends on the desired frequency resolution. To obtain the best frequency resolution, the 4^{th} level decomposition was chosen as illustrated in Fig. 3.

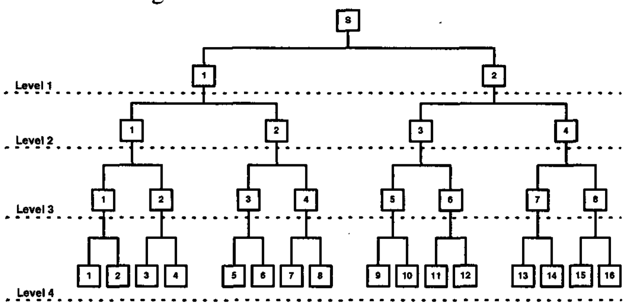


Fig. 3. 4th Level Wavelet Packet Decomposition Tree

The wavelet packet analysis was conducted on sample length of one-second duration. Thus, for one-second data, the low frequency signals contain 1920 samples and the high frequency signals contain 15360 samples. To obtain a better frequency range in the sixteen details, the samples are zero-padded symmetrically (at the beginning and the end of the signal) to achieve dyadic sample length (2048 samples for low frequency signals and 16384 samples for high frequency signals). The approximate frequency ranges for each of the details at level 4 are 64 Hz and 0.5 KHz for a signal length equal to 2048 and 16384, respectively. It should be noted that, for the signal with 2048 samples, the frequency ranges at level 4 are such that the harmonics of 60 Hz (fundamental frequency) are evenly distributed among the details. Therefore, zero-padding the signal facilitate the interpretation of analysis results in terms of signal harmonics.

2) Formation of Raw Feature Vector

After 4th level wavelet packet decomposition, the resulting 16 details were stored along with the original signal. Raw features are defined to be the maximum magnitude of spikes in each of these signals. The magnitude of the spikes is a measure of contribution of that frequency range to the original signal. To provide a better comparison among the details, the magnitude of the maximum spike is normalized by the magnitude of the corresponding spike in the original signal. The normalized magnitude of the spike was then stored in the raw feature vector. The vector thus consists of 17 elements, 16 of which represent the normalized magnitude of the spikes in the details plus the magnitude of the spike in the original signal as the 17^{th} feature. This process was performed on every input signal, resulting in a $N\times 17$ data matrix where N denotes the total number of examples in the data set (2110 examples).

3) Dimensionality Reduction

The objective of dimensionality reduction is to keep the dimensionality of the pattern recognition problem (i.e. the number of features) as small as possible while preserving good classification accuracy. Dimensionality reduction can be accomplished by means of feature selection or feature extraction. The term feature selection refers to techniques that select the best subset of the input features set. Methods that create new features based on transformations and combinations of the original feature set are called feature extraction methods. The choice between feature selection and extraction depends on the application domain.

Principal Component Analysis (PCA) is the best-known linear unsupervised feature extraction method [3]. The linear transformation is defined by the eigenvectors of the covariance matrix, which leads to vectors that are uncorrelated regardless of the form of the distribution. If the distribution happens to be Gaussian, then the transformed vectors will be statistically independent. The objective of PCA is to perform dimensionality reduction while preserving as much as the randomness (variance) in the high-dimensional space as possible. PCA performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance. The main limitation, however, is that as an unsupervised method, it does not consider class separability information. There is no guarantee that the direction of maximum variance will contain good features for discrimination.

Linear Discriminante Analysis (LDA) is another well-known linear feature extraction method, but unlike PCA it is supervised [2]-[3]. The objective of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. In LDA, interclass separation is measured by Fisher criterion, which finds the eigenvalues of the between-class scatter matrix to the within-class scatter matrix. The within-class scatter matrix is defined by:

$$S_w = \sum_{i=1}^{c} S_i \tag{1}$$

where,

$$S_i = \sum_{x\in w_i} (x-\mu_i)(x-\mu_i)^T \tag{2}$$

x denotes the data, c is the number of classes and μ_i is the mean vector of class w_i.

The between-class scatter is defined by :

$$S_B = \sum_{i=1}^{c} N_i(\mu_i-\mu)(\mu_i-\mu)^T \tag{3}$$

where N_i is the number of patterns of class i, and μ is the mean of the entire distribution.

The solution proposed by Fisher is to maximize the function that represents the difference between the means of the classes (between-class scatter) normalized by a measure of the within-class scatter. The projections with maximum class-separability are the eigenvectors corresponding to the largest eigenvalues of $S_w^{-1}S_B$. This method produces as many projections as the number of classes minus one. If the classification error estimates establish that more features are needed, some other methods must be employed to provide additional features. LDA will fail when the discriminatory information is not in the mean of the data but rather in the variance.

Raw features might be expensive to obtain and might not be numeric. Also, in some applications it may be important to extract meaningful rules from the classifier results. In such situations, feature extraction methods will not work. Hence, feature subset selection (FSS) methods must be employed. Feature subset selection requires a search strategy to select candidate subsets and an objective function to evaluate these candidates. There are a large number of search strategies among which Sequential Forward Selection (SFS) a simple greedy approach was used in this work. More details about these methods can be found in [2] and [3] where statistical pattern recognition techniques are well introduced or reviewed.

Objective functions are divided into two groups, filters and wrappers. Filters evaluate feature subsets by their information content; typically interclass distance, statistical dependence or information-theoretic measures. Wrappers are essentially pattern classifiers, which evaluate feature subsets by their predictive accuracy by statistical resampling or cross-validation. Filters are fast to be executed and their results exhibit more generality. However, they tend to select the full feature set as the optimal solution. On the other hand, wrappers generally achieve better classification rates than filters and have mechanism to avoid overfitting. The main disadvantage is slow execution.

D. Model Selection and Classifier Design

Once the extracted/selected features are obtained, the data set is organized into classes as shown in Fig. 4. There are two broad classes, load change transients and incipient abnormalities. Each class in turn contains four subgroups, Phase Notch High Frequency (PH_NHF), Neutral Notch High Frequency (NE_NHF), Phase Notch Low Frequency (PH_NLF), and Neutral Notch Low Frequency (NE_NLF). Based on these categories, three pattern analysis problems are defined. Two 4-class problems for each of load change transients and incipient abnormalities, whether the abnormality is manifested in PH_NHF, PH_NLF, NE_NHF or NE_NLF signals. The third problem is a 5-class problem for which the four classes of the load change transients are considered as one class and categories of incipient abnormalities form the remaining four classes.

Once a feature selection or extraction procedure finds a proper representation, a classifier can then be designed using a number of possible approaches. In practice, the choice of a classifier is based on which classifiers happen to be available, or best known to the user. In this study, k-nearest neighbor (KNN) classifiers were used. In these classifiers, the K closest examples in the training data set are found and the majority class is determined and assigned to the unlabeled example.

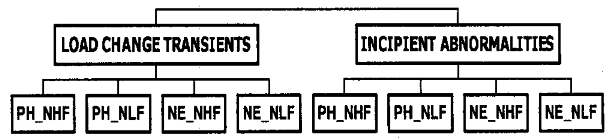


Fig. 4. Problem Formulation as a Classification Problem

III. CLASSIFICATION RESULTS

The classification results for each of the problems are shown in the following figures. The term *original* in the figures implies that all seventeen features in the original feature space are used in the classification without dimensionality reduction. The classification rate (CR) is a measure of the performance of the classifier defined by:

$$CR = \frac{\textit{number of correct assignments}}{\textit{number of total assignments}} \qquad (4)$$

A. Results of the 4-Class Classification Problem for Load change transients

Fig. 5 summarizes the classification results for the 4-class problem with load change transients. As seen in the figure, the classification rate for classifying data in the original space was 66%. Applying PCA, the classification rate rose to 68%. The two largest eigenvalues were 89.22% and 5.51% responsible for the variance of the data. Hence, using PCA, only two features were determined to represent the data in the feature space. Applying LDA, the classification rate was around 57%, which means that the discriminatory information is not only in the mean of the data.

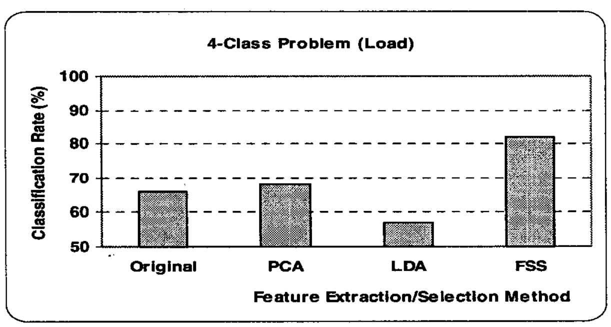


Fig. 5. Classification Results for Load 4-Class Classification Problem

Sequential forward feature selection with wrapper objective function selected features 17, 3, 11, and 12. From the frequency range mentioned earlier, it is inferred that features 3, 11 and 12 represent the contribution of the third, eleventh and twelfth harmonics, respectively. Feature 17 represents the contribution of all harmonics in the signal. The classification rate rose up to 82% when only those four features were used. Not only does FSS provide better performance than PCA or

LDA, but also selected features can be technically interpreted in terms of harmonic contents of signals.

B. Results of the 4-Class Classification Problem for incipient Abnormalities

Fig. 6 summarizes the classification results for the 4-class problem on incipient abnormalities. Classification rate in the original space was 68%. After applying PCA, the classification rate rose to 73%. The four largest eigenvalues were 82.75%, 6.87%, 2.44%, and 1.41% responsible for the variance of the data. Applying LDA, the classification rate was not improved, which means that again the discriminatory information for this class is not only in the mean. Sequential forward feature selection with a wrapper objective function selected features 17, 13, 11, and 12 which correspond to the normalized magnitude of the spikes in the original signal and its 11^{th}, 12^{th}, and 13^{th} harmonics. The classification rate increased up to 82% using these four features.

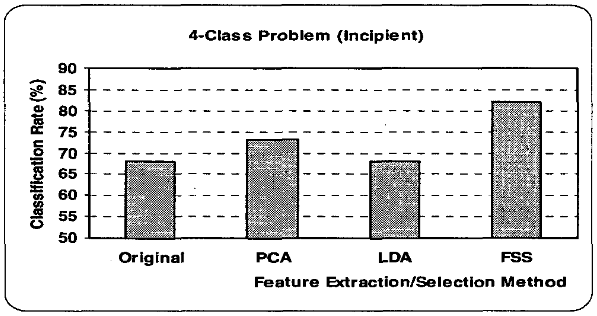


Fig. 6. Classification Results for Incipient 4-Class Classification Problem

C. Results of the 5-Class Problem for Load Change Transients and Incipient Abnormalities

Fig. 7 depicts the classification results for the 5-class problem defined earlier. In this case, the classification rate in the original space was 57%. After applying PCA or LDA, the classification rate was not improved. Sequential forward feature selection with a wrapper objective function selected features 17, 13, 2, and 1 resulting in 72% classification rate.

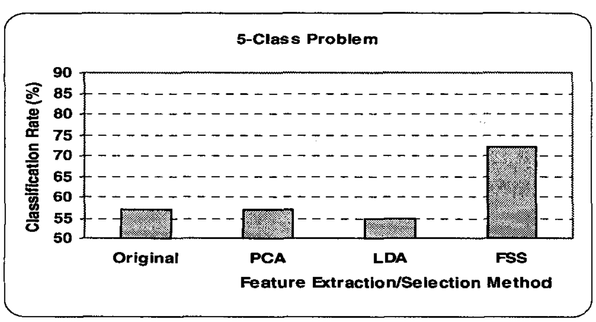


Fig. 7. Classification Results for 5-Class Classification Problem

As it is seen, the classification rate is lower than the other 4-class problems. By further investigation, it was found that the misclassified examples mostly belonged to the load change transient class whereas incipient ones were correctly classified. The first reason has to do with the number of existing examples from load change transients relative to the number of incipient abnormalities, which are dominant. Second reason lies in the fact that the best feature (the magnitude of the main frequency component) to distinguish load change transients from incipient abnormalities is not included in the set of features. Recall that all the signals were filtered by a notch at 60 HZ. In general, the magnitude of spikes in the incipient abnormalities is smaller than that of load change transients. As the order of harmonics increase, the spike magnitude becomes smaller and smaller. Therefore, it is quite reasonable to assume that the main frequency component will carry discriminatory information. It was observed that the classification rate is raised to 83% after adding this feature to the KNN classifier.

Fig. 8 shows scatter plot for two features selected by FSS method. As shown, the boundary that separates the two classes is nonlinear.

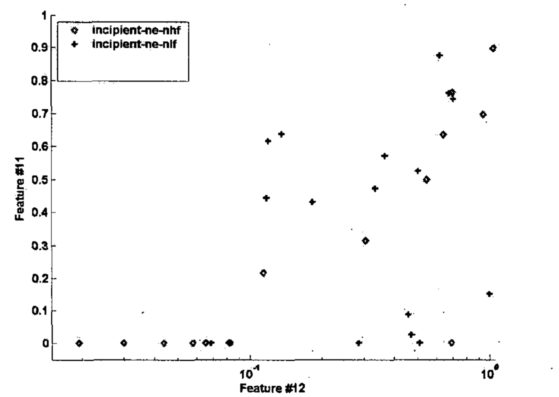


Fig.8. Scatter Plot of Two Selected Features

IV. Conclusions

Three classification problems to categorize load change transients and incipient abnormalities in the underground distribution cable were defined and solved. The classification was performed using seventeen features obtained from wavelet packet analysis. In each classification problem, methods of dimensionality reduction were employed. It was observed that the feature subset selection method had a better performance as compared to PCA or LDA feature extraction methods. The final classification results using KNN classifiers were encouraging in all three cases. Future work includes exploring additional features and utilizing powerful classifiers such as Support Vector Machines (SVM) [8] to further improve the classification rate.

Acknowledgment

The authors acknowledge Ritesh Chaturbedi for his contribution to this research.

References

[1] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd Edition, New York, Wiley, 2001.

[2] A. Webb, *Statistical Pattern Recognition*, London, Arnold, 1999.

[3] Anil K. Jain, Robert P.W. Diun and Jianchange Mao, " Statistical Pattern Recognition: A Review", *IEEE Transactions on Pattern Analysis Intelligence* Vol. 22, No.1, Jan 2000, pp. 4 -37.

[4] R. Chaturbedi, *Characterization and Detection of Incipient Underground Cable Failures*, MS Thesis, Dept. of Electrical Engineering, Texas A&M University, College Station, August 2002.

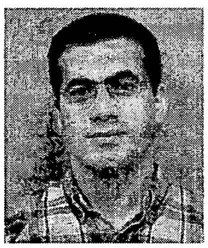
[5] Matlab reference book for wavelet toolbox, The Math Works, Version 2, 2000.

[6] Chul H. Kim and Raj Aggarwal, " Wavelet transforms in Power Systems", *Power Engineering Journal*, August 2001, pp. 193-202.

[7] A. W. Galli and O. M. Nilsen," Wavelet Analysis for Power System Transients", *IEEE Computer Applications in Power*, vol.12, no.1, January 1999, pp. 16-25.

[8] N. Cristianini, J. Shawe-Taylor, *An introduction to Support Vector Machines and other Kernel-Based Learning Methods*, London, Cambridge Press, 2000.

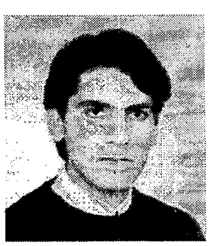
BIOGRAPHIES



Mirrasoul Jaafari Mousavi received the B.Sc. degree from the Isfahan University of Technology in 1996 and the M.Sc. degree from Sharif University of Technology in 1999 all in Electrical Power Engineering. He worked for Niroo Research Institute (NRI) from 1999 to 2001. His research interests are related to fault detection, anticipation and diagnosis, transient studies, protection and distribution automation. He is a student member of IEEE and IEEE Power Engineering Society (PES). He is currently pursuing his PhD degree in the Electrical Department at Texas A&M University.

Karen L. Butler-Purry is an associate professor in the department of electrical engineering at Texas A&M University. In 1988-1989, Dr. Butler-Purry was a Member of Technical Staff at Hughes Aircraft Co. in Culver City, California. Her research focuses on the areas of computer and intelligent systems applications in power, power distribution automation, and modeling and simulation of power systems and vehicles. Dr. Butler-Purry is a senior member of IEEE and IEEE Power Engineering Society (PES), and a member of the Louisiana Engineering Society. She is a registered professional engineer in the states of Louisiana, Texas, and Mississippi.

Ricardo Gutierrez-Osuna (M'00) received the B.S. degree in Industrial/Electronics Engineering from the Polytechnic University of Madrid in 1992, and the M.S. and Ph.D. degrees in Computer Engineering from North Carolina State University in 1995 and 1998, respectively. From 1998 to 2002 he served on the faculty at Wright State University. He is currently an assistant professor in the Department of Computer Science at Texas A&M University. His research interests include pattern recognition, machine learning, biological cybernetics, machine olfaction, speech-driven facial animation, computer vision and mobile robotics.



Massieh Najafi received his B.Sc. degree from University of Tehran in Iran in 2000. He is currently pursuing his graduate studies in the Mechanical Engineering Department at Texas A&M University. He is a research assistant in Energy System Laboratory (ESL) at Texas A&M University. He is a student member of ASME. His research interests are fault detection, sensor diagnosis, intelligent systems, and control.