Improving Detection Attacks in Electric Power System Critical Infrastructure Using Rough Classification Algorithm

Maurílio Pereira Coutinho, Germano Lambert-Torres, *Member,* IEEE,
Luiz Eduardo Borges da Silva, *Member,* IEEE, and Horst Lazarek

***Abstract*—Nowadays, National Critical Infrastructures play a fundamental role in modern society. The use of information technology (IT) to achieve service quality produces vulnerabilities and security threats. To safeguard against the threat of cyber-attacks, providers of Critical Infrastructure services also need to maintain the accuracy, assurance and integrity of their interdependent data networks. This paper presents a novel technique for improving the security of Electric Power System Critical Infrastructure by implementing anomaly detection methods to identify attacks and faults. By using Rough Sets Classification Algorithm, a set of rules can be defined to the anomaly detection process. This can be used for identify attacks and failures and, also, for improving state estimation.**

***Index Terms*—Critical infrastructure protection, electric power system, SCADA , detecting attacks, rough set theory, data mining.**

I. Introduction

THE interconnection between electricity management network, corporate network and the internet and the use of information technology (IT) to achieve service quality produces vulnerabilities and exposes the electricity cyber infrastructure to securities threats [1,2,3,4]. With the electricity market deregulation and the usage of commercial off the shelf technologies (COTS), standard TCP/IP networks, and fully networked systems the opportunity for cyber attackers increases and the threat of such attack has to be addressed [5]. Current initiatives for Security of Critical Infrastructures can be found in [6].

As part of the Electric Power System Critical Infrastructure, SCADA systems and Energy Management Systems (EMS) play a vital role in order to monitoring the safety, reliability, and protective functions of the power grid. However, these systems, that were designed to maximize functionality with little attention paid to security, represent potential vulnerability to disruption of service or manipulation of operational data that could result in public safety concerns [7].

According to [8] there are two approaches to become SCADA systems more secure: One is to identify problems at the perimeter of the system using anti-virus and Intrusion Detection Systems (IDS). The second is to model the normal data flows and control operations within the SCADA system to detect anomalies caused by attempts to change or damage the system.

Using the second approach, this paper presents the development of the technique for implementing anomaly detection to monitor Power Electric Systems, previously introduced in [6]. The problem was addressed in [6] using Rough Sets Classification Algorithm proposed by Pawlack et al [9].

This paper is organized as follow: Firstly, an overview of the Electric Power Systems Critical Infrastructure, SCADA systems and Rough Sets Classification Algorithm is presented. Then, the architecture of the Anomaly Detection System is introduced and the methodology to build the knowledge data base and how to extract the rules from such data base is described. A Six Bus Power System is used as an example.

II. Electric Power System Critical Infrastructure

In general, an Electric Power System Critical Infrastructure is highly interconnected and dynamic, consisting of several utilities. Due to its hierarchical organization, it is sub-divided into regional grids. Each sector is further split into generation, transmission, distribution, and customer service systems, supplemented with an energy trading system. The Power Grid is comprised of a myriad assets, such as Generation Plants, Transmission Lines, Transmission and Distribution Power Substations, Local, Regional and National Control Centres, Remote Terminal Units (RTUs)/Intelligent Electronic Devices (IEDs), and Communication Links [10]. Accordingly Figure 1 the computer electricity cyber infrastructure can be divided in two main components: Electric Management Systems, which allow operators to regulate power flow, and the Supervisory Control and Data Acquisition (SCADA) systems for

Manuscript received September 11, 2007. This work was supported in part by the Brazilian Research Council (CNPq) and Minas Gerais State Research Foundation (FAPEMIG).

M. P. Coutinho, G. Lambert-Torres, and L.E. Borges da Silva are with the Federal University of Itajuba (UNIFEI), Itajuba, MG, 37500-903, Brazil (phone: +55-35-36291240; fax: +55-35-3629118755; e-mail: {maurilio.coutinho, germanoltorres}@gmail.com).

H. Lazarek is with the Technical University of Dresden, Dresden, Germany.

monitoring the safety, reliability, and protective functions of the power grid. This figure illustrates the interactions between the various grid entities [11].

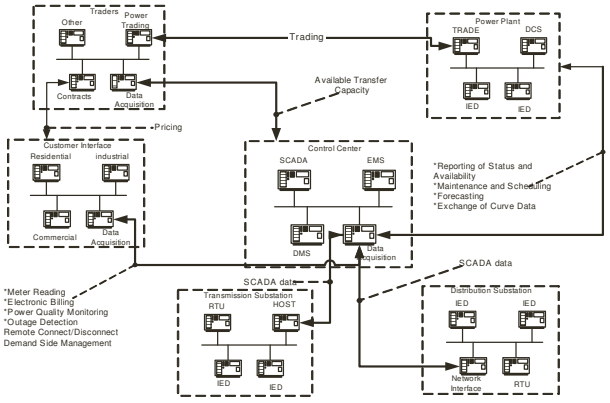

Fig. 1. Deregulated Electric Power Market[11]

III. Protecting SCADA Systems

SCADA systems are used for data collection from sensors and instruments located at remote sites and to transmit and to display the data at a Control Center for Control and Supervisory purposes. These systems can monitor and control hundreds of I/O points. The Remote Terminal Units (RTUs) are located between the remote sensors and the Control Center in order to gather the data from sensors and field devices. The sensors have Digital or Analog I/O and these signals are not in a form that can be easily communicated over long distances. The RTUs are used to digitize the sensor signals so that they can be digitally transmitted via communication protocols over long distances to the Control Center. The figure 2 illustrates such organization.

SCADA Communications can employ a diverse range of both wired (leased lines, dialup line, fiber, ADSL, cable) and wireless media (spread spectrum, cellular, WLAN or satellite). The choice depends on a number of factors that characterizes the utility existing communication infrastructure.

Analyzing the vulnerabilities of the figure 2 it can be pointed out various weak points where insider or outsider attackers can get access to the SCADA Master and the RTU. For example the Circuit Breaker can be considered an attack object because of the Internet connectivity via corporate network or via remote access using public telephone network. In the case of a success access (inside or outside) to the RTU, two possible scenarios can be visualized: (1) The attacker assumes the control of the circuit breaker or (2) the attacker corrupts the information collected by the RTU. In order to detect such scenarios anomaly detection techniques are used to identify these threats as well as the type of attack.

The Intrusion Detection System (IDS) has been studied widely in recent years. Anomaly-based IDS discovers attacks by identifying unusual behaviour (anomalies) on a host, network or application. They function on the observation that some attackers behave differently than “normal” users or events and thus can be detected by systems that identify these differences. An extended bibliography on IDS is presented in [6].

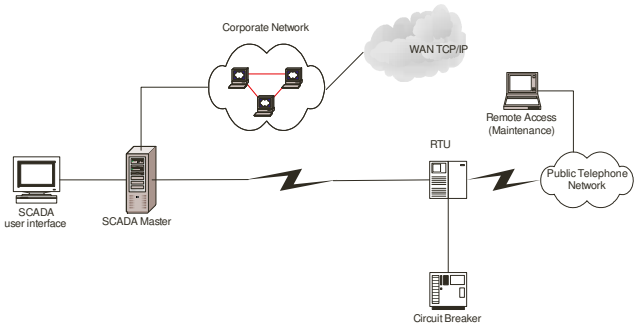

Fig. 2: SCADA System Communication Model

IV. Problem Definition

The operation of a power system is intrinsically complex due to the high degree of uncertainty and the large number of variables involved. The various supervision and control actions require the presence of an operator, who must respond efficiently to the most diverse requests, by handling various types of data and alarm information.

These data and information come from measurements of SCADA systems or from computational processes. The size of the current database in a power control center has increased tremendously over the past few years due to the use of network communications, which renders their control systems more vulnerable to manipulation by malicious intruders. In order to improve the security of SCADA systems, anomaly detection can be used to identify corrupted values caused by malicious attacks and faults.

The system operator must be apprised of the current state of the system and some forecasted position, such as load forecasting, maintenance scheduling, in order to take a control action (switching, changing taps, and voltage levels). One of the most important operator tasks is to determine the current operational state of the system. To accomplish this task, the operator receives many data from/into the system. By handling these data, the operator attempts to build an image of the operation point.

The analysis attempts to assess the operational mode in one of the 2 states: normal, or abnormal. In the first state, normal, all loads are supplied and all measurements are inside the nominal rates. In the second state, abnormal, all loads continue to be supplied, but some of the measurements are outside the nominal rates or some loads are not supplied. The operator must regularly analyze the system security, even when the operation state is normal. This analysis is conducted according to possible contingencies that could affect the power system.

According to Bigham et al [12], there are a number of ways in which anomaly-detecting methodologies could enhance the integrity and security of electricity data. Firstly, it could act as

a useful complement to existing techniques, such as state estimation, for verifying the likely correctness of electricity measurements and give operators constant feedback about changes the integrity and reliability of the data. A second application is the improvement of standard protection devices such traditional IDS and virus checker.

V. Rough Sets Classification Algorithm

The Rough Set Theory, developed by Pawlak [9], has emerged as a mathematical method to manage uncertainties from inexact, noisy and incomplete information and it has been one of the focal point research areas in artificial intelligence since its advent [13]. In [14] it is presented the basic concepts of rough set theory and point out some rough set-based research directions and applications.

Before presenting the algorithm, two major concepts in the Rough Set theory, *reduct* and *core*, must be defined. These concepts are important in the knowledge of base reduction.

Let **R** be a family of equivalence relations. The reduct of **R**, *RED*(**R**), is defined as a reduced set of relations that conserve the same inductive classification of set **R**. The core of **R**, *CORE*(**R**), is the set of relations that appear in all reduct of **R** (i.e., the set of all indispensable relations to characterize the relation **R**). The main idea behind the knowledge base reduction is a simplification of a set of examples. This can be obtained by the following procedure:

a) Calculate the core of the problem;
b) Eliminate or substitute a variable by another one; and
c) Redefine the problem using new basic categories.

The algorithm that provides the reduction of conditions can be represented by the following steps:

Step 1: Eliminate dispensable attributes
Step 2: Compute the core of the set of examples.
Step 3: Compute the reduced set of relations that conserve the same inductive classification of the original set of examples.
Step 4: Merge possible examples and compose the final set of rules.

VI. Anomaly Detection Architecture

The proposed solution for the attack scenarios pointed out in the Sections III and IV using anomaly detection is presented on figure 3 and uses intelligent techniques to extract knowledge from the SCADA system. The approach is divided in 2 steps: Firstly, the knowledge extractor should generate a set of rules that will determine the normal or abnormal behaviour of the system. The data come from RTUs and will be checked against the set of rules to define the normality of the measurements. Secondly, the anomaly detector should recognize the type of attack occurred.

In order to satisfy the limited SCADA Master computational resources, the proposed model should reduce the number of input variables and the number of examples, offering a more compact set of rules for the anomaly detector.

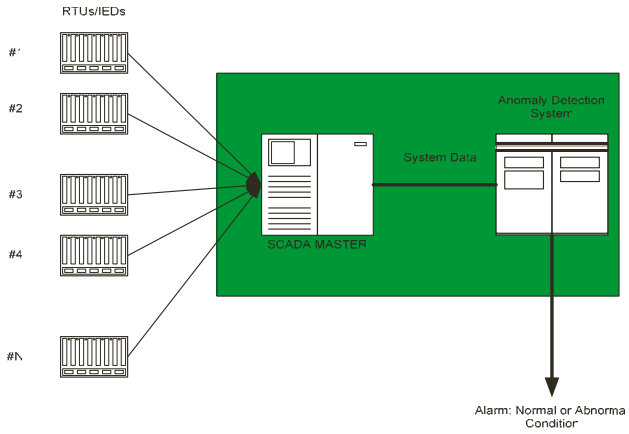

Fig. 3: Proposed Anomaly Detection Architecture

As previously defined in [6] the model uses the Rough Sets Theory to implement the data reduction. This technique is used mainly because:

- It reduces the number of rules without reducing the system knowledge base;
- It has a dynamic behaviour, because not informed rules by the expertise technician can be extracted from the system;
- It reduces the computational resources needed, like memory capacity and processor power;

However, the technique needs a huge amount of collected data to build the knowledge data base.

VII. Experiments and Results

The diagram in the figure 4 represents the test environment for the proposed architecture. The main blocks depicted in the Fig. 4 are:

- Power Flow: To solve the power flow on an electric power system. This program was adapted from [15].
- SCADA Simulator: This program simulates the functions performed by an online Supervisory Control and Data Acquisition system. The idea is to simulate the power system network, calculates all the voltages, power flows and injections on the system and then associates these quantities with a specification of where the measurements are being made on the system. This program was adapted from [15].
- State Estimator: Program for state estimation process adapted from [15].
- Rough Set Rule Extractor: This Module extracts rules from the knowledge data base, using the Rough Sets Classification Algorithm
- Anomaly Detection System: This Module uses the rules defined by the Rough Sets Rule Extractor to determine the state of the SCADA Output data.

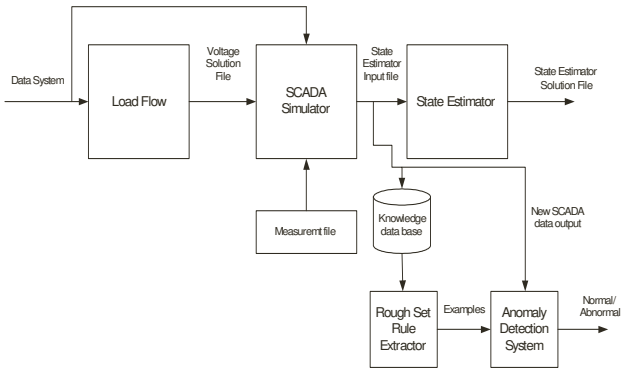

Fig.4 Test Environment Diagram.

According [15], state estimation is the process of assigning a value to an unknown system state variable based on measurements from that system according to some criteria. According [16], "numerical estimation algorithms rebuild the state of the power system in case of missing and/or corrupted data: however this approach does not address the problem of giving a normal/abnormal state assessment, and in some cases could tend to hide traces of an ongoing attack or of other anomalies". This is a risky assumption since there are often configuration errors and there is always the chance that an attacker could be mediating between the control centre and the electricity network [8].

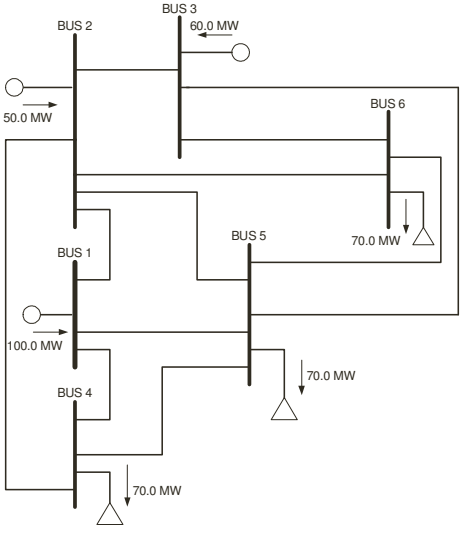

Fig. 5 Six Bus Power System [15].

To test the proposed anomaly detection model, it was used the environment test specified in the fig. 4 and a six bus network described in [15] and presented in the fig. 5. Test data was generated by introducing errors into a selection of the normal state estimation input file, generated by the SCADA Simulator Program. Xuan Jin et al [12] considered 5 types of corruption applied to the electricity data: constant bias with normally distribution deviation, loss of decimal point, sign switch, fixed at fixed value and fixed at random value. They attribute these errors to the fact that electricity measurements can be altered by random noise, attacks, software bugs, meter failures, EMI and transmission errors. The ability of the proposed anomaly detection model to identify normal and abnormal conditions were evaluated against the ability of the state estimator program to provide a reasonable output, even considering that the input data were corrupted in some way.

In order to generate the knowledge data base, the tests produced 45 examples containing the 57 measurements provided by the SCADA simulator program with some sort of errors introduced. For sake of clarity, the errors were applied only on Bus 4 and Bus 6 of the Six Bus Power System and included only sign switch error type. These examples were treated by the Rough Set Rule Extractor program and generated the following rules:

- If Active Power on Bus 4 >= -0.77 and Active Power on Bus 4 < -0.2567 and Active Power on Bus 6 >= -0.77 and Active Power on Bus 6 < -0.5133 then output condition is Normal.
- If Active Power on Bus 4 >= 0.4667 and Active Power on Bus 4 < 0.71 then output condition is Abnormal.
- If Active Power on Bus 6 >= 0.4667 and Active Power on Bus 6 < 0.71 then output condition is Abnormal.

These rules show the size of the original data reduction: The Rough Set Rule Extractor program extracted 3 simple rules from 2565 original data samples.

All previous conditions were tested with the anomaly detection program using the rules established and the results were successful checked against the knowledge data base. In order to evaluate the proposed anomaly detection model against the state estimation solution new inputs were generated by the SCADA simulator program with the same type of error previously introduced. The Base Case Load Flow produced the following values for Bus 4 and 6:

TABLE I
LOAD FLOW AND STATE ESTIMATION OUTPUTS FOR BASE CASE

LINES	LOAD FLOW [MW]	STATE ESTIMATION OUTPUT [MW]
FROM 4 TO 1	-42.53	-42.53
FROM 4 TO 2	-32.04	-32.04
FROM 4 TO 5	4.58	4.58
FROM 6 TO 2	-25.0	-25.80
FROM 6 TO 3	-42.77	-42.77
FROM 6 TO 5	-1.42	-1.42

The SCADA Simulator outputs a file that contains the power and voltage measurements for 29 meters located on selected points of the six bus power system. In the real life, the field instrumentation is connected to RTU that sends the collected data to the Control Center. In the Control Center the operator analyses the results produced by the state estimation process and takes actions in order to maintain the operational system conditions. The following output shows a sample for the Base Case.

TABLE II
STATE ESTIMATION OUTPUT WITH ERRORS INTRODUCED IN THE BASE CASE

Error on Load Bus 4	Error on Load Bus 6	FROM 4 TO 1		FROM 4 TO 2		FROM 4 TO 5		FROM 6 TO 2		FROM 6 TO 3		FROM 6 TO 5	
		Base Case	With Error	Base Case	With Error	Base Case	With Error	Base Case	With Error	Base Case	With Error	Base Case	With Error
+56.0	-70.0	-42.53	-24.08	-32.04	6.32	4.58	18.96	-25.80	-65.10	-42.7	-110.3	-1.42	-19.88
-70.0	**+56.0**	-42.53	-41.62	-32.04	-31.49	4.58	7.52	-25.80	-39.48	-42.7	-83.98	-1.42	-6.24
+63.0	-70.0	-42.53	-23.05	-32.04	8.10	4.58	19.55	-25.80	-65.44	-42.7	-110.3	-1.42	-19.90
-70.0	**+63.0**	-42.53	-41.58	-32.04	-31.84	4.58	7.46	-25.80	-38.37	-42.7	-82.48	-1.42	-5.49
+63.0	**+63.0**	-42.53	-22.18	-32.04	1.51	4.58	18.41	-25.80	-45.09	-42.7	-82.25	-1.42	-6.00
+56.0	**+56.0**	-42.53	-23.26	-32.04	0.08	4.58	17.89	-25.80	-45.84	-42.7	-83.76	-1.42	-6.72

```
1.0500  1.0787  0.2066  0.5000  1.0000  0.6000  0.6000  +0.5600  -0.7000  -0.7000...
1.0500  1.0787  0.2066  0.5000  1.0000  0.6000  0.6000  -0.7000  -0.7000  -0.7000...
1.0500  1.0787  0.2066  0.5000  1.0000  0.6000  0.6000  +0.6300  -0.7000  -0.7000...
1.0500  1.0787  0.2066  0.5000  1.0000  0.6000  0.6000  -0.7000  -0.7000  -0.7000...
1.0500  1.0787  0.2066  0.5000  1.0000  0.6000  0.6000  +0.6300  -0.7000  -0.7000...
1.0500  1.0787  0.2066  0.5000  1.0000  0.6000  0.6000  +0.5600  -0.7000  -0.7000...
```

Fig. 7: Corrupted measurements file for the Anomaly Detector

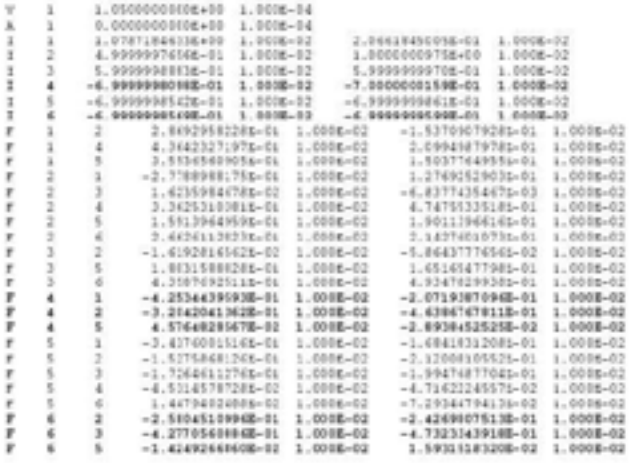

Fig. 6: SCADA Output for Base Case

In the next step six new examples were implemented in order to simulate a corruption of the SCADA output file. It was simulated a change of the sign and the value for the loads on Bus 4 and Bus 6. The table II shows the results for the state estimation process.

The SCADA output file with these corrupted examples becomes input to the proposed Anomaly Detection model. The fig. 7 shows a sample of this measurements file. The fig. 8 shows the Matlab piece of code to implement the rules established by the Rough Set Extractor program. The fig 9 presents the results for this code.

Fig. 8: Matlab Code to implement the rules

The output presented in the fig.9 demonstrates that the anomaly detector has identified the abnormal conditions in the SCADA measurements file. Comparing with the output from state estimation process, it is possible concludes that the proposed anomaly detector model identified those error conditions while the state estimation process presented a possible output that could guide the operator to take wrong actions.

```
» rules
Enter Input File Name: 'Scada_out.txt'
Enter Number of Inputs: 6
Input:      1 RESULT: ABNORMAL
Input:      2 RESULT: ABNORMAL
Input:      3 RESULT: ABNORMAL
Input:      4 RESULT: ABNORMAL
Input:      5 RESULT: ABNORMAL
Input:      6 RESULT: ABNORMAL
```

Fig. 9: Anomaly Detector Output.

VIII. Conclusions and Future Work

Critical Infrastructures, such Electric Power Systems, are vital for our modern society. Therefore they require protection from a variety of threats, and their network is potentially vulnerable to cyber attacks. The Anomaly Detection System is an important tool to increase the security of such Critical Infrastructures. This paper presents an Anomaly Detection Model using a reduced set of rules extracted from a Electric Data Base Knowledge using Rough Set Theory. A test environment was proposed and implemented and an example for power system control centers demonstrated that this technique has many advantages, such as simplicity of implementation and favorable performance. Future work includes expanding the error types introduced in the SCADA output file and the error type identification process. Besides it is the intention to compare such technique using the "Test Data for Anomaly Detection in Electricity Infrastructure" proposed in [17].

References

[1] Naedele, M., "IT Security for Automation Systems – Motivations and Mechanisms", ATP International, Vol. 1(1), 11/2003, http://www.tik.ee.ethz.ch/~naedele/publications.html

[2] Schainker, R., Douglas, J., Kropp,T., "Electric Utility Responses to Grid Securities Issues", IEEE Power & Energy Magazine, March/April 2006.

[3] Geer, D., "Security of Critical Control Systems Sparks Concern", Computer, Vol. 39, Issue 1, January, 2006, pps 20-23.

[4] Tani, M., "DOE Focuses on Cyber Security", Transmission & Distribution World, Vol 59, No. 3, March 2007, pps. 26-32.

[5] Naedele,M., "Addressing IT Security for Critical Control Systems", 40th Hawaii Int. Conf. on System Sciences (HICSS-40) Hawaii, January 2007.

[6] Coutinho, M.P., Lambert-Torres, G., da Silva L.E.B., Lazarek, H., "Detecting Attacks in Power System Critical Infrastructure Using Rough Classification Algorithm", Proceedings of the First International Conference on Forensic Computer Science, No.1, Vol.1, November 2006,pps. 93-99, Brasil.

[7] Amanullah, M.T.O, Kalam, A., Zayegh, A., "Network Vulnerabilities in SCADA and EMS", 2005 IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific, Dalian China.

[8] Bigham, J., Gamez, D., and Ning Lu, "Safeguarding SCADA Systems with Anomaly Detection", V.Gorodetsky et al.(Eds.):MMM-ACNS 2003, LNCS 2776, pp. 171-182, Springer-Verlag Berlin Heidelberg, 2003.

[9] Pawlak, Z., "Rough Sets", International Journal of Information and Computer Sciences, Vol.11, pp. 341-356, 1982.

[10] Goetz, E., "Cyber Security of the Electric Power Industry", Institute for Security Technology Studies at Dartmouth College", December, 2002 [NERC-CIP,2005] http://www.nerc.com/~filez/standards/ Reliability_Standards.html#Critical_Infrastructure_Protection.

[11] Gjermundrod, K.H., Dionysiou, I., Bakken,D., Hauser, C., Bose, A., "Flexible and Robust Status Dissemination Middleware for the Electric Power Grid", Technical Report EECS-GS-003, School of Electrical Engineering and Computer Science, Washington State University, September 25, 2003, Pullman, Washington, USA, http://www.gridstat.net/publications/GridStat-EECS-GS-003.pdf.

[12] Xuan Jin, Bigham, J., Rodaway, J., Gamez,D., Phillips, C., "Anomaly Detection in Electricity Cyber Infrastrucuture", Proceedings of CNIP, 2006, http://www.davidgamez.eu/pages/publications.html

[13] Chengdong Wu, Yong Yue, Mengxin Li, Asei Adjei, "The rough set theory and applications", Engineering Computations, Vol. 21, No.5, 2004, pp 488-511, Emerald Group Pub. Limited, UK.

[14] Pawlak,Z., Skowron,A., "Rudiments of Rough Sets", ScienceDirect, Information Sciences 177(2007)3-27, www.sciencedirect.com.

[15] Wood, A.J., Wollenberg, B.F., "Power Generation Operation and Control", 2nd Edition, John Wiley & Sons, Inc., 1996.

[16] Martinelli, M., Tronci, E., Dipoppa, G., Balducelli,C., "Electric Power System Anomaly Detection Using Neural Networks", M.Gh. Negoita et al. (Eds.), KES 2004, LNAI 3213, pp. 1242-1248, 2004, Springer Verlag Berlin Heidelberg.

[17] Bigham, J., Gamez, D., Xuan Jin, Chris Phillips, "Test Data for Anomaly Detection in the Electricity Infrastructure", International Journal of Critical Infrastructures, Volume 2, Number 4/2006, pp. 396-411.