A Rough Set Classification Algorithm for Detecting Attacks on Electric Power Systems and Other Critical Infrastructures

Maurício Pereira Coutinho, Germano Lambert Torres, Horst Lazarek, and Luiz Eduardo Borges da Silva

Abstract—Critical infrastructures (CI) play a fundamental role in modern society. Our reliance on information technology (IT) to provide quality service, however, brings vulnerabilities and security threats. To safeguard against cyber-attacks, CI providers need to ensure the integrity of their interdependent data networks. This paper presents a novel technique for improving the security of electrical power systems (a common type of CI), one capable of identifying both deliberate attacks and accidental faults. Using the Rough Set Classification Algorithm a set of rules can be defined to describe and evaluate the state of the system, including anomalous events such as attacks and failures. A test environment and simulated six-bus power system are used to evaluate the proposed Anomaly Detection System, with good results.

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

1. Introduction

Electricity management networks rely heavily on corporate networks, the Internet, and information technology to provide quality service. At the same time, this technology brings new vulnerabilities and exposes the electricity infrastructure to security threats [1,2,3,4]. Deregulation of the electricity market has increased usage of commercial, off-the-shelf technologies such as standardized TCP/IP networks and fully networked systems, further multiplying the opportunities for cyber attacks [5]. The authors have previously discussed current initiatives for improving the security of electricity systems and other Critical Infrastructures [6].

Supervisory Control and Data Acquisition (SCADA) systems and Energy Management Systems (EMS) play a vital role in monitoring the safety, reliability, and protective functions of power grids. However, these systems are designed to maximize functionality. Little attention is paid to security, a potential vulnerability of the grid. It is important to recognize that disruptions of service and manipulations of operational data are also a public safety concern [7].

Manuscript received September 11, 2007. This work was supported in part by the Brazilian Research Council (CNPq) and Minas Gerais State Research Foundation (FAPEMIG).
There are two approaches to making SCADA systems more secure [8]. First, anti-virus and Intrusion Detection Systems (IDS) can be used to identify problems at the perimeter of the system. Second, normal data flows and control operations can be modelled and monitored within the SCADA system. Operational anomalies may signal external attempts to change or damage the system.

This paper presents an implementation of the latter technique intended for Electric Power Systems (EPS), as previously introduced by the authors in public fora [6]. The problem of anomaly detection is addressed using the Rough Sets Classification Algorithm proposed by Pawlack et al. [9].

This paper is organized as follows. Sections 2 through 4 present an overview of Electric Power Systems, SCADA systems, and the nature of the security problem. Section 5 introduces the Rough Sets Classification Algorithm. Section 6 describes the architecture of the Anomaly Detection System, our method for building the knowledge base, and how to extract rules from this information. Section 7 applies the method to a six-bus power system by way of example, and Section 8 concludes.

2. Electric Power Systems

An Electric Power System (EPS) is generally highly interconnected and dynamic, encompassing several utilities. It has a hierarchical organization, sub-divided into regional grids. Each of these sectors is further split into generation, transmission, distribution, and customer service systems, as well as a supplemental energy trading system. The power grid comprises myriad material assets such as power plants, transmission lines, transmission and distribution substations, control centres (local, regional and national), Remote Terminal Units (RTUs), Intelligent Electronic Devices (IEDs), and communication links [10].

Figure 1 divides the computer network of the electricity infrastructure into two components: Energy Management Systems (EMS) for regulating the power flow, and Supervisory Control and Data Acquisition (SCADA) systems for monitoring the safety, reliability, and protective functions of the grid. This figure also illustrates the interactions between various grid entities [11].

3. Protecting SCADA Systems

SCADA systems collect data from sensors at remote sites and display the data at a control centre. These systems can monitor and control hundreds of I/O points. Remote Terminal Units (RTUs), located between the sensors and the control centre, help gather data from field devices. The sensors themselves may have a digital or analogue interface, but their signals are not usually in a form that can easily be communicated over long distances. The RTUs digitize sensor signals so that they can be transmitted to the control centre via standard communication protocols. Figure 2 illustrates this organization.

SCADA systems employ a diverse range of wired (leased lines, dialup line, fibre optics, ADSL, cable) and wireless (spread spectrum, cellular, WLAN, satellite) communication media. The exact choice depends on the utility’s characteristics, and where possible capitalises on existing communications infrastructure.

Analyzing Figure 2, we can identify several weak points where internal or external attacks might gain access to the SCADA Master and RTU. For example, the RTU might be attacked via Internet (through the corporate network) or the public telephone network. A person with access to the RTU could assume control over its circuit breaker...
or corrupt the information it collects. In order to prevent such scenarios, anomaly detection techniques are used to identify these threats as well as the type of attack.

Intrusion Detection Systems (IDS) have been studied widely in recent years. This method discovers attacks by identifying unusual behaviour (anomalies) in the host network or application; it thus assumes that some attackers will take notably different actions or cause unusual events. An extensive bibliography on IDS is presented in reference [6].

Intrusion Detection Systems (IDS) have been studied widely in recent years. This method discovers attacks by identifying unusual behaviour (anomalies) in the host network or application; it thus assumes that some attackers will take notably different actions or cause unusual events. An extensive bibliography on IDS is presented in reference [6].

Fig. 2: SCADA System Communication Model

4. Problem Definition

EPS operation is intrinsically complex due to the high degree of uncertainty and the large number of variables involved. Supervision and control of the system require a human operator, who must respond efficiently to diverse requests and alarms by handling various types of data.

These data may come from SCADA system measurements or computational processes. The typical database in a power control centre has increased tremendously in size over the past few years, largely due to an increased reliance on network communications. The control systems are thus more vulnerable to manipulation by intruders. Anomaly detection algorithms can improve the security of SCADA systems by identifying corrupted values, which may be caused by malicious attacks.

In addition to monitoring the current state of the system, the operator must react to future constraints such as load forecasts and maintenance schedules by taking a control action (switching, changing taps, modifying voltage levels). For all practical purposes, the current operational state of the system is defined entirely by data flowing to the operator.

By analysing these data, the operator attempts to classify the system as either normal or abnormal. In a “normal” state, all loads are supplied and all measurements are inside their nominal ranges. In the “abnormal” state, some measurements are outside the nominal ranges and/or some loads are not supplied. The operator must regularly verify the system’s security, even when its operational state is normal. This analysis is conducted by investigating possible contingencies that could affect the power system.

According to Bigham et al. [12], there are two ways for anomaly detection technology to enhance the integrity and security of EPS data. First, it can complement existing techniques such as state estimation by quantifying the likelihood that electricity measurements are correct. In other words, they can provide constant feedback on the integrity and reliability of the data being received. Second, they can be used to improve standard security devices such as IDS and virus checkers.

5. Rough Sets Classification Algorithm

Rough Set Theory, developed by Pawlak [9], is used to manage uncertainties from inexact, noisy and incomplete information. Since its advent, it has become a focal point of research in artificial intelligence [13]. A more recent work by Pawlak [14] repeats the basic concepts of rough set theory and points out some current research directions and applications.

Before presenting the algorithm, we must define two major concepts of Rough Set theory: \textit{reduct} and \textit{core}. These concepts are important in the knowledge of base reduction.

Let $R$ be a family of equivalence relations. A \textit{reduct} of $R$, $\text{RED}(R)$, is defined as a reduced
A set of relations that preserves the inductive classification of set $R$. The core of $R$, $\text{CORE}(R)$, is the set of relations that appear in all possible reducts of $R$ (i.e., the set of indispensable relations characterising $R$). These ideas are used in knowledge base reduction, the simplification of a set of examples into underlying rules. This is accomplished by the following procedure:

a) Calculate the core of the problem;
b) Eliminate or substitute a variable; and
c) Redefine the problem in terms of the new basic categories.

An algorithm to reduce the number of relevant conditions can be represented as follows:

**Step 1:** Eliminate dispensable attributes

**Step 2:** Compute the core of the set of examples.

**Step 3:** Compute a reduced set of relations that conserves the same inductive classification.

**Step 4:** Merge possible examples and compose the final set of rules.

6. Anomaly Detection Architecture

A solution to the problem pointed out in Sections 3 and 4 is presented in Figure 3. The proposed algorithm uses intelligent techniques (based on rough set theory) to extract knowledge from the SCADA system. Firstly, the knowledge extractor generates a set of rules that can be used to distinguish normal from abnormal behaviour. These rules are applied to the data coming from RTUs. Secondly, the anomaly detector attempts to determine whether an attack occurred and recognise its type.

In order to satisfy the limited computational resources of a SCADA Master, the number of input variables and examples should be reduced as much as possible. This will also result in a more compact set of rules for the anomaly detector.

As previously mentioned, the model uses Rough Set Theory for data reduction. This technique has several advantages:

- It can reduce the number of rules without reducing the system knowledge base;
- The behaviour is dynamic, and the expert (usually, a senior operator/engineer) can extract a set of new rules when he wants; and
- It requires few computational resources.

On the other hand, the technique needs a huge amount of data to build the knowledge base.

7. Experiments and Results

The diagram in Figure 4 represents a test environment for the proposed architecture. Its first three components (Power Flow, SCADA Simulator and State Estimator) are adapted from reference [15].

- **Power Flow:** This program solves for the power flow through a specified EPS.
- **SCADA Simulator:** This program simulates the data generated by a power system network, calculating all voltages, power flows and injections. It also associates these quantities with a measurement process.
- **State Estimator:** This program carries out a standard state estimation process.
• **Rough Set Rule Extractor**: This module uses the Rough Set Classification Algorithm to extract rules from the data.

• **Anomaly Detection System**: This module applies the rules defined by the Rough Set Rule Extractor to determine the state of the simulated SCADA data.

To test the proposed anomaly detection system (Fig. 4), we used the six-bus network model described by Wood and Wollenberg [15] and presented in Fig. 5. Test data were generated by introducing errors into a normal state input file generated by the SCADA Simulator. Xuan Jin et al. [12] considered 5 types of corruption in electricity data: (1) constant bias with normally distributed deviations, (2) loss of a decimal point, (3) sign reversal, (4) signal fixed at a constant value, and (5) signal fixed at a random value. They attribute these errors to random noise, attacks, software bugs, meter failures, Electromagnetic Interference (EMI) and transmission errors. The anomaly detection model and state estimator program were both applied to corrupted input data, and their results compared.

The knowledge data base contains 45 states, each consisting of 58 measurements provided by the SCADA simulator program with errors introduced. For the sake of clarity, in this example errors were applied only to Bus 4 and Bus 6 of the system. Furthermore, only sign reversal errors were considered. The Rough Set Rule Extractor generated the following rules:

- If Active Power on Bus 4 $\geq -0.77$, and Active Power on Bus 4 $< -0.2567$, and Active Power on Bus 6 $\geq -0.77$ and Active Power on Bus 6 $< -0.5133$, then the output condition is Normal.
- If Active Power on Bus 4 $\geq 0.4667$ and Active Power on Bus 4 $< 0.71$, then the output condition is Abnormal.
- If Active Power on Bus 6 $\geq 0.4667$ and Active Power on Bus 6 $< 0.71$, then the output condition is Abnormal.

This result demonstrates the great potential of data reduction: 3 simple rules were extracted from a database of 2565 measurements (45 states). All of the original 45 states were correctly identified by these three rules.

In order to evaluate the proposed anomaly detection model, new inputs were generated
by the SCADA simulator program (again using only sign reversal errors on Bus 4 and Bus 6).
The undisturbed state (or “base case”) has the following load flow values:

The SCADA Simulator creates a file containing power and voltage measurements at 29 selected points in the power system. (Recall that in real life, any field instruments would pass their data to the control centre via an RTU.) Table II shows six modified records derived from the base case in order to simulate corruption of the SCADA output file. Under a sign reversal error, the state estimation program generates incorrect values for all load flows.

### TABLE II
State Estimation outputs with errors introduced in the Base Case

<table>
<thead>
<tr>
<th>Error on Load Bus 4</th>
<th>Error on Load Bus 6</th>
<th>FROM 4 TO 1</th>
<th>FROM 4 TO 2</th>
<th>FROM 4 TO 5</th>
<th>FROM 6 TO 2</th>
<th>FROM 6 TO 3</th>
<th>FROM 6 TO 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Base Case</td>
<td>With Error</td>
<td>Base Case</td>
<td>With Error</td>
<td>Base Case</td>
<td>With Error</td>
</tr>
<tr>
<td>+56.0</td>
<td>-70.0</td>
<td>-42.53</td>
<td>-24.08</td>
<td>-32.04</td>
<td>6.32</td>
<td>-25.80</td>
<td>-65.10</td>
</tr>
<tr>
<td>-70.0</td>
<td>+56.0</td>
<td>-42.53</td>
<td>-41.62</td>
<td>-30.44</td>
<td>4.58</td>
<td>-25.80</td>
<td>-39.48</td>
</tr>
<tr>
<td>+63.0</td>
<td>-70.0</td>
<td>-42.53</td>
<td>-23.15</td>
<td>-37.04</td>
<td>8.10</td>
<td>-25.80</td>
<td>-65.44</td>
</tr>
<tr>
<td>-70.0</td>
<td>+63.0</td>
<td>-42.53</td>
<td>-41.58</td>
<td>-32.04</td>
<td>4.58</td>
<td>-25.80</td>
<td>-38.37</td>
</tr>
<tr>
<td>+63.0</td>
<td>+63.0</td>
<td>-42.53</td>
<td>-22.18</td>
<td>-32.04</td>
<td>1.51</td>
<td>-25.80</td>
<td>-45.09</td>
</tr>
<tr>
<td>+56.0</td>
<td>+56.0</td>
<td>-42.53</td>
<td>-23.26</td>
<td>-32.04</td>
<td>0.00</td>
<td>-25.80</td>
<td>-45.84</td>
</tr>
</tbody>
</table>

Fig. 7: Corrupted measurements file for the Anomaly Detector

| 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 -0.7000 -0.7000 -0.7000...
| 1.0500 1.0787 0.2066 0.5000 1.0000 0.6000 0.6000 +0.5600 -0.7000 -0.7000...
The SCADA output files (Fig. 6) produced from the corrupted states produce inputs for the Anomaly Detection model. Fig. 7 shows a sample of these measurements. Fig. 8 gives a Matlab program to implement the rules established by the Rough Set Extractor program. Fig 9 presents the output of this code.

```matlab
function fns()
    %getdata(input) Get Number of Inputs 'i'
    list=inpval(input,51);%extracts input data
    for i=1:length(data)
        if list(i,8)<-0.50 && list(i,10)<-0.25 && list(i,12)>0.50 && list(i,15)>0.50
            flag=1
        end
        if list(i,15)<-0.50 && list(i,10)>-0.25 && list(i,12)>0.50 && list(i,5)<0.50
            flag=1
        end
        if list(i,15)<-0.50 && list(i,10)>-0.25 && list(i,12)>0.50 && list(i,5)<0.50
            flag=1
        end
        if list(i,15)<-0.50 && list(i,10)>-0.25 && list(i,12)>0.50 && list(i,5)<0.50
            flag=1
        end
    end
end

guidance for i=1:length(data)
    if list(i,12)==0.50 && list(i,10)==0.75 && list(i,12)==0.50 && list(i,5)==0.50
        flag=1
    end
    if list(i,12)==0.50 && list(i,10)==0.75 && list(i,12)==0.50 && list(i,5)==0.50
        flag=1
    end
    if list(i,12)==0.50 && list(i,10)==0.75 && list(i,12)==0.50 && list(i,5)==0.50
        flag=1
    end
    if list(i,12)==0.50 && list(i,10)==0.75 && list(i,12)==0.50 && list(i,5)==0.50
        flag=1
    end
end

guidance
```

Fig. 8: Matlab Code to implement the rules

Fig. 9 demonstrates that the anomaly detector has correctly identified six abnormal states using SCADA measurements. Comparing this output with that of the usual state estimation process on Table II, it is possible to conclude that the latter could guide operators into taking wrong actions.

8. Conclusions and Future Work

Electric power systems are vital to modern society. While these systems require protection from numerous threats, their communication networks are particularly vulnerable to attack. The Anomaly Detection System proposed in this paper can increase the security of electric power systems and other critical infrastructures. A greatly reduced set of rules can be extracted from a knowledge data base (i.e., a large set of previously classified system states) using rough set theory. A test environment for the Anomaly Detection System was proposed and implemented, using a six-bus power system as an example. The system succeeded in extracting three rules and correctly identifying new states as abnormal. This technique is thus not only simple to implement, it has demonstrated favourable performance. Future improvements to this system will focus on expanding the range of recognised error types. It is also our intention to compare this technique to the “Test Data for Anomaly Detection in Electricity Infrastructure” method proposed in reference [17].

References

A Rough Set Classification Algorithm for Detecting Attacks...


Maurilio Pereira Coutinho received his B.Sc. and M.Sc. and D.Sc degrees in Electrical Engineering from the Federal University of Itajuba, Brazil, in 1981, 1998, and 2007, respectively. He is currently an Associate Professor at the Federal University of Itajuba (UNIFEI), Itajuba, MG, Brazil, 37500-903, (phone: +55-35-36291240; fax: +55-35-3629118755; e-mail: maurilio.coutinho@gmail.com). His current research interests are in Computer Networks, Intrusion Detection Systems and Data Protection and Data Security.

Germano Lambert Torres received his B.Sc. and M.Sc. degrees in Electrical Engineering from the Federal University of Itajuba, Brazil, in 1982 and 1986, respectively, and a PhD from École Polytechnique de Montreal, Canada, in 1990. He was a post-doctoral fellow at University of Waterloo, Canada, from 1995 to 1996. Currently he is a Professor at the Federal University of Itajuba (UNIFEI), Itajuba, MG, Brazil, 37500-903, (phone: +55-35-36291240; fax: +55-35-3629118755; e-mail: germanoltorres@gmail.com). His major field of study is Electrical Engineering with an emphasis in Power Systems

Horst Lazarek is a Professor at the Institut für Systemarchitektur, Fakultät Informatik, Technische Universität Dresden, Dresden, Germany. His current research interests are in Data Protection and Data Security. He is with the “Informatik in Entwicklungsländern” research group and he manages many Cooperation and Research Programs with South America, Africa and Russia.

Luiz Eduardo Borges da Silva received his B.Sc. and M.Sc. degrees in Electrical Engineering from the Federal University of Itajuba, Brazil, in 1977 and 1982, respectively, and a PhD from École Polytechnique de Montreal, Canada, in 1988. He was a post-doctoral fellow at University of Tennessee, USA, in 1990. He is currently a Professor at Federal University of Itajuba, Brazil. His major field of study is Electrical Engineering with an emphasis in Power Electronics and Digital Signal