ANOMALY DETECTOR ALGORITHM ON SCADA

Nicoleta Daniela IGNAT[1]

Given the fundamental role of the Critical Infrastructures (CIs) in our modern society and their exposure to security threats of increasing level combined with potential social consequences, the protection of CIs becomes a matter of extreme importance. The providers of CI services need to maintain the security of their interdependent data networks by employing a more proactive security management process in order to prevent, detect, respond to and defeat potential harmful incidents. To this end, this paper reviews the mathematical approaches to model and simulate the CIs and presents an algorithm based on Rough Set theory to identify the anomalies appeared in Supervisory Control and Data Acquisition systems (SCADA) applied for the case study of the Electric Power System CI.

Keywords: Supervisory Control and Data Acquisition (SCADA), critical infrastructure, vulnerability, assessment, algorithm, anomaly

1. Introduction

Critical infrastructures (CIs) including the energetic system, telecommunication and transport services, banking and financial systems are organizations delivering goods and services which are fundamental to the well functioning of society and the economy. An impact within a CI will be felt by many, if not all people within the economy because CI are defined as systems and assets, whether physical or virtual, that are so vital to a macro-scale, that the incapacity or destruction of such systems and assets would have a debilitating impact on national economic security, national public health or safety or any combination of those matters. Critical infrastructure is so dependable that we tend to assume it in our equations around the resiliency of our lives and businesses. CI is a given, it is an assumption.

Events such as ice storms, hurricanes, pandemics, fuel shortages, border closures, food scares, banks ceasing can affect not just a single infrastructure, but often more than one because CIs have mutual interdependencies. That is, the CIs can be so interlinked with other adjacent CI within the same sector that impacts can conduct from one to the other almost seamlessly, without any regard of borders.

[1] Asist., Depart. of Economic Engineering, University POLITEHNICA of Bucharest, Romania, e-mail: nicoleta_ignat@yahoo.com

CI interdependency is a concept not necessarily a quantifiable tangible, especially if the intent is to apply the same metric across the CI. CI interdependency derives from proxy indicators and composite metrics, but the indicators and their management will become more and more effective with time.

2. Mathematical approaches in modeling and simulations (M&S) of critical infrastructures

The emerged M&S approaches to model the interdependent infrastructures are presented in the following framework [1].

1. Complex adaptive systems (CAS) are a complex collection of numerous heterogeneous individuals/entities that interact with each other and their environment, and adapt to change and evolve their behaviour. Their collective, systemic behaviour is emergent (i.e., it follow patterns that result are not yet analytically predictable from, dynamic, nonlinear, spatiotemporal interactions among a large number of component and subsystems [2]).

2. Leontif Input-output models and simulation (LIO). Y.Y. Haimes, P. Jiang [3] extended Leontief's economic input-output models to evaluate the risk of inoperability (it measures the joint effect of the probability and degree of the inoperability of a system) in interconnected infrastructures as a result of one or more failures subject to risk management resource constraints.

3. CASCADE model and simulation (CASCADE). I.Dobson, B.A. Carreras, D.E. Newman [4] pointed out that catastrophic disruptions of large, interconnected infrastructure systems are often due to cascading failure. Thus, a CASCADE model of probabilistic cascading failure of a system with many identical components randomly loaded is introduced.

4. Complex network structure models and simulations (CNS) indicate M&S based on characterizing network anatomy. Because structure affects function [5], for instance, the topology of social network affects the spread of information and disease, and the topology of power grid affects the robustness and stability of power transmission, CNSMS is getting more and more attention for research on complex systems.

5. Graph theory models and simulation (GT). Graph is a collection of nodes and edges. In Strogatz [5], ‘small-world, ‘scale-free’, ‘random generation’ three kinds of networks are analyzed by combining graph theory with nonlinear dynamics and statistic physics. GT identifies system vulnerability and provides an aid for design and hardening networks to more stable and reliable (error-tolerant) ones. As an application, Leon [6] uses graph theory to model electric power control and protection devices and their associated connectivity.

6. Generalized Stochastic Petri Nets models and simulation (GSPN) is a suitable method to formalize and simulate dynamic aspects of complex systems,

describing the semantics and activity of workflow systems. GSPN have been used by researchers for performance analysis, workload mapping, identifying and modeling network invariants, and modeling interconnection structures. Kring and Oman [7] introduced using GSPN to identify the underlying causes (common mode failure) behind massive cascading failures of complex system, while [8] addresses to assessments and implement survivability mechanisms and mitigate common mode failures using GSPN.

7. Chaos theory models and simulation. The presumption that experiments are predictable and repeatable is fundamental to science. Thus, it is surprising when simple deterministic systems were found that were neither predictable nor repeatable. Instead, they exhibited chaos, in which the tiniest change in the initial conditions produces a very different outcome, even when the governing equations are known exactly [9], i.e. unpredictable behavior of deterministic system.

8. Hierarchical holographic modeling (HHM): is a modeling schema suggested by Haimes [10]. Its name derives from holography — the technique that can produce stunning three dimensional images which can be viewed in normal light without glasses.

9. The Rough Set Theory – a powerful technique based on the database knowledge extraction, created by Zdzislaw Pawlak [11]. This theory is based on finding a set representing the examples (data set) through two approximation sets named the upper approximation set (that must be reduced) and the lower approximation (that should be expanded). The idea is to represent the final set through a set of production rules that can detect intrusions into the system.

3. Anomalies in electric power system

The electric system can be considered as one of the most sensitive critical infrastructures for the correct functioning of other infrastructures. Despite the existence of backup resources and other procedures for continuing operations if the electricity sector collapses, communication are disrupted, trains stopped, planes grounded, economy may be impaired, banking transactions stops. This system is dynamic and interconnected, consisting of utility companies, private or public. It has a hierarchical system of control and it is subdivided into more regional system. In addition, there is the subdivision of the sector in the generation, transmission, distribution and marketing of electric power market.

The electric power grid is controlled by the exchange of control signals between control centers and the RTUs and IEDs, which in turn control circuit breakers, transformers, switches etc. The task of data acquisition and supervisory control are performed through the Supervisory Control and Data Acquisition (SCADA) systems, but the data are, in most cases, incomplete and subject to being corrupted.

SCADA systems are critical to electricity sector and interact with the control centers. This is specialized software to interface with the hardware units, such as Remote Terminal Units (RTUs) and Intelligent Electronic Devices (IEDs), which in their turn control sensors and interface with the various physical devices from the electrical system. Some examples of this type of devices include: circuit breakers, breakers switches, transformers, protection relays, etc.

There are two types of anomaly intrusion detection that can be defined: i) by identifying attacks that use the infrastructure of the data communication network and ii) by modeling the data flow and control operations in SCADA systems in order to detect anomalies caused by attempts to cause damage to the system, such as changes in the amounts of transmitted data, change of control signals, opening breakers, fraud.

In this article [12], it is proposed a complementary system, wide area measurement systems (WAMS) consisting of advanced measurement technology, information tools, and operational infrastructure that facilitate the understanding and management of the increasingly complex behavior exhibited by large power systems. The use of numerical methods in monitoring of electrical power system consumes time and resources, which can be a disadvantage for online monitoring. Nowadays, power electrical systems use an application named State Estimator (SE) [12] to deal with these issues. Since the SE cannot work well with large data losses, it assumes that its information on the network is always correct. This is a risk assumption, because there are configuration errors and there is always the chance than an attacker could be mediating between the control centre and the electrical system.

A solution for this problem is to implement an application to monitor and protect electrical power system in the case of cyber attacks using smart techniques, such as Rough Set Theory. The application must be capable of performing the online monitoring in power substation, collecting the measures from RTUs and informing the appearance of risks through the anomaly detector, as shown in the figure below.

This proposal of anomaly detector can be undertaken in two steps: implementing a classifier to detect corrupted and normal measurements and another one for implementing a classifier for type of attack or error injected. The detector should trigger an alarm in the presence of any abnormality.

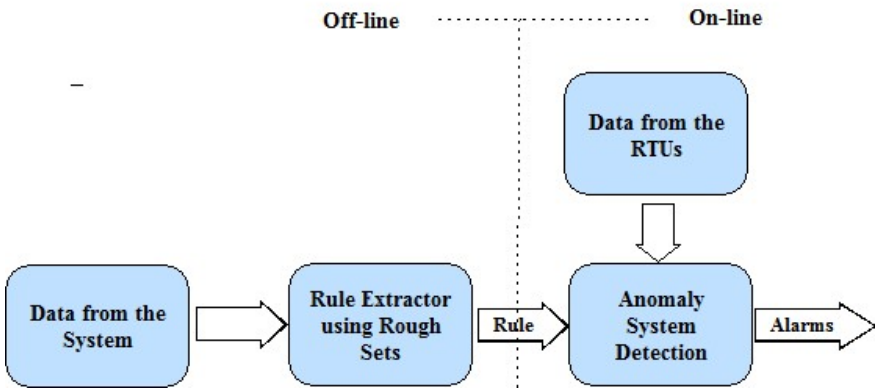

Fig. 1. Model of the proposed anomaly detector

4. Anomaly Detector Algorithm

The basis of the anomaly intrusion detector is the Rough Sets algorithm. This algorithm is defined as:

$$S : A \times O \to B \tag{1}$$

where S is the total function or information function, A is a finite set of attributes (normal or abnormal behavior), O is a finite set of objects and B is the domain of each attribute.

Let $X \subseteq O$, be a target set to represent using attribute subset A, an equivalent relationship and $S = (O, \{A\})$, a knowledge base. The target set X can be *approximated* using only the information contained within A by constructing the A-lower and A-upper approximations of X:

$$\text{A-lower: } \underline{A}X = \cup\{Y \in O / A : Y \subseteq X\} \tag{2}$$

$$\text{A-upper: } \overline{A}X = \cup\{Y \in O / A : Y \cap X \neq \emptyset\} \tag{3}$$

The elements belonging to $\underline{A}X$ set (lower approximation) certainly belong to the solution sought, while the elements of $\overline{A}X$ set (upper approximation) may belong to the solution. There are three regions created: positive region, negative region and boundery region. These regions are defined mathematically as it follows:

The positive region = $\underline{A}X$ and all points of this area are member of X.

The negative region = $O - \overline{A}X$ and all points of this area are certainly not member of X.

The boundery region = $\overline{A}X - \underline{A}X$ and the points in this area can or can't be member of X.

Because the data obtained using this method is imprecise when the boundary region is nonempty, it is necessary to use other two concepts: reduction set and core set. The reduction set of attributes, RED (A) is a reduced set of relationship that remembers the same inductive classification of set of A, while the core set of attributes, CORE (A), is defined as a set of relationship that appears along the reduction of A.

The following actions help to simplify the set of samples: i) calculation of the core set of the problem; ii) deleting or replacing a variable using another and iii) redefining the problem using new basic categories. An algorithm that follows this procedure can be represented by the following steps:

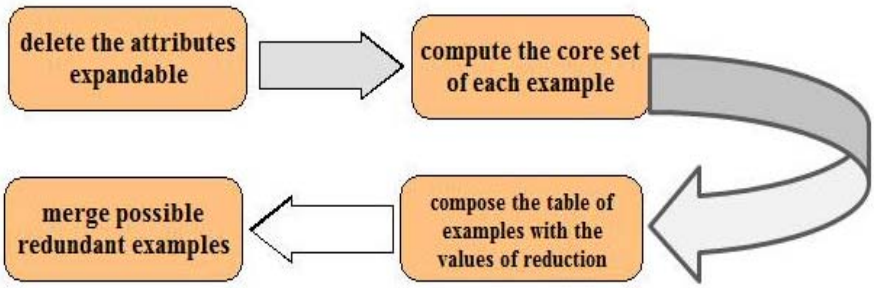

Fig. 2. Proposed algorithm

The solution for the problem of detecting possible anomalies uses intelligent techniques to extract knowledge from the SCADA system. This approach is divided into 2 steps.

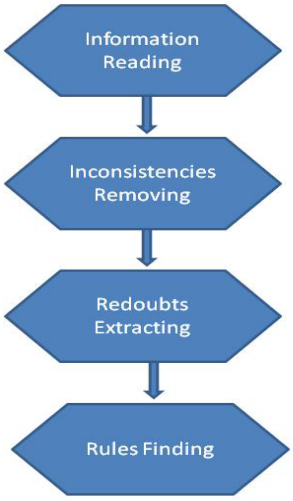

Fig. 3. Detecting anomalies using the proposed algorithm

The first step, named the knowledge extractor, generates a set of rules that determine the normal or abnormal system behaviour. The set of rules is obtained though the information collected in an offline SCADA data treatment (measurements from different parts of the system, state of the switches and the transformer taps) by the analysis of an expert, to determine its normality or not.

In the second step, the data from the real-time RTU passes through this set of rules, thereby defining normalcy or not of the collected information. The anomaly detector established by the rules should recognize the condition and may also undertake the abnormality with some kind of classification of the type of attack occurred [11].

Fig. 3 indicates how the rules extracted previously using the Anomaly Detector Algorithm should recognise anomalies in case of an attack. The large volume of data is analysed using a compact set of rules, but without reducing the knowledge base of SCADA systems. The proposed anomaly detector use a reduced set of rules extracted from an Electrical Database Knowledge with the help of Rough Classification Algorithm in order to find corrupted data.

5. Conclusion

The Electric System Critical Infrastructure is highly sensitive to a variety of threats and vulnerabilities. As the data obtained by SCADA systems are in most cases, incomplete and subject to being corrupted, protection from cyber attacks is required.

An algorithm is presented here in regards to the Anomaly Detection in SCADA systems. In order to provide a compact set of rules for the anomaly detector, a decrease in the number of input variables and the number of cases is suggested. Rough Set theory was employed to generate a model for the anomaly detector.

The anomaly detector proposed in this work eliminates the necessity for large computational resources and large memory capacity, while the number of rules is decreased without reducing system knowledge database.

The obtained results show that the proposed anomaly detector show high potential in increasing the reliability of the Electric Power System Critical Infrastructure by increasing the security of the Control Center Networks and the trustiness of the data coming from the SCADA System.

REFERENCES

[1] *N. Ignat, F. Stanculeanu "An overview on Modeling and Simulation of Critical Infrastructures", Proceedings of* 4th International Conference of Management and Industrial Engineering, 05-06 nov. 2009, Bucharest, ROMANIA

[2] *P. Coveney, R. Highfield,* 1995, cited in M. Heller, "Interdependencies in Civil Infrastructure Systems", The Bridge, **vol. 31**, no4: Frontiers of Engineering, 2001

[3] *Y.Y. Haimes, P. Jiang,* "Leontief-based model of risk in complex interconnected infrastructures", Journal of Infrastructure Systems, March 2001, pp. 1-12.

[4] *I. Dobson, B.A. Carreras, D.E. Newman,* "A probabilistic loading –dependent model of cascading failure and possible implications for blackouts", Hawaii International Conference on system sciences, IEEE, January 2003.

[5] *S.H. Strogatz,* "Exploring complex networks", Nature, **vol. 410**, pp. 268- 276

[6] *D. C. Leon, J. Alves-Foss,* Modeling Complex Control Systems to Identify Remotely Accessible Devices Vulnerable to Cyber Attack, 2002

[7] *A. Krings, P. Oman,* A simple GSPN for modeling common mode failures in critical infrastructures, Proceedings of the 36th Hawaii International Conference on System Sciences, 2002

[8] *F. Sheldon, T. Potok, A. Loebl, A. Krings, P. Oman,* Managing Secure Survivable Critical Infrastructures To Avoid Vulnerabilities, Proceedings Eight IEEE International Symposium, 25-26 March 2004, pp 293-296

[9] *J.C. Sprott,* "Chaos and Time-series analysis", Oxford University Press, first published 2003, Retrieved from:
http://sprott.physics.wisc.edu/chaostsa/

[10] *Y.Y. Haimes,* Risk Modeling, assessment, and management, Willey, Third Edition, University of Virginia Charlottesville, Virginia, 2009

[11] *C. Laing, A. Badii, P. Vickers,* "Securing Critical Infrastructures and Critical Control Systems ", Information Science Reference, 2013, pp. 34-36

[12] *N. Arghira, D. Hossu, I. Fagarasan, S. S. Iliescu, D. R. Costianu*, "Modern SCADA Philosophy in Power System Operation- a Survey", Scientific Bulletin, University POLITEHNICA Bucharest, Series C: Electrical Engineering, **vol. 73**, nr. 2, ISSN 1454-234X, 2011