Enhancing Cyber Security in Power Sector using Machine Learning

R Prabhakaran, S Asha

Abstract: ***Nowadays, our lives have become very much dependent on the power systems, whether it is in home or in offices or anywhere. Any failure in the power systems can bring our lives to a halt. To ensure no power fault, a continuous and remote monitoring, control and automation are needed. The implementation of constraints increases the efficiency of the power systems. But, to put monitoring, control and automation into practice we need network, and with this come the threat of cyber-attacks. With more open standard-based communication network, the automated power systems have become the target of the cyber-attacks. By exploiting the cyber components in networks, critical cyber components can be manipulated. Intruders can tamper the communication links by injecting false or modified data. To come up with security measures against these attacks, vulnerabilities of the power systems are being assessed to analyze the impacts of the cyber-attacks. Several techniques have been implemented so far to make the power systems less prone to threats. In this paper, technology like Machine Learning is used as anomaly discriminator and to provide security to the power system against the cyber threats.***

Index Terms: Smart Grid, SCADA, IEC 61850, cyber-attack, security.

I. INTRODUCTION

Over a decade, the digitization of the electronic enterprise has grown exponentially. The consumers are transforming each and every aspect of their lives into the digital purview. Along with this evolution, the power grids or the electric grids are also developed. The European Technology Platform has defined the smart grids as "an electricity network that can intelligently integrate the actions of all users connected to it (generators, consumers and those that do both) in order to efficiently deliver sustainable, economic and secure electricity supplies" As digital communications became a viable option, data acquisition systems (DAS) were installed to automatically collect measurement data from the substations. A smart grid is made of substations that are interconnected with lines and wires for the distribution and transmission of electricity to the consumers. In the smart grid all the in substations functionalities are executed by the supervisory control and data acquisition (SCADA) system. The modern substation automation system (SAS) constitutes Intelligent Electronic Devices (IEDs) and other equipment which allow local and remote access to the substations and communication links of the power system.SCADA system and substation automation decreases the cost of power transition increases the efficiency and smart grid. It is an automation control system that is used in industries such as energy, oil and gas, water, power and many more. The system has centralized system that monitors and controls entire sites, ranging from an industrial plant to a complex of plants across the country. The system works by operating with signals that communicate via channels to provide the user with remote controls of any equipment in a given system. It also implements a distributed database that contains tags or points through the plants. These points represent a single input or output value that is monitored or controlled by the SCADA system in the centralized control room. The points are stored in the distributed database as value-timestamp pairs.When it comes to real-time operation of the power system, communication plays a vital role. Earlier the telephone was used at the substations to communicate line loadings back to the control center as well as to dispatch operators to perform switching operations. To have a smooth flow of the communication researchers came up with IEC 61850 families of standards.Due to the extensive setting out of related technologies, the risk of cyber-threat has increased. The cyber-attackers may gain access to the components or systems in the substations through cyber intrusions. The objective of the intruders or attackers is always to maximize the loss for the substations. For that they might form different resource allocation approaches. They may send trip signals, false status reports, fabricated operations and measurements to trick the system operator. To take necessary measure against the attacks and to secure the components of the system, assessment of strength and impacts of the attacks is looked for.Among the state-of-the-art technologies, Machine Learning is one such technology which is being used in many domains like finance, health care, data science, image recognition, speech recognition, etc., and is used by corporate giants like Google, Facebook, Amazon, etc. Machine Learning is a sub-part of Artificial Intelligence. It provides the machines the ability to learn to perform and improve a specific task from experience without being programmed or human intervention. Mathematical models are made based on the data given to the machine learning algorithms, called the training data, from which the machine learns patterns based on which it makes decisions and predictions in the future.

The following are the types of learning: supervised, unsupervised, semi-supervised and

Revised Manuscript Received on July 07, 2019.

Prof.R. Prabhakaran, Assistant Professor (Senior) from the School of Computing Science and Engineering, VIT University, Chennai.
Dr. S. Asha, Associate Professor from the School of Computing Science and Engineering, VIT University, Chennai.

reinforcement. This categorization has been done based on the type of problems that they are expected to resolve, their approaches and the type of data given as input and the output.

In this paper, the potential of Machine Learning has been used to discriminate the events in the power system. The data which is being collect from the power system is used to classify the events as attack, natural or no event. The remaining of the paper have the following: section II gives the literature survey, section III shows the methodology of implementation, section IV shows the result and section V gives the conclusion and future work and section VI contains the references.

II. LITERATURE SURVEY

To impact reliability of the power system, attackers can opt for command injection attacks. These attacks send false commands or signals to take the control over and configure the control system of the power station. Only by doing so, the components of the target substation can be abnormally tripped. The attackers may also inject false supervisory control commands to the SCADA system. And ultimately, the attackers become successful in effecting the reliability of the power system. The communication protocols in the SCADA network lack authentication features to check the source of command or data packets received by the network. Intruders take the advantage of this lacuna and may replace the command packets. The authors in [4] have assumed two types of cyber-attacks that can happen in the SCADA system, namely, normal attack and penetration attack. The normal attack is set of on the substation network because of its less complicated architecture. In penetration attack the attacks may occur on the control center which is able to send control commands to the substations. Compared with the substation network, the control center network implants more complicated defence strategies, which make difficult for attackers to attack successfully. A method of assessment is the Semi-Markov Process (SMP) which has a set of states, in which transition probability matrix controls the transitions of state. SMP is used to model the behaviours of systems in an extensive range. In [5] the authors have taken a SCADA architecture which constitutes control center, backup control center, corporation network and 24 substations. Each component of this architecture gives the intruders a path to get into the SCADA system. So the authors have taken four attack scenarios for cyber components of the SCADA system. Although a number of security countermeasures such as firewall or IDS are applied in each network vulnerabilities on large scale are available to the attackers for targeting the network. These attack scenarios illustrates how the intruders can penetrate into the networks. In the scenarios the attackers are able to send trip signals for tripping the components of the network abnormally. The scenarios are: 1) Attack on Control Center LAN, 2) Intrusion in Corporation LAN, 3) Attack against Substation Network, and 4) MITM Attack on Communication Links between the Substations and Control Center. The authors have considered two Bayesian attack graph models. First attack graph is of the graph of vulnerabilities depicting the probability of successful access through LANs of the control center and substations. The second Bayesian attack graph model evaluates the probability of successful intrusion on communication links.

In the Joint Substation-Transmission Line vulnerability assessment it is assumed that the substations, transmission lines or both are susceptible to attacks [6]. Earlier all other assessments were made only either on substations or transmission lines. With the introduction of this approach vulnerabilities related to joint substation-transmission lines are discovered. The investigation of the strength of the attacking targets has been conducted through Cascading Failure Simulator (CFS) in which the target is eliminated for triggering the cascading failures. The assessment results that joint-node-link vulnerabilities holds a great part of the whole vulnerabilities. They also came up with a metric, the component interdependency graph (CIG) and proposed CIG based attack strategy, which is proven to have stronger attack performance. The multidimensional Intrusion detection system (IDS) for IEC 61850 based smart substations has come up as an effective tool in detection of any attack on the substation [7]. It has been designed to detect any activity or complexity of communications by intruders. For security, establishment of a network which is protected by security control measures such as firewalls and setting up of monitoring mechanisms in a secured zone for the detection of any gaps and failures in the security control system. The implementation of this IDS is based on Linux system. The IDS has four dimensions: 1) Access-Control Detection, 2) Protocol Whitelisting Detection, 3) Model-Based Detection, and 4) Multi-Parameter based Detection. Alter-and-hide (AaH) attacks states the alterations in the values of a substation network without being detected by the defenders in the power system [8]. The activities by these attacks are programmed in software. On successful implantation of the software in the substation network the alarms on circuit breakers get blocked and avert warning the operators in the control room. The AaH attack execution is as follows: A controller server is installed in the targeted substation which can communicate across wide area network (WAN) with SCADA server. The attackers then try to get control over the administrative user interface to set the attack in the target. Once getting the control over the administrative user interface, the attacker now installs the programmed software. Now the attacker is having the control over the substation and can manipulate any component of the substation without being caught by the defenders as all the alarms are blocked and the control center system will get no report of the malicious activities. For every disturbance event there are alarms to the system operators. The cyber inference system (CyIS) is a spatiotemporal anomaly agent which combines all topologies and alarm information so that it can detect possible AaH attacks. Researchers are working on machine learning techniques to use it in the power system domain. Supervised machine learning algorithms are used in classification, and this feature can be used in decision making. For this the supervised learning algorithms need to be studied thoroughly.

For that in [2] the authors have evaluated and compared the performance of ten supervised learning algorithms using eight performance criteria. The algorithms are support vector machine (SVM), neural network, logistic regression, naïve bayes, memory based learning, random forests, decision trees, bagged trees, boosted trees and boosted

stumps. The performance metrics are accuracy, F-score, Lift, ROC Area, average precision, precision/recall break-even point, squared error and cross-entropy. Since some of the metrics gives probabilities and not all algorithms are designed to predict probabilities, Platt Scaling and Isotonic Regression are used to compare the performance of each algorithm, both before and after calibrating the predictions.In [1] the authors have used the potential of the machine learning in detection of threat scenarios of command and data injection. A number of machine learning algorithms are evaluated using the dataset collected from remote terminal units in which both normal and attack instances are included.Since human decision making in case of cyber-attack is less certain as the attacks, sometimes, come in disguise and look like normal event. So the authors in [3] have taken various machine learning algorithms to use it as a disturbance discriminator of power system. The practical implications for deploying the machine learning system in the power system architecture also has been discussed as an enhancement in security measure.

III. METHODOLOGY

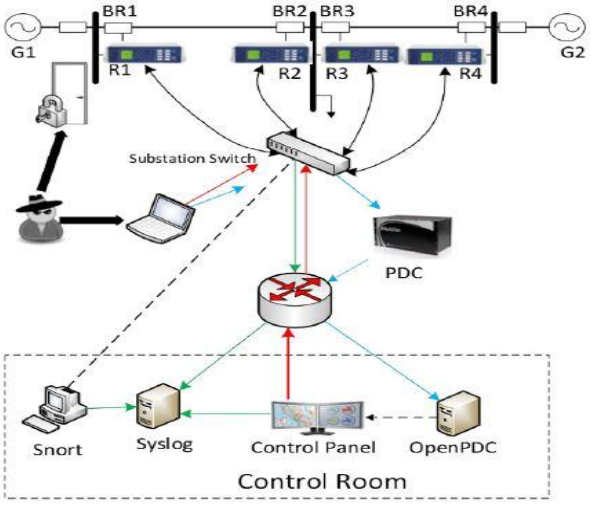
The section describes how supervised machine learning algorithms are used as disturbance discriminator in the power system. The system and the algorithm used for the evaluation are also discussed.

.

A.Power System Description

Fig. 1, shows the used power system framework. It constitutes generators, breakers, IEDs whose working is monitored by devices like Snort and Syslog. The breakers are controlled by IEDs. The IEDs relay information to the SCADA system through the substation router. With an assumption that the intruders have got access to the substation network and issuing tripping commands from the substation switch, the attack scenarios is made and simulated.

In Fig. 1, the generators are marked as G1 and G2, BR1 to BR2 are the breakers which are controlled by the IEDs R1 to R4. IEDs can switch the breakers on and off. There are two transmission lines, L1 and L2 which spans between buses B1 to B2 and B2 to B3 respectively. The IEDs use a distance protection scheme which, on detection of any faults, trips the breakers. Since there is no validation checking, the IEDs cannot differentiate between valid and faked faults.



Types of Scenarios:

1. *No event*- It is a normal scenario goes on a power system like normal operation load changes.

2. *Natural*- This scenario occurs when there is short in the power line or the breakers are disabled for line maintenance.

3. Attack- This scenario occurs when the system is being controlled by the intruders

In Figure-1 Power System Framework

B. Analytic Approach

To start, the machine learning algorithms are evaluated using Weka as machine learning tool and open-source simulated power system data provided by Mississippi State University [10]. There are 37 scenarios, which is grouped into three classes – No event (1), Normal event (8) and Attack event (28).

The dataset has 129 attributes and 4966 instances. To get more accurate result, the dataset has been reduced by using the attribute selection feature of Weka software. The dataset used for the final evaluation has 13 attributes. Three well known algorithms are tested using this dataset. For the testing, the dataset has been partitioned using the K-fold partition method with value K as 10, i.e. 10 sets of randomly selected instances from the dataset.

Classification algorithms that are tested:

1. *Naïve Bayes*- This classifier belongs to the 'probabilistic classifiers' family which applies Bayes theorem.
2. *Logistic Regression*- This belongs to the family of statistics.
3. *Multilayer Perceptron*- This classifier belongs to the family of artificial neural network. It distinguishes data which are linearly inseparable.

The system has been developed using Java on Netbeans 8.0.1. For the machine learning implementation in the system Weka packages are extensively used.

IV. RESULT AND DISCUSSIONS

Based on the evaluation of the classification algorithms, multilayer perceptron came out to have the more precision value. The result of the evaluation is shown in the graph in Figure-2.

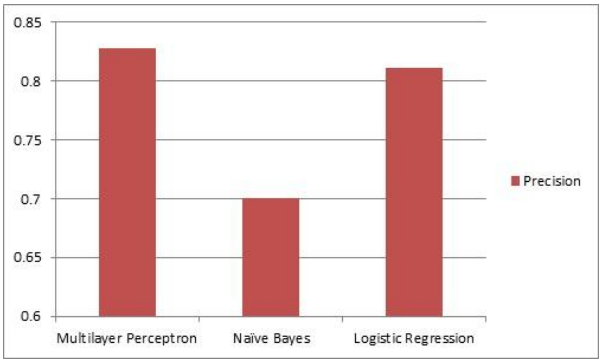


Figure-2 Graph based on the precision value of the classification algorithms

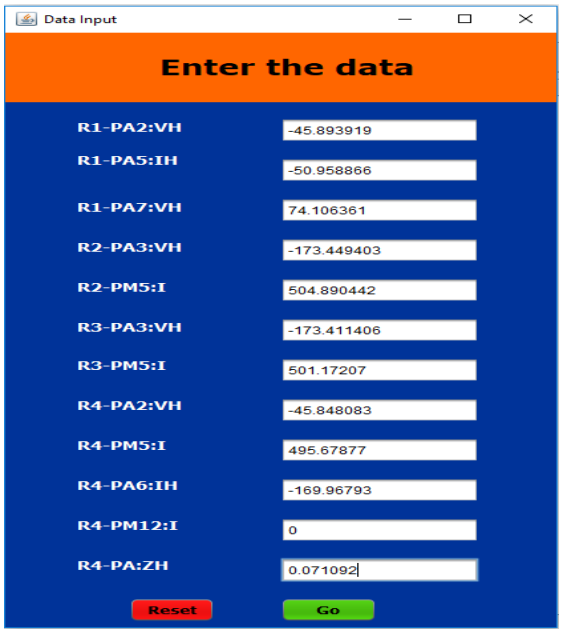


Figure-3 Input to the system

The proposed system has shown the practical use of Machine Learning technology as a decision maker. The dataset which has been taken for the project has been divided into training data and test data. A part of the data is used to train the model and the other part has been used to test the model. Based on the learning and experience, the model will predict the output of the data that will be given as input. There are three classes - No Event, Normal and Attack. These are the scenarios, one of which can be seen in any event that occur in the power system. In Fig.4 Based on the data events will be classified as one of the three. Based on the output further actions will be taken.

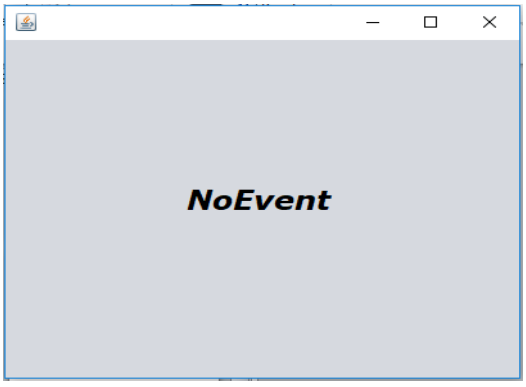


Figure-4 Classification of the event based on input to the system

V. CONCLUSION AND FUTURE WORK

For a reliable power system, it is necessary to provide a tight security system. It is also needed to make sure that the person who is allowed to access the system and the data is an authorized one. This project proposes a system using state-of-the-art technology, Machine Learning which can detect anomaly of the events that can take place in the environment where it is implanted. Based on the data received, the proposed system classifies the events occurring and based on result appropriate actions can be taken. With this system, it can be said that Machine Learning is a practicable way that can allow the control centre of the power system to make decision on whether the environment is has been attacked or not. As for the future research, the result of the system can be validated with other learning techniques, classification algorithms other attributes in the dataset. This proposed system can be taken as model on how machine learning can be applied in this field.

REFERENCES

1. Justin Beaver, Raymond Borges, Mark Buckner, “An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications”, 13th International Conf. on Machine Learning and Applications, 2013.
2. R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms,” Proceedings of the 23rd Intl. Conf. on Machine Learning, pp. 161-168, 2006.
3. Raymond C. Borges Hink, Justin M. Beaver, Mark A. Buckner Tommy Morris, Uttam Adhikari, Shengyi Pan, "Machine Learning for Power System Disturbance and Cyber-attack Discrimination" 7th International Symposium on Resilient Control Systems (ISRCS), 2014
4. Yichi Zhang, Lingfeng Wang and Yingmeng Xiang, “Power System Reliability Analysis with Intrusion Tolerance in SCADA Systems” IEEE Transactions on Smart Grid, 2015
5. Yichi Zhang, Yingmeng and Chee-Wooi Ten, “Power System Reliability Evaluation with SCADA Cybersecurity Considerations” IEEE Transactions on Smart Grid, 2015
6. Yihai Zhu, Yufei Tang, and Haibo He, “Joint Substation-Transmission line Vulnerability Assessment against the Smart Grid” DOI 10.1109/TIFS.2015.2394240, IEEE Transactions on Information Forensics and Security, 2015
7. Yi Yang, Lei Gao, Yu-Bo Yuan, Kieran McLaughlin, Sakir Sezer and Yan-Feng Gong, “Multidimensional Intrusion Detection System for IEC 61850 based SCADA Networks” DOI 10.1109/TPWRD.2016.2603339, IEEE Transactions on Power Delivery, 2016
8. Chong Wang, Chee-Wooi Ten, Yunhe Hou, Andrew Ginter, “Cyber Inference System for Substation Anomalies Against Alter-and-Hide Attacks” DOI 10.1109/TPWRS.2016.2574769, IEEE Transactions on Power Systems, 2016
9. Naiara Moreira, Elías Molina, Jesús Lázaro, Eduardo Jacob, Armando Astarloa, “Cyber-security in Substation Automation Systems” Renewable and Sustainable Energy Reviews54(2016)1552–1562, 2015
10. Mississippi State University Critical Infrastructure Protection Center, “Industrial Control System Cyber Attack Data Set”, Online: http://www.ece.msstate.edu/wiki/index.php/ICS_Attack_Dataset, Apr. 2014.

AUTHORS PROFILE

Prof.R. Prabhakaran is an Assistant Professor (Senior) from the School of

Computing Science and Engineering, VIT University, Chennai. He graduated in Computer Science and engineering – 1999 from Madras University, Chennai and Post graduated in Computer Science and engineering – 2006 from PSG College of Technology, Coimbatore, Tamil Nadu. He is pursuing his Ph.D from VIT University Chennai. His area of interest includes Network security, Biometric security, Computational Intelligence, Cloud security and Cyber security. He has published nearly 11 papers in international journal. His current research interests include power system reliability and resiliency, critical infrastructure protection and smart grid cyber security.

Dr. S. Asha

Dr. S. Asha, is an Associate Professor from the School of Computing Science and Engineering, VIT University, Chennai. She graduated from Madras University, Chennai and completed her Ph.D from Anna University Chennai. Her area of interest includes Network security, Biometric security, Computational Intelligence, Cloud security and Cyber security. She has published nearly 25 papers in international journal and conferences. Currently she is working in computational intelligence and Cyber security.