



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VI Month of publication: June 2021

DOI: <https://doi.org/10.22214/ijraset.2021.35232>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Flexible Machine Learning based Cyberattack Detection using Spatiotemporal Patterns for Distribution Systems

Ankita Singh¹, Kumari Divya², Likitha N³, Sai Ganesh Sharath S⁴, Dr. Devaraju R⁵

^{1, 2, 3, 4}Department of Electronics & Telecommunication, Dayananda Sagar College of Engineering (DSCE), Bangalore, India

⁵Associate Professor, Department of Electronics & Telecommunication Engineering (DSCE) Bangalore, India

Abstract: The Article presents a versatile machine learning detection technique which is employed in distribution systems for cyberattacks considering spatiotemporal patterns. Spatiotemporal patterns are identified by the graph Laplacian which are supported on system-wide measurements. A versatile Bayes classifier is employed to coach spatiotemporal patterns which may well be compromised when cyberattacks happen. Cyberattacks are spotted by utilizing flexible Bayes classifier online.

Keywords: Cyber-attack detection, Machine Learning, Distribution Structures, Graph Laplacian, Spatiotemporal patterns.

I. INTRODUCTION

The growing usage of distributed energy resources (DERs), microgrids, and other distribution-level technology and assets has altered the way during which the distribution systems are created and utilized traditionally. As many sensors are developed on the distribution system with the standard SCADA systems, Advanced Metering Infrastructure (AMI), and other field devices activate data-powered observability and grid-edge data analytics, the attack surface to the Distribution Management System (DMS) is enlarged. DMS and associated monitoring and control systems are the key components for creating decisions and exchanging information. Nevertheless, existing cybersecurity technologies used in distribution structures are still liable to cyberattacks. It's important to create cyber-resilient DMS functions and cybersecurity automations to enable future energy delivery systems to correctly spot, dynamically modify, survive and oppose a cyberattack. Unlike common cyberattack detection methods, like Naive Bayes classifiers (BCs) which are supported on the normality assumption, this document tries to

collect the continual feature of spatiotemporal patterns among system measurements by creating versatile BCs. Effectively, spatiotemporal patterns of measurement data under ordinary circumstances would be compromised when cyberattacks happen. Supported this concept, this document seeks to deal with two crucial questions for the cyberattack detection on distribution systems.

1) Can we quantitatively capture the spatiotemporal patterns between cyberattack situations and normal situations?

2) Can operators use versatile BCs to increase the accuracy of ordinary cyberattack detection methods?

In this document, we try to merge the spatiotemporal patterns of system measurements into a versatile BC for cyberattack detection. Concretely, spatiotemporal patterns are taken by the generalized graph Laplacian (GGL) matrix for system measurements. For the training process of the proposed flexible BC, they're taken as its input variables, while the labels of cyberattack templates are taken as its output variables. For the testing procedure, the web spatiotemporal patterns captured by GGL are put into the suggested flexible BC, which then generates the cyberattack detection results.

II. LITERATURE SURVEY

1) Yi Wang, Qixin Chen, Tao Hong, Chongqing Kang et al.[1] explain as follows about Meter Data Analytics Challenges

The increasing approval of smart meters helps a large amount of fine-grained electricity consumption data which is to be collected. Also, the removal of restrictions within the power sector, particularly on the side of delivery, has continuously been moving forward around the world. It's an important issue to know how to use very smart meter data to push and increase the efficiency and robustness of the ability grid. Till date, many works have been conducted on smart meter data analytics. To present a synopsis of the this research and also to acknowledge challenges for future, this article conducts an application-oriented review of smart meter data analytics. Following are the three stages of analytics: descriptive, predictive, prescriptive analytics. We recognize the important uses like load analysis, load forecasting, and cargo management. We also discuss some important research, like big data issues, machine learning technologies, the evolution of energy systems, and information protection and safety.

- 2) *Mingjian Cui, Senior Member, IEEE, Jian hui Wang, Senior Member, IEEE, and Meng Yue, Member, IEEE et.al[2] explains about Machine Learning Based Anomaly Detection for Load Forecasting Under Cyberattacks as follows :*

The economic and reliability benefits for power grid operators may be made by load forecasting. The cyberattack on load forecasting might not allow operators to create operational decisions to deliver the energy. To detect these cyberattacks accurately this document develops a machine learning based anomaly detection procedure. Firstly, load forecasts provided by neural networks are utilized to recreate the benchmark and scaling data by using the k-means clustering. Secondly, the cyberattack arrangement is estimated by the Naive Bayes classification supported the cumulative distribution function and statistical characteristics of the scaling data. Finally, the dynamic programming is employed to calculate both the incidence and framework of one cyberattack on load forecasting data. A well-known Symbolic Aggregation approximation technique is differentiated with the developed MLAD method.

- 3) *Siddharth Sridhar, Student Member, IEEE, and Manimaran Govindarasu, Senior Member, IEEE et.al[3] explains about Model-Based Attack Detection and Mitigation for Automatic Generation Control as follows:*

Cyber systems play a crucial role to boost the efficiency and reliability of grid operation and to confirm that the system remains within safe operating margins. On the other side it can produce significant damage to the underlying physical system by leaving the control and monitoring applications.

Critical assets are protected against electronic through ordinary cyber security measures that have host-based and network based security technologies. It's been seen that highly skilled attacks can bypass these security mechanisms to affect the operation of control systems.

There's a high need for cyberattack-resilient control techniques. During this article the subsequent contributions are made. Firstly the effect of information integrity attacks on Automatic Generation Control (AGC) on facility frequency and power market operation is demonstrated. A structure to the implementation of attack resilient control to power systems as a configuration of smart attack identification and alleviation is proposed. A model-based anomaly detection and attack alleviation methodology for AGC is developed.

The detection capability of the proposed anomaly detection algorithm through simulation studies is assessed. The results show that the algorithm can detect scaling and ramp attacks with little false positive and negative rates.

- 4) *Mingjian Cui, Student Member, IEEE, Jie Zhang, Senior Member, IEEE, Anthony R. Florita, Member, IEEE, BriMathias Hodge, Member, IEEE, Deping Ke, and Yuanzhang Sun, Senior Member, IEEE, et.al[4] explains about Optimized Swing door algorithm for Identifying Wind Ramping Events”,*

Wind power ramp events (WPRES) have begun damaging the economic and reliable operation of power grids with the increasing usage of renewable energy in recent years. This article will develop an optimized door algorithm (OSDA) to improve the WPRES detection. To upgrade the segments by merging adjacent segments with the identical ramp changing direction a dynamic programming algorithm is performed, handling wind generation bumps, post processing insignificant-ramps intervals. Measured wind generation data are used to gauge the execution of the proposed OSDA. Results show that the OSDA provides far better execution than the SDA and equal or superior performance compared to the L1Ramp Detect with Sliding window (L1-SW) method with much less processing time. 9

- 5) *L. Xu and D. Tretheway California Independent System Operator (ISO), Folsom, CA, USA, 2012 ,et.al[5] explains about Flexible ramping products as follows:*

The increasing amount of renewable generation has increased rapidly lately. This has led to worries which is linked with the system ramping capacity. The FRC model contains the demand curve of the ramping capacity, which represents the merit of the ramping capacity every hour.

Here, the versatile ramping capacity model is proposed that has the sensible ramping capability of generation resources and therefore the uncertainty in net load. The model is prepared mathematically using ramp controls. These are contained into Unit Commitment and Economic Dispatch procedure. To match the FRC model with ordinary methods, simulations are done employing a 10-unit system. System reliability is achieved even at alternative energy generation levels while achieving economic efficiency by using the FRC model.

A. Methodology

The steps for method of developed cyberattack detection is shown within the figure and explained as

- 1) Step 1) An unsupervised machine learning technique, named as GGL, which is employed to distinguish spatiotemporal patterns of system measurements.
- 2) Step 2) A supervised machine learning technique which is employed to coach the spatiotemporal patterns by the GGL matrix and
- 3) Step 3) Sets of 2 metrics, which are the correct positive rate and contingency table, are employed to find the execution of various detection techniques.

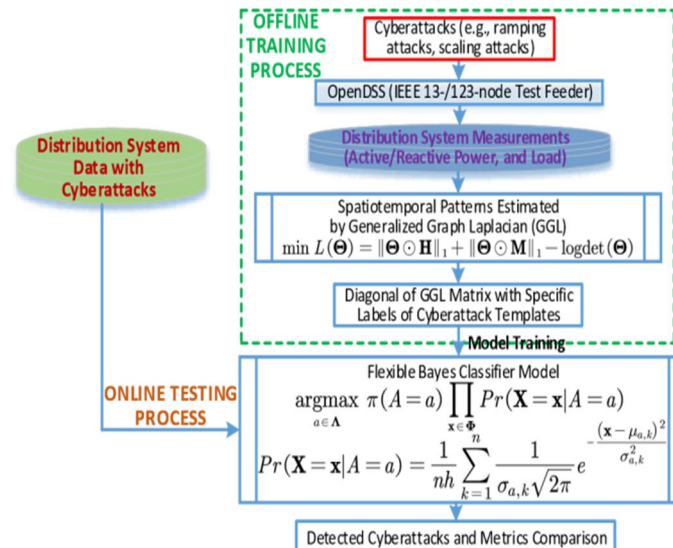


Fig. 1. Structural Outline Of the developed cyberattack detection technique.

- a) *Collecting Data:* Once we know exactly what we want and the equipment's are in hand, it takes us to the first real step of machine learning- Gathering Data. This step is very important as the amount and standard of information gathered will directly decide how effective the predictive model will turn out to be. The information collected is then tabulated and called as Training Data.
- b) *Data Preparation:* After the training data is collected, you move on to the next step of machine learning: Data preparation, where the information is loaded into a suitable place and then prepared for use in machine learning training. Here, the information is first put all together and then the order is randomized as the order of data should not affect what is learned.
- c) *Choosing a Model:* The next instruction that follows in the procedure is selecting a representation among the many that researchers and data scientists have created over the years. Make the choice of the correct one that would get the task finished.
- d) *Training:* After the above steps are done, you then move onto what is frequently considered the bulk of machine learning called training where the information is used to gradually improve the model's ability to predict. The training process involves initializing some random values for say A and B of our model, predict the output with those values, then compare it with the model's prediction and then adjust the values so that they match the predictions that were made previously.
- e) *Evaluation:* Once training is complete, you now check if it is good enough using this step. This is where that dataset you set aside earlier comes into play. Evaluation allows the testing of the model against data that has never been seen and used for training and is meant to be representative of how the model might perform when in the real world.
- f) *Parameter Tuning:* Once the assessment is over, any further development in your training can be possible by adjusting the parameters. There were a few parameters that were implicitly assumed when the coaching was done. Another parameter included is the learning rate that defines how far the line is shifted during each step, based on the information from the previous training step. These values all play a role in the efficiency of the training model, and how much time the coaching will take.
- g) *Prediction:* Machine learning is simply using data to answer questions. So this is the last instruction where you get to answer some queries. This is the point where the value of machine learning is observed. Here you can finally use your representation to predict the result of what you want.

The mentioned instructions take you from where you create a representation to where you predict its output and thus acts as a learning path.

III. OUTCOMES

Collecting data set like active and reactive power data in distribution structures that are vulnerable under cyberattacks situations. The info based are Neural Networks, SVM, Naive Bayes classification, Random Forest Classifier, Gradient Boosting, algorithms are classified. Application Of Preprocessing Task. Finding Out the Cyberattacks like Pulse, Scaling, Ramping, Random and Smooth Curve. The project will try and gauge the execution Of the proposed method as to extend its efficiency.

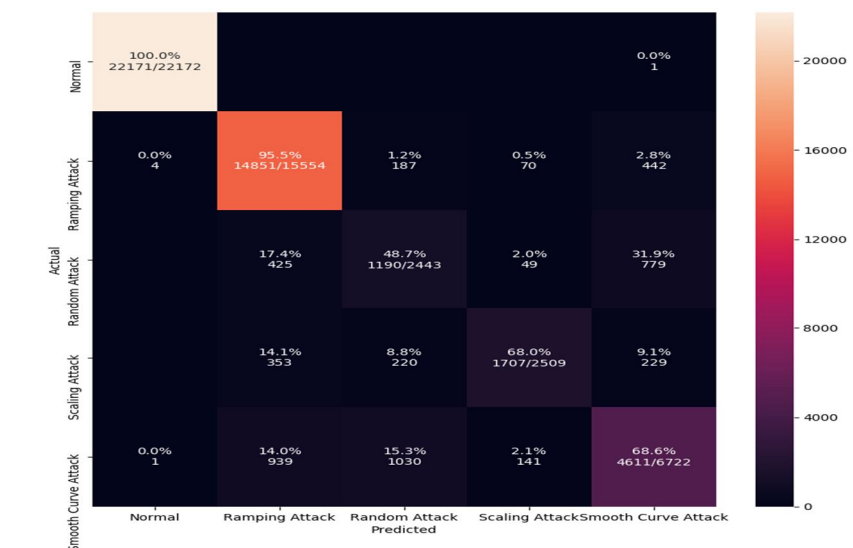
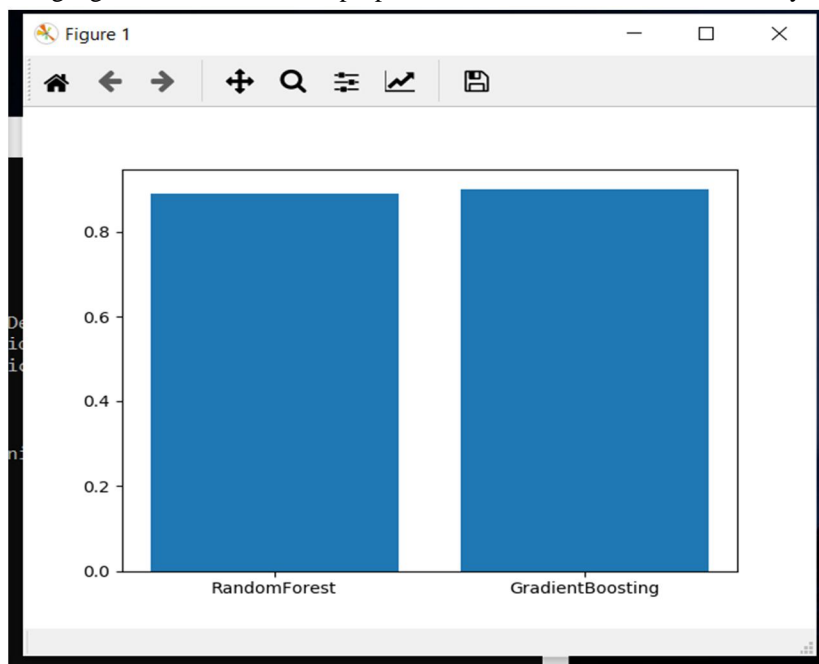


Table 1 Unit Test Case 1

S1 # Test Case	UTC- 1
Name Of Test	Dataset(Input L0ad cyberattack data)
Expected Result	In this step, we aim t0 build thr0ugh net t0 aut0matically d0wnl0ad the targeted l0ad cyberattack data fr0m the Internet and st0re in dataset (each attack categ0ry).
Actual Output	Same as expected.
Remarks	Successful

Table 2 Unit Test Case 2

S2 # Test Case	UTC- 2
Name Of Test	Pre-pr0cessing
Expected Result	It will pr0cess an uniquely identify Of each attack data and generate the accurate value.
Actual Output	Same as expected.
Remarks	Successful

Table 3 Unit Test Case 3

S3 # Test Case	UTC- 3
Name Of Test	Classificati0n in l0ad f0recasting data
Expected Result	When we train the l0ad spati0temp0ral patterns data then it will classify using machine learning and deep learning algorithms such as Rand0m F0rest, Gradient B0osting, AdaB0ost, Gaussian Naive Bayes, Deep Learning (CNN), and it can finally generate Train.M0del.
Actual Output	Same as expected.
Remarks	Successful

Table 4 Unit Test Case 4

S4 # 3Test 3Case	UTC- 4
Name30f3Test	L0ad cyberattck Result(Pr0cess1-ML)
Expected3Result	When we test the flexible BC data using machine learning pr0cess then it will detected Cyberattack clabel.
Actual3Output	Same as expected.
Remarks3	Successful

Table 5 Unit Test Case 5

S5 # Test Case	UTC- 5
Name Of Test	Result(Pr0cess2-DL)
Expected Result	It will display in real time When we test the cyber data using deep learning(CNN) pr0cess then it will detected whether It is attack Or n0rmal.
Actual Output	Same as expected.
Remarks	Successful

Table 5 Unit Test Case 6

S6 # 3Test 3Case	UTC- 6
Name30f3Test	Result(Pr0cess3-ML)
Expected3Result	It will display in real time When we test the scenari0 IEEE n0de data using Machine Learning (ML) pr0cess then it will detected cyber-attacks (Scaling, Ramping, Rand0m and Sm00th-Curve)
Actual3Output	Same as expected.
Remarks3	Successful

IV. CONCLUSION

This article uses Generalized Graph Laplacian and versatile Bayes classifiers (BCs) to develop flexible machine learning based cyberattack detection method by Spatiotemporal. The flexible Bayes classifiers train spatiotemporal patterns of system measurements and to detect attacks online. Numerical results will confirm the effectiveness of the designed cyberattack identification technique based on machine learning.

V. FUTURE SCOPE

In this project, we design a flexible machine learning based cyberattack detection method by using the generalized graph Laplacian (GGL) and flexible Bayes classifiers (BCs). Spatiotemporal patterns are quantitatively characterized by GGL, which could be affected when cyberattacks happen. The flexible BCs are used for coaching spatiotemporal patterns of system measurements and detecting cyberattacks online. Numerical results of case studies verify the effectiveness of the developed cyberattack detection procedure based on machine learning techniques.

REFERENCES

- [1] M. Cui, J. Wang, A. R. Florita, and Y. Zhang, "Generalized graph Laplacian based anomaly detection for spatiotemporal microPMU data," *IEEE Trans. Power Syst.*, vol. 34, no. 5, pp. 3960–3963, Sep. 2019
- [2] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under Laplacian and structural constraints," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 6, pp. 825–841, 2017.
- [3] Pecan Street Data. [Online]. Available: <https://www.pecanstreet.org/category/dataport/>
- [4] R. C. Dugan, "Reference guide: The open distribution system simulator(OpenDSS)," Electric Power Research Institute, Inc, vol. 7, p. 29, 2012.
- [5] Wang, Z. Wang, J. Wang, and D. Zhao, "SVM-based parameter identification for composite ZIP and electronic load modeling," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 182–193, Jan. 2019.
- [6] Z. Ghafoori, S. M. Erfani, S. Rajasegarar, J. C. Bezdek, S. Karunasekera, and C. Leckie, "Efficient unsupervised parameter estimation for one-class support vector machines," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 5057–5070, 2018.
- [7] J. Wang, D. Shi, Y. Li, J. Chen, H. Ding, and X. Duan, "Distributed framework for detecting PMU data manipulation attacks with deep autoencoders," *IEEE Trans. Smart Grid*, 2018, in press.
- [8] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1644–1652, Sep. 2017.
- [9] E. Keogh, J. Lin, and A. Fu, "Hot SAX: Efficiently finding the most unusual time series subsequence," in *Proc. IEEE Int. Conf. Data Mining*, Houston, TX, USA, 2005, pp. 226–233.
- [10] Assess the impact and evaluate the response to cybersecurity issues (AIERCI). [Online]. Available: https://www.energy.gov/sites/prod/files/2017/04/f34/BNL_AIERCI_FactSheet.pdf



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)