Evaluating Artificial Intelligence Mitigation Techniques for Countering Attack on Smart Factory SCADA Network

Love Allen Chijioke Ahakonye, Cosmas Ifeanyi Nwakanma, Jae Hyun Lee, Jae Min Lee, Dong-Seong Kim
IT Convergence Engineering, Kumoh National Institute of Technology.
(loveahakonye, cosmas.ifeanyi, antigen47, ljmpaul, dskim)@kumoh.ac.kr

Abstract

The new model of the Smart Factory (SF) compels all the industrial sectors to face an extensive digital evolution to be at the brim in a competitive scheme. As a result, Artificial Intelligence (AI) can be highlighted as a support to other technologies, such as cybersecurity. Following the trend of the vulnerability in the SF SCADA system, there is a serious need for cutting-edge mitigation solutions leveraging AI. This study aims at exploring AI techniques for the mitigation of attacks in the SCADA system. From the results of the evaluation, Weighted KNN, Fine KNN and Fine Tree are considered most ideal for SF SCADA attack mitigation.

Ⅰ. Introduction

Presently, industries have developed their business lines to offer the technology of critical industrial systems such as supervisory control and data acquisition (SCADA). This is to enable the inter-connectivity and remote accessibility of sub-stations. The smart factory (SF), SCADA system comprises the internet and industrial internet of things (IoT, and IIoT). This has facilitated the feasibility of a comprehensive implementation of remote administration which has bared the SCADA system to vulnerabilities and potential attacks. This increased the necessity for utmost safety for proficient administration and control, emerging in sustained industrial operations with premium achievement. Besides, the damage repair from attacks on SCADA and critical industrial systems can be overwhelming [1]-[4].

Mitigation techniques such as SCADA hardware security, network-based intrusion detection systems exist. Several studies have employed artificial intelligence (AI) techniques to resolve a variety of issues. AI is a technology that enables a machine simulation of human actions utilizing machine learning (ML). The automatic method of attack detection and classification in SF SCADA is based on AI techniques. However, the premise for determining the type of AI and its qualities establishes a research challenge.

Researchers have proffered ML prospects to tackle the advancement of intrusion. In [5], the authors utilized the Decision Tree, Random Forest, Naive Bayes, Logistic Regression and K-Nearest Neighbour (KNN) on a water storage facility testbed generated dataset. From their evaluation, the approach had an excellent performance in accuracy and false alarm rates with an unbalanced dataset. [6], had compared the performance of NB, KNN, ANN techniques. This technique performed poorly, hence, the ensemble technique was employed to enhance performance but this model requires minimal computational cost. Furthermore, deep learning (DL) techniques were proposed, and widely utilized for their ability to learn and predict meaningful attributes [7], [8].

Regardless of this number of studies, there is still a lack of convergence at establishing the most viable AI technique for attack detection in SCADA network. Hence, researchers face a dilemma at arriving at the AI technique of choice. Inspired by this challenge, this study seeks to achieve the following:

1) Utilization and appraisal of various AI techniques relevant to SCADA attack mitigation
2) Leveraging a current cyber-security dataset and choice of AI techniques, this work compared various ML models and their performances.
3) Preference in choosing the best proactive scheme based on computational complexity and accuracy in terms of model training time

Following section I is section II which is the overall system model. Section III is the evaluation of the AI techniques. The paper was concluded in section IV

Ⅱ. System Methodology

Modern smart factory's functionality are increasingly becoming complex. This is enabled by the possibility of massive connectivity of devices and sensors and thus, SCADA has become vulnerable to attack. Therefore, the need for efficient security is essential. To address SCADA network security vulnerability in the smart factory, this work presented a comparison of various AI algorithms developed to monitor communication traffic in encoded DNS and identify abnormal nature in the traffic. Fig. 1 depicts the system model of the study. The proposed architecture consists of 3 main phases; training, testing, and model determination phase. During the training and testing phase,the labelled cyber-security dataset which is extracted from normal web browsing activities and domain name service channelling mechanisms is split into training and testing sets and imported into the diverse ML algorithms after implementing a five (5) fold cross-validation. The test set was used to validate the performance of the training set. The choice and decision for the best model is regulated by the performance metrics (PM). The PM used in evaluating the models is the accuracy of the selected model, training time and the model's mis-classification error (MCE).

Fig 1. System Model of the Evaluation of AI Techniques for SF SCADA Vulnerability Mitigation

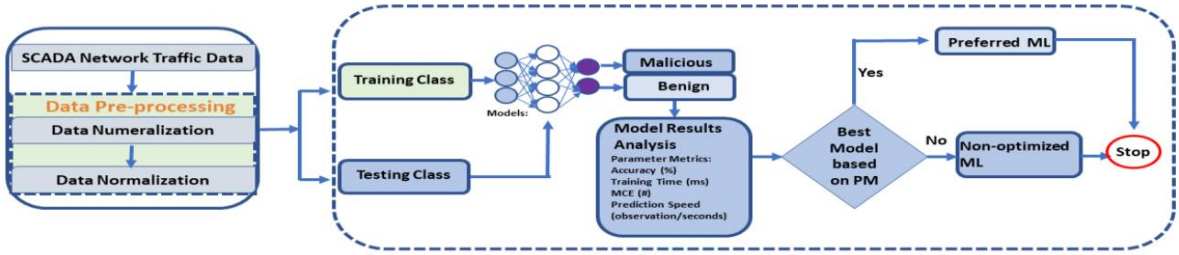

Table 1.
COMPARISON OF ACCURACY, TIME, MCE AND PREDICTION SPEED

Parameter Metrics	Weighted KNN	Fine KNN	Fine Tree
Accuracy (%)	99.2	99.2	98.1
Training Time (s)	8.8344	13.583	4.5778
MCE (#)	2134	2192	5096
Prediction Speed (obs/sec)	250000	440000	1100000

Table 2.
COMPARISON OF ACCURACY, TIME, MCE AND PREDICTION SPEED SHOWING 3 LEAST PERFORMED AI TECHNIQUES

Parameter Metrics	Bagged Tree	QDR	Coarse Tree
Accuracy (%)	99.4	92.9	95
Training Time (s)	85.488	3.8003	3.5817
MCE (#)	1512	19225	13357
Prediction Speed (obs/sec)	1600000	990000	1200000

Ⅲ. Performance Evaluation of various AI Techniques

A. Analysis and Comparison of various ML Algorithms for Intrusion Detection and Mitigation in SCADA

In this study, leveraging on a cyber-security dataset [9], MATLAB 2019Rb was used to examine all ML prospects. The algorithms include classes of Decision Trees, Discriminant Analysis, Logistic Regression, Naive Bayes, Support Vector Machines, K-Nearest Neighbours and Ensemble classifiers.

B. Summary of Performance Evaluation

A total of twenty-two (22) models were simulated, the result of the analysis for the top three best-performed algorithms is as shown in Table I and the least performed in Table II. The comparison is based on the parameter metrics in terms of model training time, accuracy and MCE. For validation, 30% of the dataset was reserved.

Ⅵ Conclusion

In this study, a relative analysis of various ML AI techniques was done. This was to determine the best ML prospect for attack mitigation in SF SCADA. From the evaluation, it is evident that several ML exhibit significant performance in terms of accuracy, MCE, training time and prediction speed. Nevertheless, considering the complexity of SF operations, the accuracy, detection time and number of misclassification errors are the basis for the choice of the preferred ML. Based on a trade-off of these parameter metrics, the Weighted KNN had the best performance occasioned by its versatility, interpretability, high accuracy and swift calculation time. However, it is not suitable for large dataset and sensitive to nosiy data.

ACKNOWLEDGMENT

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the IITP by MSIT, Korea.

References

[1] D.-S. Kim and H. Tran-Dang, "Industrial Sensors and Controls in Communication Networks," Computer Communications and Networks. Springer International Publishing, Cham, 2019.

[2] C. I. Nwakanma, F. B. Islam, M. P. Maharani, J.-M. Lee, and D.-S. Kim, "Detection and Classification of Human Activity for Emergency Response in Smart Factory Shop Floor," Applied Sciences, vol. 11, no. 8, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/8/3662

[3] G. C. Amaizu, C. I. Nwakanma, S. Bhardwaj, J.- M. Lee, and D.-S. Kim, "Composite and Efficient DDoS Attack Detection Framework for B5G Networks," Computer Networks, vol. 188, p. 107871, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621000438

[4] C. I. Nwakanma, M. S. Hossain, J.-M. Lee, and D.-S. Kim, "Towards Machine Learning Based Analysis of Quality of User Experience (QoUE)," International Journal of Machine Learning and Computing, vol. 10, no. 6, pp. 752–758, 2020. [Online]. Available: https://doi.org/10.18178/ijmlc.2020.10.6.1001

[5] M. Teixeira, T. Salman, M. Zolanvari, R. Jain, N. Meskin, and M. Samaka, "SCADA System Testbed for Cybersecurity Research Using Machine Learning Approach," Future Internet, vol. 10, no. 8, p. 76, Aug 2018. [Online]. Available: http://dx.doi.org/10.3390/fi10080076

[6] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Intrusion Detection in SCADA based Power Grids: Recursive Feature Elimination Model with Majority Vote Ensemble Algorithm," IEEE Transactions on Network Science and Engineering, pp. 1–1, 2021.

[7] J. Li, Y. Qu, F. Chao, H. P. H. Shum, and L. Ho, Edmond S. L.and Yang, Machine Learning Algorithms for Network Intrusion Detection. Cham: Springer International Publishing, 2019, pp. 151–179. [Online]. Available: https://doi.org/10.1007/978-3-319-98842-96

[8] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network Intrusion Detection system: A Systematic Study of Machine Learning and Deep Learning Approaches," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, p. e4150, 2021

[9] "CIRA-CIC-DoHBrw-2020 Dataset," 2020. [Online]. Available: https://www.unb.ca/cic/datasets/dohbrw-2020.html