A Noise-Removal Machine Learning Approach for Enhanced Threats Detection in ICS Networks

Urslla Uchechi Izuazu, Vivan Ukamaka Ihekoronye, Dong-Seong Kim, Jae Min Lee
uursla8@kumoh.ac.kr, ihekoronyevivian@gmail.com, (dskim, ljmpaul)@kumoh.ac.kr

***Abstract*—The centralized and vulnerable nature of industrial control system (ICS) networks, attracts malicious actors seeking to exploit vulnerabilities, compromise data and disrupt critical processes. Current threat detection methods overlook real-world noise disturbances experienced in industrial processes, leading to sub-optimal model performance. This study introduces a security framework that can be deployed at the supervisory part of ICS, to handle noisy traffic from industrial processing, ensuring an effective attack detection. Experimental simulations validate its effectiveness when compared with state-of-the-art based on key performance metrics.**

***Index Terms*—Attack, ICS, Intrusion detection, Machine learning**

I. Introduction

Industrial control systems (ICS) are responsible for monitoring, and controlling physical processes within diverse industrial settings, such as production plants, smart grids, and intelligent transport systems. These systems incorporate components like distributed control systems (DCSs), programmable-logic- controllers (PLCs), supervisory control segments, data acquisition (SCADA) systems, human-machine interfaces (HMIs), and sensors. These components collectively oversee essential control functions across industrial sectors [1]. In a modern ICS, a three-layered structure is employed: The application layer for enterprise administration, the network layer for supervision, and the physical layer for field operations, all integrated as a unit. This integration enhances operational efficiency but also renders the network susceptible to threats due to design oversights and insecure communication protocols. Attacks are often launched at the supervisory segment, as it play a central role in coordinating the entire industrial process, providing malicious actors with substantial information. A widely known attack instance is the Stuxnet malware2010 [2], which led to the infiltration of Iran's Natanz uranium enrichment facility, leading to a shutdown of the facility. This instance and more, highlight ICS as a good target for hackers, making its security a pressing international concern. Current machine learning (ML) approaches [3], [4], [5] often assume a perfect scenario void of disturbances, and fail to consider the inherent noise in real-life industrial environments arising from vibrations in industrial processes, or faulty sensor measurements. In reality, the deployment of these models may introduce performance degradation, resulting in catastrophe. To address this issue, this paper proposes an intrusion detection system (IDS), that addresses the challenge posed by noise and provides a robust mechanism for threat detection and classification in ICS networks.

This study aims to achieve the following specific objectives:

i. To design an ML-based intrusion detection framework for threat detection, with an emphasis on addressing inherent noise present in real-world industrial environments.

ii. To fortify the model against noise interference through the incorporation of a regularization technique, to enhance robustness and optimal performance.

ii. To ascertain the proposed model performance using the ICS-flow dataset, and renowned key evaluation metrics.

The remainder of this paper is structured as thus: Section II presents the Methodology and System Design, while Result Discussion, Performance Evaluation, and Conclusion are captured in Sections III and IV, respectively.

The paper is structured as follows: Section II presents the Methodology and System Design, while Result Discussion, Performance Evaluation, and Conclusion are captured in Sections III and IV, respectively.

II. Methodology and System Design

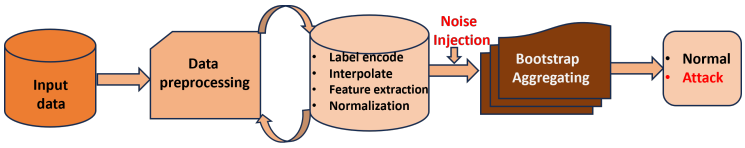

Fig. 1. Workflow of the Proposed System

The proposed model utilizes an ensemble learning approach that combines multiple decision trees, using an established anomaly detection mechanism, as depicted in Fig. 1. The ICS network traffic flow undergoes a sequential preprocessing pipeline. Initially, categorical labels are encoded, and missing values are interpolated to avoid data loss. Subsequently, feature selection and normalization are applied. To enhance the model's robustness to noise, a regularization technique was introduced; "Gaussian noise", with a range of 0.1-0.4, follows a probability density function resembling the normal distribution [6]. A noise value of 0.2 was injected into the traffic flow, and used for training multiple decision tree (DT) algorithms. A technique known as bootstrap aggregating (Bagging) [7]. Here, each tree is trained on a random subset of the training data through bootstrap sampling, which involves drawing samples with replacements from the original dataset. The final prediction is then obtained by aggregating the predictions of individual trees, ensuring that the collective intelligence of the trees is better equipped to handle and adapt to any form of

variations/noise tending to reduce optimal flow analysis and accurate detection of threats within the network.

The ICS-flow dataset was obtained from a bottle-filling factory control system and contains 45,719 network traffic flows, representing normal ("0") and various attacks (IP-Scan, Port-Scan Replay, DDoS, MitM denoted as ("1")) [8].

The experiment was done within a Python environment, utilizing a system configuration that includes an Intel(R) Core(TM) g5-7400 CPU running at 3.00GHz, a Tesla K80 GPU, and 8GB of RAM.

III. Result Discussion and Performance Evaluation

To assess the effectiveness of the proposed model, 3 scenarios were examined and compared. In scenario 1, we assumed an ideal case, where a DT model was trained without the introduction of noise. Moving to scenario 2, A realistic element of noise was added to assess its impact on model performance, reflecting the challenges encountered in real-world ICS. In scenario 3, we employed the bootstrap aggregating method to investigate its performance on noisy traffic signals. Table III shows the various performances of the scenarios, using the following key metrics: Accuracy, precision, f1-score, Matthews correlation coefficient (MCC), and recall.

TABLE I
Comparison of the Proposed Bagging Method with DT models "with" and "without" noise injection

Scenario	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MCC (%)
1 DT (without noise)	87.7	87.4	87.7	86.9	75.8
2 DT (with noise)	64.3	73.3	64.3	67.4	54.0
3 Bagging (with noise)	**89.5**	**89.1**	**89.3**	**89.5**	**76.2**

Based on the simulation result, scenario 1 under ideal noise-free conditions, displayed a good marginal performance across all metrics except for the MCC score, which is attributed to the imbalances in the distribution of true and false positives and negatives. In Scenario 2, a performance dip was observed, emphasizing the drastic effect of noise on the model performance. However, scenario 3, utilizing our proposed approach, showcased superior results, outperforming both scenarios, achieving an accuracy of 89.5%, 89.1% precision, 89.3%, 89.5%, and 76.2% recall, f1-score and MCC respectively, highlighting its resilience against noise-induced challenges.

Fig.2 shows the out-of-bag (OOB) score of our proposed bagging methods, with variations in the number of trees in the ensemble. The OOB score serves as an indicator of the model's performance on unseen data, offering insights into how the accuracy evolves with varying numbers of trees in the ensemble. Notably, a consistent improvement is observed with an increasing number of trees until it stabilizes at 100,

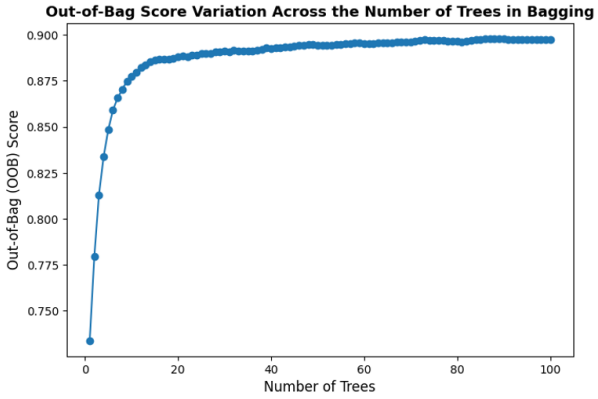

Fig. 2. Performance Accuracy of proposed Bagging model

highlighting the point at which further tree additions no longer significantly enhance the model's predictive power

IV. Conclusion

This study introduces a security framework for real-time attack detection in ICS. Our approach incorporates a regularization technique, acknowledging the noise inherent in industrial activity, thus enhancing robustness in threat detection. Experimental results using the ICS-Flow dataset demonstrate the framework's effectiveness in distinguishing threats from normal network operations in the presence of noise. Future work includes model optimization and evaluation on additional metrics, like training time, which is an important factor when dealing with time-critical systems like ICS networks.

Acknowledgment

This research was supported by the MSIT, Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2024-2020-0-01612) supervised by the IITP and by Priority Research Centers Program through the NRF of Korea funded by the MEST(2018R1A6A1A03024003)

References

[1] D. Bhamare, M. Zolanvari, A. Erbad, R. Jain, K. Khan, and N. Meskin, "Cybersecurity for industrial control systems: A survey," *computers & security*, vol. 89, p. 101677, 2020.
[2] M. Baezner and P. Robin, "Stuxnet," ETH Zurich, Tech. Rep., 2017.
[3] M. Gaiceanu, M. Stanculescu, P. C. Andrei, V. Solcanu, T. Gaiceanu, and H. Andrei, "Intrusion detection on ics and scada networks," *Recent Developments on Industrial Control Systems Resilience*, pp. 197–262, 2020.
[4] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Efficient classification of enciphered scada network traffic in smart factory using decision tree algorithm," *IEEE Access*, vol. 9, pp. 154 892–154 901, 2021.
[5] H. Li, B. Wang, and X. Xie, "An improved content-based outlier detection method for ics intrusion detection," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, pp. 1–15, 2020.
[6] X. Zhang, Y. H. Wu, D.-P. Covei, X. Hao *et al.*, "Complex boundary value problems of nonlinear differential equations: Theory, computational methods, and applications," in *Abstract and Applied Analysis*, vol. 2013. Hindawi.
[7] M. H. L. Louk and B. A. Tama, "Dual-ids: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system," *Expert Systems with Applications*, vol. 213, p. 119030, 2023.
[8] A. Dehlaghi-Ghadim, M. H. Moghadam, A. Balador, and H. Hansson, "Anomaly detection dataset for industrial control systems," *arXiv preprint arXiv:2305.09678*, 2023.