Anomaly Detection for Industrial Control Systems Based on Neural Networks with One-Class Objective Function

Emmanuel Aboah Boateng, J.W. Bruce

Department of Electrical and Computer Engineering

1. INTRODUCTION

Industrial control systems (ICS) describe the use of network connectivity to integrate hardware and software in order to control critical infrastructure such as chemical plant, electricity distribution, and nuclear facilities. The prevalence of internet of things technology and networked sensors in many ICS have exposed critical infrastructure to several malicious activities and cyber threats.

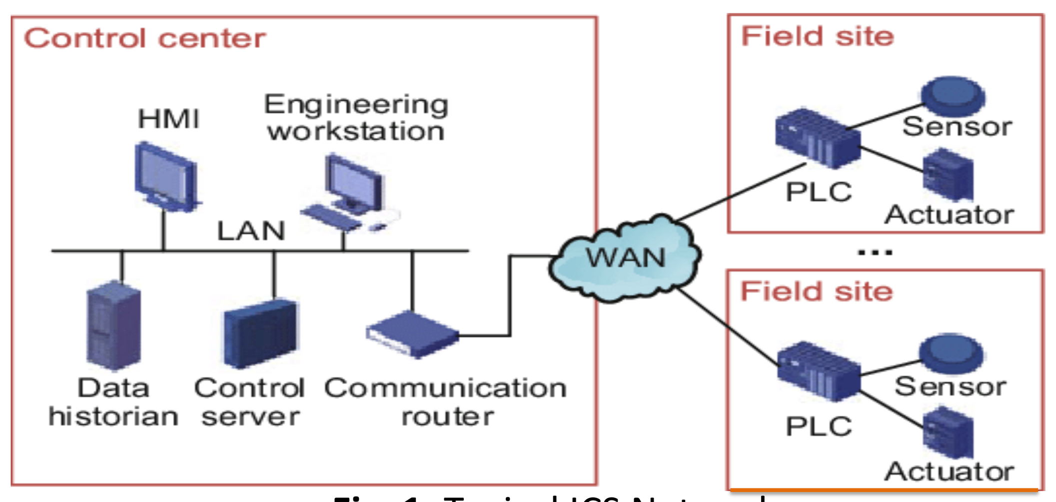


Fig. 1: Typical ICS Network

Because ICS control critical infrastructure, ICS attacks can cause irreparable damage to enterprises and even loss human life. Programmable logic controllers (PLCs) which monitor and control the physical processes of ICS have unique architecture which makes it difficult to apply traditional techniques for ICS protection. This research work, therefore, proposes a novel approach using neural networks with one-class objective function for anomaly detection in ICS. This approach was evaluated on a real-world ICS dataset: the Secure Water Treatment (SWaT) dataset.

2. OBJECTIVES

- Present a neural network with one-class objective function for anomaly detection in ICS
- Evaluate the network on a real-world dataset (SWaT dataset)
- Compare our approach against previous works

3. DATASET

The SWaT dataset is a current and widely used open-source dataset for ICS security research. It was collected from a scaled-down water treatment plant by [1] as shown in Fig. 2.

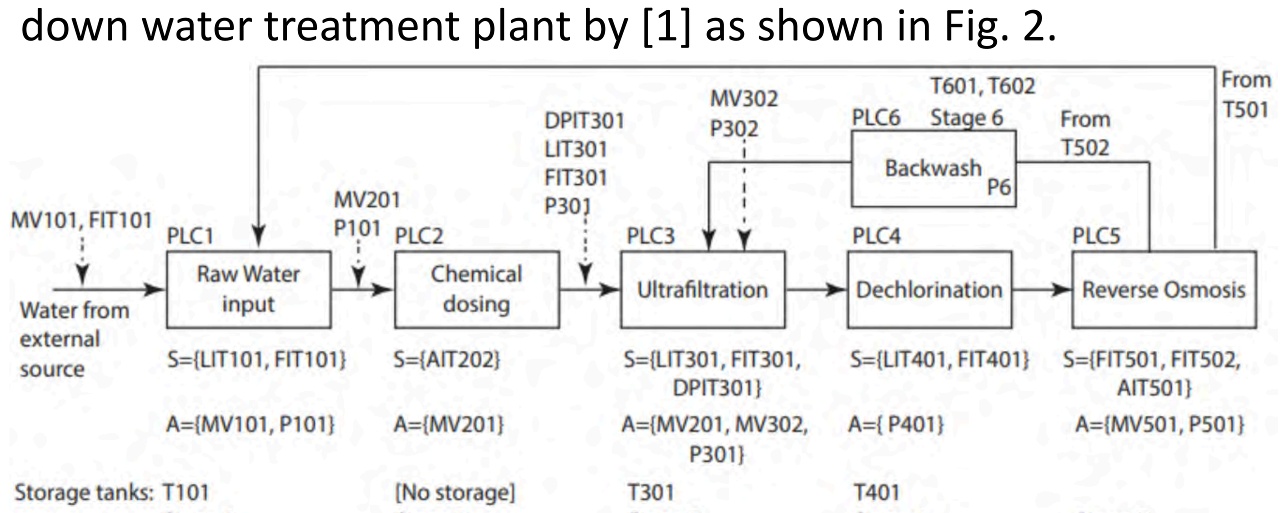


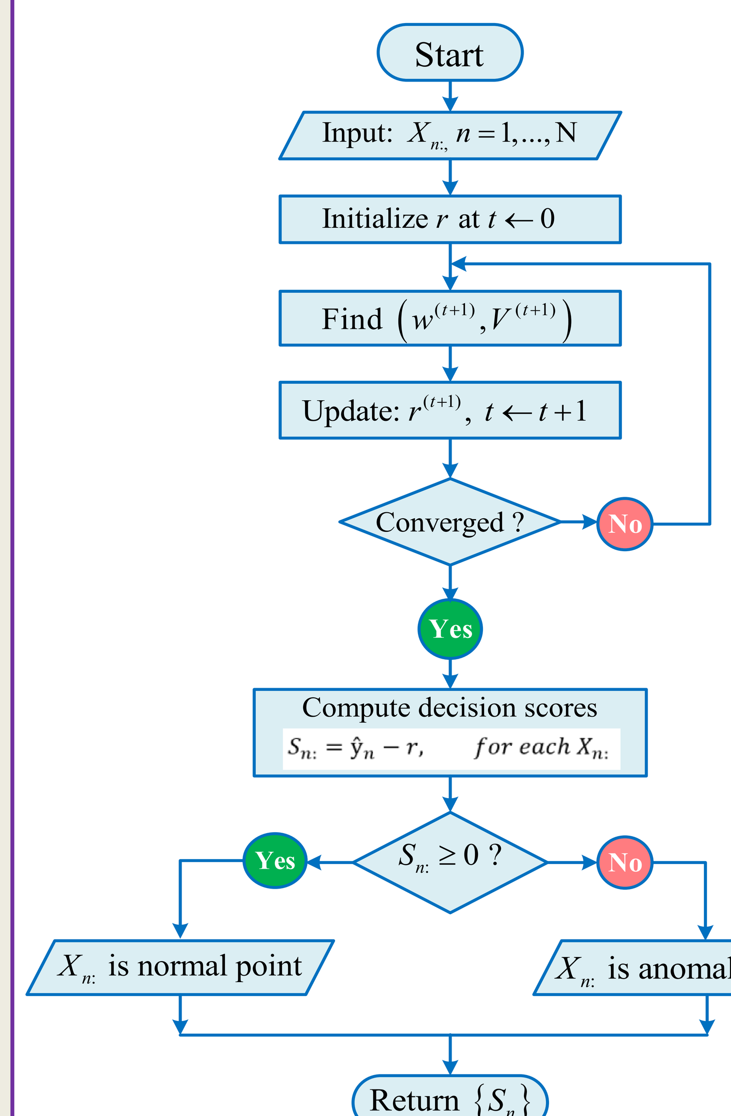
Fig. 2: SWaT Testbed [1]

The testbed consisted of 25 sensors and 26 actuators. The data was recorded for 11 days in which 36 different attacks were injected to compromise about 6% of the dataset.

4. METHODOLOGY

A. Machine Learning Algorithm

This research work employs an unsupervised machine learning technique which combines the abilities of neural networks to learn complex relationships with a one-class objective function which then separates the anomalous instances from the normal conditions.



Notations

X_n = training data
r = bias of hyperplane
w = scalar output
V = weight matrix
$S_{n:}$ = Scoring function

Fig. 3: Algorithm of the model

From Fig. 3, r is first initialized, and the model uses backpropagation to learn the parameters (w, V) of the neural network. The model then updates r and once convergence is achieved, the scoring function $S_{n:}$, labels the data points as normal and anomalous instances. Fig. 4 represents the model architecture of the neural network with one-class objective function.

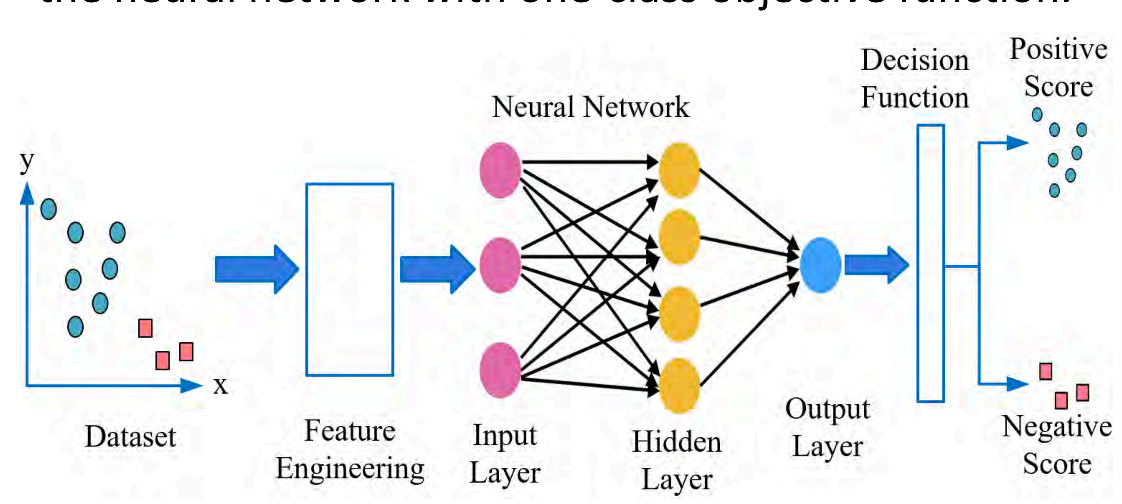


Fig. 4: Model Architecture

Based on the optimization problem of one-class SVM and the minimization algorithm proposed in [2], the objective function can be formulated as:

$$\min_{w,V,r} \frac{1}{2}\|w^T w\| + \frac{1}{\alpha}\|V^T V\| + \frac{1}{N\nu}\sum_{n=1}^{N} \max(0, r - \langle w, g(VX_{n:})\rangle) - r \quad (1)$$

Where the parameters of Equation 1 have already been defined in Fig. 3 and α is a hyper-parameter for controlling the weight matrix, V.

B. Anomaly Detection Framework

The SWaT dataset was first preprocessed by normalizing all the data points. Only the normal instances were used for training the model in order to enable the network to learn the normal pattern. The performance of the model was then evaluated on a second log of the SWaT dataset containing both normal and anomalous instances as shown in Fig. 5.

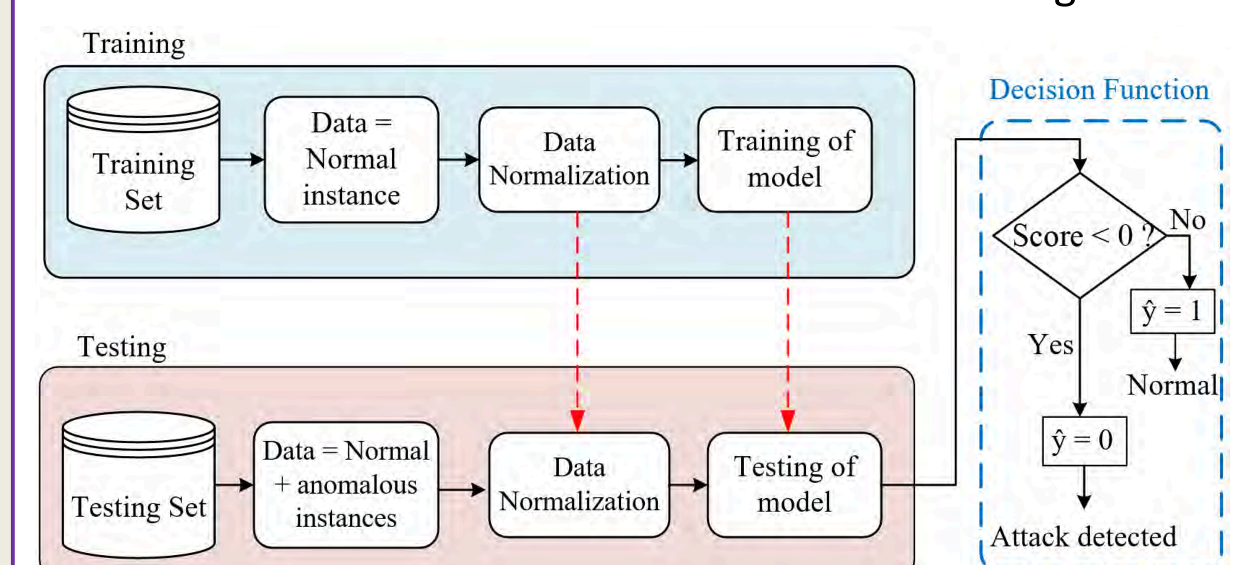


Fig. 5: Framework of Attack Detection Approach

It is worth noting that the dataset was pre-processed similar to what has been done in previous works. As a result, we were able to compare the performance of our approach to other applied approaches in literature that have been developed using the SWaT dataset.

5. RESULTS AND DISCUSSION

Several simulations were run and the hyper-parameters of the architecture with the best model is shown in table 1.

Table 1: Hyper-parameters of the best model

Hidden layers(k)	nu	Alpha (α)	Activation Funct. g(.)	r value
32	0.016	9	Sigmoid	0.1

The performance metrics of evaluation were precision, recall and F1-score. Table 2 summarizes the results of our approach as compared to other state-of-the-art techniques.

Table 2: Results comparison between different detection methods on the SWaT dataset

Method	F1-score	Precision	Recall	Complexity
NN [3]	0.812	0.976	0.696	Low
SVM [4]	0.796	0.925	0.699	High
ID-CNN [5]	0.860	0.867	0.854	High
RNN [4]	0.802	0.982	0.678	High
TABOR [6]	0.823	0.862	0.788	Average
KNN [7]	0.350	0.348	0.348	Average
FB [7]	0.360	0.358	0.358	Average
AE [7]	0.520	0.516	0.516	Average
EGAN [7]	0.510	0.406	0.677	High
DIF [8]	0.882	0.935	0.835	Average
NN-one class	0.800	0.950	0.710	Average

Our technique achieved improved F1-score of 80% and recall of 71%. As compared to other approaches with similar computational complexities such as TABOR, AE, FB, KNN and DIF, our model performed better in terms of precision.

Table 3: Recall values of the different approaches

Attack No.	NN	RNN	SVM	TABOR	ID-CNN	DIF	NN-One class
17	0.98	0.99	1.00	0.99	1.00	1.00	0.96
18	0.71	0.88	0.88	0	1.00	0.82	0.02
19	0.92	0	0	0	0.017	0.34	0.69
20	0.29	0	0.01	0	0.02	1.00	1.00
21	0.99	0	0	0.99	1.00	0.17	0.03
22	0	0	0	0.20	0.06	0	0
23	0.03	0.94	0.94	1.00	1.00	1.00	1.00
24	0.87	0	0	0	0	1.00	1.00
25	0.83	0	0	0.99	1.00	0	1.00
26	0.78	0	0	0	0.30	1.00	1.00
27	0.33	0	0.91	0	0.94	1.00	0.93
28	0.84	0	0	0.88	0.89	0.43	0.88
29	0	0	0	0.60	0.99	0	0.62
30	0	0	0	0.26	0	0.95	0.95
31	0.81	0	0.12	0.89	0.88	0.93	1.00
32	0.84	1.00	1.00	0.99	0.90	1.00	1.00
33	0.77	0.92	0.93	0.99	1.00	1.00	1.00
34	0.84	0.94	0	0.40	0.91	1.00	1.00
35	0.78	0.93	0.93	0.99	1.00	1.00	1.00
36	0	0	0.36	0	0.64	0.63	0.79

From Table 3, it can be realized that our model was able to detect most of the last 20 attacks. Our model had the highest recall on the attacks 20, 23 – 26 and 30 - 36, i.e., achieving 100% recall in most cases.

6. CONCLUSION

The viability of anomaly detection in ICS based on neural network with one-class objective function is demonstrated. The model framework was evaluated on the SWaT dataset. In comparison with previous works, our technique showed significant improvement in terms of attack detection capability and computational complexity, and this shows that the technique is suitable for use in real ICS scenario.

7. REFERENCES

[1] J. Goh, et. al, ‘‘A data set to support research in the design of secure water treatment systems,’’ *in Proc. Int. Conf. Crit. Inf. Secur.*, 2016, pp. 88–99

[2] R. Chalapathy, et. al, "Anomaly detection using one- class neural networks." *arXiv preprint arXiv*:1802.06360 (2018)

[3] D. Shalyga, et. al., ‘‘Anomaly detection for water treatment system based on neural network with automatic architecture optimization,’’ 2018.

[4] J. Inoue, et. al, ‘‘Anomaly detection for a water treatment system using unsupervised machine learning,’’ *in Proc. IEEE ICDMW*, 2017, pp. 1058–1065.

[5] M. Kravchik, et. al, ‘‘Detecting cyber attacks in industrial control systems using convolutional neural networks,’’ *in Proc. CPS-SPC*, 2018, pp. 72–83.

[6] Q. Lin, et. al, ‘‘TABOR: A graphical model-based approach for anomaly detection in ICS,’’ *in Proc. ACCCS*, NY, USA, 2018, pp. 525–536.

[7] D. Li et. al, ‘‘MAD-GAN,’’ *in Artificial Neural Networks and Machine Learning*, Springer, 2019

[8] M. Elnour et. al., A dual-isolation-forests-based attack detection framework for industrial control systems. *IEEE Access*. 2020 Feb 19; 8:36639-51.

8. ACKNOWLEDGEMENTS

Support from funds provided by CMR and the state of Tennessee to TN Tech in recognition of the University Carnegie Classification, R2, is appreciated.