IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 3, March 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003085 |

# Identification of Attacks in Water Infrastructures using Machine Learning Techniques

**Tharsanee R M[1], Shreekanthan C M[2]**

Assistant Professor, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India[1]

Information Processing Specialist, Tata Consultancy Services, Chennai, India[2]

**ABSTRACT:** Water distribution networks are often inclined to experience attacks in the form of cyber-physical threats which leads to the disruption in the normal operation of the system and also causes damage to the assets present in the network. This work proposes an algorithm which is used to identify anomalous phenomena in different components that are available in the water distribution network. This algorithm is designed in such a way that it contains a number of modules which are used to facilitate the anomaly- detection mechanisms for monitoring the data in real time. Artificial neural networks, one of the deep learning techniques, are used to predict the variation in the behavior of the system with respect to its normal operation. In order to uncover the global anomalies, this technique is adopted which can decompose the space occupied with high dimensionality by the data. The algorithm is trained in such a way that attack scenarios are included as part of training and testing datasets. The proposed algorithm is able to successfully detect the attacks which are simulated in BATADAL datasets - Battle of the attack detection algorithms with maximum sensitivity as well as specificity.

**KEYWORDS:** water distribution systems, machine learning, artificial neural networks, cyber attacks.

## I. INTRODUCTION

Water Distribution Systems have started to adopt the smart technologies for the distribution of water so that efficiency as well as the reliability of the distribution can be improved. These types of Smart water distribution systems can be categorized as a type of Cyber physical system which can easily perform the monitoring of a system online, collect and transmit the data faster, compute and automatically perform functional operations as a highly integrated process. A typical smart water system is a combination of sensors which are distributed in nature coupled with actuators which are remote [1]. These sensors and actuators will be linked together with the PLCs which are nothing but the Programmable Logic Controllers. These PLCs in turn are managed by the Supervisory Control and Data Acquisition Systems which are commonly called as SCADA systems. Though this combination of introducing cyber elements into the physical infrastructure is having numerous benefits, it also brings in a new threat to the system. These threats are in the form of cyber attacks by adversaries to various water infrastructures including water treatment plants or distribution systems [2]. Cyber attacks generally happen in various components of the water distribution system such as attacks on the sensors of the system that is involved in monitoring, attacks on the SCADA system, attacks on the PLCs, attacks may also happen in the communication routes used for wireless communication.

The paper is organized as follows: Section II describes the existing methods of attack detection in cyber physical systems. In section III proposed methodology is summarized with brief description on datasets, anomalies and algorithm. In section results and discussion is presented with suitable justifications and section V concludes the paper.

## II. EXISTING SYSTEM

Cyber attacks are generally intended to affect the performance of the system, to provide access to the system properties to unauthorized users, in advanced cases it may also result in causing physical damage to the water infrastructure [3]. Public safety will be put to trouble when regular working of water treatment plants are altered to contaminate the quality of water, also sensors indicating the water quality may be suppressed to restrict it from providing warnings when quality of water is altered [4]. Thus it becomes necessary to implement appropriate security measures in order to restrict such cyber attacks in smart water distribution systems. Emphasis should be given to impose security measures to the various components in the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003006 |

water distribution system including sensors, actuators and SCADA modules, in order to make them resilient to cyber attacks [5]. Though attacks happening to the components in the water distribution system at least once are unavoidable, it is highly essential to mitigate necessary processes to identify abnormal behaviors in the system so that possible interruptions in the services provided by the system can be avoided. There are several works in the literature with respect to the detection of cyber attacks in cyber physical systems like power grids and industrial control systems with less importance to identification of attacks in water infrastructures [6,7]. Thus, the proposed work is focused towards detection of attacks in water infrastructures using machine learning techniques. The main objectives of this work are to find the components that are being attacked, to identify the existing attack with high speed and to eliminate the instances of false alarms for the attacks. The attacks mainly in the SCADA systems are of our prime focus. To establish the attack detection, the machine learning algorithms are trained for some SCADA observations for datasets with attack scenarios.

## III. PROPOSED METHODOLOGY

### 3.1 Datasets used

C-Town network is a medium sized water distribution network which is employed as a benchmark dataset for the implementation of the attack detection algorithm in the proposed work[8]. The number of nodes in this C-Town network is upto 388 and the connections between the nodes ranges upto 429. The major components of this network are water tanks whose water levels control the water pumps, pumping stations along with the control valves. This C-Town network has been primarily used in the Battle of the Attack Detection algorithms and it is a kind of a Water grid technology which is composed of sensors and actuators remotely to control the various components of the network namely tanks, valves as well as the stations[9]. In addition to the sensors and actuators used in the network there are PLCs and Scada systems which are used for the operation of the system as in real time. There are three different kinds of datasets that are used in the implementation. Firstly, a dataset which has past data of the network for a specific period of time is taken. The main aim of using this dataset is to understand the normal behavior of the system without any attack scenarios. Secondly, a dataset is used which consists of network data with attack scenarios. This dataset can be used mainly to implement the attack detection in the network and thus validate its performance. Thirdly, a dataset which was recorded for a three month period is taken which is used to test the performance of the dataset under multiple attack scenarios.

### 3.2 Anomaly in SCADA Network

A data dependent approach is used in the proposed work to detect the anomalies. An anomaly is generally referred to as a data point that deviates from the usual behavior of the system [10]. Anomalies in SCADA observations are intended to be detected using the machine learning techniques. From the observations that are present in the SCADA system, it is required to find a subset of observations that can be designated to exhibit normal behavior. In this work, to understand the normal behavior of the system we consider the first dataset which contains the past data for a period of one year. There are two types of anomalies in general namely simple outliers and contextual anomalies. Simple outliers can be detected by just comparing the data with the already known statistical past data to identify if it deviates from the normal behavior. Detection of simple outliers is much easier compared to the detection of contextual anomalies which requires a trained algorithm to understand the patterns in the behavior. This trained algorithm is the machine learning algorithm which is capable of learning complex behavior of the system and thus in turn detects the anomalies with precision. Artificial neural networks are used in the proposed work for the effective detection of cyber physical attacks in the SCADA observations made by the water distribution system.

### 3.3 Proposed Algorithm

The proposed system consists of an algorithm which is implemented in three different modules; the first module is intended to check if the observations in the SCADA system are as per the rules specified by the actuators. The second module is focused towards the identification of simple outliers for the given set of observations. The third module is targeted towards finding the contextual anomalies in the dataset using Artificial Neural Networks.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003006 |

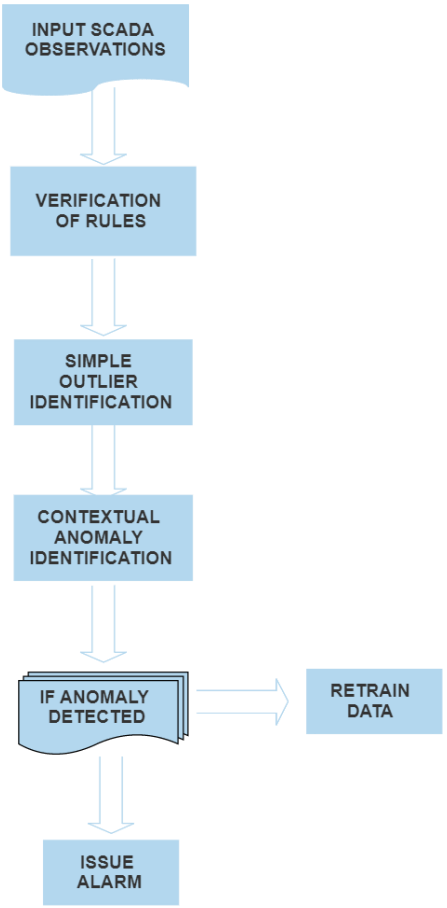


Figure 1. Architecture of proposed system

### 3.3.1 First Module: Verification of rules

The first module is to check the rules defined by the statistical fences, for instance, the rules can be as simple as checking the status of the pumps and valves. The predefined rules in the system may be disrupted if an attack occurs in the system that stops the normal operational behavior of the components in the system. Each time a rule is violated then the signal is notified for each anomalous data point.

### 3.3.2 Second Module: Simple Outlier Identification

In this module, simple outliers in the data observations are detected by comparing the observed data against the statistical data already available in the dataset. To accomplish this, it is required to set an upper limit and lower limit in the data which can act as a fence. These fences are essential to check if the current observations exceed the set boundaries defined by the fences that are generated from the historical data.

### 3.3.3 Third Module: Contextual Anomaly Identification

This is the module in which Artificial Neural Networks is used to forecast patterns from the observed data to identify the anomalous patterns in the data. The algorithm understands the behavior of the system from historical data and tries to predict the attacks from the deviations observed. These ANNs are multi-layered and consist of several artificial neurons connected to one another consecutively. These neurons are mainly used to perform kind of computations which are non linear in nature. The sum of the weighted outputs from each neuron in one layer is fed to every other neuron which is available in the next layer as in the form of a feed-forward fashion.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003006 |

## IV. RESULTS AND DISCUSSION

Performance comparison is vital to understand the behavior of the proposed algorithm. The second dataset is validated by applying the proposed algorithm in order to evaluate the performance of the algorithm.

### 4.1 Measure-to-Detect (MTD) metric

In order to evaluate the performance of the system a metric named Measure-To-Detect is jointly used with the Traditional Confusion Matrix (TCM) measure. In this metric MTD is used to detect the time taken for the identification of the anomalies and TCM refers to the amount of anomalies identified as TRUE correctly. The Metric Value (MV) is defined as shown in equation 1,

$$MV = (MTD + TCM) / 2$$

### 4.2 Module Performance

Figure 2 depicts the performance exhibited by each module in detecting anomalies in the given dataset. For most of the known anomalies, the proposed algorithm was able to identify the attacks as well with more efficiency. The third module was able to detect the attacks with more speed and reliability. The modules also exhibit good performance to detect the attacks in the specific components of the system. The whole algorithm is tested for the data which is the SCADA observations collected for a period of one year and the anomalies were detected accurately during the mentioned period.

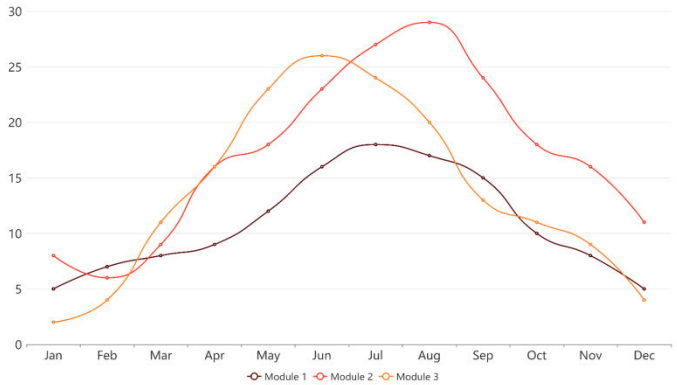


Figure 2. Performance of Proposed System

## V. CONCLUSIONS

Smart Technologies introduced in the water infrastructures have imposed cyber threats to the system in addition to the already existing physical threats to the system in the form of damages caused to the pipeline, altering the quality of water and compromising the physical equipment. This proposed work is aimed to employ machine learning techniques for the detection of cyber physical threats to the water distribution networks. Artificial Neural Networks is used to identify both the regular pattern in the normal behavior of the system and it is also empowered to study the anomalous behavior. Thus making it to compare and mitigate the attacks to the water distribution system. The performance of the system is evaluated using Measure-to-Detect metric and it is apparent that the proposed system works well to detect the threats to the water infrastructure with more accuracy.

## REFERENCES

1. Abokifa, A. A., Haddad, K., Lo, C. S., and Biswas, P. (2017). “Detection of Cyber Physical Attacks on Water Distribution Systems via Principal Component Analysis and Artificial Neural Networks.” Proceedings of World Environmental and Water Resources Congress 2017, 676–691.
2. Almalawi, A., Fahad, A., Tari, Z., Alamri, A., Alghamdi, R., and Zomaya, A. Y. (2016). “An Efficient Data-Driven Clustering Technique to Detect Attacks in SCADA Systems.” IEEE Transactions on Information Forensics

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 3, March 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1003006 |

and Security, 11(5), 893–906.

3. Arad, J., Housh, M., Perelman, L., and Ostfeld, A. (2013). "A dynamic thresholds scheme for contaminant event detection in water distribution systems." Water Research, 47(5), 1899– 1908.
4. Chandola, V., Banerjee, A., and Kumar, V. (2009). "Anomaly Detection: A Survey." ACM computing surveys (CSUR), 41(3), 15–58.
5. Housh, M., and Ohar, Z. (2017a). "Integrating physically based simulators with Event Detection Systems: Multi-site detection approach." Water Research, 110, 180–191.
6. Laszka, A., Abbas, W., Vorobeychik, Y., and Koutsoukos, X. (2017). "Synergic Security for Smart Water Networks: Redundancy, Diversity, and Hardening." Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, 21–24.
7. Maglaras, L. A., and Jiang, J. (2014). "Intrusion detection in SCADA systems using machine learning techniques." Science and Information Conference, IEEE 2014, 626–631.
8. Mathur, A. (2017). "SecWater: A Multi-Layer Security Framework forWater Treatment Plants." Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, 29–32.
9. Ohar, Z., Lahav, O., and Ostfeld, A. (2015). "Optimal sensor placement for detecting organophosphate intrusions into water distribution systems." Water Research, 73, 193–203
10. Pasqualetti, F., Dorfler, F., and Bullo, F. (2013). "Attack detection and identification in cyberphysical systems." IEEE Transactions on Automatic Control, 58(11), 2715–2729.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details