ISSN: 2229-6948(ONLINE)
DOI: 10.21917/ijct.2021.0381

# CYBERSECURITY IN IIOT AND IOMT NETWORKS USING MACHINE LEARNING ALGORITHMS - A SURVEY

**Pallavi Arora[1], Baljeet Kaur[2] and Marcio Andrey Teixeira[3]**

*[1]Department of Computer Science and Engineering, I.K. Gujral Punjab Technical University, India*
*[2] Department of Electronics and Communication Engineering, Guru Nanak Dev Engineering College, India*
*[3]Department of Informatics, Federal Institute of Education, Science, and Technology of São Paulo, Brazil*

*Abstract*

***Rapid advancements in micro-computing, mini-hardware manufacturing, and machine-to-machine (M2M) communications have allowed for innovative Internet of Things (IoT) solutions to redefine numerous networking applications. With the emergence of IoT branches such as the Internet of Medical Things (IoMT) and the Industrial Internet of Things (IIoT), healthcare and industrial systems have been changed by IoT. This paper presents an overview of the technologies that are being used to secure IoMT as well as IIoT frameworks seen within the research articles.***

***Keywords:***

***Machine Learning, Healthcare, Cybersecurity, Internet of Things (IoT)***

## 1. INTRODUCTION

The Industrial Internet of Things (IIoT) is based on integrating Internet of Things (IoT) technologies into industrial control systems (ICSs). ICSs are utilized to monitor industrial machinery and activities and are an important component of crucial structures. They allow direct device tracking as well as interaction, actual data gathering including interpretation, and archiving of all industrial system occurrences.

The incorporation of Internet of things into such devices boosts system cognition as well as safety while optimizing and perhaps automating industrial operations. The supervisory control and data acquisition (SCADA) system has been the most extensive subsystem of an ICS. With its human-machine interaction (HMI), it provides a graphical user interface (GUI). The HMI enables operators to easily monitor system status, communicate with IIoT devices, and receives warnings signalling problematic behavior.

The Internet of Medical Things (IoMT) is a subset of IoT networks tailored to the healthcare industry. It can reduce unnecessary hospital visits and the strain on healthcare systems by linking people to their doctors and allowing medical data to be transmitted through a secure network. However, the security of IoMT devices and healthcare systems is a significant problem. Healthcare data in IoMT systems should be safeguarded at all phases, including data gathering, transmission, and archiving.

With innovations in both security protection measures and new forms of attacks against IIoT and IoMT systems, a comprehensive assessment of existing security approaches and threats is necessary. This paper presents detailed review on how machine learning methods can be used for cybersecurity in ICS and healthcare domain.

### 1.1 CONTRIBUTION

The following are the primary contributions of this paper:

- We gave an insight of IIOT and IoMT communication protocols.
- Examining research publications that have developed effective machine learning-based intrusion detection systems (IDSs) targeting SCADA frameworks.
- We discussed the usage of machine learning algorithms in healthcare security.

The remainder of the paper is organized as follows: section 2 acquaints with the IIoT and IoMT communication protocols. Section 3 provides a brief overview regarding machine learning in IDS. Section 4 provides a review of the literature concerning the use of machine learning techniques for IIoT and IoMT security. Finally, section 5 provides the paper concluding thoughts.

## 2. IOT COMMUNICATION PROTOCOLS

### 2.1 IIOT COMMUNICATION PROTOCOLS

In SCADA systems, various IIoT data transfer protocols are utilized. However, most of these protocols were developed without taking into consideration cyber threats or security methods to mitigate them. The Modbus transmission interface, that is currently the most commonly used standard process in all of these networks, served as the cornerstone for SCADA.

Building Automation and Control Network (BACnet), Distributed Network Protocol version 3 (DNP3), and Message Queuing Telemetry Transport (MQTT) are emerging mechanisms that have recently gained popularity. All four protocols are free to use. Modbus was developed in 1979 as a supplier exclusive SCADA technology [1]. The BACnet [2]and DNP3[3] standard protocols originally introduced through 1995 and 1993, correspondingly. MQTT concept first used in 1999 [4].

- *Modbus***:** Modbus is among the oldest and perhaps extensively utilized SCADA protocols. Its transmission is master-slave in nature and sequential in character. The master (for example, an HMI) has been the equipment that requests relevant input, whereas the slave (for example, a PLC) has been the device that gives data. The master can also send information into the slave storage locations. A typical Mod-bus connection system consisted of one master and maximum to 247 slaves, one having its own unique identifying number (ID).
- *DNP3:* Another common network interface used in SCADA systems is DNP3. It was initially intended to be extremely

trustworthy, however it lacks adequate security features. consequently, the majority of DNP3 equipment lack verification, confidentiality, and permissions. DNP3 is concerned with the four OSI layers: networking, application, information connection, and tangible. Information is transmitted in a juncture mode, although it is most commonly done in master-slave setups with numerous slaves and masters. It is one of the oldest and most widely used protocols in SCADA systems. The most recent version of this protocol, issued by IEEE in 2012, enables safe authentication based on the IEC 60870-5 standard [5]. Authentication is given using digital signatures through this standard, which has been designed for control systems. Using public key infrastructure (PKI) in IIoT devices, on the other hand, is not yet viable.

- *BACnet:* BACnet was designed primarily for building industrial automation systems. BACnet, like most commercial networking systems, never was properly secured. The specification supports a number of communication protocols including Ethernet, token-passing, master-slave, and point-to-point communication [6].

## 2.2 IIOT COMMUNICATION PROTOCOLS

IoMT communications protocols make it easier for relatively close medical devices to connect using various information systems (IT)-related systems. Some of these protocols are listed as follows:

- *Bluetooth*: Bluetooth is a common wireless communication technology centered just on IEEE 802.15.1 specification. It is intended for minimal energy, fairly less priced gadgets and operates at 2.4 GHz. BLE is used in IoT networks that make use of battery-powered, minimal energy, fairly less priced devices. It enables equipment to rapidly connect to the network and build basic equipment interconnections or stellar networks. This primarily utilized in Internet of Things (IoT) projects which need narrow range communication, reduced lag, as well as limited frequency, such as Human Interface Devices (HID), sporting as well as wellness monitors, including portable healthcare equipment [7] [8].
- *LoRaWAN:* Semtech developed LoRa (Long Range) as a tangible level architecture to support reduced energy and broad connectivity. This may handle long-distance and low-power transmissions of up to 10 km in sparsely populated areas. Because LoRa specifies the physical layer, it was necessary to develop higher network communication standards on top of it. The LoRaWAN protocol seems to be a MAC based technique that is mostly used as a communication protocol.
- *MQTT:* MQTT (Message Queue Telemetry) was created by IBM well with goal of simplifying lighter M2M connections. That an asynchronous disseminate data interchange format that extend the functionality of the TCP layer at the application level, allowing programs and consumers to communicate over networks. The goal is to guarantee enough bandwidth and energy efficiency for devices with limited computation and storage capacity.

# 3. MACHINE LEARNING IN IDS

Machine Learning is the domain of research that enables computers to innovate and gain from experience whilst being explicitly programmed. Machine learning is concerned with the creation of algorithms that can learn from data. The learning process begins with observations or data in order to seek for patterns in data and generate better predictions based on the examples presented. The fundamental goal is to enable computers to learn without human intervention and change activities accordingly. Machine Learning Algorithms are widely classed as follows:

- **Supervised machine learning algorithms:** Using tagged examples, they may use what they have learnt in the past to anticipate future events. The algorithm analyses a training dataset to create an inferred function that may be used to predict output values. The algorithm will give predictions for new inputs after a sufficient amount of training. The computer is given a fresh collection of examples, so that the supervised learning algorithm can analyze the training data and generate accurate results from labelled data.
- **Unsupervised machine learning algorithms:** They are utilized when the training data is neither tagged nor categorized. Unsupervised learning investigates how computers may infer a function from unlabeled data to explain a hidden structure. The objective of the computer in this case is to group unsorted data based on trends, analogies, and distinctions without any prior training data.

# 4. LITERATURE REVIEW

## 4.1 MACHINE LEARNING ALGORITHMS IN CYBERSECURITY ICS

Chihta Lin et al. [9] built an ICS testbed to investigate two operating cases: water level management and air pollution control. They created an intrusion detection system that can learn on its own. Their findings demonstrate that their method can identify many types of network attacks.

Mohammad et al. [10] conducted many tests on the IDS to assess the productivity and effectiveness of machine learning methods. J48, random forest, decision tree, naive Bayes, and Bayes network are among the algorithms tested. The KDD intrusion detection dataset is used in the evaluation. Their tests have shown that there is no one ML method that can efficiently tackle all sorts of attacks. The KDD dataset, on the other hand, is out of date and cannot cover all types of cyberattacks.

In [11] the authors outline some real-world cyberattacks carried out on an electrical infrastructure. A heterogeneous ecosystem comprising SCADA assets (e.g., PLCs, HMIs, and process control data centers) manages a simulated power grid in this testbed. They describe several attacks and explore some of the obstacles that an attacker might face while carrying those attacks.

Keliris et al. [12] created a supervised methodology with protective strategic plan which recognizes attacks in real time by taking into account an ICS operational behavior. They used a standard chemical-based procedure on their hardware controllers and looked at a variety of threat avenues. A taught SVM system was used to detect abnormalities in actual time and distinguish among disruptions and suspicious attack.

The authors [13] present a deep learning-based multi-channel attack detection approach for information security in social networks. They also present a voting method for determining whether traffic is an attack or not, that further accomplishes better accuracy by voting to obtain the majority decision of multi-channel classifiers. The suggested technique is flexible, and the neural network selection may be altered according to the attributes of the feature vectors, making their method suitable for various jobs.

The authors [14] initially introduced typical IIoT protocols and their related vulnerabilities in this research. They then conducted a cyber-vulnerability evaluation and explored how machine learning may be used to mitigate these risks. Following that, an overview of the existing intrusion detection systems based on machine learning methods is provided. Finally, they discussed the case study, which details a legitimate test platform which the authors built in order to conduct cybercrime then develop effective intrusion detection system (IDS).

This study [15] focused on the creation of such a cybersecurity test platform for SCADA networks, and investigates the viability of using machine learning techniques to detect computer hackers in actual environments. The test platform is built with instrumentation that has been utilized in practical commercial situations. The results demonstrate the effectiveness of machine learning models in identifying attacks in real time.

## 4.2 SECURITY IN HEALTHCARE

The primary goal is to employ machine learning in healthcare to augment patient care and get better results. Machine learning has made it simpler to identify and properly diagnose many illnesses. Predictive analysis using efficient multiple machine learning algorithms aids in more accurate illness prediction and treatment of patients. The healthcare sector generates huge volumes of healthcare data on a regular basis, which may be utilized to extract information for forecasting disease that may occur in the future to a patient based on treatment history and health data. This concealed information in healthcare data will be used subsequently to make effective health decisions for patients.

Attacks on Internet-connected medical equipment has the potential to inflict serious physical fatalities to patients. In this study, authors [16] design and build a unique mobile agent-based intrusion detection system to protect the network of medical equipment that are linked together. The presented technique is multilevel, independent, and decentralized and uses machine learning and regression methods to identify network-level intrusions and abnormalities in sensor data. They mimic a hospital network architecture and conduct extensive tests on several subsets. Our simulation findings show that proposed method gets significant detection accuracy while using less resources.

The study in [17] seeks to create a technique based on an artificial neural network approach to anticipate suspicious devices based on bandwidth utilization. All of the investigations have been carried out in the healthcare field. The devices' mobility pattern was separated into six sections, each with its own specialized slice. The security module tracked all clients connected to slices on a frequent basis, and machine learning was used to detect and deactivate troublesome or suspicious devices.

The work in [18] describes an ensemble learning-based intrusion detection system that employs a fog-cloud architecture to detect and mitigate hostile activity in an IoMT context. The ensemble design incorporates Decision Tree, Naive Bayes, and Random Forest as first-level individual learners. The classification findings are then utilized by XGBoost to distinguish normal and attack occurrences at the following level. The experimental results demonstrate that the suggested framework can achieve a detection rate of 99.98%, an accuracy of 96.35%, and can reduce false alarm rate upto 5.59%.

The usage of Machine Learning (ML) techniques to well-known intrusion detection systems (IDS) is critical for dealing with rapidly evolving cybersecurity attacks through ensuring an effective and efficient detection procedure. Most ML-enabled IDS techniques in the context of the Internet of Things (IoT) employ centralized approaches in which IoT devices exchange their data with data centers for additional analysis. To address privacy problems associated with centralized methods, the usage of Federated Learning (FL) has received attention in a variety of industries, including healthcare and transportation systems. This research [19] presents a complete review of the usage of FL for IDS in IoT. They specifically assessed FL behavior by taking into account various data distributions, training rounds, and aggregation approaches.

Table.1. A brief comparison of the works described in Subsection 4.1

| Author | Objective | ML technique used | Dataset | Comments |
|---|---|---|---|---|
| [9] | Developing an intrusion detection system capable of self-learning. | J48, SVM, Naïve Bayes, Decision Tree. | Testbed | Water level management and air pollution control. |
| [10] | To evaluate several machine learning classifiers on the basis of the KDD intrusion dataset. | J48, Random Forest, Random Tree, Decision Table, MLP, Naive Bayes, and Bayes Network | KDD Intrusion Dataset | KDD dataset is obsolete and does not encompass all forms of cyberattacks. |
| [11] | To give a realistic explanation of various representative cyber-attacks against SCADA systems. | - | Testbed | Focused more on planning and executing attacks rather than detecting and mitigating it. |
| [12] | To create a real time machine learning classifier capable of detecting various different abstract | SVM | Testbed | Proposed model was able to identify all payloads previously undetected |

| | | | | |
|---|---|---|---|---|
| | types of attacks and distinguishing them from process disruptions. | | | with minimal delays while avoiding false alarms. |
| [13] | To present a neural network-based end-to-end intelligent threat detection technique in social networks. | Deep Learning | NSL-KDD | Proposed attack detection approach outperforms many other attack detection methods that employ feature detection and Bayesian or SVM classifiers |
| [14] | To investigate the efficiency of machine learning (ML)-based security solutions in dealing with each type of attack. | Random Forest, Decision Tree, KNN, Logistics Regression, SVM, ANN, Naive Bayes. | Testbed | The security features of these systems are investigated using backdoor code injection and SQL intrusion attacks. |
| [15] | To demonstrate the development of a SCADA system testbed for cybersecurity research, as well as research on the viability of using machine learning algorithms to identify computer security in timely manner. | Random Forest, Decision Tree, Logistic Regression, Naïve Bayes and KNN | Testbed | Main emphasis is laid on reconnaissance attacks. |

Table.2. Brief comparison of the works described in Subsection 4.2

| Author | Objective | ML technique used | Dataset | Comments |
|---|---|---|---|---|
| [16] | To create a new mobile agent-based intrusion detection system to protect the network of interconnected medical devices. | SVM, DT, NBC (Naive Bayes Classifier), KNN, and RF. | - | Simulating a hospital network architecture and conducting thorough experiments for various subsets of the Internet of Medical Things. |
| [17] | To deliver a low-cost, artificial intelligence-powered security solution for the IoT-enabled healthcare environment. | Deep Learning | MATLAB neural network | The effective use of artificial neural networks identifies and blocks potentially malicious devices. |
| [18] | To develop an intrusion detection system (IDS) based on an ensemble learning method to detect and mitigate cyber-attacks in an IoMT environment | Ensemble Learning | ToN_IoT | The proposed ensemble model outshines certain current state-of-the-art approaches. |
| [19] | To develop FL-enabled IDS for IoMT. | Logistic regression | ToN_IoT | FL based solution for IDS is still in its early stages, and existing approaches generally rely on unrealistic settings and data distributions. |

## 5. CONCLUSION AND EMERGING PERSPECTIVES

Safety is paramount for IIoT and IoMT gadgets. Here remains currently a considerable deficit for maintaining adequate protection for all these devices, which is why concentrating on these IoT applications is so important. In order to provide a secure platform, machine learning technologies and dataset analysis have indeed been increasingly used in IT systems. Although IoT-based healthcare technology is still in its early stages of development, its incorporation into the medical business has substantially enhanced medical delivery and patient treatment levels. Securing the data-driven in the medical care environment would most likely contribute to making correct judgments for patient treatment delivery; empowering clinicians to accurately diagnose the patient health condition; and enhancing emergency response times.

## REFERENCES

[1] Modbus, "The Modbus Organization", Available at https://modbus.org/, Accessed at 2021.

[2] Bacnet, "BACnet Website", Available at http://www.bacnet.org/, Accessed at 2021.

[3] DNP, "Overview of DNP3 Protocol", Available at https://www.dnp.org/About/Overview-of-DNP3-Protocol, Accessed at 2021.

[4] MQTT, "MQTT - The Standard for IoT Messaging", Available at https://mqtt.org/, Accessed at 2021.

[5] IEEE Standards, "IEEE SA - The IEEE Standards Association - Home", Available at https://standards.ieee.org/, Accessed at 2021.

[6] A. Antonini, A. Barenghi, G. Pelosi and S. Zonouz, "Security Challenges in Building Automation and SCADA", *Proceedings of International Carnahan Conference on Security Technology*, pp. 1-6, 2014.

[7] Cypress Perform, Available at "Bluetooth Low Energy (BLE)", Available at https://www.infineon.com/dgdl/Infineon-Component_BLE_V2.0-Software%20Module%20Datasheets-v03_66-EN.pdf?fileId=8ac78c8c7d0d8da4017d0eae085e299e, Accessed at 2015.

[8] Texas Instruments, "Bluetooth low energy Protocol Stack Introduction", Available at http://software-dl.ti.com/lprf/simplelink_cc2640r2_sdk/1.00.00.22/exports/docs/blestack/html/ble-stack/index.html#:~:text=The%20Bluetooth%20low%20energy%20protocol%20stack%20(or%20protocol%20stack)%20consists,two%20sections%20are%20implemented%20separately, Accessed at 2021.

[9] C.T. Lin, S.L. Wu and M.L. Lee, "Cyber Attack and Defense on Industry Control Systems", *Proceedings of IEEE International Conference on Dependable and Secure Computing*, pp. 1-3, 2017.

[10] M. Almseidin, M. Alzubi, S. Kovacs and M. Alkasassbeh, "Evaluation of Machine Learning Algorithms for Intrusion Detection System", *Proceedings of IEEE 15th International Symposium on Intelligent Systems and Informatics*, pp. 277-282, 2017.

[11] L. Rosa, T. Cruz, P. Simoes, E. Monteiro and L. Lev, "Attacking SCADA Systems: A Practical Perspective", *Proceedings of IEEE International Conference on Integrated Network and Service Management*, pp. 741-746, 2017.

[12] A. Keliris, H. Salehghaffari, B. Cairl, P. Krishnamurthy, M. Maniatakos and F. Khorrami, "Machine Learning-Based Defense against Process-Aware Attacks on Industrial Control Systems", *Proceedings of IEEE International Conference on Integrated Networks*, pp. 1-10, 2016.

[13] F. Jiang, "Deep Learning based Multi-Channel Intelligent Attack Detection for Data Security", *IEEE Transactions on Sustainable Computing*, pp. 1-10, 2018.

[14] M. Zolanvari, M.A. Teixeira, L. Gupta, K.M. Khan and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things", *IEEE Internet of Things*, Vol. 6, No. 4, pp. 6822-6834, 2019.

[15] 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, pp. 76-88, 2018.

[16] G. Hatzivasilis, O. Soultatos, S. Ioannidis, C. Verikoukis, G. Demetriou and C. Tsatsoulis, "Review of Security and Privacy for the Internet of Medical Things (IoMT)", *Proceedings of IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 457-464, 2019.

[17] A. Jain, T. Singh and S. Kumar Sharma, "Security as a Solution: An Intrusion Detection System using a Neural Network for IoT Enabled Healthcare Ecosystem", *Interdisciplinary Journal of Information, Knowledge, and Management*, Vol. 16, pp. 331-369, 2021.

[18] P. Kumar, G.P. Gupta and R. Tripathi, "An Ensemble Learning and Fog-Cloud Architecture-Driven Cyber-Attack Detection Framework for IoMT Networks", *Computer Communications*, Vol. 166, pp. 110-124, 2021.

[19] E.M. Campos, "Evaluating Federated Learning for Intrusion Detection in Internet of Things: Review and Challenges", *Proceedings of IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 1-13, 2021.