# Securing the Internet of Things and Wireless Sensor Networks via Machine Learning: A Survey

Marwa Mamdouh, Mohamed A. I. Elrukhsi, Ahmed Khattab
Department of Electronics and Communications Engineering,
Faculty of Engineering Cairo University
Giza, Egypt
marwa_mamdouh90@cu.edu.eg, e.m.apollo@gmail.com, akhattab@ieee.org

***Abstract*— The Internet of Things (IoT) is the network where physical devices, sensors, appliances and other different objects can communicate with each other without the need for human intervention. Wireless Sensor Networks (WSNs) are main building blocks of the IoT. Both the IoT and WSNs have many critical and non-critical applications that touch almost every aspect of our modern life. Unfortunately, these networks are prone to various types of security threats. Therefore, the security of IoT and WSNs became crucial. Furthermore, the resource limitations of the devices used in these networks complicate the problem. One of the most recent and effective approaches to address such challenges is machine learning. Machine learning inspires many solutions to secure the IoT and WSNs. In this paper, we survey the different threats that can attack both IoT and WSNs and the machine learning techniques developed to counter them.**

***Keywords*— Security; Machine Learning; Neural Networks; Internet of Things; Wireless Sensor Networks.***

## I. INTRODUCTION

The Internet of Things (IoT) is based on Wireless Sensor Networks (WSNs) in which devices are connected to the Internet and communicate with each other without human involvement. Security in WSNs and IoT is a crucial and not easy problem. On one hand, the battery-powered WSN and IoT devices are typically inexpensive and resource-limited. Therefore, advanced security techniques that consume high power are not appropriate. On the other hand, the highly dynamic nature of such distributed networks makes centralized security and key management algorithms not applicable. In addition to the nature of wireless communication which is prone to attacks, the individual nodes are susceptible to physical capture.

Machine learning is a promising solution to address the security threats in WSNs and IoT [1, 2]. Machine learning, a type of artificial intelligence, uses different learning algorithms to train the devices without explicit programming. The nature of machine learning is appropriate for WSNs and IoT [1] for the following reasons: (1) mathematical models cannot be built for complicated IoT and WSN environments. (2) Some applications use data sets which need to be correlated. (3) Machine learning can work with the dynamics and unexpected behavior of WSNs and IoT. (4) Machine learning algorithms do not need human intervention that suits the nature of WSNs and IoT. However, two main challenges face machine learning in WSNs and IoT: The resource and computation limitation of the nodes, and the need for large data sets for learning.

In this paper, we present an overview of the security attacks in WSNs and IoT (Section II), and the different machine learning algorithms (Section III). We survey the existing works that address the security problems in IoT and WSNs using machine learning (Section IV). Finally, we draw our conclusions and insights (Section V).

## II. SECURITY ATTACKS IN IOT AND WSNS

In this section, we first present a high level classification of the security attacks that target WSN and IoT systems. Then, we elaborate on the most critical attacks which are tackled by the existing literature.

### A. Classification of Attacks

*1) Goal-Oriented Attacks:* Such attacks threaten data confidentiality and can be either passive or active. Passive attackers obtain sensitive information such as encryption keys without the awareness of the legitimate users. They use such information in the decryption of weakly encrypted data. Eavesdropping and traffic analysis are examples of passive attacks. Meanwhile, active attackers monitor the network and obtain sensitive information to have the ability to control the network and change this information. Example active attacks are: Denial of Service (DoS), Sybil attacks, hole attacks, jamming and spoofing.

*2) Performer-Oriented Attacks:* These attacks can be classified into inside and outside attacks depending on the location of the attacker with respect to the network. In inside attacks, the attacker is one of the legitimate nodes. Hence, it can access important information such as decryption keys. Therefore, it is difficult to detect this type of attacks. Internal attacks could cause misrouting, packet drop, eavesdropping and modification of data. Meanwhile, outside attackers may send large amounts of data to cause congestion in network or exhaustion of the resources such as DoS attacks.

*3) Layer-Oriented Attacks:* Such attacks are categorized according to which layer of the protocol stack is targeted. As shown in Fig. 1, every layer is prone to different attacks. For instance, the data link layer can be attacked by: (1) data flooding in which legitimate nodes that use carrier sensing protocols to access the channel will suffer from high probabilities of collision, (2) unfairness attacks where malicious nodes send large numbers of packets without waiting for reasonable time to allow other nodes to access the channel, and (3) exhaustion attacks in which malicious nodes send large numbers of request-to-send messages to deplete the batteries of other nodes.

### B. Critical IoT and WSN Attacks

In what follows, we elaborate on the security attacks that are mostly encountered in IoT and WSNs.

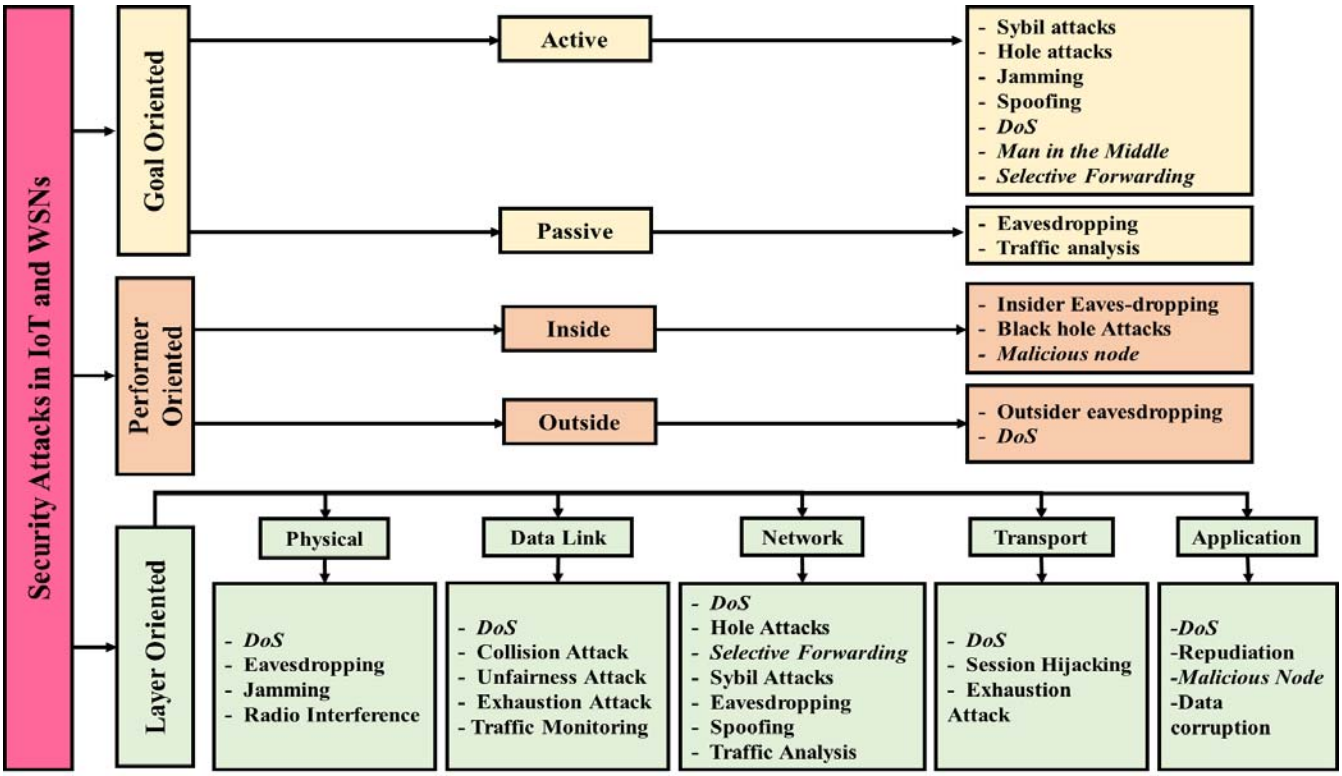


Fig. 1. Classification of security attacks in WSNs and IoT.

*1) Denial of Service (DoS) Attack:* DoS attacks target the availability of service. It can make some of the users unavailable or prevent the legitimate users to communicate. Moreover, it can cause the users of the network to take wrong decisions. DoS attacks may force the IoT devices to be always ON to deplete their batteries.

*2) Man in the Middle Attack:* The attacker disguises itself as a device that is already in direct connection with another device. Then, it can eavesdrop on the ongoing communication, inject false or tampered information, or fully intercept the communication.

*3) Selective Forwarding Attack:* Attacker nodes select some packets to pass in the network and drop the rest as all packets pass through these nodes. This causes what is called a hole in the network when all the packets from some nodes are dropped.

*4) IoT Device Vulnerability Threats:* Such threats are caused by introducing new devices to the network. These devices might have security vulnerability or been infected by a malware. The vulnerabilities in these devices can be exploited by the attackers. For example, network address translation (NAT) hole punching can be created to allow remote attackers to gain access to the network for malicious reasons such as to infiltrate data, compromise other devices, or inject tampered or false information to the network.

## III. Overveiw of Machine Learning Techniques

Machine learning algorithms are typically categorized to supervised, unsupervised and reinforcement learning. In this section, we briefly overview each of these techniques.

### A. Supervised Learning

In supervised learning, known inputs and their corresponding outputs are given for learning. Such information helps the machine to identify the output for other inputs. The key supervised learning algorithms are:

*1) k-nearest Neighbour:* In this algorithm, the reading of an unknown node is the average of the *k* nearest neighbours identified by the Euclidian distance [5]. For example, if the reading of a WSN node is missing, it is predicted from the average readings of the neighbouring nodes in a certain area. It is a simple algorithm in its computations. However, it gives inaccurate results in case of large training sets and high dimensions.

*2) Support Vector Machine (SVM):* SVM is used for classification by finding a hyper plane between two classes. To find such a plane, SVM aims at maximizing the margin (distance between nearest points) and differentiates between the two classes with minimal errors. SVM uses a kernel function if no linear hyper plane can be found for classification. This function adds new features to have the classes linearly separated. SVM gives high accuracy, and hence, is widely used to address security problems in IoT and WSNs.

*3) Neural Network (NN):* By analogy to the nervous system and how brain works, NN uses neurons. Such a biology-inspired system consists of different layers. NN can solve nonlinear and complex problems. However, it is complex in its computations. Therefore, it is difficult to use in distributed IoT and WSN systems.

*4) Bayesian:* This supervised algorithm does not need a large data set for classification. It uses probability

distributions for learning. Then, it gets the new probabilities based on the current knowledge. Bayesian classifiers need prior knowledge about the environment which limits their use in IoT and WSN systems.

### B. Unsupervised Learning

In unsupervised learning, no outputs are given. Only inputs are used in learning. Based on these inputs, the system classifies them into groups called clusters. A new input then can be classified to the right group.

*1) Principal Component Analysis (PCA):* PCA gets important information from data sets and defines it as new orthogonal values to define new coordinates. This method reduces the amount of needed data. Therefore, it turns large data sets to smaller ones.

*2) k-means Clustering:* In this algorithm, a data set is categorized into clusters. First, it chooses *k* random centroids. Then, nodes are grouped into clusters of the nearest centroid. The algorithm recalculates the centroids by taking the average of the nodes in each cluster, then repeats the previous steps until it converges.

### C. Reinforcement Learning

Reinforcement learning does not have known inputs with corresponding outputs. This relationship is learnt by a reward scheme. If its performance in doing tasks is high, it is given a reward. This algorithm interacts with the surrounding environment for learning. Q-learning is an example of reinforcement learning.

## IV. Machine Learning Security in IoT and WSNs

In this section, we overview the different works that apply machine learning to secure IoT and WSNs.

### A. Counter DoS Attacks

SVM and NN machine learning were used to detect DoS attacks at the WSN medium access control layer in [6]. SVM and NN depend on two variables to train their machines: collision rate and arrival rate. All nodes send both rates to calculate the probability of DoS attack. In NN, if this probability is greater than a threshold, the node is a victim of a DoS attack. Therefore, it shuts itself down and resumes working when the attack is over. In SVM, two classes classify the probability of a DoS attack to either Low or High. It was shown that SVM accuracy is better with a shorter time to detect the attack.

In [7], the authors used multi-layer deep learning with 4 layers of neurons. They assume a two-layer network architecture: An IoT layer that contains the sensor/actuator infrastructure, and a fog layer which hosts the shared computation and storage of the network. This architecture aims at elevating the training and the attack detection burdens to the fog nodes to compensate the IoT edge node resource limitations.

Finally, a comparison of the ability of the different classification algorithms to identify the normal data in DoS scenarios is presented in [8].

### B. Counter Selective Forwarding Attacks

In [9], the authors presented a one-class SVM to detect selective forwarding and black hole attacks. A simple intrusion detection system is presented to save memory and energy. SVM depends on two variables for classification: bandwidth and hop count. This algorithm assumes that nodes do not consume more energy to add security. None of the works discussed in [10] apply machine learning to detect selective forwarding attacks.

### C. Counter Man in the Middle Attacks

Artificial NN was used in [11] with three input neurons (device ID, sensor value, and delay), two hidden neurons and five layers with three hidden layers. Using already available packages in R programming language, the proposed NN monitors the health of the node. If the values deviate from the expected value, this indicates false information or a man in the middle attack.

A watermark technique secures the communication between an IoT device and the cloud in [12]. It adds pseudorandom noise and calculates a bit stream at the IoT device. If the bit-stream extracted at the receiver does not match, the attack alarm starts. If the bit-stream was static, it can be cracked. Therefore, the authors proposed using a recurrent NN machine learning approach called Long Short Term Memory (LSTM) to generate dynamic bit-stream from features such as spectral flatness, mean, variance, skew, and kurtosis.

Reinforcement learning was used to secure mobile edge caching, which is dedicated devices located at the network edge to support storage and computational services to the network users including the IoT [13].

### D. Machine Learnig-Based IoT Device Identification

IoT SENTINEL classifies newly installed devices in home or small office networks to either trusted, strict, or restricted devices [14]. The gateway is responsible for monitoring the traffic generated by the newly installed devices, creating device finger prints that are delivered to an IoT security service provider, which in turn classifies the device according to its type and traffic generated using a machine learning-based classification model.

Random Forest supervised machine learning was used in [15] to map a traffic stream to a device type to overlook new devices in large enterprises. If the traffic could not be mapped to one of the types in the created white list, then this device is not authorized.

Machine learning was used in [16] to differentiate between IoT and non-IoT devices based on data traffic. Sessions coming from each device are used for classification. Features are extracted from the different layers. For training, eight IoT devices, smart phones and PCs are used. Another classifier is used to know the IoT device type by monitoring a few consecutive sessions.

### E. Machine Learning Security with Bio-Inspiration

The algorithm presented in [17] applied machine learning to discover if a benevolent node became a malicious one. Furthermore, bio-inspiration is applied as an immune system to revoke the effect of malicious nodes. First, the *k*-means algorithm generates two clusters: normal cluster and faulty

cluster. SVM is then applied to create a decision block that has three regions: normal region, fault region and critical region at boundaries. Based on an anomaly detection algorithm, the mean and standard deviation of the normal nodes presented by SVM dataset is calculated. Then, anomaly is detected, and the immune system is activated. By analogy to biological systems, virtual antibodies are generated, and finally malicious nodes are shut down.

### *F. Machine Learning for Secure IoT Access Control*

The IoT system access control problem was changed from being centralized to distributed in [18] to negate the single point of failure and to increase the privacy by keeping the information at the edge points. A blockchain scheme (which is a distributed list of all the transactions in the network) is used to facilitate the communication between non-trusted members without a trusted intermediary. Machine learning was used to upgrade and improve the control policy using reinforcement learning and introduces the concept of "*SmartContract*" which is an executable code to grant or deny the permission for a request. It provides access tokens to the allowed requests.

## V. Concluding Remarks

In this paper, we have presented a survey of the different machine learning algorithms used to secure WSNs and IoT. Table I summarizes the different machine learning approaches used to secure IoT and WSNs. While many approaches result in high accuracy, SVM is less complex than NN in classification. There are still many challenges in securing IoT and WSNs because machine learning should compromise between a high level of security and a low computational complexity to be suitable for the resource-limited IoT and WSN devices.

TABLE I. Machine learning algorithms to secure IoT and WSNs.

| Paper | Attack | Learning Approach | Complexity | WSN/ IoT |
|---|---|---|---|---|
| Security Enhancement [6] | DoS | SVM | Low | WSN |
| | | NN | Moderate | |
| Distributed attack detection [7] | DoS | NN | High | IoT |
| Selective forwarding attacks [9] | Selective Forwarding | SVM | Moderate | WSN |
| Machine Learning to secure IoT [11] | Man in the Middle | ANN | High | IoT |
| Dynamic Watermarking [12] | Man in the Middle | RNN LSTM | High | IoT |
| IoT Sentinel [14] | Traffic Monitoring | Not Specified | Moderate | IoT |
| Unauthorized IoT Devices [15] | Traffic Monitoring | Supervised Learning | Moderate | IoT |
| ProfilIoT [16] | Traffic Monitoring | Supervised Learning | Moderate | IoT |
| Bio-Inspiration [17] | Malicious node | K-means & SVM | Moderate | WSN |
| Dynamic Access Control Policy [18] | Different attacks | Blockchain | High | IoT |

## References


[1] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surv. Tutor.*, vol. 16, no. 4, pp. 1996–2018, 2014.

[2] I. Butun, S. D. Morgera, and R. Sankar, "A survey of intrusion detection systems in wireless sensor networks," *IEEE Commun. Surv. Tutor.*, vol. 16, no. 1, pp. 266–282, 2014.

[3] K. Shabana, N. Fida, F. Khan, S. Jan, and M. Rehman, "Security issues and attacks in Wireless Sensor Networks," *Int. J. Adv. Res. Comput. Sci. Electron. Eng.*, vol. 5, no. 7, pp. 81-87, 2016.

[4] K. Chelli, "Security issues in wireless sensor networks: attacks and countermeasures," in *Proc. of World Congress on Engineering*, 2015.

[5] F. Chen, P. Deng, J. Wan, D. Zhang, A. Vasilakos, and X. Rong, "Data mining for the internet of things: literature review and challenges," *Int. J. Distrib. Sens. Netw.*, vol. 11, p. 431047, 2015.

[6] A. B. Raj, M. V. Ramesh, R. V. Kulkarni, and T. Hemalatha, "Security enhancement in wireless sensor networks using machine learning," in *Proc of IEEE HPCC-ICESS*, 2012.

[7] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," *Future Gener. Comput. Syst.*, 2017.

[8] V. Singh, S. Puthran, and A. Tiwari, "Intrusion detection using data mining with correlation," in *Proc. of International Conference for Convergence in Technology (I2CT)*, 2017.

[9] S. Kaplantzis, A. Shilton, N. Mani, and Y. A. Sekercioglu, "Detecting selective forwarding attacks in wireless sensor networks using support vector machines," in *Proc. of Int. Conf. on Intell. Sensors, Sensor Net. and Information (ISSNIP)*, 2007.

[10] N. M. Alajmi and K. M. Elleithy, "Comparative analysis of selective forwarding attacks over Wireless Sensor Networks," *Int. J. Comput. Appl.*, vol. 111, no. 14, 2015.

[11] J. Cañedo and A. Skjellum, "Using machine learning to secure IoT systems," in *Proc. of 14th Annual Conference on Privacy, Security and Trust (PST)*, 2016.

[12] A. Ferdowsi and W. Saad, "Deep Learning-Based Dynamic Watermarking for Secure Signal Authentication in the Internet of Things," *ArXiv171101306*, 2017.

[13] L. Xiao, X. Wan, C. Dai, X. Du, X. Chen, and M. Guizani, "Security in Mobile Edge Caching with Reinforcement Learning," *ArXiv180105915*, 2018.

[14] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A.-R. Sadeghi, and S. Tarkoma, "IoT Sentinel: Automated device-type identification for security enforcement in IoT," in *Proc. of IEEE Int. Con. on Distributed Computing Systems (ICDCS)*, 2017.

[15] Y. Meidan et al., "Detection of Unauthorized IoT Devices Using Machine Learning Techniques," *ArXiv170904647*, 2017.

[16] Y. Meidan et al., "ProfilIoT: a machine learning approach for IoT device identification based on network traffic analysis," in *Proc. of the Symposium on Applied Computing*, 2017.

[17] H. Rathore, V. Badarla, S. Jha, and A. Gupta, "Novel approach for security in wireless sensor network using bio-inspirations," in *Proc. of Int. Conf. on Comm. Sys. and Net. (COMSNETS)*, 2014.

[18] A. Outchakoucht, E.-S. Hamza, and J. P. Leroy, "Dynamic Access Control Policy based on Blockchain and Machine Learning for the Internet of Things," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 7, pp. 417–424, 2017.