Review of Machine Learning Techniques for Detection of Routing Attacks in Wireless Sensor Network

Shraddha Sarode
Assistant Professor
SVKM's NMIMS, Sunandan Divatia School of Science
Mumbai, India

ABSTRACT
Wireless sensor networks are widely applied in many fields like transportation, urban terrain tracking, healthcare, precision agriculture, etc. However, this deployment has introduced new security concerns. These security concerns involve two kinds of attacks on wireless sensor networks active and passive. Passive attacks are launched to observe the network without disrupting network functionality. Active attacks can disrupt the function of the network and can be initiated on layers of communication protocol. Active attacks that are related to network layer routing attacks are presented. For the last decade, machine learning algorithms have been used in many important applications, including detection of routing attacks. The objective of this paper is to review machine learning algorithms that can be used to detect routing attacks in wireless sensor networks. In this paper, evaluation parameters and challenges of applying machine learning algorithms in wireless sensor networks are also discussed. These challenges can serve as potential future research directions.

Keywords
Machine learning, wireless sensor network security, routing attacks, detection, security.

1. INTRODUCTION
Wireless sensor network is a self-organizing infrastructure less network of tiny, low price, low power, battery operated nodes. These nodes are typically used to sense environmental parameters like temperature, pressure, humidity, vibrations etc. Each node wirelessly transmits sensed data to central sink node or base station. Being battery operated wireless data transmission is a power expensive operation and cannot be repeated frequently. If an attacker was to inject malicious data packets into a wireless sensor network, it would drastically affect the sensing capability of the network. There by hampering the monitoring of asset under construction.

Machine learning has capability to derive meaning from huge heaps of data. Machine learning has already established in image, text and speech recognition [1]. Google's machine learning diabetic retinopathy algorithm has received CE mark clearance and is being adopted by ministry of public health, Thailand [2].

Machine learning focuses on learning from data without the need of program. Machine learning is already used for outlier detection, localization, Coverage & connectivity, fault detection etc. of wireless sensor network [1]. It is also effectively adopted for predicting future events [3] based on current wireless sensor network data. In this paper machine learning techniques which are applied for detection of routing attacks in wireless sensor network are discussed.

The rest of the paper is organized as, in section II outlines routing attacks in wireless sensor networks, in section III, review of machine learning algorithms to detect routing attacks in wireless sensor network is given, Section IV discusses comparison of machine learning algorithms based on different parameters. After that conclude the discussion with some future direction.

2. RELATED WORK
Wireless sensor network is made up of multiple nodes connected with each other. Basic use of these connections is for communication which is information exchange between nodes. This suggests that to exchange information certain path must be taken to reach destination node. The process of selecting an optimal path to exchange particular information is known as routing. This process is carried out in network layer to exchange information. This information is transmitted through the packets.

Communication of sensor nodes is threatened by various types of security attacks by preventing one or more network devices that perform routing functions such as an unauthorized attacker monitors, listen to and modifies the data stream in the connections. These security attacks are divided of two kinds, active and passive [4] based on damage or access level. Active attacks can be launched using any of the layers of communication protocol with different intentions like disruption of network or dropping of packets.

Support vector machine is used for detection of black hole attack in [12] for selective forwarding & wormhole attack in [6], [13] for Sybil attack in [10]. SVM was used for full dataset containing black hole, grey hole attacks along with flooding and scheduling attacks after that it was also tested for reduced dataset containing only grey hole and flooding attack [14]. It is also applied to detect Gray hole attack in [14]. In [15] the authors reviewed that one-class SVM was used detect selective forwarding and black hole attacks. Lower dimensional data is converted into higher dimensional data in SVM to classify non-linear data using kernels. Radial Basis Function is very common choice for this kernel trick, which is used to detect sinkhole attack in [16].

Naïve Bayesian is applied in detection of Sybil attack, sinkhole attack and hello flood attack [13]. It is used as a part of enhanced code-based roundtrip time-based method prevent black hole and worm hole attacks [17]. Naïve Bayesian is widely used for defending this kind of attack where its accuracy rate is ranging from 98 to-99 for detection of hello flood, Sybil and sinkhole attack [6].

C4.5 is used in [13] to detect black hole attack. Decision trees are also used to detect sinkhole attacks in [16], [1]. These algorithms are fast and make accurate predictions. J48 has been applied in [7] to detect grey hole and black hole attack.

Decision tree was built with J48 in [14] for full dataset containing black hole, grey hole attacks along with flooding and scheduling attacks after that it was also tested for reduced dataset containing only grey hole and flooding attack. Random forest which is averaging of output given by multiple decision trees that are created using random selection of variables. This technique is used by authors to detect flooding attacks [7].

Artificial neural network can be applied to detect faulty sensor nodes [1]. Neural networks are also used to detect flooding attacks [12] k-nearest neighbour is also a classification, machine learning technique. The size of the input dataset affects the performance of the k-NN, it is known as lazy learner, because it doesn't learn a discriminative function from the training data but "memorizes" the training dataset instead. Authors have used k-NN algorithm to classify sensor nodes in [19].

Partitioning based clustering technique k-means is used for detection of black hole attack and sinkhole attack in [13]. Authors have reviewed that k means clustering along with LEACH protocol is used for detection of black hole attack [20], [28]. This algorithm is also used by authors to cluster legitimate and attacker nodes in wireless sensor network in [8]. k- medoid is also a partitioning based unsupervised clustering technique was reviewed by authors which has been used for black hole attack detection [20].

3. TYPES OF ROUTING ATTACKS IN WIRELESS SENSOR NETWORK

In this section attacks on the network layer are discussed, also known as routing attacks because these attack includes injecting control in the sensor node itself. Routing attack involves black hole attack, sybil attack, grey hole attack, selective forwarding attack, sinkhole attack, hello flood attack, wormhole attack. [5]

3.1 Black hole Attack

In this attack, there is one single malicious node which misleads other nodes in the wireless sensor network. There are many activities performed by this node that are, this node accumulates the information received and without forwarding packets, it drops them later on. To accumulate this information as well as to intercept the packets this node declares itself in the optimal path selected for communication by using routing protocols, such as ad-hoc on-demand distance vector (AODV) [4]

The specialized communication pattern and multi hop nature of the sensor network make it susceptible to this attack. When energy is selected as a metric to decide cluster head specifically in low energy adaptive clustering hierarchy (LEACH) protocol the malicious node gets selected as its energy is higher than other nodes. This results into receiving data from cluster members, accumulate that data and later on do not forward the data to the base station. [6]

3.2 Gray hole Attack

This attack is a variation of black hole attack; it also has one malicious node. This malicious node involves interception of packets, data fabrication, dropping of packets, launch of other active attacks. [4] This node does not necessarily drop all packets but it may forward some packets. Gray hole attack node can behave normally which makes it random and difficult to detect. Malicious node tries to become cluster head in Gray Hole attack by advertising itself using LEACH protocol. [7]

3.3 Worm hole Attack

Two malicious nodes play role in this type of attack. A tunnel is created between these two nodes. This tunnel is used to pass the messages between conniving nodes because this route of tunnel is posed as a low latency link between two malicious nodes. Such tunneling between these two conniving sensor nodes is called wormhole. This attack misdirects packets by introducing false routes, it has the ability to change network topology. The tunnel can be formed by sending copied packets over a wired network or using boosting long-distance antennas transmitting over low-latency routes [8]. This attack becomes effective when coupled with selective forwarding and Sybil attack, where it is very difficult to detect.

3.4 Selective forwarding

It is a special case of black hole attack, similar to black hole attack this also has single malicious node. In this attack malicious sensor node selects few packets to forward, unlike black hole attack it does not drop all packets. The reason behind this is that if other legitimate nodes found that this sensor node is not forwarding packets then alternate route can be selected to route packets, to avoid this sensor node, which also limits suspicion of being malicious sensor node. [9]

3.5 Sinkhole Attack

In this attack sensor node tries to advertise its updated routing information. This information is then used by neighboring nodes to transmit packets through this compromised sensor node. This process of gathering traffic is called sinkhole attack. All the information exchange happens through this conniving sensor node. Wireless sensor network becomes vulnerable to this kind of attack because of the many to one communication pattern. Sinkhole attack can be used to launch other attacks like selective forwarding attack, acknowledge spoofing attack and drops or altered routing information. [4]

3.6 Hello Flood Attack

In this type of attacks HELLO packets are routed by malicious node usually with high transmission power which results in making the legitimate nodes believe that malicious node is their neighbour. This attack is used to consume most of the network assets like computational and battery power, bandwidth and thus utilizing the node's resources or to interrupt a routing function to deprive the operation of wireless sensor network. Generation of false or misleading routes, packet loss, confusion in routing are some of the impacts of this attack. [5]

3.7 Sybil Attack

This attack involves masquerading technique to pose single sensor node as set of multiple sensor nodes. In this attack routing table is modified, it provides fault sensor readings, also leads to packet loss/ corruption. In such an attack, a malicious sensing node shows various IDs to other nodes that are component of the network. Many legitimate and compromised nodes are used together to get entry in the wireless sensor network. The Sybil attack mainly focused on fault tolerant schemes such as dispersity, topology maintenance and multipath routing. [10]

4. MACHINE LEARNING TECHNIQUES FOR HANDLING ROUTING ATTACKS IN WIRELESS SENSOR NETWORK

Various applications of wireless sensor network are discussed in [11]. With these increasing applications, attacks on wireless

sensor networks are bound to increase as well. As detection of routing attacks is considered, this section will present machine learning techniques that are used in detection of these attacks.

4.1 Support Vector Machine (SVM)

It is extremely popular algorithm for its capability to work with non-linear data. It is supervised learning algorithm. Linear discriminant function is used to separate two classes. When the data is two dimensional then equation of this discriminant function represents straight line, when the data is three dimensional the equation of the discriminant function represents plane and it represents hyperplane when more than three dimensions are present in input data. Weight and bias are two different terms used in equation decide orientation and position of this separating function in d dimensional space. There are support vectors and feature vectors in data. Support vectors are those points which lie closest and or on the two-class separating line. The distance between the decision line and the support vectors is computed. This distance is called the margin. This decision line is surrounded by margin. This algorithm tries to maximize margin in order to avoid misclassification.

4.2 Naïve Bayesian (NB)

This algorithm is inspired from Bayes theorem. It is also a supervised learning algorithm. The basic assumption in this is that it considers all the features to be independent. It is a technique that works well for large datasets. Wireless sensor network has large number of sensor nodes; they can be classified using naïve Bayesian algorithm [8]. Basic assumption of naïve Bayesian classifier is that all terms are independent of each other. To classify attacks classifier uses prior probabilities and posterior probability. These prior probabilities of attack and features along with likelihood are later on used to calculate posterior probability. Prior probabilities are calculated based on occurrence of different features given in the dataset. Likelihood is a conditional probability of feature given attack or class. [25] To avoid zero probabilities while calculating this likelihood, in case of features being absent Laplacian correction is adopted in naïve Bayesian algorithm. Laplacian correction takes one as an initial count for occurrence of feature. Based on number of attacks involved in the dataset, appropriate Naïve Bayesian model can be selected such as multinomial and Bernoulli binomial model.

4.3 Decision Tree (DT)

It is a tool used to take decisions about certain process. It has tree like structure specifying different features at different levels and nodes of tree. Leaf nodes contains decision labels or outcomes. Variables are split based on some criteria like entropy and Gini index. To form this decision, tree many algorithms like ID3 (iterative dichotomiser3), C4.5(J48) etc. J48 is an open-source java implementation of the C4.5 algorithms in the WEKA data mining tool.

4.4 Random forest (RF)

This algorithm is a supervised learning algorithm which involves creation of multiple decision trees is called as forest. These individual decision trees are created with combination of variables. They work efficiently for heterogeneous and big datasets. These trees are built using bagging technique. This algorithm averages output from different trees and gives result which helps to get accurate result and also treats missing values well.

4.5 Artificial neural network (ANN)

It is also a supervised learning algorithm; it is arranged in the form of layers. Each layer consists of processing units called neurons. Most of the times there are three layers namely, input, hidden and output. In artificial neural network each neuron is connected to all neurons in the next layer. Information can be processed in forward and forward-backward (feedback artificial neural network) direction. Activation functions are mainly used to get the output from neurons. Many activation functions like sigmoid, ramp etc. are used widely. Artificial neural network can be used for large and non-linear datasets. For large datasets it increases complexity and training time

4.6 K nearest neighbors (k-NN)

It is a simple and easy to implement algorithm. It can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. K nearest neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). The algorithm gets significantly slower as the number of data points and variables increase. It calculates distances between the samples in the dataset. K value suggests minimum neighbours or closest samples for selected data point. Many distance metrics can be used to find the nearest feature vectors, such as Euclidean distance, Minowski distance, hamming distance etc. can be used. Selection of these distance metrics changes the output of an algorithm. When the k-NN algorithm is used for classification, the object is assigned to the most common class among its k nearest neighbours.

4.7 Principal Component Analysis (PCA)

This algorithm is generally used for reducing dimensionality of dataset. Original data is reduced by removing variables with less information contained and is constructed using Eigen vectors. Eigen values gives the overall idea about the component being principal along with trace value. In [24] principal component analysis was used to reduce the dataset.

4.8 k-Means Clustering

This algorithm starts with selection of first k random centroids. Sensor nodes are clustered after calculating distances between the centroids and them. [25] The algorithm keeps on calculating mean of sensor nodes to find centroids for each cluster and updates clusters again by calculating distances. This process is repeated until clusters stop changing for at least two iterations.

5. DISCUSSION

From the related work, it is clear that support vector machine algorithm is adopted in all routing attacks and k-NN and PCA are used in literature but the serve different purposes. The Bayesian, Random Forest and k-NN algorithms are applied to classify sensor nodes where as PCA is applied for dimensionality reduction.

In this paper, a review is provided of machine learning techniques for detection of various routing attacks. Machine learning algorithms can be compared based on datasets because based on different data algorithms can adopt different functions or number of iterations. Performance metrics, time to train the algorithm [21], interpretation of output etc. can also be used to compare machine learning algorithms. When there are large datasets available naïve Bayesian and decision tree methods are preferred. In case of non-linear high dimensional data, kernel trick of SVM is preferred. Artificial

neural network takes more time in training in case of high dimensional data because number of variables are equal to the number of neurons in the input layer [21]. k-NN algorithm can be used to impute missing values in dataset [17]. The Bayesian, Random Forest and k-NN algorithms are applied to classify sensor nodes [19]. Performance of machine learning techniques can be evaluated using different evaluation parameters, for supervised learning algorithms, accuracy, precision, recall and F1 measure [22], receiver operating curve (ROC) [27] can be used. Other performance parameters are Jitter (delay), Goodput, Throughput [18], Detection Rate [23], Packet dropping ratio, False Probability [17], Packet loss, Packet delivery ratio [29] etc. Advantage of machine learning algorithms is that they can tolerate missing values, imprecise data etc. Disadvantage of machine learning algorithms is that they work on past data. [26]

6. EVALUATION

In the related work evaluation parameters like Goodput, Throughput and Jitter (During the data transmission, there is a possibility of small recurrent delay) [18], Packet dropping ratio or Packet dropping count & False Probability [17], Data drop rate, Energy consumption, Network lifetime, Detection Rate [23], Packet loss and Packet delivery ratio [29] have been used.

7. CONCLUSION

In this paper machine learning techniques to detect routing attacks in wireless sensor network are reviewed. Support vector machine is effective in most of the attacks and used very frequently because of its dynamic nature of working with large and non-linear datasets. Based on review it has been observed that neural networks require more time in training in case of large datasets. Though they are effective in high dimensional datasets. Neural network algorithms, can be computationally expensive so applying them with certain feature engineering techniques would increase their performance.

In future, independent component analysis or singular value decomposition can be adopted for dimensionality reduction. Reinforcement learning can also be used in the future for detection of attacks in wireless sensor networks. Since support vector machine is used in many of the applications, this algorithm can be implemented with real time data for routing attack attacks detection. To improve the performance of already existing techniques ensemble learning methods can also be adopted.

8. REFERENCES

[1] Kumar, D. P., Amgoth, T., & Annavarapu, C. S. R. (2019). Machine learning algorithms for wireless sensor networks: A survey. *Information Fusion*, *49*, 1-25.

[2] *Improving the Effectiveness of Diabetic Retinopathy Models*, https://ai.googleblog.com/2018/12/improving-effectiveness-of-diabetic.html

[3] Das, K., Majumdar, S., Moulik, S., & Fujita, M. (2020, September). Real-Time Threshold-based Landslide Prediction System for Hilly Region using Wireless Sensor Networks. In *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)* (pp. 1-2). IEEE.

[4] Riaz, M. N., Buriro, A., & Mahboob, A. (2018). Classification of attacks on wireless sensor networks: A survey. *International Journal of Wireless and Microwave Technologies*, *8*(6), 15-39.

[5] Virmani, D., Soni, A., Chandel, S., & Hemrajani, M. (2014). Routing attacks in wireless sensor networks: A survey. *arXiv preprint arXiv:1407.3987*.

[6] Humaira, F., Islam, M. S., Nur, F. N., & Hussain, K. A. (2020, July). A Comprehensive Study on Machine Learning Algorithms for Wireless Sensor Network Security. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

[7] Quincozes, S. E., & Kazienko, J. F. (2020, June). Machine Learning Methods Assessment for Denial of Service Detection in Wireless Sensor Networks. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)* (pp. 1-6). IEEE.

[8] Xiao, Z., Liu, C., & Chen, C. (2009, December). An anomaly detection scheme based on machine learning for WSN. In *2009 First International Conference on Information Science and Engineering* (pp. 3959-3962). IEEE.

[9] Kaplantzis, S., Shilton, A., Mani, N., & Sekercioglu, Y. A. (2007, December). Detecting selective forwarding attacks in wireless sensor networks using support vector machines. In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information* (pp. 335-340). IEEE.

[10] Mounica, M., Vijayasaraswathi, R., & Vasavi, R. (2021). Detecting Sybil Attack In Wireless Sensor Networks Using Machine Learning Algorithms. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1042, No. 1, p. 012029). IOP Publishing.

[11] Khan, R. A., & Pathan, A. S. K. (2018). The state-of-the-art wireless body area sensor networks: A survey. *International Journal of Distributed Sensor Networks*, *14*(4), 1550147718768994.

[12] Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H. P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, *16*(4), 1996-2018.

[13] Gunduz, S., Arslan, B., & Demirci, M. (2015, December). A review of machine learning solutions to denial-of-services attacks in wireless sensor networks. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 150-155). IEEE.

[14] Al-issa, A. I., Al-Akhras, M., ALsahli, M. S., & Alawairdhi, M. (2019, April). Using machine learning to detect DoS attacks in wireless sensor networks. In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) (pp. 107-112). IEEE.

[15] Mamdouh, M., Elrukhsi, M. A., & Khattab, A. (2018, August). Securing the internet of things and wireless sensor networks via machine learning: A survey. In *2018 International Conference on Computer and Applications (ICCA)* (pp. 215-218). IEEE.

[16] Patil, B., & Agarkhed, J. (2020, October). An Exploratory Machine Learning Technique for Investigating Intrusion in Wireless Sensor Networks. In *2020 IEEE Bangalore Humanitarian Technology*

Conference (B-HTC) (pp. 1-6). IEEE.

[17] Narayanan, K. L., Krishnan, R. S., Julie, E. G., Robinson, Y. H., & Shanmuganathan, V. (2021). Machine Learning Based Detection and a Novel EC-BRTT Algorithm Based Prevention of DoS Attacks in Wireless Sensor Networks. *Wireless Personal Communications*, 1-25.

[18] Kumar, N. S., Suryaprabha, E., & Hariprasath, K. (2021). Machine learning based hybrid model for energy efficient secured transmission in wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*, 1-16.

[19] Kim, T., Vecchietti, L. F., Choi, K., Lee, S., & Har, D. (2020). Machine Learning for Advanced Wireless Sensor Networks: A Review. *IEEE Sensors Journal*.

[20] Ahmad, B., Jian, W., Ali, Z. A., Tanvir, S., & Khan, M. S. A. (2019). Hybrid anomaly detection by using clustering for wireless sensor network. *Wireless Personal Communications*, *106*(4), 1841-1853.

[21] Yu, D., Kang, J., & Dong, J. (2021). Service Attack Improvement in Wireless Sensor Network Based on Machine Learning. *Microprocessors and Microsystems*, *80*, 103637.

[22] Aledhari, M., Razzak, R., & Parizi, R. M. (2021). Machine learning for network application security: Empirical evaluation and optimization. *Computers & Electrical Engineering*, *91*, 107052.

[23] Otoum, S., Kantarci, B., & Mouftah, H. T. (2019). On the feasibility of deep learning in sensor network intrusion detection. *IEEE Networking Letters*, *1*(2), 68-71.

[24] Poornima, I. G. A., & Paramasivan, B. (2020). Anomaly detection in wireless sensor network using machine learning algorithm. *Computer Communications*, *151*, 331-337.

[25] Jiawei Han, Micheline Kamber, and Jian Pei, Data Mining: Concepts and Techniques, 3rd Ed., Han, Kamber& Pei, University of Illinois at Urbana-Champaign &Simon Fraser University, 2011

[26] Amutha, J., Sharma, S., & Sharma, S. K. (2021). Strategies based on various aspects of clustering in wireless sensor networks using classical, optimization and machine learning techniques: Review, taxonomy, research findings, challenges and future directions. *Computer Science Review*, *40*, 100376.

[27] Al-Akhras, M., Al-Issa, A. I., Alsahli, M. S., & Alawairdhi, M. (2020, November). POSTER: Feature Selection to Optimize DoS Detection in Wireless Sensor Networks. In *2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH)* (pp. 263-265). IEEE.

[28] Wazid, M., & Das, A. K. (2016). An efficient hybrid anomaly detection scheme using K-means clustering for wireless sensor networks. *Wireless Personal Communications*, *90*(4), 1971-2000.

[29] Raghav, R. S., Thirugnansambandam, K., & Anguraj, D. K. (2020). Beeware routing scheme for detecting network layer attacks in wireless sensor networks. *Wireless Personal Communications*, *112*(4), 2439-2459.

[30] Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. *Journal of Sensors*, 2016.

[31] *Intelligence and Security Informatics Data Sets*, https://www.azsecure-data.org/other-data.html

[32] *Labelled Wireless Sensor Network Data Repository (LWSNDR)*, https://www.uncg.edu/cmp/downloads/lwsndr.html