Research on Intrusion Detection Method of Industrial Internet Based on Machine Learning

Yanfa Xu*
Shandong Vocational College of Science and Technology, 261053, China

*Corresponding author: yanfa_xu@sdvcst.edu.cn

Abstract. The mobile Adhoc network (MANET) is being used more and more widely, and the related network security issues have also begun to receive widespread attention. Researching the MANET network's possible attack methods, the paper proposes an intrusion detection performance evaluation model based on machine learning technology and proposes a comprehensive evaluation index. It compares seven machine learning algorithms' performance in MANET network intrusion detection, sufficient for building security. The MANET network is of great significance. Use the GloMoSim simulation tool to simulate the MANET network's normal behavior and the three intrusions of black hole, flood, and packet loss, and analyze the performance of seven machine learning algorithms in various attack situations in various attack situations detail. Our analysis results show that the evaluation model can better reflect the performance of various machine learning algorithms. Multilayer perceptrons, logistic regression, and support vector machines have higher detection rates and lower false alarm rates.

Keywords: MANET, intrusion detection, machine learning, industrial Internet.

1. Introduction
At present, the importance of industrial control networks has become increasingly prominent. With the computer and network technology development, industrial control has widely adopted open industrial communication protocols, network facilities, and available software and hardware, and even exchanged data with the Internet and enterprise management information systems. Attacks on industrial control networks have also increased rapidly. Cyber threats in the field of industrial control have greatly endangered the regular operation of industrial control, making the vulnerability of industrial control systems gradually appear, and attacks on industrial control systems have multiplied. MANET is composed of some mobile devices with communication capabilities, has dynamic network topology and self-organization capabilities, and is widely used in many fields. Compared to wired networks, MANET is more vulnerable to intruders. MANET uses wireless channels for communication. Within the range of wireless signals, attackers can easily eavesdrop and initiate attacks. At present, many methods have been proposed to detect abnormal activities in the network, but most of the research focuses on the innovation and improvement of a single algorithm (based on neural networks, clustering, support vector machines, etc.), and there is a lack of matching on a unified data set. A comparative study of different detection algorithms.
Aiming at three types of intrusion behaviors in MANET network, this paper proposes an intrusion detection performance evaluation model based on machine learning technology, establishes a characteristic model that characterizes network traffic, and proposes a comprehensive evaluation index account the detection rate and false alarm rate. Training time and detection time and compared the performance of seven well-known machine learning algorithms on this basis. Use simulation tools to simulate, sample, and statistically process network traffic data and generate a series of data sets to simulate the actual situation [1] virtually. Abnormal network nodes launch various attacks while the network is running. These attacks will be recorded in the form of statistical data and used in machine learning algorithms.

2. MANET and its security issues

2.1. MANET
Mobile Adhoc network was born in the late 1960s and early 1970s. In 1973, the US Department of Defense Advanced Research Projects Agency DARPA funded a research project, Packet Radio Network (PRNET), a packet-switched computer communication network that can transmit data through wireless channels [2]. These works have created pioneering theories and practices for MANET.

MANET is different from traditional wired networks. In MANET, each node acts as a router to transmit data to other nodes. Through mutual communication and cooperation, these mobile nodes are jointly responsible for the network's regular operation.

2.2. MANET network security
MANET is an open network. Attackers have more opportunities to invade the network and launch various attacks, such as floods, black holes, and denial of service [3]. Figure 1 shows a MANET network with attack nodes.

![Figure 1. Schematic diagram of the MANET network with attack nodes](image)

We evaluate different types of machine learning algorithms by simulating three types of attacks: (1) Flood attack: The attacker broadcasts a large number of forged data packets, exhausting network bandwidth resources and making the network unable to work typically. The abnormal node simulated in this experiment sends a large number of routing request data packets to the entire network to achieve the purpose of consuming network resources. (2) Packet loss attack: In this attack, the abnormal node discards the data packet sent to it. The abnormal node simulated in this experiment discards routing...
reply packets and data packets. (3) Blackhole attack: The abnormal node declares to other nodes that it has a minimum routing path, tricks other nodes to establish a routing connection with it, and then performs abnormal behavior. This experiment simulates that abnormal nodes absorb data packets sent by other nodes and discard them.

3. Related work

Intrusion detection technology is an active technology to ensure information security, which can effectively make up for passive defense shortcomings in traditional security protection technologies such as firewalls. Therefore, for complex industrial control system network attacks, intrusion detection technology is an effective means of detecting and preventing intrusions. Intrusion detection technology includes feature detection and anomaly detection [4]. Feature detection has a high detection accuracy rate, and the problem is that it cannot detect unknown attacks; anomaly detection has strong versatility, can detect unknown attacks, and has broad application prospects. Its primary defect is the false detection rate. Higher.

There are currently many ways to implement intrusion detection systems, such as intrusion detection systems based on statistics, patterns, rules, and status. This article focuses on using supervised machine learning algorithms. Some scholars have proposed an intrusion detection model in the MANET network, which is very similar to the multi-agent intrusion detection system under the traditional wired network. These works guide the subsequent establishment of an efficient intrusion detection system in MANET. Some scholars have proposed an intrusion detection system based on data mining technology. They use the attribute correlation between nodes to train a neural network, and the model obtained represents normal behavior [5].

Some scholars use the support vector machine algorithm to analyze and extract routing protocol interactions and design a distributed intrusion detection system. Some scholars analysed the denial-of-service attack against the MAC layer, proposed a solution based on the protection flow, and pointed out the proper position to join the protection flow. Recently, some scholars have evaluated the effects of five supervised learning algorithms in MANET network intrusion detection but have not analysed in detail the performance of different algorithms under different attack nodes and attack ratios. This article comprehensively considers different attack situations and proposes a performance evaluation model, which comprehensively considers the detection rate, false alarm rate, training time, and algorithm detection time. It proposes a comprehensive evaluation index to compare seven well-known machine learning algorithms. Conduct a comprehensive quantitative assessment.

4. Intrusion detection method based on machine learning

4.1. Machine learning algorithm

Machine learning is the core of artificial intelligence. At present, there have been a large number of researches on intrusion detection technology based on machine learning, including intrusion detection methods based on cluster analysis, data mining, behavior statistics, neural network, and other technologies. However, these technologies have slow processing speed, difficult to build models, and false alarms. The high rate and the difficulty of obtaining pure training data cannot meet the high real-time and availability of industrial control systems. Research on intrusion detection technology based on machine learning is of great significance for establishing intelligent and efficient intrusion detection models and improving the detection accuracy of abnormal behavior in industrial control networks. Aiming at the characteristics of industrial control system network data, this paper proposes an intrusion detection method based on semi-supervised machine learning, improving the detection accuracy of network attack traffic.

4.2. Feature model and evaluation criteria

This paper proposes a new evaluation model to analyze the performance of different machine learning algorithms. The evaluation process is shown in Figure 2.
This experiment set up three network types to evaluate the detection effect of various algorithms under different attack situations. The number of attack nodes is 30, 10, and 5, respectively. In each experiment, network traffic data is collected, and the characteristic data is counted every 15s, which is used to represent the network activity status in the sampling interval. After obtaining the original training set, by adjusting the proportion of attack data, a new training set with 75%, 25%, and 5% of attacks is obtained to evaluate the algorithm's performance more comprehensively [8]. All experimental data are collected from the network layer. In terms of features, we have expanded the features used by Mitrokotsa and introduced five new features, including packet loss rate (PDR), packet waiting to send rate (PWR), hop count (HC), route number (Route Number), and lousy link Number (Broken Links), defined as follows:

\[
PDR = \frac{\text{Packetsdropped}}{\text{packetsoriginated}}
\]  

\[
PWR = \frac{\text{Packetswaited}}{\text{packetsoriginated}}
\]  

\[
HC = \sum_{\text{nodes}} \sum_{\text{routes}} (\text{hopsgothrough})
\]  

\[
RN = \sum_{\text{nodes}} (\text{routeaccomplishes})
\]  

\[
BLN = \sum_{\text{nodes}} (\text{linkfailure})
\]  

By adding these new features, we hope to capture more statistical information in the MANET network. All the features used in the experiment are defined as follows: Hop Counts: The sum of the number of hops in all routes completed by all nodes. PDR: The discard ratio of packets. PWR: The ratio
of the packet waiting to be sent in the queue. Broken Links: The total number of terrible network links. Retries: The number of retries in the network due to harmful links.

The experiment uses a standard confusion matrix to evaluate the detection effects of different algorithms. Detection rate (TPR) and false alarm rate (FPR) are commonly used evaluation criteria. Detection rate means the ratio of correctly detected attack samples to the total attack samples, and false alarm rate means the standard samples that are incorrectly judged as attacks account for the total standard samples; the ratio is defined as follows:

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{TN + FP}$$  (6)

This process is repeated ten times, and finally, the ten times are averaged to obtain the final analysis result. Based on the statistically obtained TPR, FPR, model detection, and modelling time, we propose a comprehensive evaluation index Total Score, which is defined as follows:

$$Total\ Score = \frac{TPR \times (1 - FPR)}{TPR + 1 - FPR} \times \frac{\text{detectingtime}}{100} - \frac{\text{trainingtime}}{1000000}$$  (7)

This indicator gives a comprehensive evaluation of each machine learning algorithm. The higher the TPR, the lower the FPR value, the lower the detection time and training time, the higher the total index value.

5. Experiment

5.1. Simulation environment

This experiment uses GloMoSim to simulate a MANET network. We simulate a 2000×2000m2 network with 50 nodes randomly distributed. The wireless transmission range of each node is 250m, and the channel capacity is 2Mb/s. Each node sends a fixed bit rate (application-layer data packets; the packet size ranges from 128 bytes to 1024 bytes. The nodes move according to the Random Way Point model, and the waiting time is set to 0, 30, 50, 100, and 400s. Each experiment simulates a 1000s network running time.

By modifying the node's maximum moving speed and the pause time, we created a series of experimental data sets with a statistical sampling interval of 15 s. A type of attack is simulated, and the statistical data after each sampling is marked as four types: NORMAL, FLOODING, BLACK-HOLE, and PACKETDROP. Finally, we mix all the labelled data to get the training set. The machine learning tool uses the open-source project WEKA.

5.2. Experimental results

We conducted 180 experiments to simulate the network operation under different parameters and generated the original data set. The number of attack nodes was set to 30, 10, and 5, respectively. The traffic data contained three attack types and one standard type. In the simulation time of 18000s, 50 nodes were sampled, and 11664 records were obtained. The original data set generated by the specific experiment is shown in Table 1. The intrusion data generated under each number of attack nodes, adjust the proportion of attack data to 75%, 25%, and 5%, and obtain nine different training sets for detection and analysis. We adjust the corresponding learning parameters for each machine learning algorithm, conduct a series of experiments, and select the parameter combination with the lowest detection error. Finally, the arithmetic average of the TPR and FPR obtained from the experiment on the nine training sets is performed to obtain the experimental result graph. The average TPR of each machine learning algorithm under different numbers of attack nodes and attack ratios. The average TPR of packet loss attacks under different attack nodes and attack ratios. The average TPR of flood attacks under different
numbers of attack nodes and attack proportions. The average TPR of black hole attacks under different attack nodes and attack proportions. It can be seen that multi-layer perceptron, support vector machine, and logistic regression have higher TPR. With the reduction of attack nodes and the proportion of attacks in the network, the TPR of these three machine learning algorithms is generally higher than other algorithms. It can be seen that when the number of attack nodes is large, and the attack ratio is large, the FPR values of all algorithms are lower.

However, as the number of attack nodes or the proportion of attacks decreases, multilayer perceptrons, support vector machines, and logistic regression still have lower FPR. Multilayer perceptron, support vector machine, and logistic regression have the highest average TPR and the lowest average FPR. The average training time of all algorithms. Among the above three algorithms, the multilayer perceptron spends 8.744 s to train the classification model, while the support vector machine and logic.

| Number of attack nodes | Packet loss | Flood | Blackhole | normal |
|-----------------------|-------------|-------|-----------|--------|
| 30                    | 980         | 990   | 990       | /      |
| 10                    | 989         | 947   | 990       | /      |
| 5                     | 924         | 990   | 990       | /      |

6. Conclusion
This paper studies MANET networks' possible attack modes, propose an intrusion detection performance evaluation model based on machine learning technology, establishes a characteristic model that characterizes network traffic, and proposes a comprehensive evaluation index. The effect of MANET network intrusion detection is analyzed. The experiment simulated three kinds of intrusion behaviors: packet loss, flood, and the black hole. The simulation tool sampled the original flow data and obtained the training set, which was used to train the classification model of the machine learning algorithm. Experimental results show that the intrusion detection performance evaluation model can better characterize the MANET network's intrusion behavior, provide better classification effects for machine learning algorithms, and give the detection effects of different algorithms under different network attack situations. The proposed comprehensive evaluation index can better comprehensively reflect the performance indicators of the algorithm. Future work mainly focuses on further studying these three algorithms, such as using Boosting and other technologies to build robust classifiers and modeling and simulation, to establish an efficient MANET intrusion detection system.

References
[1] Mishra, P., Varadharajan, V., Tupakula, U., & Pilli, E. S., (2018). A detailed investigation and analysis of using machine learning techniques for intrusion detection. IEEE Communications Surveys & Tutorials, 21 (1), 686 - 728.
[2] Sultana, N., Chilamkurti, N., Peng, W., & Alhadad, R., (2019). Survey on SDN based network intrusion detection system using machine learning approaches. Peer-to-Peer Networking and Applications, 12 (2), 493 - 501.
[3] Verma, A., & Ranga, V., (2020). Machine learning based intrusion detection systems for IoT applications. Wireless Personal Communications, 111 (4), 2287 - 2310.
[4] Biswas, S. K., (2018). Intrusion detection using machine learning: A comparison study. International Journal of Pure and Applied Mathematics, 118 (19), 101 - 114.
[5] Park, S. T., Li, G., & Hong, J. C., (2020). A study on smart factory-based ambient intelligence context-aware intrusion detection system using machine learning. Journal of Ambient Intelligence and Humanized Computing, 11 (4), 1405 - 1412.