Intrusion Detection in Ad Hoc Network Using Machine Learning Technique

Mahendra Prasad[1(✉)], Sachin Tripathi[1], and Keshav Dahal[2]

[1] Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, India
je.mahendra@gmail.com, var_1285@yahoo.com

[2] School of Computing, Engineering and Physical Sciences, University of West of Scotland, Paisley, UK
Keshav.Dahal@uws.ac.uk

Abstract. Ad hoc network is a temporary self-organizing and infrastructure less network. So, it is mostly applied in the military field and disaster relief. Due to wireless communication and self-organizing property ad hoc network is more vulnerable to several intrusions or attacks than the traditional system. Blackhole attack is an important routing disruption attack that malicious node advertises itself as part of a path to the destination. In this paper, we have simulated blackhole attack in ad hoc network environment and collected data of essential features for attack behaviors classification. Then, many machine learning techniques have applied for classification of benign and malicious packet information. It suggests a new approach for select features, essential information collection, and intrusion detection in ad hoc network using machine learning techniques. We have shown comparative results of different machine learning techniques. Our results indicate that this approach can use with different classifiers and can extend it with other intrusions.

Keywords: Intrusion detection · Blackhole attack · Ad hoc network · Machine learning

1 Introduction

In present day, an ad-hoc network has been employed in many applications such military field, disaster relief, and other emergency services [1]. It is a peer-to-peer network that successfully transfers packets through multi-hop without any infrastructure. Due to dynamic nature, ad hoc network requires a unique security scheme to protects the network and detects intrusions or attacks. A popular routing dispersion attack method is known as blackhole attack that contains malicious nodes. These advertise them self as a part of the destination node and try to engage many possible connections [2]. A conventional approach of blackhole provides some solutions that are (1) find more than one routes and send packets, (2) unicast ping packets to the destination using these routes [3]. Intrusion Detection System (IDS) is a popular detection method that detects attacks. It enlarged the detection capacity and decreased the false alerts. An IDS can detect any violation

R. Patgiri et al. (Eds.): BigDML 2019, CCIS 1317, pp. 60–71, 2020.
https://doi.org/10.1007/978-3-030-62625-9_6

of security policies such as confidentiality, authenticity, integrity and availability [4]. A detection method can classify into benign and attacks by employing Machine Learning (ML) technique. There are many ML algorithms exist with pros and cons depend on nature of dataset. The ML works on learning method whether supervised, unsupervised or semi-supervised [5]. We have enumerated our major work below.

1. A blackhole attack is simulated in Network Simulator (NS-3) with many normal and malicious nodes. Where a malicious node can accommodate its all malicious activities.
2. We have analyzed the main features of the node and collected essential information of packets or message sending by nodes and organized them.
3. Finally, ML techniques have applied in the supervised mode of training on the collected data to detect malicious information or packets and measured their performance.

The rest of paper is arranged in the following sequence. Section 2 reviews recent literatures and Sect. 3 elaborates key topics as preliminaries, while Sect. 4 discuss proposed methodology. In Sect. 5, it descries simulation of the proposed method. Section 6 provides performance measures and comparative results. Finally, we conclude proposed method and future direction in Sect. 7.

2 Related Work

An IDS is a quick monitoring system and produce an alert when finding any intrusion. In this section, mainly introduce detection methods which related to IDS in ad hoc network. Kalkha et al. [6] proposed an approach based on a new routing algorithm to identify and avoid malicious node in the wireless sensor network. They applied the Hidden Markov Model to identify the most likely malicious path and detection module analyses the shortest path from source to destination. Omar et al. [7] proposed a threshold based multi-hop acknowledgment method that considered as blackhole node when the reputation of node increase or decrease to the threshold. Chatterjee et al. [8] suggested triangular encryption due to its low computation overhead and simulated it in network simulator NS-2. Panos et al. [9] proposed a dynamic threshold cumulative sum based mechanism that detects abrupt changes in normal behavior.

Mitrokotsa et al. [10] described a model selection and classification method for intrusion detection in ad hoc network. They had worked on selected features which are RREQ Sent, RREQ Received, RREP Sent, RREP Received, RERR Sent, RERR Received, Data Sent, Data Received, Number of Neighbors, PCR (Percentage of change in route entries) and PCH (Percentage of change in number of Hop). They also analyzed the cost and effect of the model. Subsequently, examined tuning of classifiers when unknown attacks appear in the system. They shown approx 90% detection rate vs FN to FP cost and approx 05% false alarm rate vs FN to FP cost of blackhole attack. Sen et al. [4] introduced an IDS using an evolutionary technique in ad hoc network. They have explored evolutionary computation techniques specifically genetic programming and grammatical evaluation. Then, employed multi-objective evolutionary technique to discover optimal trade-off.

Feng et al. [11] proposed an IDS method for anomalies detection in ad hoc network based on the learning method. They have applied deep learning to detect Denial-of-Service (DoS) and privacy attacks by grab packet information in ad hoc network. Subba et al. [12] proposed hybrid IDS in ad hoc network for unsupervised data. Their method elect cluster leader that provides intrusion detection service. Hybrid IDS comprises a lightweight and heavyweight module that detects intrusions and incomplete information anomalies. These works are simulated in network simulator and applied machine learning techniques to detect intrusions. We have proposed ML-based detection method and demonstrated a promising effect against blackhole attack in ad hoc network.

3 Preliminaries

3.1 Ad Hoc Network

Ad hoc is an infrastructure less and temporary self-organizing network. It establishes for special services such as battlefield, rescue services, etc. where no preexisting infrastructure or infrastructure failed [10, 11]. The application of this network is dynamic nature and quickly deployed. It is composed of nodes at different places and transfers messages to nodes in radio range. Neighbor nodes in network help for transferring message from source to destination [4, 12] using a routing protocol. Ad hoc On-demand Distance Vector (AODV) comes under distance vector routing protocols and applies in ad hoc network. AODV uses Route-Request (RREQ) packets when a node requires to build a route towards the destination. An immediate node sends RREQ to neighbors in range and establish a route and answers the source node by Route-Reply (RREP) packet. Due to the mobility of node every new diffusion establish a route [13]. A dynamic nature in the wireless environment of network intruders can easily adapt.

3.2 Blackhole Attack

Blackhole attack contains malicious nodes that can engage data packets by a false route reply packet. Malicious nodes falsely claim that have shortest route to the destination. When they receive data packets simply drop them. A malicious node tries to engage as a much possible active connection to the network resources. When the source establishes a malicious route node sends a false route reply message and acknowledge that it has an active route to the destination node [2, 14]. A conventional method suggests mitigating blackhole attack in ad hoc network. Unlike other approaches detect malicious node after the carried out information while some approach identified malicious nodes before the routing process and isolate them [14]. Our aim to detect malicious information during the routing process using the ML technique in ad hoc network. Then, it sends an alert to the network administrator.

3.3 Machine Learning Techniques

Despite the improvement of security schemes, continuous changing attack methods that need robust detection technique. The most acceptable technique is detecting in the context of attack sample whether the sample is normal or malicious. When analyzes sample

is malicious then isolate it before harm network resources [5]. The whole detecting process is based on learning method that can learn by a group of sample then provide a decision. ML techniques have been categorized into three categories which are supervised, unsupervised, and semisupervised learning. These are adopted by ML techniques that are applied to detect blackhole attack in this work. We have simulated many ML techniques on blackhole attack samples such as Ada Boost, Bayes Net [15], Decision Table, Hoeffding Tree, J48, KStar, Multi-Layer Perceptron (MLP) [16], Naive Bayes [15], Random Forest, Random Tree, and Stochastic Gradient Descent (SGD) [17]. These ML techniques work in only labeled dataset or supervised mode of training.

MLP is more suitable for linearly separable binary class problem [16]. It consist with minimum three layer namely input layer, hidden layer, and output layer. Naive Bayes and Bayes Net classifier are effectively used for condition monitoring that can applied for multi class [15]. SGD addresses the problem of high computational cost by some modification in gradient decent algorithm. It is only differ by how much data compute gradient for objective function and much faster convergence [17].

3.4 Intrusion Detection System

An IDS is immediate detecting method by intruders carry out information against the system. The primary aim to detect intrusions in communication and generate an alert to network administrator [11]. This is a powerful system to detect malicious information in the learning mode of training. Traditionally, intrusion was detected by conventional approaches such as encryption and decryption, authentication, firewall, etc. It is categorized in three categories such misuse detection system, anomaly detection system, and hybrid detection system. A misuse detection system is executed by matching the sample which is stored in the database and provide the decision. Anomaly detection system checks any deviation of sample form baseline if get then mark as malicious. Hybrid detection system uses both detection method property and reduces the drawback of detection system [13]. It is much powerful detection system than others and gives an acceptable decision.

4 Proposed Method

This section elaborates the proposed method of blackhole attack detection. We assume that the ad hoc network comprises N bidirectional communication nodes in the network space that share packets or information over a shared wireless medium. This network space contains $N - M$ normal nodes and M malicious nodes. Malicious nodes tune their behaviors and perform malicious activities. This method starts with feed data and simulates blackhole attack with malicious nodes. Subsequently, it gathers basic information of nodes which are in ad hoc network in a specified format. Then, this process selects essential features and collect data that build a dataset. Finally, we have applied many ML techniques for classification of information and provided the valid decision. A sequence of work is described in Algorithm 1.

Algorithm 1 Blackhole attack detection

1: input initial coordinate of nodes in the form of X and Y.
2: simulate some nodes with malicious activities as blackhole attack that attracts packet and drops it and others as normal.
3: trace pcap file of each node at each stage of message transfer and receive.
4: export packet informations in required file.
5: select essential features.
6: data collection using selected features.
7: apply various ML techniques to classify normal and malicious information.
8: store outcome as a confusion matrix.
9: compute different statistical measures.
10: evaluate comparative results.

We have described details of simulation procedure such as essential feature selection, data collection process, statistical measures, and different ML techniques results in the next section. It is also shown simulation results and tabled comparative results of ML techniques.

5 Experiments

5.1 Simulation

We have simulated blackhole attack in network simulator NS-3 [18]. Despite of NS-2, it is more priorities the use of the standard tool for input and output of file format therefore external tool also can be used. It is not a purely new simulator but also simulates predecessor simulator concepts, program, and data. The NS-3 provides network simulation in C++ and python program. To execute this work, the simulator enters into the main loop that executes events in predefined order from the data structure. This process continues until the event stack empty or predefined time has reached. In this simulation, network contains 25 nodes in network space including five malicious nodes. Experimental parameters of the simulator environment are topology space 1000 $\times$ 1000 m2, random node movement, radio range 250 m, etc. Figure 1 shows nodes position and radio range at a stage of nodes communication.

In recent days, WEKA (Waikato Environment for Knowledge Analysis) is recognized as a landmark system of data mining and ML. It allows researchers easy access to state-of-the-art technology in ML [19] and it has explored learning algorithms in many languages on various platforms which can operate on different types of data formats. WEKA is not only providing a toolbox of learning algorithms but also provides a framework for researchers can deploy a new learning algorithm. The task of WEKA is collecting dataset and providing results on selected ML techniques would be in various statistical parameters. We have executed our collected dataset on 11 different ML algorithms under 10-fold cross-validation and analyzed comparative results.

5.2 Data Generation

We have traced the output (Packet Capture in short pcap) files which have enough information to compute the required parameters. Any publicly available tool can analyze

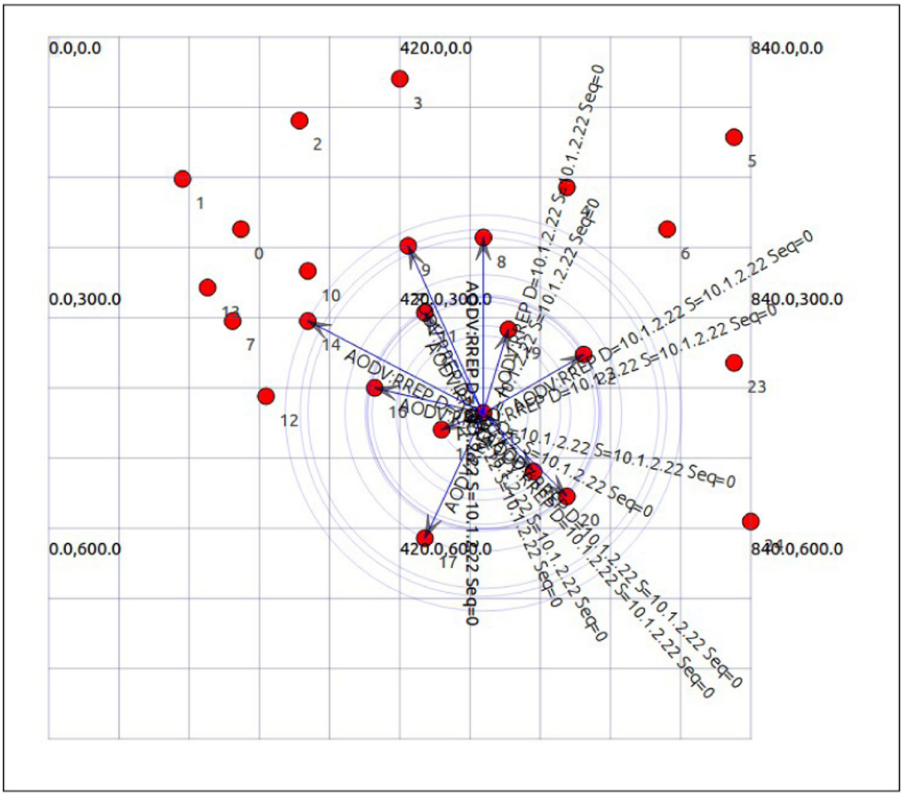

Fig. 1. Initial node position and radio range.

*.pcap traces file and gathers information for the further process. NS-3 supports standard output format for traced data which is in pcap format. We have used Wireshark and tcpdump [18] packet analysis tool to export data into standard or required file format.

5.3 Features Detail

It is a difficult task to select features that distinguish normal node and malicious node information. Features may depend on network structure and mode of data transmission. This work has analyzed the whole characteristics of nodes and gathered information into a proper format. A continuous data type provides simple information as numbers, and discrete data type may provides the string information.

From Table 1, duration indicates the transferring time of the packet from source to destination. The flag shows the status of packets and hopcount shows the intermediate nodes. Size of packets defines in packet size that includes header length in themselves. Messages are divided into many categories which are mainly Route Request, Route Reply, Route Acknowledgment, etc. Neighbor node is a number of node surrounding the node in communication range. When the sender and originator of message are same, then land indicates by Zero otherwise One. Unicast and broadcast are two different types

of message transferring modes. Message sequence number, originator sequence number, and stream index are generated sender or receiver for uniquely identified packets. The flow of message through the nodes can define the highest flow, lowest flow, average flow. Number failed connection and failure rate can compute using the Route Error message.

Table 1. Information of adopted features to aim of blackhole attack detection

S.No.	Feature name	Type
1	Duration	Continuous
2	Protocol	Discrete
3	Packet size	Continuous
4	Flag	Discrete
5	Header length	Continuous
6	Hop count	Continuous
7	Life time	Continuous
8	Message type	Discrete
9	Destination sequence number	Continuous
10	Message transfer mode	Discrete
11	Number of neighbors	Continuous
12	Land	Discrete
13	Message sequence number	Continuous
14	Stream index	Continuous
15	Highest flow	Continuous
16	Average flow	Continuous
17	Lowest flow	Continuous
18	Average hop count	Continuous
19	Number of failed connection	Continuous
20	Failed connection rate	Continuous
21	Label	Discrete

Finally, label the message or sample using the unique id of node generated or transferred message in the network.

5.4 Data Collection

We have collected distinct 711 (80 malicious and 631 benign) samples on 13 basic features including binary labels (named as Dataset-1). A quantity of benign sample is much higher than the malicious sample that can decrease the performance of the system. The size of the dataset is small that can lead the problems like bias or overfitting.

Although, we extend this work by increasing the features and simulation time that provide a new dataset. It contains 12,604 (2,654 malicious and 9,950 benign) samples that have 21 features including binary labels named as Dataset-2.

6 Result Analysis

6.1 Performance Measures

The outcome of the algorithm is collected in the form of confusion matrix that computes different statistical parameters by True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP is sample predicted as normal whenever the actual sample is also normal. TN is sample predicted as attack whenever the actual sample is also attack. FN is sample predicted as attack whenever the actual sample is normal. FP is sample predicted as normal whenever the actual sample is attack [11].

$$TPR(Recall) = \frac{TP}{TP + FN} \tag{1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

$$Avg.Performance = \sum_{i=1}^{c} \frac{S_i}{S} * Performance_{\mathrm{i}} \tag{6}$$

Where TPR is true positive rate and FPR is the false positive rate. Recall and TPR is the same whenever Precision provides correct prediction by test samples. F-measure provides a harmonic mean of Precision and Recall. Accuracy is the proportion of true prediction and total samples. $S = (S1 + S2 + \ldots + Sc)$ is the total sample and c is a number of class that computes the average performance of the system.

6.2 Performance Comparison

This section summarizes results of different ML techniques which executed on collected data of blackhole attack in ad hoc network. Table 2 shows results of computed statistical parameters such as TPR, FPR, Precision, F-measure, and accuracy. ML techniques such as Ada Boost, Bayes Net, Decision Table, Hoeffding Tree, J48, KStrar, MLP, Naive Bayes, Random Forest, Random Tree, and SGD are executed on Dataset-1.

Table 2. Performance of machine learning techniques for Dataset-1

Technique	TPR	FPR	Precision	F-measure
Ada Boost	0.931	0.249	0.933	0.932
Bayes Net	0.880	0.223	0.913	0.892
Decision Table	0.956	0.180	0.956	0.956
Hoeffding Tree	0.923	0.512	0.915	0.912
J48	0.951	0.181	0.952	0.951
KStar	0.887	0.538	0.878	0.882
MLP	0.932	0.205	0.938	0.935
Naive Bayes	0.885	0.298	0.904	0.892
Random Forest	0.927	0.271	0.929	0.928
Random Tree	0.907	0.481	0.898	0.901
SGD	0.934	0.205	0.938	0.936

Table 2 shows results on mentioned parameters that can easily recognize the best technique. When an attack is detected in the system then sends an alarm to the network administrator to isolates that node. While FN higher value indicates, normal packet information is falsely predicted as an attack. FP is the opposite of FN which indicates attack falsely detected as normal meanwhile the system allows the attack to enter and harm network resources. Decision Table classifier shows lower FPR and higher TPR, Precision, and F-measure. An opposite of this, KStar technique is shown higher FPR and lower Precision and F-measure. While Naive Bayes technique is shown lower TPR or Recall. Decision Table is producing a better detection system whenever KStar and Naive Bayes both are quantitatively given poor results.

Table 3 shows confusion matrix of SGD, Bayes Net, and MLP of Dataset-2 that use for statistical parameters computation. Table 4, 5 and 6 show the performance on different statistical parameters of SGD, Bayes Net, MLP respectively. The performance of the system is measured with the help of statistical parameters. TP and TN are correct prediction parameters which higher value improve system performance, and lower value

decreases the system performance. FP and FN are the incorrect predictions of parameters which lower value improve the performance of the system and higher value decrease the system performance. These parameters are used to compute performance measures such as TPR, FPR, Precision, F-measure, and Accuracy.

Table 3. Confusion matrix for Dataset-2

SGD			Bayes Net			*MLP*		
Class	Benign	Malicious	Class	Benign	Malicious	Class	Benign	Malicious
Benign	9841	2439	Benign	9145	529	Benign	9912	154
Malicious	109	215	Malicious	805	2125	Malicious	38	2500

Table 4. Statistical parameters of SGD for Dataset-2

Parameters	Benign	Malicious	Avg. performance
TP	9841	215	–
TN	215	9841	–
FP	2439	109	–
FN	109	2439	–
TPR	0.989	0.081	0.798
FPR	0.918	0.010	0.728
Precision	0.80	0.664	0.773
F-measure	0.885	0.144	0.730
Accuracy	0.798	0.798	0.798

Table 5. Statistical parameters of Bayes Net for Dataset-2

Parameters	Benign	Malicious	Avg. performance
TP	9145	2125	–
TN	2125	9145	–
FP	529	805	–
FN	805	529	–
TPR	0.919	0.80	0.894
FPR	0.199	0.080	0.174
Precision	0.945	0.725	0.90
F-measure	0.932	0.761	0.896
Accuracy	0.894	0.894	0.894

Table 7 shows performance of detection system where SGD, Bayes Net, MLP are performed on Dataset-2. In these detection systems, MLP provides a higher detection rate, precision, F-measure, accuracy, and lower false alarm rate. The training complexity

Table 6. Statistical parameters of MLP for Dataset-2

Parameters	Benign	Malicious	Avg. performance
TP	9912	2500	–
TN	2500	9912	–
FP	154	38	–
FN	38	154	–
TPR	0.996	0.942	0.985
FPR	0.058	0.004	0.047
Precision	0.984	0.985	0.985
F-measure	0.990	0.963	0.985
Accuracy	0.985	0.985	0.985

of detection systems of Dataset-2 as SGD (0.55 s), Bayes Net (0.15 s), and MLP (1.77 s). MLP took more training time to other ML techniques.

Table 7. Overall performance of detection system for Dataset-2

Classifier	TPR	FPR	Precision	F-measure	Accuracy
SGD	0.798	0.728	0.772	0.730	0.798
Bayes Net	0.894	0.174	0.90	0.896	0.894
MLP	0.985	0.047	0.985	0.985	0.987

7 Conclusion

In this paper, we have proposed a machine learning based intrusion detection system in the ad hoc network where intrusion as a blackhole attack. Blackhole attack is applied in the network and simulated with many malicious nodes. The main features of nodes are identified and collected information of traced pcap file using tcpdump. This information makes a set of distinct samples which is with known labels. Machine learning techniques are applied to this set of data which work in the supervised mode of training. Experiments show the simulated blackhole attack such activities, and various machine learning techniques provide their detection accuracy. Where MLP is shown the better result to other classifiers, it has shown 98.5% detection rate and 4.7% false alarm rate whenever it took more training time. These promising results encourage us to extend this work to identify more useful features and collect more information. Moreover, this work may simulate with other intrusions.

References

1. Khanna, N., Sachdeva, M.: A comprehensive taxonomy of schemes to detect and mitigate blackhole attack and its variants in MANETs. Comput. Sci. Rev. **32**, 24–44 (2019)
2. Khamayseh, Y.M., Aljawarneh, S.A., Asaad, A.E.: Ensuring survivability against black hole attacks in MANETs for preserving energy efficiency. Sustain. Comput. Inform. Syst. **18**, 90–100 (2018)
3. Al-Shurman, M., Yoo, S.-M., Park, S.: Black hole attack in mobile ad hoc networks. In: Proceedings of the 42nd Annual Southeast Regional Conference, pp. 96–97. ACM (2004)
4. Sen, S., Clark, J.A.: Evolutionary computation techniques for intrusion detection in mobile ad hoc networks. Comput. Netw. **55**(15), 3441–3457 (2011)
5. Ucci, D., Aniello, L., Baldoni, R.: Survey of machine learning techniques for malware analysis. Comput. Secur. **81**, 123–147 (2018)
6. Kalkha, H., Satori, H., Satori, K.: Preventing black hole attack in wireless sensor network using HMM. Procedia Comput. Sci. **148**, 552–561 (2019)
7. Hammamouche, A., Omar, M., Djebari, N., Tari, A.: Lightweight reputation-based approach against simple and cooperative black-hole attacks for MANET. J. Inf. Secur. Appl. **43**, 12–20 (2018)
8. Chatterjee, N., Mandal, J.K.: Detection of blackhole behaviour using triangular encryption in NS2. Procedia Technol. **10**, 524–529 (2013)
9. Panos, C., Ntantogian, C., Malliaros, S., Xenakis, C.: Analyzing, quantifying, and detecting the blackhole attack in infrastructure-less networks. Comput. Netw. **113**, 94–110 (2017)
10. Mitrokotsa, A., Dimitrakakis, C.: Intrusion detection in MANET using classification algorithms: the effects of cost and model selection. Ad Hoc Netw. **11**(1), 226–237 (2013)
11. Feng, F., Liu, X., Yong, B., Zhou, R., Zhou, Q.: Anomaly detection in ad-hoc networks based on deep learning model: a plug and play device. Ad Hoc Netw. **84**, 82–89 (2019)
12. Subba, B., Biswas, S., Karmakar, S.: Intrusion detection in mobile ad-hoc networks: Bayesian game formulation. Eng. Sci. Technol. Int. J. **19**(2), 782–799 (2016)
13. Liu, G., Yan, Z., Pedrycz, W.: Data collection for attack detection and security measurement in mobile ad hoc networks: a survey. J. Netw. Comput. Appl. **105**, 105–122 (2018)
14. Woungang, I., Dhurandher, S.K., Obaidat, M.S., Peddi, R.D.: A DSR-based routing protocol for mitigating blackhole attacks on mobile ad hoc networks. Secur. Commun. Netw. **9**(5), 420–428 (2016)
15. Muralidharan, V., Sugumaran, V.: A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. Appl. Soft Comput. **12**(8), 2023–2029 (2012)
16. Yi-Chung, H.: Pattern classification by multi-layer perceptron using fuzzy integral-based activation function. Appl. Soft Comput. **10**(3), 813–819 (2010)
17. Sharma, A.: Guided stochastic gradient descent algorithm for inconsistent datasets. Appl. Soft Comput. **73**, 1068–1080 (2018)
18. Riley, G.F., Henderson, T.R.: The ns-3 network simulator. In: Wehrle, K., Güneş, M., Gross, J. (eds.) Modeling and Tools for Network Simulation, pp. 15–34. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12331-3_2
19. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD Explor. Newslett. **11**(1), 10–18 (2009)