Journal of Universal Computer Science, vol. 22, no. 4 (2016), 521-536
submitted: 1/10/15, accepted: 30/3/16, appeared: 1/4/16 © J.UCS

Feature Selection for Black Hole Attacks

Muneer Bani Yassein
(Currently at School of Computing, Edinburgh Napier University
10 Colinton Road, Edinburgh EH10 5DT, UK
Department of Computer Science, Jordan University of Science and Technology, Jordan
m.baniyassein@napier.ac.uk)

Yaser Khamayseh
(Department of Computer Science, Jordan University of Science and Technology
Irbid, Jordan
yaser@just.edu.jo)

Mai AbuJazoh
(Department of Computer Science, Jordan University of Science and Technology
Irbid, Jordan
mmabujazoh12@cit.just.edu.jo)

Abstract: The security issue is essential and more challenging in Mobile Ad-Hoc Network (MANET) due to its characteristics such as, node mobility, self-organizing capability and dynamic topology. MANET is vulnerable to different types of attacks. One of possible attacks is black hole attack. Black hole attack occurs when a malicious node joins the network with the aim of intercepting data packets which are exchanged across the network and dropping them which affects the performance of the network and its connectivity. This paper proposes a new dataset (BDD dataset) for black hole intrusion detection systems which contributes to detect the black hole nodes in MANET. The proposed dataset contains a set of essential features to build an efficient learning model where these features are selected carefully using one of the feature selection techniques which is information gain technique J48 decision tree, Naïve Bayes (NB) and Sequential Minimal Optimization (SMO) classifiers are learned using training data of BDD dataset and the performance of these classifiers is evaluated using a learning machine tool Weka 3.7.11. The obtained performance results indicate that using the proposed dataset features succeeded in build an efficient learning model to train the previous classifiers to detect the black hole attack.

Keywords: Black hole attack; Intrusion Detection System (IDS); information gain; decision tree J48; Naïve Bayes (NB); Sequential Minimal Optimization(SMO); BDD dataset
Categories: C.2.0, C.2.3, C.4, I.6.0

1 Introduction

Maintaining the security in Mobile Ad-Hoc Network (MANET) is the most difficult issue where this type of networks is vulnerable to different types of attacks due to its unique characteristic. In MANETs, there is no centralized infrastructure to monitor the nodes behaviour and to manage the nodes membership where any node can join and leave the network arbitrarily which allow for the malicious nodes to join to the network easily and without prior detection. Moreover, the routing algorithms in MANET are based on the cooperative algorithms which require the mutual confidence between neighbouring nodes. Therefore, it causes to violate the security principles of networks. In addition to that, the medium in this type of networks is open for all the nodes which make the Eavesdropping easier than in wired networks [Djenouri,05][Singh, 14] [Aarti, 13].

This paper discusses one of the potential attacks that may expose the MANETs which is black hole attack. This attack occurs when a malicious node joins the network with the aim of intercepting the data packets which are exchanged across the network and dropping them without delivering them to the destination which lead to the Denial of Service (DoS) problem. Generally, in this kind of attacks there are two cases to obtain the data packet by the malicious node; the first one when the malicious node exploit the routing protocol such as AODV protocol to send a route reply control message (RREP) to the source node immediately upon receiving a route request control message (RREQ) in order to introduce itself as the intermediate node that has the best routing path to the destination node. The source node uses this false path to send its data packets. Second case when the malicious node has the ability to intercept the data packets without sending any RREP to the Source node. In both cases, once the black hole node receives this data packet, it drops the data immediately. Therefore, the source node becomes unable to send its data to the destination node which affects the performance of the network and its connectivity [Jhaveri, 12] [Kurosawa, 07][Sanaei, 13].

To protect the network from black hole attack, there are several protective measures like firewall which have been put in place in order to detect the unsafe activities. However, these protective measures do not guarantee the full protection for the network. Therefore, there is a need for a second line of defence that has the ability to detect new vulnerabilities which appears day by day. For this purpose a lot of systems are designed in the past but the Intrusion Detection Systems (IDSs) are the most important systems from them [Djenouri, 05][Huan, 05][John, 94].

In order to make MANET more secure against black hole attacks, in this paper we study the behaviour of the black hole nodes by monitoring a set of features for each node in the network. The audit data are recorded in a new dataset; namely BDD; for black hole intrusion detection systems which contributes to detect and avoid the black hole attacks. Feature selection process is applied on the BDD dataset in order to select the most important and relevant features that contribute to the detection and prevention of black hole attacks where feature selection is a technique for selecting the most relevant and important features by removing and discarding the redundant and irrelevant features from the data in order to build an efficient and effective learning model [Djenouri, 05][Huan, 05][John, 94].The proposed dataset contributes to improving the performance of learning models by speeding up the training and testing process. Moreover, using this dataset makes the detection rate of black hole attacks be faster and more accurate. Therefore, the network will be more secure from this kind of attacks.

2 Related works

Black hole attack is one of the most popular attacks that adversely affect on the performance of the network. Therefore, different techniques were proposed in order to prevent and detect the black hole attacks. Most of these techniques can be divided into two categories which are: secure existing protocol or using Intrusion Detection Systems (IDS)[Kurosawa, 07][BaniYassein, 14]. The most of techniques that use Intrusion Detection Systems (IDS) to detect the black hole attack are depend in their work on KDDCUP'99 dataset or NSL KDD dataset. In 1999 the KDD cup was organized and the researchers from the entire world were invited in order to design creative methods to construct IDS on training and testing datasets which popularly called as KDDCUP'99dataset [KDDCUP, 99].The KDDCUP'99dataset is extensively used as one of a few datasets which are publicly available for network-based anomaly detection systems. It based on the 1998 DARPA Intrusion Detection Evaluation Program which was prepared by MIT Lincoln Labs in order to evaluate the research on the intrusion detection. The Lincoln Labs work to collect huge collection of raw TCP dump data which are simulated over a period of 9 weeks on a local area Network where the LAN was as a true U.S. Air Force environment, but with add multiple of attacks to it. The KDDCUP'99dataset is built upon the training data which is about 4 gigabyte of compressed binary TCP dump data that was collected from the seven weeks of network traffic. The extensive amount of training data is one of the biggest challenge in the network based instruction detection. Therefore, these data are summarized to about 50000000 connection records before feeding it to the machine learning algorithm. On the other hand, the two week of the test data produced about two million connection records. In order to make the intrusion detection evaluation be more realistic and efficient, the training data contains 22 different attacks out of the 39 attacks which appeared in the test data. These attacks fall into one of the following categories: Denial of service (DOS), Probe, User to Root (U2R) or Remote to Local (R2L). For each connection in the data set there are 41 features and each connection is labelled as either attack or a normal connection. Some of these features are derived features which contribute to distinguish the attacks connections from the normal. Although the KDDCUP'99 dataset is widely used for evaluation the instruction detection systems, it suffers from some problems that could affect on the performance of the evaluated systems. According to the results of the statistically analysis on this dataset in [Tavallaee, 09], the first important problem in KDDCUP'99 dataset is the huge number of redundant records. Tavallaee et al. analyzed the KDD training set and the KDD test set and they found that there are about 78% of the records are repeated in the training set and there are about 75% of the records are repeated in the test set. The huge amount of redundant records in training set lead to make the learning algorithms biased to the more frequent records. Therefore, it will cause to prevent the algorithm to learn from the infrequent records, which are usually more harmful to the network. On the other hand, the frequent records in the test set effect on the evaluation results, where these results will be biased to the methods that have the best detection rates on the frequent records. Moreover, the huge amount of records in KDD train and test sets will make it difficult to run the experiments on the complete set. Therefore, random parts of the train set are used as test set which will make the comparison between the evaluation results of different research work very difficult. In order to solve these problems the authors proposed anew dataset which is NSL-KDD. The proposed dataset contains selected records of the KDDCUP'99 dataset without any redundant records, so it is a new version of KDDCUP'99 dataset but without suffers from any of the problems that mentioned previously. The proposed dataset has the following advantages over the original KDD dataset:

- The training set in the proposed dataset doesn't include any redundant records so the learning algorithms will not biased to the more frequent records.
- There are not any redundant records in the test set in the proposed dataset, so the evaluation results will not biased to the methods that have the best detection rates on the frequent records.
- In the train and test sets the number of records are reasonable, which make it is possible to run the experiments on the complete set without the need to randomly select a small parts from the train set to use it as test set. As a result, the results of the evaluation for the different work will be comparable.

Although the proposed dataset solved some of the inherent problems of the KDDCUP'99 dataset, but it still suffers from some of the other problems [McHugh, 00]. In spite of these problems the proposed dataset can be used as an effective dataset to help the researchers to compare between the different IDS where there is a lack of available data sets for network-based IDSs.

By studying the previous works we noticed that the most of existing IDSs depend on its work to detect the black hole attack on the KDDCUP'99 dataset or on the NSL KDD dataset. KDDCUP'99 dataset is an old dataset that is created in 1999. Whereas, the NSL KDD dataset is created in 2001 and contains selected records of the KDDCUP'99 dataset without any redundant records. Both datasets are not created especially to detect the black hole attack and suffered from many problems. Accordingly, in this paper we study the behaviour of the MANETs with presence the black hole attacks in order to create a new dataset (BDD) which is the only labelled dataset that available for black hole attack.

3 Proposed Dataset

Using an Intrusion Detection Systems (IDS) is one of the popular methods that are used to protect the network from various types of attacks. Intrusion detection is the process of monitoring and analyzing the activities which occurring in the network in order to detect the abnormal activities. The abnormal activities constitute violations on the network security and affect on the network performance. IDS can be classified as: host-based IDS and network-based IDS. The host-based IDS is run on each node in the network where it monitors and analyses the activities of the nodes internally such as, file systems, system logs and disk resources; whereas the audit data is obtained in the network-based IDS from the network traffic which flow through it, where this kind of IDS is run on a gateway of the network. Basically, there are two main types of IDS which are: anomaly-based detection and signature-based detection (misuse detection model). The normal behaviours of the nodes are kept in the anomaly detection systems where through this system, the captured data is compared with the normal behaviours in order to determine if there are normal or abnormal activities. On the other hand, the misuse detection systems keep signatures or patterns of known attacks in order to compare them with the captured data where any matched between the captured data and the saved patterns is considered as intrusion. The main drawback of this kind, that it cannot detect new kinds of attacks [Bhoria, 13][Jain, 12][Shrivas, 13].In this paper, host-based IDS is used where the activities of each node in the network are collected; whereas the anomaly-based detection is used as a detection technique.

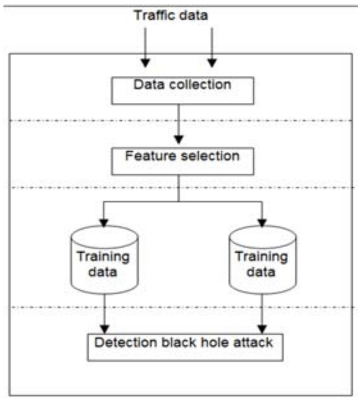

Figure 1: Proposed intrusion detection model.

As shown in Figure 1, the proposed system architecture of IDS comprises the following component:

3.1 Data Collection Module

In this research, the audit data is collected from each node in the MANET during the normal scenario and the black hole attack scenario in order to create a normal profile and attack profile. The normal profile is created using the data which are collected from the MANET with absence the black hole attack; whereas the attack profile is created using the data that are collected from the network with presence the black hole attack. To collect the audit data, GloMoSim 2.03 simulator is used for simulating normal and black hole attack scenarios in MANET. In our experiments, we try to collect the audit data from the maximum amount of possible cases. Table 1 shows the simulation parameters that are used to extract the audit data from different scenarios.

Parameter	Value
Simulator	GloMoSim 2.03
Simulation time	800,1000 and 1200 second
Simulation area	1000m × 1000m
Number of nodes	20, 25, 30, 35, 50, 75,100, and 125
Number of black nodes	0,1,2,6 and 10
Mobility model	Random waypoint
Minimum speed	0 , 0.5 meter/second
Maximum speed	20 , 2 meter/second
Pause time	0,10

Table 1. Simulation parameters for extract audit data.

Figure 2 shows the features that monitored and recorded in BDD dataset for each node in MANET.

One of the main behavioural characteristics of black hole node that it introduces itself as intermediate node that has the best path to the destination node, it sends RREP message with high destination sequence number and low count of hops to the destination node. Thus, before collected the previous features, the thresholds for destination sequence number and count of hop to the destination that send by RREP is calculated in order to use these thresholds to calculate "*Number of high destination sequence number*" and "*Number of low count of hops to destination*" features. These thresholds are calculated by collecting the data of RREP messages which are sent from each node in the network and recording the collected data in a new dataset (namely RPD). Each record in this dataset is labelled as normal node if the RREP is sent from normal node or black hole node if the RREP is sent from black hole node.

J48 decision tree classifier which will be discussed later in this research is applied to RPD dataset in order to classify the sender of RREP message to black hole node or normal node. According to this classifier, if the node send RREP with destination sequence number > 6292 and with number of hop count = 2, then this RREP is sent from black hole node, otherwise the RREP is sent from normal node.

Number of RREQ sent	Number of RREP sent	Number of RREP forward	Number of RREP sent with route
Node average speed	Number of RERR sent	Number of RERR re-sent	Number of routes selected
Number of data sent	Number of data packets originated	Number of data packets received	Number of packets dropped
Number of hop counts	Number of CTRL packets sent	Number of broken link retries	Number of broken links
Number of RREQ received	Number of RREP received	Number of data packets in node	Number of low count of hops to
Number of high destination	Number of intermediate	The node which send the maximum	Maximum number of reply from the
Number of bytes sent	Number of bytes received	Number of act as source	Number of act as destination
AVG replying fast			

Figure 2: The collected features for each node.

3.2 Feature Selection Module

Before feeding the large amount of audit data to the classifier, it should be summarized using feature selection process. Through the features selection process, the most important and relevant features, that contribute to detect and prevent black hole attack, are selected by removing and discarding the redundant and irrelevant features. Using feature selection process led to reduce the computation time and model complexity and contribute to build an efficient and effective learning model [Huan, 05][John, 94][Upendra, 13]. One of the successful approaches that are used for features selection is information gain measure which is the most important features selection measure for constructing the decision tree classifier. Information gain (IG) is

a measure of the changes in the entropy from the previous state to the state that takes some information, where entropy (modified information) is the expected information that needed in order to classify a tuple in the dataset [Jain, 12]. In order to calculate the IG value for each attribute in the dataset, the entropy measure should be calculated. Entropy is calculated as the following equation:
Suppose X takes n values, V_1, V_2… V_n and $P_{(X=V1)} = p_1$, $P_{(X=V2)} = p_2$......$P_{(X=Vn)} = p_n$ then:

$$H(x) = p_1 \log_2 p_1 - p_2 \log_2 p_2 - \ldots\ldots p_n \log_2 p_n \qquad (1)$$

$$= -\sum_{i=1} p_i \log_2 (p_i)$$

In the BDD dataset, there are two class labels which are “Normal” and “Black”. Therefore the entropy equation for our dataset is:

$$H(x) = p_1 \log_2 p_1 - p_2 \log_2 p_2$$

$$= -\sum_{i=1}^{2} p_i \log_2 (p_i)$$

Where p_1 indicates to a set of samples that belongs to a target class “Normal”, p_2 indicates to a set of samples that belongs to a target class “Black”. To calculate the encoding information that would be gained by the attribute A_i the following equation is used:

$$Gain(X, A_i) = H(X) - \textstyle\sum_{v\in Values(A_i)} P(A_i = v) H(X_v)(2)$$

In this research, we used a learning machine tool Weka 3.7.11 to apply a feature selection process on the BDD dataset. WEKA is open source data mining software that stands for Waikato Environment for Knowledge Analysis. It was created by researchers at the University of Waikato where the first implementation of the Weka in its modern form was in 1997 [Dash, 13]. Weka is used for data pre-processing and classification where through using this tool, the IG for all the features in our dataset have been calculated and arranged according to theirs IG value. High IG value for a specific feature indicates that this feature is more relevance to detect black hole attack than the other features that have less IG values. All of the ranked features were analysed and studied in this research based on the behavioural characteristics of black hole node in order to ensure the effectiveness of the selected features in the detection and prevention black hole attack. According to the analysis result, the following four features are selected as most relevant features:

(1) Number of RREQ sent feature.
(2) Number of RREP that forward feature.
(3) Number of high destination sequence number feature.
(4) Number of low count of hops to destination feature.

Moreover, another two features are selected as relevant features for the black hole attack in spite of they haven’t a very high IG value, where the results of the analysis of these two features show that they can play a significant role in detection and prevention the black hole attack, these two features are:

(5) Number of act as source feature.
(6) Number of act as destination feature.

According to the behavioural characteristics of the black hole node the previous 6 features are the most relevance feature to distinguish between the normal node and the black node. The black hole node is a strange node that joins to the network with the aim of dropping the receiving data packets instead of sending them to the desired destination node. Therefore, the black hole node doesn’t send any RREQ message to discover a route to any other node, where it doesn’t generate data packets to send them. In addition to that, black hole node doesn’t re-broadcast the received RREQ message to its neighbours nodes where it sends RREP message immediately when it receives any RREQ message in order to gain the data packets and drop it. These two abnormal characteristics can contribute in detection the black hole attack, so they are summarized in “Number of RREQ sent” feature in our dataset. “Number of RREP that forward” feature is another attribute that contribute to distinguish between the normal and black nodes where one of the characteristics of the black hole node that it doesn’t re-broadcast the received RREQ messages. Thus, it doesn’t unicast any received RREP message from the replying node to the source node where replying node unicast back the RREP message through the route that is used by the RREQ message of this RREP message. As we mentioned previously, black hole node introduces itself as intermediate node that has the best path to the destination node. For that, it sends RREP message to the source node with high destination sequence number and low hop counts. Monitoring the RREP destination sequence number and replying hop count contribute to detect if the replying node is a black hole node or a normal node so the features “Number of high destination sequence number” and “Number of low count of hops to destination” are selected as relevance features. Moreover, we selected “Number of act as source” and “Number of act as destination” features as relevance features in our dataset, in spite of their IG values are low. The reason of that is due to the significant role played by these two features in detection and prevention the black hole attack where according to the behavioural characteristics of the black hole node, it doesn’t act as source node or as destination node in the network.

After applied the features selection process, BDD dataset become contains only six features which are the most relevant features that contribute to detect and prevent black hole attack. In addition to that, BDD dataset provides labelled data for researchers whom working in the field of black hole detection where the BDD is the only labelled dataset that available for black hole attack.

3.3 Detection module

Through this module, the test data is compared with normal profile of the network by a predefined classifier in order to classify these data to normal or attack. If there is any deviation from the normal behaviour of the network, then the event is considered as attack. Otherwise it is considered as normal. The following subsections describe the detection process in detail.

3.4 Classification

Classification is the process that is used to learn a model (classifier) from the training set, which is a set of labelled data, in order to classify a test data into one of the class labels. Classification-based anomaly detection techniques are based on their work on two phases; training phase and testing phase. The training phase using an available labelled training set to learn a specific classifier where this classifier is used in the testing phase to classify a test instance to normal or abnormal [Ganapathy, 13]. In this research the following classifiers, which are available on the Weka collection of machine learning algorithms, are learned in order to classify the testing data of BDD dataset as normal or black:

- **J48 decision tree classifier**: one of the popular classifiers that are used in Weka data mining tool, it is a simple implementation of C4.5 decision tree algorithm. Through the training phase, J48 classifier creates a binary tree which is based on the attribute values of the training set. While through the training phase, J48 classifier apply the binary tree, that has been built, for each tuple in the testing set in order to classify these tuple [Patil, 13] [Chandolikar, 12].
- **A Naive Bays(NB) classifier**: is a simple probabilistic classifier that is based on applying the Bayes' theorem to classify the new instance with strong independent assumptions between the attributes. This classifier calculates a set of probabilities by counting the frequency of the values in the dataset. This classifier derived the probability for a specific feature, which appears as a member of the set of probabilities, by calculating the frequency of each feature value within a class of a training set [Upendra, 13] [Patil, 13][Chandolikar, 12].
- **Sequential Minimal Optimization (SMO)**: a new algorithm for training the Support Vector Machines (SVM). SMO is developed to solve the training SVM problem which is a very large quadratic programming (QP) problem. SMO solve this problem by break the QP problem into series of QP sub-

problems. These QP sub-problems are solved analytically where the using the time-consuming numerical QP optimization such as inner loop is not allowed. In SMO, the memory amount grows linearly with the training set size. Therefore, SMO can handle a very large training set [Platt, 98].

3.5 Evaluation Measuring Performance

To evaluate the classifier's performance, the confusion matrix should be constructed. This matrix contains information about actual and predicted classifications which done by the classifiers, where these data is used to evaluate the performance of the classifiers. Table 2 shows a sample of confusion matrix for two classes.

	Predictive Class Positive	Predictive Class Negative
Actual class positive	True Positive (TP)	False Negative (FN)
Actual class Negative	False Positive (FP)	True Negative (TN)

Table 2: Confusion matrix.

In this confusion matrix, TP value indicates for the number of positive samples that are correctly classified by the classifier, FP refers to the number of positive samples that is incorrectly classified, similarly FN refers to the number of negative samples that are classified as Positive samples instead of Negative sample. Whereas, TN is the number of Negative samples which are correctly classified by the classifier.

Based on the above Table, the performance of each one of the previous classifiers can be evaluated using some of very popular statistical measures [Jain, 12][Shrivas, 13] [Chandolikar, 12] which are:

- **Accuracy**: this measure is used to determine the percentage of samples that are classified correctly. From the above confusion matrix, accuracy is calculated as following :

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} (3)$$

- **Precision**: is a ratio of number of relevant samples that are retrieved to the total number of samples (relevant and not relevant) that are retrieved. Precision is calculated as following :

$$\text{Precision} = \frac{TP}{\text{TP+FP}} (4)$$

- **True positive rate** (**TPR**): it also called **sensitivity** and **Recall**; it is a ratio of the number of relevant samples that are retrieved to the total number of relevant samples.

Recall= $\frac{TP}{(TP+FN)}$(5)

- **False positive rate** (FPR): it also called **False alarmrate;** it is calculated as following:

False alarm rate = $\frac{FP}{TN+FP}$(6)

- **F-measure**: which is a harmonic mean of Recall and Precision, it is calculated as following :

F-measure = $\frac{2*\text{precision} * \text{Recall}}{\text{precision} + \text{Recall}}$(7)

4 Experimental Results

To assess the effectiveness of the previous classifiers in intrusion detection process, we performed series of experiments using Weka 3.7.11. 10-fold cross-validation approach is used to partition the BDD dataset into training set and testing set. This approach led to randomly divide the BDD dataset into 10 subsamples where 9 subsamples are used as a training data and one subsample is used as test data and this process is repeated 10 times [Chandolikar, 12].

We have performed classification using J48, NB and SMO classifiers on our proposed BDD dataset with the six features which are selecting on the feature selection process. For each one of the previous classifiers, the confusion matrix is generated for class attack which having two possible values; normal or black. Table 3 shows the confusion matrixes for the three classifiers.

	J48 classifier		NB classifier		SMO classifier	
Actual vs. Predicted	Normal	Black	Normal	Black	Normal	black
Normal	1188	1	1188	1	1189	0
Black	0	100	58	42	80	20

Table 3: Confusion matrix of J48, NB and SMO classifiers.

Based on the above Table information, Weka calculated the following statistical performance measures: Accuracy, Precision, True Positive Rate, False Positive Rate and F-Measure where these performance measures are used to evaluate the performance of each one of the three classifiers in detection the black hole attack. Table 4 shows the values of the statistical performance measures for the three classifiers.

Parameter	Classifier		
	J48 classifier	NB classifier	SMO classifier
Accuracy	99.9224 %	95.4228 %	93.7936 %
Precision	0.999	0.955	0.942
Recall	0.999	0.954	0.938
TP Rate	0.999	0.954	0.938
FP Rate	0.00	0.535	0.738
F-Measure	0.999	0.946	0.918

Table 4: Comparison of Results forJ48, NB and SMO classifiers.

According to the Table 4, the experimental results show that using the proposed dataset features succeeded in builds an efficient learning model to train J48, NB and SMO classifiers to detect the black hole attack whereJ48 classifier gave 99.9224 % accuracy, NB classifier gave 95.4228 % accuracy and SMO classifier gave 93.7936 % accuracy. From the experimental result we conclude that J48 classifier have much better performance in detection black hole attack than other two classifiers, where this classifier gave the highest accuracy and 0 false positive rate in detection the black hole attack. Moreover, reliance on the BDD features to build new black hole detection and prevention technique may lead to increase the detection accuracy and decrease the false alarm for this technique.

5 Conclusions

The main focus of this research is security issue in mobile ad hoc network. In MANET, security is essential and even more challenging due to the MANET characteristics such as: open media nature, node mobility, self-organizing capability and dynamic topology. Black hole problem is one of denial of service attacks where the black hole node joins the network with the aim of intercepting the data packets which are exchanged across the network and dropping them without delivering them to the destination. Therefore, this problem affects the performance of the network and its connectivity. In order to make MANET more secure against black hole attacks, we studied the behaviour of the black hole nodes by monitoring a set of features for each node in the network. The audit data are recorded in the new dataset which is BDD dataset. Feature selection process is applied on the BDD dataset in order to select the most important and relevant features that contribute to detect the black hole attack. Information gain which is one of the successful approaches that are used for features selection is used in this research to select the most relevant features. After applied the features selection process, the BDD dataset becomes contain six relevant features which are Number of RREQ sent feature, Number of RREP that forward feature, Number of high destination sequence number feature, feature of low count of hops to destination feature, Number of act as source feature and Number of act as destination feature.

The experimental results show that using the proposed dataset features succeeded in build an efficient learning model to train J48, NB and SMO classifiers to detect the black hole attack. Therefore, reliance on the BDD features to build new black hole detection and prevention technique may lead to increase the detection accuracy and decrease the false alarm for this technique. Moreover, the proposed BDD dataset provides labelled data for researchers whom working in the field of black hole attack detection where the BDD dataset is the only labelled dataset that available for black hole.

References

[Aarti, 13] Aarti, D.,Tyagi, S.: "Study of MANET: Characteristics, Challenges, Application and Security Attacks"; International Journal of Advanced Research in Computer Science and Software Engineering, 3,5(2013), 252-257.

[BaniYassein, 14] BaniYassein, M., Khamayseh, Y. and Nawafleh, B.: "Improved AODV Protocol to Detect and Avoid Black Hole Nodes in MANETs"; FUTURE COMPUTING 2014, The Sixth International Conference on Future Computational Technologies and Applications (2014), 7-12.

[Bhoria, 13] Bhoria,P.,Garg,K.: "Determining Feature Set of DOS Attacks"; In International Journal of Advanced Research in Computer Science and Software Engineering, 3, 5(2013): 875-878.

[Chandolikar, 12] Chandolikar, N.,Nandavadekar V.: "Comparative Analysis of Two Algorithms for Intrusion Attack Classification Using KDD Cup Dataset"; In International Journal of Computer Science and Engineering (IJCSE), 1, 1(2012), 81-88.

[Dash, 13] Dash R.: "Selection ofthe Best Classifier from Different Datasets Using WEKA"; In International Journal of Engineering Research & Technology (IJERT), 2, 3 (2013), 1-7.

[Djenouri, 05] Djenouri, D., Khelladi, L.,Badache, N.: "A Survey of Security Issues in Mobile ad hoc Networks"; IEEE communications surveys;7,4 (2005), 2-28.

[Ganapathy, 13] Ganapathy, S., Kulothungan, K., Muthurajkumar, S., Vijayalakshmi, M., Yogesh, P.,Kannan,A.: "Intelligent Feature Selection and Classification Techniques for Intrusion Detection in Networks: A Survey"; In Journal on Wireless Communications and

Networking; 1(2013),1-16.

[Huan, 05] Huan, L.,Yu, L.: "Toward Integrating Feature Selection Algorithms for Classification and Clustering"; In Knowledge and Data Engineering, IEEE Transactions, 17,4 (2005), 491-502.

[Jain, 12] Jain, K.,Upendra: "An Efficient Intrusion Detection Based on Decision Tree Classifier using Feature Reduction"; In International Journal of Scientific and Research Publication, 2, 1 (2012), 66-70.

[Jhaveri, 12] Jhaveri, Rutvij, H., Sankita, J.,Devesh,C.: "DoS attacks in mobile ad hoc networks: A survey"; Advanced Computing & Communication Technologies (ACCT) 2012, Second International Conference on IEEE 2012; 2, 1 (2013): 535-541.

[John, 94] John, H.,Kohavi, R., Pfleger, K.: "Irrelevant Features and the Subset Selection Problem"; In International Conference on Machine Learning (1994),94: 121-129.

[Kaur, 13] Kaur, E.,Singh, A.: "Performance Evaluation of MANET with Black Hole Attack Using Routing Protocols"; International Journal of Engineering Research and Applications (IJERA), 3, 4(2013), 1324-1328.

[KDDCUP, 99] KDDCUP: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html(1999, accessed August 2014).

[Kurosawa, 07] Kurosawa, S.,Nakayama, H., Kato, N.,Jamalipour, A.,Nemoto,Y.: "Detecting Blackhole Attack on AODV-based Mobile Ad Hoc Networks by Dynamic Learning Method"; International Journal of Network Security; 5, 3 (2007),338-346.

[McHugh, 00] McHugh J.: "Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory"; In ACM Transactions on Information and System Security (TISSEC), 3, 4 (2000): 262-294.

[Ochola, 14] Ochola, E.,Eloff M.: "A Review of Black Hole attack on AODV Routing in MANET";http://icsa.cs.up.ac.za/issa/2011/Proceedings/Research/Ochola_Eloff.pdf?origi=publication_detail. (2011, accessed December 2014).

[Patil, 13] Patil,T., Sherekar, S.: "Intelligent Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification"; In International Journal of Computer Science and Application

[Platt, 98] Platt J.: Sequential minimal optimization: "A fast algorithm for training support vector machines"; http://research.microsoft.com/en-us/um/people/jplatt/smoTR.pdf (1998, accessed July 2014).

[Royer, 99] Royer, M., Chai-Keong, T.: "A Review of Current Routing Protocols for ad hoc Mobile wireless Networks"; Personal Communications, IEEE , 6,2 (1999), 46-55.

[Sanaei, 13] Sanaei, M., Ghani,I., Hakemi,A.,Jeong,S.: "Routing Attacks In Mobile Ad Hoc Networks: An Overview"; In Science International, 25,4 (2013), 1031-1034.

[Shrivas13] Shrivas,K,Singhai, K.,Hota, S.: "An Efficient Decision Tree Model for Classification of Attacks with Feature Selection"; In International Journal of Computer Applications, 48, 1(2013), 42-48.

[Singh, 14] Singh, P.,Singh, G.: "Security Issues And Link Expiration In Secure Routing Protocols In Manet"; International Journal of Advanced Research in Computer and Communication Engineering ;3,7 (2014), 2278-1021.

[Tavallaee, 09] Tavallaee, M., Bagheri, E., Lu, W.,Ghorbani A.: "A Detailed Analysis of the KDD CUP 99 Data Set"; In Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defence Applications (2009), 1: 53-58.