I. J. Computer Network and Information Security, 2019, 1, 1-10
Published Online January 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2019.01.01

Identification of Compromised Nodes in MANETs using Machine Learning Technique

Rodney Sebopelo
Computer Science Department, North-West University, Mafikeng, South Africa
E-mail: rsebopelo@gmail.com

Bassey Isong and Naison Gasela
Computer Science Department, North-West University, Mafikeng, South Africa
E-mail: {bassey.isong, naison.gasela}@nwu.ac.za

Received: 03 August 2018; Accepted: 15 November 2018; Published: 08 January 2019

***Abstract*—**Mobile ad hoc networks (MANETs) are faced with various security challenges emanating from malicious attacks. Their dynamic nature make nodes more vulnerable to attacks from either malicious nodes or intruders since there is no fixed infrastructure resulting in each node acting as router to transmit data. Currently, several solutions have been proposed and implemented in different ways aimed at eliminating or reducing these malicious attacks. However, the attacks still persist. Therefore, this paper proposes an efficient security mechanism based on machine learning as a solution that detects and identifies malicious attacks in real-time basis by classifying packets data as either normal or abnormal. To achieve this, we conducted experiments using logistic regression (LR) and a support vector machine (SVM) to choose the best predictive model utilizing the Iris data set. The results obtained show that LR performed better than SVM with an accuracy of 100% detection rate. Thus, LR is better suited for the identification of malicious attacks in MANETs. Furthermore, we proposed and designed a framework to detect malicious attacks in real-time in MANETs based on packet behavior using the LR model and the components were presented. We believe that, if this framework is implemented in MANETs, it could go a long way to reduce the rate of attacks in the infrastructure less network.

***Index Terms*—**Security, Attacks, Intrusions, MANETs, Machine Learning.

I. Introduction

In recent years, research in the network paradigm called mobile ad hoc networks (MANETs) has received significant attention dominated by security challenges. MANET is an independent collection of mobile nodes characterized as dynamic, self-configuring and self-deployable that can communicate with each other without any central supervision [1, 2]. MANETs are wireless communication technology in which nodes communicate with each other without the need of a fixed or physical infrastructure regardless of the geographical location. This is made possible because each node acts as a router to facilitate the transmission of data from the source to the destination. Thus, due to their ad hoc or dynamic nature, a fixed infrastructure like a base station is needless. Moreover, having a strict layered architecture in MANETs exacerbates the ability to meet up with the dynamics of a wireless network setting and the absence of a central control point makes the networks more vulnerable to security or routing attacks compared to other networks [2]. According to Yang *et al.* [2], MANETs are more prone to security attacks because they have unique features such as open network architecture, shared wireless medium, and stringent power constraints. Moreover, Kumar [3], added that the dynamic and topological nature of the network may change rapidly and unpredictably overtime and such dynamic nature and mobility of nodes makes a MANET network vulnerable to attacks. Karpijoki *et al.* [4,5] also stated that attacks on MANETs may include actions such as dropping or amending packets, or gaining substantiation or authorized access by inserting forged packets into a data stream, having their effective output compromised for features like changing topology, restricted battery power, lack of centralized control and unreliability. Furthermore, Zhou [6,7] reported that the dynamic nature of the network topology has attracted many different application areas such as military tactical networks, wireless sensor networks, and many others. These applications however, have in turn introduced some design issues and challenges that need to be addressed.

Nevertheless, for a network to be secured, it has to meet the requirements for secured protocols to ensure that the confidentiality, availability, authenticity and integrity of the network is preserved. Hence, MANETs have to meet such requirements. Currently, several solutions have been proposed, developed and deployed which are either preventive or detective in nature [17, 18, 19, 20, 21]. Yet, MANET is still faced with the never-ending routing attacks. Therefore, this paper proposes a detective approach using machine learning (ML) as an efficient and viable solution. Our approach is to supplement the

approach in [17]. It involved detecting and identifying compromised nodes or packet dropping nodes in MANETs so that mitigation action can be taken early to prevent compromising the confidentiality and integrity of transmitted data. We used Iris data set [31] on two ML algorithms of support vector machines (SVM) and logistic regression (LR). The objective is to determine the more effective and efficient algorithms in the detection of intrusions or malicious attacks in MANETs. We utilized two metrics: packet delivery ratio (PDER) [17] and packet modification and misroute rate (PMMR). The results obtained show that LR performed better than SVM in terms of detection accuracy. The results allowed us to design a MANET intrusion detection framework based on LR model to detect malicious attacks or compromised nodes using learned information.

The rest of the paper is organized as follows: Sect. II presents security issues in MANETs, Sect. III discusses relevant related works, Sect. IV presents an overview of machine learning, Sect. V is the research methodology used and Sect. VI present the results and analysis. Sect. VII is the result discussions, Sect. VIII presents the proposed compromised nodes identification framework, Sect. IX presents the validity threats and Sect. X is the paper conclusion.

II. Security in MANETs

Security is one of the most key components for basic network functioning such as packets and routing protocols. The performance of a network can be disrupted when crucial security measurements are not taken into account when designing sensitive applications. This is the position MANETs found itself since early routing protocols failed to take into account several security measures. In particular, MANET has a strict layered architecture which has been found inept in coping with the dynamics within the wireless network environment. Thus, a single layer alone cannot handle security issues in isolation and consequently, MANET is vulnerable to several security attacks such as denial of service (DoS), black hole, gray hole, worm hole, flooding, impersonation, routing table runoff, packet modifying, selfish node and so on [17, 18, 20]. Furthermore, several methods have been adopted to solve the various attacks, nevertheless, these attacks are still at large due to the dynamic nature of the MANETs. Thus, it is vital and crucial that effective security countermeasures have to be adopted that take the wireless nature and dynamic changing topology of MANETs into consideration. To provide secure network communication in the network layer where attacks take place, there are security requirements that need to be addressed to ensure secure packet transmission such as confidentiality, availability, authentication, integrity and non-repudiation[8,9,10,11].

III. Related Works

As stated in the sections above, MANETs are faced with a series of security challenges that come in different patterns due to recent technological advancements. However, several improvements have also been made in MANETs over the years [3, 4, 5, 6, 7, 12 13, 14, 15, 16]. Unfortunately, these improvements are yet to produce a secure and efficient nature of MANET mobile devices. Consequently, it can be concluded that there are other security loopholes that are still lacking to secure MANET. Some of these solutions are discussed as follows:

Patel *et al.* [17] proposed an SVM for the detection and identification of packet dropping nodes in MANETs. SVM was used to categorize activities as either normal or malicious. The technique was implemented with ad hoc on-demand vector (AODV) routing protocol and the method was evaluated by the use of metrics namely PDER, packet modification rate (PMOR) and packet misroute rate (PMISR). In a similar work, Sukla *et al.* [18] proposed the technique of addressing the problem of packet forwarding misbehavior and proposed the mechanism to detect and remove two types of attacks: black hole and grayhole attacks that exhibit packet forwarding misbehavior. The method employed the technique of finding the chain of nodes cooperating misbehaviors which drop an important fraction of data packets. The study developed an algorithm to detect these attacks. Also, Nikos *et al.* [19], projected a two-phase detection technique to identify unauthorized nodes and compromised nodes for a specific service in MANET. This approach is enabled with the operations of the network that can be found in both link and network layers. Zero knowledge techniques were utilized to identify nodes which are not based on encryption algorithms that are either symmetric or asymmetric. Furthermore, Jhaveri *et al.* [20] used AODV to detect black and grey hole attacks on nodes. In this case, the sender node checks through the routes available to find out whether the received destination for its message is not damaged. To avoid any of the malicious node that might disrupt the transmission, the sender will have to broadcast a “check” and provide a destination message reply that would pursue the same route requested. Shila *et al.* [21] introduced an approach to investigate selective forwarding attack, gray hole in the wireless network using an algorithm based on routing AODV. The algorithm consisted of a phase’s counter-threshold, the technique that uses a detection threshold and the data packet counter to detect the attacks, and also query-based on acknowledgement of the intermediate mobile nodes to localize adversaries. In relation to the above discussed methods, our approach is an extension of the approach of Patel *et al.* [17]. To this end, we utilized LR to efficiently detect or identify compromised nodes in MANETs.

IV. Machine Learning Overview

ML gives the computer the ability to learn and adapt on the logical, binary and other operations that gathers information from a set of examples [22]. It is a powerful collection of techniques used for data mining and knowledge discovery. Its techniques are employed when

designing efficient and accurate prediction algorithms. ML requires the notion of sample complexity to measure and evaluate the sample size for the algorithm and learn different families of concepts [23]. There are several ML algorithms that exist such as SVM and LR. The choice of these algorithms is based on the fact that they have been widely used in the construction of predictive models in both networks and non-networks setting as well as their predictive capability and accuracy. They have used for intrusion detection such as in [17,21] and so on.

SVM is a supervised learning algorithm that can be used for both classification and regression challenges[24]. It belongs to a class of kernel functions rooted in the statistical learning theory and structural risk minimization [25]. The SVM technique is considered as one of the best linear classifiers that improves robustness and is not sensitive to the scarcity and correlation of characteristics of data compared to other classifiers. On the other hand, LR is a regression method that is used to make predictions based on a dependent variable [26]. In LR, the maximum likelihood ratio or estimation is used to determine the statistical significance of the training samples. This model is widely used in cases where predictions such as the presence or absence of characteristic or where the result is based on the values of set of predictor variables[27]. LR model is made up of predictors that are continuous and categorical, consisting of the binomial possible type of dependent variable that can accept only two values "0" and "1".

In this paper, SVM and LR will be used as classifiers to identify which model is more effective in the detection of a compromised node based on their performance.

V. Methodology

The methodology utilized in the selection of a predictive model in MANETs based on packets behaviors is captured in Fig. 1.

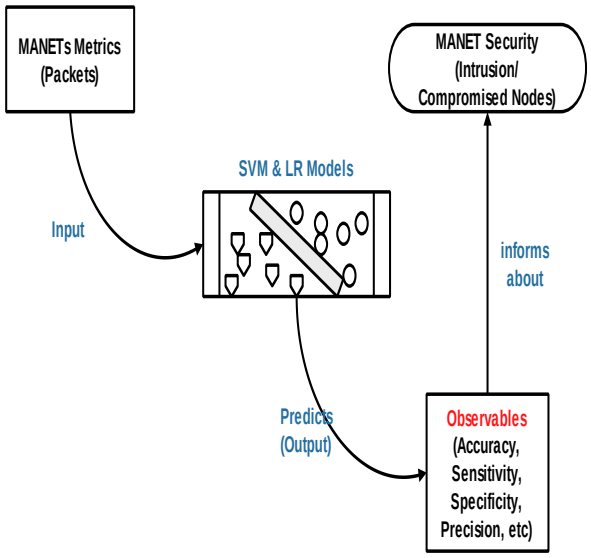

Fig.1. Research Framework

A. Model Parameters

The model parameters are given in terms of the variables utilized: dependent and independent. The independent variables are the MANET packets data (x_i) in the form of metrics such as PDER and PMMR which are inputs to the models. These metrics were employed in [17] to analyze the behavior and performance of each node in the network. However, in [17], PMMR were known as PMOR and PMISR metrics. We utilized them to determine the effectiveness of the models based on classification accuracies. Their computation is as follows:

$$PDER = \frac{No.\ of\ \ packets\ transmitted}{Total\ \ no.\ of\ \ incomming\ \ packets} \quad (1)$$

$$PMMR = \frac{No.\ of\ \ packets\ modified/misrouted}{Total\ \ no.\ of\ \ incomming\ \ packets} \quad (2)$$

In equation 1 and 2, the larger the number of PDER the better the performance of the mobile node. PMMR, shows changes observed in the packet content due to attacks during transmission. It could be done accidentally or intentionally to some of the network nodes. It can also be as a result of nodes sending packets to wrong destinations during communication. These metrics will be used to evaluate the performance of LR and SVM. The dependent variable are the values to be predicted. They include the targets (y_i) of MANET data in which class's value depends on the metric (x_i). The dependent variable outcome can account for dichotomous variables solutions that allow the results to be classified as intrusion or non-intrusion. In this paper, our model will classify MANETs packets as normal or abnormal packets.

B. Model Construction

This section presents the construction of the models: SVM and LR.

Support vector machine: For the SVM model, consider the problem of separating MANET packets data into two separate classes' normal and abnormal packets. The equation is of the form:

$$F\left(x\{w,b\}\right) = w.x + b = 0 \quad (3)$$

where *f* is the function that forms the separation of the two classes, and *x* represents the real valued n – dimensional input pattern of metrics (PDER and PMMR) of packets data and *n* denotes the subsets of MANET data, *w* is an orthogonal dimensional vector or weight vector that is to be computed by SVM convex optimization [28] and *b* is a bias term used to identify the perpendicular of a distance from the origin to the hyper planes.

$$w.x + b \geq +1 \ if\ \ y_i = 1 \quad (4)$$

$$w.x + b \leq -1 \ if\ \ y_i = -1 \quad (5)$$

Equations 4 and 5 represent the two parallel hyper planes that separate the two classes into normal or abnormal packets. Equation (4) denotes a linear separating hyper plane of MANET data with support vectors of the normal packet. The label $y_i = 1$ represents dependent variable of class normal packets. Linear equation (5) denotes the side of separating hyper plane of abnormal packets. The label $y_i = -1$ represents dependent variable of abnormal packets, and *x* indicates the input data of binary value.

Logistic regression: LR is given by:

$$f(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}} \quad (6)$$

Where the independent variables in $t \epsilon \mathbb{R}$ denotes the metrics (PDER and PMMR), where $\mathbb{R}$ denotes the real numbers. Dependent variable *f(t)* may denote either the "normal packet" or a binary "0", or "abnormal packet" or binary "1". Thus, f *(*PDER and PMMR*)* = "0" or "1" representing "Normal" or "Abnormal" respectively. However,

$$t = \beta_0 + \beta_i x \quad (7)$$

Where *t* denotes the linear combination of the metrics and coefficients. β_0 denotes the *y* intercept (the value of the criterion when the metric is equal to zero).

And,

$$f(t) = \frac{1}{1 + e^{-(\beta_0 + \beta_i x)}} \quad (8)$$

Where *f(x)* is the probability of the classes denoted as Normal or Intrusion and $\beta_i x$ is the regression coefficient.

C. Evaluation and Data Description

Model evaluation: To evaluate the performance of the models constructed, we employed cross validation and confusion matrix. Cross validation is a statistical method which compares and evaluates the learning algorithms by dividing the data set into two segments: a training set and a test set [30]. To evaluate both SVM and LR models using cross validation, we used the training set to train the models while the test set evaluates or validate the models. We employed the k-folds cross validation. Confusion matrix is used to evaluate the performance of the models constructed. It is an error matrix which summarizes prediction results on a problem of the classification in a table [29]. Thus, a confusion matrix will be used to compute the number of correct and incorrect predictions for MANET packet data. Parameters to be computed are the accuracy, sensitivity, specificity, precision, false positive (FP), false negative (FN), true positive (TP), true negative (TN) and miscalculation. These parameters will be used to assess the performance of SVM and LR models in both training and testing. Each row in the matrix represents the actual conditions of MANET data instances while columns denote the predicted conditions of MANET data, which compares the performance of SVM and LR algorithms.

Data description: In order to build the models, we used historical data that has been widely used for classification and regression analysis [24, 31]. Albeit it is not related to MANETs, it has been used for typical test cases of many statistical classification techniques in ML that include SVM, LR, Bayes' networks and so on. We used the Anderson's Iris data set obtained in [31] to quantify the structure of the iris flowers of three related species: *Iris setosa*, *Iris virginica*, and *Iris versicolor*. The data set contains 150 observations with 50 samples for each species of Iris and 4 features measured in centimeters: namely, the length and width of the sepal and petal. The dataset is also widely used for cluster analysis since it contains two clusters that are easily separated.

Based on the metrics employed: PDER and PMMR, the following manipulation were carried out on the two classes of the Iris which are the Iris setosa (50) and Iris versicolor (50) having four features each: the Sepal length, Sepal width, Petal length and Petal width. Thus, we computed average on each as follows:

Av_SepalIris = (Sepal length + Sepal width) / 2 = PDER
Av_PetalIris = (Petal length + Petal width) / 2 = PMMR

The given prepared metrics, will serve as input to the SVM and LR models.

VI. Results and Analysis

This section present the results of the classification performed using the selected ML algorithms as well as PDER and PMMR as metrics to evaluate the performance of the models. The models were simulated using Pycharm – Python. Each model was computed using three iterations.

Table 1. SVM Performance

Evaluation Parameter	Step 1 (%)	Step 2 (%)	Step 3 (%)
Accuracy: (% of correct prediction)	100	92	94
True Positive	12	9	14
True Negative	13	16	11
False Positive	0	0	0
False Negative	0	0	1
Misclassification	0	0	4
Sensitivity	94	94	93
Specificity	84	86	90
Precision	92	90	90

A. SVM Analysis

This subsection presents of the classification results of SVM model. Table 1 presents the information captured or collected randomly using the confusion matrix. It shows information about the performance of the SVM in the classification of MANETs packets as either normal or abnormal when delivered or modified in the network. Fig.2 shows the scattered plot. The indication of the results is that SVM algorithm produced an optimal separating hyperplane or a decision boundary between the two classes of MANET packet data.

The separation is based on the linear classifier method which is also based on a linear kernel function. SVM performed excellently in the classification without misclassification. The indication is that, when MANET packet data is transmitted in the network, the number of packet data received were correctly classified in real-time as either normal or abnormal. This is very important because knowing early when a node is compromised is critical to mitigating the situation before posing a real threat to the network and confidential information.

B. LR Analysis

LR provides knowledge of the relationships and strength among the MANET packets data travelling in the network. The classification result is presented in a scatter plot shown in Fig. 3. LR produced an excellent classification of MANETs packets data based on their category by learning from the delivered packets data. Table 2 shows the performance parameters obtained from the confusion matrix with a high accuracy in all the rounds. The indication is that MANETs packets delivered were classified correctly as normal or abnormal. That is, intrusion or degradation on the network performance or incorrect routing information were properly identified.

Table 2. LR Performance

Evaluation Parameter	**Step 1 (%)**	**Step 2 (%)**	**Step 3 (%)**
Accuracy (% of correct predictions)	100	100	100
True Positive	12	9	17
True Negative	13	16	8
False Positive	0	0	0
False Negative	0	0	0
Misclassification	0	0	0
Sensitivity	100	100	100
Specificity	100	100	100
Precision	100	100	100

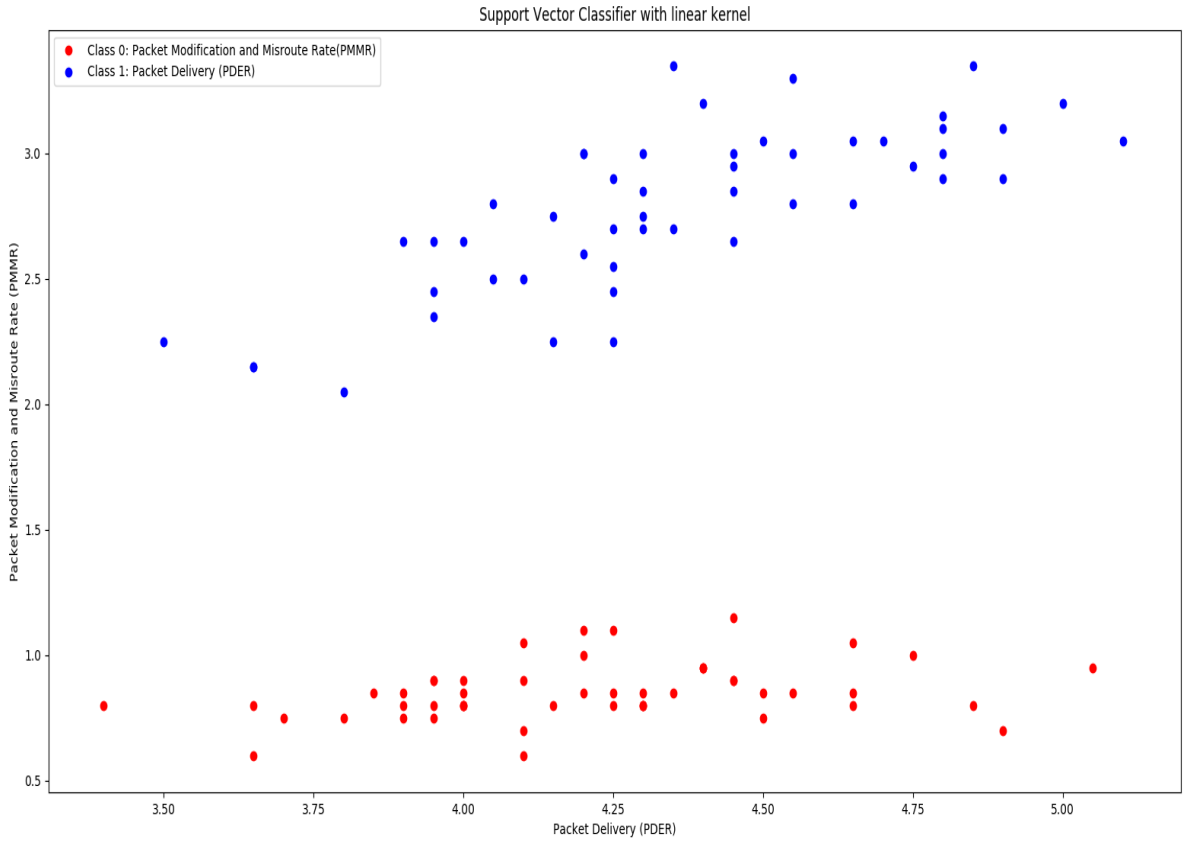

Fig.2. SVM Classification Plot

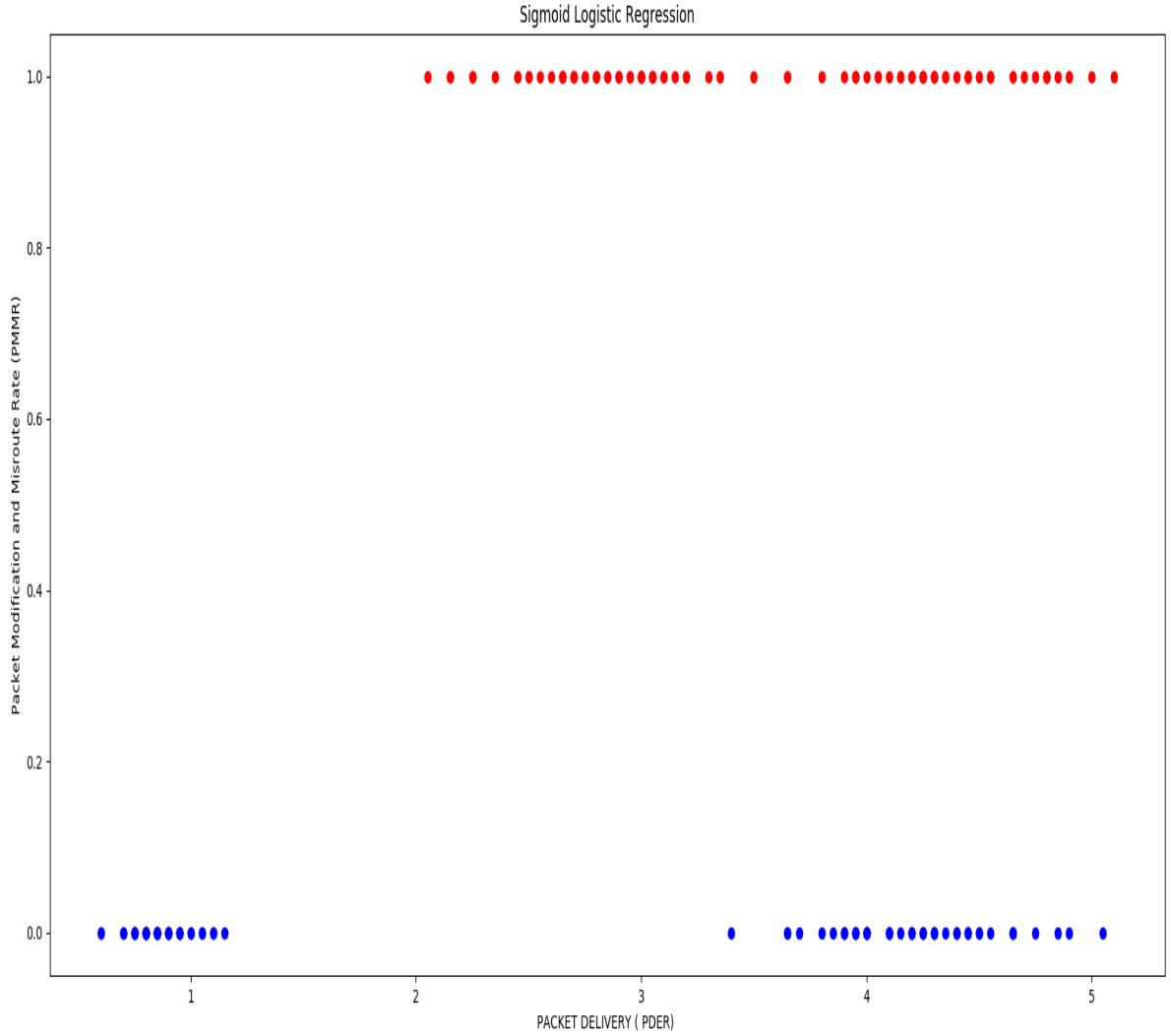

Fig.3. LR Classification Plot

VII. Discussion

In the above subsections, we presented the results obtained from the SVM and LR models using MANETs packet data. The objective is to identify the best performing model in terms of high accuracy and detection rate with low false alarm for communication and deliverance of MANET data packets in the network. Shown in Table 1 and Table 2 are the performance results of SVM and LR respectively. The results in Table 2 indicate that the rates or percentages of accuracy in the classification for "normal" MANET packet data by LR were excellently high in all the iterations as compared to the SVM model results in Table 1. In LR, the classification accuracy remained constant at 100% despite the increases or decreases in the percentage of true positive and true negative. This shows that, increase in the percentage of "normal" MANET data packets resulted in a decrease in the percentage of "abnormal" packets detected in the network. The rationale is that, the learning ability of "abnormal" packets data was negatively impacted since the amount of attack data to be learnt were insufficient. On the other hand, the high percentage of "normal" MANET packets data enhanced the learning process to understand more actual packets behaviour that is normal. Moreover, in Table 2 the parameters: FP, FN and misclassification were all 0% which indicates the low false alarm rate of the LR model.

In Table 1, SVM presented a predictive accuracy rate of 92% on the average as compared to 100% of the LR algorithm. The indication is that, it was difficult for the SVM algorithm to detect or identify all attacks on MANET packets which was not the case in LR. In a nutshell, from results shown in Table 1 and Table 2, it is convincingly clear that LR algorithm has the higher predictive accuracy in the classification of MANET packet data into "normal" and "abnormal" packets. LR also maintained consistency with the average of 100% in terms of precision, specificity, sensitivity, misclassification, and FP rate as computed in the confusion matrix. Thus, LR outperformed the SVM algorithm with detection capability of 100% accuracy for MANET packets. To this end LR model is chosen for usage in the design and development of MANET intrusion detection framework based on packets behavior.

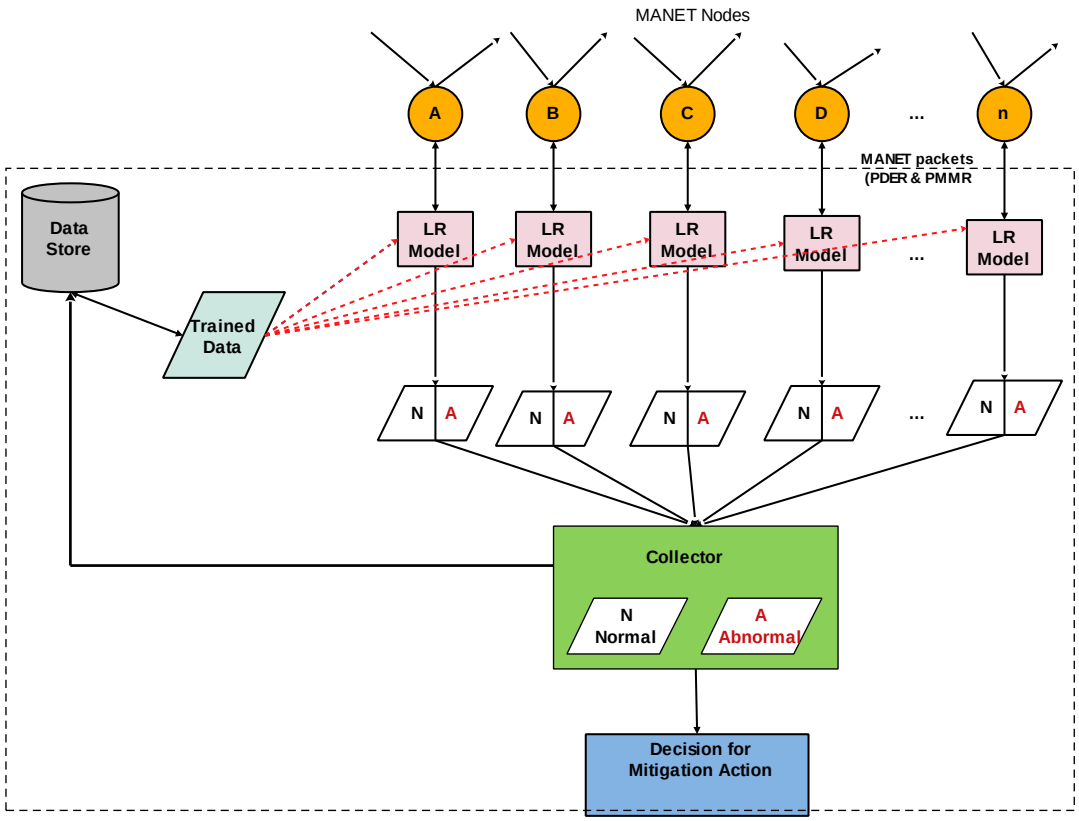

Fig.4. MANET Intrusion Identification Framework

VIII. Proposed Compromised Node Identification Framework

This section presents the proposed framework for predicting or identifying compromised nodes in MANET using the chosen ML algorithm – LR prediction model. The framework is provided with a regression method that is used to make probabilistic predictions based on classification analysis [32]. The model is integrated into the MANET which will accept packets as inputs and will be classified as "normal" and "abnormal" packets. The choice of LR stemmed from its high predictive power or predictive accuracy in the classification of MANET packet data during the training and testing as well the consistency it maintained in its good performance.

The framework is shown in Fig. 4. The framework operates by using historical data as training to provide the knowledge about the relationships and strengths in MANET packets data travelling in the network. For predictions regardless of the network characteristics, LR predicts or classifies the MANETs packets data based into their category by learning from the trained data. The components of the framework are discussed as follows:

A. The Network

This malicious attack detection framework is incorporated into the MANET as shown in Fig. 4 having several nodes. Each node in the MANET will receive and send a number of mobile packets when communicating with other nodes. The packets sent and received by each node is computed as PDER and PMMR in the network. They constitute the important measures used in the computation of packets' delivery and modified rates. Thus, PDER and PMMR will be used to identify if an intrusion occurs in the network or not as well as which node is compromised.

B. The Predictive Model

The LR model in the framework is the engine room that will automatically identify in real-time whether an intrusion occurred or not. The LR model is built using the trained data which are MANET historical packets data. The inputs are the MANET packet data computed as PDER and PMMR. Based on the trained data, once packets (PDER and PMMR) are received in the nodes, they are sent to the LR model to fit the model. The model operates by learning from the trained information to classify the packets according to their category: "NORMAL" or "ABNORMAL" packets. However, to construct the model using the data set (Iris) collected, Table 3 shows the descriptive statistics of the data, Table 4a and Table 4b shows the parameters for the two possible models, LRM1 and LRM2 constructed using the SPSS tool. We computed the LR parameter for LRM1 and LRM2 using (1) enter method and (2) the forward stepwise method with a cut-off value of 0.5 in 20 iterations. The statistics computed are the R^2 (Cox & Snell), the regression co-efficient (β), statistical significance value (ρ), the odds ratio (Exp(β)), - 2 Log likelihood and the constant (α).

Table 3. Descriptive statistics

Metric	N	Min	Max	Max	Std. Deviation
PDER	100	3.40	5.10	4.2825	0.35974
PMMR	100	0.60	3.35	1.8235	1.00311
Intrusion	100	0	1	0.50	0.503

Table 4a. LRM1 Enter Method

Metric	β	ρ-value	Exp(β)
PDER	-13.270	0.999	0.000
PMMR	31.253	0.995	3.739E+13
Constant (α)	2.982	1.000	19.727
-2 Log likelihood	0.000		
R^2	0.750		

Table 4b. LRM2 Forward Stepwise method

Metric	β	ρ-value	Exp(β)
PDER	-13.270	0.999	0.000
PMMR	31.253	0.995	3.739E+13
Constant (α)	2.982	1.000	19.727
-2 Log likelihood	0.000		
R^2	0.750		

Table 4c. LRM2 Forward Stepwise method

Metric	β	ρ-value	Exp(β)
PMMR	36.231	0.994	5.433E+15
Constant (α)	-58.173	.994	0.000
-2 Log likelihood	138.629		
R^2	0.750		

Moreover, the classification results for the LRM1 and LRM2 are summarized in Table 5 and the predictors for each model are shown in Table 6.

Table 5. Summary of Classification Results

	Predicted			
	Intrusion			
	Non-Intrusion		Intrusion	
LR Model	LRM1	LRM2	LRM1	LRM2
Non-Intrusion	50	50	0	0
Intrusion	0	0	50	50

Table 6. Predictors

Model	PDER (β)	PMMR (β)	Constant (α)
LRM_1	-13.270	31.253	2.982
LRM_2	-	36.231	-58.173

Therefore, based on Table 6, the LR or predictive model for intrusion or malicious attack detection in the MANET can be formulated as follows:

For LRM1, the model is:

(1/MANET_INTRUSION) = 2.982 + (-13.270) PDER + (31.2553)PMMR

For LRM2, the model is:

(1/MANET_INTRUSION) = -58.173 + (36.231) PMMR.

The choice of any of these models, LRM1 and LRM2 will depend on the value of the R^2 and the log likelihood statistics. For prediction, based on the threshold or cut-off value of 0.5, packets are classified as:

$$Prediction = \begin{cases} 0, Normal, MANET\_Intrusion \geq 0.5 \\ 1, Abnormal, MANET\_Intrusion < 0.5 \end{cases}$$

The overall algorithm is shown on Table 7.

As shown in Table 7, "authentic node" means that the node is not compromised in anyway in the network.

Table 7. Intrusion Identification Algorithm

LR intrusion identification algorithm
Collect all the MANET data packets in the network computed as PDER and PMMR Based on trained data, fit the model with PDER and PMMR If (MANET_INTRUSION $\geq$ 0.5) then Node is authentic (NORMAL) else (MANET_INTRUSION < 0.5) Node is suspicious (ABNORMAL)

C. Data Collector

This is the receiver of the output of the LR model once the MANET packets are classified into two classes: normal or abnormal. The classified data are then taken for decision making and storage in the data for reuse when needed.

D. Data Store

MANET data predicted or classified as "normal" and "abnormal" are stored in the data store for reuse as input for training the model in future.

E. Decision Making

Once the LR model output is received, decision can be made as to which node was compromised or not. This will then be followed by a mitigation action to identify the source of the attacks and steps taken to guard against further malicious attacks in the network that will compromise the integrity and confidentiality of information.

IX. Validity Threats

There are several threats that could invalidate the results analyzed and reported in this paper. As discussed in earlier sections, the Iris data [31] used is not related to MANET and the nature of the data values is quite different from the values of PDER and PMMR in a real-world MANET. Moreover, the adjustments made to the data in order to obtain the values that correspond to

PDER and PMMR packets rate of MANET might affect the predictive capacity of the ML algorithms. Also, there might be other suitable metrics better than PDER and PMMR that we did not see or utilize in this paper. Thus, we cannot generalize the results report in this paper to intrusion detection in MANETs. However, we are confident that the approach and the framework discussed in this paper will produce a high accuracy in the detection of malicious or intrusion attacks in the MANETs if adopted.

X. Conclusions

MANET is an infrastructure less and self-organizing network. This characteristic makes MANETs vulnerable to various types of attacks such as DoS, gray hole, worm hole, black hole and other attacks which may compromise the confidentiality and integrity of information transmitted in the network. In this paper, we have presented and discussed the results obtained from SVM and LR models built using MANETs network packet data. The objective was to identify the best predictive model in the detection of malicious attacks in MANET based on packets behaviour. That is, a model with high accuracy and detection rate with low false alarm for communication and deliverance of MANET packets in the network. The results obtained shows that LR outperformed SVM with a predictive accuracy of 100% while SVM showed 92%. The indication is that the SVM algorithm had difficulty in identifying or differentiating between “normal” and “abnormal” MANET packets. LR also maintained consistency in terms of precision, specificity, sensitivity, misclassification and FP rates. Based on these results, we proposed and designed a framework to detect malicious attacks in MANETs using the LR model. The components of the model and function were presented. Thus, based on its mode of operation, if adopted for use in the MANET environments, it could go a long way to help reduce the number of attacks and vulnerabilities faced by MANETs. Moreover, the future work is to implement the idea discussed in this paper in a real-world MANET network in order to assess its effectiveness and performance.

Acknowledgment

This research was supported by FRC and the Department of Computer Science at the NWU-Mafikeng and CSIR, South Africa.

References

[1] J. Hoebeke, I. Moerman, B. Dhoedt, and P. Demeester, "An overview of mobile ad hoc networks: applications and challenges," *Journal-Communications Network*, vol. 3, pp. 60-66, 2004.

[2] H. Yang, H. Luo, F. Ye, S. Lu, and L. Zhang, "Security in mobile ad hoc networks: challenges and solutions," *IEEE wireless communications*, vol. 11, pp. 38-47, 2004.

[3] R. P. Kumar, A. Excellencia, and P. Kanimozhi, "Providing a New EAACK to Secure Data in MANET‖," ed: IJREAT.

[4] V. Karpijoki, "Security in ad hoc networks," in Proceedings of the Helsinki University of Technology, Seminars on Network Security, Helsinki, Finland, 2000.

[5] M. Al-Shurman, S.-M. Yoo, and S. Park, "Black hole attack in mobile ad hoc networks," in *Proceedings of the 42nd annual Southeast regional conference*, 2004, pp. 96-97.

[6] L. Zhou and Z. J. Haas, "Securing ad hoc networks," *IEEE network*, vol. 13, pp. 24-30, 1999.

[7] R. S. Singamsetty, "Detection of malicious nodes in mobile ad hoc networks," University of Toledo, 2011.

[8] P. Yau and C. J. Mitchell, "Security vulnerabilities in ad hoc networks," in The Seventh International Symposium on Communication Theory and Applications, July 13–18, 2003, Ambleside, Lake District, UK, 2003, pp. 99-104.

[9] B. Kaur, "Security Architecture for MANET and Its Application in M-Governance," in *Communication Systems and Network Technologies (CSNT), 2013 International Conference on*, 2013, pp. 491-496.

[10] M. Kumar, A. Bhushan, and A. Kumar, "A study of wireless ad-hoc network attack and routing protocol attack," *International Journal of Advanced Research in Computer Science and Software Engineering ISSN*, vol. 2277, 2012.

[11] S. Boora, Y. Kumar, and B. Kochar, "A Survey on Security Issues in Mobile Ad-hoc Networks," *IJCSMS International Journal of Computer Science and Management Studies*, 2011.

[12] D. Djenouri, L. Khelladi, and N. Badache, "A survey of security issues in mobile ad hoc networks," *IEEE communications surveys*, vol. 7, pp. 2-28, 2005.

[13] J.-S. Li and C.-T. Lee, "Improve routing trust with promiscuous listening routing security algorithm in mobile ad hoc networks," *Computer communications*, vol. 29, pp. 1121-1132, 2006.

[14] W. Zhang, R. Rao, G. Cao, and G. Kesidis, "Secure routing in ad hoc networks and a related intrusion detection problem," in *Military Communications Conference, 2003. MILCOM'03. 2003 IEEE*, 2003, pp. 735-740.

[15] S. B. S. G. Varaprasad, "Identification of Critical Node for the Efficient Performance in Manet."

[16] B. Sivakumar and G. Varaprasad, "Identification of critical node for the efficient performance in Manet," *Editorial Preface*, vol. 3, 2012.

[17] N. J. Patel and R. H. Jhaveri, "Detecting Packet Dropping Misbehaving Nodes using Support Vector Machine (SVM) in MANET," *International Journal of Computer Applications*, vol. 122, 2015.

[18] S. Banerjee, "Detection/removal of cooperative black and gray hole attack in mobile ad-hoc networks," in *proceedings of the world congress on engineering and computer science*, 2008.

[19] N. Komninos, D. Vergados, and C. Douligeris, "Detecting unauthorized and compromised nodes in mobile ad hoc networks," *Ad Hoc Networks*, vol. 5, pp. 289-298, 2007.

[20] R. H. Jhaveri, S. J. Patel, and D. C. Jinwala, "A novel approach for grayhole and blackhole attacks in mobile ad hoc networks," in *Advanced Computing & Communication Technologies (ACCT), 2012 Second International Conference on*, 2012, pp. 556-560.

[21] D. M. Shila and T. Anjali, "Defending selective forwarding attacks in WMNs," in *Electro/Information Technology, 2008. EIT 2008. IEEE International Conference on*, 2008, pp. 96-101.

[22] T. O. Ayodele, "Introduction to machine learning," in *New Advances in Machine Learning*, ed: InTech, 2010.

Authors' Profiles

Rodney Sebopelo is currently an MSc student in the Department of Computer Science, Mafiikeng campus, North-West University. His research interests include Networking, Cloud Computing, and Machine Learning.

Bassey Isong received B.S. degree in computer science from the University of Calabar, Nigeria, in 2004, M.Sc. degrees in Computer Science and Software Engineering from Blekinge Institute of Technology, Sweden, in 2008 and 2010 respectively and a Ph.D. degree in Computer Science from the North-West University, South Africa, in 2014. He is currently a Senior Lecturer in the Department of Computer Science, North-West University, Mafikeng, South Africa. He is a member of the IEEE Computer, Communication and Education Societies. His research interests include and not limited to Software Engineering, Cloud Computing, Software Defined Networks, Internet of Things, Cybersecurity, Machine Learning and Computer Science Education.

Naison Gasela is an Associate Professor and HoD of the Department of Computer Science, North-West University, Mafikeng Campus. He is a member of the ACM and IEEE. His research interests include and not limited to Artificial Intelligence, Machine Learning, Software Engineering, Algorithms and Computer Network.

How to cite this paper: Rodney Sebopelo, Bassey Isong, Naison Gasela,"Identification of Compromised Nodes in MANETs using Machine Learning Technique", International Journal of Computer Network and Information Security(IJCNIS), Vol.11, No.1, pp.1-10, 2019.DOI: 10.5815/ijcnis.2019.01.01

