ANOMALY DETECTION IN MOBILE ADHOC NETWORKS (MANET) USING C4.5 CLUSTERING ALGORITHM

S. Radha Rammohan
Department of Information Technology, Salalah College of Applied Sciences, Directorate General of College of Applied Sciences, Ministry of Higher Education, Post Box No. 1905, Postal Code-211 Sultanate of Oman, Oman

ABSTRACT
Mobile adhoc networks, abbreviated as MANETs are more vulnerable to various attacks than other types of networks. This is because of natural properties of MANETs such as dynamic topology, ever changing infrastructure and lack of central controller. Hence detecting anomaly behaviours in MANETs is a challenging task and this topic has attracted many researchers. In this paper, clustering method based anomaly detection method is proposed. The method uses C4.5 clustering algorithm to classify the events as normal or abnormal. The proposed method works in three phases viz. training, detection of anomaly behaviour and identifying the type of attack. The proposed method uses 141 features of network to identify the attacks. The experiments were conducted on NS2 simulator and results show that the proposed method is effective in detecting anomalies. Moreover, False Alarm Ratio (FAR) is also low which ensure that the detected anomalies are real attacks.

Key words: MANETs, Anomaly, Attacks, detection, C4.5, clustering, Denial of Service and False Alarm Ratio (FAR)

http://www.iaeme.com/issue.asp?JType=IJITMIS&VType=7&IType=1

1. INTRODUCTION
Mobile Adhoc Networks, abbreviated as MANETs, is a type of wireless networks that continuously changes its locations and configures itself. Since MANETs contain autonomous nodes, the network topology may change rapidly with an unpredictable manner[1]. Many of the existing and real time applications prefer wireless networks
as a preferred choice. Few important applications of MANET include mobile conferencing, home and community networking, battle field operations, emergency rescues, sensor dust, tactical operations, environmental monitoring, etc[1][2].

Figure 1 MANET

Manets provide wireless communication among several mobile nodes. A simple MANET is formed by grouping several mobile wireless nodes without backing fixed network infrastructure. In MANETs, mobile nodes cooperate with each other by forwarding packets[3]. This forwarding allows mobile nodes to communicate with other mobile nodes that reside outside the reach of radio range.

With continuously changing network topology, communicating in the open medium with other mobile nodes and non-availability of proper strategies to manage and monitor MANET communications, these networks are often become victims of attacks[3]. Hence MANETs attracted many researchers and it is a popular research topic. As said earlier, MANETs are always endangered with attacks than wired networks. There are various issues that need to be addressed while designing a framework for avoiding attacks. Non-availability of centralized monitoring sever makes MANET difficult to detect attacks[4]. In MANET architecture, resources are very minimal. Providing secure communication in MANET is difficulty because of ever-changing topology and size of the network. Several other issues in MANET include scalability, limited power supply, bandwidth problems and lack of predefined boundary. Hence it is difficult to provide security in MANETs[4].

Providing security solution in MANET is providing security to each mobile node in the network as well as to protect the entire MANET. Protecting the network is nothing but restricting unauthorized access in the mobile adhoc network. It is some sort of defense-in-depth approach that protects the entire MANET. Hence the security solution must contain techniques relating to the prevention, detection and reaction to unauthorized access[4][5]. In order to evaluate a nodes access in network, a framework need to be designed that will classify an access as normal or abnormal. To create this sort of framework, the traditional way is to detect anomaly in audit trail data generated by operating systems. It is the record of activities on a system in chronological order. Inspecting these records manually is not feasible due to its large size. Hence some data mining technique need to be used for analyzing the audit data[5].

M.Alikhani and M.Ahmadi Livani [6] proposed dynamic anomaly detection using incremental approximation PCA in AODV based MANETs. The proposed method
Anomaly Detection in Mobile Adhoc Networks (MANET) using C4.5 Clustering Algorithm

contains three phases viz. training, detection and updation. The simulation results show that average anomaly detection rate and false alarm rate is high when compared to WPCA. Yi-an Huang et al [7] proposed cross-feature analysis for detecting anomalies in adhoc routing, where existing security mechanisms used in wired networks have been redesigned for MANETs. This method uses cross-feature analysis that captures inter-feature correlation patterns in normal traffic. The resulting model is easy to understand and can be examined by humans.

Christoforas Panos et al [8] proposed anomaly detection engines for mobile adhoc networks, which provides evaluation and comparison of various anomaly detection strategies. The paper provides limitations of several popular detection engines such as high rates of false alarms, processing over head and lack of adaptability under dynamic network conditions. Kyong-Heon Han et al [9] proposed cross layer anomaly detection in mobile adhoc networks, which uses multi layer detection frame work to detect anomalies. Detection is done by exploiting the information available across different layers of protocol stack. The method uses fixed width clustering algorithm for efficiently detecting anomalies in MANETs. The method is also capable of detecting various types of UDP flooding attack and sinkhole attacks. Bo Sun et al [10] proposed zone-based intrusion detection system for mobile adhoc networks. The paper suggests that local detection scheme alone cannot detect anomalies completely. Hence a cooperative mechanism is needed to achieve desirable performance. The method uses Markov Chain based local anomaly detection model which includes feature extraction, data preprocess, detection engine construction and parameter tuning.

All the above methods work on individual nodes to detect anomalies. The above methods are capable of detecting various attacks but centralized and distributed frame work that can efficiently detect the anomalies is the need. Hence this paper proposes a collaborative mechanism to detect anomalies in MANETs.

The paper is organized as follows: Section 1 briefs with introduction to MANETs and literature survey. Section 2 describes anomaly detection techniques, challenges in MANETs and various types of attacks prevailing in MANETs are discussed in section 3. Section 4 explains proposed method and experimental results, conclusion is given in section 5 and 6 respectively.

2. ANOMALY DETECTION TECHNIQUES

Anomaly detection is the process of identifying data points or items or observations that do not confirm the expected pattern of a given group. Anomaly detection is widely used in behaviour analysis and it is like outlier analysis. This anomaly is applicable to domains such as fraud detection, intrusion detection, fault identification, event detection systems and system health monitoring [3][11].

Anomaly based detection can be classified into three group of techniques. Statistical based techniques monitor the traffic activities and creates a profile indicating the behaviour. Profile construction is done using some metrics such as packet rate, traffic rate, communication rate and IP address [11]. Each node will have normal profile. When anomaly occurs, profile relating to current scenario is created and anomaly privilege is calculated. If the privilege value exceeds the threshold, then the activity is considered as abnormal.

Knowledge based anomaly detection strategies examine the audit data and based on some set of predefined rules, an event is classified as normal or abnormal. These strategies work in three phases; training, rules extraction and classification according
to rules. This model is capable of detecting illegal behaviour patterns. This method is robust and flexible whereas development of high quality knowledge is often difficult and time consuming[11].

Machine based anomaly detection techniques works by creating a model that classify pattern of observed behaviour. The model needs to learn the behaviour of all nodes in the network. This requires resources like energy and bandwidth. In certain cases, machine learning strategy is same like statistical based strategy[12].

3. CHALLENGES IN MANETS

The dynamic nature of MANETs makes it difficult to design a good technique for anomaly detection. There are several challenges that an anomaly detection technique has to address[12][13]. Few important ones are:

- Cooperativeness – During routing process, MANET believes that intermediate nodes are non-malicious and are cooperative to the security threats. Malicious attacker takes advantages of this sovereignty and enters into the routing path by becoming one of the intermediate nodes.
- Adversary inside the network – As per the MANET concept, nodes can freely join and leave the network. This property of MANET makes difficult to distinguish between genuine and malicious nodes. Such malicious nodes are called compromised nodes.
- No predefined boundary – There is no physical boundary for a node in MANET. Hence nodes can freely roam in MANET. This allows malicious nodes to perform eavesdropping impersonation, DOS and tempering type of attacks.

3.1 Attacks in MANET

There are several types of attacks that can be registered in MANETs. Few among them are:

- Sink hole / Black hole attack – A node advertise itself that it is having the shortest path to other nodes. This makes the network protocol to direct all the traffic to the specific node. This specific node may not forward any traffic at all.
- Eavesdropping attack – The malicious node secretly listens the communication happening in the other nodes, gathers information.
- Sybil attack – It is a type of attack where a malicious node creates multiple identities. These fake identities are either fabricated or stolen from other genuine nodes. Multiple identities make a malicious node to appear in more than one location in the internet simultaneously.
- Node capture attack – An attacker completely captures the genuine node and the readings sensed by genuine node is either manipulated or made inaccurate. Attacker may extract security information from captured node and misuse it.
- Network partition – An attacker divides the network into many partitions and node in one partition cannot communicate with the node in other partition even though a path exists between them.
- Denial of service attack – An attacker prevents nodes from sending and receiving data from other nodes.

Apart from the above types of attacks, there are other types of attacks such as routing loop attack, selfishness attack, sleep deprivation attack, location disclose attack, warm hole attack, routing table overflow attack, Byzantine attack, etc. These
are several attacking techniques ranging from simple cache poisoning to malicious flooding[14].

4. PROPOSED ANOMALY DETECTION METHOD

Anomaly detection is the process of searching events which do not conform to an expected pattern. The patterns detected are said to be anomalies. The other names for anomalies are outliers, changes, deviations, intrusions, surprises, etc.

Let \( f = \{ f_1, f_2, f_3, \ldots, f_i \} \) be the feature set containing the features and \( C_i \) be the classifier. The classifier \( C_i \) can be learned from a set of training data. The classifier predicts the most likely value for the given feature set. By applying classifier \( C_i \), the probability of the given values of the features are calculated. The average probability is calculated as follows:

\[
\text{average}(P) = \frac{\sum P_i}{L}
\]

If the feature is below the average probability, it is assumed to be anomaly. Else the feature is normal. The next step is to identify the type of attack that an anomaly feature exhibit. This is done by matching the anomaly behaviour with predefined identification rules. In general, the node which runs the anomaly detection technique is said to be monitoring node and node whose behaviour is tracked is said to be monitored node. If the attack is packed dropping, the monitoring node is 1-hop neighborhood of the monitored node[14]. For flooding type of attacks, only attack type is identified.

Few notations used in anomaly detection techniques are as follows:

- \#(*, m) \rightarrow number of incoming packets on the monitored node
- \#(m,*) \rightarrow number of outgoing packets on the monitored node
- \#([(m),*]) \rightarrow number of outgoing packets for which monitored node is the source
- \#(*,[m]) \rightarrow number of incoming packets for which the monitored node is the destination
- \#([s], m) \rightarrow number of incoming packets to the monitored node for which ‘s’ is the source
- \#(m,[d]) \rightarrow number of outgoing packets from the monitored node for which ‘d’ is the destination
- \#(m, n) \rightarrow number of outgoing packets from monitored node for which ‘n’ is the next hop
- \#([s], M, [m]) \rightarrow the number of packets from source ‘s’ to destination ‘M’ through ‘m’

4.1 Anomaly detection using clustering

One way of detecting anomalies is to use clustering technique. Clustering is used to find the group of similar data points in such a way that each group of data is well distinguished. Several clustering algorithms exist and in this, C4.5 clustering algorithm contains certain advantages. C4.5 algorithm is well suited for classification...
type of problems[15]. It deals with both continuous and discrete attributed, finding missing values and pruning trees. It is a supervised learning algorithm and requires training[16].

In clustering algorithm, node which is close is said to be cluster head and other nodes as citizens. The size of the cluster is defined as the number of nodes in the cluster. To select a cluster head, following algorithm is used:

- Each node in the cluster contributes a random value to the input
- Compute the integer from 0 to \( S_c - 1 \) where \( S_c \) is the size of the cluster
- The above selection can be done using the selection function

The next step is to convert the data vector into a suitable form which the clustering algorithm can understand and use. For this, data conditioning operation is performed. Once clustering is over, similar clusters are combined to form one single cluster. Two clusters \( C_1 \) and \( C_2 \) are said to be similar if the inter-cluster distance between their centers is less than the fixed radius. A new cluster \( C_3 \) is formed by merging similar clusters \( C_1 \) and \( C_2 \) with center as mean of fixed radius of \( C_1 \) and \( C_2 \).

The procedure for training the normal profile is as follows

**Algorithm training**

**Input**:
- Set of normal features \( X_N(t) \)

**Output**:
- Normal profile \( P(0) = (\varphi(0), \mu(0)), d_{max}(0) ) \)

**Begin**

1. expand \( X_N(0) \) \( \rightarrow \) \([0, 1] \)
2. \( X_N(0), (\varphi(0), \mu(0) ) \)
3. //compute column-centered matrix, column-means vector & first principal component//
4. for each \( x_{i0} \in X_N \) do
   1. calculate \( d_p ( x_{i0}, \varphi(0) ) \)
   2. //projection distance//
5. \( d_{max}(0) = \max\{ d_p ( x_{i0}, \varphi(0) ) \} \)

**End**

4.2 Detection algorithm

To categorize a cluster as normal or abnormal, average inter-cluster distance of nearest neighbour is used. The below method is used.

Let \( C \) is the cluster set, modulus of \( C \) is the number of clusters in \( C \), the inter-cluster distance is calculated by using the formula
Anomaly Detection in Mobile Adhoc Networks (MANET) using C4.5 Clustering Algorithm

\[ D_i = \sum_i \sum_j \| x_i - x_j \|^2 = 2n_r \sum_i \| x_i - \bar{x} \|^2 \]

The above formula is used for only one cluster. For k clusters

\[ W_k = \sum_{r=1}^{kj} \frac{1}{2n_r} D_r \]

The average inter-cluster distance is calculated as below

\[ \frac{1}{k} \sum_{j=1}^{k} d(c_i, c_j) \text{ if } k \leq |c| - 1 \]

\[ \frac{1}{|c| - 1} \sum_{j=1}^{|c| - 1} d(c_i, c_j) \text{ if } k > |c| - 1 \]

If the value of AD for a cluster is more than the standard deviation of AD, then the cluster is abnormal else normal. The routine used to detect anomaly is as follows:

**Algorithm anomaly_detect**

**Input :**
- Normal profile \( p(t-1) = (\phi(t-1), \mu(t-1), d_{max}(t-1)) \)
- Feature Vector \( f(t) \)

**Output**
- Set of normal behaviour \( X_N(t) \)
- Set of anomalies \( X_{AB}(t) \)

**begin**

For each \( f(t) \) do

\( d_p(x_{it}, \phi(t-1)) \)

If \( d_p(x_{it}, \phi(t-1)) \leq d_{max}(t-1) \) then

\( X_{N}(t) = X_{N}(t) \cup x_{it} \)

Else

\( X_{AB}(t) = X_{AB}(t) \cup x_{it} \)

Endif

**Endfor**

5. EXPERIMENTAL RESULTS

The proposed method on MANET is simulated using NS2 simulator. The number of features used are 141 and few among them are source IP address, destination IP address, route request from node I, reply request from node I, total packet received, total packet sent, etc. By using cbrgen.tel program, CBR traffic model with 512 byte data for 1000 seconds are captured. The experimental parameters are as follows:

| Table 1 Parameters List |
The protocol that is used in our experiment is Adhoc on Demand Distance Vector Routing protocol. All the events are labeled either as normal or abnormal. The experimental results show that our model detects anomaly in a better way. Furthermore, false alarm rate is also very low. The experimental results are tabulated below:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation time</td>
<td>1000 sec</td>
</tr>
<tr>
<td>Mobility model</td>
<td>Random Walk Point</td>
</tr>
<tr>
<td>Classifier</td>
<td>C4.5</td>
</tr>
<tr>
<td>Execution time for training</td>
<td>1000 sec</td>
</tr>
<tr>
<td>Execution time for testing</td>
<td>1000 sec</td>
</tr>
<tr>
<td>Pause time</td>
<td>5 sec</td>
</tr>
<tr>
<td>Maximum connections</td>
<td>30</td>
</tr>
<tr>
<td>Maximum bandwidth</td>
<td>2 MBPS</td>
</tr>
<tr>
<td>Topology</td>
<td>1000 m x 1000 m</td>
</tr>
<tr>
<td>Transmission rate</td>
<td>250 MBPS</td>
</tr>
</tbody>
</table>

From the above table, it is clear that detection rate is higher and false detection alarm rate is low. Moreover, the values of detection rate, partial detection rate and misclassification rate are well suited for detecting rushing, neighbourhood and flooding type of attacks.

6. CONCLUSION

In this paper, detecting anomaly behaviours in MANET using C4.5 clustering algorithm is proposed. The method contains three steps with training the clusters in the first step, followed by detecting abnormal activities in the next step and finally identifying the attack. In the training phase, normal network features are used to create the network profile. All the activities of the network are classified as either normal or abnormal. Abnormal data set is analyzed and the type of attacks is identified. For a good anomaly detection algorithm, false rate alarm should be minimum. Experimental results of the proposed method showed that false rate alarm is very low. Moreover, the anomalies detected are highly accurate.

REFERENCES


Anomaly Detection in Mobile Adhoc Networks (MANET) using C4.5 Clustering Algorithm


[10] BO SUN, “INTRUSION DETECTION IN MOBILE AD HOC NETWORKS” A Dissertation Submitted to the Office of Graduate Studies of Texas A&M University, May 2004


