Constructing an Efficient Mobility Profile of Ad-Hoc Node for Mobility-Pattern-Based Anomaly Detection in MANET

Chaoli Cai  
Dept. of Computer Science  
Western Michigan University, U.S.  
Email: ccai@cs.wmich.edu

Song Ci  
Dept. of Electrical and Computer Engineering  
University of Nebraska-Lincoln, U.S.  
Email: cisong@ieee.org

Sghaier Guizani  
Dept. of Electrical Engineering  
University of Quebec a Trois Riviers, CA  
Email: guizans@umoncton.ca

Ala Al-Fuqaha  
Dept. of Computer Science  
Western Michigan University, U.S.  
Email: alfuqaha@cs.wmich.edu

Abstract—Numerous approaches have been proposed for intrusion detection, especially for anomaly detection, in ad hoc networks. However, little research work has been done in actually implementing such a scheme based on statistical methods. In this paper, we present an efficient anomaly detection algorithm based on a statistical method originated from pattern recognition, which can effectively identify abnormal behavior such as mobility pattern of MANETs. In the proposed algorithm, the mobility pattern of a specific node is characterized by a multi-leaf tree structure, second-level nodes stands for the possible starting points and leaf nodes stand for the destination node of each possible path. Since our algorithm is using statistical method, a normal profile for each node is generated through extensive experiments, where the specific tree generated may have several starting points and ending with several destination points. For each path between any two nodes (parent and children), we can get the distribution of every different mobility pattern. By comparing the mobility patterns with the training data, we can distinguish abnormal nodes from normal behavior nodes in mobile Ad Hoc networks. Simulation results demonstrate that our proposed detection algorithm can achieve good performance in terms of false alarm rate and detection rate for nodes with regular mobility patterns.

I. INTRODUCTION

A wireless mobile ad-hoc network consists of a group of "peer" mobile nodes which are capable of communicating with each other without a fixed infrastructure. The unique nature of wireless ad-hoc networks such as arbitrary node movement and lack of centralized control make them very vulnerable to a wide variety of outside and inside attacks. How to provide effective security protection for MANETs has become one of the main challenges in deploying MANET in practice.

In general, two complementary approaches exist to protect a system: prevention and detection. Intrusion prevention techniques, such as encryption and authentication, can deter attackers from malicious behavior. The history of security research has showed that prevention based techniques alone can only reduce the intrusions rather than totally eliminate them. No matter how many intrusion prevention measures are deployed, sooner or later, an attacker can exploit some security holes to break into a system. Therefore, intrusion detection systems (IDSs), served as the second line of defense, are indispensable for a reliable system.

Generally, there are two intrusion detection techniques: misuse based detection and anomaly based detection. A misuse based detection technique encodes the known attack signature and system vulnerabilities. If it finds a match against current activities, an alarm is generated. Misuse detection techniques are not effective in detecting novel attacks. An anomaly based detection technique creates normal profiles of system states and node behaviors and compares them against current activities. If a significant deviation is observed, an alarm is triggered. Anomaly detection can detect unknown attacks. Obviously, under the ad-hoc environment, anomaly detection technique is more crucial than misuse detection techniques since it is impossible for us to know all the attack patterns in advance. However, the normal profiles are usually very hard to build due to the mobility of nodes. Therefore, how to establish normal profiles for mobile nodes is crucial in designing an efficient intrusion detection algorithm.

There are recent research efforts in this area. Some of these are prevention techniques [3]. The authors addressed the guidelines for secure routing protocol (SecAODV) which can prevent most routing disruption attacks such as rushing attacks, man-in-the-middle attack, resource consumption attack, etc. The authors also discussed the scalability and introduced the possible solution - threshold. Therefore, not only can they improve the scalability but reduce the false positive due to congestion as well. Other research addressed the architecture problems in MANETs [11]. In that paper, the author gives a general idea on how to build cooperative, distributed intrusion detection architecture. The architecture is organized as a dynamic hierarchy. To maintain communications efficiency,
the hierarchy is automatically reconfigured in an on-demand fashion using clustering techniques. The utility of the architecture is illustrated via multiple attack scenarios. Another group of research works focus on the misuse detection technique [9]. In that paper, the authors focus on signature based (misuse) intrusion detection technique and investigate the ability of different ad-hoc network routing protocols. The authors clearly show that reactive routing protocols (AODV, TORA, DSR) are less effective than proactive ones, such as DSDV. The authors also investigate the relationship between the probability of detecting an attack and percentage of nodes that have to be part of the IDS.

As discussed above, misuse IDS in ad hoc networks is not effective since there exist so many unknown attacks. Therefore, a new research area is to develop anomaly detection techniques [1]. According to this scheme, each node that participates in a mobile ad hoc network analyzes locally available network data for anomalies. Intrusion attempts are detected by employing a distributed cooperative mechanism, in which all participating nodes cast votes according to the data they have previously analyzed. The authors avoid the reliance on known attack patterns by using an anomaly detection model. However, two main problems exist in that paper. First, trace data (feature or audit data source) design is not complete, that is, what information a routing protocol should include making the IDS effective. Second, detection model design is not clear, that is, when to initiate the intrusion response? Furthermore, such IDS models suffer from performance penalties and high false alarm rates.

Our work is based on the following assumption: the ad hoc node and all its secrets except GPS information can be captured by an attacker who in turn can do whatever he wants without being captured. But the attacker does not know the mobility pattern of the authentic node in advance and can not change it, since a node’s mobility pattern is a reflection of the routines of its life and different ad hoc nodes have different favorite routes and habitual movement patterns. In this paper, we propose a novel approach to construct the normal profile of a node, from which an efficient anomaly detection algorithm is designed. The sequence of points traversed by a node and its pattern distribution are used as the feature value. When an intrusion occurs, the attacker tends to have a different mobility pattern. Therefore, we can detect anomaly by comparing the mobility pattern.

The rest of the paper is organized as follows. In section 2, we present our assumptions in developing detection techniques for mobile ad-hoc networks. Section 3 describes the threat model, network model and mobility model. Section 4 presents the details of constructing the normal profile especially the mobility pattern of each node based on the statistical method. Section 5 presents the simulation study of our proposed detection approach. In section 6, we conclude the paper and point out ideas for future research work.

II. ASSUMPTIONS

Our anomaly detection algorithm relies on the following assumptions:

First, we assume each ad hoc node has a specific mobility database which describes its normal activities such as mobility patterns. All node mobility databases are stored in an absolute secret place which is hard to be compromised. And there exits at least one device such as GPS inside each ad hoc node which can provide the accurate location of the node at any time. It will be hard for an attacker to hide or fabricate his location if he uses the captured node. Even if an attacker finds some magical method to fabricate his location, he still does not know how to manipulate such GPS device which is inside the node.

Second, we assume ad hoc nodes can be compromised and all secret information associated with the compromised nodes is open to attackers. This assumption is reasonable since nowadays tamper-resistant hardware and software are still expensive to ad hoc nodes. This assumption justifies our research in anomaly detection, that is, all prevention approaches will be helpless once ad hoc node is captured and compromised.

Finally, we assume most ad hoc nodes have fairly regular routes. Therefore, it is viable to create the normal profile for each node. Actually, all research jobs on intrusion detection are based on the following two assumptions: 1) the subject activity is observable and 2) the normal and abnormal activities demonstrate distinct behavior. We exclude such scenarios that a node has totally random behavior. For this type of nodes, our current approach based on mobility pattern is inaccurate.

III. MODEL DESCRIPTION

A. Threat Model

Since mobility is the most typical and crucial feature in MANETs, this paper will focus on the detection of attacks targeted at MANET mobility patterns, more specifically on detecting one of the most important active attacks: changing the normal mobility patterns from any starting point to the subsequent points within any geographical area. Fig. 1 depicts an example of the attack of this type:

![Fig. 1. An Example of Threat Model.](image)

In the above example, A1 is one of the starting points within geographical area A, and B1 is one of the following subsequent points within geographical area B. Path A1, O, B1 is the normal mobility pattern from A1 to B1. If this ad hoc node has been compromised in A1, then even if it can follow path A1, m’, m’’, B1 to the same following points B1 within
B, we will also treat this path as an abnormal case since there exists distinct behavior on progress starting from A1 to B1.

### B. Network Model

We use two dimension structured graph network topology such as hexagonal cell configurations to define our network model. Therefore, in our system, the network is modeled as a generalized graph \( G = (V, E) \). The vertex set \( V \) represents all the geographical areas. If two areas are adjacent to each other, there exists an edge between their two vertices.

### C. Mobility Model

Since we exclude the scenario that a node has totally random behaviors, that means, in this paper we do not consider the random walk model, in which the ad hoc node will move to any one of the adjacent cells with the equal probability after leaving a cell. In reality, this is not true. So we assume each ad hoc node normally traverse with a destination in mind. In our model, the mobility of a node can be represented by a multi-leaf tree structure. Each route starting from the second-level of the tree to the leaf can be represented by a sequence of symbols, A1, B3, C2, ... , where A, B, C, ... denotes the identity of the geographical area, and 1, 2, 3, ... denotes the location within that cell. The process to generate such a tree structure is the process so-called collecting training data.

### IV. Pattern Based Anomaly Detection

#### A. Feature Extraction

The first step in intrusion detection is to extract effective features. Features are security related measures that could be used to construct suitable detection algorithms. Effective features must be selected to reflect the subject activities. Based on this, in our environment, we build the normal profiles of ad hoc nodes with regular move patterns in wireless ad hoc networks. Under the assumption that each node will have its normal routines, locality number within the geographical area traversed by each node is an ideal candidate feature for our usage. We will also associate the distribution of mobility pattern for each sub-route. All these features we selected will be relatively stable. Therefore, a string which combines with the number and the alphabet could represent the path taken by a node. This string will feed into our model for multi-leaf tree construction.

#### B. Terminology Definition

Before we go further, some terminology definitions will be given as follows.

We classify velocity and direction to \( n \) non-overlap velocity and direction ranges, e.g., \( V_1, V_2, ..., V_n \); and \( D_1, D_2, ..., D_n \). Mobility pattern will be defined as follows:

\[
P_{ij} = V_i, D_j
\]

So, as to our classification, there are totally \( n^2 \) mobility patterns. It can be represented as a 2D table. We will address this in detail and give an example in Section V.

#### C. Multi-Leaf Tree Structure

We will use multi-leaf tree structure to model the normal profile of ad hoc nodes, see the following example (Fig. 2):

Fig. 2 is a typical profile associated with a specific ad hoc node. After some experiments, we’ve done the job of collecting training data; hence we build such a multi-leaf tree structure.

- The second-level nodes, such as the above, A1, C3, ..., G2 are the possible starting points within the corresponding geographical area; \( P(A1) \) is the probability of this node starting from A1;
- B1 and B2 are two possible children nodes whose parent node is A1, that is, staring from A1, after certain period, there are two possible subsequent distinct destinations associated with their probability denoted by \( P(B1|A1) \);
- From the level of B1 and B2 above, for each node, we attached a rectangle box which specify the probability of pattern distribution such as \( Pd_{A1→B1} \). \( \text{Pd}_n \) denotes the \( n^{th} \) pattern distribution;
- We keep going, until we get to the final destination which is the leaf node of the above graph. Each specific route staring from the second-level nodes to the leaf node is a possible complete route associated with all of its probability and pattern distribution for all of its intermediate nodes;
- All the above information can be gathered by statistical method;

Therefore, we can have the following equations and relations:

1) \( \sum_{i=1}^{n} P_i = 1; (i: \text{possible starting points}) \)
2) \( \sum_{j=1}^{m} P(\text{Pd}_j) = 1; (\text{Pd}_j: \text{possible pattern distribution} \ j \text{ associated with each node}) \)
3) For each node starting from the third level, \( \text{Pd}_i \neq \text{Pd}_j \) if \( i \neq j \)
4) \( \sum_{i=1}^{n} P(k|\triangle) = 1; (k: \text{all of the children nodes associated with the same parent node, } \triangle: \text{parent node}) \)
5) Let:
   - \( S_1 \) = the 1\textsuperscript{st} set of pattern distributions;
   - \( S_2 \) = the 2\textsuperscript{nd} set of pattern distributions;
   - \( \ldots \)
   - \( S_n \) = the \( n^{th} \) set of pattern distributions;
And all these $n$ children nodes belong to the same parent node, therefore we have the relation:

$S_1 \neq S_2 \neq \ldots \neq S_n$

**D. Anomaly Detection Algorithm**

In this section, we will give the flow chart of our anomaly detection algorithm.

**E. The Distance Measure**

In our model, we introduce a threshold, $P_{thr}$, which is a design parameter. When $\text{Distance}(S) \leq P_{thr}$, string $S$ is evaluated as normal, otherwise string $S$ is identified as anomalous.

As to the distance measure, we will use Mahalanobis Distance. The Mahalanobis distance is a very useful way of determining the “similarity” of a set of values from an “unknown” sample to a set of values measured from a collection of “known” samples.

For example, if two distributions have means $m_1$ and $m_2$, and standard deviations $s_1$ and $s_2$, the difference between two independent samples from each distribution would have mean $m_1 - m_2$ and standard deviation $\sqrt{s_1^2 + s_2^2}$ so a metric measuring the distance between the 2 distributions could be $\sqrt{(m_1 - m_2)^2/(s_1^2 + s_2^2)}$, and we will use this formula for Distance Measure in our model.

**V. SIMULATION STUDY**

**A. Data Sets**

Without losing of generality, in this paper we tested one normal profile which stands for the regular movement behavior of a specific node. It is created by extensive experiments and based on the statistical methods.

As to the velocity and direction, we have the data range assumption like this (Table. 1):

Where:

- $D_1 = [0, \Pi/3)$, $D_2 = [\Pi/3, 2\Pi/3)$, $D_3 = [2\Pi/3, \Pi)$
- $D_4 = [\Pi, 4\Pi/3)$, $D_5 = [4\Pi/3, 5\Pi/3)$, $D_6 = [5\Pi/3, 2\Pi)$

Therefore, in our simulations, there are totally 18 mobility patterns.

**Note that the mobility data set we generated is general enough for most ad hoc nodes. However, it may not be suitable for nodes with totally random movement behaviors.**

**B. Performance Metric**

We use the following two metrics to evaluate the performance of our proposed detection algorithm:

- False Alarm Rate: It is measured over normal itineraries. Suppose $m$ normal itineraries are measured, and $n$ of them are identified as abnormal, false alarm rate is defined as $n/m$.
- Detection Rate: It is measured over abnormal itineraries. Suppose $m$ abnormal itineraries are measured, and $n$ of them are detected, detection rate is defined as $n/m$.

**C. Simulation Results**

1) When $P_{thr} = 0.2$: Here we will give the simulation results in terms of false alarm rate and detection rate when $P_{thr} = 0.2$.

**TABLE I**

<table>
<thead>
<tr>
<th>Sample Mobility Pattern Definition.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_1$ (1~3)</td>
</tr>
<tr>
<td>$V_2$ (4~6)</td>
</tr>
<tr>
<td>$V_3$ (7~9)</td>
</tr>
</tbody>
</table>

**Fig. 4. False Alarm Rate. ($P_{thr} = 0.2$)**

**Fig. 5. Detection Rate. ($P_{thr} = 0.2$)**

With the above simulation results when $P_{thr} = 0.2$, we can clearly see that the detection rate is very high, it starts from 0.8 and increases with the increase of the speed range. The
reason is that the weight of high speed range is more crucial than that of low speed range, so any little change to high speed range will lead to big pattern deviation, but low speed range will not do so.

The false alarm rate is fairly low, it starts from 0.14 and it decreases with the increase of the speed range. The reason is similar to above. The possibility of pattern change with high speed range will be less than that of low speed range, hence the possibility of error with high speed range will be also less than that of low speed range.

2) When \( P_{thr} = 0.1 \): Here we will give the simulation results in terms of false alarm rate and detection rate when \( P_{thr} = 0.1 \).

![Fig. 6. False Alarm Rate. (\( P_{thr} = 0.1 \)](image)

![Fig. 7. Detection Rate. (\( P_{thr} = 0.1 \)](image)

With the above simulation results when \( P_{thr} = 0.1 \), we can see that the trend of the curve is similar with that when \( P_{thr} = 0.2 \). The only difference are those the detection rate which starts from 0.9 will be better and the false alarm rate which starts from 0.24 will be worse as to the results of Fig. 8 and Fig. 9. From this, we can conclude that the design parameter threshold is the tradeoff between false alarm rate and detection rate. If we increase the value of threshold, then we may decrease the false alarm rate but in the meantime, we decrease the performance of detection rate, and vice versa.

VI. CONCLUSION

Based on mobility pattern distribution, this paper presents a novel approach to construct the mobility profile of ad hoc nodes in wireless ad hoc network. Each node’s itinerary is modeled as a multi-leaf tree structure. An intrusion detection algorithm is then developed to detect potential internal attackers - masquerades. Simulation results demonstrate that our approach can achieve desirable performance in terms of false alarm rate and detection rate for nodes having normal mobility patterns.

Our future research work will focus on the following three facets: 1) Our model right now can only apply to the ad hoc nodes which have normal mobility patterns; it can not fit into the ad hoc nodes which have totally random mobility behaviors. More features such as system call history may be accommodated into the system to make it suitable to all scenarios; 2) although the profile of ad hoc nodes will not change often than that of other wireless networks such as cellular mobile networks, suitable mechanisms should also be deployed to keep the profile up-to-date; 3) in this paper, we treat threshold as a design parameter. For a more reasonable value of threshold, suitable mechanisms should be deployed to establish the corresponding connection between the mobility level and the threshold.

REFERENCES


