Feature Analysis for Intrusion Detection in Mobile Ad-hoc Networks

Vetrichelvi[1] Rajaram[2] Vanitha[3] Dr.G.Mohankumar[4]

[2]Department of ECE, Park College of Engineering and Technology, Coimbatore, TamilNadu(TN), 641-659, India

Abstract
As Mobile Ad-hoc network (MANET) has become a very important technology, research concerning its security problem, especially, in intrusion detection has attracted many researchers. Feature selection methodology plays a central role in the data analysis process. The proposed features are tested in network operating conditions. PCA is used to analyze the selected features. This is because, redundant and irrelevant features often reduce performance of the detection system. STR belongs to hierarchical routing protocol and does not attempt to consistently maintain routing information in every node. Furthermore, through the use of tree's intrinsic routing function, the STR protocol exhibits hybrid behavior and better performance of normalized routing load in large, mobile, ad hoc networks

Keywords:
Feature selection, intrusion detection, MANET, PCA, STR

1 Introduction

Mobile Ad-hoc Network (MANET) is an unstructured wireless network that can be established temporarily, Each node is selfish and independent in the decision making. In MANET, nodes can add-in to the network or detach from it at any time. Thus, there is no central control on the network for the nodes to follow . Intrusion detection models were introduced by Denning in 1987 and rather are a new technology [5,].

Intrusion detection systems can be categorized into two models: Signature-based intrusion detection [2] and anomaly-based intrusion detection. Signature-based in-trusion detection uses signatures of the attacks to detect the intrusion. This type of detection monitors the net-work forwarding a match between the network traffic and a known attack pattern. On the other hand, anomaly-based is performed by learning the normal behavior of the network and comparing it versus the behavior of the monitored network. The advantage of the anomaly-based detection is its ability to detect new attacks without any prior knowledge about it [5].

This paper purposes a neighbor monitoring intrusion detection based on the traffic of the node, where feature selection is used to improve its performance. The proposed approach uses the anomaly-based intrusion de-tection method. In Ad-hoc networks, packets that are sent from each node can be used for network condition monitoring. Using the traffic data, behavior of the node's neighbor can be monitored. In the reported work, 16 fea-tures in the network traffic are monitored. This paper, in-tends to show the difference between the normal operating state of a network and the operating state of the network once it experiences a DoS attack. The reported work intends to reduce the dimensionality of the network features. This reduction may lead to increase in intrusion detection speed, since the IDS would have fewer features to analyze. Network features such as movement and number of the nodes are also considered in the reported work. This paper is organized in the following way: Section 2 presents solutions for intrusion detection systems (IDS)in mobile Ad-hoc networks. Section 3 describes the tech-niques and protocols used in this work.

These techniques include PCA and based IDS. Dynamic source routing (DSR) is the protocol used in the proposed scheme.

2 Related Works

The solution for IDS in MANET was proposed by Zhang et al. [26]. In their paper, they proposed two important solutions, i.e. anomaly-based and signature-based detection. The work detects attacks using anomaly-based intrusion detection. Implementing this kind of detection, moving speed of the nodes, their distance, rate for the rout change, hop counter parameters were used. In signature-based intrusion detection, a pre-prepared rule is used to detect an attack.

In a work reported by Hu et al., an approach based on digital signatures was used to detect rushing and worm-hole attacks [11, 12].

In a reported work by Gilham et al., a rule-based in-trusion detection system named IDES is introduced [19]. IDES learns users' behavior and uses misuse detection approach. Alerts are generated once a suspicious activity that deviates significantly from the established normal usage proposed is detected.

In a work reported by Kim et al., they have developed a real-time intrusion detection system which combines on-line feature extraction method with least squares support vector machine classier. They have used DARPA99

Manuscript received September 5, 2010
Manuscript revised September 20, 2010

(KDD 99) dataset for the experiments and there is no simulation environment used in this work. Thus this method can detect new attacks. In a work reported by Richeldi et al.a genetic algorithm is proposed to select efficecetive features. This method is very slow. Another feature selection is proposed by Chen et al., in which they utilize three feature selection algorithms. They use an SVM classifier [4] and two multi-labels [3]. In a work reported by Wang et al., Markov Blanket algorithm is applied on the feature selection part of an intrusion detection method. In this approach, Markov Blanket algorithm can decrease the number of features.

3. The Proposed Techniques and Protocols

There are many techniques that can be used for monitor-ing the nodes and analyzing the results. In the proposed method of approach, a proposed-based monitoring technique and the Principal Component Analysis (PCA) technique are implemented. Network performance is dependent on the selected routing protocol. For the same reason, the DSR protocol is used in this work

3.1 Proposed based Neighbor Monitoring IDS

This paper proposes a proposed based intrusion detection system for wireless Ad-hoc networks.
In the pro-posed IDS, each node builds a proposed for every one of its neighbors . The proposed includes all features listed in Table 1. The data packet size indicates the packet type. They are all traffic related features [10].
A node can use a proposed by keeping it to monitor its neighbor node's behavior. This paper simulates this technique on a number of selected features in a simulation environment.

3.2 Principal Component Analysis

PCA is used to analyze results of the scenario-based Ad-hoc network simulations [10]. Simulation output is in comma-separated vector (CSV) format.
PCA is a classic technique in statistical data analysis, feature extraction and data compression. Goal is a smaller set of variables in a set of multivariate measurements with less redundancy.
The starting point for PCA is a random vector x with n elements. There are available samples x(1) : : : x(T) from this random vector. No explicit assumptions on the probability density of the vectors are made in PCA, as long as and the second-order statistics are known or can be estimated from the sample [6].

3.3 DSR Protocol

DSR is an active routing protocol that is implemented based on source routing . The header of the packet has a list of nodes addresses to pass it in source routing. source route discovers the path to the source node. Using this method prevents the route to follow a cyclic path. Middle nodes of this routing do not need to collect latest nodes status such as sequence number unlike AODV. All nodes can listen to the packets in the DSR routing network. Nodes can update the routing infor-mation in cache table based-on available paths in packet header for further usages. This routing protocol does not need to use HELLO packets . This paper intends to use this protocol because of the aspects such as Nodes ability to sni® packets in the network [8].

Table 1: Networks feature

Features	1. My address 2. Destination address 3. Route REQuest (RREQ) from node I 4. Route REPly (RREP) from node I 5. Route error from node I 6. Total packet received from node I 7. My received sent packet 8. ACK packet from node I 9. Traffic sent from node I 10. Total received RREQ, 11. Total RREP 12. Total received (Route Request Error) RRER 13. Total Traffic received, 14. Total ACK received, 15. Timestamp 16. DSR header

4. Implementation

This paper implements the networks to monitor the fea-tures and evaluate the selected feature and analyze net-works. There are more than 80 scenarios that show nor-mal and attack networks with parameters as described in Table 2. Since these 80 scenarios have similarities in their behavior, only a selected number of them are re-ported in this paper. For example, in a scenario with 20 nodes, voice over IP (VOIP-PCM) traffic, fast movement is simulated. An experimental network is implemented for experimenting with different scenarios. Networks test run was for 180 seconds. Radio signal radius is 250 meters for all nodes. The implementation area is 2000m * 2000m in size. The network topology is WLAN (infrastructure). All the scenarios were implemented in the simulation en-vironment.

4.1 Data Collection

This paper intends to answer the following three ques-

tions. Question number one: why features listed in Table 1 are good candidates? Question number two: is this feature a proper one? Or how it can be evaluated? Question number three: which feature is appropriate?
This section will explain why the selected features are the right candidates. Selected features are described here.

1) My address:

   It shows the source addresses. This feature is applied to specify the intruder or the misbehaved node.

2) Destination address:

   Destination address is selected to specify nodes that are under attack.

3) RREQ from node I:

   Some Denial of Service (DoS) attack methods attack other nodes by sending a lot of route request. Therefore, route request feature is selected to specify how many Route Requests is sent.

4) My sent packet:

   This feature shows selfish behavior of the nodes. Assume a node that wants to send a packet to destina-tion B, but there is no straight path from the node A to B. Therefore, it sends a Route Request to its neigh-bors. Let node X receives the packet and forwards it to other nodes. Since the node A is a neighbor of the node X, the forwarded packet is sent to node A as well. This feature specifies neighbor behavior. This is needed because neighboring node may not forward the packet. Other features are also selected to specify neighbor's behavior. Nodes monitor the whole received traffic. This traffic. is received from neighboring nodes. Each node monitors these features from the network traffic.

4.2 Scenarios

In this section, three scenarios are presented. This work uses a pro-le-based feature selection.

4.2.1 Scenario 1

A normal state of operation for a network is a state that presents the normal daily operation of the network once it is not under any kind of attacks. Different types of net-work traffic are generated in network. Node movements in some networks are made to be fast or slow. Nodes are made static or dynamic. In this network, distances between nodes are variable.

A sample network may include 20 nodes with VOIP (GSM) traffic. The nodes in this network are pervaded in 2000*2000 square meter area and these nodes can move as fast as 200 meters per second. Distant nodes are very few, and nodes are within the transmission radius of several nodes. Other networks are implemented using same parameters as in Table 2.
Network operation is tested and simulated in various operating conditions and its operating traffic. was stored .

Table 2: Network parameters

Number of Nodes	20-50
Traffic type	Voip (gsm)-IMAP video conference high quality-video conference (low quality)
Nodes movement	Random way point (fast-static-low)
Nodes distances (density)	50 Far-close

4.2.2 DoS Attack

DoS attack is divided into two categories. In the one intruder attacks other nodes in the Ad-hoc net-work services and does not let them to provide their service Most of the network resources are wasted trying to generate routes to the unknown destinations or routes that are not going to be used for any communication. This implies that the existing version of DSR is vulnerable to such type of malicious behavior from an internal node (which is named a compromised node

4.2.4 Scenario 3

Another kind of DoS attack is simulated in this scenario. It changes intruder's path manager parameters to send a large number of RREDs. This is implemented for two reasons. First reason is that the invisible cache is destroyed quickly. Thus, this node sends one RREQ for each data packet. Second reason is because the intruder tries to establish a random short time connection with other Ad-hoc nodes. This method of attack is tested in previous scenarios

5 Analysis

Analysis for PCA outputs is presented in this section. The scenarios in Section 4.2 are simulated in various conditions (relevant to Table 2). Important features for various normal and abnormal states in Ad-hoc networks are described in the following sections. It is also shown that some features in normal state of operation are universal features and their values only experience small changes once the state of the operation for the network changes.

5.1 Experiment 1

There are 20 networks tested in this experiment. Since all the 20 networks have shown similar behaviors during the simulations, only 6 selected experiments are reported in this paper (Table 3). Figure 1 shows that the 4th and the 11th feature, where the latter one is the most important one and has the max-imum variance. This feature holds maximum information value among all other features in the network. Features are tested in 20 network scenarios.

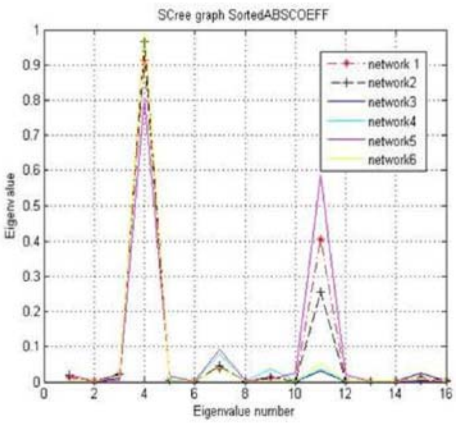

Figure 1: Network's six normal states of operation are presented with different colors

When network operates in a normal state, nodes send RREQ to the other nodes with a rate. But in different operating conditions it shows a low variation in its value. 14 As for example, RREP sent from all adjacent nodes, have higher variations with respect to the times-tamp (time).

6. Simulation Model

We used NS2 (Network Simulation 2) to evaluate the performance of *STR*. The simulation modeled a net-work in a 2500m×2500m area with 50 mobile nodes. Radio transmission range is 250 meters. The mobility of each node is arranged from 2m/s to 8m/s, and the pause time of the mobile nodes is zero. Traffic sources are continuous bit rate (CBR) with the rate of 15kbit/s. The source-destination pairs are randomly selected over the network. Four important performancemetrics. are evaluated: (1) Normalized routing load-the number of routing control packets transmitted per data delivered at the destination. Each hop-wise transmission of a routing control packet is counted as one transmission. (2) Route discovery delays-the delay between a route requests being issued and a reply with a valid route being received. (3) Route discovery load-the route discovery packets being used to find a valid route to the destination. (4) Packet delivery ratio-the ratio between the number of received data packets and those originated by the sources. Figure shows the packet delivery ratio of *STR* under various offered load with different speeds. Packet delivery ratio declines both with speeds and offered load. We note that at low speeds, packet delivery ratio is sensitive with offered load. Packer delivery ratio declines while offered load is increasing. This is because most data packets are sent to their father node before they go to the destination in STR. And it will probably lead to traffic congestion in some father nodes and cause packet loss. As mobility increase, the packet delivery ratio is less sensitive with offered load. Since most packet loss ascribe to the fiercely change of network topology.

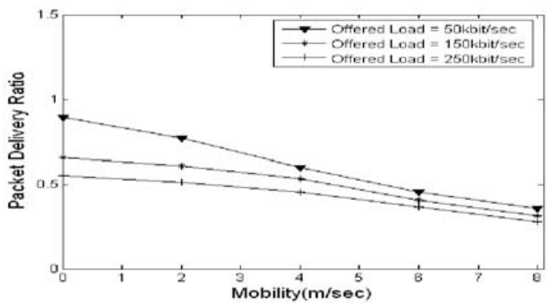

Figure 2

7. Conclusions

In this paper, based neighbor monitoring intrusion detection approach in MANET was presented. This approach is based on feature selection method and it ap-plies PCA theory to determine network operating conditions. Best network parameters (features) to identify nor- mal state of network operation are presented. 16 features are selected for test in these networks. PCA shows that it is not necessary to monitor all the features to identify the operating condition of the networks. *Sub area Tree Routing* (*STR*), a novel routing protocol for multi-hop wireless ad hoc networks based on the idea of establishing sub area trees and dividing the whole network into many logical subareas, has been proposed. This pro-tocol constructs a hierarchical network structure with two tiers in which different routing strategy is adopted, and routing mechanism combines the advantages of proactive routing and on-demand routing. maintaining routing in-formation, especially in large, mobile, adhoc environment.

References

[1] A. Agah, and S. K. Das, Preventing DoS attacks in wireless sensor networks: A repeated game theory approach," International Journal of Network Security, vol. 5, no. 2, pp. 145-153, 2007.

[2] F. Anjum, D. Subhadrabandhu and S. Sarkar. Signature-based intrusion detection for wireless Ad-hoc networks," Proceedings of Vehicular Technology Conference, vol. 3, pp. 2152-2156, USA, Oct. 2003.

[3] W. Chen, J. Yan, B. Zhang, Z. Chen, and Q. Yangm Document transformation for multi-label feature selection in text categorization," Proceedings of Seventh IEEE International Conference on Data Mining, pp. 451-456, USA, 2007.

[4] H. Deng, Q. A. Zeng, and D. P. Agrawal, \SVM based intrusion detection system for wireless ad hoc networks," Proceedings of the IEEE Vehicular Technology Conference, pp. 2147-2151, USA, 2003.

[5] D. E. Denning, An Intrusion Detection Model," IEEE Transactions in Software Engineering, vol. 13, no. 2, pp. 222-232, USA, 1987.

[6] H. A. Edelstein, Introduction to Data Mining and Knowledge Discovery, Crows Corporation, Third Edition, 1999.

[7] D. M. Farid, and M. Z. Rahman, Learning intrusion detection based on adaptive Bayesian algorithm," 11th International Conference on Computer and Information Technology (ICCIT2008), pp. 652-656, 2008.

[8] A. HyÄvarinen, J. Karhunen, and E. Oja, Independent Component Analysis, John Wiley & Sons Inc., USA, 2001.

[9] C. E. Perkins, “Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers,” Proceedings of ACM SIGCOMM, pp. 234-244, 1994.

[10] G. Pei, M. Gerla, and T. W. Chen, “Fisheye state routing in mobile ad hoc networks,” in Proceedings of the 2000 ICDCS Workshops, Taipei, Taiwan, pp. D71-D78, April 2000.