CROSS-LAYER FEATURES BASED INTRUSION DETECTION SYSTEM FOR WIRELESS AD HOC NETWORK

MR.Dipak K .Patani*

MR.Chetan P .Undhad*

Prof.Chetan Singhadiya**

Abstract

A Mobile Ad hoc Network (MANET) is a network of mobile nodes which dynamically grouped to gather and establish arbitrary and temporary network topology. Ad hoc network is vulnerable to many kind of attack because of infrastructure less architecture. Cross layer based intrusion detection system (IDS) for wireless ad hoc networks using association rule mining and classification is our main focus in this paper. Specifically, features of MAC layer and network layer to profile normal behaviours of mobile nodes are used. The proposed CIDS is able to effectively detect an attack and is able to localize the attack source. False positive rate is reduced through the module 2 of the CIDS where intelligence gathered from neighbour nodes is used to make a collaborative decision by the monitor node. Our proposed solution will lead new track and work in the field of CIDS and eliminating other network attack like jelly fish.

Keywords: *Association rule mining, CIDS, classification, jelly fish attack, MANET.*

----------***----------

* P.G. Student, Computer Department, NOBLE Collage of Engineering, Junagadh.

** Assistant Professor, Computer Department, NOBLE Collage of Engineering, Junagadh.

1. MOTIVATION

An ad hoc network is a group of nodes connected gather by wireless links. Associations between nodes are established when they are in the vicinity of each other. All mobile nodes agree to relay each other's packets, and function as routers. While self-organizing nature of ad hoc networks provides convenient method for communication among mobile nodes. Main problem in ad hoc network is the lack of central authority which will restrict any node from doing misbehaviour by defining the privilege to individuals or which can monitor the traffic so it is very hard to detect the attack.

Attack on Ad hoc network may affect the wired network also. It is Because of hijacking of the legitimate node (station) by the malicious node by forming the ad-hoc network with that legitimate node.

Attack in ad hoc network can be applied on each individual layer of the network protocol stack. Because of lack of infrastructure the IDS used in wired network can not be used in wireless ad hoc network. Also the resource constrained environment may create problem for development of IDS. In this paper, we propose to use a rule-based data mining and classification techniques for anomaly detection to detect attacks on ad hoc networks with reduced feature set. Anomaly detection techniques are usually prone to high false positive alarm rates.

Our approach is to specify a reduced feature set across the MAC layer and the network layer to profile normal user behaviours. The proposed method aims at easing the complexity of the proposed IDS, and extending its detection ability to both layers.

2. RELATED WORK

In paper [1] author has used classification based technique for routing anomaly detection in the Ad hoc network. Their work was followed by the classification based approach on certain features at both MAC and Network layer. Authors have proposed new approach for the classification based on average probability(AP).Association rule based IDS[2] has been proposed for detecting the misbehaviour of the node. Authors have used Bayesian network to improve performance of the system. In paper [3], monitor based packet drop detection using cross layer

detection is given. In this technique mobile node (who is in the vicinity of the attacker) with sufficient energy will work as a monitor and detect the attacker by sending packet and detecting the reply from that malicious node. Author of paper [4] has given a Support vector machine (SVM) based method for sinking behaviour identification using cross layer features.

Our focus will be on anomaly detection with data mining based approach. Few methods based on data mining of database have been already proposed by researchers in [1][2]. Association rule mining is the process of capturing rules from given data based on support threshold and confidence threshold for selected rules with respect to minimum support and minimum confidence.

❖ **Classification of IDS**

Most of the IDS suggested previously by the researcher in [1]-[7] can be classified into below four category.

1) Agent based intrusion detection system.
2) Group based intrusion detection system.
3) Cluster based intrusion detection system.
4) Cross layer Intrusion detection system.

Because of changing topology and movement of the node some time it is very hard to relies on a single-layer detection method because there is not enough evidence using single layer detection. As a result, the concept of multi-layer or cross-layer detection mechanism is raised and discussed in [3] and [4]. IDS proposed previously have architecture consist of four modules: data gathering, profile generation, Anomaly detection and decision tacking system.

***Data gathering*:** this module collects audit data (network activities) from a given network in normal condition within its observable radio transmission range.

***Profile generation*:** this module has two subsystems:

1) Data preparation: here the collected data are prepared for creating normal behaviour profile. Processes like filtering, aggregation, data suppression are applied here.

2) Profiler (Profile generator), the second phase is made up of several techniques like clustering, classification rule mining or SVM where normal profile is made by the pre processed data. A normal profile is an aggregated rule set of multiple training data segments.

***Anomaly Detection*:** This phase detect anomaly in the network with the help of derived rule set in this module, test data profiles are compared with the expected normal profiles. Any rules with deviations beyond a threshold interval are considered as anomalies. Suppose some rule generated from test data was not previously available in normal profile then it will be detected as anomaly, it is considered as an anomaly rule; if the rule is in the rule set, but its support and confidence level is beyond the interval [*minimum − threshold, maximum+ threshold*], the pattern described by the rule becomes unusual, and is consider as an anomaly rule [2].

Decision tacking system: when any anomaly rule trigger that will be attended locally as well as globally by giving alert to the neighbours when the support and confidence of anomaly rule goes above tolerated level. Here are some attack those are possible at different layer.

Layer	**Type of Attack**
Physical	Jamming, Tampering
Data link	Collision, Exhaustion, Unfairness, Jamming
Network & routing	Neglect and Greed, Homing, Misdirection, black hole, packet drop.
Transport	lFlooding, Desynchronization

Table 1. Classification of attack at different layer. [5]

3. PROJECT BODY

3.1. Problem statement:

In the mobile ad hoc network nodes can directly communicate with all the other nodes within their radio ranges; whereas nodes that not in the direct communication range use intermediate node(s) to communicate with each other. Here are some basic features of MANET which cause threat to the security of the MANET.

Unreliability of wireless links between nodes: Because of the limited energy supply for the wireless nodes and the mobility of the nodes, the wireless links between mobile nodes in the ad hoc network are not consistent for the communication participants. And any malicious node can communicate with the node which is in the in the vicinity of that node

Non uniform topology: Due to change in the position of mobile nodes, the routing information will be changing all the time.

Lack of power supply: because of energy constrained environment in ad hoc network the node will be rely on battery power. Any malicious node can easily attack the node by attacking the power of the node by making node busy in bogus communication and some time leads the victim node to the off state. This cause DOS (denial of service) type of attack.

Because of above listed futures ad hoc network is vulnerable to some serious type of attacks there for we need to pay more attention to the security issues in the mobile ad hoc networks. **So our main objectives are:**

How to prevent and detect malicious activity of the node in the ad hoc network?

How to increase the throughput of the system?

How to reduce overall delay in routing mechanism?

3.2. Evaluation metrics

The evaluation metrics consist of following things.

Throughput	The overall output of the system that goes down because of malicious activity of node
Delay in routing	Total delay produced in packet or data delivery because of malicious node.
Overhead	Extra work that nodes have to compute for routing because of malicious activity of the node.

3.3. Feature of interest

There are many type of features available at both MAC as well as network layer. According o the need of efficiency and effectiveness different feature set can be taken.

At network layer IP packets can be categorised in to two type 1) control packets (i.e. Route Request, Route Reply, Route Error) and 2) data packets. We combine all routing control packets into one category as routing Control packet (CtrlPkt) , and name IP data packet as routing data packet (DataPkt). Thus the payload in a MAC data frame contains either a CtrlPkt or DataPkt. In summary, we present our proposed feature set and its value space as below. We are also interested in packet delivery ratio and delay to the packet because of queue buffering.

Table 1. The proposed feature set

Feature		Value Space
Direction of the packet	:	SEND, RECV, DROP
Source address (SA)	:	*sai*, $\forall i \in$ node set *S*
Destination address (DA)	:	*daj* , $\forall j \in$ node set *S*
Type of data	:	RTS, CTS, DATA, ACK,
Packet type	:	DataPkt , CtrlPkt
Delay	:	in form of second
Package drop	:	percentage of packet drop
Route related features	:	route add - remove, total route change, avg route length

3.4. Proposed architecture.

Our work follows existing data mining techniques like association rule mining and classification for detecting anomaly. Some intrusion detection techniques suggested in literature use probabilistic analysis where the resulting models are not straightforward to be re-evaluated by human experts [2]. Some data mining models require temporal sequence from data stream, which is domain specific and highly inefficient when a large feature set is involved [2]. Because of large feature set data availability of MANET we can use different combination of feature to correlate them.

In given paper, a cross feature based anomaly detection algorithm is given in phase 2. Formally speaking, in the cross-feature analysis approach, aim is to establish correlations between each

feature and all other features. i.e., try to solve the classification problem $\{f1, f2, \ldots, fi-1, fi+1, \ldots, fL\} \rightarrow fi$ where $\{f1, f2, \ldots, fL\}$ is the feature vector [1]. Note that in the machine learning area, the terminology *class* in a classification system represents the task to be learned based on a set of features, and the *class label*s are all possible values a class can take [1]. Thus, the anomaly detection problem can be transformed into a set of classification sub-problems, where each sub-problem chooses a different feature as a new class label and all other features from the original problem are used as the new set of features. The outputs of each classifier are then combined to provide an anomaly detector. The new model of CIDS can be used as an effectively to reduce anomaly.

Our main object is to identify the normal and abnormal profile with selected feature set and detect the anomaly in the network. It consists of two phases CIDS Phase -1, CIDS Phase-2.

CIDS Phase -1

CIDS phase -1 works for collection of data and use it as a training data set to generate normal behaviour. By applying the association rule mining we will mine the packet level event, which contains <*Timestamp, Dir, SA, DA, PktType*>. An example association rule is *(sa5, DataPkt → da12, RECV), (0.4, 1),* which describes an event pattern related to the *RECV* flows of the monitoring node. Here the support of rule is 0.4(40%). That is , 20% of transaction records matches the event of "node 5 sends data packets to node 12", and confidence 100% suggest that when node 16 receives data packets, they are 100% of the time from node 7. and then prune the rules with MFI criteria. MFI is defined as a frequent item set for which none of its immediate supersets are frequent [2]. This pruning process dramatically reduces the size of normal profile, yet still captures the frequent association patterns from a data set. In our experiments, the MFI pruning can reduce the number of association rules by 20 to 40%. Once association rules are extracted from multiple segments of a training data set, they are then aggregated into a rule set, which is considered as a normal profile. After extracting the Rule set the process continuously train the rules to certain amount of time. This process gives the average rule set values (support and confidence) at the normal behaviour (normal profile) of network without attack. when the attach is there in the network all the nodes which has higher threshold value for packet sending rate then avg(threshold) will be kept in non-trusted region.

We also train a classification model for CIDS phase-1, Here we are defining classes $Ci : \{f1, \ldots, fi-1, fi+1, \ldots, fL\} \rightarrow \{fi\}$. For normal events, the prediction by Ci is very likely to be the same as the true value of the feature; however, for anomalies, this prediction is likely to be different. Because Ci is trained from normal data, and their feature distribution and pattern are assumed to be different from those of anomalies. This implies that when normal vectors are tested against Ci, it has a higher probability for the true and predicted values of fi to match. Such probability is significantly lower for abnormal vectors. With the help of degree matching we can distinguish between normal and abnormal behaviour. We name the model defined above a **sub-model with respect to** fi [1]. Obviously, relying on one sub-model with respect to one labelled feature is insufficient. Therefore the model building process is repeated for every feature and up to L sub-models are trained [1].

Data: feature vectors of training data $\boldsymbol{f1, \ldots, fL}$.

Result: classifiers $\boldsymbol{C1, \ldots, CL}$.

begin

$\forall i$, train $\boldsymbol{Ci : \{ f1, \ldots, fi-1, fi+1, \ldots, fL\} \rightarrow fi;}$

return $\boldsymbol{C1, \ldots, CL;}$

end

Where $Ci(x)$ is the predicted value from sub-model with respect to fi.

Algorithm 1: Cross-Feature Analysis: Training Procedure [1].

By analysing the flow of network we can predict that the network is under some malicious activity or not with the help of normal traffic rate observed in association rules. When there is abnormal traffic at some point those nodes will be considered as possibly malicious and kept in non-trusted region. Now the monitor node will execute the CIDS module-2, and try to detect whether the node in non trusted region is exactly malicious or not. For that the classification technique of data mining is used. Monitor node is selected on the cluster based technique. Group of nodes which are within certain vicinity of victim node will form a cluster among themselves. Where one node will work as cluster head and that node will execute CIDS module 2. Selection of the cluster head will be based on the energy level of the cluster node. Some node might be

selfish and will refuse to work as monitor. Here we assume that each node with sufficient energy will be having equal chance to work as cluster head and elected randomly.

CIDS phase-2

An event is classified as anomaly if and only if the *average probability is* below the threshold. Assume that $p(fi(x)|x)$ is the estimated probability for the true class of the labelled feature average probability is the average output value of probabilities associated with true classes over all classifiers. The optimized version is shown in Algorithm given below.

Data: classifiers $C1, \ldots, CL$, event $x = (f1, \ldots, fL)$, decision threshold θ;
Result: either normal or anomaly;

begin

AvgProbability $\leftarrow i\ p(fi(x)|x)/L$;

if *AvgProbability* $\geq \theta$ **then** return "normal";

else return "anomaly";

end

Algorithm 2. Verification of malicious node. [1]

3.5. Methodology.

We are trying to show the effect of malicious activity on through put of network. We have implemented the jelly fish attack with ns-2 and try to check the trough put which goes down due to malicious activity of the nodes. The jelly fish attack can be implemented by creating malicious node and increasing the delay of packet forwarding or by dropping the packet from the malicious node.

Table 2. Parameters for the simulation are as follows.

Area	2000m X 2000m
Nodes	50
Packet size	512

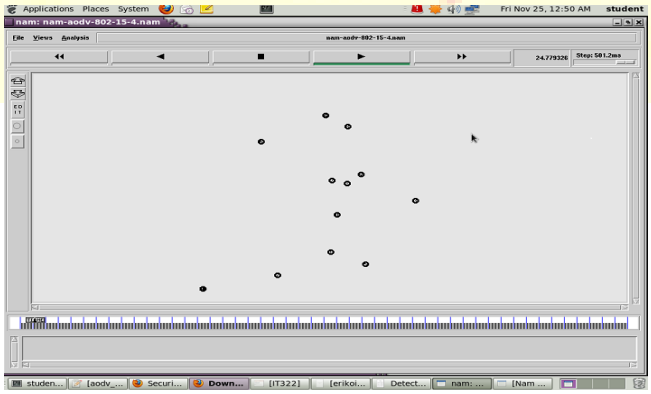
Transmission protocol	UDR
Application traffic	CBR
Transmission rate	10 Mbps
Pause time	24.73
Maximum speed	31 sec
Propagate model	Radio propagation
Maximum malicious node	50
Type of attack	Delayed forwarding
Examination	AODV

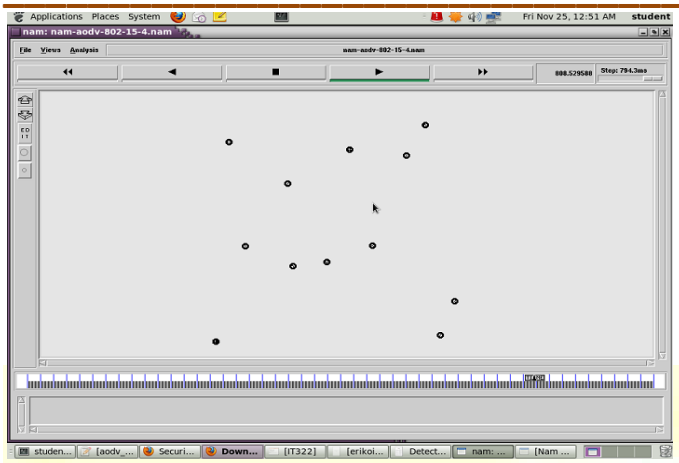
3.6. Expected results of IDS:

Proposed algorithm is the mixture of two analysis techniques association rule mining and classification. All threat will be detected in the first phase of the CIDS phase-1 where network flow is tested. With the help of CIDS phase -2, we are verifying the malicious activity of the targeted node. So the overall effect of malicious node is taken in to consideration by analysing traffic at network and MAC layer which will result into greater probably of Intrusion Detection.

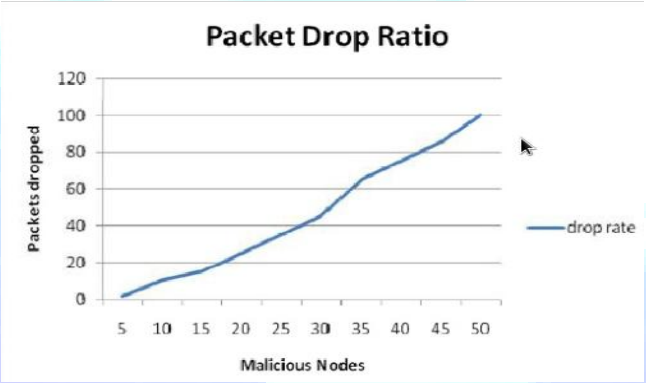
4. RESULTS AND ANALYSIS:

Screen shot of 12 nodes randomly moving the screen in random direction and getting communicated with each other when the come at the vicinity of each other.





Screen shot of 12 nodes randomly moving at time 800ms.



It is seen from the graph that when malicious node is introduced in the network the overall throughput will goes down by the malicious activity of node because packet drop ration increase with increase of malicious activity.

5. CONCLUSION

In this paper we have presented a cross-feature based anomaly IDS for ad hoc networks using unsupervised association rule mining and classification technique. Here we have tried to reduce the false alarm rate by using classification technique, and also try to reduce the number of redundant alerts.

6. REFERENCES:

1. Y. Huang, W. Fan, W. Lee, "Cross-Feature Analysis for Detecting Ad-Hoc Routing Anomalies," *In proc of.* 23rd International Conference on Distributed Computing Systems, 2003.

2. Y. liu, Y. Li,H. Man, " Short Paper: A Distributed Cross-Layer Intrusion Detection System for Ad Hoc Networks", *In proc of.* The First International Conference on Security and Privacy for Emerging Areas in Communications Networks,2005.

3. G. Li,J. He, Y. Fu, " A Group-based Intrusion Detection Scheme in Wireless Sensor Networks" The 3rd International Conference on Grid and Pervasive Computing – Workshops, IEEE, page no -286-291,2008.

4. J. John, B. Lee,A. Das, "Cross-Layer Detection of Sinking Behavior in Wireless Ad Hoc Networks Using SVM and FD", IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 8, NO. 2,pg. no 231-245, MARCH-APRIL 2011.

5. S.Bose and A.Kannan ,"Detecting Denial of Service Attacks using Cross Layer based Intrusion Detection System in Wireless Ad Hoc Networks", IEEE-International Conference on Signal processing, Communications and Networking,page no.82-188, 2008.

6. G. Helmer, J. Wong, V. Honavar, L. Miller and Y. Wang, "Lightweight Agents for Intrusion Detection." *Journal of Systems and Software*, Elsevier. Vol. 67. No. 1. pp. 109-122, 2003.

7. G. Thamilarasu, A. Balasubramanian2, S. Mishra2 and R. Sridhar, "A Cross-layer based Intrusion Detection Approach for Wireless Ad hoc Networks", MASS 2005 Workshop - WSNS05,IEEE,2005.

8. H. Yang, H Luo, F. Ye; S. Lu, L. Zhang, "Security in mobile ad hoc networks: challenges and solutions," IEEE Wireless Communications, Volume: 11 Issue: 1, page no. 38-47, Feb 2004.