http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 7 (1): 143-148, 2008
ISSN 1812-5638

Wireless Node Misbehavior Detection Using Genetic Algorithm

[1]P.C. Kishore Raja, [2]M. Suganthi and [1]R. Sunder
[1]Department of Electronics and Communication Engineering,
Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, Tamilnadu 602105, India
[2]Department of Electronics and Communication Engineering,
Thiagarajar College of Engineering, Madurai 625015, India

Abstract: This study presents behavior-based wireless network intrusion detection using genetic algorithm that assumes misbehavior identification by observing a deviation from normal or expected behavior of wireless node. The feature set is constructed from MAC layer to profile the normal behavior of wireless node. The wireless node behavior is learned by using genetic algorithm and current wireless node behavior can be predicted by genetic algorithm based on the past behavior. A 3-tuple value i.e., entropy index, newness index, mismatch index is calculated for constructed feature set in a session. The 3-tuple value of a wireless node behavior in a session are compared with expected non-intrusive behavior 3-tuple value to find intrusions. The performance of wireless intrusion detection is evaluated using detection probability and false alarm probability.

Key words: Wireless network, intrusion detection, genetic algorithm

INTRODUCTION

Wireless networking applications continue to proliferate at an incredible pace as wireless features, functions, security and throughput improve. At the same time, vulnerability of wireless networks keeps with technology. In a wireless network, one cannot make the assumption that wireless nodes are trusted. In infrastructure network, wireless nodes associate themselves with an access point, which is connected to wire line network that solves centralized network management function. In case of ad-hoc networks, network does not have a centralized network management function. All leads to increase in vulnerability that ranges from passive eavesdropping to active interfering. There is a need of security measures for wireless networks. One of security measure is intrusion detection.

In this study, we propose to use a behavior based intrusion detection technique using genetic algorithm to detect intrusions on wireless ad hoc networks. It is in contrast to signature based intrusion detection techniques, which may be impractical for ad hoc networks due to difficulties of specifying, distributing and updating signatures of attacks. Another challenge to behavior based intrusion detection in ad hoc networks is resource constraints. Our approach is to specify a reduced feature set of MAC layer to profile normal wireless node behaviors.

RELATED WORK

Most of current work (Liuy *et al*., 2005) on IDS for wireless networks employ either distributed and cooperative architecture or distributed and hierarchical architecture. Zhang and Lee (2000) proposed the first distributed and cooperative anomaly based IDS framework. In this framework, local anomaly detection engine is built on a rule based classification algorithm RIPPER and local response is activated when a node locally detects a anomaly or intrusion with high confidence. When a node detects an anomaly or intrusion with weak confidence, if then initiates a global intrusion detection procedure through a cooperative detection engine. Huang and Lee (2003) extended their previous work on local anomaly detection and developed a cross feature analysis technique to explore the correlations between features using classification decision tree induction algorithm C4.5. Their detection engine uses features extracted from routing table and also they incorporated statistical features. However, system is unable to localize the attack.

Tseng *et al*. (2003) developed distributed IDS using specification based detection techniques to detect on attacks on AODV routing protocol. Generally specification based detection a technique of any kind has to balance trade off between model complexity and accuracy.

Corresponding Author: P.C. Kishore Raja, Department of Electronics and Communication Engineering,
Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, Tamilnadu 602105, India
Tel: 91-44-27162321

FEATURE OF INTEREST

In wireless networks, MAC layer manages and maintains communication between mobile nodes by coordinating access to a shared radio channel and utilizing protocols that enhance communications over a wireless medium. The proactive mechanisms are employed in wireless networks before any data communication. These mechanisms cannot give prefect prevention. This work concentrates on reactive mechanism, which detects intrusion or anomaly behavior in wireless networks. This works assumes that wireless networks use both CSMA/CD and CSMA/CA. We extract feature set from MAC layer (Liuy *et al*., 2005) to characterize wireless node behavior in wireless ad hoc network. Table 1 describes extracted wireless feature set from MAC.

Network Allocation Vector (NAV) is a counter resides at each node that represents the amount of time that channel will be occupied by the current sending node. RTS/CTS/DATA/ACK are four features defined over these four types of packets. The transmission and reception traffic rate (transmit-traffic-rate and receive-traffic-rate) are busy/idle indicators. The retransmission rates of RTS (Retransmit RTS) packets and DATA (Retransmit DATA) packets are channel congestion indicators. Active Neighbor node count indicates the number of active neighbor nodes of the monitoring node. The remaining extracted features from network layer are shown in Table 1.

Table 1: Statistical wireless feature set and its values

Feature	Unit	Range
Nav	Second	[0,1]
		[1,3]
		[3,5]
		[5,∞]
Transmit-traffic-rate	Byte	[0,102.4k]
		[102.4k,204.8k]
		[204.8k,3027.2k]
		[3027.2k,∞]
Receive-traffic-rate	Byte	[0,102.4k]
		[102.4k,204.8k]
		[204.8k,3027.2k]
		[3027.2k,∞]
Retransmit rts	Count	[0,3]
		[3,5]
		[5,7]
		[7,∞]
Retransmit data	Count	[0,7]
		[7,∞]
Neighbor-node-count	Count	[0,7]
		[7,15]
		[15,23]
		[23,29]
Forward-node-count	Count	[0,0]
		[1,1]
		[2,2]
		[3,3]
		[3,29]

WIRELESS INTRUSION DETECTION ARCHITECTURE

The goal of intrusion detection is seemingly simple: to detect intrusions and also to identify unauthorized use, misuse and abuse of wireless nodes by both internal attackers and external penetrations. In other words, Intrusion detection is a process of identifying and responding to malicious activity targeted at computing and network resources. A network intrusion is a sequence of activities by a malicious individual that results in unauthorized security threats to a target network. Generally, Intrusion detection is classified as (1) Profile based intrusion detection and (2) Signature based detection. The designing an IDS in wireless networks is tougher challenge due to vulnerabilities and lack of physical infrastructure. Without centralized audit point such as routers and gateways, an IDS for wireless networks is limited to using only the current traffic coming in and out of the node an audit data (Fig. 1).

This study describes wireless intrusion detection architecture to monitor and detect the malicious activity of wireless node. The entire architecture consists of wireless traffic capturing module, preprocessor module, detector module, knowledge base training module and decision module. The first step is to collect the wireless feature set using NS2 in a session. This feature set is fed into preprocessing module. In the preprocessing module, wireless feature set is encoded into alphabets. The detection module has two jobs. First job is to learn the

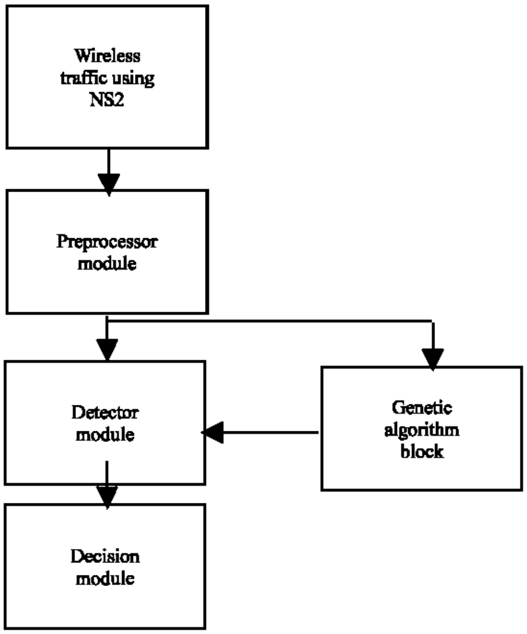

Fig. 1: Wireless intrusion detection modules

encoded wireless feature set and second job is to detect the intrusions from learned wireless feature set. The detection module uses genetic algorithm to perform training and detection.

A wireless node behavior exhibits some regularity in wireless node's event sequence in wireless traffic. Any deviation from wireless normal node behavior treated as signing of intrusion. In this study, genetic algorithm is used to learn the wireless node behavior in a computing environment due to its robustness and adaptability to changes in the environment. The genetic algorithm involves two steps. The first step involves coding the vectors with a string of bits, which form the input population of a genetic algorithm. The second step is finding a fitness function to test each individual of population against some evaluation criteria. In the learning process each event sequence of wireless node behavior forms a gene. Fitness is calculated for a collection of genes. If genes with required fitness cannot be found in the current generation, new set of genes are evolved through crossover and mutation. The process of evolution continued until genes with required fitness are found. The detection process involves defining vectors for event data and testing whether the vector indicates an intrusion or not.

Preprocessor module: Different steps involved in mapping wireless feature set into genes to learn regularities in wireless node behavior profile. In preprocessing, each wireless node feature set forms a gene. Alphabets are assigned for each wireless feature set in a session. Wireless node feature set in a session is divided into string of alphabets of size n called behavior gene. We have used 27 alphabets (Raghavan and Balajinath, 2001) to represent MAC layer feature set of wireless node in evolution of behavior.

In order to the wireless intrusion detection system to recognize intrusive behavior, it must first learn encoded feature set extracted from MAC layer of wireless node and wireless node behavior profile is formed in each wireless node in order to describe its normal behavior. Detector involves a process of establishing profiles of normal wireless node behavior and past behavior with current one. Detection depends on an assumption that extracted features set of MAC layer of wireless node exhibit predicable, consistent patterns of usage. The approach also accommodates adaptations to changes in wireless node behavior profile over time. In this approach, fitness function required for reproduction is based on the observation that extracted features set of MAC layer of wireless node can be best captured by observing the trend in total behavior entropy. This total behavior entropy gives a measure of amount of randomness in the wireless node behavior profile. It gives the frequency of entries in wireless node behavior profile. Frequent change in the total behavior results in large entropy value and the entropy value remains approximately the same for normal behavior.

Fitness function: Determination of appropriate fitness function to measure the fitness of behavior gene is important to improve the accuracy of the prediction. The fitness function a gene is given by:

$$\text{Fitness} = 1-|\sigma x-\phi| \qquad (1)$$

Where:

σx = Wireless node behavior entropy of predicated gene,

ϕ = Average wireless node behavior entropy of m previous behavior genes.

Entropy is defined as:

$$\text{Entropy} = \Sigma\, I-(P\,(i) * \log\,(P\,(i))/\log\,(n)) \qquad (2)$$

Where:

p(i) = Fraction of number of time alphabet I occurred to size of behavior gene and n is number of alphabets in the behavior genes

Parameters for wireless node behavior characterization: The input of the genetic algorithm consists of m behavior genes. A behavior gene is a set of n encoded feature set extracted from MAC layer of wireless node in the session. The three genetic operators are reproduction, crossover and mutation that are applied on the input population and output is a gene that describes the normal behavior of wireless node profile. The current gene is then taken and a 3-tuple value match index, entropy index and newness index is calculated from the actually occurred behavior and the predicated behavior. The 3 tuple value is described as follows:

Match index: The match index is a measure of regularity in the wireless node behavior. It is given by:

Match index = Count of encoded feature predicated correctly in a wireless node feature set/Size of the feature set

Entropy index: The entropy is a measure of wireless node behavior dynamics in the wireless feature set profile. It is given by:

$$\text{Entropy index} = \Sigma\, I-(P\,(i) * \log\,(P\,(i))/\log\,(N)) \qquad (3)$$

Where:

p (i) = Probability of occurrence of wireless node feature i in the wireless node feature set,

N = No. of unique wireless node feature in the wireless node feature set.

Newness index: The newness index is a measure of the number of new wireless node feature of wireless node, which have not occurred earlier in the feature set. It is given by:

Newness index = 1–No. of new wireless node feature of wireless node as well as in feature set/length of wireless node feature set.

This 3 tuple value calculated for the current feature set extracted from network layer and MAC layer of wireless node is compared with the threshold values to determine whether feature set extracted from network layer and MAC layer of wireless node is intrusion or not. Having a single threshold instead of having three-threshold values. The three-index value are combined to form and checked against the single threshold performs the detection. New index is extracted from three index called New mismatch index. This is because of the direction of inequality operators. If the match index of feature set extracted from network layer and MAC layer of wireless node is less than threshold of the match index, sample is considered intrusive. Also if the entropy index or newness index of current sample is greater than the threshold of corresponding index, sample is declared intrusive.

$$\text{Mismatch index} = 1 - \text{match index.}$$

$$\text{Threshold} = \alpha_1 * \text{MMI} + \alpha_2 * \text{EI} + \alpha_3 * \text{NI} \quad (4)$$

The weights α_1, α_2, α_3, need to be chosen for the finding the threshold. These weights are fixed by observing the values of the three indices that determine whether total behavior gene is intrusive or not.

SIMULATION ENVIRONMENT

The simulation is conducted on the platform of Network Simulator (NS2) (Liuy *et al*., 2005). Table 2 shows the NS2 parameters in our simulation. In the simulation, each node starts its move from a random location to a random destination with a randomly selected speed that uniformed distributed between (0, maxspeed). Once the destination is reached, the node stays there for as long as specified by pause time. then another destination location is chosen. Dynamic network topology and different mobility scenarios are modeled by varying the maxspeed and the pause time. To prevent all flows start from the beginning at the same time, each source node chooses its starting time for sending packets from the range of (0 s time).

We extracted the feature set described in Table 1 from NS2 simulation. These feature set is encoded and fed into intrusion detection module. The performance of wireless intrusion detection was tested using 30 wireless nodes. In experimental study, effect of wireless node feature set in the accuracy of predication, selection of length of initial observation period to learn wireless node behavior, evaluation of performance using false alarm rate and accuracy of intrusion detection are studied. Accuracy of behavior gene prediction decides the value of 3-tuple which in turn affects the accuracy of intrusion detection.

Operation of wireless node behavior based IDS: The wireless node feature set is extracted from simulated wireless networks. It is encoded and used as input stream to genetic algorithm based intrusion detection module. Learning module learns the trend in entropy value across wireless node feature set. The 3-tuple value < match index, entropy index, newness index > is calculated in a session. Expected normal behavior 3-tuple value is compared with calculated 3-tuple value to calculate the deviation in wireless node behavior. Intrusion detection module computes probability of current wireless node feature set being intrusive from deviation in wireless node behavior.

Performance of wireless intrusion detector: The performance of wireless intrusion detector is evaluated through two parameters.

Accuracy of intrusion: In this approach, intrusion is a set of actions which are deviating from the normal wireless node behavior. Probability of a feature set being intrusive is same as accuracy of detection.

$$\text{Accuracy} = (1-n/N) \times 100 \quad (5)$$

Where:

n = Count of feature value that are in total feature set,

N = Initial size of total feature set.

False alarm rate: False alarm rate is a measure of count of instances in which a genuine wireless node is classified as an intruder.

$$\text{False alarm rate} = (n/N) \times 100 \qquad (6)$$

Where:

n = Count of feature value that are in total feature set,
N = Initial size of total feature set.

RESULTS AND DISCUSSION

The 30 wireless node feature set is collected. Abnormal feature set value is artificially created. The normal values and abnormal values are appended to form a wireless feature set. Each wireless feature set has 7 values to form one behavior gene. The genetic algorithm for finding the normal behavior of a wireless node receives the initial 350 wireless feature set as initial population. The wireless feature set that occurs after the first 350 are used for testing. For the genetic algorithm, the cross probability is set as 0.6, the mutation probability is set as 0.001 and the number of generations is set as 5. These values have been obtained after experimental analysis. Table 2 shows least probability of false alarm rate. Initial values are assigned to 3 tuple parameter. The probability of current wireless feature set was computed. The Fig. 2-4 shows the entropy index, network index, newness index, sequence index with intrusive samples.

The performance of wireless intrusion detection is specified in terms of detection probabilities and false alarm probability. The detection probability against false alarm describes probability for different threshold values (Fig. 5).

Table 2: NS2 simulation environment

Parameters	Value/Choice
Topology	500×500 m
Node movement	Random waypoint model
Max movement speed	10 m sec^{-1}
Radio range	250 m
Node set count	30
Total No. of flows	25
Average transmission rate per flow	2 packets sec^{-1}, 512 b packet^{-1}
Training execution time	2000 sec
Testing execution time	200 sec
Feature sampling interval	5 sec

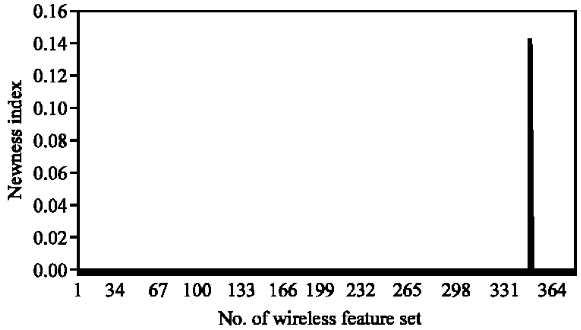

Fig. 2: Filtering out intrusive samples using newness index with a threshold value (0.1)

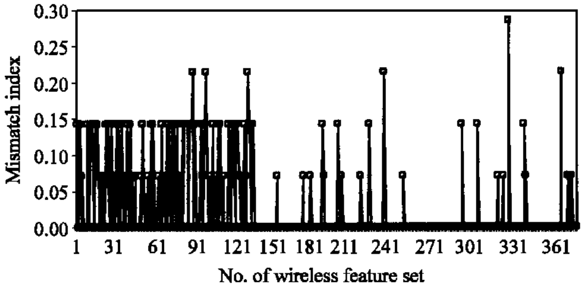

Fig. 3: Filtering out intrusive samples using match index with a threshold value (0.1)

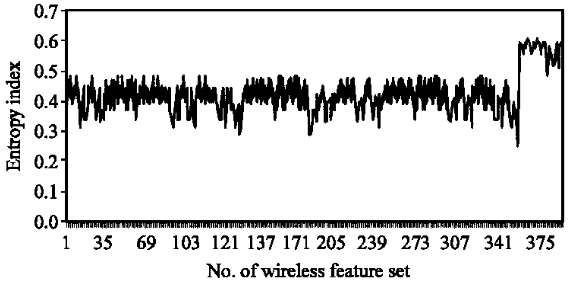

Fig. 4: Filtering out intrusive samples using entropy index with a threshold value (0.5)

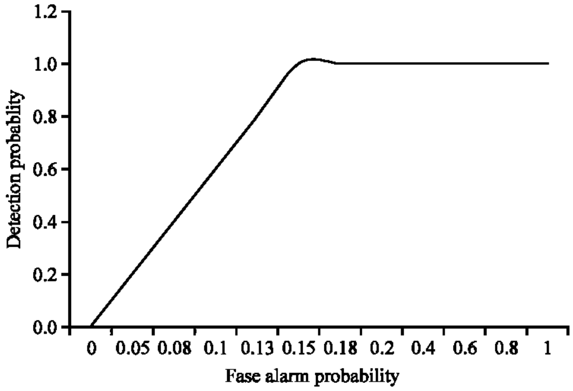

Fig. 5: Performance evaluation of wireless intrusion detection

CONCLUSION

We described a novel idea of wireless intrusion detection architecture. As we know that wireless security vulnerabilities will keep pace with the technology. Since the traditional perimeter defenses are inadequate for wireless network. The proposed work offers new kind of defense against intrusion.

REFERENCES

Huang, Y. and W. Lee, 2003. A cooperative intrusion detection system for ad hoc networks. In: Proceedings of the 1st ACM workshop on Security of Ad Hoc and Sensor Networks, pp: 135-147.

Liuy, Y., Y. Liy and H. Many, 2005. MAC layer anomaly detection in ad hoc networks. 6th IEEE Information Assurance Workshop, 15-17 June, USA.

Raghavan, S.V. and B. Balajinath, 2001. Intrusion detection through learning behavior model. Int. J. Comput. Commun., 24: 1202-1212.

Tseng, C., P. Balasubramanyam, C. Ko, R. Limprasittiporn, J. Rowe and K. Levitt, 2003. A specification-based intrusion detection system for AODV. In: Proceedings of the 1st ACM Workshop on Security of Ad Hoc and Sensor Networks, pp: 125-134.

Zhang, G.Y. and W. Lee, 2000. Intrusion detection in wireless ad-hoc networks. 6th International Conference on Mobile Computing and Networking, pp: 275-283.