International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-2, Issue-9)

Research Article | **September 2013**

Detection of Anomaly Network Traffic for Mobile Ad-hoc Network Using Fuzzy Logic

RajyaLakshmi G.V [1], Anusha.K [2]
School Of Information Technology and Engineering,
VIT University, Vellore – 632014, India

Abstract:

Mobile ad-hoc Network (MANET) is a self organized wireless network which is a dynamic topology with large number of nodes (mobile devices) responsible for its creation, operation and maintenance. Due to distinguishable characteristics of MANET there is a minor difference between legitimate traffic and illegitimate traffic. Many techniques have been developed for this issue. A hybrid technique which is a combination of entropy of network features and Support Vector Machine is compared with individual methods. In this paper, we propose effective anomaly based model analyzed by extracting more network features from the MANET and used fuzzy logic for classify the network traffic that is attack traffic or legitimate traffic.

Keywords: MANET, Wireless Networks communication, Fuzzy Logic, Anomaly Detection Technique, IDS.

I. Introduction

Network security is becoming of important significance in MANET. Anomaly based intrusion detection system finding problems in the traffic data patterns and raises alarms if there is any unusual behavior in the traffic data pattern. When compare to Signature based intrusion detection system, Anomaly based detection system can detect a new attack at the time of attack with the assumption. Here a hybrid approach uses Anomaly based detection system as entropy of network features and Support Vector Machine for detecting changes in the behavior of network traffic.

Network Entropy [1] is used to measure ambiguous or any disorder in the network. And also it is used for identifying unusual behavior in the MANET. In this paper Entropy of different network features are calculated from the network and observed at the time of attack. When the entropy values of these network features deviates from fixed predefined range, observed that there is a attack but with high false alarm rate. Support Vector Machine (SVM) is one of the best categorization methods. This algorithm based on linear and non linear regression. This model is using different network features for classifying. Support Vector Machine does not give good results when network features are give it as directly. Also SVM computation methods are complex where as FUZZY uses simple system to calculate the membership functions. To evaluate the anomaly traffic detection techniques, we are going to extract network features from the Mobile Ad hoc Network. The network traffic is affected by different attacks, like Denial of Service attack, port scanning, etc. Normalized relative network entropy has been developed to analyze network behaviors. Support Vector Machine is applied to solve the Denial of Service (DoS) attack. Several major kinds of classification method including Bayesian networks, decision tree induction, k-nearest neighbor classifier, genetic algorithm, case based reasoning and fuzzy logic techniques. In this paper we are going to propose fuzzy logic for traffic classification.

II. Literature Survey

Qian Quan, Che Hong-Yi and Zhang Rui [1], Entropy is a measurement of the disorder of a system. If the system tends to be disorder, its entropy increase towards 1; if the system tend to be order, its entropy decrease towards 0. Entropy based intrusion detection which recognizes the network behavior only depends on the packets themselves and do not need any security background knowledge or user interventions, shows great advantages in network security areas. Here they compare two entropy methods, network entropy and normalized relative network entropy (NRNE), to classify different network behaviors. The experimental results show although the two methods are efficient, the improved relative network entropy, NRNE is better which takes more attributes into consideration simultaneously and we can get an overall view of the abnormal network behavior**.**

Thair Nu Phyu [2], Classification is a data mining (machine learning) technique used to predict group membership for data instances. Here they presented the basic classification techniques. Several major kinds of classification method including decision tree induction, Bayesian networks, k-nearest neighbor classifier, case-based reasoning, genetic algorithm and fuzzy logic techniques. The goal of this survey is to provide a comprehensive review of different classification techniques in data mining. Decision trees and Bayesian Network (BN) generally have different operational profiles, when one is very accurate the other is not and vice versa. On the contrary, decision trees and rule classifiers have a similar operational profile. The goal of classification result integration algorithms is to generate more certain, precise and accurate system results. Numerous methods have been suggested for the creation of ensemble of classifiers. Although or perhaps because many methods of ensemble creation have been proposed, there is as yet no clear picture of which method is best.

Thakare S.P.1 and Ali M.S.2 [3], It is possible to develop an anomaly based intrusion detection system which detects the intrusion behavior within a network. By introducing fuzzy decision-making module system will be more accurate for attack detection, using the fuzzy inference approach. An effective set of fuzzy rules for inference approach will be identified automatically by making use of the fuzzy rule learning strategy, which is more effective for detecting intrusion in a computer network. The strategy for implementation is as follows: The definite rules will be generated by mining the single length frequent items from attack data as well as normal data. Fuzzy rules will be identified by fuzzifying the definite rules. These rules will be fed to fuzzy system, which will classify the test data.

W. Lee and S. Stolfo [4], Novel Network Data Mining approach that applies the K-means clustering algorithm to feature datasets extracted from flow records. Training data are divided into clusters of time intervals of normal and anomalous traffic. While the data mining process is relatively complex, the resulting cluster centroids can be used to detect anomalies in new on-line monitoring data with a small number of distance calculations. This allows deploying the detection method for scalable real-time detection. Applying the clustering algorithm separately for different service improves the detection quality. .Training data containing unlabeled flow records are separated into clusters of normal and anomalous traffic. The corresponding cluster centroids are used as patterns for computationally efficient distance-based detection of anomalies in new monitoring data.

Kriangkrai Limthong, Kensuke Fukuda, Yusheng Ji, and Shigeki Yamada [5], Naïve Bayesian classifier used for network traffic anomaly detection. We performed experiments on real network traffic acquired from a campus network. In addition, we selected five different types of test bed anomalies from DARPA in order to evaluate the efficiency and accuracy of naïve Bayesian classifier by using seven time interval values. The results of our study illustrated three separate features; the number of packets, the sum of packet size, and the number of flows, used in the naïve Bayesian classification method. In the classification step they applied naïve Bayesian for classifier to distinguish between normal and attack traffic.

III. Methodologies

A hybrid approach is used for detecting anomalies in the network which is a combination of both entropy and fuzzy based anomaly detection methods. By combining the entropy of network feature and fuzzy based anomaly detection benefits of both the techniques are used and demerits of both the techniques can be removed. Entropy based techniques has the advantage that it can better present the properties of the network traffic and fuzzy based anomaly detection is good for classification. In this firstly normalized entropy values of network features are calculated using the algorithm which is given below in entropy based approach , then these normalized entropy values are sent to fuzzy based approach for classification. Now this fuzzy classification model can classify the network traffic in attack traffic or legitimate traffic.

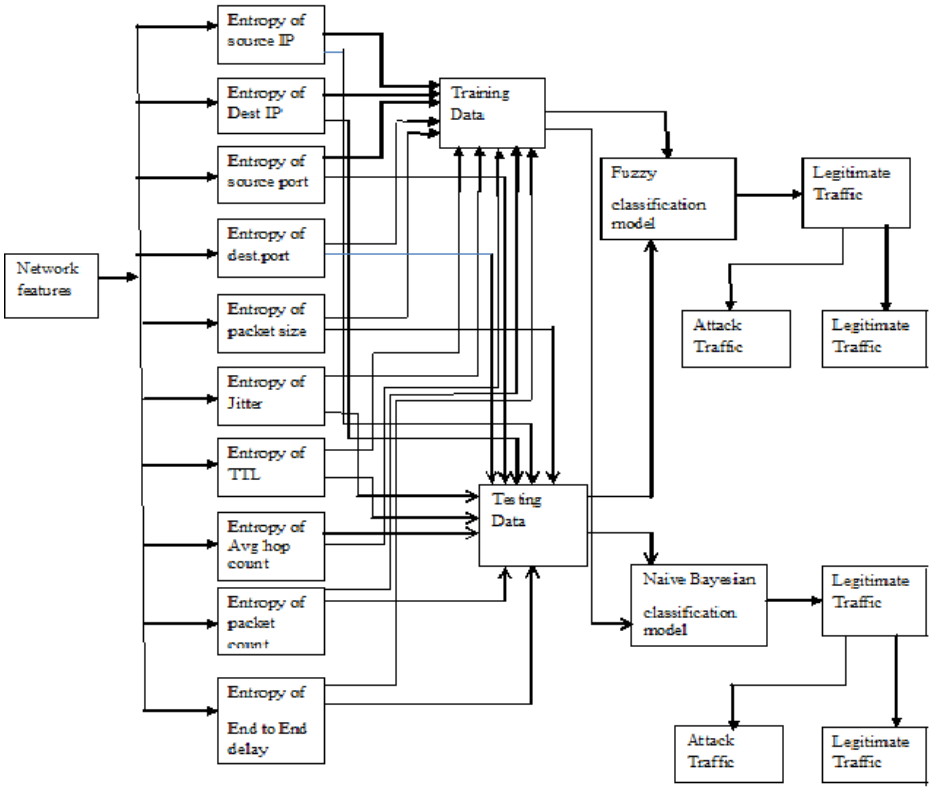

Fig 1: Overall architecture

IV. Implementation

A. *Entropy based approach*

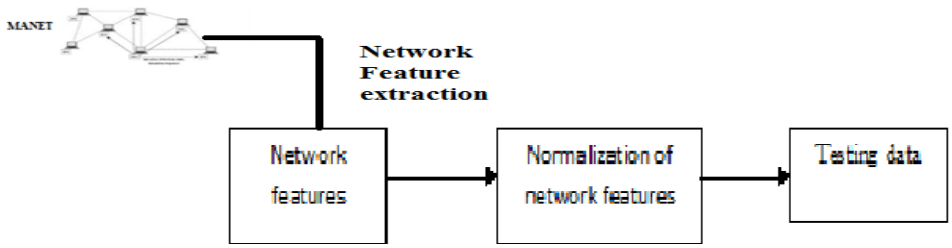

Fig 2 Normalized network entropy

Entropy is a measurement of uncertainty or randomness. Entropy of different network features are calculated to create anomaly based intrusion detection system. Here we are extracting network feature from the dataset.

B. Network feature extraction:

- Normalized entropy of packet size
- Normalized entropy of source IP address
- Normalized entropy of source port number
- Normalized entropy of destination IP address
- Normalized entropy of destination port number
- Normalized entropy of packet type (TCP , ICMP , UDP)
- Normalized entropy of packet count
- Normalized entropy of payload
- Normalized entropy of Time to Live (TTL)
- Normalized entropy of number of different source and destinations pair.

C. Normalized entropy calculation:

Input : Network feature
Output: Normalized entropy for each network feature
Extract features from packet header (for example: TTL)
Calculate frequency of all TTL value
Calculate probability for each TTL value

$$Pi=mi / T$$

Here mi = frequency of ith TTL value
T = total number of packets in that time interval
Calculate entropy for Each TTL value

$$hi=\sum -pi \log pi$$

Normalize the entropy, in the time interval by

$$H= \sum hi / \log(F)$$

F is the total number of distinct TTL value.

Finally normalized entropy values will be given as input for the next method that is Fuzzy based Anomaly Detection method and then traffic will be classified from the output.

D. *Fuzzy Based Anomaly Detection*

By using Fuzzy based Anomaly Detection method traffic will be detected and also classified the traffic that falls into the one category that is attack traffic or the normal traffic. We can get more accurate values than other methods while using fuzzy logic.

From the output, traffic is classified as legitimate traffic or attack traffic.
Steps involved in the process

- Input will be fuzzified.
- Fuzzification will be done by membership functions.
- Deriving inference rules.
- Decide the type of traffic by using Defuzzification

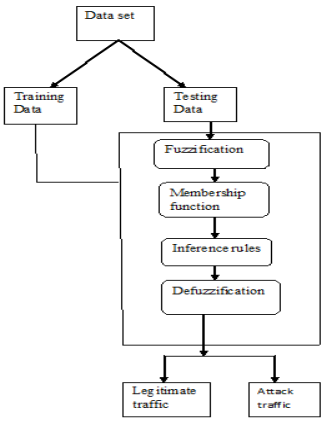

Fig 3 Fuzzy based anomaly detection

E. Fuzzification process:

The normalized entropy values will be given as input for this method which decides the type of traffic. The normalized entropy values are

- Source IP - IP address of the source
- Destination IP -IP address of the destination
- Source port - Port number of the source
- Destination port - Port number of the destination
- Packet byte size - Size of the packet
- Packet type - Type of the packet (TCP, UDP, ICMP)
- Time To Live - This field controls number of hops a datagram can travel through in a network
- Payload - It is the part of the transmitted data
- Packet count up - Total number of packets
- No of different source & destination pair - Total number of source and destination pair.

These are the testing data which will be obtained from the Entropy based Intrusion Detection System.

Membership Functions:

Membership function defines the fuzziness in a fuzzy set irrespective of the elements in the set, which are discrete or continuous. Membership function can be thought of as a technique to solve a problems on the basis of experience rather than a knowledge .Available membership functions are trapezoidal, Gaussian , triangular methods. Here we are using trapezoidal method.

Inference rules:

Fuzzified output will be compared using inference rules then from the output traffic will be classified. For example we can take TTL value. If the TTL value is constant at particular time Then the traffic is not legitimate otherwise the traffic is legitimate.

Defuzzification:

This is about output of the Fuzzification process. Output of traffic classification process:

- Legitimate traffic
- Attack traffic

F. Naive Bayesian classifier

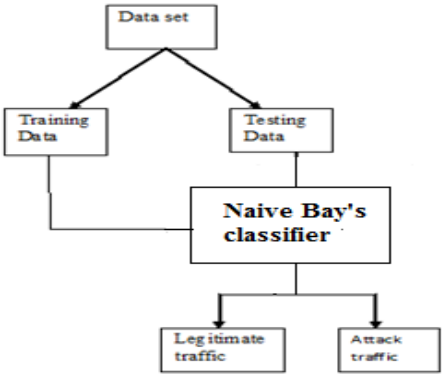

Fig 5: Naïve Bayesian classifier for traffic classification

A naive Bayesian classifier is a simple probabilistic classifier based on applying Bayesian theorem with strong (naive) independence assumptions. A naive Bayesian classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter.

A naive Bayesian classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features. For some types of probability models, naive Bayesian classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayesian models uses the method of maximum likelihood in other words, one can work with the naive Bayesian model without believing in Bayesian probability or using any Bayesian methods.

By using Naïve bay's classifier traffic will be classified as normal traffic and attack traffic. Here also normalized entropy values of every network features with attack are given as input for this method that is known as testing data. Training data is normalized entropy values of normal traffic scenario network features. Then by using probability formulas or method of naïve bays classifier traffic will be classified as legitimate traffic and attack traffic.

V. Simulation

We use QualNet 5.0.2 network simulator to simulate the scenario. The scenario which has 1500*1500 canvas area contains network components. The protocol used here is AODV. And we used TRAFFIC GEN application for traffic generator. Also we used Ns2 network simulator to simulate the scenario and compare the network scenario with attack and without attack. And also we have generated the graph and compare with attack and without attack. We have included more network features, by adding more network features the network traffic will be analyzed more effectively than previous.

A. Comparison between Fuzzy based classifier and Naïve Bayesian classifier

It is a Computation less classifier, because here we are using membership function and inference rules to generate the result. So it is simple to use. But in Naïve Bayesian we are using probability distribution methods to compute the result. So here computation is more when compare to fuzzy based classifier.

Moreover in fuzzy we won't use probabilistic method to give the result. Here we are using linguistic method for discrete and continuous variables. Therefore by using we can get accurate result than Naïve Bayesian classifier.

B. Naïve Bayesian classifier will use probabilistic method to compute the result.

Fuzzy limits from 0 to 1. Therefore we can easily calculate average value which is between 0 to 1 by using membership function. But in Naïve Bayesian classifier we can't calculate in between values from 0 to 1 and also can't calculate average traffic by using membership function. So we can only classify the traffic as normal traffic or attack traffic.

C. Training data set values

- Normalized entropy value of TTL - 0.98
- Normalized entropy value of Average hop count -1.66
- Normalized entropy value of Server address -1.635
- Normalized entropy value of Client address -2.241
- Normalized entropy value of Packet count -2.817
- Normalized entropy value of Packet size -3.381
- Normalized entropy value of Throughput -3.396
- Normalized entropy value of End to End delay -4.03

These are the normalized entropy values of each network features for the normal traffic scenario. This dataset is called as training dataset. And these dataset is fixed as threshold values for traffic classification and given as input for Fuzzy based classification method.

D. *Naïve Bayesian classifier for traffic classication*

TABLE 1
Naïve Bayesian classifier Testing Dataset

Network Feature	Threshold value	Attack traffic
TTL	<= 0.98	0
TTL	>0.98	1
Avg hop count	<=1.66	0
Avg hop count	>1.66	1
Server Address	<=1.635	0
Server Address	>1.635	1

Client Address	<=2.241	0
Client Address	>2.241	1
Packet count	<=2.817	0
Packet count	>2.817	1
Packet size	<=3.381	0
Packet size	>3.381	1
Throughput	<=3.936	0
Throughput	>3.936	1
Avg End to End delay	<=4.03	0
Avg End to End delay	>4.03	1

Given testing data X, *posteriori probability of a hypothesis* H, P(H|X), follows the Bayes theorem

P (Ci / X) = P (X / Ci) * P (Ci)

Naïve Bayesian classifier follows this probability method for classification.

Class:

C1: Attack traffic = "Yes"

C2 : Attack traffic = "No"

Probability of two classes P(Ci) :

P (Attack traffic = "Yes") = 8/16 = 0.5

P (Attack traffic = "No") = 8/16 = 0.5

P (X / Ci) Calculation:

P(TTL = "<= 0.98" | Attack traffic = "No") = 0/8 = 0

P(TTL = " > 0.98" | Attack traffic = "yes") = 1/8 = 0.125

P(Avg hop count = "1.66" | Attack traffic = "No") = 0/8 = 0

P(Avg hop count = "1.66" | Attack traffic = "yes") = 1/8 = 0.125

P(Server Address = "<=1.635" | Attack traffic = "No") = 0/8 = 0

P(Server Address = "> 1.635" | Attack traffic = "yes") = 1/8 = 0.125

P(Client Address = "<=2.241" | Attack traffic = "No") = 0/8 = 0

P(Client Address = ">2.241" | Attack traffic = "Yes") = 1/8 = 0.125

P(Packet Count = "<=2.817" | Attack traffic = "No") = 0/8 = 0

P(Packet Count = ">2.817" | Attack traffic = "Yes") = 1/8 = 0.125

P(Packet Size = "<=3.381" | Attack traffic = "No") = 0/8 = 0

P(Packet Size = ">3.381" | Attack traffic = "Yes") = 1/8 = 0.125

P(Throughput = "<=3.936" | Attack traffic = "No") = 0/8 = 0

P(Throughput = ">3.936" | Attack traffic = "Yes") = 1/8 = 0.125

P(Avg End to End delay = "4.03" | Attack traffic = "No") = 0/8 = 0

P(Avg End to End delay = "4.03" | Attack traffic = "Yes") = 1/8 = 0.125

P(X/Ci) = P(X/Attack traffic="Yes")

=0.125*0.125*0.125*0.125*0.125*0.125*0.125*.125 = 1

P(X/Ci) * P(Ci)= P(X/Attack traffic="Yes") = 1*0.5 = 0.5

P(X/Ci) = P(X/Attack traffic="No")

=0*0*0*0*0*0*0*0= 0

P(X/Ci) * P(Ci)= P(X/Attack traffic="No") = 0*0.5 = 0

Therefore X belongs to class ("Attack traffic=Yes"). Because it has greater value than the ("Attack traffic value=No") value, that is normal traffic value. So for the testing data it has attack traffic in the network traffic scenario.

VI. Conclusion & Futurework

We designed anomaly traffic detection based model can be analyzed with network traffic more effectively by extracting more network features and also we included fuzzy logic and Naïve Bayesian classifier for classifying the anomaly network traffic and we proposed that better method as fuzzy based classification than Naïve Bayesian classification. Naïve Bayesian based classification doesn't give accurate and efficient results than fuzzy based classification method. Moreover Naïve Bayesian classifier uses probability method for classification. In future work, we can extract more network features for anomaly based model can be analyzed still more effectively. And we can calculate same fuzzy based classification using intuitionistic fuzzy for more efficient and accurate result and also in future we can compare many machine learning based algorithm with this and propose best classification algorithm.

References

[1] Qian Quan, Che Hong-Yi and Zhang Rui, "Entropy Based Method For Network Anomaly Detection", 15th IEEE Pacific Rim International Symposium on Dependable Computing, 2009.
[2] Thair Nu Phyu, "Survey of Classification Techniques in Data Mining", Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol: I March 18 - 20, 2009.
[3] Thakare S.P.1 and Ali M.S.2, "Network intrusion detection system & fuzzy logic", BIOINFO Security Informatics ISSN: 2249-9423 & E-ISSN: 2249-9431, Volume 2, Issue 1, 2012
[4] W. Lee and S. Stolfo, "Data mining approaches for intrusion detection, in *proceedings of the 7th USENIX Security Symposium*, San Antonio, TX, 1998.
[5] Kriangkrai Limthong, Kensuke Fukuda, Yusheng Ji, and Shigeki Yamada, "Impact of Time Interval on Naïve Bayes Classifier for Detecting Network Traffic Anomalies" ICSEC , September 2011.
[6] Andreas Kind, Marc Ph. Stoecklin, and Xenofontas Dimitropoulos, "Histogram-Based Traffic Anomaly Detection" IEEE TRANSACTIONS ON NETWORK SERVICE MANAGEMENT, VOL 6, NO 2, JUNE 2009.
[7] Sandeep A.Thorat, Amit K. Khandelwal ,Bezawada Bruhadeshwar and K. Kishore, "Payload Content based Network Anomaly Detection", Centre for Security Theory and Algorithmic Research (CSTAR).
[8] Marianne A. Azer, Sherif M El-Kassas, Magdy S. El-Sodani, "A Survey On Anomaly Detection Methods For Ad Hoc Networks", Ubiquitous Computing And Communication Journal, Vol 2, NO 3.
[9] Yu Gu, Andrew McCallum, Don Towsley, "Detecting Anomalies in Network Traffic Using Maximum Entropy Estimation".
[10] Milan Kumari, Sunila Godara, "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction", IJCST Vol. 2, Iss ue 2, June 2011
[11] Neminath Hubballi, Santosh Biswas, Sukumar Nandi, " Fuzzy mega cluster based anomaly network intrusion detection"
[12] Vijayan R, Mareeswari V and Ramakrishna K: "Energy based trust solution for detecting selfish nodes in MANET using fuzzy logic", International journal of research and review in computer science, vol.2 No.3, June 2011.
[13] S.Ganapathy, P. Yogesh, and A.Kannan: "Intelligent agent based intrusion detection system using enhanced multiclass SVM", Computational Intelligence and Neuroscience, volume 2012, article ID 850259.
[14] Monita Wahengbam, Ningrinla Marchang: "Intrusion detection in MANET using fuzzy logic", 2012 IEEE.
[15] QualNet Network simulator 5.0.2 "Programming guide".