A Survey on Decision Tree Based Method in IDS for Attack Detection in Mobile Ad Hoc Network

Archan A. Thakkar[1] Prof. Avani Parmar[2]

[1]P.GStudent [2]Faculty

[1,2]Department of Computer Engineering

[1,2]Hasmukh Goswami College of Engineering, Vahelal, Ahemdabad, Gujarat, India

Abstract— Mobile Ad hoc Network (MANET) is a collection of mobile nodes that are capable of communicating with each other without any infrastructure. Mobile Ad Hoc Network suffers from security attacks. Security is a burning issue in MANET. There are many algorithm used to protect whole network. The field of Intrusion Detection has received increasing attention in present years. The Intrusion Detection techniques using data mining have attracted more and more interest. In data mining techniques Decision Tree is an important method of classification. ID3 is one of the popular decision tree algorithm. The goal of our survey is to present a comprehensive review of the recent literature on the ID3 algorithm and survey on how they are improved than the others and which factors are to altered to further impove them.

Key words: Security Attacks, Intrusion Detection System, Decision Tree, ID3 algorithm

I. INTRODUCTION

Mobile Ad hoc Network (MANET)[1] is a self organized system which doesn't have any predefined infrastructure. MANET mobile nodes are connected by wireless links. MANET are suitable for application in which no infrastructure exists such as milatry field, emergency rescue, vehicular communication and mining operation. In ordinary wireless network, two nodes cannot communicate with each other when the distance between the two nodes is beyond the communication range. MANET solves this problem by allowing intermediate parties to relay data transmissions.

II. OVERVIEW OF ATTACKS:

Security is a major problem in Mobile Ad hoc Network(MANET) because of its characteristics like open medium, changing its topology dynamically and lack of central monitoring and no clear defense mechanism. There are following types of attacks [1]:

A. Rushing Attack:

Rushing Attack is an effective denial of service attack. This attack makes route discovery process difficult. Whenever attacker receives a route request packet from node it floods the packet quickly throughout the network before other nodes which also receives the same route request packet can react.

B. Blackhole Attack:

In Blackhole attack source first send route request packet for route establishment. Blackhole attacker node directly replie through Route Reply to source node. Source node selects blackhole node, it will then eavesdrop or directly drop the received data packets.

C. DDoS Attack[2]:

In Distributed Denial of Service (DDoS) attack different system compromise together and send malicious packets to the target system. By this target system flooded with useless packets which results in failure of the system thereby denying services to users.

III. INTRUSION DETECTION SYSTEM

Intrusion Detection System is a system to monitor events in computer or network and analyze the security violation in network. Intrusion Detection System can be classified based on which events they monitor, how they collect information. Intrusion Detection System can be classified in Network based IDS and Host based IDS [8].

A. Host Based Intrusion Detection System (HIDS):

Host based IDS runs on the machine it monitors. It is dependent on operating system.

B. Network Based Intrusion Detection System (NIDS):

Network based IDS are often runs on dedicated machine that observe the network flows sometimes in conjuction with a firewall.

Intrusion Detection System can also be classified based on its decision techniques.

A. Misuse based/ Signature based IDS:

Signature based IDS are also known as misuse based IDS. In misuse detection looks for a specific signature to match an instruction. They are provided with the signature o pattern but SIDS are of little use for as yet unknown attack methods. So Signature based IDS cannot detect unknown attack, it will only detect known attack.

B. Anomaly based/ Statistical based IDS:

Anomaly based IDS have the advantage that they can detect new types of intrusion as deviation from normal data. In this problem set of normal data to train from is given, and a new piece of test data is also given. The objective of Intrusion Detection System is to determine whether the test data belong to normal or to an anomalous behavior. However this anomaly detection scheme suffers from a high rate of false alarams.

IV. DECISION TREE

In a decision tree method classification[4] is done based on some decision. The classification is done from the root node to the leaf node. Leaf nodes specify category information. Each node in decision tree holds some attributes and branches corresponds some value according to attributes. Internal nodes are test nodes and test result corresponds to each branches and leaf nodes defines the distribution

All rights reserved by *www.ijsrd.com*

situation of various types. Following are most essential and commonly used classification algorithm used in decision tree.

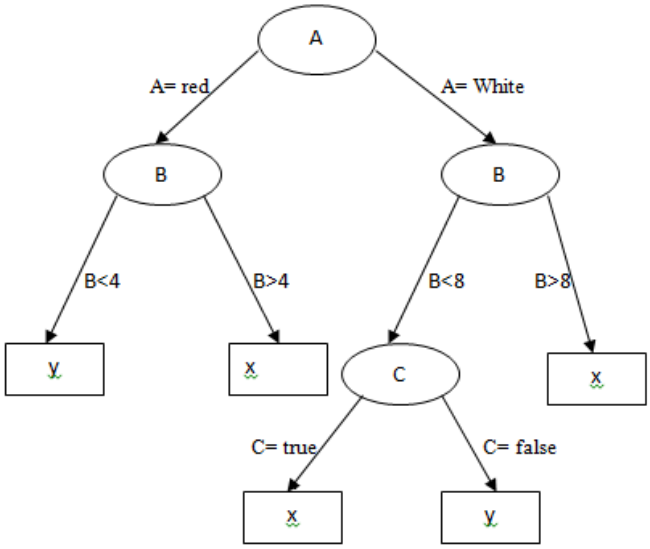

Fig. 1: Decision Tree

A. ID3 Algorithm:

ID3(Iterative Dichotomizer 3) algorithm[4] is a decision tree algorithm. ID3 algorithm is based on the Entropy and Information Gain. In ID3 algorithm number of levels in decision tree is equal to the number on input variable. Splitting process continue until no more split exist.

1) C4.5 Algorithm:

C4.5 algorithm[4] is extension of the ID3 algoithm. The attribute selection in this algorithm is based on the assumption. This algorithm don't exploit statistical correlation between different features of the training points. This algorithm is preferred where the storage is a major concern because in this algorithm there is no extra computation and storage required.

V. RELATED WORK

Data Mining is the process of mining useful, meaningful information from large volume of data. There are many effective techniques such as correlation analysis, evolution analysis, classification technique. Decision Tree is used in which classification rules are generated. ID3 algorithm is a decision tree algorithm used for classification.

Mrutyunjaya panda [5] describes the basic ID3 algorithm. ID3 is a n attribute based learning algorithm that constructs a decision tree based on training set of data. In ID3 attribute has highest information gain is selected and generates decision tree according to the information gain. The attribute which has highest information gain is selected as a root node. This process continues iteratively and generates a decision tree.

Feng Yang [6] proposes a study uses a Taylor formula. ID 3 algorithm has some disadvantages. In ID 3 algorithm, decision tree is generated based on the information gain value which is favorable for features with the large number of property value but sometimes more values are not always best. In ID3 algorithm, decision tree algorithm changes with increasing the training set. So to overcome this shortcoming of the ID 3 algorithm in [6] uses the Taylor formula in the equation of the Information Gain. By using Taylor formula in attribute selection criteria it simplifies the complexity of the decision tree and also reduce the generation time of the decision tree and also reduce the data calculation and thus improve the efficiency of the decision tree classifier.

Giovanni Vigna [7] proposes a WebSTAT, a STAT based Intrusion Detection System that analyzes web requests looking for malicious behavior. The goal is to perform early detection of malicious activity and possibly prevent more serious damage to the protected site. WebSTAT operates on multiple event streams and it is able to correlated both network-level and operating system-level events with entries contained in server logs. This integrated approach supports more effective detection of web-based attacks and generates a reduced number of false positives.

K.Hanumantha Rao [8] proposes a new approach which combines k-means method and ID3 algorithm. K-means method is a clustering method. K-means method is used because it is a data driven method and it is greedy approach so this method at least guarantees a local minimum of the criterion function. First stage of proposed algorithm is K-means clustering, in this stage K-means clustering is performed on training dataset to obtain k disjoint clusters. In next stage of proposed algorithm the K-means method is combined with the ID3 decision tree learning by building an ID3 decision tree using the instances in each K-means cluster. Computer is connected to Network; by connecting the network will get packets to search the packet whether it is normal or anomaly.

WU Sen [9] proposed a improved classification algorithm based on minsup and minconf concept based on ID3. Here minsup describes minimum support and minconf describes minimum rule confidence. Minimum support reduce data size for subtree. Minimum support to test class attributes which is set by users or experts. Another parameter is minimum rule confidence. This parameter is also set by users or experts. After computing rule computing of each branch in decision tree, comparison of the result with the minimum rule confidence, and discard branches of the decision tree whose rule confidence are less than minimum rule confidence. This method simplifies decision tree which improves ability of tree to correctly classifies data set. Main disadvantage of this algorithm is that, if minsup and minconf parameters are set strictly then a lot of useful information may be ignored and efficiency of the algorithm will reduce.

Guangun Zhai [10] proposes improvement on basic ID3 algoithm. In basic ID3 algorithm, attributes are selected based on maximum Information Gain. In this if the attributes selected is not suitable then decision tree have longer path which result in that rules set may have redundancy. To remove this shortcoming of ID3 algorithm in [] proposes a modification in the formula of the Information Gain. Two new attributes are added in main formula which is n and m where n is a number of sample and m is a number of values of attribute A. modified formula of Information Gain is as follows:

Gain'(A) = Gain(A)*(n-m)

The decision tree created by improved algorithm has less node. So detecting time of time of the improved algorithm is less which increases the efficiency of the algorithm.

All rights reserved by www.ijsrd.com

Hongwu LUO, [11] proposes improvement on ID3 algorithm. In ID3 algorithm splitting attribute selection to choose the attribute which has many values in ID3 algorithm. In this proposed algorithm which modifies the Information Gain module of the ID3 algorithm. Taylor formula is used in improvement of ID3 algorithm. In ID3 algorithm Logarithm function is used in information entropy. For large data set repeating Logarithm calculation lead to a large number of computation so it decreases the speed of generation of decision tree. By using Taylor formula Logarithm operation will be reduced. In improved algorithm Attribute weight parameter is included. By this when calculating the information gain, a priority value is given to each condition attribute, which reduces the non-essential attributes and also avoids dependence on attribute with more values and so result derived is more correct which increases the efficiency of the algorithm.

VI. CONCLUSION

Mobile Ad hoc Netwok(MANET) has many challenges, but broad range of applications are eliciting more and more interest from the research community as well as from industry. By using decision tree techniques of data mining like ID3 in Intrusion Detection System can provide more security in MANET.

REFERENCES

[1] H. Nguyen , U. T. Nguyen, "A study of different types of attacks on multicast in mobile ad hoc networks " Ad Hoc Networks 2008.
[2] Rui Zhong, Guangxue Yue , " DDoS Detection System based on Data Mining", Processings of the second International symposium on Networking and Network Security (ISNNS'10)
[3] V. Jaiganesh, "Classification Algorithms in Intrusion Detection System: A Survey", IJCTA | Sept-Oct 2013
[4] Mrutyunjaya panda, " A comparative study of data mining algorithm for Network Intrusion Detection", First International Conference on Emerging Trends in Engineering and Technology 2008,IEEE
[5] Feng Yang, Hemin Jin, Huirnin Qi, " Study on the Application of Data Mining for Customer Groups Based on the Modified ID3 Algorithm in the E-commerce", 2012 International Conference on Computer Science and Information Processing (CSIP), 2012 IEEE.
[6] Giovanni Vigna ," A Stateful Intrusion Detection System for World-wide web servers", Computer Security Applications Conference,2003. Proceedings. 19th Annual 8-12 Dec. 2003, IEEE
[7] K. Hanumantha Rao, "Implementation of Anomaly Detection Technique Using Machine Learning Algorithms", International Journal of Computer Science and Telecommunications [Volume 2, Issue 3, June 2011]
[8] WU Sen," Improved Classifi'cation Algorithm by Minsup and Minconf Based on ID3", 2006 IEEE.
[9] Guangqun Zhai , Chunyan Liu, " Research and Improvement on ID3 Algorithm in Intrusion Detection System", 2010 Sixth International Conference on Natural Computation (ICNC 2010), 2010 IEEE.
[10] Hongwu LUO, " An Improved ID3 algorithm based on attribute importance-weighted", 2010 IEEE.

All rights reserved by *www.ijsrd.com*