



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Differentiating Network Attacks using C4.5 Algorithm with Multiboosting

V.Balaji¹, Varalakshmi.K²

[#] Department of Computer Science and Engineering,

SRM University, SRM Nagar,

Kattankulathur – 603203,

Kancheepuram District, Tamil Nadu,

INDIA.

Abstract— In most of the security system intrusion detection system is deployed to detect and possibly prevent the attacks. Data mining methods can be used in intrusion detection system. There are several data mining methods like classification tree algorithms (C4.5), Support vector machine and so on. For a detection system to detect an attack efficiently and accurately, it must be able to reduce false positives (identifying an attack when there is no attack in the data stream). In the proposed system multiboosting technique is used with the C4.5 classification algorithm to reduce false positives and classification errors while classifying the attacks. The attacks that are going to be detected by the system are Denial of Service attacks (DoS), Remote to Local attack (R2L), User to Root attack (U2R).

Keywords— Multiboosting technique, C4.5 algorithm, Denial of Service attacks (DoS), Remote to Local attack (R2L), User to Root attack (U2R)

I. INTRODUCTION

Networks are subject to attacks from malicious sources. Attackers can steal the confidential information passively or damage the data in a system or in a network during data flow. Consequence of cyber attack depends on the size of data and cost of data storage and network infrastructure of organization. DoS attacks cause huge financial and infrastructure loss. The potential attacks that are to be detected are probe attack, R2L attack, DoS attack, U2R attack which is described below.

A. DoS ATTACKS:

A denial of service (DoS) attack aims to deny access by legitimate users to shared services or resources. This can occur in a wide variety of contexts, from operating systems to network-based services. It can be launched in two forms. The first form aims to crash a system by sending one or more carefully crafted packets that exploit software vulnerability in the target system. The second form is to use massive volumes of useless traffic to occupy all the resources that could service legitimate traffic. When the traffic of a DoS attack comes from multiple sources, it is called a Distributed Denial of service (DDoS) attack. By using multiple attack sources, the power of a DDoS attack is amplified and the problem of defense is made more complicated. Yahoo, eBay, Amazon.com, E*Trade, ZDnet, Buy.com, the FBI, and several other Web sites have fell victim to DDoS attacks resulting in substantial damage and inconvenience. For the DoS attack to be detected, traffic features such as the “percentage of connections having same destination host and same service” and packet level features such as the “source bytes” and “percentage of packets with errors” are significant.

B. PROBE ATTACK:

Probe attacks are conducted by the attackers by using programs to automatically scan a large amount network IP addresses in order to find vulnerabilities that can be exploited. Once any vulnerability is found, the attackers can thus gain the access to the system and start to gather information without authorization. One of the most common probe attacks is called port scanning, which allows attackers to scan all ports on network hosts and discover which ones are available for connections. The popular scanning methods include TCP scanning, UDP scanning, SYN scanning, ACK scanning, FIN scanning, ICMP scanning, protocol scanning, and idle scanning. For example, portsweep attack discovers exploitable communication channels on remote hosts by systematically requesting connections to multiple TCP ports. Basic connection level features such as the

“duration of connection” and “source bytes” are significant while features like “number of files creations” and “number of files accessed” are not expected to provide information for detecting probes.

C. U2R (USER TO ROOT) ATTACK:

U2R attacks are the attacker pretends as a legitimate user of the system without authorization and then exploits the system’s vulnerabilities to get root access of that system. For example, the attacker may exploit a system’s vulnerabilities to gain root privileges and install a backdoor program onto a system for future access. The result may cause the system crash or make the system execute the attacker's program as if it is part of the system’s original programs. Features such as “number of file creations” and “number of shell prompts invoked,” can be used for detecting U2R attack.

D. R2L RREMOTE TO LOCAL) ATTACK:

R2L attacks are unauthorized attackers through networks gain local access as users of local machines. The attacks can be launched from anywhere on the internet. Once the attacker has access to the information systems, they can then exploit the machine’s vulnerabilities and cause serious damages such as stealing important data or crashing the information systems. Any service that needs password to access possibly becomes an attacked target. The R2L attacks are one of the most difficult to detect as they involve the network level and the host level features. We therefore select both the network level features such as the “duration of connection” and “service requested” and the host level features such as the “number of failed login attempts” among others for detecting R2L attacks.

To strengthen the network security there are various algorithms used in the intrusion detection such as encryption, VPN, firewall, etc. These algorithms are statically providing a protection over the network. But the intrusion detection is dynamic, so that we have to provide a dynamic protection over the network. Intrusion Detection system identifies and deals with the malicious use of computer and network resources.

II. LITERATURE SURVEY

In 1980, the intrusion detection system was initially suggested by J.P.Anderson [2]. Intrusion detection is categorized [1] into two, host based and network based. Host based IDS is installed on each system for monitoring of malicious activities locally. Network based IDS is installed on network elements like routers to monitor the network traffic.

Any organization that needs a security system must design according to the type of threats that it faces. DoS attack is the most important threat that any organization faces. Tao Peng et. al. [3] discusses different types of DoS attack and the prevention and detection method. Prevention methods are ingress filtering/egress filtering, router based packet filtering, source address validation enforcement (SAVE) protocol. Detection method includes DoS attack specific detection, anomaly based detection. Preventive and detection method deployed entirely depends on the needs of the organization.

Classification algorithms that classify the attacks must do it accurately. It must reduce false positives and false negatives to improve accuracy. Classification errors can be reduced by various algorithms like adaboost, bagging and wagging. Geoffrey I. Webb [4] proposes multiboosting technique, which is a combination of adaboost and wagging techniques, to reduce classification errors. When compared with other techniques multiboosting performs better in reducing classification errors in classification algorithms.

Attackers periodically change their DoS attack tactics to evade the security systems. So, it is necessary to continuously update the attack detection systems. Wenk Lee et. al. [5] proposed a data mining techniques: classification, meta learning, association rules, and frequent episodes that can learn rules for accurately capturing the behavior of intrusions and normal activity. These rules then can be used to update the security system to detect new attacks.

Dos attacks can cause heavy data load, system crashes, huge financial and infrastructure losses to an organization. So, it is necessary to detect the DoS attack as earlier as possible to minimize the losses. Yu Chen et al. [6] proposes the Distributed Change-point Detection (DCD) for the detection of DoS attacks. The method detects the attack by detecting the abrupt changes in the network traffic. The method achieves higher accuracy by reducing false positive alarms less than 1% of the total traffic.

Dos attacks can be detected in an intrusion detection system (IDS). IDS is of two types: Host intrusion detection system which are run on individual systems, Network intrusion detection systems which are run on

network elements like routers. Sapna S. Kaushik et. al. [7] discusses four types of attack: DoS attack, probe attack, User to Root (U2R) attack, Remote to Local (R2L) attack that can be detected in an intrusion detection system. It is found that probe attack is detected when an attacker adds some data into the data field sent by the original sender, R2L attack is detected when connection duration exceeds a maximum threshold and DoS attack is detected when a packet does not reach the destination thus service to the destination is denied. Attacks can be detected in a number of ways in an IDS which depends on the organizational needs

Most of the DoS attack detection systems detects the attack after much of the damage is done to the systems. It is necessary to detect DoS attack at an early stage so that enough precautionary measures can be done to minimize the damages. Hoai-Vu Nguyen et al. [8] proposes a k-NN classifier method which detects the DoS attack by classifying the network status into normal, pre-attack and attack. This method has many advantages like easy implementation, short time computation and high accuracy of 91% which is necessary for early detection of DoS attacks.

Domain Name System (DNS) provides name to address mapping services. Attackers perform 2 type of DoS attacks on DNS systems to deny internet services to the users. In first method, called direct DoS attack, attackers overwhelm the system by sending large number of packets. In the second method, called amplification attack, attackers increase the effect of normal DoS attack. Samaneh Rastegari et al. [9] proposes 3 neural network techniques: Back Propagation (BP), Radial Basis Function (RBF), Self Organizing Maps(SOM) to classify the DoS attacks. Back propagation neural network technique gives higher accuracy in detecting and classifying the DoS attacks than the other 2 techniques.

A DoS attack detection system must be able to differentiate the attack from the normal data traffic to perform accurately. It must reduce false positive and false negative earlier to reduce damages to the systems. Kejie Lu et. al. [10] proposes anti-DDoS technique which uses a machine learning algorithm to detect DDos attack. The algorithm is based on temporal and spatial characteristic of data traffic, both incoming and outgoing traffic. The detection method has many advantages like ability to detect attacks having high and low data rate, robust against time-varying patterns and deployable in large scale networks.

Any DoS attack detection system must satisfy most of the goals in order to effectively detect and prevent the DoS attacks. These goals are, the ability of the router to adjust the data traffic to reduce the attack impact on the downstream systems, reducing false positives and false negatives, adaptability of the security systems to the network environment. But most of the security system does not satisfy most of the goals. Laura Feinstein et al. [11] proposes a method that is based on the statistical properties of specific fields in the packet header. The method has advantages like reducing attack impact on the systems instantly, adaptability and filtering rules that can reduce false positives and false negatives.

L. Zenghui et al. [12] developed a frame work for constructing the intrusion detection models which captures the actual behavior of intrusions and normal activities. Even though this model can detect the probe attacks and U2R attacks, it fails to detect the DOS attacks and R2L attacks. To improve the detection rate, T. Lappas et. al. [13] presented a new idea on data mining can support IDS's by utilizing the bi-clustering tools. S. Sun et al. [14] achieves high detection rate by applying support vector clustering algorithm. But still some types of attacks cannot be detected, specifically it missed R2L attacks. C4.5 algorithm was used to improve the performance level. C4.5 classification tree is similar to Support vector machine (SVM) in the detection of R2L attacks. But the performance of C4.5 algorithm is high in comparison with SVM.

In proposed system intrusion detection is network based i.e. the data is collected from the network called Internet packets. After capturing the packet from the internet, features are extracted from the packet header. Protocol information are extracted from the packet header. After identifying this information, certain features would be used to classify the various attacks using the dataset, which contains minimum features by which an attack can be identified. Data mining techniques would be used for processing large amount of data. Data mining classification technique is used here to process the large number of packets and it is used to classify the various attacks. In the proposed system, c4.5 algorithm is used to classify the various attacks transmitted over the internet. To improve the efficiency of intrusion detection system and to reduce the false positives and classification errors, multiboosting algorithm is used.

III. PROPOSED SYSTEM

The proposed system uses mainly the c4.5 algorithm, for classifying the attacks, and the multiboosting technique will be used to reduce the false positives.

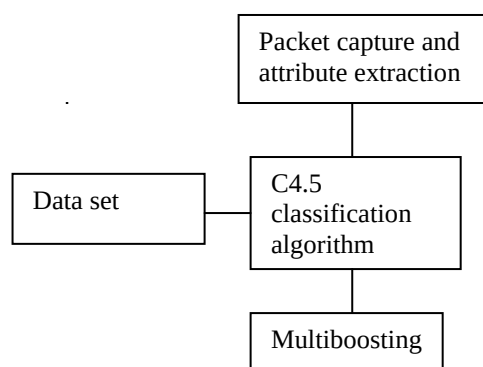


Figure 1 Overall architecture.

a. Packet capture and attribute extraction:

When a packet is sent, it passes through 7 layers of the network. At each layer, protocol adds its header to the packet coming from the above layer in the sender side. In the receiver side the packet is decapsulated, removing the header added at the sender, in the reverse order. In this module network packets will be captured using Jpcap(Java packet capture) software and packet attributes contained in the packet header will be extracted. Jpcap is a Java class package that allows Java applications to capture and/or send packets to the network.

b. Data set:

The data set contains predefined attack features and normal data packet features in a tabular form with each column representing the attribute value for the attack packet and normal packet and each row correspond to the each attack type and the normal data features. This dataset is used in C4.5 classification algorithm for classification of attacks.

c. C4.5 Classification algorithm:

C4.5 classification algorithm is a data mining technique which is based on tree construction. Each internal node consists of a test on an attribute. Leaf node contains the class label when the node cannot be further splitted.

Each node is splitted using the splitting attribute which is selected by computing information gain on the attributes in the dataset. The attribute which has highest information gain is considered as splitting attribute. When a node cannot be further splitted, the leaf node belongs to a separate class of attack.

d. Multiboosting algorithm:

Multiboosting is used to reduce false positives and the classification errors in the C4.5 classifier algorithm. Multiboosting combines the features of both Adaboost and wagging. Adaboost has superior bias(classification errors) and variance reduction capability while wagging has good variance reduction capability.

REFERENCES

- [1] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lily Suriani Affendey, "Intrusion Detection Using Data Mining Techniques", IEEE 2010.
- [2] J. P. Anderson, "Computer security threat monitoring and surveillance," technical report, James P. Anderson Co., Fort Washington, Pennsylvania, 1980.
- [3] Tao peng, Christopher leckie, and Kotagiri ramamohanarao , "Survey of Network-Based Defense Mechanisms Countering the DoS and DDoS Problems", *ACM Computing Surveys*, Vol. 39, No. 1, Article 3,2007.
- [4] Geoffrey I. Webb , "Multiboosting: A technique for combining Boosting and Wagging", Prepublication draft of paper published in *Machine Learning*, 40, 159–196, 2000.
- [5] Wenke Lee Salvatore, J. Stolfo Kui, W. Mok, "A Data Mining Framework for Building Intrusion Detection Models", *IEEE Symposium on Security and Privacy*, 1999.
- [6] Yu Chen, Kai Hwang, Wei-Shinn Ku, "Collaborative Detection of DDoS Attacks over Multiple Network Domains", *IEEE transactions on parallel and distributed systems* , TPDS-0228-0806.
- [7] Sapna. S. Kaushik, Dr. Prof. P. R. Deshmukh , "Detection of Attacks in an Intrusion Detection System", *International Journal of Computer Science and Information Technologies*, Vol. 2 (3), pg. 982-986, 2011.

- [8] Hoai-Vu Nguyen and Yongsun Choi, "Proactive detection of DDoS attacks utilizing k-NN classifier in an anti-DDoS framework", *International Journal of Electrical and Electronics Engineering*, 2010.
- [9] Samaneh Rastegari, M. Iqbal Saripan and Mohd Fadlee A. Rasid, "Detection of Denial of Service attacks against Domain Name System Using Neural Networks", *International Journal of Computer Science Issues*, Vol. 6, No. 1, 2009.
- [10] Kejie Lu , Dapeng Wu , Jieyan Fan , Sinisa Todorovic , Antonio Nucci, "Robust and efficient detection of DDoS attacks for large scale internet", Science Direct, Computer Networks, 5036-5056, 2007.
- [11] Laura Feinstein, Dan Schnackenberg, Ravindra Balupari, Darrell Kindred, "Statistical approaches to DDoS attack detection and response", *DARPA Information Survivability Conference and Exposition*, 2003.
- [12] L. Zenghui, L. Yingxu, "A Data Mining Framework for Building Intrusion Detection Models Based on IPv6," Proceedings of the 3rd International Conference and Workshops on Advances in Information Security and Assurance. Seoul, Korea, Springer- Verlag, 2009.
- [13] T. Lappas and K. P. , "Data Mining Techniques for (Network) Intrusion Detection System," January 2007.
- [14] S. Sun, Y. Wang, "A Weighted Support Vector Clustering Algorithm and its Application in Network Intrusion Detection," *etcs*, vol. 1, pp.352-355, 2009 First International Workshop on Education Technology and Computer Science, 2009.

ACKNOWLEDGMENT

At this moment, I would like to thank my supervisor Ms. K. Varalakshmi, our Head of Department Dr. E. Poovammal and Coordinator Mrs. C. Malathy and also other staff members those who supported me to carry out this work.