JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Implementation of Botnet Detection System using Machine Learning

Prof. Dr. Snehal Bhosale[1]
Prof. R. U. Shekokar[2]
Prerna Gajanan Honwalkar[3]

[1]*(Prof. Dr. Snehal Bhosale, RMD Sinhgad School of Engineering, RMDSTIC, Pune, India, Maharashtra)*
[2]*(Prof. R. U. Shekokar, RMD Sinhgad School of Engineering, RMDSTIC, Pune, India, Maharashtra)*
[3]*(Prerna Gajanan Honwalkar, RMD Sinhgad School of Engineering, RMDSTIC, Pune, India, Maharashtra)*

Abstract: *The field of information and computer security is rapidly developing in today's world and the number of security risks is continuously being explored every day. Botnet detection is an active area of research as no single technique is available that can detect the botnet attack very efficiently. The paper gives the system that can detect botnet activity based on traffic behavior analysis by classifying network traffic behavior using machine learning. Traffic behavior analysis methods do not depend on the packets payload, which means that they can work with encrypted network communication protocols. Network traffic information can usually be easily retrieved from various network devices without affecting significantly network performance or service availability.*

Keywords: Security, Botnet Activity Detection, Machine learning, Traffic behavior, TCP Packet, TCP Packet Filtering, Feed Forward Neural Network.

I. Introduction

Botnets have become one of the most malicious threats over the Internet among many of the security threats like servers, and networks compromised with malware, Trojan horses, phishing, ad-wares, deceive, ransomware, and viruses. Among the mentioned attacks Botnet comprises 80% of the attacks on the internet in the modern world. The botnet infects unprotected machines and keeps track of the communication with the command and control server to send and receive malicious commands. The attacker uses botnet to initiate dangerous attacks such as DDoS, fishing, data stealing, and spamming. One such powerful and harmful attack is the denial of service (DoS) attack.

II. Literature survey

A DDoS attack is exemplified by the direct attempt of attackers to prevent legitimate users from using a specific service. A botnet is a collection of zombie networks whose tendency is to propagate bot continuously [1]. A software program controls the computers and for specific purposes, known as bots. Bot attack is difficult to handle as botnet rapidly propagates in order to get off the detection process. Due to this dynamic behavior, the value of botnet information degrades quickly.

Botnet analysis is used to detect the type of attack. The botnet detection can be done using various machine learning algorithms. Author of [2] makes the recent survey of the majority of available techniques. Different machine learning techniques may give different results but the model with comparatively better result can be taken as the best-fitted model. Like Matija Stevanovic et al. [3] examines the effectiveness of detecting botnet network traffic using three methods that target protocols widely considered as the main carriers of botnet Command and Control (C&C) and attack traffic, i.e. TCP, UDP and DNS. The network traffic classification is done using Random Forests classifier. The method has been evaluated through the series of experiments using traffic traces originating from 40 different bot samples and diverse non-malicious applications. The evaluation indicates accurate and time-efficient classification of botnet traffic using mentioned technique.

Mohit Goyal et al. [4] perform behavioral analysis to detect the bot-nets using http based C&C Servers in the IOT environment. The main features that are used for botnet detection are Duration gap in each request and Variation in number of IPs attached to particular URL. The presence of Malwares can be detected using supervised machine learning algorithms like Logistic Regression, SVM (Support Vector Machine) and Artificial Neural Network (ANN). Among which the neural networks outperformed all other methods.

Lakshya Mathurb et al. [5] analyzed and experimented with five different classification techniques to find out two most suitable techniques for detection of botnet. Author used Logistic Regression, Multi Class classifier, Random SubSpace, Randomizable Filtered Classifier and Random Committee. Among which Logistic Regression and Multi Class classifier gives higher accuracy. The features used for detection are Ts (Flow start time), Te (Flow end time), Td (Flow duration), Sa (Source IP address), Da (Destination IP address), Sp (Source port), Dp (Destination port), Pr (Protocol), Ra/flg (Flg Flags), Ipkt (Input Packets), ibyt (Input Bytes).The CTU-13 Dataset and ISOT Dataset is used for the experimentation.

Authors [6] proposed new features to distinguish C&C channels from benign traffic. Detection method uses a random forest classifier implemented over Apache Spark, a Big Data processing framework with more than 99% of accuracy. The proposed features can be extracted before the communication end, which enables a premature response.

By studying the mentioned technique, we have implemented a system using machine learning model to detect botnets with better precision and reduce false positives by studying existing work done in the botnet detection area. The proposed system can be able to analyze the traffic coming from multiple IP in addition with packet length, delta time and packet count. The system detect botnet activity based on traffic behavior analysis by classifying network traffic behavior using machine learning technique like feed forward neural network. The algorithm helps to decide which source packets are to accept or which is to decline.

III. Proposed System

A botnet is a number of internet-connected devices, each of which is running one or more bots used to perform Distributed Denial-of-Service (DDoS) attacks. Such attacks are typically attempts to exhaust victim’s bandwidth or disrupt legitimate users’ access to services. Traditional architecture of internet is vulnerable to DDoS attacks and it provides an opportunity to an attacker to gain access to a large number of compromised computers by manipulating their vulnerabilities to set up attack networks or Botnets.

The proposed system helps to detect botnet activity by classifying network traffic behavior using machine learning. The system is able to detect spoofing attack, header attack and attack with multiple IP with by analyzing certain features like Average Packet Length, Delta Time, Packet count and Hardware address. Figure 1 shows the architecture of the proposed system followed by detailed working of the system.

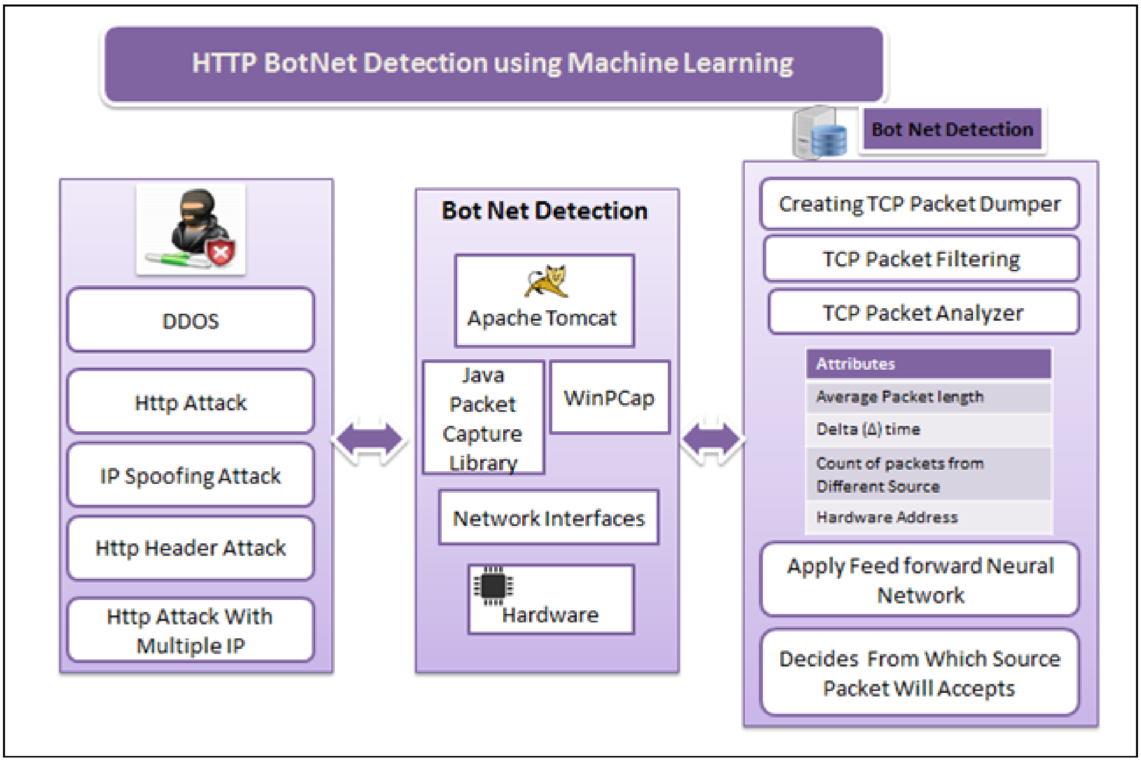

HTTP BotNet Detection using Machine Learning
Bot Net Detection
DDOS
Http Attack
IP Spoofing Attack
Http Header Attack
Http Attack With Multiple IP
Bot Net Detection
Apache Tomcat
Java Packet Capture Library
WinPCap
Network Interfaces
Hardware
Creating TCP Packet Dumper
TCP Packet Filtering
TCP Packet Analyzer
Attributes
Average Packet length
Delta (Δ) time
Count of packets from Different Source
Hardware Address
Apply Feed forward Neural Network
Decides From Which Source Packet Will Accepts

Figure 1: System Architecture

This project makes use of JPCap library to capture packets and analyze and filter by its type. The H/w mentioned in the architecture diagram is Networks interface card. Application consist of following modules.

1. **Creating TCP Packet Dumper.**

- Using JpCap library we dump all the TCP requests (packets) into the dump file. Dump file size is maximum 2048KB.
- Dump file is created on the basis following attribute:
- packetIndex, Timeval, sourceAddress, sourceHardwareAddress, sourcePort, destinationAddress, destinationHardwareAddress, destinationPort, sequenceNumber, acknowledgementNumber, flagsPresent, packetPriority, packetLength, offset, TimeToLive.

2. **Apply TCP Packet Filtering.**

   We can apply the TCP packet filter on the dump files only for the incoming TCP request.

3. **Web Application Firewall**

   a) **TCP Packet Analyzer**

   It performs the analysis on dump files and gives the following attribute to Feed Forward Neural Network.

   - Hardware Address
   - Count of Packet from the Different IP
   - Average Length of TCP packet
   - Average difference in consecutive requests (Δ time).
   - Request Ratio
   - Number of self-packets/total packets

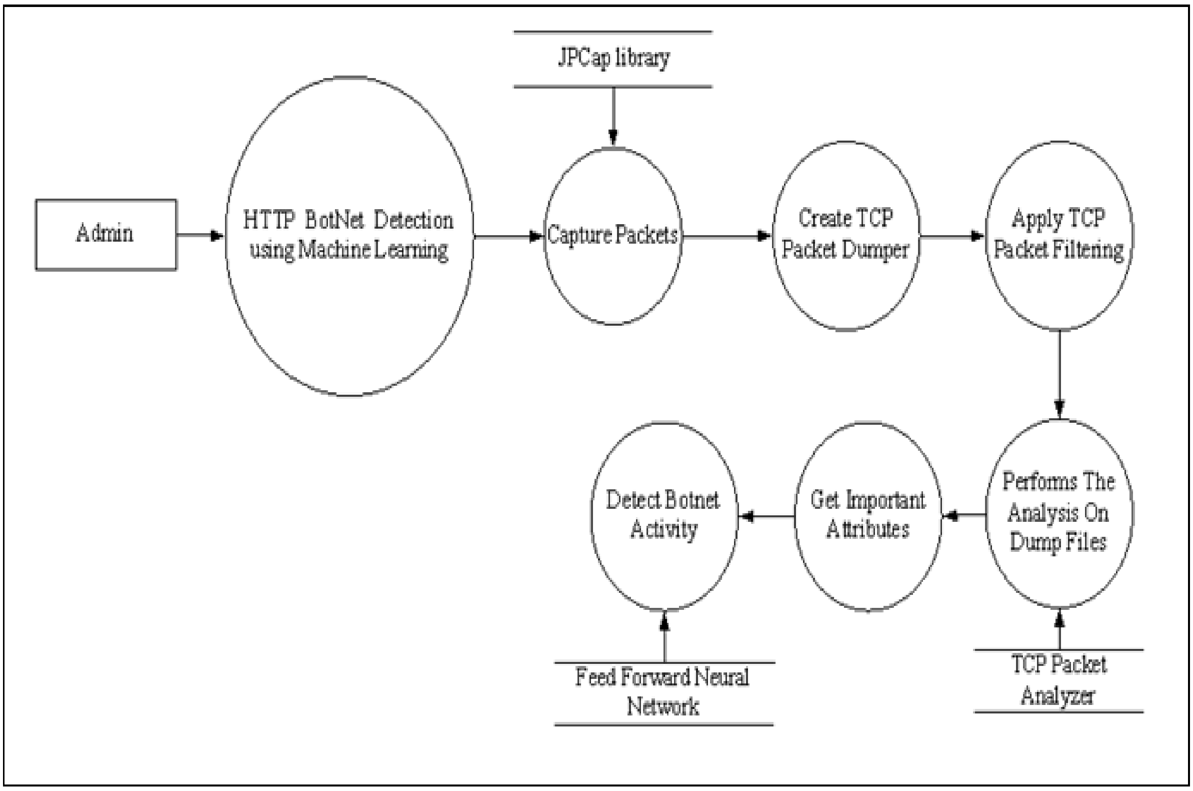
   b) **Apply Feed Forward Neural Network**

A Feed Forward Neural Network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers typically consist of Convolutional layers, pooling layers, fully connected layers and normalization layers. Feed Forward Neural Network will be used to train the data analytics engine on the TCP Packet Analyzer attributes like count of packets coming from the particular IP, time stamp and length of packet etc. for recognizing the request getting from the bot or not.

4. Detecting Bot

On the basis of Feed Forward Neural Network result, the decision is taken to allow the packets from the particular machine or not which can be Bot. By using Feed Forward Neural Network we are able to successfully detect botnet activity with high accuracy. Figure 2 shows the overall dataflow of the proposed system.

Figure 2:


JPCap library
Admin
HTTP BotNet Detection using Machine Learning
Capture Packets
Create TCP Packet Dumper
Apply TCP Packet Filtering
Performs The Analysis On Dump Files
Get Important Attributes
Detect Botnet Activity
Feed Forward Neural Network
TCP Packet Analyzer

Dataflow Diagram

IV. Algorithm Used

The feed forward neural network is the simplest type which consists of a set of processing elements called “neurons”. The information moves in only one direction, forward, from the input layer, through the hidden layer and to the output layer. There are no cycles or loops in the network. An example of a simple FNN with a single hidden layer is shown in following figure 3. As shown, each neuron computes the sum of the inputs weight at the presence of a bias and passes this sum through an activation function (like sigmoid function) so that the output is obtained.

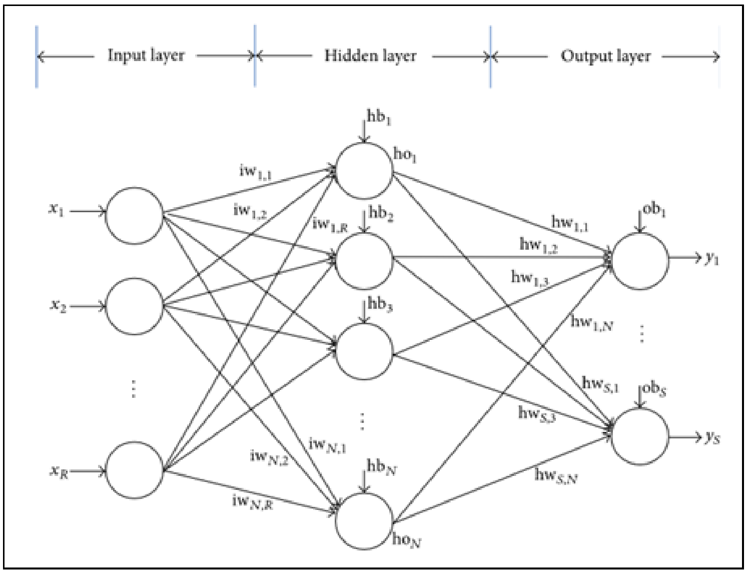


Figure 3: Feed Forward Neural Network

V. Experimental Details

We have calculated/detected the average of packet length and average of difference in time for different IP addresses like 192.168.0.1, 192.168.0.100 etc.

Table 1: Traffic Analysis for different IP address

	Values		
Row Labels	**Sum of Packet Count**	**Sum of Average Delta**	**Sum of Average Length**
08:9e:01:d7:86:20	1	0	290
20:cf:30:cc:03:b4	16	0.771604	100.1875
30:b5:c2:9d:10:1a	15213	0.013158	961.6003418
48:5a:b6:99:71:61	1	0	290
74:27:ea:aa:4f:48	3	40.553955	190.6666667
Grand Total	**15234**	**41.338717**	**1832.454508**

The graphical representation for the traffic for different IP addresses is given in figure 3 below.

Source MAC	Packet Count
08:9e:01:d7:86:20	1
74:27:ea:aa:4f:48	3
30:b5:c2:9d:10:1a	15213
20:cf:30:cc:03:b4	16

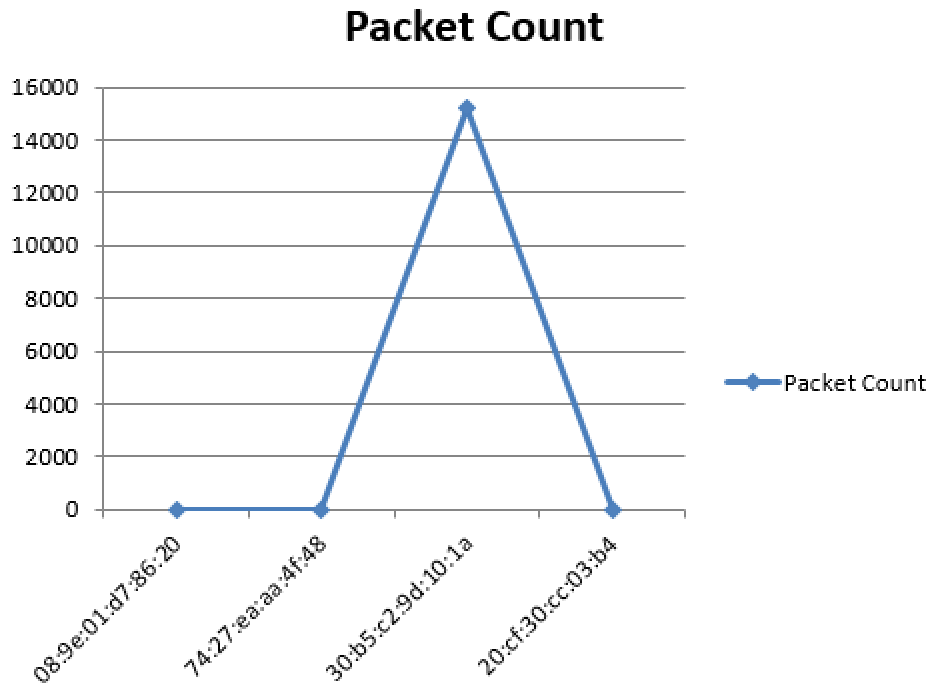


Figure 4: Graph for Packet length & Average Diff in time for different IP address

VI. Conclusion

The results can help us know the potential of applying Machine Learning algorithms for analyzing network traffic behavior. Our proposed model allows detecting bot activity in both the command and control and attack phases based on the observation of its network flow characteristics for specific time intervals. We emphasize the detection in the command and control phase because we would like to detect the presence of a bot early before any malicious activities can be performed, and we use the concept of time intervals to limit the duration we would have to observe any particular flow before we may raise our suspicions about the nature of the traffic.

References

[1] Anchit Bijalwan,"Botnet Forensic Analysis Using Machine Learning",Security and Communication Networks / 2020 / ArticleVolume 2020 |Article ID 9302318 | https://doi.org/10.1155/2020/9302318

[2] Prof. Chaitali Bhalerao1, Riddhika Kulkarni2, Swapnil Bobade3, Krutika Bharekar4, Vaishnavi Zadbuke5,"A Survey on HTTP BotNet Detection Techniques using Machine Learning",JETIR, Volume 8 | Issue 2 | Year February-2021.

[3] Matija Stevanovic; Jens Myrup Pedersen,"An analysis of network traffic classification for botnet detection", 2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA).

[4] Mohit Goyal; Ipsit Sahoo; G. Geethakumari,"HTTP Botnet Detection in IOT Devices using Network Traffic Analysis", 2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC).

[5] Lakshya Mathurb, Mayank Rahejab, Prachi Ahlawat,"Botnet Detection via mining of network traffic flow ",International Conference on Computational Intelligence and Data Science (ICCIDS 2018).

[6] Paulo Angelo Alves Resende, André Costa Drummond ,"HTTP and contact-based features for Botnet detection", 2018 John Wiley & Sons.