Volume 6, Issue 7, July 2016 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Analyzing the Security Issues in Dual Stacking

Abhishek Singh Charak, Jasbir Singh
Department of Computer Science & I.T, University of Jammu,
Jammu and Kashmir, India

Abstract— As the IPv4 addressing space is exhausting quickly, resulting in the internet user to migrate from IPv4 to IPv6, but the migration of such a huge systems is hard to accomplish. Therefore several techniques are needed which ascertain smooth translation of IPv4 to IPv6 and vice-versa. The solution to this is categories into three types, i.e. Dual Stack, Tunneling and Translation., and these could be misused to trespass the firewall system or intrusion detection system because presently most of these are running on IPv4 and thus ipv6 malicious packet will trespass the security. In this paper, a methodology have shown which by using naive bayes classifier, which classifies the normal hexadecimal strings and malicious hexadecimal strings in transmitted packets which are transmitted through the firewall, the "Naive Bayes Classifier" have shown very good performance in classifying the malicious and normal packet based on text categorization.

Keywords— Naive Bayes Classifier, Cisco routers and Firewall, FileZilla, Ubuntu

I. INTRODUCTION

A malicious file is defined as a program that performs a undesired function, such as compromising a system's security or stealing a confidential information without user permission. Large no. of malicious programs are created every day and most cannot be accurately detected until a signature is not generated for them. For that time being, system protected on signature-based algorithm are vulnerable to attacks. In 2015, there were 1,966,324 registered notifications about attempted malware infections that aimed to steal money via online access to bank accounts and also ransomware programs were detected on 753,684 computers of unique users [1].

One of the problem faced by the virus community is to come up with a method for detecting new malicious files that are not detected yet and not present in the malicious category [2]. By using naive bayes classifier we have tried to scan the file which is transmitted through the network. The naive bayes classifier classifies files based on its training over the provided data set.

The naive bayes classifier can be used to classify different kinds of transmitted packet. A packet consisting of different hexadecimal strings that are being classified by the filter. In this paper the packets of file is classified into two classes. i.e. malicious class or normal class. The method discussed here use supervised forms of machine learning. That means the filter that is formed, first needs to be trained by previously classified hexadecimal strings provided by the user. This is so because naive bayes classifier cannot know in advance that which packet belongs to which class. In this method we treated each executable’s features as a text document and classification is based on it. The main assumption in this approach is that the binaries contain similar features such as signatures, machine instructions, etc.

As IPv6 comes to replace the depleted IPv4 and the IPv6 comes with more number of network address bits from 32 bits (IPv4) to 128 bits. This has evolve the need for more securing the IPv6 based networks [3]. Bayesian text classification has been used for decades, and it has remained relevant throughout years. Naive bayes classifier have outperformed the newer and more sophisticated method been developed.

There is difficulty in detecting the Internet worm and remote attack scripts which appear as the new danger, by applying the data mining in the intrusion detection system. But, according to the latest result of study, there was the good performance in the virus and normal programming classification by using the malicious class or normal class. The method naive bayesian classifier [4].

Naive bayesian classifier as one of algorithm which shows the most excellent performance in the text categorization, modified all the malicious codes and abnormal codes to a hexadecimal value and classified by treating like the general text [5].

II. LITERATURE REVIEW

As the development of IPv6 gaining importance, the issues regarding the protocols implementation emerged as one of the important aspect. The IPv6 introduces vulnerabilities in addition to those inherent in ipv4, In [6] the author presents an analysis of network attacks that are common in IPv4 and makes a comparative analysis of how these attacks may impact the IPv6 network. The paper also establishes guidelines and principles for mitigating these attacks. In [7] the author have shown most security related products (firewall/IDS/IPS) have not been programmed to inspect IPv6 packets in depth thus can allow malicious packets to pass through by taking advantage on the encapsulation of IPSEC in IPv6, some tools have been developed to do so like “VoodooNet” or “v00d00n3t”.In [8] the author have demonstrated that in

IPv4/IPv6 transition mechanism, there are a dual stack, tunneling, and translation. Among them, tunneling may be misused in order for an malicious code to avoid a firewall system or intrusion detection system. thus they had given a methodology which classifies the normal traffic and malicious traffic in IPv4/IPv6 tunneling environment, by using naive bayes classifier. As the worms over network includes signatures or machine instructions these Network packet can be supposed as general documents. Accordingly, 'naive bayes classifier' can be utilized for the network traffic analysis. In [9] the author have shown Due to shifting from IPv4 to IPv6, has raised the necessity for effective and efficient malware detection techniques for ipv6 networks. Because of the evolvable and polymorphic malware, current malware detection technology cannot cope with the exponential growth of malwares. For malware detection in dual stack IPv4/IPv6 networks, the proposed integrated approach consist of three modules, the first module is a malware portable executable (PE) file analyzer which generates a features of a malware from its executable file; the second module is a feature selector which selects the most important and informative features; and third module is an adapted evolving classification function that uses genetic algorithm to detect the malware in evolvable manner.

III. RESEARCH METHODOLOGY

In this paper a transmission mechanism is implemented in GNS3 (Graphical Network Simulator), using Cisco routers, the packets are generated by using FileZilla FTP software installed on two different machines over the network and the filtering is done by using Cisco ASA firewall and naive bayes classifier. The Step wise implementation of the system is as shown:-

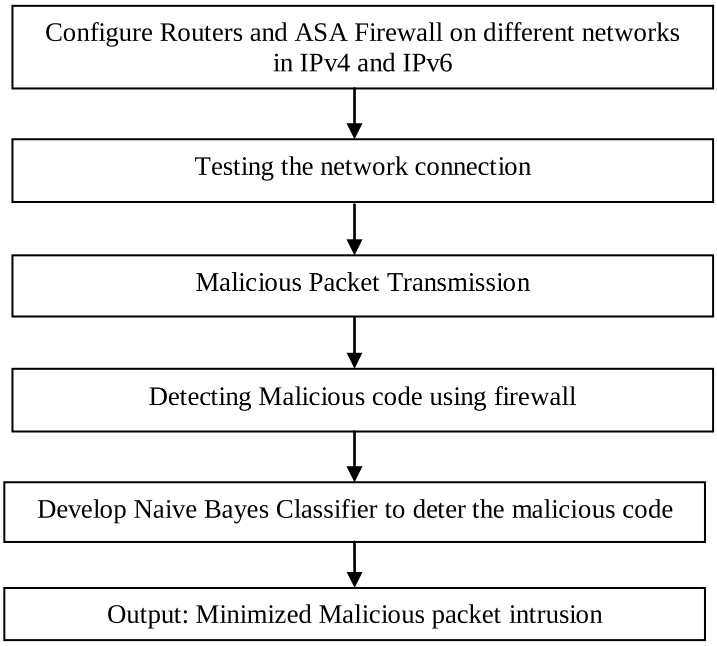
Configure Routers and ASA Firewall on different networks in IPv4 and IPv6
Testing the network connection
Malicious Packet Transmission
Detecting Malicious code using firewall
Develop Naive Bayes Classifier to deter the malicious code
Output: Minimized Malicious packet intrusion

Figure: Proposed framework

Naive Bayes Classifier

Naive bayes classifier is a statistical algorithm which is widely used for the text categorization. It can classify the text by using the statistical information provided in the training document, i.e. to which class the text belong. The classifier is formed as each Instance x is combined with the attribute value, and is applied to the learning process that the target function f(x) can acquire the value from finite set V. When the target function learning set is given and the new instances which are shown in the line of attribute values *<a1,a2,...an>* , naive bayesian classifier predicts the target value and classifies the new instance.

Bayesian approach to classify the new instance and gives the highest probability value , when the attribute value *a1,a2,...an* which explains the instance is given:

$$v_{MAP} = \underset{v_j \in V}{\arg\max}\ P(vj|a_1, a_2, \cdots, a_n)$$

Here, the following formula can be expressed by using the Bayesian Theory.

$$v_{MAP} = \underset{v_j \in V}{\arg\max} \frac{P(a_1, a_2, \cdots, a_n | vj) P(vj)}{P(a_1, a_2, \cdots, a_n)}$$

$$= \underset{v_j \in V}{\arg\max}\ P(a_1, a_2, \cdots, a_n | vj) P(vj)$$

Here, *P(a1,a2,...an|vj)* and *P(vj)* can be calculated based on the training set. As each target value calculates the frequency which occurs from the learning data set, *P(vj)* can be easily estimated. But, it is not easy to estimate *P(a1,a2,...an|vj)* ,unless we have a big learning data set.

© 2016, IJARCSSE All Rights Reserved

To calculate the authentic estimate, the number that increases the no. of instances by possible target values must be estimated. Because naive bayes classifier is based on the theory that the attribute values are conditionally independent to the provided target value, the given instance's target value can be indicated by multiplying the joint probability of *a1,a2,… a*n by each attribute's probability.

That is, the formula is *P(a1,a2,…an|vj) = P(ai|vj)* When it is substituted to ***vMap.*** Naive Bayes Classifier can be calculated.

Naive Bayes Classifier :

$$v_{NB} = \underset{v_j \in V}{\arg\max} \; P(v_j) \prod_i P(a_i | v_j)$$

In here, *VNB* indicates the target value and result value by Naive Bayes Classifier. *P(ai|vj)* is estimated from the learning set in NBC is the number which multiplies the number of attribute value by the number of target values. Because it is the probability mode, poor performance may be shown if there is no sufficiently suitable data.

IV. RESULTS AND ANALYSIS

The naive bayes classifier for the detection of malicious packet transmitting over the network could be implemented on the following topology's ubuntu client machine or it can be implemented on the DMZ system over the network as shown in figure:

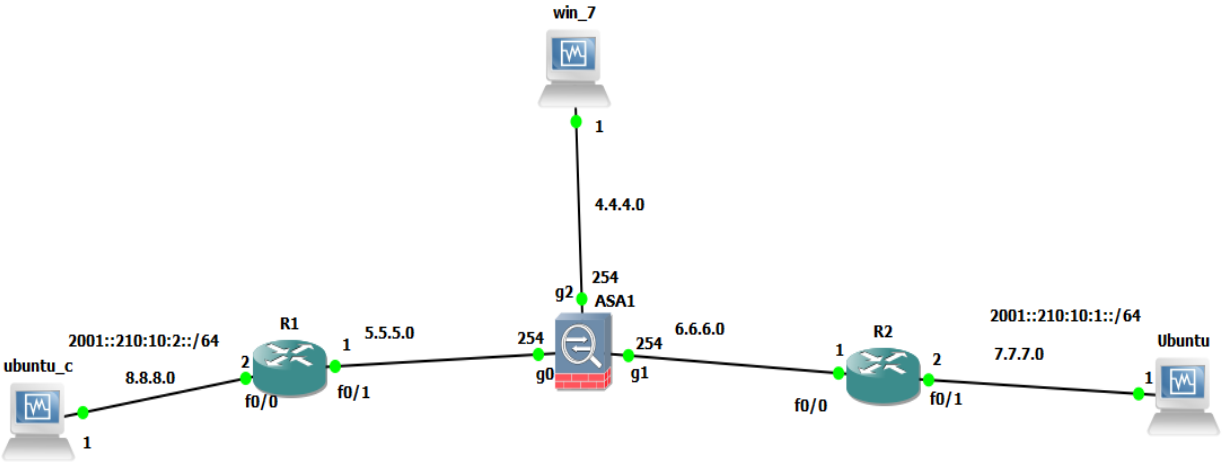

Figure: Network topology for dual stacking implemented on GNS3 network simulator

The Naive Bayes classifier is trained on Matlab r2015a with two datasets, the first dataset contains the strings of malicious file and benign file and the second dataset contains the strings of signature files taken from the Macafee antivirus and benign files, here the malicious strings are denoted by 0 and the benign strings are denoted by 1, the strings in these two files are converted to decimal forms and then trained on the naive bayes classifier, the output of the naive bayes classifier will be formed by first dividing the dataset into two xlsx sheets i.e. the first sheet contains the converted strings and the second sheet determines the type of strings '0' or '1', and these sheets will be input to the confusionmat which will return the confusion matrix determined by the known and predicted groups in first and second sheet respectively, the no. of rows of first and second xlsx sheet must be same to get the results.

The result will be a square matrix known to be confusion matrix with size equal to the no. of distinct element in first and second xlsx sheets. i.e. 0 and 1 in this case. Confusion matrix is a count of observations known to be in one group but taken in another group i.e. it is the false positive and false negative rate.

The resultant matrix for malicious and benign strings 344 in number is:

```
cMat1 =


   111     2
   209     4
```

Here the matrix depicts that out of 111 benign strings 2 are misinterpreted as malicious files i.e. 98.2% true positive and out of 209 malicious strings 4 are misinterpreted as benign strings i.e. 98.1% false positive.

The resultant matrix for signature and benign string 364 in number:

```
cMatSign =


   132     0
   202     3
```

Here out of 132 benign strings 0 string is misinterpreted as malicious file that is 100% true positive rate and out of 202 malicious strings 3 are misinterpreted as benign file i.e. 98.5% false positive rate.

V. CONCLUSION AND FUTURE SCOPE

Migration from IPv4 to IPv6 has evolved a need for more securing the networks which are running on the IPv4 network and communicate with IPv6 with Dual Stacking, Tunneling and NAT. In this study the normal and malicious packet are classified based on data transmission. As the naive bayes classifier can classifies the text without classifying the general text or Encapsulated text. Naive bayes classifier can analyze the text which may or may not be IPv6, and give a deduction of whether it belong to a malicious class or benign class. We are working on a software firewall containing the naive bayes classifier to deduce the maliciousness of the packets and this will be implemented on a system to provide a additional layer of security for the network.

REFERENCES

[1] Kaspersky Security Bulletin 2015 "Overall Statistics For 2015".

[2] Matthew G. Schultz , Data Mining Methods for Detection of New Malicious Executables". ids.cs.columbia.edu/sites/default/files/binaryeval-ieeesp01.pdf.

[3] Niranjan Ravi1, Muppidathi @ Saravanan A2, Manoranjan Periyasamy3 . " Implementation of IPv6/IPv4 Dual-Stack Transition Mechanism ". International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 11, November 2014.

[4] Study on new malicious code detection mechanism using 'Naive Bayes Classifier', 2003

[5] Matthew G. Schultz, eleazar Eskin, Erez Zadok, and Salvatore J. Stolfo. "Data Mining Methods for Detection of New Malicious Executables." To appear in IEEE Symposium on Security and Privacy, May 2001.

[6] Junaid Latief Shah, Javed Parvez ," Security Issues in Next Generation IP and Migration Networks "IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 1, Ver. III (Jan – Feb. 2015), PP 13-18 www.iosrjournals.org

[7] Jalan Gadong,Brunei Darussalam, "Information Technology Protective Security Services".

[8] Kyu-Cheol Oh, Ki-Ho Lee, You-Jae Won," Malicious Code Detection Method over IPv4/IPv6 Tunneling Using Naive Bayesian Classifier ", Proceedings of the 5th WSEAS Int. Conference on Information Security and Privacy, Venice, Italy, November 20-22, 2006

[9] Altyeb Altaher*, Sureswaran Ramadass and Ammar ALmomani, "An intelligent approach for malware detection in dual stack IPv4/IPV6 networks " International Journal of Physical Sciences Vol. 7(10), pp. 1607-1612, 2 March, 2012.

[10] Kevin P. Murphy ." Denotes advanced material that may be skipped on a first reading." https://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall06/reading/NB.pdf

[11] Robert Moskovitch · Dima Stopel · Clint Feher ·Nir Nissim · Nathalie Japkowicz · Yuval Elovici." Unknown malcode detection and the imbalance problem". Published online: 11 July 2009 © Springer-Verlag France 2009