Stochastic Modelling and Computational Sciences

AN EFFICIENT APPROACH FOR DETECTING AND PREVENTING DOS/DDOS ATTACK ON WEB SERVER

Chiragkumar Dilipbhai Patel
Lecture, Computer Engineering Department, K. D. Polytechnic, Patan, Gujarat, India
patelchiraag@gmail.com

ABSTRACT
In today's world the growth of Internet is increase tremendously. With the increase in internet a greater number of people connects to internet, so more number of chances of attacks is also happen. To defend various kind of attack is challenging task in current internet. Out of many attacks DoS/DDoS attack are perform measure role to slow down the internet (Web Server). User does not get resources at right time. Users have to wait for resources. Many techniques are introduced like machine learning, artificial intelligence base technique to defend this attack. In this paper we focus on naive bayes data mining classification with the use of KDD Dataset for detection of DoS/DDoS attack. We also develop threshold base mechanism with the use of CAPTCHA technique to prevent the web server after detecting DoS/DDoS attack.

Keywords: DoS/DDoS, KDD Dataset, CAPTCHA, Naive bayes, Threshold base algorithm

1. INTRODUCTION
Web Server is a part of client server architecture in internet. Client send request to web server and server check its database and give reply to the client according the request of client. Client use http protocol for sending the request [1]. Sever use various kind of technology to reply the client request like script, database services and many more. Different kind of Web server is available on internet to provide different kind of services like banking, communication etc.

Now a day's different kind of web server are use to serve different kind of client request. Attackers are interested to slow down the server or close the service of server. Due to attacker client are suffering to get them reply back within specific time. Attackers send more number of forge requests (DOS attack) to server, so server becomes busy. Attackers also use tools to launch such type of attacks [3]. Tools are easily available on internet. One of the solutions to prevent from this attack is to use intrusion detection system. IDS were first introduced by James Anderson in 1980. IDS are powerful tool for intrusion detection. Two kinds of IDS are used to detect and prevent the kind of attacks in networks. One is hosting based IDS. In HIDS, the IDS monitor every activity of the host and make one file like logs and many more. IDS use this logs file to detect the intrusion. HIDS is located at different host in network. Second kind of IDS in Network based IDS [8]. NIDS collect the live traffic in network and analyze it for intrusion detection.

IDS use different kind of technique to detect the attack. According the technique use, it classifies in two categories. (1) Anomaly based IDS: in this IDS system it continuously monitors the system behavior. If the behavior of the system is changed from the normal behavior, system is considered under attack otherwise system is in normal state. (2) Signature Based IDS: Signature based IDS system use database in which various kind of pre-define pattern of attacks are stored. These IDS collect the traffic and compare pattern with the stored pattern. If it match then classify as an attack otherwise treat as normal traffic [1, 8]. Antivirus system are working like signature based IDS. Signature based IDS do not find the latest attack. For this continuous updating of database is required.

In this paper we use KDD dataset for traffic. KDD dataset is widely used in network related research. The KDD Cup 1999 dataset is generated from the 1998 DARPA intrusion detection evaluation program. The KDD Cup 1999 dataset is used by MIT Lincoln labs [?]. They acknowledge and approved this dataset for different engineers and specialists. This dataset has 41 features. Out of these 41 features, 1 to 9 is utilized as basic features

of packet. 10 to 22 utilized as content features, 23 to 31 are utilized as traffic features with 2 seconds for traffic window and 32 to 41 for host primarily based features. Each packet contains 100 bytes [8].

We are using Naive Bayes data mining classification algorithm on KDD dataset. Naive Bayes is based on bayes theorem. Bayes theory calculates posterior prob- ability using prior probability and likely hood [9]. Naive bayes find posterior probability of every attribute of KDD dataset and classify in two class either normal or attack. KDD dataset have total 22 different attacks.

We are using threshold base mechanism after classification. Threshold base algorithm is use to detect and prevent the DOS attack. If the calculate value is cross the predefine threshold, then detect as an attack. Otherwise consider as normal packet. If attack is identified then CAPTCHA mechanism is used to identify the traffic is automated tool generated or normal. If the attacker does not pass the CAPTCHA test, then IP address of that attacker system is blocked [5].

In this work we present architecture to detect and prevent DoS/DDoS attack on web server using Naive bayes and threshold base algorithm. Rest of paper is organized as follows. Section 2 described the related work done by various researchers, Section 3 give details idea of proposed system, in section 4 simulation and its result and section 5 conclude the paper.

2. LITERATURE SURVEY

Vijay Katkar et al. [1] present architecture of offline signature-based network intrusion detection system for detection of DoS/DDoS against HTTP servers using distributed processing and Naive Bayesian classifier. But this system is offline; it will not work on real time. This system detects only known attack.

Samad Kolahi et al. [2] present mechanism, which use various Linux tools for UDP DDoS attack. They focus on TCP throughputs, round-trip time and CPU utilization before and during the attack on Web Server with Linux Ubuntu 13. They use Access Control Lists, Threshold Limit, IP Verify and Load Balancing parameter for detection of DoS/DDoS.

Abdulaziz Aborujilah et al. [3] define mechanism, which focus on studying the effects of (DoS) attack on CPU power performance and in network bandwidth. This system uses; three parameters for evolution. The parameters are packet limit, sending delay, sending duration.

Indu Mandwi et al. [4] Addressed on Jamming attacks in wireless networks. Authors used encryption-based system to prevent the DoS attacks. Author's de- eloped system, which contain three schemes that prevent real-time packet classic- fiction by combining cryptographic primitives with physical layer attributes.

P. Pandiaraja et al. [6] developed proxy-based system. Proxy is protected, filter, monitoring the application against DDoS attacks. Author used Hidden Semi- Markov Model to describe the time varying traffic behaviors and special behavior of the traffic. In this model authors change the http header and used encryption and decryption-based threshold base algorithm to find the DoS attacks.

Pooja Bhoria et al. [7] statistically analyzed NSL KDD dataset with 6 folds cross validation technique. Authors apply various feature selection technique to improve the IDS performance. They increased the classification accuracy and decrease the compilation time. Feature selection techniques are useful to determine the rule for finding the various DoS attacks.

G. V. Nadiammai et al. [8] proposed efficient data adapted decision Tree tech- nique. This scheme improves performance of intrusion detection system. The proposed Efficient Data Adapted Decision Tree technique reduce the actual size of dataset and helps the DBA to analyze the real time traffic with a smaller number of false alarm rate.

Dr. Saurabh Mukherjee et al. [9] developed Feature Vitality Based Reduction Method (FVBRM). The feature reduction is performed on 41 attributes to get 24 using FVBRM on NSL KDD dataset. The developed model is

comparing with existing scheme like Correlation based feature selection (CFS), Information Gain (IG) and Gain Ratio (GR). The results are compared using identified performance matrices like classification accuracy, recall, sensitivity.

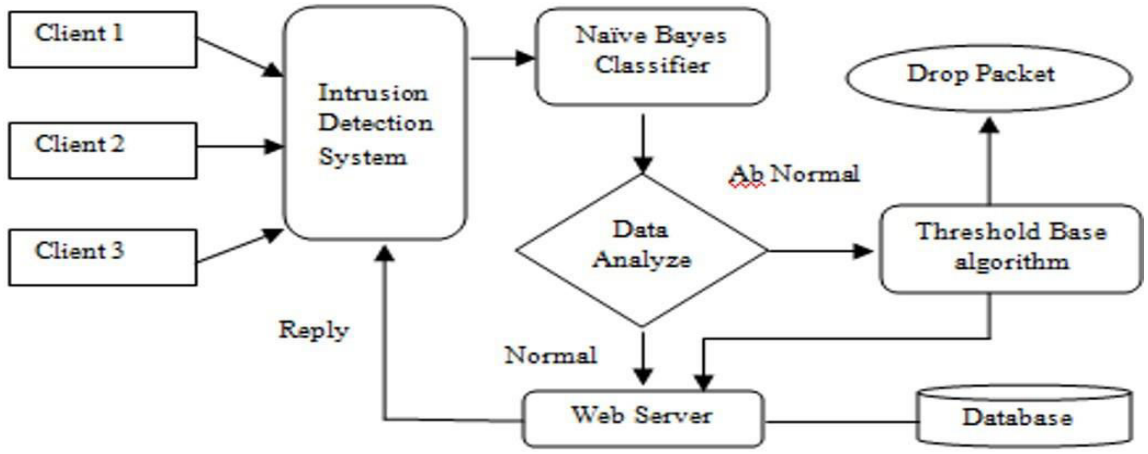

Fig.1: System architecture

3. PROPOSED ARCHITECTURE

As shown in Fig. 1. System architecture the client sent request to the web server. We put the intrusion detection system before the web server which checks all the incoming traffic using naive bays classifier algorithm. The classified data are analyzed. If any abnormal data is found it give to the threshold base algorithm. If data is cross the limit of threshold, then packet is drop. Threshold base algorithm is also used CAPTCHA mechanism to identify automated tool generated request. Otherwise, request is passing to the web server. Finally, web server gives reply to client.

Here, we propose System architecture to detect and mitigate the DoS/DDoS Attack against web server. System is divided into three phases.

(1) Analyze the Incoming Traffic to the web server.

(2) Differentiate DoS/DDoS attack traffic from normal traffic using Naïve Bayesian classification.

(3) Prevent the effect of DoS/DDoS attack using threshold base algorithm and CAPTCHA mechanism.

First phase we have to analyze the incoming traffic coming to the web server. This we can do by collection all the traffic and made one dataset. This dataset is use for analysis purpose. We are using KDD dataset for analysis purpose.

Second phase is to differentiate attack traffic from normal traffic. In today's days, attackers use BOTNET machines for attack. So, attackers use thousands of such machines for attack against victim (web server). So, my focus is to differentiate that traffic from normal traffic. Here we used KDD Dataset which is part of DARPA. To differentiate traffic, we use naive Bayesian classification technique on KDD dataset.

Third phase is to mitigate the effect of DoS/DDoS attack. After identifying and detecting the attack, next step is to mitigate the effect of this attack. For this we use threshold base mechanism. So, my focus is to increase the availability of web server for legal users by mitigate the effect of DoS/DDoS attack.

```
=== Summary ===

Correctly Classified Instances      205693               95.1291 %
Incorrectly Classified Instances     10532                4.8709 %
Kappa statistic                          0.9037
Mean absolute error                      0.0043
Root mean squared error                  0.0634
Relative absolute error                  9.968  %
Root relative squared error             43.4058 %
Coverage of cases (0.95 level)          95.6911 %
Mean rel. region size (0.95 level)       4.4203 %
Total Number of Instances           216225
```

Fig. 2: Experiment result

4. EXPERIMENT AND RESULTS

We use windows 8.1 with weka tool version 3.7 installed in it. In first phase we use KDD dataset in weka tool. We select Naive bayes classification algorithm to classify the data without any feature selection and selection attributes. This time Naive bayes give accuracy of 95.12 % as shown in Fig. 2.

Now in weka tools we use select attributes facility to improve the classification accuracy. We use CfsSubsetEval attribute evaluator with Best First search method. Weka tool give us best 14 attributes out of total 41 attributes of KDD dataset as given in Table 1.

Table 1: List of 14 attributes

Attribute Number	Attribute Name	Attribute Number	Attribute Name
2	Protocol Type	23	Count
3	Service	24	SRV Count
5	SRC Bytes	30	Diff SRV Rate
6	DST Bytes	36	DST Host Same SRC Port Rate
7	Land	37	DST Host SRV Diff Host Rate
8	Wrong Fragment	40	DST Host Rerror Rate
12	Logged In		
14	L Root Shell		

We again do the classification using naive bayes algorithm, but this time we select above 14 attributes shown in Table 1. Now weka tool give the classification accuracy 98.57 % as shown in Fig. 3.

```
=== Summary ===

Correctly Classified Instances      213142               98.5742 %
Incorrectly Classified Instances      3083                1.4258 %
Kappa statistic                          0.9711
Mean absolute error                      0.0013
Root mean squared error                  0.0313
Relative absolute error                  3.0373 %
Root relative squared error             21.4623 %
Coverage of cases (0.95 level)          99.3377 %
Mean rel. region size (0.95 level)       4.4529 %
Total Number of Instances           216225
```

Fig. 3: Classification with pre-processing

Following Table 2, Show the Comparison of our work with literature survey [8]. Fig. 4 Show the graphical version of the above table. We can observe that Naive Bayes algorithm with best search method give highest sensitivity, specificity, Accuracy and false alarm rate. Now, based on the above result we use these 14 attributes in our last module threshold base algorithm. In threshold base module we are developed two modules. First Module detects the attack that is DoS/DDoS detection module. Second

Table 2: Result comparison with [8]

Algorithms	Sensitivity (%)	Specificity (%)	Accuracy (%)	FAR (%)
C4.5	86.57	82	93.23	1.56
SVM	83.82	64.29	87.18	3.2
C.4.5+ACO	89.26	85.42	95.06	0.87
SVM+ACO	87.42	67.95	90.82	2.42
C4.5+PSO	92.51	88.39	95.37	0.72
SVM+PSO	90.06	70.8	91.57	1.94
EDADT	96.86	92.36	98.12	0.18
BEST+NA¨IVE BAYES	98.6	98.8	98.57	0.04

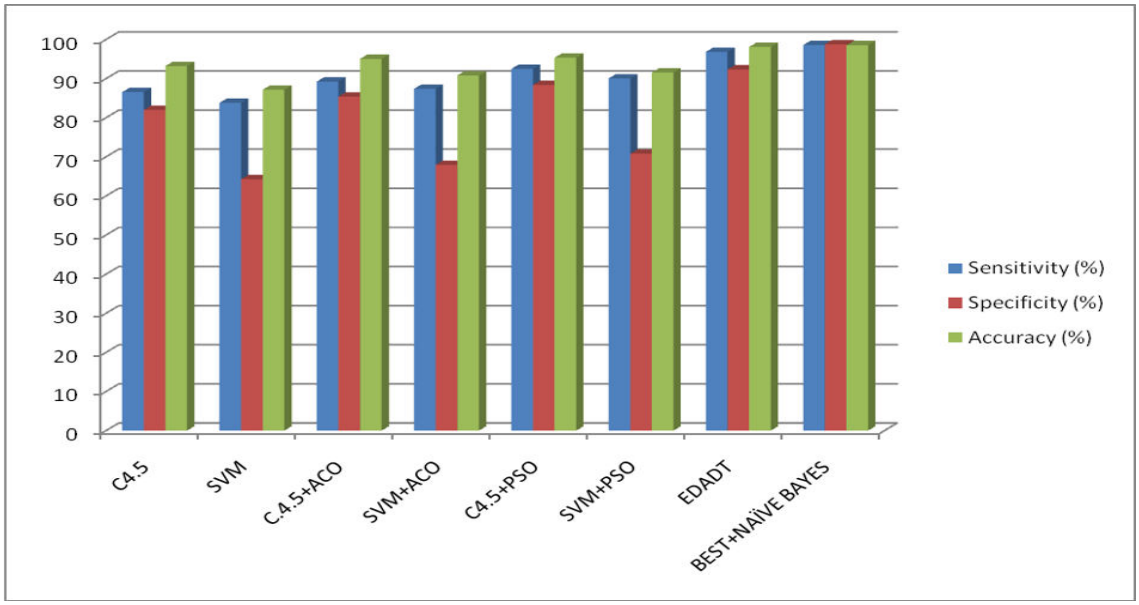

Fig. 4: Result comparison with [8]

Module is DoS/DDoS prevention module attack. In detection module we detect the traffic for fix interval of time, during this time whatever traffic packet are collect we analyze the above 14 field for that packet. If the 14-field value match with the rule design for DoS/DDoS attack and number of packets from that IP is cross the predefined threshold value then consider that IP address is suspicious. Next DoS/DDoS prevention module sends one CAPTCHA page to suspicious IP node to check whether that node is automated or manual. If suspicious IP address node does not pass the CAPTCHA test node, then suspicious IP address of that node is blocked for future traffic on web server. If collected traffic packet does not cross the threshold limit and suspicious IP node pass the CAPTCHA then packet goes to web server. Next web server gives reply back to client according Its request. Fig. 5 show the developed module for threshold base module. We developed this part in visual studio with SQLite database.

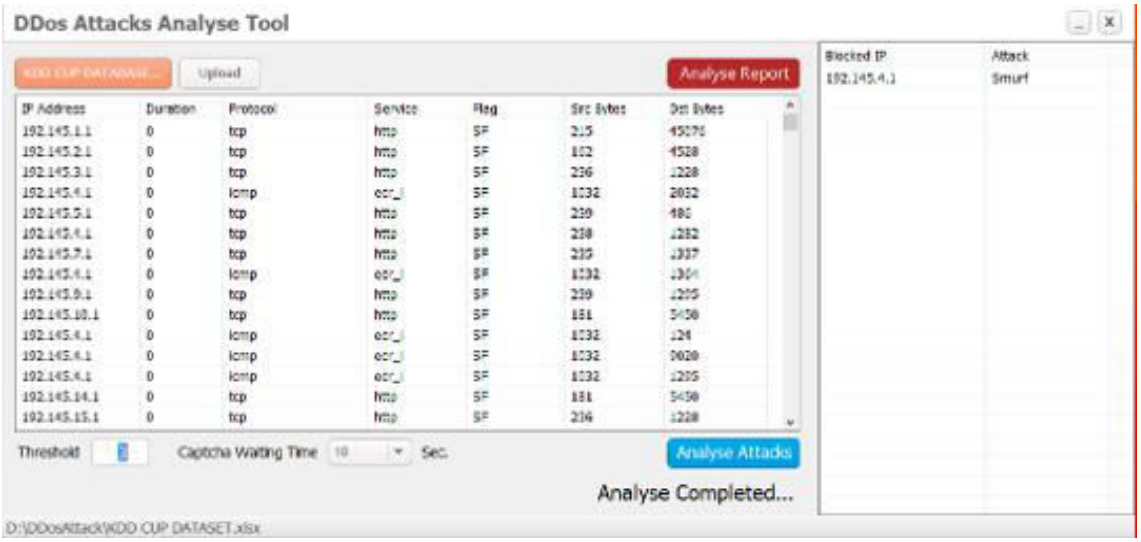

Fig. 5: Snapshot of threshold base module in Visual Studio

Following Table 3 give the exact attributes name out of 14 attributes that we identified using naive bayes classification and best method algorithm. Listed attributes name is used to design the DoS/DDoS attack detection rule for our threshold base mechanism.

Table 3: List of selected attributes for DoS detection rule

DoS Attack Name	Name of Attributes used for designing rule
Smurf	Protocol Type, Service, SRC Bytes, Flag, DST Host Count
Neptune	Protocol Type, Service, Flag, Serror Rate, Srv Serror Rate
Back	Protocol Type, Service, Flag, SRC Bytes, DST Bytes, Same SRV Rate, SRV Count
Land	Protocol Type, Service, Flag, Land, Srv Count, Dst Host Srv Serror Rate
Pod	Protocol Type, Service, Flag, SRC Bytes, Wrong Fragment, DST Host Count, DST Host Diff Srv Rate
Teardrop	Protocol Type, Service, SRC Bytes, Wrong Fragment, DST Host Count

5. CONCLUSION

This paper introduces offline architecture based on intrusion detection system for detection and prevention of DoS/DDoS attack on http-based web server. Presented architecture is based on Naive Bayes classification algorithm and threshold base mechanism with the use of CAPTCHA which improve the reliability of sys- tems. Discussed architecture, use only 14 attributes out of available 41 attributes of KDD dataset to detect and prevent the DoS/DDoS attack. So, less overhead and better efficiency is achieved compare to other systems. In future we try to implement this architecture with real time traffic and design more rule to detect various kind of other attacks as well.

REFERENCES

1. Vijay Katkar, Amol Zinjade, Suyed Dalvi, Tejal Bafna, Rashmi Mahajan, “Detection of DoS/DDoS attack against HTTP Servers using Naive Bayesian” Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on 26-27 Feb. 2015, pp. 280 – 285.

2. Samad Kolahi, Kiattikul Treseangrat, Bahman Sarrafpour, “Analysis of UDP DDoS Flood Cyber Attack and Defense Mechanisms on Web Server with Linux Ubuntu 13”, Communi- cations, Signal Processing, and their Applications (ICCSPA), 2015 International Conference on 17-19 Feb. 2015, pp. 1 – 5.

3. Abdulaziz Aborujilah, Shahrulniza Musa, "Detecting TCP SYN based Flooding Attacks by Analyzing CPU and Network Resources Performance", Advanced Computer Science Appli- cations and Technologies (ACSAT), 2014 3rd International Conference on 29-30 Dec. 2014, pp. 157 - 161.
4. Indu Mandwi, Preeti Karmore, "Implementation of Packet-Hiding Algorithm for Prevent- ing Selective Jamming Attacks", Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on 9-10 Jan. 2015, pp. 1 – 6.
5. Khundrakpam Johnson Singh, Tanmay De, "DDoS Attack Detection and Mitigation Tech- niques Based on Http Count and Verification using CAPTCHA", Computational Intelligence and Networks (CINE), 2015 International Conference on 12-13 Jan. 2015, pp. 196 – 197.
6. P. Pandiaraja, J. Manikandan, "Web Proxy based Detection and Protection Mechanism against Client Based HTTP Attacks", Circuit, Power and Computing Technologies (IC- CPCT), 2015 International Conference on 19-20 March 2015, pp. 1 – 6.
7. Ms. Pooja Bhoria, Dr. Kanwal Garg, "Determing Feature set of DoS Attacks", International Journal of advanced research in Computer Science and Software Engineering, Volume 3 Issue 5, May 2013, pp. 875-878.
8. G.V. Nadiammai, M. Hemalatha, "Effective approach toward Intrusion Detection System using Data Mining Techniques", Egyptian Informatics Journal December 2014, pp. 37-50.
9. Dr. Saurabh Mukherjee, Neelam Sharma, "Intrusion Detection using Naive Bayes Classifier with Feature Reduction", Published by Elsevier Ltd. In 2011, pp. 119-128.
10. 10. Chiragkumar Dilipbhai Patel: Lecturer in the Department of Computer Engineer- ing, K. D. Polytechnic, Patan, Gujarat, India Email address: patelchiraag@gmail.com