VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belagavi-590018

A PROJECT WORK

PHASE –I

REPORT ON

“ANALYSIS & DETECTION OF DISTRIBUTED DENIAL OF SERVICE USING MACHINE LEARNING TECHNIQUES”

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

By

DHIVAKAR A K	**1EP17CS019**
PRUTHVI S	**1EP17CS063**
SANTHOSH E	**1EP17CS076**
SHREYA M	**1EP17CS082**

Under the guidance of

Dr.EMILIN SHYNI C
Professor,
Department of CSE,
EPCET

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
East Point College of Engineering and Technology
Jnana Prabha, Bidarahalli, Virgo Nagar Post, Bengaluru, Karnataka 560049
2020-2021

EAST POINT COLLEGE OF ENGINEERING & TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project work phase-I entitled "**ANALYSIS & DETECTION OF DISTRIBUTED DENIAL OF SERVICE USING MACHINE LEARNING TECHNIQUES**" is a bonafide work being carried out by **DHIVAKAR A K [1EP17CS019], PRUTHVI S [1EP17CS063], SANTHOSH E [1EP17CS076] & SHREYA M [1EP17CS082]** in the partial fulfillment of the requirements for VII semester for the award of degree of **Bachelor of Engineering** in **Computer Science and Engineering** of **Visvesvaraya Technological University, Belagavi,** during the academic year **2020-2021**. It is certified that all the corrections/suggestions indicated for the project has been incorporated in this report. The report has been approved as it satisfies the academic requirements prescribed by the University.

Signature of Guide **Dr. Emilin Shyni C**	**Signature of Co-ordinater** **Mrs. Divya U H**	**Signature of HOD** **Dr.Sateesh T K**
Professor, **Dept. of CSE,**	**Asst. Professor,** **Dept. of CSE,**	**Professor &** **Head of the Dept. CSE,**
EPCET, Bengaluru	**EPCET, Bengaluru**	**EPCET, Bengaluru**

Name of the Reviewers **Signature with date**

1. …………………………………………… ....

2. …………………………………………… ....

ACKNOWLEDGEMENT

Firstly, we thank the **Management and Principal of East Point College of Engineering and Technology**, Bangalore for providing us an opportunity to work on this project. It gives us immense pleasure to express our deep sense of gratitude whose words of advice have always been a constant source of inspiration for us.

We express our gratitude to **Dr. Prakash S**, Principal, EPCET who has always been a great source of inspiration

We would like to express our heartfelt thanks to **Dr. T. K. Sateesh**, Professor and Head of Department of Computer Science and Engineering, EPCET for his valuable advice and encouragement to us in completing this project work.

We are obliged to **Dr. Emilin Shyni C,** Professor, Dept. of CSE, who have rendered her valuable assistance as the project guide.

We would like to thank our **Parents** and **Friends** for their support and encouragement during the course of our project. Finally, we offer our regards to all the faculty members of CSE department and all those who supported us in any respect during the major project.

DHIVAKAR A K [1EP17CS010]
PRUTHVI S [1EP17CS063]
SANTHOSH E [1EP17CS076]
SHREYA M [1EP 17CS078]

ABSTRACT

Distributed Denial of Service attacks fall under the category of critical attacks that compromise the supply of the network. DDos attacks continue to grow rapidly so to detect and mitigate these attacks became a challenging task. The work will be carried out on an active portal built on local server and a brand-new dataset will be generated with Intrusion Detection System. The work incorporates various machine learning algorithms: Support Vector Machine, Decision Tree, KNN and Logistic Regression for classification.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Distributed Denial of Service (DDoS) attacks are a subclass of denial of service (DoS) attacks. A DDoS attack involves multiple connected online devices, collectively known as a bot-net, which are used to overwhelm a target website with fake traffic.

Distributed Denial of Service attacks fall on the classification of basic assaults that bargain the accessibility of the system. The high rate assault floods the casualty with an enormous number of solicitations while in the low rate situation the casualty is undermined by moderate and traded off solicitations which brings about the fatigue of assets.

Driving Force behind the idea

Security is the most important feature of any software. We see a lot of theft of data these days. The main idea behind our project is to reduce the traffic induced by the hackers and provide maximum data security to our customers.

The software that we will be developing will help in detection of Ddos attacks and provide a solution when the attack is happening.

We tend to use machine learning techniques in a cloud platform so as to make our software work more efficiently.

1.1. PROBLEM STATEMENT

DDoS attacks can be done by two different ways: directly and indirectly. Direct attacks damage the victim machines directly by knowing a weakness in the system of the victim machines. On the other hand, indirect attacks the attacker prey on other elements with which the victim machines are associated and hinder their work.

“In order to prevent this DDoS attack implement classification algorithms like Random Forest, SVM & Naïve Bayes on newly generated dataset to increase the detection rate in Distributed denial of service (DDoS) and classify the traffic data.”

As in classification we use the training dataset to get better boundary conditions which could be used to determine each target class. Once the boundary conditions are determined, the next task is to predict the target class.

1.2. EXISTING SYSTEM

Here the task was carried out using three classification algorithms -SVM, Random Forest, Naive Bayes

Accuracy was used in order to evaluate the performance

The main disadvantage of this system is that using accuracy as evaluation metrics system failed to provide accurate results when the False Negatives and False Positives are crucial.

1.3. PROPOSED SYSTEM

Here in our proposed system we intend to develop an active portal built by us which will be hosten on a local server and detect the attack and classify the generated dataset and use four machine algorithms to find the accuracy of the attack - Decision Tree, Random Forest, SVM, Logistic Regression.

The main advantage of our system is that we are calculating Jaccard Index and F1 score which are performance calculators that are used to calculate accuracy and precision of every algorithm that is being used respectively.

1.4. AIM OF THE PROJECT

The main aim of our project is to process the newly generated dataset for DDOS detection and to increase the accuracy of the detection rate in DDOS attack.

1.5. OBJECTIVES

- To identify whether the traffic data is suspicious or normal.
- To reduce false negative rate in distributed denial of service (DDOS)
- To get good detection rate.

CHAPTER 2

LITERATURE SURVEY

The study was carried on own cloud Platform. To generate a dataset required for experimentation tor hammer tool was used. Several algorithms are used in order to evaluate performance. Among all algorithms SVM showed greater performance[1].

In cloud environment they have proposed an adaptive detection scheme to tackle a security issue that arose in a cloud environment. This method is the combination of the adaptive DR-LOF and SAX which helps to reduces false alarm rates. The proposed IDS has showed promising results in Detecting 7 types of DoS attack in addition to that it identifies the pinpoint to cause that anomaly. The 7 types of DDoS attacks are categorized into 3 types they are volume, protocol, and application based attacks. The Experimental results of proposed method clearly shows that DDoS attack more accurately found by using WEKA package. By using this scheme, the detection rates were increased and false alarm rate also reduced[6].

Now a day’s software defined networks (SDNs) and cloud computing have been widely adopted by researchers and industry. Technology development have helped attackers in increasing the attacks too, for instance, the of Denial of Service (DoS) attacks are advanced to distributed DoS (DDoS) attacks which are identified by conventional firewalls. They have presented the state of art of the DDOS attacks in SDN and cloud computing scenarios. In this they have worked on outline of DDoS attack situations and their recognition instruments in SDN and cloud computing conditions but SDN might be a victim of DDoS attack. The main focus is on the analysis of SDN and cloud computing architecture[7].

Cloud computing is a revolution in IT technology. It provides scalable, virtualized on-demand resources to the users with greater flexibility, less maintenance and reduced infrastructure cost. Technologies and legacy protocols may contain bugs and Vulnerabilities that can open doors for the attackers. Attacks such as DDoS (Distributed Denial of Service) may result in serious damage and also affect the cloud performance. In a DDoS attack, the attacker to send a large number of packets from these already-captured zombies to a server. This occupies a major portion of network bandwidth and consume much of the servers’ time. Thus, in this work, they have designed a DDoS detection system based on the C.4.5 algorithm to mitigate the DDoS threat. The proposed algorithm, coupled with signature detection techniques, generates a decision tree to perform automatic, effective detection of signatures attacks for DDoS flooding attacks[8].

Kaspersky Lab decided to counter this threat directly, and announced the company's intention to create specialized anti-DDoS technologies and solutions for businesses. The company has spent years building and testing these technologies in laboratory settings, as well as in the field protecting multi-million-dollar networks, including successfully protecting several Russian banks targeted by hacktivists in 2013. The full solution, Kaspersky DDoS Protection, is now being introduced in selected global markets, and will continue to be launched in more markets as we customize the solution to meet the individual legal requirements of each country. The first element of Kaspersky DDoS Protection is their experts, who have acquired a unique level of expertise, including a detailed understanding and analyzing of how DDoS attacks work. The second element is a sensor, either installed in Kaspersky DDoS Protection cloud or at customer premises which analyses the traffic which goes to the client's resource. Underpinning Kaspersky DDoS Protection is its scrubbing centers which are located in main Internet backbone lines. The dedicated emergency team of Kaspersky Lab experts is available 24/7 to monitor anomalies in the client's traffic. Another key way of controlling DDoS-traffic is to filter it on the provider side. One more layer of protection is the DDoS Intelligence system which is designed to intercept and analyze commands sent to bots from command and control (C&C) servers[4].

They introduced an intrusion detection system (IDS) plays a critical role in computer protection systems. Numerous approaches such as machine learning, data mining, and statistical techniques have been examined for IDS task. Recent studies reveal that combining multiple classifiers, i.e., classifiers ensemble, may possess better performance compared to single classifier. In this paper, we conduct a comparative study of the performance of five renowned ensemble techniques, i.e., bagging, stacking, boosting, rotation forest, and voting, based on three base classifiers, i.e., decision tree (C4.5), convolutional neural network (CNN), and support vector machine (SVM). Based on the experimental results, boosting and stacking perform better than bagging, rotation forest, and voting scheme. In particular, boosting-C4.5 and stacking possess the best performance in terms of performance metrics such as accuracy, precision, recall, and AUC value[10].

Cloud computing is a revolution in IT technology that provides scalable, virtualized on-demand resources to the end users with greater flexibility, less maintenance and reduced infrastructure cost. These resources are supervised by different management organizations and provided over Internet using known networking protocols, standards and formats. The underlying technologies and legacy protocols contain bugs and vulnerabilities that can open doors for intrusion by the attackers. Attacks as DDoS (Distributed Denial of Service) are ones of the most frequent that inflict serious damage and affect the cloud performance. In a DDoS attack, the attacker usually uses innocent compromised computers (called zombies) by taking advantages of known or unknown bugs and vulnerabilities to send a large number of packets from these already- captured zombies to a server. This may occupy a major portion of network bandwidth of the victim cloud infrastructures or consume much of the servers' time. Thus, in this work, we designed a DDoS detection system based on the C.4.5 algorithm to mitigate the DDoS threat. This algorithm, coupled with signature detection techniques, generates a decision tree to perform automatic, effective detection of signatures attacks for DDoS flooding attacks. To validate our system, we selected other machine learning techniques and compared the obtained results[10].

The study was carried on own cloud Platform. To generate a dataset required for experimentation tor hammer tool was used.Several algorithms are used in order to evaluate performance. Among all algorithms SVM showed greater performance[5].

CHAPTER 3

Requirement specification

✧ HARDWARE REQUIREMENTS:

❖ System	: Pentium IV 2.4 GHz.
❖ Hard Disk	: 40 GB.
❖ Floppy Drive	: 1.44 Mb.
❖ Monitor	: 14' Colour Monitor.
❖ Mouse	: Optical Mouse.
❖ Ram	: 512 Mb.

✧ SOFTWARE REQUIREMENTS:

❖ Operating system	: Windows 7 Ultimate.
❖ Platform	: Anaconda
❖ Languages used	: Python
❖ Tool used in conda/IDE	: Jupyter
❖ Packages installed	: Numpy Pandas Sklearn

CHAPTER 4

SYSTEM ANALYSIS

4.1. ARCHITECTURE

The below figure shows the proposed architecture for the project.

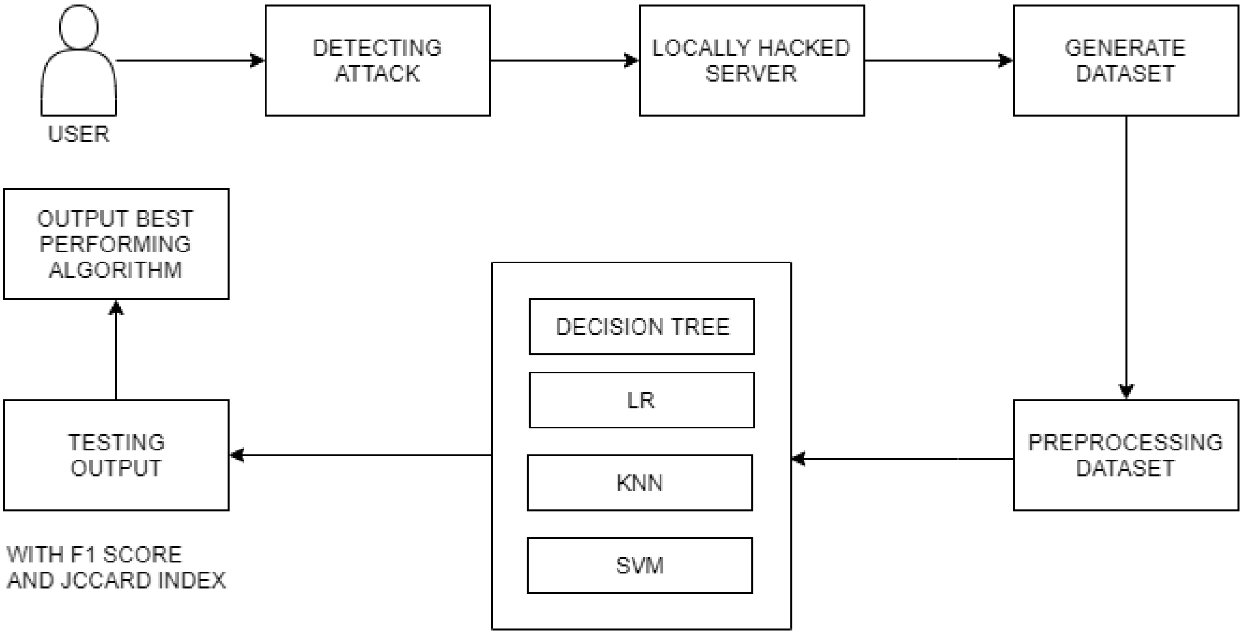

Figure 4.1: System Architecture

4.2. MODULES TO BE IMPLEMENTED

4.2.1. ATTACK DETECTION

A DDoS attack will be performed on an active portal hosted on the local server using DDoS attack tool (Tor's Hammer) and a new data set will be generated.

4.2.2. PROCESSING DATASET

The newly generated dataset will be in the form of ARFF which will need processing, henceforth in this module we will process the data and covert the ARFF file to CSV file.

4.2.3. TRAINING AND TESTING DATA

The newly generated dataset will be divided into training and testing datasets. We will train the model using training dataset and we will test the model using the below algorithms.

(i) **Random Forest:** Estimates the test error without incurring the cost of repeated model training associated with cross-validation.

(ii) **Support Vector Machine:** It deals with the wide variety of classification problems includes high dimensional and not linearly separable problems.

(iii) **Decision Tree:** It follows the concept of divide and conquer tree building.

(iv) **Logistic Regression:** It is used to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

(v) KNN: KNN searches for the nearest neighbors in the given boundary each time we want to make a prediction**.**

4.2.4. ALGORITHM SELECTION

Based on the F1 score and Jaccard Index calculated, the algorithm with highest accuracy and precision will be selected to detect the DDoS attack.

EXPECTED OUTCOME:

Module No	**Module Name**	**Input**	**Expected Outcome**
1	Data Preprocessing	Unprocessed dataset	All the missing value in the entries must be assigned to null
2	Decision Tree	Training and testing data	100%
3	Logistic Regression	Training and testing data	100%
4	KNN	Training and testing data	100%
5	SVM	Training and testing data	100%

Table 4.2: Expected Outcome

CONCLUSION

Machine learning classification algorithms will be applied to the data set namely Support Vector +-Machine, Decision tree, KNN and Logistic Regression. After considering all the algorithms and implementing them in the model whichever algorithm shows greater results compared to other algorithms will be selected.

We have done our literature survey, theoretical work and research about our project. We intend to develop a system according to our proposed architecture which we expect to give desired results. For future work, we plan to collect more normal data as well as more variants of attack data in order to expand our analysis.

REFERENCES

[1] Wani, Abdul Raoof, Q. P. Rana, and Nitin Pandey. &quot;Cloud security engineering dependent on client confirmation and symmetric key cryptographic techniques.&quot; Reliability, Infocom Advances and Optimization, sixth International Conference on. IEEE, 2017.

[2] Wani, Abdul Raoof, Q. P. Rana, and Nitin Pandey. &quot;Analysis and Countermeasures for Security and Privacy Issues in Cloud Computing.&quot; System Performance and Management Investigation. Springer, Singapore, 2019. 47-54.

[3]E.Cambiaso, G. Papaleo, and M. Aiello, "Scientific classification of Slow DoSAttacks to Web Applications," in Recent Trends in Computer Networks and Distributed Systems Security, vol.335 of Communications in Computer and Information Science, pp. 195–204, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[4] Kaspersky Labs, Global IT security dangers review 2014 - dispersed forswearing of administration (DDoS) assaults, 2014

[5] Wani, Abdul Raoof, Q. P. Rana, and Nitin Pandey. &quot;Analysis and Detection of DDoS Attacks on Cloud Computing Environment utilizing Machine Learning Techniques.&quot;IEEE 2019

[6]Mrs.G.Madhupriya,Dr.S.MercyShalinie,Ms.S.RajaRajeswari.”DETECTING DDoS ATTACK IN CLOUD COMPUTING USING LOCAL OUTLIER FACTORS” Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018)

[7] SHI DONG , KHUSHNOOD ABBAS AND RAJ JAIN.”A Survey on Distributed Denial of Service (DDoS) Attacks in SDN and Cloud Computing Environments”IEEE Access 2019

[8] Marwane Zekri, Said El Kafhali, Noureddine Aboutabit and Youssef Saadi.”DDoS AttackDetection using Machine Learning Techniques in Cloud Computing Environments”IEEE 2017

[9] Xiao, Le, et al. &quot;A protocol-free detection against cloud oriented reflection DoS attacks.&quot; Soft Computing 21.13 (2017): 3713- 3721.

[10] Tama, Bayu Adhi, and Kyung-Hyune Rhee. &quot;Data mining techniques in DoS/DDoS attack detection: A literature review.&quot; Information (Japan) 18.8 (2015): 3739.

[11] M. Zekri, S. El Kafhali, M. Hanini, and N. Aboutabit, “Mitigating economic denial of sustainability attacks to secure cloud computing environments,” Transactions on Machine Learning and Artificial Intelligence, vol. 5, no. 4, pp. 473–481, 2017.

[12] L. M. Ibrahim, “Anomaly network intrusion detection system based on distributed time-delay neural network (dtdnn),” Journal of Engineering Science and Technology, vol. 5, no. 4, pp.457–471, 2010.

[13] http://www.alldatasheet.com

[14] Wikipedia