International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

ML Based Classification And Prediction Techniques For DDOS Attacks

G.Kadiverlu
Computer Science and Engineering
(JNTUH)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
drsasikumar@sphoorthyengg.ac.in

Dr. Subba rao Kolavennu
Computer Science and Engineering
(JNTUH)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
profrao99@gmail.com

G Sathwika Reddy
Computer Science and Engineering
(JNTUH)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
sathwikareddy731@gmail.com

Sripathi Pranathi Reddy
Computer Science and Engineering
(JNTUH)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
sripathipranathi20@gmail.com

D Manogna Sree
Computer Science and Engineering
(JNTUH)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
dmanogna.2002@gmail.com

K Vaishnavi
Computer Science and Engineering
(JNTUH)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
kunchakurivaishnavi53@gmail.com

***Abstract*—** DDOS attacks, also known as distributed denial of service attacks, have emerged as one of the most serious and fastest-growing threats on the internet. DDOS attacks are an example of cyber attacks that target a specific system or network in an attempt to render it inaccessible or unusable for a period of time. DoS attacks are motivated by the desire to significantly degrade the performance or completely consume a certain resource and cause the failure of processing or exhaust the system resources by exploring a system flaw. A number of ensemble classification techniques are presented in this paper, which combines the performance of various algorithms. They are then compared to existing ML algorithms in terms of their effectiveness in detecting different types of DDoS attacks using accuracy, F1 scores, and ROC curves. The result shows high accuracy and good performance.

Keywords— Cyber Attacks · DDOS Attacks.

INTRODUCTION

Intrusion is a severe issue in security and a prime problem of security breach, because a single instance of intrusion can steal or delete data from computer and network systems in a few seconds. Intrusion can also damage system hardware. Furthermore, intrusion can cause huge losses financially and compromise the IT critical infrastructure, thereby leading to information inferiority in cyber war. Therefore, intrusion detection is important and its prevention is necessary. Different intrusion detection techniques are available, but their accuracy remains an issue; accuracy depends on detection and false alarm rate.

Thus, support vector machine (SVM), random forest (RF), and extreme learning machine (ELM) are applied in this work; these methods have been proven effective in their capability to address the classification problem. Intrusion detection mechanisms are validated on a standard dataset,

KDD. This work used the NSL–knowledge discovery and data mining (KDD) dataset, which is an improved form of the KDD and is considered a benchmark in the evaluation of intrusion detection

RELATED WORK

In the literature review section we briefly explained all the related model and the closest rival to our proposed study. We studied the latest research papers of the past two years for this research work and also Gozde Karatas *et al.* [2] proposed a machine learning approach for attacks classification. They used different machine learning algorithms and found that the KNN model is best for classification as compared to other research work. Nuno Martins *et al.* [1] proposed intrusion detection using machine learning approaches. They used the KDD dataset which is available on the UCI repository. They performed different supervised models to balance un classification algorithm for better performance. In this work, a comparative study was proposed by the use of different classification algorithms and found good results in their work. Laurens D'hooge *et al.* [6] proposed a systematic review for malware detection using machine learning models. They compared different malware datasets from online resources as well as approaches for the dataset. They found that machine learning supervised models are very effective for malware detection to make a better decision in less time.

A cybercrime in which the attacker floods a server with internet traffic to prevent users from accessing connected online services and sites.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

HARDWARE DESCRIPTION.

A. Intel i3 processor

Intel Core is a line of streamlined midrange consumer, workstation and enthusiast computer central processing units(CPUs) marketed by Intel Corporation. These processors displaced the existing mid- to high-end Pentium processors at the time of their introduction, moving the Pentium to the entry level. Identical or more capable versions of Core processors are also sold as Xeon processors for the server and workstation markets.

The lineup of Core processors includes the Intel Core i3, Intel Core i5, Intel Core i7, and Intel Core i7, along with the X-series of Intel Core CPUs

Although Intel Core is a brand that promises no internal consistency or continuity, the processors within this family have been, for the most part, broadly similar.

The first products receiving this designation were the Core Solo and Core Duo Yonah processors for mobile from the Pentium M design tree, fabricated at 65 nm and brought to market in January 2006. These are substantially different in design than the rest of the Intel Core product group, having derived from the Pentium Pro lineage that predated Pentium 4.

The first Intel Core desktop processor—and typical family member—came from the Conroe iteration, a 65 nm dual-core design brought to market in July 2006, based on the Intel Core microarchitecture with substantial enhancements in micro-architectural efficiency and performance, outperforming Pentium 4 across the board (or near to it), while operating at drastically lower clock rates. Maintaining high instructions per cycle(IPC) on a deeply pipelined and resourced out of order execution engine has remained a constant fixture of the Intel Core product group ever since.

Fig:1 intel i3

B. 4GB RAM

A randon-access memory device allows data items to be read or written in almost the same amount of time irrespective of the physical location of data inside the memory, in contrast with other direct-access data storage media (such as hard disks, CD-RWs, DVD-RWs and the older magnetic tapes and drum memory), where the time required to read and write data items varies significantly depending on their physical locations on the recording medium, due to mechanical limitations such as media rotation speeds and arm movement.

RAM contains multiplexing and demultiplexing circuitry, to connect the data lines to the addressed storage for reading or writing the entry. Usually more than one bit of storage is accessed by the same address, and RAM devices often have multiple data lines and are said to be "8-bit" or "16-bit", etc. devices.

Fig:2 4GB RAM

C. Hard Disk

A **hard disk drive** (**HDD**), **hard disk**, **hard drive**, or **fixed disk**, is an electro-mechanical data storage device that stores and retrieves digital data using magnetic storage with one or more rigid rapidly rotating platters coated with magnetic material. The platters are paired with magnetic heads, usually arranged on a moving actuator arm, which read and write data to the platter surfaces. Data is accessed in a random-access manner, meaning that individual blocks of data can be stored and retrieved in any order. HDDs are a type of non-volatile storage, retaining stored data when powered off. Modern HDDs are typically in the form of a small rectangular box.

Introduced by IBM in 1956,HDDs were the dominant secondary storage device for general-purpose computers beginning in the early 1960s. HDDs maintained this position into the modern era of servers and personal computers, though personal computing devices produced in large volume, like mobile phones and tablets, rely on flash memory storage devices. More than 224 companies have produced HDDs historically, though after extensive industry consolidation most units are manufactured by Seagate, Toshiba, and Western Digital. HDDs dominate the volume of storage produced (exabytes per year) for servers. Though production is growing slowly (by exabytes shipped), sales revenues and unit shipments are declining because solid-state drives(SSDs) have higher data-transfer rates, higher areal storage density, somewhat better reliability, and much lower latency and access times.

Fig: 3 Hard Disk

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

SOFTWARE DESCRIPTION

A. *Python*

Python is a high-level, general-purpose programming language. It is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured, object oriented and functional programming.

It was created by Guido van Rossum, and released in 1991.It works on different platforms. It has a simple syntax to the English language. It has syntax that allows developers to write programs with fewer lines than some other programming languages. It is used for web developed, software development, system scripting.

B. *Jupyter*

Jupyter is a project to develop open-source software, open standards, and services for interactive computing across multiple programming languages. It supports over 40 programming languages, including Python and R, and other data languages like Julia and Scala. It is also very easy to install with a simple pip command.

It is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning.

There are three major components in the platform:

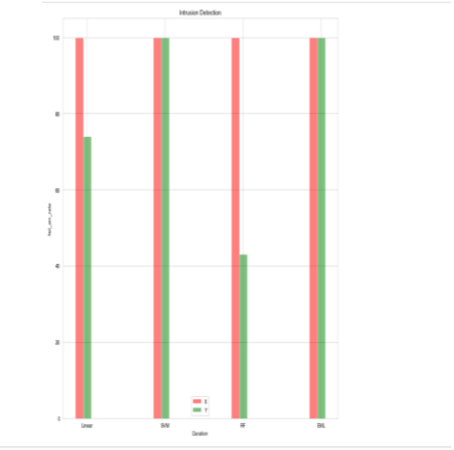
• **Cells** – are the individual units of the notebook, and they can contain either text or code.

• **Runtime Environment** – is responsible for executing the code in notebook. And it can be figured to support different languages.

• **File System** – allows you to upload, store, and download data files, code files, and outputsfrom your analysis.

WORKING

This system is developed using supervised learning method. In that classification method is used. There are 4 modules i.e., we have to load the data set. Data pre-processing and feature engineering are done. In data pre-processing , it will handle null values, duplicate values and categorical values. In feature engineering, we are performing three activities i.e., outlier detection, feature selection, data scaling. Now the data set is classified into training set and test set. Model creation will be done i.e., every algorithm is trained data set. The input in training set is known as features or x-train and output is known as y-train or label. After all algorithms are trained with training dataset. Now algorithms are trained with test dataset. The input in test dataset is known as x-test and output is known as y-test. Now again algorithms are trained with new input. That input is input and output from training set and input from test set. The different algorithms are trained and generates output called y-predict.
The both outputs y-test and y-predict are compared to known which algorithm generates accurate results.



International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

CONCLUSION

By using ML based classification and prediction techniques for DDOS Attacks, most of the DDoS attacks are detected. Firstly, we selected the dataset from Kaggle. Then, Python and jupyter notebook were used to work on data wrangling. We divided the dataset into two sets i.e., train set and test set. After we normalize the dataset for algorithm. After data normalization, we applied the proposed, supervised machine learning approach. We observed that both Support Vector Machine(SVM) and Extreme Machine Learning(EML) are 100% accurate.

FURTHER SCOPE

Looking to the future, for functional applications, it is important to provide a more user-friendly, faster alternative to deep learning calculations, and produce better results with a shorter burning time. It is important to work on unsupervised learning toward supervised learning for unlabeled and labeled datasets. Moreover, we will investigate how non-supervised learning algorithms will affect the DDoS attacks detection, in particular, we non-labeled datasets are taken into account.

REFERENCE

[1] G. Karatas, O. Demir, and O. K. Sahingoz, "Increasing the performance of machine learning-based idss on an imbalanced and up-to-date dataset," IEEE Access, vol. 8, pp. 32 150–32 162, 2020.
[2] A. Agarwal, M. Khari, and R. Singh, "Detection of ddos attack using deep learning model in cloud storage application," Wireless Personal Communications, pp. 1–21, 2021.
[3] R. Saini and M. Khari, "An algorithm to detect attacks in mobile ad hoc network," in International Conference on Software Engineering and Computer Systems. Springer, 2011, pp. 336–341.
[4] K. S. Sahoo, B. K. Tripathy, K. Naik, S. Ramasubbareddy, B. Balusamy, M. Khari, and D. Burgos, "An evolutionary svm model for ddos attack detection in software defined networks," IEEE Access, vol. 8, pp.2020.