

**OPEN ACCESS**

Volume: 11

Special Issue: 1

Month: July

Year: 2023

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Impact Factor: 3.025

Received: 08.05.2023

Accepted: 13.06.2023

Published: 01.07.2023

Citation:

Gudada, Priyanka V,  
and Rahul R. Bhat.

“A Machine Learning  
Based Classification and  
Prediction Technique  
for DDoS Attacks.”

*Shanlax International  
Journal of Arts, Science  
and Humanities*,  
vol. 11, no. S1, 2023,  
pp. 169–76.

DOI:

[https://doi.org/10.34293/  
sijash.v11iS1-July.6333](https://doi.org/10.34293/sijash.v11iS1-July.6333)

# A Machine Learning Based Classification and Prediction Technique for DDoS Attacks

**Priyanka V Gudada**

*Department of Masters of Computer Applications,  
RajaRajeswari College of Engineering*

**Rahul R Bhat**

*Department of Masters of Computer Applications,  
RajaRajeswari College of Engineering*

## Abstract

*The result is of exploiting holes in safeguarding Internet the number of Internet of Things (IoT) devices of cyber-attacks and data breaches has skyrocketed across various corporations, companies, and sectors. Because of its capacity to extract and learn deep characteristics of known assaults and detect novel attacks without using machine learning has reduced the necessity for manual feature engineering. in cyber-attack detection. Despite via means of improved Machine Learning (ML) techniques for intrusion detection, the assault remains a huge danger to the Internet. The primary goal the This study's objective is to locate and on the network. The expansion of social networks is now increasing on a daily basis.*

*However, detecting the assaults is a difficult task. By examining Those details in the KDDCUP Dataset, this project will dynamically detect the attack. The feature scaling approach utilised to normalise a variety of independent variables or data components. The feature reduction PCA technique utilised to locate the directions of highest variance in high-dimensional data and project it onto a new subspace with the same or less dimensions than the original one. Finally, the ML classification technique utilised to categorise the data utilised to assaults and the typical event, the final report is created.*

**Keywords:** DDOS, Machine Learning, Cyber Attacks, PCA Algorithm.

## Introduction

The IoT (Internet of Things) is regarded as a fast evolving paradigm in computer history. IoT has expanded tremendously in several technology domains during the last few years. It has resulted from the convergence of hundreds of billions of devices from various systems (such as smart automobiles, smart health care, smart grid, smart home, and so on) with the internet. However, because IoT merges the digital and physical worlds, this convergence has resulted in numerous cyber-attacks against IoT equipment. Because of the heterogeneity, enormous scale, limited hardware resources, and worldwide accessibility of IoT systems, IoT security has become a challenge. With the advancement of information technology, IoT technology has swiftly grown and is now extensively employed in various industries such as industry, agriculture, military, and so on.

Although IoT is highly utilised and technologically diversified, numerous devices are continually being integrated with the IoT, either as IoT terminals or as IoT branches. As an open environment based on the Internet, the IoT (Internet of Things) has complex and diversified security concerns in its gadgets, which are continually attacked and destroyed from the outside. Consequently, it is imperative to enhance the identification of security concerns in IoT. Among the present security technologies are security gateways, firewalls, code signatures, encryption technology, furthermore, yet they are all the same passive security defence measures that cannot perform active detection and reaction. The goal of IoT intrusion security detection is to assess whether the IoT is in a harmful environment by collecting data and analysing attack behaviour. To accurately categorise and predict data. To address the sparsity issue. To enhance the overall forecast findings' performance.

## Literature Survey

Tsai, Chih-Fong To investigate and comprehend the present state of employing machine learning approaches to tackle intrusion detection issues. This chapter examines 55 related publications from 2000 to 2007 that centred on the growth of single, hybrid, and ensemble classifiers. The classifier design, datasets used, and other experimental settings of related works are compared. The current successes and limits focused on the use of machine learning in creating systems for detecting intrusions. and analysed. a number of prospective study avenues are also suggested. We examined current research on utilising machine learning to detect intrusions approaches. This research, in particular, examines recent works published between 2000 and 2007. In addition, throughout the review, we cover a wide range of Machine Learning approaches employed in the intrusion detection area, such as single, hybrid, and ensemble classifiers.[1]

Y. Yang the IoT (Internet of Things) is pervasive in our daily lives. They are used in our homes, hospitals, and outside to manage and monitor variations in the environment, prevent fires, and perform a range of other useful functions. However, all of these advantages may come at the expense of significant hazards to privacy and security. Numerous research projects have been undertaken to countermeasure those concerns and develop a better solution to eliminate or at least minimise their effects on the user's privacy and security needs to be able to secure IoT devices. The survey is divided into four sections. The first section will look at the most crucial constraints of IoT devices and solutions. The second will describe how to classify IoT threats.[2]

R.S. Miani IoT, or the "Internet of Things," is a new paradigm that combines the Internet with physical devices from many areas such as environmental monitoring, industrial operations, human health, and home automation. It increases the prevalence of Internet-connected gadgets in our everyday lives, bringing with it, additional to numerous benefits, security difficulties. Systems for detecting intrusions (IDS) have been an essential tool for network and information system security for multiples of two decades. However, because of features such as constrained-resource devices, unique protocol stacks, and standards, implementing classic IDS approaches to IoT is problematic. We provide a summary of IDS research activities for IoT in this article. Our goal is to uncover emerging trends, unresolved challenges, and future research opportunities. The IDSs proposed in the literature were categorised based on the following characteristics: detection technique, IDS deployment strategy, security threat and validation approach. We also reviewed the many options for each characteristic, delving into parts of works that either offer unique IDS schemes for IoT or provide attack detection methodologies for IoT threats that may be included in IDSs.[3]

P. Torres Examines that A botnet is a network of hacked computers that may be remotely managed to carry out coordinated assaults or perpetrate fraud. Because botnets are constantly changing, traditional detection methods are always one step behind. Recently, network traffic behaviour analysis has emerged as a solution to the Botnet detection challenge. The behavioural

analysis technique seeks to generalise the typical patterns that Botnets follow throughout their life cycle to be able to detect undetected Botnet activity. This paper examines the ability to Recurrent Neural Networks (RNN) to identify network traffic behaviour by modelling it as a sequence of states that change over time. The recent success of RNN in solving sequential data issues makes them a feasible choice for sequence behaviour analysis. A RNN's performance is measured by considering two crucial aspects: network traffic imbalance and optimal sequence length.[4]

Flavio Bonomi described that Fog computing extends the Cloud Computing concept to the network's edge, allowing for a new generation of applications and services. The Figure has the following distinguishing characteristics: a) low latency and position awareness; b) widespread geographical distribution; c) mobility; d) a very high number of nodes; e) a prominent role for wireless access; f) a strong presence of streaming and real-time applications; and g) heterogeneity. In this study, we suggest that the aforementioned properties make the Fog an ideal platform for a variety of essential (IoT) services and applications, including Connected Vehicle, Smart Grid, SmartCities, and, more broadly, Wireless Sensors and Actuators Networks (WSANs).[5]

### **Existing System**

Present-day systems are used to identify intrusion attacks on IoT devices. It use the deep learning auto encoder approach to detect intrusion attacks in IoT devices, whether they are normal or malicious.

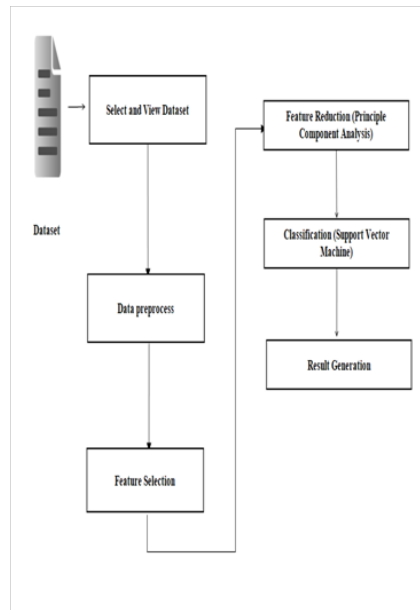
When utilising a deep learning encoder, essential variables will be misunderstood, and information will be trained using the incorrect use case.

Many academics are utilising deep learning and machine learning approaches to find the most common DDoS attacks that have the most impact on social networking. Some the majority current work in this field is covered below.

Encryption and accessibility control are similar in that they both refer to privacy and prevention. The main distinction is that encryption generally concerns with data secrecy. Data can be accessed by either a trustworthy or an untrusted party. Encryption guarantees that only authorised and trustworthy parties have access to the data. Access control, on the other hand, attempts to limit data access. Data constraints frequently occur among trusted parties. As a result, encryption solutions must be more powerful than access control mechanisms. Encryption places severe constraints on data secrecy.

### **Proposed System**

The model that is suggested is introduced to address all of the shortcomings of the current system. In the proposed system, feature scaling is a mechanism for normalising the range of independent variables data features. It will assist in selecting the best feature by utilising the PCA method. As a result, the system's performance will improve. It speeds up training and allows for more easy weight decay and Bayes optimisation. The use of machine learning for DDOS detection Leo Breiman invented Random forest is one of the most well-known machine learning methods for categorisation. Different decision trees are generated by the random forest. The tree classification method builds each tree from the initial data using an alternate bootstrap test. This investigation utilised the NSL-KDD dataset. The test was conducted utilising a laptop running Windows 10 64-bit, with an Intel (R) Core (TM) i5-2450 M CPU running at 2.50 GHz and 8.00 GB of RAM. The total number of examples utilised for training was 22,544, and the dataset had 42 characteristics. Using random data, the model was trained forest. The model's construction time was 8.71 seconds, while its testing time was 1.28 seconds. This experiment was conducted out with the help of the Weka 3.8 tool.



**Figure 1 Proposed Architecture**

Fig 1: This shows the proposed architecture and flow diagram.

## Various Types of Ddos Attack

### Icmp Flood

The hacker utilises ICMP echo request packets to deliver a service request to a genuine user. The attack can eat incoming and outgoing bandwidth and cause servers to reply to packets, resulting in overall system delay.

### Syn Flood

It may target any device It is associated with the system via the internet. The sender despite several SYN queries from the SYN flood, does not respond to the host's SYN-ACK and sends SYN queries with a faked IP address. The host waits for the acknowledgement, which results in denying services.

### Ping of Death

It transmits a larger-than-allowable-size packet into the system to the intended source. The maximum length of an IP packet is 65535 bytes. This exploit has the potential to overflow a packet-allocated memory buffer.

### Slowloris

It makes a request to the connection between a single computer and a server. It establishes a connection to the target server but only delivers a portion of the request, and it always sends HTTP headers in an incomplete manner. Because the target server has all failed connections open, the maximum concurrent connection results to connection denial.

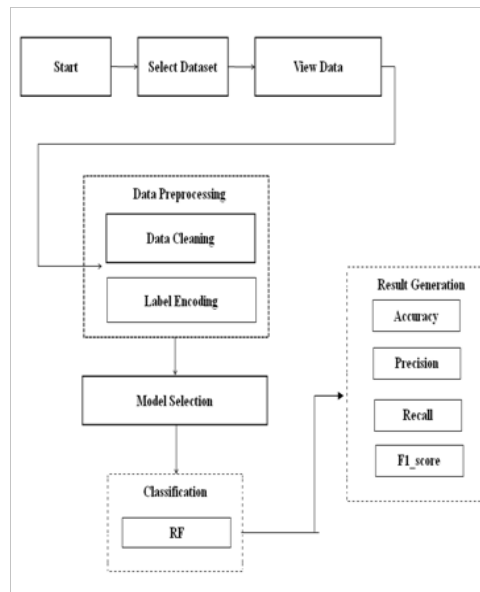
## NTP Amplification

The attacker concentrates on assaulting the publicly available source via UDP packets in this case. The attacker acquires a list of open NTP servers from which he may quickly construct a high-volume, high-bandwidth assault.

## HTTP Flood

This attack is carried through via HTTP request. To carry out this attack the assailant makes a request to the finish line user over http, after which the attack is carried out.

## Implementation



**Figure 2 Implementation**

Fig 2 Shows the implementation and the how the flows works.

## Data Selection and Loading

The selection process for the proper data kind and source, in addition acceptable devices for data collection, referred to as data selection. Data selection comes before data collection and is the process through which data relevant to the analysis is determined and obtained from data gathering. After data is received and integrated from many sources, cleaned and formatted, and then loaded into a storage system, such as a cloud data warehouse, data loading refers to the “load” component. The KDD CUB dataset is being utilised in this study to detect intrusion attacks.

## Data Preprocessing

Data pre-processing is the process of removing unwanted data from a dataset. Remove missing data: Null values, such as missing values, are evaluated in this stage. eliminated using the imputer library.

## Splitting Dataset into Training Dataset and Testing Dataset

The act of separating accessible data into segments referred to as data splitting. typically two areas for cross-validates. One piece of the data is used to build a predictive model, while the other is used to evaluate the effectiveness of the model. When examining data mining techniques, it is essential to divide the data into training and testing sets. the division of a data collection into a training set and a testing set, the majority of Data is utilised for instruction and a smaller piece of information is utilised for testing.

## Classification

Random forests, also known as random decision forests, are a technique for ensemble learning that builds several decision trees to do classification, regression, and other tasks during training and then the mode of the classes (classification) or the mean/average prediction (regression) of the individual trees.

## Prediction

It is the method of anticipating DDoS attacks based on the dataset. With the help of this initiative, data from a dataset by improving the overall prediction outcomes.

## Results

The total classification and forecast will be used to create the Final Result. When it comes to the suggested technique is assessed using metrics such as,

- Accuracy.
- Precision.
- Recall.
- F1\_score.

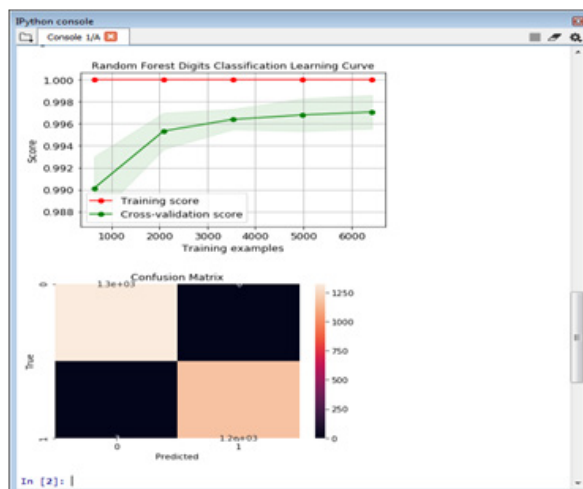
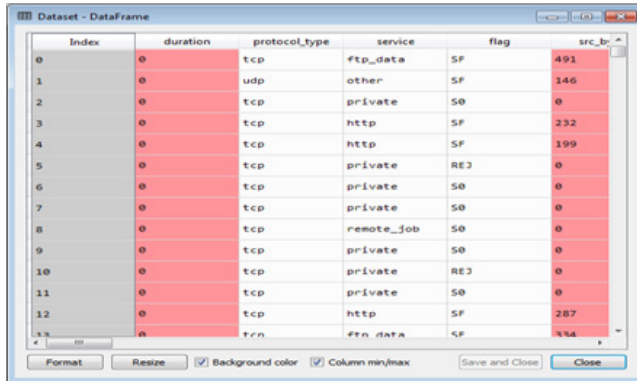


Figure 3 IPython Console

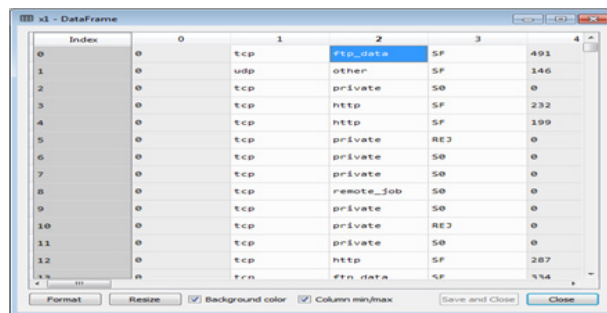
Fig 3: This shows graphical representation of the project .



Index	duration	protocol_type	service	flag	src_b
0	0	tcp	ftp_data	SF	491
1	0	udp	other	SF	146
2	0	tcp	private	SF	0
3	0	tcp	http	SF	232
4	0	tcp	http	SF	199
5	0	tcp	private	REJ	0
6	0	tcp	private	SF	0
7	0	tcp	private	SF	0
8	0	tcp	remote_job	SF	0
9	0	tcp	private	SF	0
10	0	tcp	private	REJ	0
11	0	tcp	private	SF	0
12	0	tcp	http	SF	287
13	0	tcp	ftp_data	SF	494

**Figure 4 Dataset-Dataframe**

Fig 4: This shows the data sets used in this project.



Index	0	1	2	3	4
0	0	tcp	ftp_data	SF	491
1	0	udp	other	SF	146
2	0	tcp	private	SF	0
3	0	tcp	http	SF	232
4	0	tcp	http	SF	199
5	0	tcp	private	REJ	0
6	0	tcp	private	SF	0
7	0	tcp	private	SF	0
8	0	tcp	remote_job	SF	0
9	0	tcp	private	SF	0
10	0	tcp	private	REJ	0
11	0	tcp	private	SF	0
12	0	tcp	http	SF	287
13	0	tcp	ftp_data	SF	494

**Figure 5 X1-Dataframe**

Fig 5: This shows the data sets used in this project for the classification.

**Conclusion**

Machine learning classifier which used in this work to assess intrusion attacks using IoT device data. The algorithm Principle Component Analysis (PCA) is applied to the information to extract the feature. A machine learning method used in categorization is the Support Vector Machine (SVM), which forecasts the outcome based on accuracy, precision, recall, and f1\_score.

This study’s dataset was derived from a single network. This research can be put into action using a bigger or smaller network region. In the near future, systems will rely on analytics and data networks to utilise machine learning to create predictions or deep learning.

**Reference**

1. S. Sarkar, S. Chatterjee, S. Misra, “Assessment of the suitability of fog computing in the context of internet of things,” IEEE Deals made on CloudComputing, volume. 6, no. 1, pp. 46-59, 2018.
2. H. Subir, G. Amrita, C. Mauro, “Limca: an ideal clustering algorithm for internet of things lifetime maximisation,” Wireless Networks, vol. 4, pp.1-19, 2018.
3. S. K. Choi, C. H. Yang, J. Kwak, “System Ksii Transactions on Internet & Information Systems, “Hardening and security monitoring for iot devices to mitigate iot security vulnerabilities and threats,” vol. 12, no. 2, pp.906-918, 2018.

4. S. U. Haq, Y. Singh, "On iot security modelstraditional and block chainJournal of Computer Science International Sciences & Engineering, vol. 6, no. 3, pp.26-31, 2018.
5. P. Bajpai, A. K. Sood, R. J. Enbody, "The art of mapping iot devices innetworks," Network Security, vol. 4, pp. 8-15, 2018.
6. S. Y. Hashemi, F. S. Aliee, "Why dynamic protection for internet use ofthings?," Journal of Computing Science & Engineering, vol. 12, no. 1, pp.12-23, 2018.
7. G. X. Cui, D. K. Li, "Overview on Deep Learning Based on AutomaticEncoder Algorithms," Computer Systems & Applications, vol. 9, pp. 7, 2018.
8. M. Peter, K. Gergana, M. Radoslav, P. Nevena. "Curve fitting problem:torque – velocity relationship with polynomials and boltzmann sigmoidfunctions," acta of bioengineering & biomechanics, vol. 20, no. 1, pp. 169-184,2018.
9. D. Rathore, A. Jain, "Create a hybrid technique for intrusion detection that makes use of SOM and ensemble cluster classification. network," International Journal ofAdvanced Computer Research, vol. 2, no. 3, pp. 181-186, 2019.
10. R. F. Molanes, K. Amarasinghe, J. Rodriguez-Andina, et al., "Deeplearning and reconfigurable platforms in the IOT: Challenges andopportunities in algorithms and hardware," IEEE industrial electronicsmagazine, volume. 12, no. 2, pp. 36-49, 2018.