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

MITIGATING DDOS ATTACKS IN IOT NETWORK ENVIRONMENT

K. VENKATESWARLU [1], GOLLA SUSHMA[2], RITESH[3], VALLEPU SRIKANTH[4], SINGARA CHAITRAN[5]

[1]Assistent Professor, Department of CSE-IOT, Malla Reddy College of Engineering Hyderabad, TS, India.

[2,3,4,5] UG students, Department of CSE-IOT, Malla Reddy College of Engineering Hyderabad, TS, India.

INTRODUCTION TO DDOS:

DDoS attacks, which include widespread disruption of service, are a serious issue that affects many websites regularly.By learning more about DDoS attacks and how they work, you can better defend your website from this threat.

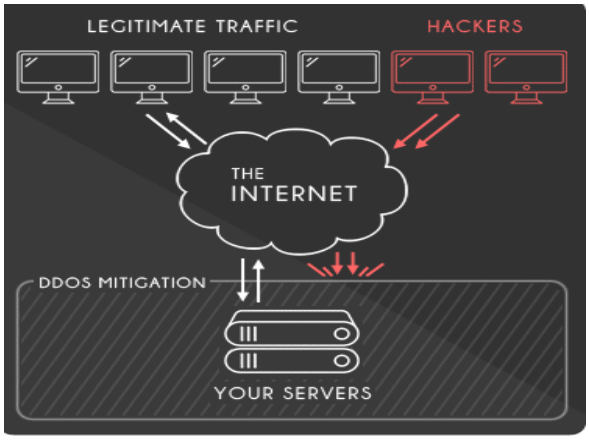
Hidden Lessons

A distributed denial of service (DDoS) attack is an attempt to bring down a website, server, or other internet-connected device by overwhelming it with an excessive volume of data.

To defend your website against DDoS attacks, it is essential to have it hosted by a company that offers DDoS mitigation and protection services.

Whether or whether your website is part of a high-risk category, you still need to take steps to protect it since it might be hit if an attacker finds a way to link to the more well-known sites that are already being targeted.

If you own a local business, why would you want it struck?



Source: Liquid Web

One of the most common misconceptions held by website owners is that their site is safe since no one would bother to hack into it.

However, the fact is that no website is safe from these kinds of assaults. You have no idea who would want to take your website down or why.

Some types of websites are obviously more common targets than others. Some categories of websites are more vulnerable to attacks such as distributed denial of service attacks and hacking.

Sites with a lot of visitors: The prestige of the group launching the attack is greatly enhanced if they are able to disrupt the services of a major website like

Amazon.com, Sony, Microsoft, or others, or obtain access to their private data.

The financial sector is always under assault, as are other sites providing economic services. DDoS assaults are often used as a cover for more malicious hacking operations, particularly those that target highly sensitive data.

Certainly, numerous sites connected to the more well-known attack targets may also be hit. You should take precautions to safeguard your site regardless of whether or not it falls into a high-risk category.

Also, if your website is hosted on a shared server, it will be affected directly if any of the other websites on that server are attacked. If a large enough attack is conducted against any site owned by the same holding company, even if you're on a virtual private server (VPS) or dedicated web server, your site may be impacted. This is because even if a hosting company can handle a lot of traffic, a distributed denial of service attack (DDoS) might still bring down the whole network.

Keeping this in mind, it's simple to understand how vulnerable any modern website is to attacks like this. Websites are especially vulnerable due to the ease with which cybercriminals and other bad actors may control them.

Ensure the Security of Your Sites Currently

If you don't have a distributed denial of service (DDoS) plan in place before an attack starts, stopping it might take a long time. This is crucial since it is usually difficult to make adjustments or upgrades to the systems after your website is under attack because they get bogged down by the assault web traffic. Many attacks on defenceless systems persist until the attacker gives up.

As a result, it's critical that you start planning for protection against DDoS attacks right now. For most sites, this level of protection is all that is needed, so look for a web host that offers it.

However, bigger websites should invest in a DDoS mitigation service that can handle the most severe assaults. No matter what kind of website you're running, you should always be ready for an attack.

Review of Selected Literary Works: Yang Xiang (2011) [1] discusses the destructive effects of low-rate distributed denial-of-service attacks. The low-cost ddos assaults are identified using two novel approaches: generalised decline and information range tactics. This study also compares and contrasts the new techniques with the old ones, namely the Shannon entropy and the kullback-liebler distance. To speed up the discovery process, we employed the alpha value of generalised entropy and the details range measure. These two new indicators would make it easy to distinguish between genuine and fake visitors to a website. In order to track out the origin of an attack, the IP trace back technique is used.By assessing the offender, this method may put an end to the assault.This research demonstrates that the suggested method is used to identify low-traffic assault targets and further reduce strike rates.

IP traceback has been researched and found to be the best method to identify the attacker by M.Vijaylakshmi, Dr.S.Mercy Shalinie, A.Arun Pragash (2012) [2]. This statistic was used to locate the router closest to the incoming web traffic and so identify the source of the DDoS attacks. Each incoming packet is marked as important and then sent to the network as part of the router's mapping procedure. When an attacker sends forged IP addresses as part of an attack, this approach may be

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

used to identify them. Attacks of this kind often target the layers between the network and the application. We also made advantage of aggressive traffic shaping and reactive filtering mechanisms. In this article, we employed an NTRO sensor-wise and secure environment test bed to measure the system's performance. Finding the foe is the paper's main reward. DDoS Attacks are studied by onowar H.Bhuyan (2012) [8], who looks at the problems and many difficulties of the detection strategies.

Python: A Quick Start Guide

Python is a widely used high-level programming language that is interpreted, interactive, object-oriented, and rich in features. Python is a great resource for programmers since it serves as both a high-level language and a debris language. Between 1985 and 1990, Guido van Rossum was successful. The Python job's resource code, like that of the Perl job, is made available under the GNU Public Licence (GPL).

Python's versatility stems from the fact that it can be put to use in a number of contexts, from the most serious to the most spur-of-the-moment of tasks, and in a number of display paradigms, including the procedural, object-oriented, and practical. Python's design philosophy encourages the use of generous indentation to make code more readable.

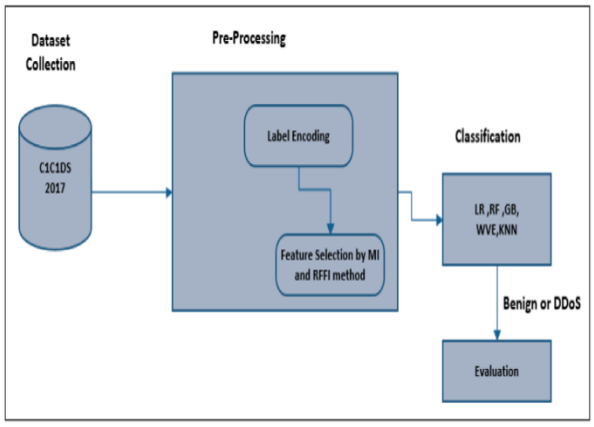
In this chapter, you will learn the fundamentals of the Python programming language. This guide will get you up and running with Python displays in no time at all making use of basic but effective techniques.

Machine Learning: An Overview of Algorithms

The term "artificial intelligence" is used to describe the way in which computers figure out how to complete tasks without being explicitly programmed to do so. In the early days of computers, it was straightforward to programme algorithms that would guide the machine through all of the steps needed to solve a problem. This allowed computers to be used for the tasks that humans had previously considered impossible. Therefore, there was no need for the computer to acquire any fresh information. It's not easy to build the algorithms required for more involved jobs like face recognition. This is in part due to the fact that humans can't provide precise instructions for how they recognise faces. However, there is a great deal of information associated with faces. The challenges of directly creating the necessary algorithms have been outweighed by the ease with which we can currently aid computers in learning how to recognise faces for themselves using publicly available data. The goal of machine learning researchers is to help computers figure out how to solve problems for which there is no clear-cut solution.

DESIGN METHODOGLOGY:

BLOCK DIAGRAM:

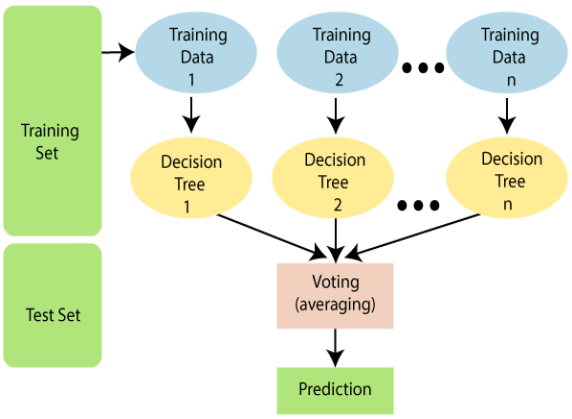


ALGORITHMS

RFC

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

The supervised learning approach that Random Forest is a part of is one of the most widely used in machine learning. It is applicable to ML issues involving both classification and regression. It relies on ensemble learning, the method of integrating many classifiers to address a challenging issue and enhance the model's accuracy.



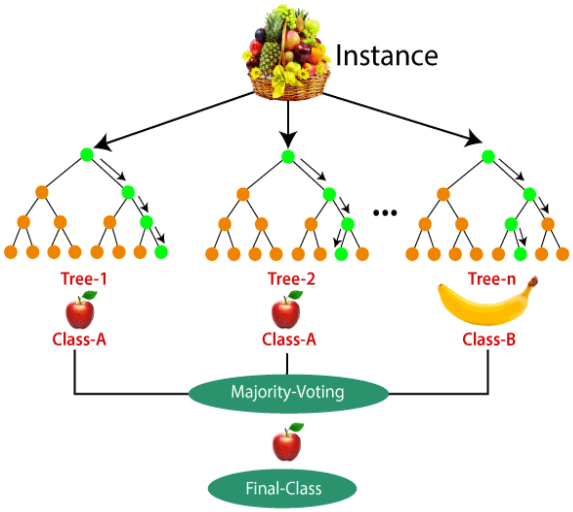
Note: To better understand the Random Forest Algorithm, you should have knowledge of the Decision Tree Algorithm.

Assumptions for Random Forest

Given that the random forest uses a combination of trees to make its classification predictions, it is likely that some decision trees will provide the right result while others would not. However, after combining all of the trees, the proper result is predicted. Two such assumptions for an improved Random forest classifier are shown below.

So that the classifier can make a reliable prediction, rather than a guess, there should be some real values in the feature variable of the dataset.

There can't be much overlap between the trees' forecasts.



Applications of Random Forest

Random forest was largely used in the following four places:

The financial sector relies heavily on this method to identify potential funding risks.

In the field of medicine, this method may be used to detect disease trends and assess potential dangers associated with such trends.

Using this method, we may pinpoint areas with a similar land use pattern.

Advertising and marketing: This formula may be used to identify common trends in both fields.

Gains from Unplanned Forestation

Category and Regression tasks are both feasible for Random Forest to complete.

It is capable of handling large, high-dimensional datasets.

It improves the layout's precision and eliminates the overfitting issue.

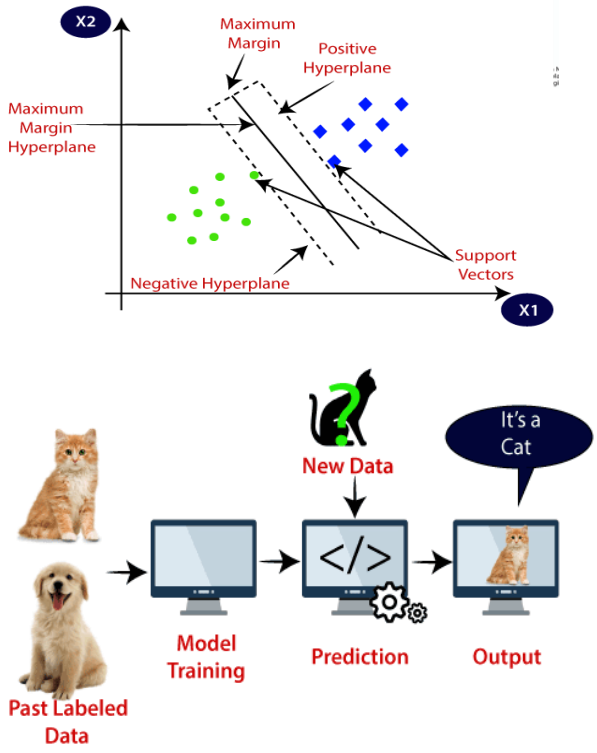
Challenges posed by Unplanned Forestry

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

Even though arbitrary forest may be used for both classification and regression tasks, it performs poorly for the latter.

SVM Support Popular solutions for Monitored Knowing include the Support Vector Machine (SVM), which may be utilised to solve both classification and regression issues. However, in Machine Learning, it is often used for Category issues.

The SVM formula's goal is to find the optimal line or decision boundary that can divide the n-dimensional space into classes, making it easier to later assign the new data element to the proper category. A hyperplane defines the narrowest feasible set of options.



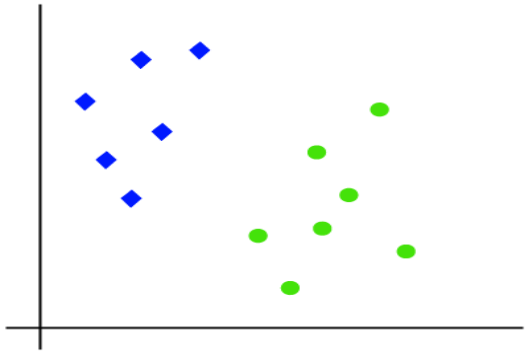
Face recognition, picture classification, and text classification are just few of the uses for the SVM algorithm.

There are two distinct types of SVM:

Linear Support Vector Machines: If a dataset can be split into two groups using a single straight line, we say that the data is linearly separable, and we use a classifier known as Direct SVM for this kind of information.

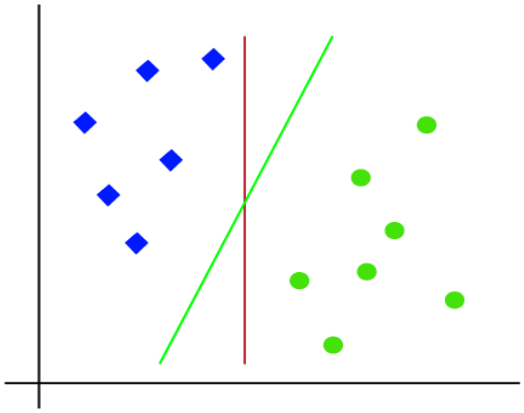
When a dataset cannot be categorised along a straight line, it is said to be non-linear, and the classifier used to categorise it is known as a non-linear support vector machine (SVM).

Direct SVM: The SVM formula's operation may be grasped by an illustration. Consider a data collection labelled "green" and "blue," with corresponding functions labelled "x1" and "x2." We need a classifier that can decide if a pair of collaborators (x1, x2) should be labelled as green or blue. Think about the diagram below:

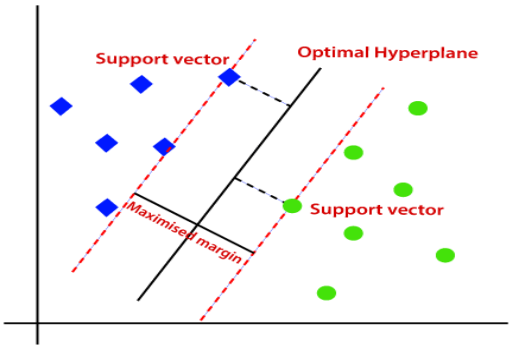


So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:

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

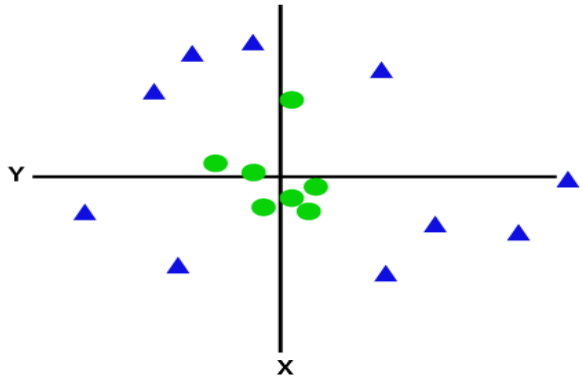


This optimum line or boundary for making a choice is known as a hyperplane, and it may be located with the help of the SVM formula. The SVM method locates the intersection of the two paths that is most closely aligned. We refer to these points as "support vectors." Margin refers to the space outside of the hyperplane that the vectors occupy. SVM's goal is to maximise this profit margin. The ideal hyperplane is the one that has the greatest possible margin.



Non-Linear SVM:

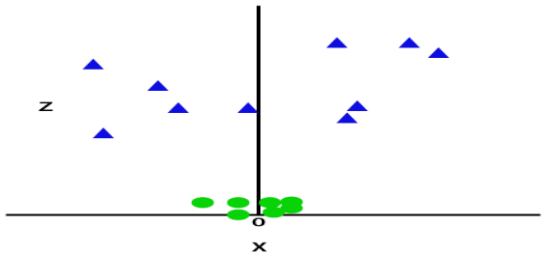
If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:



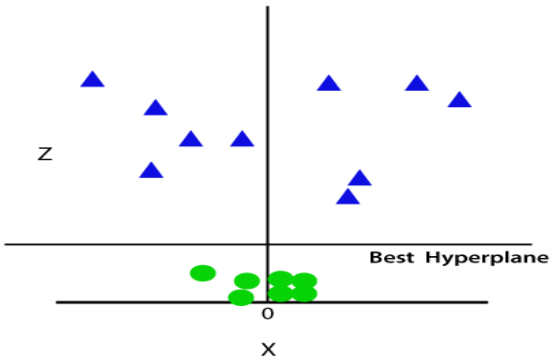
So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:

$z=x^2+y^2$

By adding the third dimension, the sample space will become as below image:

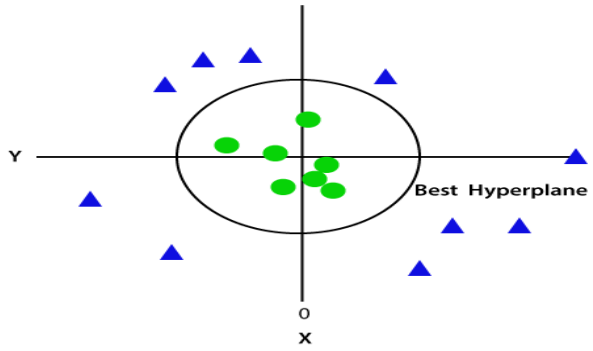


So now, SVM will divide the datasets into classes in the following way. Consider the below image:



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

Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with z=1, then it will become as:



Hence we get a circumference of radius 1 in case of non-linear data.

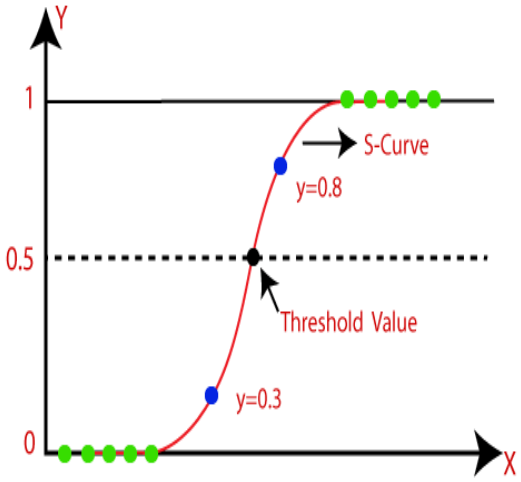
LR

One of the most well-known AI algorithms, logistic regression is a part of the Managed Discovering methodology. Using a given collection of independent factors, it may make predictions about the categorical dependent variable.

The outcome of a dependent variable of interest may be predicted using logistic regression. Therefore, the final output must be a single, unambiguous number. Rather of giving the precise value as 0 and 1, it gives the probabilistic values that fall between those two extremes, such as "Yes" and "No" or "0" and "1," etc.

Outside of their respective applications, Linear Regression and Logistic Regression are quite similar. Regression problems may be solved with direct regression, whereas category problems using logistic regression.

Instead of a straight line, the "S" shaped logistic function is fitted in logistic regression. This function expects two possible values, either 0 or 1.



Note: Logistic regression uses the concept of predictive modeling as regression; therefore, it is called logistic regression, but is used to classify samples; Therefore, it falls under the classification algorithm.

Assumptions for Logistic Regression:

The dependent variable must be categorical in nature.

The independent variable should not have multi-collinearity.

Advantages of the Decision Tree

It is simple to understand as it follows the same process which a human follow while making any decision in real-life.

It can be very useful for solving decision-related problems.

It helps to think about all the possible outcomes for a problem.

There is less requirement of data cleaning compared to other algorithms.

Disadvantages of the Decision Tree

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

The decision tree contains lots of layers, which makes it complex.

It may have an overfitting issue, which can be resolved using the **Random Forest algorithm.**

For more class labels, the computational complexity of the decision tree may increase.

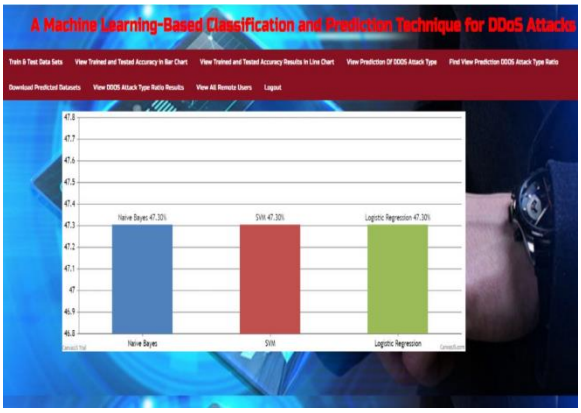


Fig.1. Output results.

Fig.2. Output graphs.

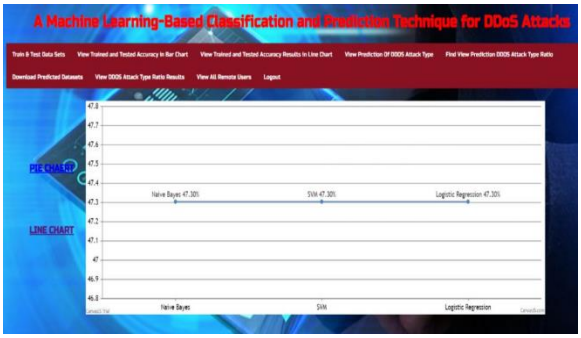


Fig.3. Accuracy levels.

Fig.4. DDos Attack detection.

CONCLUSIONS:

Finding evidence of a distributed denial of service attack is a common challenge. Lack of cloud solution is triggered by this kind of attack, hence detection is essential. This kind of attack may be identified using a machine learning model. The purpose of this research is to locate a DDoS attack that is very effective. The CICDDoS 2019 and CICIDS 2017 datasets were used in this study. Experiments used a variety of information from both databases linked to DDoS attacks. We use both the MI and RFFI methods to determine which functions are most important. Machine learning algorithms (RF, GB, WVE, KNN, LR) are then fed the selected functions. When compared to other methods, RF's overall prediction accuracy of 0.99993 when using 16 functions and 0.999977 when using 19 features is superior. Using MI and RFFI as attribute choosing strategies, we find that RF, GB, WVE, KNN, and LR all perform quite well. Future work on DDoS and other strike detection could use semantic networks and wrapper function choosing techniques like sequential function selection.

REFERENCES:

Malik, N.; Sardaraz, M.; Tahir, M.; Shah, B.; Ali, G.; Moreira, F. Energy-efficient load balancing algorithm for workflow scheduling in cloud data centers using queuing and thresholds. Appl. Sci. 2021, 11, 5849. [Google Scholar] [CrossRef]

Yan, Q.; Yu, F.R. Distributed denial of service attacks in software-defined networking with cloud computing. IEEE Commun. Mag. 2015, 53, 52–59. [Google Scholar] [CrossRef]

Lau, F.; Rubin, S.H.; Smith, M.H.; Trajkovic, L. Distributed denial of service attacks. In Proceedings of the SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics.'Cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions'(Cat. No. 0), Nashville, TN, USA, 8–11 October 2000; IEEE: Piscataway, NJ, USA, 2000; Volume 3, pp. 2275–2280. [Google Scholar]

Sambangi, S.; Gondi, L. A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression. Proceedings 2020, 63, 51. [Google Scholar]

Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. Radiographics 2017, 37, 505–515. [Google Scholar] [CrossRef]

Hasan, A.; Moin, S.; Karim, A.; Shamshirband, S. Machine learning-based sentiment analysis for twitter accounts. Math. Comput. Appl. 2018, 23, 11. [Google Scholar] [CrossRef][Green Version]

Malik, S.; Tahir, M.; Sardaraz, M.; Alourani, A. A Resource Utilization Prediction Model for Cloud Data Centers Using Evolutionary Algorithms and Machine Learning Techniques. Appl. Sci. 2022, 12, 2160. [Google Scholar] [CrossRef]

Aljamal, I.; Tekeoğlu, A.; Bekiroglu, K.; Sengupta, S. Hybrid intrusion detection system using machine learning techniques in cloud computing environments. In Proceedings of the 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA), Honolulu, HI, USA, 29–31 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 84–89. [Google Scholar]

Kushwah, G.S.; Ranga, V. Optimized extreme learning machine for detecting DDoS attacks in cloud computing. Comput. Secur. 2021, 105, 102260. [Google Scholar] [CrossRef]

Makuvaza, A.; Jat, D.S.; Gamundani, A.M. Deep Neural Network (DNN) Solution for Real-time Detection of Distributed Denial of Service (DDoS) Attacks in Software Defined Networks (SDNs). SN Comput. Sci. 2021, 2, 1–10. [Google Scholar] [CrossRef]

Manimurugan, S.; Al-Mutairi, S.; Aborokbah, M.M.; Chilamkurti, N.; Ganesan, S.; Patan, R. Effective attack detection in internet of medical things smart environment using a deep belief neural network. IEEE Access 2020, 8, 77396–77404. [Google Scholar] [CrossRef]

Intrusion Detection Evaluation Dataset (CIC-IDS2017). Available online: https://www.unb.ca/cic/datasets/ids-2017.html (accessed on 30 September 2021).

DDoS Evaluation Dataset (CIC-DDoS2019). Available online: https://www.unb.ca/cic/datasets/ddos-2019.html (accessed on 27 April 2022).

Khan, S.; Kifayat, K.; Kashif Bashir, A.; Gurtov, A.; Hassan, M. Intelligent intrusion detection system in smart grid using computational intelligence and machine learning. Trans. Emerg. Telecommun. Technol. 2021, 32, e4062. [Google Scholar] [CrossRef]

Sandhu, R.S.; Samarati, P. Access control: Principle and practice. IEEE Commun. Mag. 1994, 32, 40–48. [Google Scholar] [CrossRef]

Khan, M.S.; Khan, N.M.; Khan, A.; Aadil, F.; Tahir, M.; Sardaraz, M. A low-complexity, energy-efficient data securing model for wireless sensor network based on linearly complex voice encryption mechanism of GSM technology. Int. J. Distrib. Sens. Netw. 2021, 17, 15501477211018623. [Google Scholar] [CrossRef]

Sardaraz, M.; Tahir, M. SCA-NGS: Secure compression algorithm for next generation sequencing data using genetic operators and block sorting. Sci. Prog. 2021, 104, 00368504211023276. [Google Scholar] [CrossRef]

Zhong, Z.; Xu, M.; Rodriguez, M.A.; Xu, C.; Buyya, R. Machine Learning-based Orchestration of Containers: A Taxonomy and Future Directions. ACM Comput. Surv. (CSUR) 2021. [Google Scholar] [CrossRef]

Bindra, N.; Sood, M. Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset. Autom. Control. Comput. Sci. 2019, 53, 419–428. [Google Scholar] [CrossRef]

Kshirsagar, D.; Kumar, S. An efficient feature reduction method for the detection of DoS attack. ICT Express 2021, 7, 371–375. [Google Scholar] [CrossRef]