electronics

MDPI

Article

Security Analysis of DDoS Attacks Using Machine Learning Algorithms in Networks Traffic

Rami J. Alzahrani [1,2,*] and Ahmed Alzahrani [1]

[1] Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; asalzahrani@kau.edu.sa
[2] Department of Computer Science, Faculty of Computer Science and Information Technology, Al-Baha University, Al-Baha 65799, Saudi Arabia
* Correspondence: ralzahrani0654@stu.kau.edu.sa

Abstract: The recent advance in information technology has created a new era named the Internet of Things (IoT). This new technology allows objects (things) to be connected to the Internet, such as smart TVs, printers, cameras, smartphones, smartwatches, etc. This trend provides new services and applications for many users and enhances their lifestyle. The rapid growth of the IoT makes the incorporation and connection of several devices a predominant procedure. Although there are many advantages of IoT devices, there are different challenges that come as network anomalies. In this research, the current studies in the use of deep learning (DL) in DDoS intrusion detection have been presented. This research aims to implement different Machine Learning (ML) algorithms in WEKA tools to analyze the detection performance for DDoS attacks using the most recent CICDDoS2019 datasets. CICDDoS2019 was found to be the model with best results. This research has used six different types of ML algorithms which are K_Nearest_Neighbors (K-NN), super vector machine (SVM), naïve bayes (NB), decision tree (DT), random forest (RF) and logistic regression (LR). The best accuracy result in the presented evaluation was achieved when utilizing the Decision Tree (DT) and Random Forest (RF) algorithms, 99% and 99%, respectively. However, the DT is better than RF because it has a shorter computation time, 4.53 s and 84.2 s, respectively. Finally, open issues for further research in future work are presented.

Keywords: cyber security; IoT; machine learning; intrusion detection system; IoT security; DDoS attack

Citation: Alzahrani, R.J.; Alzahrani, A. Security Analysis of DDoS Attacks Using Machine Learning Algorithms in Networks Traffic. *Electronics* **2021**, *10*, 2919. https://doi.org/10.3390/electronics10232919

Academic Editor: Myung-Sup Kim

Received: 28 October 2021
Accepted: 23 November 2021
Published: 25 November 2021

1. Introduction

Distributed denial of service (DDoS) attacks are the most critical threats to many areas of our life such as IoT, smart cities, healthcare, information technology and commercial parts [1]. DDoS attacks continue to threaten the network security of all business sectors despite their size because of their continuous increases in complexity, volume and frequency [2]. The authors of [3] have classified DDoS attacks into two parts: (i) The first part is named reflection-based DDoS attacks. In this part, cyberspace gadgets are utilized to transmit attack traffic such as HTTP calls to the target, and the attacker's identity is hidden. These requests are sent through the source IP address targeting the IP addresses in the reflector servers (bots). Therefore, all of these concurrent demands are forwarded to the victim. Typically, these attacks are passed out to misuse the application protocols (i.e., TCP, UDP individually or integration of them). MSSQL or SSDP can be used in TCP-based attacks, while CharGen, NTP or TFTP can be used in UDP [3]. A collection of these protocols is used with the confirmed attacks, which consists of the following protocols: DNS, LDAP, NetBIOS, SNMP, or PORTMAP [3]. (ii) The second part is exploitation-based DDoS attacks, which similarly uses both TCP and UDP. The SYN flood attack is a TCP-based attack, while the UDP flood and UDP-Lag are UDP-based attacks [3]. Figure 1 provides a detailed DDoS attack taxonomy [3].

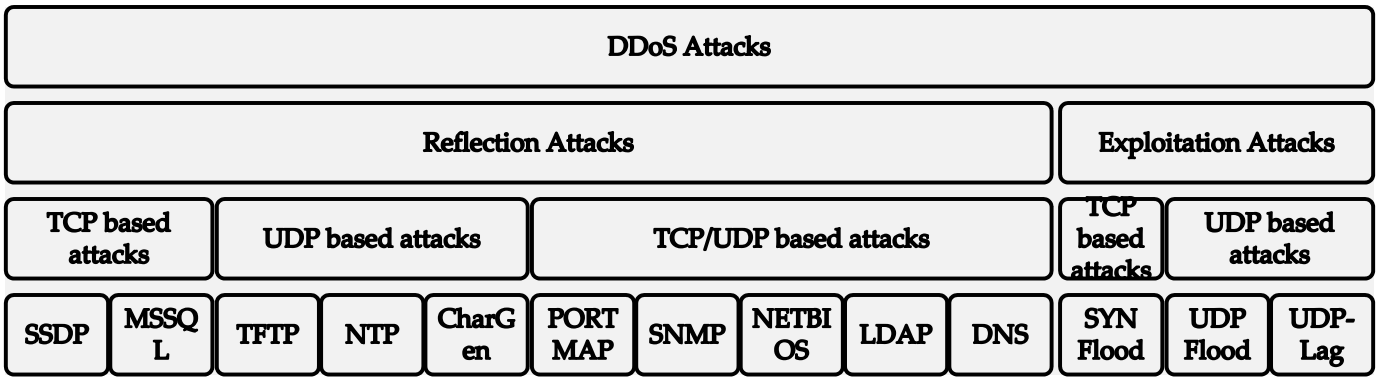


Figure 1. The Taxonomy of DDoS attack [3].

According to a CISCO report [3], there will be a huge growth in the number of DDoS attacks in the near future. According to the statistics presented in [3], by 2022, the amount of DDoS attacks will be doubled to 14.5 million, in contrast to 2017. Figure 2 shows the global increase in the number of DDoS attacks between 2017 and 2022. Because of the increasing size and traffic of DDoS attacks rapidly, there is a serious threat to service providers, and the highest reported attack was 1.7 Tb/s [2]. Recently, the cost of downtime caused by DDoS attacks was significantly high, and it has cost USD 221,836.80 [2]. Comparing between 2017 and 2018, the number of attacks against IPS devices and firewalls was nearly doubled from 16% to 31%, respectively [2]. During this time, DDoS attacks have also increased from 11% to 34% against cloud-based services and third-party data centers.

The DDoS attacks are still on the top of threats due to the accessibility of business applications, services and networks. There is a similarity between DDoS attacks and non-malicious availability issues such as system administrators performing maintenance or technical problems with the network [4,5]. These issues lead to significant challenges to accurately identify and powerfully defend these types of attacks. The network performance for gaining access to files or inaccessibility of a specific website can be slow when trying to recognize a DDoS attack [6].

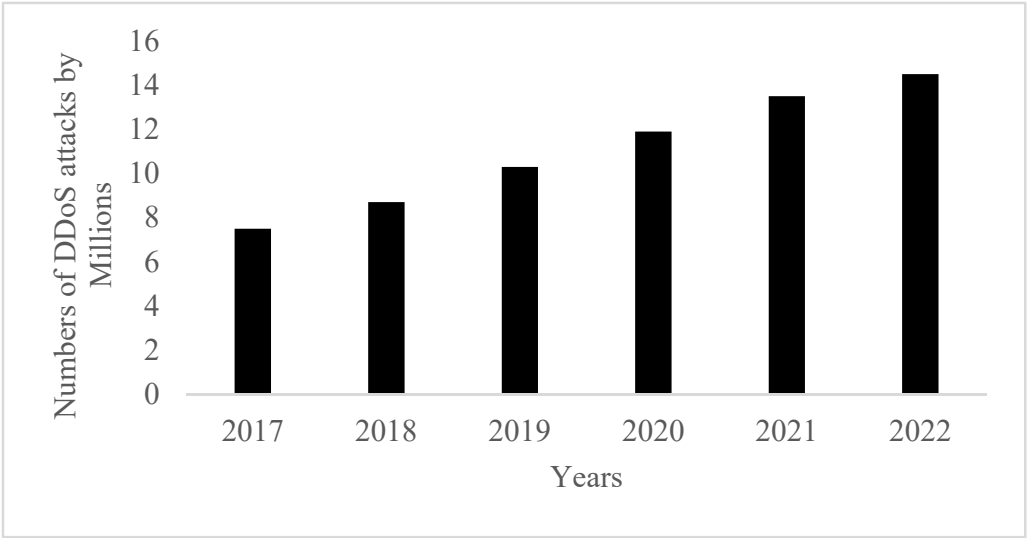


Figure 2. Global DDoS attacks forecast 2017–2022.

Criminals demonstrating attack capabilities, gaming, and extortion were the highest motivations behind these attacks in 2017 [2]. Continuously, the attackers are beefing up their computing capacity to make DDoS attacks [7]. The main contribution of this research

is to implement different machine learning (ML) algorithms in WEKA tools to analyze the detection performance for DDoS attacks using the most recent CICDDoS2019 datasets. This research has used six different types of ML algorithms: K-NN, SVM, NB, DT, RF, and LR.

There is a need to design and develop intelligent security solutions for the protection of IoT devices and against attacks generated from compromised IoT devices.

1.1. Motivation

Cybercriminals have used DDoS attacks to turn down the servers that are being targeted and penetrate venture networks that have the ability to overwhelm results. Many organizations face problems managing modern cyberattacks because of the increasing numbers of DDoS attacks' size and complexity. With the latest technologies, because of resource restrictions such as limited memory and processing capacity, smart gadgets and IoT are particularly vulnerable to a wide range of DDoS attacks, so the cybercriminals are aware of these modern technologies and their weaknesses [8]. Many organizations in 2016, such as Netflix, CNN and Twitter, were disconnected for nine hours because of an attack on their internet service providers. This technical problem caused many issues, for example, financial losses, productivity losses, brand harm, insurance rating decreases, client and provider unstable relationships, and exceeding the IT financial plan [9].

Cybercriminals might use a DDoS attack to stop clients from accessing a server or a website [1]. To secure data processing, information technology, and commercial parts, we have to build an IDS system to expose and prevent DDoS attacks. If security teams employ modern and innovative technologies such as ML, automation and AI, the cybersecurity costs will be reduced significantly [10]. This project will use different supervised machine learning (ML) algorithms to analyze the detection performance for DDoS attacks.

1.2. Main Contribution

In this research, a detailed review of network threats from IoT network and their devices with corresponding ML- and DL-based attack detection techniques is presented. This work aims to contribute to the research conducted in this field. The key contributions of this research are described as follows:

- This research covers a review of ML- and DL-based IDSs, involving their pros, cons and detections methods.
- Covering and comparing different datasets available for network- and IoT-security-related research. This is done by presenting which ML was used and the resulting accuracy found.
- Presentation of the current research challenges and their future directions for research in this field.

This paper is structured as follows: Section 2 shows the related work of different DL models and an experiment with datasets containing DDoS attacks. Section 3 presents in detail the evaluation of the performance of the research paper. Section 4 describes the measurements of evaluation. Section 5 presents some challenges and future work. Finally, Section 6 shows the conclusion of the research paper.

2. Related Work

Numerous studies on the application of DL in intrusion detection (ID) of DDoS attacks are presented here. This part summarized different deep learning models and an experiment with datasets containing DDoS attacks. The detection methods used for IDSs can be divided into three methodological types [11]: signature-based detection techniques, anomaly-based detection techniques and hybrid-based detection techniques.

2.1. Signature-Based Detection Techniques

This type of detection techniques contains a repository of attack signatures and compares the network traffic against this repository of signatures. When the match is found, a detection alert is raised. This approach can detect known attacks for which signatures

are stored in the repository, but it cannot detect zero day or new attacks, even if it is not effective against existing attack mutations [12].

This research, Ref. [13] proposed the use of an artificial immune system (AIS) to overcome the shortcomings of signature-based approaches. This technique created detectors based on attack signature utilizing the immune cell paradigm, which can determine if a packet is legitimate or malicious based on its classification as a self or non-self element. The system has the ability to adopt new patterns as a result of constant system monitoring. However, in a resource constrained IoT environment, the feasibility of such a detection technique is questionable.

The researchers in [14] solved the resource constraint problem in signature-based IDS by using a separate Linux machine with an adapted version of the Suricata-based signature IDS. On the other hand, the researcher gave no indication of how to keep attack signature up to date. The researchers in [15] expanded the research in [14] by presenting changes to signature matching techniques. Another study by [16] addressed IoT processing power limits by combining auxiliary shift values with a multiple pattern detection technique to reduce the number of matching operations necessary between attack signatures and network traffic packets. The system used signature repositories of the open source IDS (Snort) and the open source antivirus (ClamAV).

In this research [17] a signature-based IDS proposed to detect DDoS attacks in IoT networks. In a hybrid deployment, it consists of two units: (i) IDS detectors and (ii) IDS routers. The IDS router is a firewall and detection device that is hosted in the border gateway. The sensors monitor the internal traffic was employed by the IDS detectors. The results presented that the scheme identifies version number change and hello flooding attacks.

2.2. Anomaly-Based Detection Techniques

This type of detection techniques relies on the monitored environment's baseline typical behavior profile [18]. This usual baseline is then utilized to compare the actions of the system at any given time. Any deviations from the authorized threshold are recorded by using an alarm, but no classification for the sort of attack detected is provided. There have also been attempts to use behavioral detection models based on ML models that learn normal and attack events; however, establishing normal profiles is preferable to learning normal and attack events, which cannot include new attack events in real world networks [19]. Anomaly-based detection approaches are more effective in discovering novel attacks compared to signature-based detection techniques. Ml algorithms are used in anomaly-based detection strategies to create a baseline normal profile of monitored systems. Due to the significant computing resources required to train and test ML algorithms, their implementation in resource and energy constrained IoT environments remains a challenge.

This research, Ref. [20] proposed a lightweight IDS scheme for IoT. There are two levels to this scheme: training and evaluation. The technique is trained to make the system lightweight using features derived from the packet inter-arrival time of the received data during the training phase. The scheme uses the support vector machine (SVM) classifier to detect an intrusion or abnormal traffic during the evaluation stage. In terms of detection speed and classification accuracy, the lightweight IDS method performs effectively.

This research, Ref. [21] proposed a real-time scheme to detect wormhole attack in RPL-based IoT. It detects malicious users and nodes using routing information and Received Signal Strength Indicator (RSSI). In both centralized and dispersed installations, the real-time IDS systems are examined. It achieves a detection rate of 90%.

2.3. Hybrid-Based Detection Techniques

This type of detection technique utilizes a combination of both previous techniques to avoid the shortcomings and optimize the benefits of detecting existing and new attacks.

In this research [22] SVELTE is an IDS for IP-connected IoT systems that employ RPL as a routing protocol in 6LoWPAN networks, according to the inventors. They attempted to balance a compromise between the storage costs of signature-based detection and the computing costs of anomaly-based detection strategies.

In this research [23] SDN was utilized to track compliance with the manufacturer usage description (MUD) behavioral profile and build ML methods for detecting volumetric attacks such as DoS, reflective TCP/UDP/ICMP flooding and ARP spoofing to IoT devices. As a result, they found that their scheme was effective in detecting volumetric attacks.

The authors of this research [24] proposed a novel method for detecting DDoS traffic on device classes that was based on individual device traffic characteristics. The authors of this research examined the categorizations of machine type communication (MTC) traffic generated by IoT devices. The purpose of their methodology was to evaluate whether the observed IoT device created legitimate or DDoS traffic by comparing traffic variations generated by the IoT device to the legitimate traffic class to which the device initially belongs.

This research, [25] investigates the potential for using such features to classify devices, regardless of their operation or purpose. This kind of classification is necessary for a dynamic and heterogeneous environment. A total of 41 IoT devices were employed in this study. The concept of supervised ML has improved the logistic regression method. A classification model was created using Logitboost. A number of 13 network traffic features created by IoT devices were used to create multiclass classification model. Research has demonstrated that it is possible to classify devices into four classes with high performance and accuracy based on the traffic flow features of such devices. Model performance shows high results according to measures such as precision, F-measure, true-positive ratio, false-positive ratio and kappa coefficient.

In this research [26] the authors proposed a DDoS traffic detection model for various IoT device classes that uses a boosting method using logistic model trees. Because the characteristics of network traffic from each device class may differ slightly, a distinct version of the model will be developed and applied for each device class. Their study results showed high accuracy and an effective way in detecting DDoS activity.

The following studies presented deep learning detection in DDoS attacks and the techniques used. Table 1 reviewed the various DL models.

The authors in this study, [27] suggested a deep learning strategy for detecting and preventing flood attacks which are known as DoS-based Hello on the IoT healthcare network. They confirmed this type of attack by using the Deep Belief Network (DBN) model, which involved transferring many Hello packets to slow down the network. The DBN technique has utilized the bypass-linked attacker update-based rider optimization algorithm (BAU-ROA) to produce different effective results and work further optimally. To improve the execution of ROA, a metaheuristic algorithm called BAU-ROA is developed—a high-performing optimization method with a straightforward calculation approach and fewer computation parameters. Experiments have discovered that the BAU-ROA algorithm outperforms other optimization algorithms when it comes to the operational procedure of DBN.

The study [28] has solved the problem of the sampling-based technique utilized in network security, which was insufficient in the early stages of IoT network's SDN exposure by a proposed deep learning model. Stacked autoencoders (SAE) are a technique used in this study that contains a decoder that lowers their layer and the decoder that increases their layer as asymmetrical. They used the SAE technique, a deep learning technique for optimizing IDS on sampling produced by adaptive questioning and sFlow approaches, and the examinations looked at its effect on accuracy. Consequently, and after two samples, minor CPU consumption remained seen, and practical consequences for sFlow and adaptive questioning were obtained through accuracy averages of 91% and 89%, respectively.

In this research [29], the DDoS attacks were detected in the SDN controller level using a DL technique named LSTM implemented for cloud and fog computing security. The advantages of using the LSTM model are as follows: it is appropriate for exercise using network packets acquired at varied period intervals, and it spreads the information about the previous packet on the present packet. The experiment revealed that an LSTM DL model with three hidden levels and 128 units was suitable. LSTM was used to analyze the Botnet datasets named ISCX 2012 and IDS CTU-13; the experiments presented that the accuracy obtained was 98.88%, and the model was successfully implemented.

The researchers in this study, [30], implemented a DCNN technique for the exposure of DDoS attacks on the OSN. Because the shallow machine learning algorithms were unable to execute traffic analysis as required, the usage of DCNN was found appropriate when dealing with a smaller sample of the dataset. It has been noticed in the results of the experiment that the accuracy of DCNN was 99% which is the best performance compared to KNN, SVM and Naïve Bayes. The shallow machine learning algorithms show an accuracy rate of 93%, while the others show 88% and 79%, respectively.

Another study [31] proposed an IDS with a security framework that utilized the association of both the nonsymmetric deep autoencoder (NDAE) DL technique and RF. They used both models to guarantee the security of SDN. In NDAE, only the encoder part is the based, dissimilar the decoder and encoder constructing of the traditional autoencoder. The utilized DL technique remained ideal to overwhelm the problems that become apparent as a result of the shallow ML categorization including extended training periods that come up with a need for high memory and processor necessity. Consequently, the reason behind choosing the NDAE is the greater accuracy by means of minimal CPU consumption and low training period. CICIDS2017 and NSL-KDD datasets remained utilized to assess the execution of the technique employed to expose DDoS attacks. By applying the NDAE hybrit technique to the NSL-KDD and CICIDS2017 datasets, the result showed that the precision amount of 99.60%, 99.24%, respectively, was gained. As a result, it was indicated that the implemented model is appropriate for usage in an IDS.

In this research [32], the CNN and FNN techniques, both learning models, were recommended for analyzing network traffic and utilizing DDoS IDS. The NSL KDD dataset has been utilized to develop the implementation. They observed both FNN and CNN techniques to gain a high precision compared with SVM, J48, naive bayes, RF and RT, which are shallow ML algorithm strategies for detecting network anomalies and determining anomaly kinds.

This research, [33] recommended using the ANN with signature-based technique to expose DDoS attacks in the IDS, which monitors destructive movements in the network. As a consequence of the implementation, once comparing both ANN and signature-based techniques, the result showed that the joined employment of these two methodologies resulted in a 99.98% accuracy rate. From the above studies, as a result of the investigation, it is understood that the DL technique has a great stage of accomplishment in network traffic analysis and detection of DDoS attacks.

This research, Ref. [34] presented a comprehensive assessment of current and previous studies in IoT traffic characterization in terms of IoT application and design. The core attention of the papers in IoT has clearly been stated in the survey offered, with the traffic characterization towards security concerns being the primary focus. In this study, they compared the performance of four ML algorithms: DT, KNN, NB, and gradient-boosting (GRB) classifiers with regard to several factors such as accuracy, precision, recall and F1 score. They used the BoT-IoT dataset. This study's performance evaluation results suggest that DT and GRB performed better in terms of accuracy. These strong results will contribute to the IoT's networks increased security.

Table 1. Summary of the utilize of deep learning in DDoS.

Study	Model	Dataset	Application Area	Feature	Result
[27]	DBN	Generated dataset	IoT Network	BAU-ROA optimization	Produces a better outcome than other optimization methods
[28]	SAE	Generated dataset	SDN of IoT	sFlow-based and adaptive polling	sFlow 91% accuracy and Adaptive polling 89% accuracy
[29]	LSTM	ISCX 2012 and IDS CTU-13	Fog and cloud	LSTM has 128 units and three hidden layers	The accuracy rate was 98.88%
[30]	DCNN	Generated dataset	Traffic classification	To analyse few number samples in the dataset	The accuracy of the model was 99% and better than shallow ML algorithms in terms of performance
[31]	NDAE and RF	NSL-KDD and CICIDS2017	SDN	DL and shallow learning algorithms are combined in a hybrid model.	Accuracy was 98%. As a consequence, it has higher accuracy than others because it has a lower false-positive rate (FPR) less than 5%
[32]	FNN and CNN	NSL-KDD	Traffic categorization	An efficient feature modelling ability	Higher accuracy than shallow machine learning algorithms
[33]	ANN	Generated dataset	IDS	An integration of signature-based detection and ANN	The accuracy value was 99.98%
[34]	DT, K-NN, NB and GRB	BoT-IoT dataset	IoT Network	Analysis normal and attack traffic	Better accuracy in DT and GRB with 99.96% and 99.88%, respectively.

3. Evaluation of Performance

This study demonstrates the detecting execution of the six supervised ML classifiers, which are K_Nearest_Neighbors (K-NN), super vector machine (SVM), naïve bayes (NB), decision tree (DT), random forest (RF) and logistic regression (LR).

The experiments in this study use a hardware specification of Intel® Core™ i7-8650U CPU @ 1.90 GHz processor, 16 GB RAM with the operating system Windows 10, 64 bit. In this research, the ML technique in WEKA tool is being tested for forecasting DDoS attacks. This study uses the WEKA version 3.9.4 tool for data pre-processing, categorization, regression, assembling, visualization and association rules. The Java code has been used for writing WEKA, and it is an open source tool established in New Zealand at the University of Waikato. All the algorithms that have been used are supported in WEKA. WEKA has a graphical user interface and a command-based interface which make it attractive to be used in this research. It requires file formats such as CSV and ARFF. In machine learning, the dataset is required to train selected algorithms to gain knowledge.

3.1. CICDDoS2019 Dataset

This study used the CICDDoS2019 dataset collected from the University of New Brunswick Canadian Institute for Cybersecurity. To forecast DDoS attacks, this complete dataset contains 50,063,112 instances with 80 features and 11 class labels. Table 2 presents the classes label with the number of instances for each class.

3.2. The Characteristics Utilized in the Implementation

This study used the chosen 24 features that have been utilized in the study [3] to forecast DDoS attacks. The RFR was utilized to determine the significance of individual features in the dataset. Table 3 presents a list of the features used here, along with a short explanation.

Table 2. The amount number of instances in the dataset.

DDoS Attribute (Class Label)	Number of Instances
DNS	5,071,011
LDAP	2,179,930
MSSQL	4,522,492
NetBIOS	4,093,279
NTP	1,202,642
SNMP	5,159,870
SSDP	2,610,611
SYN	1,582,289
TFTP	20,082,580
UDP	3,134,645
UDP_Lag	366,461

Table 3. The feature set utilized in the IDS.

Feature	Description
Fwd Packet Length Max	Maximum packet size in the forward (outgoing) direction
Fwd Packet Length Min	Smallest packet size in the forward route
Min Packet Length	Minimum of a packet's length
Max Packet Length	Maximum of a packet's length
Average Packet Size	A packet's average size
FWD Packets/s	Number of forward packets (p/s)
Fwd Header Length	The extent of a forwarded packet's header
Fwd Header Length 1	Number of bytes in a header in the forward direction
Min_Seg_Size_Forward	Minimum segment size in the forward direction
Total Length of Fwd Packet	Packet size in the forward direction
Fwd Packet Length Std	The standard deviation of a packet in the forward direction
Flow IAT Min	The minimum amount of time passes between two packets in a flow
Subflow Fwd Bytes	The average number of bytes in a sub-flow in the forward direction
Destination Port	Address to accept the sent TCP or UDP packets
Protocol	TCP or UDP for data transference
Packet Length Std	The packet extent standard variation
Flow Duration	The flow's duration in µs
Fwd IAT Total	In the forward route, the total time among two packets
ACK Flag Count	The number of packets with ACK
Init_Win_Bytes_Forward	In the forward route, the number of bytes in the early window
Flow IAT Mean	Mean time amongst two packets in the flow
Flow IAT Max	Maximum time amongst two packets in the flow
Fwd IAT Mean	Mean time amongst two packets in the forward route
Fwd IAT Max	Maximum time amongst two packets in the forward route

3.3. Multibel Categorization Utilized in the Implementation

The 11 class labels utilized in the implementation for attack exposure are presented in this study. Figure 4 shows all the classes labels that are employed in the implementation. Based on the 24 characteristics provided in Table 3 above, these attacks are predicted. Table 4 presents the 11 class labels used and briefly produce an explanation of exploitation-based and reflection-based DDoS attacks.

Using the WEKA tool, the dataset CICDDoS2019 has been imported and analyzed with CSV format by changing the dataset attribute from Numeric to Nominal. Then, we have chosen 24 features, described in Table 3. Figure 3 shows the chosen feature in the WEKA tool interface.

Table 4. The description of the 11 chosen classes of DDoS attacks.

	Type	Class	Description
Reflection based attack	UDP Attacks	NTP	The attacker uses publicly available NTP servers to overload the aim with UDP traffic in an expansion attack known as NTP [35].
		TFTP	The buffer excess defenselessness in TFTP and server is taking advantage of TFTP attack [36].
	TCP Attacks	MSSQL	An injection of the MSSQL allows malicious SQL declarations to be executed [37].
		SSDP	An SSDP attack uses universal plug and play (UPnP) networking protocols to direct a massive volume of traffic to a victim, causing their computational properties to be overwhelmed [38].
	TCP/UDP Attacks	DNS	A DNS attack takes use of DNS flaws [39].
		LDAP	LDAP injection is an attack utilized to achieve web-based applications that structure LDAP declarations based on client information [40].
		NETBIOS	A security flaw in NetBIOS permits an attacker to read data [41].
		SNMP	An SNMP attack produces a huge amount of traffic that is pointed towards numerous networks' victims [42].
Exploitation based attack	TCP Attack	SYN Flood	SYN flood is a type of DoS attack in which an attacker forwards a series of SYN requests to a target system in order to exhaust server resources and render the system unusable to real traffic [43].
	UDP Attacks	UDP	Flooding with UDP packets is an attack that sends a high amount of UDP packets to a victim to overload their capability to proceed and reply. As a result, the firewall that protects the victim server is overburdened [44].
		UDP-Lag	UDP-Lag is a type of attack that deactivate the client-server relationship [3].

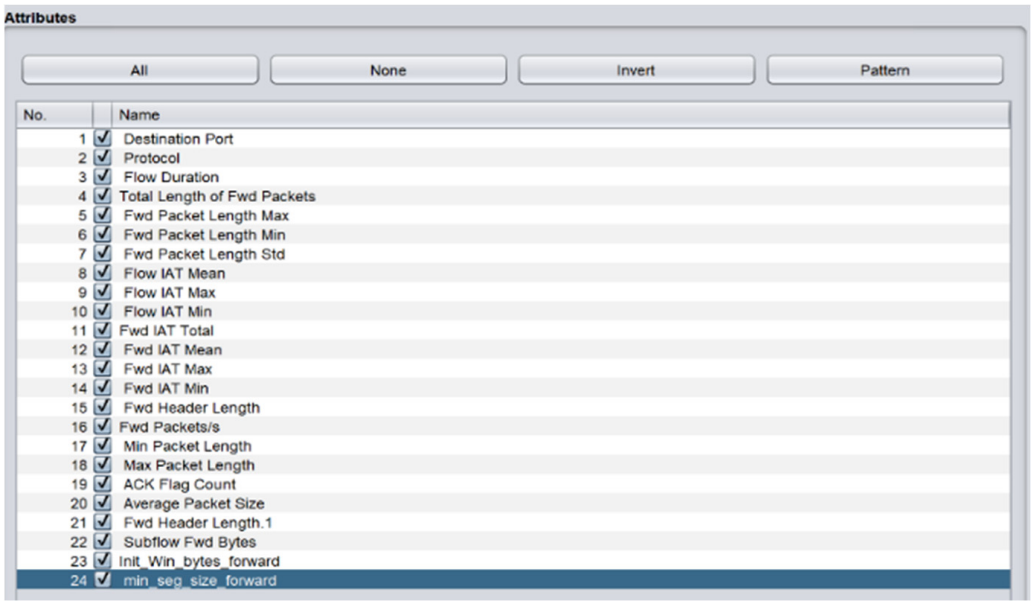


Figure 3. The 24 features chosen.

4. Measurement of Evaluation

An IDS should predict DDoS attacks with high detection accuracy. There can be significant inclusion for a community when the system does not guarantee success to expose an attack [8]. Table 5 shows a list of the measurement of evaluation.

$$TPR = \frac{\sum tp}{\sum DDoS\ attacks\ in\ dataset} \quad (1)$$

$$FPR = \frac{\sum fp}{\sum Benign\ traffic\ in\ dataset} \quad (2)$$

$$p = \frac{tp}{tp + fp} \quad (3)$$

$$r = \frac{tp}{tp + fn} \quad (4)$$

$$f - measure = \frac{2 \times p \times r}{(p + r)} \quad (5)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Table 5. List of Used Notations.

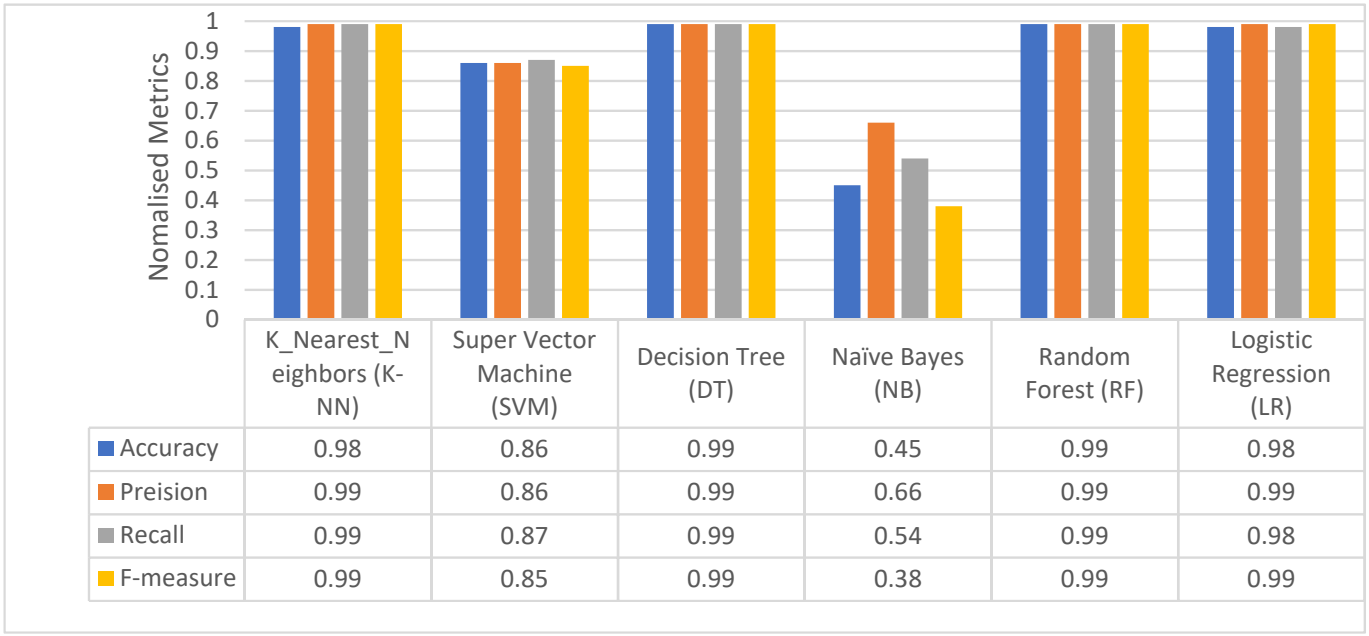
Symbol	Meaning
True Positive (tp)	It is the amount of DDoS attacks that have been recognized as attacks.
True Negative (tn)	It is the amount of legitimate network traffic instances benign recognized as legitimate.
False Positive (fp)	It is the amount of legitimate network traffic examples benign misidentified as attacks.
False Negative (fn)	It is the amount of DDoS attacks that cannot be defined as legitimate.
TPR	The amount of DDoS attacks exposed as attacks is split by the total amount of DDoS attacks in the dataset and is calculated as shown in Equation (1).
FPR	Calculated by dividing the amount number of benign instances imperfectly distributed as DDoS attacks by the whole amount number of benign instances in a dataset, and it is calculated as shown in Equation (2).
Precision (p)	Defined as the amount number of tp among all instances that are forecast to be positive, and it is calculated as shown in Equation (3).
Recall (r)	The percentage of tp from all instances that are essentially positive and is calculated as shown in Equation (4).
f-measure	The weighted harmonic means of precision and recall and is calculated as shown in Equation (5).
Accuracy	It is obtained by the equation below, and it displays the model's exact forecast rate, and it is calculated as shown in Equation (6).

Table 6 summarizes the presented experiment result for the six types of performance of the selected algorithms.

Table 6. Performance metrics for each algorithms.

Selected Algorithm	Accuracy	Precision	Recall	F-Measure	Computation Time
K_Nearest_Neighbors (K-NN)	0.98	0.99	0.99	0.99	3.5 s
Super Vector Machine (SVM)	0.86	0.86	0.87	0.85	7.29 s
Naïve Bayes (NB)	0.45	0.66	0.54	0.38	1.3 s
Decision Tree (DT)	0.99	0.99	0.99	0.99	4.53 s
Random Forest (RF)	0.99	0.99	0.99	0.99	84.2 s
Logistic Regression (LR)	0.98	0.99	0.98	0.99	5.53 s

Figure 4 shows the performance metrics of selected algorithms. The best accuracy was found in the DT and RF algorithms.



	K_Nearest_Neighbors (K-NN)	Super Vector Machine (SVM)	Decision Tree (DT)	Naïve Bayes (NB)	Random Forest (RF)	Logistic Regression (LR)
Accuracy	0.98	0.86	0.99	0.45	0.99	0.98
Preision	0.99	0.86	0.99	0.66	0.99	0.99
Recall	0.99	0.87	0.99	0.54	0.99	0.98
F-measure	0.99	0.85	0.99	0.38	0.99	0.99

Figure 4. The performance metrics of selected algorithms.

In Table 7, the studies on DDoS attack traffic detection using ML algorithms and the classification model we propose are shown comparatively. When Table 7 is examined, it is seen that different datasets were used to detect attack traffic. Some of the researchers used public datasets containing network traffic data from conventional network topologies [44] such as KDD Cup'99 [45] and UNB-ISCX [46]. The use of these datasets is positive for comparing the performance of ML algorithms used in the detection of attack traffic.

Table 7. The comparison of the related studies.

Datasets	Feature Selection	ML Algorithms	Accuracy
CIC DoS dataset [47]	No feature selection	RT, J48, REP Tree, SVM, RF, MLP	95%
KDD Cup'99 [48]	No feature selection	SVM and DNN	92.30%
UNB-ISCX [49]	No feature selection	Semi-supervised ML algorithm	96.28%
CICDDoS2019 (Our approach)	Feature selection RFR	SVM, K-NN, DT, NB, RF and LR	99%

The results show that ML models are quite successful in detecting attack traffic. The work in this paper aims to contribute to the research conducted in this field. The experimental results showed that using the random forest regressor (RFR) feature selection methods increases the accuracy of ML methods in detecting attack traffic.

For attacks such as DDoS attacks that need to be intervened without wasting time, it is important to detect the attack traffic by using system resources as efficiently as possible. Therefore, the most effective features should be selected when creating ML models.

It can be seen from Table 7 that the performance of ML models in studies using feature selection algorithms is better than in other studies. It can be said that model classification performance contributes positively to the classification of attack traffic when used in conforming to feature selection algorithms. However, the presented studies are run by applying different models on different datasets, and it is difficult to make general evaluations on comparative results.

Table 8 shows six different types of supervised machine learning algorithms that this research has been used in the experiment.

Table 8. Pros and cons of different ML-based methods [50].

ML Method	Pros	Cons
KNN	— Simple to understand and easy to implement. — It works easily with multiclass dataset.	— It is difficult to figure out what the best value for K is and how to find missing nodes.
SVM	— Due to their simplicity, SVMs are extremely sclable and capable of executing tasks such as anomaly-based intrusion detection in real time, as well as online learning. — SVMs are thought to be appropriate for data with a large number of feature attributes. — SVMs consume less memory and storage.	— The usage of an optimal kernel function in SVM, which is utilized to separate data that is not linearly separable, is still a challenge. — SVM-based models are challenging to understand and interpret.
NB	— It is simple and quick to forecast the test dataset's class. It also does well with multi-class prediction. — When the assumption of independence is met, an NB classifier outperforms conventional models such as logistic regression while using less training data.	— Conditional independence of the assumption class, which may result in accuracy loss. For some attributes, the assumption of independence may not be valid. Practically dependencies exist among variables.
DT	— Easy and simple to utilize.	— It requires bigger storage. — It is computationally complex. — It is easy to utilize only if few DTs are used.
RF	— It generates a more reliable and accurate output which is resistant to overfitting. — It requires substantially fewer inputs and does not require the process of feature selection.	— Because RF creates multiple DTs, it may be impractical to employ in real-time applications that require big datasets.
LR	— It is easier to put into practice, interpret, and train with. — It does not make any assumptions about class distributions in feature space. — It is very fast at classifing unknown records. — It performs well when the dataset is linearly separable and has good accuracy for many simple datasets.	— If the number of observations is smaller than the number of features, LR should be avoided; otherwise, overfitting may occur. — The assumption of linearity between the dependent and independent variables is a major limitation of LR.

5. Challenges and Future Work

Memory and other limited resources and computing abilities, as well as a diversity of standards and protocols, characterize the Internet of Things. These variables add significantly to the difficulties in researching IoT security issues, including anomaly mitigation utilizing IDS. In spite of the extensive study on anomaly detection in IoT networks, there are numerous key outstanding challenges that require additional investigation. The following are a few of these issues:

1. There are no publicly available IoT network traffic datasets. Because assessing and validating anomaly prevention strategies on a real network will be difficult, efforts to create an IoT dataset are essential. This will make evaluating and validating suggested anomaly mitigation techniques in the IoT much easier.
2. There are not any standard authentication apps for IoT. The validation of implemented structures is critical since it guarantees that they are developed acceptably. The

implemented structures are put to the test in a variety of ways, including simulations and tests. However, because of a lack of standard authentication applications, most of implemented IDS structures in the IoT are not evaluated in contrast to other IDS structures in the IoT. As a result, efforts must be made to produce standard authentication, which will assure duplication, reproducibility, and research continuity.

3. RNN and CNN are examples of supervised and unsupervised ML techniques, and both can be discovered using the CICDDoS2019 dataset.
4. It is possible to gather and examine real-time packets against the classified training dataset. It is possible to use a technique for splitting the data and comparing it with the performance of the classifiers utilized fold cross authentication.

6. Conclusions

In this research, DDoS attacks are serious challenges to many areas of our life. This leads us to try to find a comprehensive intrusion detection system to decrease the number of attacks facing many sectors. This study has used CICDDoS2019, which is the newest and complete dataset accessible by Canadian Institute for Cybersecurity. It has also examined six diverse ML algorithms: SVM, K-NN, DT, NB, RF and LR. The following measurements accuracy, precision, recall, true-positive ratio, false-positive ratio and F-measure have been used in the evaluation. The result of the experiment shows that the best accuracy is found when using DT and RF algorithms 99% and 99%, respectively. Both DT and RF have achieved the same result in precision 99%, recall 99% and F-measure 99%. However, the DT is better than RF because it has less computation time of 4.53 s and 84.2 s, respectively. The results show that ML models are quite successful in detecting attack traffic. Our work aims to contribute to the research conducted in this field. This paper contributes that as shown in the experiments, the random forest regressor (RFR) feature selection methods increases the accuracy of ML methods in detecting attack traffic. The implementation of this study can be employed into our real-life system in different domains in IoT. Finally, the limitations and future possibilities for network anomaly mitigation systems in the IoT are explored.

Author Contributions: Conceptualization, R.J.A.; Funding acquisition, R.J.A.; Methodology, R.J.A. and A.A.; Resources, R.J.A.; Supervision, A.A.; Visualization, R.J.A.; Writing—original draft, R.J.A.; Writing—review & editing, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: I am glad that I completed this work successfully. This work would not have been possible without the help of my supervisor, Ahmed Alzahrani. I would like to thank him for his expert advice and usual support.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Covington, M.J.; Carskadden, R. Threat implications of the Internet of Things. In Proceedings of the 2013 5th International Conference on Cyber Conflict, Tallinn, Estonia, 4–7 June 2013; pp. 1–12.
2. Conner, B. Worldwide security. *Netw. Secur.* **2003**, *2003*, 16. [CrossRef]
3. Sharafaldin, I.; Lashkari, A.H.; Hakak, S.; Ghorbani, A.A. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In Proceedings of the International Carnahan Conference on Security Technology, Chennai, India, 1–3 October 2019. [CrossRef]
4. Anstee, D.; Escobar, J.; Sockrider, C. 10th Annual Worldwide Infrastructure Security Report. 2015. Available online: https://www.netscout.com/blog/cloud-crosshairs (accessed on 14 March 2021).

5. Mouli, V.R.; Jevitha, K. Web Services Attacks and Security- A Systematic Literature Review. *Procedia Comput. Sci.* **2016**, *93*, 870–877. [CrossRef]
6. Oliveira, R.A.; Laranjeiro, N.; Vieira, M. Assessing the security of web service frameworks against Denial of Service attacks. *J. Syst. Softw.* **2015**, *109*, 18–31. Available online: https://www.sciencedirect.com/science/article/pii/S0164121215001454 (accessed on 26 October 2021). [CrossRef]
7. Abhishta; Joosten, R.; Nieuwenhuis, L.J.M. Analysing the Impact of a DDoS Attack Announcement on Victim Stock Prices. In Proceedings of the 2017 25th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP 2017), St. Petersburg, Russia, 6–8 March 2017; pp. 354–362. [CrossRef]
8. Subbulakshmi, T.; Balakrishnan, K.; Shalinie, S.M.; Anandkumar, D.; Ganapathisubramanian, V.; Kannathal, K. Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset. In Proceedings of the 3rd International Conference on Advanced Computing, ICoAC 2011, Chennai, India, 14–16 December 2011; pp. 17–22. [CrossRef]
9. Gupta, B.; Joshi, R.C.; Misra, M. Defending against Distributed Denial of Service Attacks: Issues and Challenges. *Inf. Secur. J. A Glob. Perspect.* **2009**, *18*, 224–247. [CrossRef]
10. Samtani, S.; Kantarcioglu, M.; Chen, H. Trailblazing the Artificial Intelligence for Cybersecurity Discipline. *ACM Trans. Manag. Inf. Syst.* **2020**, *11*, 1–19. [CrossRef]
11. Zarpelão, B.B.; Miani, R.S.; Kawakani, C.T.; de Alvarenga, S.C. A survey of intrusion detection in Internet of Things. *J. Netw. Comput. Appl.* **2017**, *84*, 25–37. [CrossRef]
12. Kumar, R.; Kumar, P.; Tripathi, R.; Gupta, G.P.; Kumar, N.; Hassan, M.M. A Privacy-Preserving-Based Secure Framework Using Blockchain-Enabled Deep-Learning in Cooperative Intelligent Transport System. *IEEE Trans. Intell. Transp. Syst.* **2021**. [CrossRef]
13. Liu, C.; Yang, J.; Chen, R.; Zhang, Y.; Zeng, J. Research on immunity-based intrusion detection technology for the Internet of Things. In Proceedings of the 2011 7th International Conference on Natural Computation, ICNC 2011, Shanghai, China, 26–28 July 2011; Volume 1, pp. 212–216. [CrossRef]
14. Kasinathan, P.; Pastrone, C.; Spirito, M.A.; Vinkovits, M. Denial-of-Service detection in 6LoWPAN based Internet of Things. In Proceedings of the International Conference on Wireless and Mobile Computing, Networking and Communications, Lyon, France, 7–9 October 2013; pp. 600–607. [CrossRef]
15. Kasinathan, P.; Costamagna, G.; Khaleel, H.; Pastrone, C.; Spirito, M.A. Demo: An IDS framework for internet of things empowered by 6LoWPAN. In Proceedings of the ACM Conference on Computer and Communications Security, Berlin, Germany, 4–8 November 2013; pp. 1337–1339. [CrossRef]
16. Oh, D.; Kim, D.; Ro, W.W. A Malicious Pattern Detection Engine for Embedded Security Systems in the Internet of Things. *Sensors* **2014**, *14*, 24188–24211. [CrossRef]
17. Ioulianou, P.; Vasilakis, V.; Moscholios, I.; Logothetis, M. A Signature-based Intrusion Detection System for the Internet of Things. Jun 2018. Available online: https://eprints.whiterose.ac.uk/133312/ (accessed on 28 March 2021).
18. Keshk, M.; Turnbull, B.; Moustafa, N.; Vatsalan, D.; Choo, K.-K.R. A Privacy-Preserving-Framework-Based Blockchain and Deep Learning for Protecting Smart Power Networks. *IEEE Trans. Ind. Inform.* **2019**, *16*, 5110–5118. [CrossRef]
19. Mitchell, R.; Chen, I.-R. A survey of intrusion detection techniques for cyber-physical systems. *ACM Comput. Surv.* **2014**, *46*, 1–29. [CrossRef]
20. Jan, S.U.; Ahmed, S.; Shakhov, V.; Koo, I. Toward a Lightweight Intrusion Detection System for the Internet of Things. *IEEE Access* **2019**, *7*, 42450–42471. [CrossRef]
21. Deshmukh-Bhosale, S.; Sonavane, S.S. A Real-Time Intrusion Detection System for Wormhole Attack in the RPL based Internet of Things. *Procedia Manuf.* **2019**, *32*, 840–847. [CrossRef]
22. Raza, S.; Wallgren, L.; Voigt, T. SVELTE: Real-time intrusion detection in the Internet of Things. *Ad Hoc Netw.* **2013**, *11*, 2661–2674. [CrossRef]
23. Hamza, A.; Gharakheili, H.H.; Benson, T.A.; Sivaraman, V. Detecting Volumetric Attacks on IoT Devices via SDN-Based Monitoring of MUD Activity. In Proceedings of the 2019 ACM Symposium on SDN Research, SOSR 2019, San Jose, CA, USA, 3–4 April 2019; pp. 36–48. [CrossRef]
24. Cvitić, I.; Peraković, D.; Periša, M.; Botica, M. Novel approach for detection of IoT generated DDoS traffic. *Wirel. Netw.* **2019**, *27*, 1573–1586. [CrossRef]
25. Cvitić, I.; Peraković, D.; Periša, M.; Gupta, B. Ensemble machine learning approach for classification of IoT devices in smart home. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 3179–3202. [CrossRef]
26. Cvitic, I.; Perakovic, D.; Gupta, B.; Choo, K.-K.R. Boosting-based DDoS Detection in Internet of Things Systems. *IEEE Internet Things J.* **2021**, 1. [CrossRef]
27. Srinivas, T.A.S.; Manivannan, S. Prevention of Hello Flood Attack in IoT using combination of Deep Learning with Improved Rider Optimization Algorithm. *Comput. Commun.* **2020**, *163*, 162–175. [CrossRef]
28. Ujjan, R.M.A.; Pervez, Z.; Dahal, K.; Bashir, A.K.; Mumtaz, R.; González, J. Towards sFlow and adaptive polling sampling for deep learning based DDoS detection in SDN. *Future Gener. Comput. Syst.* **2019**, *111*, 763–779. [CrossRef]
29. Priyadarshini, R.; Barik, R.K. A deep learning based intelligent framework to mitigate DDoS attack in fog environment. *J. King Saud Univ.-Comput. Inf. Sci.* **2019**. [CrossRef]
30. Hasan, Z.; Hasan, K.Z.; Sattar, A. Burst Header Packet Flood Detection in Optical Burst Switching Network Using Deep Learning Model. *Procedia Comput. Sci.* **2018**, *143*, 970–977. [CrossRef]

31. Krishnan, P.; Duttagupta, S.; Achuthan, K. VARMAN: Multi-plane security framework for software defined networks. *Comput. Commun.* **2019**, *148*, 215–239. [CrossRef]
32. Zhu, M.; Ye, K.; Xu, C.Z. Network Anomaly Detection and Identification Based on Deep Learning Methods. In *International Conference on Cloud Computing*; Springer: Cham, Switzerland, 2018; Volume 10967 LNCS, pp. 219–234. [CrossRef]
33. Alzahrani, S.; Hong, L. Detection of distributed denial of service (ddos) attacks using artificial intelligence on cloud. In Proceedings of the 2018 IEEE World Congress on Services, SERVICES 2018, San Francisco, CA, USA, 2–7 July 2018; pp. 37–38. [CrossRef]
34. Alzahrani, R.J.; Alzahrani, A. Survey of Traffic Classification Solution in IoT Networks. *Int. J. Comput. Appl.* **2021**, *183*, 37–45. [CrossRef]
35. Rudman, L.; Irwin, B. Characterization and analysis of NTP amplification based DDoS attacks. In Proceedings of the 2015 Information Security for South Africa, Johannesburg, South Africa, 12–13 August 2015. [CrossRef]
36. Liu, Q.; Zhang, Y. TFTP vulnerability finding technique based on fuzzing. *Comput. Commun.* **2008**, *31*, 3420–3426. [CrossRef]
37. Rehman, S.U.; Khaliq, M.; Imtiaz, S.I.; Rasool, A.; Shafiq, M.; Javed, A.R.; Jalil, Z.; Bashir, A.K. DIDDOS: An approach for detection and identification of Distributed Denial of Service (DDoS) cyberattacks using Gated Recurrent Units (GRU). *Future Gener. Comput. Syst.* **2021**, *118*, 453–466. [CrossRef]
38. Wang, X.; Sun, Y.; Nanda, S.; Wang, X. Looking from the Mirror: Evaluating IoT Device Security through Mobile Companion Apps. 2019. Available online: https://www.usenix.org/conference/usenixsecurity19/presentation/wang-xueqiang (accessed on 26 October 2021).
39. Hudaib, A.A.Z.; Hudaib, E.A.Z. DNS advanced attacks and analysis. *Int. J. Comput. Sci. Secur.* **2014**, *8*, 63–74. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.736.2315&rep=rep1&type=pdf (accessed on 26 October 2021).
40. Alonso, J.M.; Bordon, R.; Beltrán, M.; Guzmán, A. LDAP injection techniques. In Proceedings of the 2008 11th IEEE Singapore International Conference on Communication Systems, ICCS 2008, Guangzhou, China, 19–21 November 2008; pp. 980–986. [CrossRef]
41. Sarıkoz, B.G. An Information Security Framework for Web Services in Enterprise Networks. 2015. Available online: https://open.metu.edu.tr/handle/11511/24441 (accessed on 26 October 2021).
42. Gondim, J.J.; Albuquerque, R.D.O.; Orozco, A.L.S. Mirror saturation in amplified reflection Distributed Denial of Service: A case of study using SNMP, SSDP, NTP and DNS protocols. *Future Gener. Comput. Syst.* **2020**, *108*, 68–81. [CrossRef]
43. Yaacoub, J.-P.A.; Noura, H.N.; Salman, O.; Chehab, A. Robotics cyber security: Vulnerabilities, attacks, countermeasures, and recommendations. *Int. J. Inf. Secur.* **2021**, 1–44. [CrossRef]
44. Lau, F.; Rubin, S.H.; Smith, M.H.; Trajković, L. Distributed denial of service attacks. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Nashville, TN, USA, 8–11 October 2000; Volume 3, pp. 2275–2280. [CrossRef]
45. Chica, J.C.C.; Imbachi, J.C.; Vega, J.F.B. Security in SDN: A comprehensive survey. *J. Netw. Comput. Appl.* **2020**, *159*, 102595. [CrossRef]
46. Yusof, M.A.M.; Ali, F.H.M.; Darus, M.Y. Detection and Defense Algorithms of Different Types of DDoS Attacks. *Int. J. Eng. Technol.* **2018**, *9*, 410–444. [CrossRef]
47. Perez-Diaz, J.A.; Valdovinos, I.A.; Choo, K.-K.R.; Zhu, D. A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning. *IEEE Access* **2020**, *8*, 155859–155872. [CrossRef]
48. Karan, B.v.; Narayan, D.G.; Hiremath, P.S. Detection of DDoS Attacks in Software Defined Networks. In Proceedings of the 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS, Bengaluru, India, 20–22 December 2018; pp. 265–270. [CrossRef]
49. Ravi, N.; Shalinie, S.M. Learning-Driven Detection and Mitigation of DDoS Attack in IoT via SDN-Cloud Architecture. *IEEE Internet Things J.* **2020**, *7*, 3559–3570. [CrossRef]
50. Asharf, J.; Moustafa, N.; Khurshid, H.; Debie, E.; Haider, W.; Wahab, A. A Review of Intrusion Detection Systems Using Machine and Deep Learning in Internet of Things: Challenges, Solutions and Future Directions. *Electronics* **2020**, *9*, 1177. [CrossRef]