International Journal of Smart Home
Vol.14, No.2 (2020), pp.15-24
http://dx.doi.org/10.21742/ijsh.2020.14.2.02

Detection of Distributed Denial of Service Attacks based on Machine Learning Algorithms

Md Abdur Rahman

Associate Professor of Computer science, Department of Mathematics, Jahangirnagar university, Savar, Dhaka, Bangladesh
marahmanju@juniv.edu

Abstract

Distributed Denial of Service (DDoS) attacks make the challenges to provide the services of the data resources to the web clients. In this paper, we concern to study and apply different Machine Learning (ML) techniques to separate the DDoS attack instances from benign instances. Our experimental results show that forward and backward data bytes of our dataset are observed more similar for DDoS attacks compared to the data bytes for benign attempts. This paper uses different machine learning techniques for the detection of the attacks efficiently in order to make sure the offered services from web servers available. This results from the proposed approach suggest that 97.1% of DDoS attacks are successfully detected by the Support Vector Machine (SVM). These accuracies are better while comparing to the several existing machine learning approaches.

Keywords: *Machine learning, Machine learning algorithms, DDoS attack, Benign Attempts, Confusion matrix*

1. Introduction

Network infrastructures encounters enormous attacks. The Denial-of-Service (DoS) attacks is based on the congestion and one of major threats which break records continuously. These deny the victim to receive services in internet through inundating with malicious traffics.

These attacks were observed firstly in 1998 [1]. Several well-known web sites, such as Amazon, eBay, and Yahoo, etc. were encountered by DDoS attacks in the year of 2000. These web sites were attacked through the internet although these were highly secured web sites to provide services to web clients. These events prove that DDoS attack became a major threat which has to be detected and protected to access data sources and get services by users. The year 2015 and 2016 was the worst year, because DoS attacks was recorded by 500 Gbps and 800 Gbps respectively [2].

DDoS attacks can be operated on cloud platforms. In this case, attackers use the virtual machines to attack the web site by using VM bots [3]. Attackers rent virtual machines in order to attack because of huge computational ability than using their physical machines.

Nowadays, there is no debate about the increasing popularity of Internet of Things (IoT), and it is used in the vehicles, wearable devices, and even in home. It need networking to connect public facilities, household appliances, medical equipment, interconnected vehicles, etc. [4][5][6][7]. One great work used Support Vector Machines (SVM) for the detection of

Article history:
Received (July 23, 2020), Review Result (August 26, 2020), Accepted (September 29, 2020)

Print ISSN: 1975-4094, eISSN: 2383-725X IJSH

DDoS attacks [8]. Dao et al. has proposed DDoS attack detection algorithm for SDN devices [9]. Some researchers have proposed about how to boost and improve the IoT security with SDN technology. One paper proposed a distributed architecture was proposed with security for IoT based SDN domain (Flauzac et al. [10]). Another paper investigated the potential threats on the Open Flow control channel Li et al. [11]. Ahmed and Kim [12] provided guidance for the mitigation of DDoS attacks in IoT.

Hu et al. categorized whether the process is normal or intrusive class using K-nearest Neighbor Classifier [13]. The processes of the same class will make the cluster together. This work had used machine learning exclusively for the detection of attacks. However, this classifier is expensive for computation while the simultaneously increasing the number of processes.

DoS attack prevents the authentic clients from accessing information from the web server. Two types of DoS attacks are recorded: network level attacks and application level attacks [14]. Network level DoS attacks disable the connectivity of valid users to access network resources, and application level DoS attacks disrupt the services from server resources temporarily or indefinitely. More than 30% of network attacks are accomplished by DoS attack [15]. These attacks interrupt of accessing to a simple webpage to very large servers.

The defense mechanisms for DDoS can be categorized into two main parts: source side defences and destination side defences. It is difficult to recognize the attack from the source side. However, D-WARD [19][20] system was developed to compare incoming traffic with and outcoming traffic on the source side in order to detect DDoS attacks.

Destination side defense systems can detect and respond the DDoS attacks at the node of victim. Several systems [16][17][18] can monitor received packages while detecting the attack. Then it can discard the connection. Meanwhile, network have been vulnerable by attack bundles. As a result, it is much difficult to stop the attack by the attackers.

Several works used four types machine Learning algorithms: unsupervised, supervised, reinforcement, and semi-supervised learning to train the machine learning models for evaluating the results [21][22][23]. Moreover, Agrawal et al., 2011 used SVM for the prediction of total zombies in cyber attacks [24]. One recent work developed an anomaly-based application layer of Bio-Inspired for early and fast detection of DDoS attacks from HTTP flood [29]. Patgiri et Al. have used two machine learning algorithms: Support Vector Machine and Random Forest, and followed a thorough experiment to detect intrusion. The performance of these two algorithms is compared to detect intrusion [30].

The rest of this paper is structured as follows: In Section 2, different machine learning techniques are presented for providing basic information. In Section 3, we present our proposed framework. In Section 4, we discuss simulation results after applying different ML approach, and then compare the results. In the last section, we focus a summary of outcome and future work.

2. Concepts of machine learning

We have used different machine learning (ML) approaches such as Logistic Regression (LR), Decision Tree, and Support Vector Machines (SVMs) etc. in our system for detection DDoS attack from the benign attempts. These ML approaches are discussed below:

2.1. Support Vector Machines

Support Vector Machines (SVMs) are the machines that are using to plot training vectors in feature space, and these vectors are labelled by its class. This classification problem is

similar to quadratic optimization problem. They use a technique that is able to avoid curse of dimensionality. The predominant feature of SVMs is that it can classify the dataset through determining set (collection) of support vectors made by the set of training inputs, and then a hyper-plane is generated in high-dimensional feature space.

2.2. Decision tree

Decision tree is used in machine learning for classification. It is efficient way that follows a divide-and-conquer strategy to construct decision tree recursively. The decision tree has the root, internal nodes, branches, and leaves like a tree. Each tree represents a rule which based on the data attributes. Leaves are labeled as the decision for classification. Let the classes are denoted by $C_1, C_2, \ldots, C_n$, and each leaf of decision tree is identifying a specific class from class C_i.

2.3. Logistic regression

Logistic Regression is one of the most effective classification approaches. It is possible to determine the application layer DDoS attack from the effective features after feature extraction. In this paper, we have used logistic regression, however the performance is not suitable for our dataset. The logistic regression can be explained as follows: suppose there are k independent features $x_1, x_2, x_3, \ldots, x_k$, then the probability of DDoS attack detection is expressed as follows:

$$P = P(y = 1 | x_1, x_2, x_3, \ldots, x_k) \tag{1}$$

and logistic regression is as

$$p = \frac{e^y}{1 + e^y} \tag{2}$$

where,

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_k x_k \tag{3}$$

γ_0 is the coefficient, and $x_1, x_2, x_3, \ldots, x_k$ are the features.

3. Proposed framework

We observe datasets in which objects in a specific group are related to each other, and different from objects of other groups. As our datasets contains two types of data: DDoS attack and benign. As our goal is to observe anomalies, we observe if there are any data which is different from normal data. The advantage of using machining learning techniques is that they can separate DDoS attack objects from datasets from the benign objects.

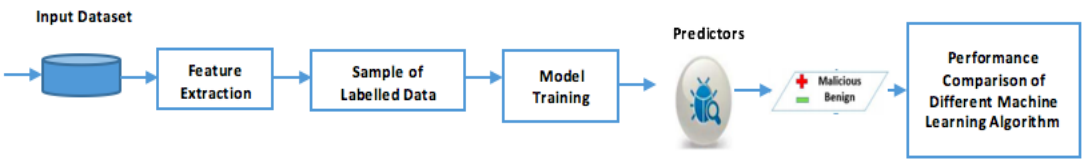

Figure 1. Block diagram of the proposed system

This framework contains of several components: feature selection, data pre-processing, data analysis for different machine learning algorithms, training the dataset to algorithms, testing the dataset and then comparing the results with different algorithms. The block diagram of the proposed system is shown in [Figure 1]. The machine learning DDoS attack detection system consists of the following phases:

Preprocessing: it is required to make collected data to an understandable format so that it must be complete, consistent and free from lacking in specific behavior.

Training: several machine learning algorithms is trained using the machine readable data. The predominant thing is that our machines experiences through these data which have DDoS attacks and normal data. The features of these training data fall into two classes: benign and attack.

Testing: after machine learns from the training dataset, then it can make predictions from new dataset based on its learning. This is for measuring the performance on testing data.

We have developed a classifier that can classify malicious packet from the benign packet. This model works as detectors which firstly detect it and then stop or minimize the strength of an attack. Indeed, this detector receives the request from the web clients, then it can identify the malicious packet if this request is falling into the DDoS class. This is detected as this request does not behave normally. In this paper, we focus on different Machine Learning Algorithms such as Support Vector Machine (SVM), decision tree, and logistic regression. This framework offers robust techniques to the DDoS attack detection.

Also, our classifier identifies irregularities in the network. The predominant thing is that this classifier has to permit authentic packets for passing through the network so that these must be reached to the destination without any interruption or delays. For providing the services to the legitimate clients, these detectors must check each request precisely.

After loading datasets using pandas, we have chosen two attributes, initial window forward bytes and backward bytes, to observe the main trends of attacks and benign attempts. The following figure presents a visualization of the relation between these attributes [Figure. 2] in

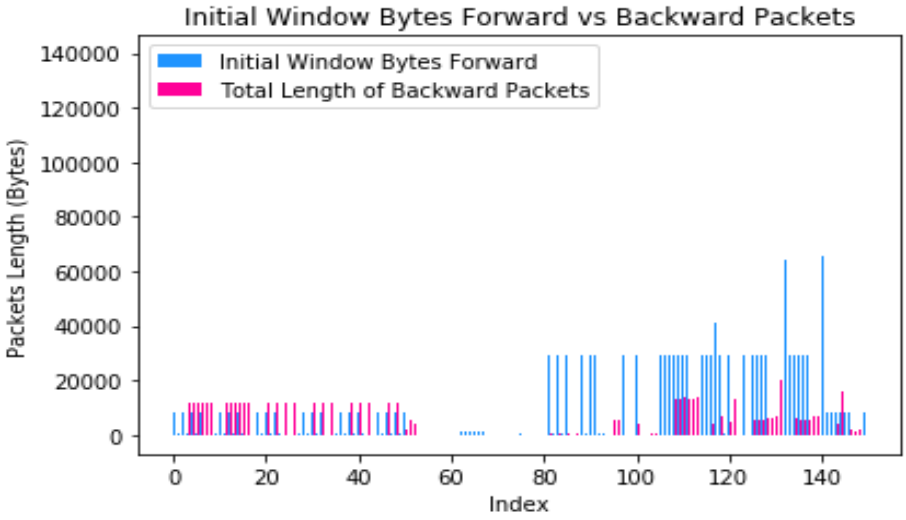

Figure 2. forward and backward packets indicates the data for DDoS attacks and benign attempts

Table 1. Dataset of Canadian institute of cybersecurity

Destination	Flow Duration	Total Fwd Pkts	Total Bwd Pkts	Total Length of	Total Length of	Initial Window bytes	Initial Window bytes	Label

				Fwd Pkts	Bwd Pkts	Fwd	Bwd	
53	83718	4	2	184	300	-1	-1	BENIGN
445	10706606	29	24	8142	4220	8192	2050	BENIGN
80	39723	3	5	26	11601	8192	229	DDoS
443	118945	19	25	1169	43577	29200	61	BENIGN
80	80803000	9	6	62	11607	256	229	DDoS

4.1. Data collection

In the simulation, we use the dataset of Canadian Institute of Cybersecurity which includes the two types of objects: BENIGN and DDoS attacks. This attack is accomplished over various network and sessions. These sessions became to attack and non-attack phases. Table I represents the sample of dataset for this proposed work.

We examine the features of the dataset and observe nine parameters. However, four attributes of them have a good indication for the detection of DDoS attack. Some records of datasets have some similarities. That means that intruder sends the data which have little variation while requests from different web clients have much variations as legitimate web clients are different users, and they request servers for different requirements. We observe that their data has huge variations than attackers shown in [Figure 2].

The plot shows the little variations of data in the first part which indicates very suspicious known as DDoS attack in which the attacker sends the almost same length of forwarding data for attacking the servers shown in [Figure 2]. On the other hand, the data values of features are very distinguishable in the second part of the data. It has great variations of data in the second part which is the indication of normal data. This is because different web clients request to the servers for varieties of resources. That is the main reason to have huge variations of data.

4.2. Prediction accuracy

This model can make the predictions based on the data for which the correct labels are assigned. We observed categorical data in the label so that each object is identified whether this comes from the web clients request or attacker.

It is important to know what different types of data are sent by the web clients and attackers. At the same time, there should be data which is responded by the web servers. We have examined closely about the differences among these two categories objects. From the dataset, the selected data of objects can be set into a NumPy array in python. The dataset is divided into training set and the test set. This training data have been feed to the proposed model which is SVM(s) classifiers. Using this data, this model learns about the situations. Then, it can predict for each test data for comparing it against its label. The accuracy for prediction correctly is measured which expresses how fine this model works. For this classifier, the accuracy of test set is almost 0.971, which indicates that the percentage of prediction is 97.1%. According to the mathematical expectations, our model meets the 97.1% correct through its forecasts.

4.3. Evaluation through confusion matrix

It is imperative to keep in remember that misclassifying a DDoS attack makes the negative impression of the model, although it can classify the benign attempt correctly. We focus to calculate the true positives and true negatives for the confusion matrix to know the success or failure of this work.

In terms of detection accuracy, training time, running time, scalability, Support Vector Machines (SVMs) achieve expected outcome and beat other techniques like decision tree, logistic regression, etc. while testing three different Machine Learning classifiers on the dataset. More importantly, we observed that SVMs have high detection accuracy of the DDoS attack among them, which is 97.1%. But, minimum accuracy is recorded for logistic regression algorithm (59.3%). [Table 2] represents the performance score board of SVM, decision tree, and logistic regression. [Figure 3]. represents a bar diagram to show performance of three machine learning algorithms.

Table 2. Performance of different machine learning algorithms

Method	Accuracy	Precision	F1-Score	Sensitivity	FP	FN
Logistic Regression	0.593	0.647	0.616	0.589	0.409	0.410
Decision Tree	0.827	0.865	0.833	0.803	0.159	0.196
SVM	0.971	0.980	0.971	0.962	0.021	0.037

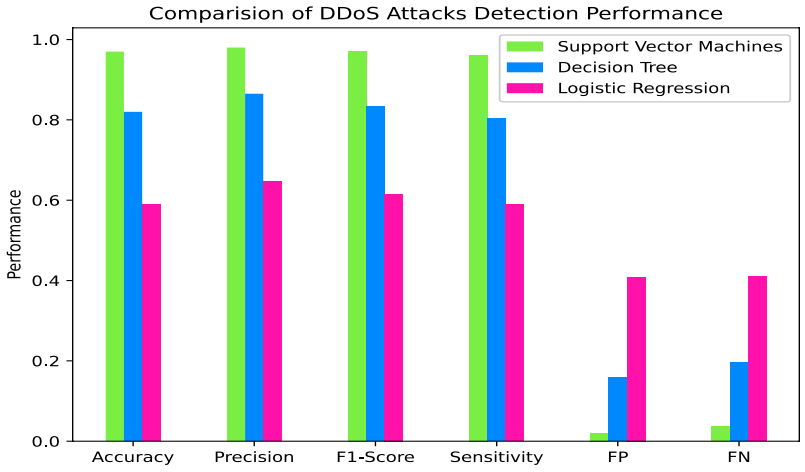

Figure 3. A bar diagram to show performance of three machine learning algorithms

We have compared the proposed approach with some exiting works in which our approach has high detection accuracy of the cyber attack while comparing random forest classifiers. As our accuracy is 97.1% which is higher than other classifiers expressed in Table 3. The closest accuracy is observed in the work of Mellor el al. [27] which is 96%. Moreover, Agarwal et al. [24] used SVM and it has MSE score 0.81.

The true positive rate is the attempts of DDoS attacks correctly identified by the algorithm. classified correctly, and true negative rate is not the attempts of DDoS attacks correctly identified by the algorithm. Also, the false positive rate is the proportion of benign attempts classified as DDoS attacks, and false negative rate is similar to the proportion of DDoS attacks classified as benign attempts.

Table 3. Comparison between existing works with proposed work

Classifiers	Authors	Dataset	Accuracy
Random Forest	Bharathidason et al., 2014 [26]	Multiple dataset	61-96%
Random Forest	Mellor et al., 2013 [27]	776 Land Cover Maps	96%
Random Forest	Almseidin et al., 2017 [28]	KDD Intrusion	93.7%
Random Forest	Bindra et al., 2019 [25]	CIC IDS2017	96.2%
SVM	Agrawal et al., 2011 [24]	--	0.81 (MSE)
SVM	Proposed Approach	Dataset of Canadian Institute of Cybersecurity	97.1%

Let us denote the number of benign attempts classified as benign as $\text{benign}_{\text{benign}}$, the number of benign attempts classified as DDoS attacks as $\text{benign}_{\text{DDoS}}$, the number of DDoS attacks classified as benign as $\text{DDoS}_{\text{benign}}$, and the number of DDoS attacks classified as DDoS attacks as $\text{DDoS}_{\text{DDoS}}$. We then define fp, the false positive rate, as

$$fp = \frac{benign_{DDoS}}{benign_{DDoS} + benign_{benign}} \tag{4}$$

and fn, the false negative rate, as

$$fn = \frac{DDoS_{benign}}{DDoS_{benign} + DDoS_{DDoS}} \tag{5}$$

Following this definition, fp = 0.021 will correspond to two of every 100 benign attempts being classified as DDoS, and fn = 0.037 would correspond to three of every 100 DDoS attacks being classified as benign attempts. These terms fp and fn which are used in this work for showing the evaluations.

5. Conclusion

We have applied different machine learning algorithms for detecting the patterns of DDoS attacks. We also validate their performance for ranking the best ML algorithms for serving these purposes. In terms of detection accuracy, training time, running time, scalability, support vector machines (SVMs) achieve expected outcome and beat other techniques like decision tree, logistic regression, etc. while testing three different Machine Learning classifiers on the dataset. More importantly, we observed that SVMs have high detection accuracy of the DDoS attack among them, which is 97.1%. But, minimum accuracy is recorded for the logistic regression algorithm (59.3%). The range of the accuracy of our three classifiers is approximately 59.3% to 97.1%. These results encourage to do additional research for the detection of DDoS attack to protect servers to serve their assigned services. This work has some imitations. We will address to solve this issue in our next work.

References

[1] S. C. Lin and S. S. Tseng, “Constructing detection knowledge for DDoS intrusion tolerance”, Expert Systems with Applications, vol.27, no.3, pp.379-390, **(2004)** DOI: 10.1016/j.eswa.2004.05.016

[2] D. Anstee, D. Bussiere, G. Sockrider, and C. Morales, "Worldwide infrastructure security report", Arbor Netw., Burlington, MA, USA, Tech. Rep 9, **(2014)**

[3] R. Miao, R. Potharaju, M. Yu, and N. Jain, “The dark menace: Characterizing network-based attacks in the cloud”, Proceedings of the ACM Conference on Internet Measurement Conference, pp.169-182, **(2015)** DOI: 10.1145/2815675.2815707

[4] H. Ma, L. Liu, A. Zhou, and D. Zhao, “On networking of Internet of Things: Explorations and challenges”, IEEE Internet Things J., vol.3, no.4, pp.441-452, Aug, **(2016)** DOI: 10.1109/JIOT.2015.2493082

[5] P. G. Neumann, “Risks of automation: A cautionary total-system perspective of our cyber future”, Commun. ACM, vol.59, no.10, pp.26-30, Oct, **(2016)** DOI: 10.1145/2988445

[6] X. Liu, S. Zhao, A. Liu, N. Xiong, and A. V. Vasilakos, “Knowledge aware proactive nodes selection approach for energy management in Internet of Things”, Future generation computer systems, vol.92, pp.1142-1156, **(2019)** DOI: 10.1016/j.future.2017.07.022

[7] Y. Liu, A. Liu, S. Guo, Z. Li, Y.-J. Choi, and H. Sekiya, “Context-aware collect data with energy efficient in cyber physical cloud systems”, Future generation computer systems, vol.105, pp.932-947, **(2020)** DOI: 10.1016/j.future.2017.05.029

[8] R. T. Kokila, S. T. Selvi, and K. Govindarajan, “DDoS detection and analysis in SDN based environment using support vector machine classifier”, In 2014 IEEE Sixth International Conference on Advanced Computing (ICoAC), pp.205-210, **(2014)** DOI: 10.1109/ICoAC.2014.7229711

[9] N.-N. Dao, J. Park, M. Park, and S. Cho, “A feasible method to combat against DDoS attack in SDN network”, In 2015 IEEE International Conference on Information Networking (ICOIN), pp.309-311, **(2015)** DOI: 10.1109/ICOIN.2015.7057902

[10] O. Flauzac, C. Gonzalez, A. Hachani, and F. Nolot, “SDN based architecture for IoT and improvement of the security”, In 2015 IEEE 29th international conference on advanced information networking and applications workshops, pp.688-693, **(2015)** DOI: 10.1109/WAINA.2015.110

[11] C. Li, Z. Qin, E. Novak, and Q. Li, "Securing SDN infrastructure of IoT–fog networks from MitM attacks", IEEE Internet of Things Journal, vol.4, no.5, pp.1156-1164, **(2017)** DOI: 10.1109/JIOT.2017.2685596

[12] M. E. Ahmed and H. Kim, “DDoS attack mitigation in Internet of Things using software defined networking", In 2017 IEEE third international conference on big data computing service and applications (BigDataService), pp.271-276, **(2017)** DOI: 10.1109/BigDataService.2017.41

[13] W. Hu, Wenjie, Y. Liao, and V. R. Vemuri, “Robust Support Vector Machines for Anomaly Detection in Computer Security”, In ICMLA, pp.168-174, **(2003)**

[14] S. Ranjan, R. Swaminathan, M. Uysal, and E. W. Knightly, “DDoS resilient scheduling to counter application layer attacks under imperfect detection” In INFOCOM, Citeseer, (2006) DOI: 10.1109/INFOCOM.2006.127

[15] Mcafee lab threat report, http://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q1-2015.pdf

[16] S. Noh, C. Lee, K. Choi, and G. Jung. “Detecting distributed denial of service (ddos) attacks through inductive learning”, in International Conference on Intelligent Data Engineering and Automated Learning, Springer, pp.286-295, **(2003)** DOI: 10.1007/978-3-540-45080-1_38

[17] T. Shon and J. Moon. “A hybrid machine learning approach to network anomaly detection”, Information Sciences, vol.177, no.18, pp.3799-3821, **(2007)** DOI: 10.1016/j.ins.2007.03.025

[18] N. B. Amor, S. Benferhat, and Z. Elouedi, “Naive bayes vs decision trees in intrusion detection systems”, in Proceedings of the 2004 ACM symposium on Applied computing, ACM, pp.420-424, **(2004)** DOI: 10.1145/967900.967989

[19] J. Mirkovic, G. Prier, and P. Reiher, "Attacking ddos at the source”, In 10th IEEE International Conference on Network Protocols”, pp.312-321, **(2002)** DOI: 10.1109/ICNP.2002.1181418

[20] “Source-end ddos defense”, In Second IEEE International Symposium on Network Computing and Applications, pp.171178, **(2003)** DOI: 10.1109/NCA.2003.1201153

[21] Application of Machine Learning. https://medium.com/app-affairs/9-applications-of-machine-learning- from-day-to-day-life-112a47a429d0, **(2018)**

[22] A. Dey, “Machine learning algorithms: A review”, International Journal of Computer Science and Information Technologies, vol.7, no.3, pp.1174–1179, **(2016)** DOI: 10.21275/ART20203995

[23] Logistic Regression, https://machinelearningmastery.com/logistic-regression-for-machine-learning/, December, **(2017)**

[24] P. K. Agrawal, B. B. Gupta, and S. Jain, "SVM based scheme for predicting number of zombies in a DDoS attack". 2011 European Intelligence and Security Informatics Conference, Athens, pp.178-182, **(2011)** DOI: 10.1109/EISIC.2011.19

[25] N. Bindra and M. Sood. "Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset", Automatic Control and Computer Sciences, vol.53, no.5, pp.419-428, **(2019)** DOI: 10.3103/S0146411619050043

[26] S. Bharathidason and C.J. Venkataeswaran, "Improving classification accuracy based on random forest model with uncorrelated high performing trees", Int. J. Comput. Appl., vol.101, no.13, pp.26-30, **(2014)** DOI: 10.5120/17749-8829

[27] A. Mellor, A. Haywood, C. Stone, and S. Jones, "The performance of random forests in an operational setting for large area sclerophyll forest classification", Remote Sens., vol.5, no.6, pp.2838–2856, **(2013)** DOI: 10.3390/rs5062838

[28] M. Almseidin, S. Alzubi, and K. M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system", 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), pp.277-282, **(2017)** DOI: 10.1109/SISY.2017.8080566

[29] I. Sreeram and V. P. K. Vuppala, "HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm", Applied computing and informatics, vol.15, no.1, pp.59-66, **(2019)** DOI: 10.1016/j.aci.2017.10.003

[30] R. Patgiri, U. Varshney, T. Akutota T., and R. Kunde, "An investigation on intrusion detection system using machine learning", In 2018 IEEE Symposium Series on Computational Intelligence (SSCI), pp.1684-1691, **(2018)** DOI: 10.1109/SSCI.2018.8628676

Authors

Md Abdur Rahman

Mr. Md. A. Rahman has been currently working as an Associate Professor of Computer Science at the department of Mathematics, Jahangirnagar University, Dhaka, Bangladesh. He received B.Sc. degree from same university in 2001, and MS in Computational Science from Texas A&M University-Commerce, TX, USA, in 2014. From 2012 to 2014, he was a Graduate Teaching Assistant at Computer Science and Information Systems in Texas A&M University-Commerce, TX, USA. His research interests include multi-agent path planning (robotics), deep learning, smart agriculture, and networking. Also, he worked to develop single page web application for several prestigious projects of United States of America; these sound experience helps him to handle different type of projects in an excellent way. Currently, Mr. Rahman is focusing to do research on cyber security and smart agriculture using machine learning and deep learning. Moreover, Mr. Rahman is working as an executive member of International Conference on Information Technology (www.icit.org) since 2018.

This page is empty by intention.