BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 197-205
DOI: 10.2478/bjlp-2022-004020

Detection and Prevention of DOS Attacks in Cloud Data Using J48 Algorithm Compared with Random Forest Algorithm for Improved Prediction Rate

S.Yuga Sai Sekhar
Research Scholar, Department of Computer Science and Engineering,Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105

P.Sriramya
Project Guide, Corresponding Author, Department of Computer Science and Engineering,
Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The aim of this research is to detect the Denial of Service (DOS) attacks using two machine learning algorithms, the Novel J48 algorithm and Random forest algorithm and compare accuracy to evaluate efficiency of two machine learning algorithms. **Materials and Methods:** Considering Multiple Novel J48 algorithms as group 1 and random forest algorithms as group 2 process was implemented to predict DoS attacks and to get a prediction rate to compare algorithms. The algorithm should be efficient enough to detect the exact type of DoS attack . The sample size considered for implementing this work was N=20 for each of the groups considered. The sample size calculation was done with spss. The pretest analysis was kept at 80%. Sample size is estimated using G-power. **Results:** Based on statistical analysis, the significance value for calculating accuracy was found to be 0.048. The Novel J48 Algorithm gives a slightly better accuracy rate with a mean Flow_Packets_Sec percentage of 89.69% and Random forest algorithm has a mean Flow_Packets_Sec of 75.29% with a significant value of two tailed tests is 0.048 ($p<0.05$) with 95% confidence interval. **Conclusion**: Through this, prediction is done for detection of DoS attacks and the Novel J48 algorithm gives a slightly better prediction rate value than the Random forest algorithm.

Keywords

Novel J48 Algorithm, Random Forest Algorithm, Prediction rate, DataSet, Machine Learning, DOS attacks

INTRODUCTION

The aim of this research is to predict DoS attacks with machine learning algorithms using Novel J48 algorithm compared with Random forest algorithm. The main importance of this study is prediction of DoS attack and its type. This helps to detect DoS attacks more efficiently and we can also detect the type of DoS attack. (D. Zhang et al. 2021) Nowadays

data security is important in everybody's life. Confidential organizations like ISRO, DRDO, NASA should secure their files from the hackers. (Shah and Kasbe 2021) So it is important to detect the attacks. DoS attacks are used to slow down the service and hack the devices. So,detection of DoS attacks is important. It is used in many companies to detect the vulnerability of their systems. Many companies hire hackers to detect the vulnerabilities. This research is used in cyber security to find vulnerabilities of a device.

When a DoS assault is launched against an IoT network and the network is flooded with massive amounts of traffic(Alenezi and Reed 2013), services are unavailable, network defences are ineffective, and the availability factor is compromised (Alenezi and Reed 2013; Tavares and Ruiz 2021). Although most IDPS use one or more detection methodologies DoS and DDoS Attack Detection Using Deep Learning and IDS 657 classified into two categories, signature-based or anomaly-based(Nikolskaia and Minbaleev 2020), the existence of an Intrusion Detection and Preventing System (IDPS) has little chance of withstanding a DoS attack (Scarfone and Mell 2007). Every year more than 30 articles are published. This is the trending topic at present; many articles are cited under this topic of research. Scholars are working on this to improve data security and reduce the data breach.

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020). Till now various techniques are used to detect DoS attacks.(H. Zhang et al. 2017). But all the techniques give the poor accuracy of detecting the type of DoS attacks. In existing research there are different Vulnerabilities/ loopholes which can give false results which leads to the low accuracy of detecting DoS attacks. (Djanie et al. 2019) These techniques gave bad results. The aim of this research is to detect DoS attacks using different algorithms. By comparing them we came up with the Novel J48 algorithm which gives accurate results. This can predict the type of DoS attacks with good accuracy. By this we can easily handle the attack and save the device from data breach. (Bakhtiar, Pramukantoro, and Nihri 2019)

MATERIALS AND METHODS

The research was done in a data analytic lab in Saveetha School of Engineering. This research requires the data samples of server workflow. The data set of a server workflow should be taken during the DoS attack on a system. This research includes two algorithms in which group 1 is the Novel J48 algorithm compared with group 2 of random forest algorithm.We took 20 samples of each algorithm to get the accurate results. The data set was taken from a device that iterated 10 times to get the desired accuracy with G power of 80%.(anaN et al. 2018)

Random Forest Algorithm

Random forest regression is the existing algorithm in this work. Novel exploratory data analysis is applied to analyze input data and to summarize their main characteristics. The training dataset goes through novel exploratory data analysis to extract the main feature for data extraction. Random forest algorithm is a tree-based algorithm that uses quality features of multiple decision trees. Random forest is a supervised algorithm where a training dataset is given as input and predicted values are found. The decision tree algorithms have disadvantages like low accuracy in executing and inaccurate predictions. These disadvantages can be solved using (Li and Wang 2022) the Random forest algorithm. Here in this algorithm, the data is divided into tree sets, and the program is executed in different types of ways where accuracy is found.(Li and Wang 2022). Fig. 1 gives the pseudocode for the Random Forest algorithm.

J48 Algorithm

The J48 algorithm is one of the best machine learning algorithms to examine the data categorically and continuously. When it is used, for instance, it occupies more memory space and depletes the performance and accuracy in classifying medical data(Hilal et al. 2021).It is based on a top-down strategy, a recursive divide and conquer strategy. You select which attribute to split on at the root node, and then you create a branch for each possible attribute value, and that splits the instances into subsets, one for each branch that extends from the root node.(anaN et al. 2018) Fig. 2 represents the pseudo code for the J48 algorithm.

First we should get the data set of a server workflow of DoS attack and filter the data set with the required parameters. Take the sufficient data Store in a data frame. Exploratory analysis (Daud et al. 2018) should be done with this data frame. We should get the accuracy and detection rate using a random forest algorithm.

For the second sample we should use the same data set of server workflow of device attacks and filter the data set with required parameters. Take the required amount of data and store it in a data frame. perform the exploratory analysis on the data set. get the accuracy detection rate by using the J48 algorithm. The Pseudo code for J48 algorithm is given in Fig. 2.

Google co lab(version 2.1 x) is used for research. open the colab and Mount the drive. Upload the required dataset to the colab workspace. store the dataset in required variables. write the code for the expression detection using random forest and J48 algorithms. run the code. the output accuracy and detection rate will be displayed in the output space. now use various samples and get the outputs. calculate the prediction rate with SPSS.

The data set is taken from an article "DoS Detection" from the source code `` kaggle.com. We used colab software to perform the random forest and Novel J48 algorithms for the stimulation. Each algorithm has twenty samples. The detection rate is calculated by SPSS software. The independent variables in this research are data packets and data flow with different rates. The dependent variables are accuracy and detection rate.The detection rate has been calculated with sample outputs.

Statistical Analysis

In this research the working Statistical tool called International Business Machines (IBM)Statistical Package for Social Sciences (SPSS) V22.0. The accuracy values are determined using descriptive and group statistics provided by this software. The significant values of independent sample tests are determined. According to the comparison of Novel J48 Algorithm and Random Forest Algorithm on all platforms, Novel J48 Algorithm looks to outperform Random Forest Algorithm. In this dataset Packet_length and FTA_flow are independent variables that stay constant when other parameters are changed. So, this SPSS tool formulates the values and gives the graph and the prediction rate.

RESULTS

The accuracy of the Novel J48 algorithm is 96.67 percent, whereas Random Forest accuracy is 90.91 percent, according to the results obtained. The outcomes are measured in terms of accuracy for the specified inputs. The results are obtained using the IBM SPSS programme. Statistical significance was found between the Predictive algorithm and has an accuracy of $p = 0.048$ (<0.005, 2-tailed), which is more accurate than the value, according to the data.

In Table 2, for each sample size the Accuracy is calculated for both of the respective algorithms and filled in the respective columns. It was observed that the increase in sample sizes increased the accuracy of both the algorithms. At last Average Accuracy is calculated and stored. The average Accuracy of the Novel J48 Algorithm is 96.67% and Random Forest is 90.91%. In performing statistical analysis of 5 samples, the Novel J48 model obtained 0.7094 standard deviation with 0.2244 standard error while Random Forest obtained 0.8178 standard deviation with 0.2586 standard error.

In Table 3, Independent t-tests were used to compare the accuracy of two algorithms and a statistically significant difference was noticed. Fig. 3 shows the Statistical significance difference was observed between the Novel J48 algorithm and has an accuracy of 96.67%, Random Forest Algorithm 90.91%, p = 0.048 (<0.05, 2-tailed). SPSS is applied for a dataset fixing confidence interval as 95% and level of significance as 0.05. The mean difference and standard error difference for two algorithms is tabulated. The accuracy of our Novel J48 algorithm in this suggested paper is 96.67 percent. Because we use a well-balanced and limited dataset to create our Random Forest, it has a higher accuracy than the other techniques.

DISCUSSION

This proposed paper observed that the Novel J48 algorithm has the best performance with an accuracy rate of 96.7% when compared to the Random Forest algorithm with an accuracy rate of 90.91%. For the Detection of DOS attacks , the performance of Novel J48 algorithm and Random Forest algorithm has been measured for the dataset taken from GitHub and a database collected from Kaggle repository.

In this research we found that the prediction rate is more accurate when we use "Avg_packets ", "Flow_Packets_Sec '' as parameters. Moreover, the performance of the Novel J48 algorithm is more satisfying compared to the Random forest algorithm. In the Random forest algorithm the output graph fluctuates when the input differs, whereas the Novel J48 algorithm gives the perfect and stable graph. (H. Zhang et al. 2017) This demonstrates that J48 is a basic decision tree classification approach. Weka technology was used to extract efficient results from a bank dataset. experiment.(Hermawan et al. 2021) Also visible is the Naive Bayes classifier good outcomes. The Novel J48 algorithm gives the perfect prediction rate in many of the cases, except when the packet attack rate is more than 1000 at a single attempt.(Sahu and Mehtre 2015)

Novel J48 algorithm is relatively expensive as the complexity and time taken are more.sometimes calculation can be far more complex compared to other algorithms.The future scope of this project is to reduce the cost and time complexity of Novel J48 algorithm. The main objective is to increase speed and reduce the complexity by changing the pseudo code and the data algorithms.

CONCLUSION

In this study, DOS detection using two distinct algorithms, Novel J48 Algorithm and Random Forest Algorithm. J48 algorithm performed better than the Random forest algorithm and certain other techniques in making generalizations from the testing phase to the validation set. The Novel J48 Algorithm showed a higher accuracy rate (96.67%) and accomplished better at a more accurate level than that of the Random Forest algorithm (90.91%).

DECLARATION

Conflict of Interests
No conflict of interests in this manuscript

Authors Contribution
Author SYSS was involved in data collection, data analysis, and manuscript writing. Author PSR was involved in conceptualization, data validation, and critical review of manuscripts.

Acknowledgement
The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary Infrastructure to carry out this work successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. Inoaura Technologies, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical And Technical Sciences.
4. Saveetha School of Engineering

REFERENCES

Alenezi, Mohammed, and Martin J. Reed. 2013. "Denial of Service Detection through TCP Congestion Window Analysis." *World Congress on Internet Security (WorldCIS-2013)*. https://doi.org/10.1109/worldcis.2013.6751036.

anaN, N. Sarav, N. Sarav anaN, Periyar University, and V. Gaya thri. 2018. "Performance and Classification Evaluation of J48 Algorithm and Kendall's Based J48 Algorithm (KNJ48)." *International Journal of Computer Trends and Technology*. https://doi.org/10.14445/22312803/ijctt-v59p112.

Bakhtiar, Fariz Andri, Eko Sakti Pramukantoro, and Hilman Nihri. 2019. "A Lightweight IDS Based on J48 Algorithm for Detecting DoS Attacks on IoT Middleware." *2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech)*. https://doi.org/10.1109/lifetech.2019.8884057.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Daud, Nurraina, Nor Laila Mohd Noor, Syed Ahmad Aljunid, Nurulhuda Noordin, and Nur Islami Mohd Teng. 2018. "Predictive Analytics: The Application of J48 Algorithm on Grocery Data to Predict Obesity." *2018 IEEE Conference on Big Data and Analytics (ICBDA)*. https://doi.org/10.1109/icbdaa.2018.8629623.

Djanie, Djanie, Tutu, and Dzisi. 2019. "A Proposed DoS Detection Scheme for Mitigating DoS Attack Using Data Mining Techniques." *Computers*. https://doi.org/10.3390/computers8040085.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

Hermawan, Dicky Rahma, Mohamad Fahrio Ghanial Fatihah, Linda Kurniawati, and Afrida Helen. 2021. "Comparative Study of J48 Decision Tree Classification Algorithm, Random Tree, and Random Forest on In-Vehicle CouponRecommendation Data." *2021 International Conference on Artificial Intelligence and Big Data Analytics*. https://doi.org/10.1109/icaibda53487.2021.9689701.

Hilal, Anwer Mustafa, Fahd N. Al-Wesabi, Masoud Alajmi, Majdy M. Eltahir, Mohammad Medani, Mesfer Al Duhayyim, Manar Ahmed Hamza, and Abu Sarwar Zamani. 2021. "Machine Learning Based Decision Tree J48 with Grey Wolf Optimizer for Environmental Pollution Control." *Environmental Technology*, December, 1–23.

Li, Xinxin, and Lina Wang. 2022. "Financial Data Center Configuration Management System Based on Random Forest Algorithm and Few-Shot Learning." *Computational Intelligence and Neuroscience* 2022 (January): 9051629.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Nikolskaia, Kseniia, and Aleksey Minbaleev. 2020. "Legal Regulation of Incidents Related to DDoS Attacks." *2020 International Conference Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS)*. https://doi.org/10.1109/itqmis51053.2020.9322874.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral*

Health & Preventive Dentistry 18 (1): 379–86.
Sahu, Shailendra, and B. M. Mehtre. 2015. "Network Intrusion Detection System Using J48 Decision Tree." *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. https://doi.org/10.1109/icacci.2015.7275914.
Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.
Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.
Scarfone, K. A., and P. M. Mell. 2007. "Guide to Intrusion Detection and Prevention Systems (IDPS)." https://doi.org/10.6028/nist.sp.800-94.
Shah, Priyanka, and Tanmay Kasbe. 2021. "Detecting Sybil Attack, Black Hole Attack and DoS Attack in VANET Using RSA Algorithm." *2021 Emerging Trends in Industry 4.0 (ETI 4.0)*. https://doi.org/10.1109/eti4.051663.2021.9619414.
Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.
Tavares, Juliano C. C., and Linnyer B. Ruiz. 2021. "Towards a Novel Edge to Cloud IoMT Application for Wildlife Monitoring Using Edge Computing." *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)*. https://doi.org/10.1109/wf-iot51360.2021.9596017.
Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.
Yasin, Waheed, Hamidah Ibrahim, Nur Izura Udzir, and Nor Asilah Wati Hamid. 2014. "Intelligent Cooperative Least Recently Used Web Caching Policy Based on J48 Classifier." *Proceedings of the 16th International Conference on Information Integration and Web-Based Applications & Services*. https://doi.org/10.1145/2684200.2684299.
Zhang, Dan, Zehua Ye, Gang Feng, and Hongyi Li. 2021. "Intelligent Event-Based Fuzzy Dynamic Positioning Control of Nonlinear Unmanned Marine Vehicles Under DoS Attack." *IEEE Transactions on Cybernetics* PP (December). https://doi.org/10.1109/TCYB.2021.3128170.
Zhang, Heng, Yifei Qi, Huan Zhou, Jian Zhang, and Jing Sun. 2017. "Testing and Defending Methods Against DOS Attack in State Estimation." *Asian Journal of Control*. https://doi.org/10.1002/asjc.1441.

TABLES AND FIGURES

Table 1. The below table shows the 10 iterations of the Random Forest algorithm and J48 algorithm with different iterations and their extracted accuracies.

S.NO	J48	RANDOMFOREST
1	99.89	93.89
2	95.92	95.23
3	97.3	89.8

4	90.78	92.38
5	92.89	95.89
6	93.78	90.45
7	100	92.34
8	97.9	86.39
9	98.4	87.3
10	89.8	90.49

Table 2. Independent Variable:The statistical calculations for independent samples T test between Novel J48 Algorithm and Random Forest Algorithm. This independent sample test consists of significance as 0.048, significance (2-tailed).

Accuracy	Levene's test for equality of variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tail-ed)	Mean Differenc e	Std. Error differenc e	95% Confidence Interval of the Difference	
								Lowe r	Uppe r
Equal variance s assumed	4.16 5	0.04 8	9.17 8	38	<.00 1	0.67	0.073	0.52 3	0.81 8
Equal variance s not assumed			9.17 8	34.15 8	<.00 1	0.67	0.073	0.52 2	0.81 8

Table 3. Group Statistics:Novel J48 Algorithm and Random Forest are the two machine learning algorithms used in this statistics.Sample size N = 20. Mean for Novel J48 Algorithm is 96.725 and Random Forest is 90.025. Std. Deviation for Novel J48 Algorithm is 2.6681 and Random Forest is 1.8815. Std Error Mean for J48 is 0.5966 and Random Forest is 0.4207.

	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	J48	20	96.725	2.6681	0.5966
	Random Forest	20	90.025	1.8815	0.4207

Algorithm 1 Random Forest

Precondition: A training set $S := (x_1, y_1), \ldots, (x_n, y_n)$, features F, and number of trees in forest B.

```
function RandomForest(S, F)
    H ← ∅
    for i ∈ 1, ..., B do
        S^(i) ← A bootstrap sample from S
        h_i ← RandomizedTreeLearn(S^(i), F)
        H ← H ∪ {h_i}
    end for
    return H
end function
function RandomizedTreeLearn(S, F)
    At each node:
        f ← very small subset of F
        Split on best feature in f
    return The learned tree
end function
```

Fig. 1. Pseudo Code for Random Forest Algorithm

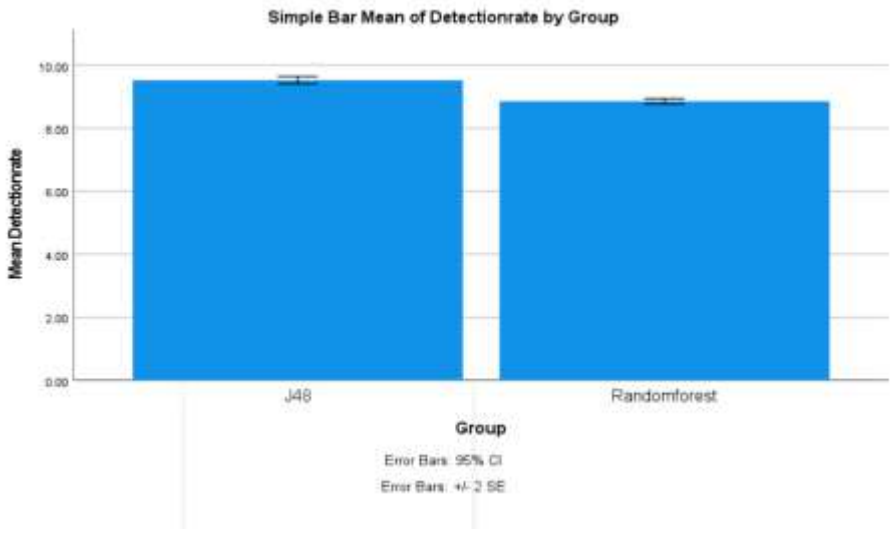

Fig. 3. Bar Graph showing Comparison on mean accuracy of J48 (96.7%) and RFA (90.1%). X-axis: J48, RFA, Y-axis: Mean Accuracy with ± 1 SD.

Algorithm of J48 (D)
Input: a dataset D
begin
Tree = {}
If (D is "pure") | |(other stopping criteria met)then terminate;
For all attribute a $a \in D$ D do
Compute criteria of impurity function if we split on a;
a_{best}= Best attribute according to above computed criteria
Tree = Create a decision node that tests a_{best} in the root
D_v= Induced sub-datasets from D based on a_{best}
For all D_v do
begin
Tree $_v$= J48(D_v)
Attach Tree $_v$ to the corresponding branch of Tree
end
return Tree
end

Fig. 2. Pseudo Code for J48 Algorithm(Yasin et al. 2014)