

A Novel Sampling Technique for Detecting Cyber Denial of Service Attacks on the Internet of Things

Bassam Kasasbeh¹ ^a and Hadeel Ahmad²  ^b

¹Department of Data Science & Artificial Intelligence, Al Hussein Technical University, Amman, Jordan

²Department of Computer Science, Applied Science Private University, Amman, Jordan

Keywords: Internet of Things (IoT), Denial of Service (DoS) Attacks, Intrusion Detection Systems (IDS), Multi-Class Imbalanced Data, Machine Learning.

Abstract: Internet of Things (IoT) devices are vulnerable to a wide range of unique security risks during the data collection and transmission processes. Due to a lack of resources, these devices increased the attack surface and made it easier for an attacker to find a target. The Denial of Service (DoS) attack is one of the most common attacks that can target all layers of the IoT protocol. Therefore, Intrusion Detection Systems (IDS) based on machine learning (ML) are the best ways to confront these risks. However, an imbalanced dataset for cyber attacks makes it difficult to detect them with ML models. We propose an undersampling technique that clusters the data set using Fuzzy C-means (FCM) and picks similar instances with the same features to ensure the integrity of the dataset. We used accuracy, precision, sensitivity, specificity, F-measure, AUC, and G-means to determine how good the results were. The proposed technique had 97.6% overall accuracy. Furthermore, it got 96.94%, 96.39%, 99.59%, 98.08%, and 97.16% True Positive Rates (TPR) in the Blackhole, Grayhole, Flooding, Scheduling, and Normal (no attacks) classes, respectively. The results show that the proposed method for detecting DoS attacks in the IoT has succeeded.

1 INTRODUCTION


Nowadays, the rapid growth of emerging and Internet-based decentralized technologies like the Internet of Things (IoT) and cloud computing has led to an explosion of information in virtually every technical and commercial field that exists today (Stoyanova et al., 2020). IoT refers to the interconnected networks of devices that enable the seamless exchange of information between various physical devices. These devices could be industrial robots, medical and healthcare devices, wearables, smart TVs, smart city infrastructures, or driverless cars (Mbarek et al., 2020). In addition, many objects and intelligent sensors are linked to exchange data through the IoT without human intervention (Jan et al., 2019).


Additionally, a Wireless Sensor Network (WSN) enables access to many IoT objects via a wide range of sensors and actuators accessed over the internet. However, IoT sensors typically have limited memory, power, and a tiny battery, making it difficult to com-

pute, store, access, and analyze IoT data. Additionally, a growing volume of diverse data and objects necessitates a platform to accommodate it all (Masengo Wa Umba et al., 2022)(Jiang et al., 2020).

IoT devices will be more common than mobile devices and will have access to sensitive personal information (Meneghello et al., 2019). Unfortunately, this means that it will be easier to attack and that there will be more vulnerabilities in the IoT in general (Islam et al., 2022). Therefore, Intrusion Detection Systems (IDS) are needed to protect IoT communications because security is essential to most IoT applications. IDS is software or hardware that monitors data traffic to detect malicious activities and protect end users from intrusions threatening an information system's real-time availability, integrity, and privacy (Abiodun et al., 2021)(Islam et al., 2022).

Therefore, it is essential to figure out what affects how well the IDS works in IoT apps and come up with a plan to make the detection process better and more efficient. Class imbalance in datasets is one of the most critical challenges requiring more research in IDS (Tabbaa et al., 2022). The dataset is deemed imbalanced when the classes are represented in differ-

^a  <https://orcid.org/0000-0002-3240-3002>

^b  <https://orcid.org/0000-0001-8595-9891>

ent proportions (Ahmad et al., 2022). For example, when looking at datasets to study the effects of cyber attacks on network traffic flow, most of the data is normal (no attack behaviour), and only a tiny amount is attack data (Churcher et al., 2021).

Thus, it is essential to combat Denial of Service (DoS) attacks. Even though work on detecting DoS attacks has become more popular in the past few years, it is still a big problem for IoT apps today (Adat and Gupta, 2018). DoS attacks on IoT apps usually have significant effects, mainly because limited sensor devices make them (Adat and Gupta, 2018). IDS is regarded as one of the most effective methods for detecting DoS attacks. IDSs monitor system activity in order to identify and block malicious traffic. Attacks can be easily detected by figuring out network traffic's normal pattern and size (Almomani et al., 2016).

This paper uses an undersampling technique for multiclass data balancing to build a classification IDS for IoT apps called Multiclass Similarity-Based Selection (MSBS). We used the WSN-DS dataset, a multiclass imbalanced dataset with five cyber-DoS attacks labeled blackhole attack, grayhole attack, flooding attack, scheduling attack, and normal (no attacks). The proposed technique balances the dataset by reducing the sample size of the majority classes.

We compared the proposed method to Random Undersampling (RUS) (Leevy et al., 2021) and the multi-label approach for Tomek Link undersampling (MLTL) (Pereira et al., 2020) to test it. In order to evaluate the three undersampling techniques, we used the widely used machine learning algorithms named K-Nearest neighbours (kNN), Logical Regression, and Naive Bayes. In addition, the evaluation parameters accuracy, precision, sensitivity, specificity, F-measure, area under the curve (AUC), and G-means were used to compare the classification performance between the proposed technique and the other two undersampling techniques.

The following is a summary of the main contributions of this study:

1. In the paper, an undersampling technique was developed for IDS to find cyber-DoS attacks, and its effectiveness in a big data environment was confirmed.
2. It has been shown that the MSBS undersampling technique is better at finding cyber-DoS attacks in IoT apps than the other undersampling techniques in the literature.
3. The proposed method is evaluated with three distinct machine learning classification algorithms to assess its effectiveness. The results showed that

the proposed method performed significantly better than the methods described in the literature.

The rest of the paper is organized: Section 2 provides review-related work. Section 3 provides an overview of the WSN-DS dataset used to classify cyber-DoS attacks and the proposed MSBS undersampling technique. Section 4 presents the results and discussion. Finally, conclusions and suggestions for future research are presented in Section 5.

2 LITERTIAL REVIEW

In recent years, numerous IDSs have been proposed in the literature and are used to monitor IoT devices against various cyber-DoS attacks.

(Almomani et al., 2016) created a specialized dataset for WSN networks that he called WSN-DS. This dataset was based on the network traffic in wireless sensor nodes and included four types of cyber-DoS attacks: blackhole, grayhole, flooding, and scheduling. Using this dataset, the authors trained an artificial neural network (ANN) to detect and classify DoS attacks without considering the dataset's balancing. Experiments show that DoS attacks were more accurately detected when one hidden layer was used.

(Kumari and Mehta, 2020) developed an ensemble-based intrusion detection model using various ML classification algorithms, including Decision Tree, J48, and Support Vector Machine (SVM) The nine most relevant and significant intrusion detection features from the KDD99 dataset were determined using particle swarm optimization. The proposed model produced results that were 90% more accurate.

(Pokharel et al., 2020) present a hybrid IDS model of Naive Bayes and SVM. A real-time historical log dataset was normalized and preprocessed for this study. After enhancement, the proposed model achieved 95% accuracy and precision. In addition, it has been demonstrated that classifier performance improved when session-based features were added.

(Kumari and Mehta, 2020) evaluate Bayesian networks and RandomTree classifiers with ensemble learning. On the KDDcup99 dataset, the ensemble IDS model was compared to base classifiers for accuracy, precision, and recall. This study concludes that the proposed model has a better effect on precision and recall than the accuracy rate and claims that IDS presents a sound effect for the whole dataset, no matter the sample size. Furthermore, the Bayesian network performs better on small datasets, while RandomTree does better on large ones.

(Vinayakumar et al., 2019) proposed a scalable, hybrid DNN framework called Scale-Hybrid-

IDS-AlertNet to monitor network traffic and host-level events to alert for possible network attacks. DNN models that did well on KDD Cup 99 were benchmarked on the NSL-KDD, UNSW-NB15, Kyoto, WSN-DS, and CICIDS 2017 datasets. Experimental tests show that the DNN outperforms traditional machine-learning classifiers. However, complex DNN models require much computation.

In this paper, we proposed a novel undersampling technique for multiclass datasets to detect cyber-DoS attacks in IoT apps. The proposed technique works based on the similarity between the minority and majority instances to balance the dataset. This technique was applied to a high imbalance dataset of cyber-DoS attacks called WSN-DS (Almomani et al., 2016). In order to verify the success of the proposed technique in classifying cyber-DoS attacks, the WSN-DS dataset was balanced using RUS (Leevy et al., 2021) and MLTL (Pereira et al., 2020). RUS is an undersampling method that achieves balance by randomly removing instances from the majority classes. The MLTL technique, on the other hand, identifies and eliminates so-called Tomek links from the multiclass dataset. A pair of instances is a Tomek Link if they are neighbors but belong to different classes.

3 METHODOLOGY

This paper proposes a technique for the multi-class imbalance dataset problem. The key to this technique is to identify and then prioritize instances that are similar to one another. The proposed technique solves the class imbalance issue by employing a Fuzzy C-means (FCM) clustering strategy, which yields a robust set for the sampling phase (Aydilek and Arslan, 2013). The clustering step ensures that the instances will be grouped according to the degree to which their characteristics are relatively similar to ensure that the data will be analyzed in the most efficient manner possible (Ahmad et al., 2022). Then, we find the minority class in each cluster and find the euclidean distance between each instance in the minority class and every other instance to find instances with similar features. We do this for each cluster, and then we put the instances from each cluster together to produce a balanced dataset. This process ultimately aids in reducing the removal of relevant and essential instances that occur when employing the random undersampling technique. As a result, the efficiency of the IDS is improved, and at the same time, the dataset becomes balanced. Machine learning algorithms, including logistic regression, Naive Bayes, and K-nearest neighbour (kNN), are used during the classification stage to

classify cyber-DoS attacks. Figure 2 shows the MSBS framework.

3.1 Dataset

(Almomani et al., 2016) collected a dataset representing WSN features under a variety of different attack scenarios by using the LEACH protocol (Almomani and Al-Kasasbeh, 2015). This dataset is well known as WSN-DS. Since the LEACH protocol is widely used in WSNs and IoT, it was selected as the study's protocol (Behera et al., 2018). Twenty-three features were culled from the WSN-DS dataset using the LEACH routing protocol. Each sensor node's state in a wireless sensor network can be described with these characteristics. The total number of records in the WSN-DS dataset is 374661. This dataset simulates four different types of cyber-attacks: Grayhole attack (14596 records), Blackhole attack (10049 records), Scheduling attack (6638 records), and Flooding attack (3312 records). The other 340066 records show no attack behavior. Figure 1 shows the distribution of the WSN-DS dataset instances.

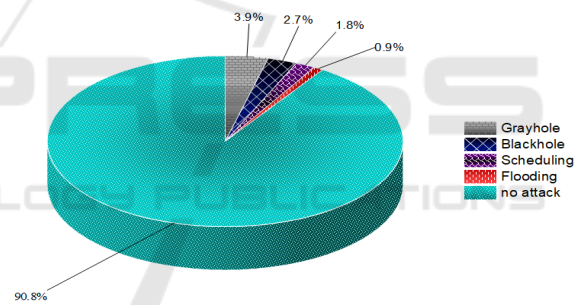


Figure 1: Distribution of the Dataset instances.

3.2 Fuzzy C-Means Clustering Algorithm

Clustering refers to organizing data into groups according to predetermined standards or parameters. One of the popular clustering techniques is the FCM clustering algorithm. FCM is a machine learning (ML) clustering approach that splits the dataset into two or more groups. Using FCM, each record in the dataset is clustered based on how similar it is to the others, representing the distance between the instance and the cluster center. To determine the optimal number of clusters, we tested the dataset using FCM with two clusters, three clusters, . . . up to ten clusters. This procedure aims to get as many clusters as possible while ensuring that each cluster has all types of attacks. After that, we found that the optimal number

of clusters is 4. Table 1 shows the distribution of each cluster.

3.3 Dataset Resampling

In order to deal with imbalanced datasets, it is necessary to modify classification algorithms to achieve improvements or equalization of classes within the training information. This procedure is known as data preprocessing (Ahmad et al., 2022). The primary goal of knowledge preprocessing is to either increase the number of instances that fall into the minority class or reduce the frequency of those that belong to the majority classes to get the same number of instances in each class (Patel et al., 2020).

After grouping the data into separate clusters based on their similarities, the next step in the sampling process is determining the minority class in each cluster. In the first cluster (c1), for example, Table 1 shows that the scheduling attack constitutes a minority class with 1885 instances. Then, it compares the distance between each minority instance (scheduling attack in this cluster) and the other samples in the same cluster to determine which instance from each class is closest to each minority instance. Thus, after applying this procedure, cluster C1 has 1885 instances. Next, we apply this procedure to all other clusters. Table 2 shows the number of samples after completing the selection process according to similarity. Lastly, this method combines the instances chosen from each cluster into a balanced dataset based on their similarities. Algorithm 1 shows the details of the proposed technique steps.

3.4 Performance Evaluation

We compared the classification efficacy of the proposed technique (MSBS) by using accuracy, precision, sensitivity (True positive rate), specificity (True negative rate), F-measure, AUC, and the geometric mean (G-Mean). These performance metrics directly result from the confusion matrix data used as the basis for their calculation. The confusion matrix comprises four primary elements: the True Positive, the True Negative, the False Negative, and the False Positive. To examine the model's efficacy regarding the influence of a highly imbalanced dataset, we opted to undertake an exhaustive study that includes all pertinent performance characteristics for a typical classification process. Equations 1 through 8 reflect all performance measures:

$$Accuracy = \frac{True\ Positive + True\ Negative}{\#of\ all\ samples} \quad (1)$$

Algorithm 1: Pseudo-code for the MSBS.

Input: Imbalanced multiclass Dataset **D**,
Number of classes $N_{classes}$, Number of
clusters $N_{clusters}$, Classes label **CL**

Output: Balanced multiclass Dataset **D'**

```

1 for  $i \leftarrow 1$  to  $N_{clusters}$  do
2    $N_{samples} = size(D[i])$ 
3   for  $j \leftarrow 1$  to  $N_{classes}$  do
4     for  $k \leftarrow 1$  to  $N_{samples}$  do
5        $SC[j] = count(CL[j])$ 
6     end
7   end
8    $MinorityClass[i] = min(SC)$ 
9   for  $r \leftarrow 1$  to  $N_{samples}$  do
10    Calculate the distance for every
        sample in minority class and the
        other samples.
11  end
12  Sort samples by distance in descending
        order.
13  Select observations near minority class
        samples.
14 end

```

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (3)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (4)$$

$$F - measure = \frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision} \quad (5)$$

$$AUC = \frac{1}{2} (Sensitivity + Specificity) \quad (6)$$

$$G - Mean = \sqrt{Sensitivity \times Specificity} \quad (7)$$

$$ErrorRate = \frac{False\ Positive + False\ Negative}{\#of\ all\ samples} \quad (8)$$

4 RESULTS AND DISCUSSION

This section uses the proposed method on the imbalanced WSD-DS dataset (Almomani et al., 2016). The study was carried out using the Google Colab platform. Several machine learning libraries were

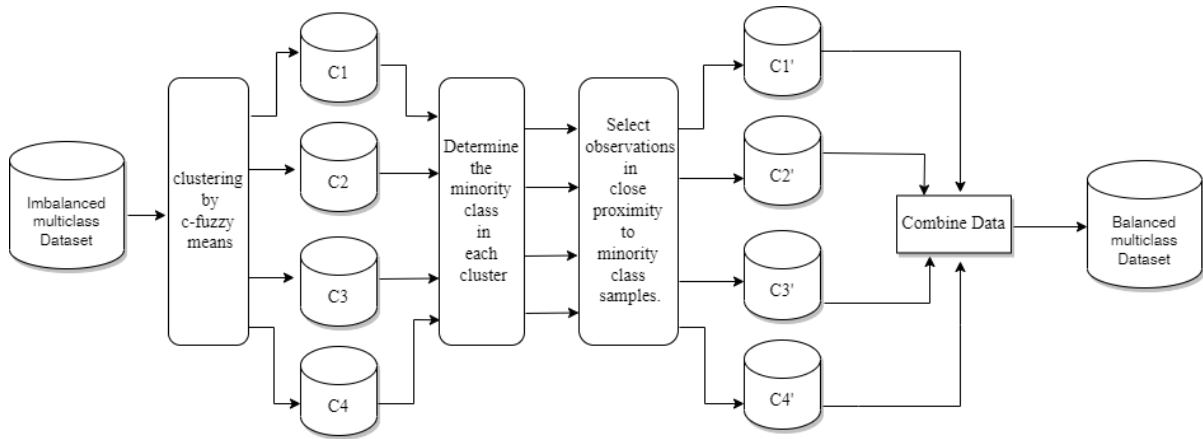


Figure 2: MSBS Framework.

Table 1: The FCM distribution of each cluster.

Cluster	Total number of instances	Grayhole	Blackhole	Scheduling	Flooding	No attacks
C1	54201	11458	7331	1885	2482	30946
C2	292956	104	238	1154	100	291459
C3	24564	2991	2377	924	700	17572
C4	2940	43	103	2675	30	89

used, including scikit-learn, sklearn, pandas, and matplotlib. The proposed method was tested with three different ML algorithms, and the results were interpreted. We analyzed and compared the proposed method with RUS and MLTL undersampling techniques in this section. We used three popular ML classifiers: kNN, logistic regression, and Naive Bayes. In this analysis, we use 70% of the total dataset for training and 30% for testing.

4.1 Comparison and Analysis Results with Different Machine Learning Techniques

In Tables 3, 4, and 5, given below, the results of the three undersampling techniques are compared using kNN, logistic regression, and Naive Bayes, respectively are shown. The results of an evaluation performed with the kNN algorithm are presented in Table 3. This evaluation shows how our proposed method differs from the other techniques by showing that accuracy is improved by 13.3% when using RUS and 6.84% when using MLTL. It was established that the proposed method was superior in terms of the outcomes of precision, sensitivity, and F-measure, with very similar percentages of superiority in accuracy. The problem with our proposed approach is that it only marginally improves specificity by 2.82% compared to RUS and 1.86% compared to MLTL, which means that the proposed method has few false posi-

tive results. However, the proposed method gave the highest G-mean and AUC compared to other undersampling techniques.

The outcomes of the comparison using logistic regression are displayed in Table 4. Our proposed method is superior to the other undersampling techniques. On the other hand, the difference in sensitivity between the RUS technique and the MLTL technique, which is 16.01% and 6.92%, respectively, is considered to be a slight difference when compared to our proposed method. However, it is important to note that the MLTL technique has more specificity than other technologies, which means that the logistic regression makes our proposed method less likely to give false positives.

In Table 5, we see a comparison of the results using Naive Bayes, which shows that all undersampling techniques are vastly superior to the other machine learning algorithms employed. The accuracy of our proposed method reached 97.2%, demonstrating its obvious and substantial superiority. Compared to other undersampling methods, the proposed method is 6.7% more accurate than the RUS method and 4.4% more accurate than the MLTL method. Furthermore, the results show that the proposed method is much better regarding sensitivity, F-measure, AUC, and G-means. In contrast, the results from Naive Bayes show much convergence in terms of accuracy and specificity.

Table 2: Number of instances after the selection process.

Cluster	Total number of instances	Grayhole	Blackhole	Scheduling	Flooding	No attacks
C1	9425	1885	1885	1885	1885	1885
C2	500	100	100	100	100	100
C3	3500	700	700	700	700	700
C4	150	30	30	30	30	30

Table 3: kNN results.

Techniques	Accuracy	Precision	Sensitivity	Specificity	F-Measure	AUC	G-MEAN
RUS	0.827	0.829	0.613	0.957	0.828	0.892	0.890
MLTL	0.877	0.878	0.693	0.966	0.878	0.922	0.920
MSBS	0.937	0.937	0.791	0.984	0.937	0.961	0.960

Table 4: Logistic Regression results.

Techniques	Accuracy	Precision	Sensitivity	Specificity	F-Measure	AUC	G-MEAN
RUS	0.897	0.926	0.706	0.983	0.911	0.94	0.939
MLTL	0.941	0.956	0.766	0.998	0.948	0.97	0.969
MSBS	0.965	0.966	0.819	0.992	0.965	0.979	0.978

Table 5: Naive Bayes results.

Techniques	Accuracy	Precision	Sensitivity	Specificity	F-Measure	AUC	G-MEAN
RUS	0.909	0.921	0.912	0.977	0.916	0.945	0.944
MLTL	0.943	0.947	0.945	0.986	0.946	0.966	0.965
MSBS	0.976	0.981	0.977	0.994	0.979	0.986	0.985

4.2 Analysis of Cyber-DoS Through a Naive Bayes Model

Figure 3 shows the True Positive Rate (TPR) for each type of cyber-DoS attack as classified by the Naive Bayes classifier, which is the best classifier for classifying the various cyber-DoS attacks described in the previous section. This figure shows that the proposed technique is significantly superior to the MITI technique in classifying flood attacks. The TPR for this attack reached 99.6% compared to 4.56% for the MITI technique. Also, the scheduling attack showed that the proposed technique was better because its TPR reached 98.1%, which was 4.72% higher than the MITI technique. Figure 3 also shows that the normal mode (no attack behaviour) ranked third among all attack types in the TPR. Unfortunately, the MITI technique gained the upper hand at this time, as the gap between the proposed method and the MITI method reached 0.8%.

Moreover, at the level of a blackhole attack, the proposed method stood out because its TPR reached 96.9%, whereas the MITI method only reached 93.33%. Finally, even though it is the weakest type of TPR cyber-DoS attack, the grayhole attack in the proposed technique was distinguishable from the other attacks. It reached 96.4%, which is 5.49% higher than the MITI technique.

Figure 4 shows the False Positive Rate (FPR). We note the distinction of the proposed technique in achieving a distinct relative in this aspect compared to other techniques in classifying four distinct types of cyber-DoS attacks, namely blackhole, grayhole, scheduling, and no attack behaviour. However, the MITI technique appears superior in classifying flooding attacks, with a 65.4% difference from the proposed technique.

Figure 4 also shows that the scheduling attack has a remarkably high FPR, reaching 4.81% with the RUS technique, 3.43% with the MITI technique, and 1.43% with the proposed technique. On the other hand, the flooding attack shows the highest likelihood of obtaining an FPR percentage, reaching 1.29% using the RUS technique, 0.23% using the proposed technique, and 0.08% using the MITI technique.

Figure 5 shows the error rate (EER) in classifying all cyber-DoS attacks in this dataset, with EERs of 0.27%, 0.83%, 0.99%, 1.12%, and 1.52% for flooding attacks and no attack behaviour, Grayhole, and Blackhole, respectively. We also note from this figure that the proposed technique had the lowest EER in classifying the normal state (no attack behaviour), which achieved an EER that was 36.35% less than the MITI method and 59.39% less than the RUS technique. Finally, It shows that the scheduling attack had the highest EER among all cyber-DoS attacks.

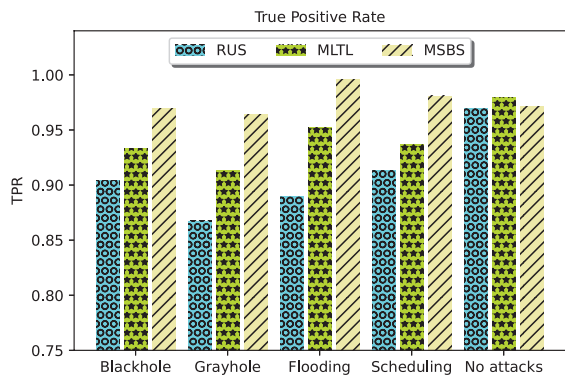


Figure 3: The True Positive Rate of all cyber-DoS attacks.

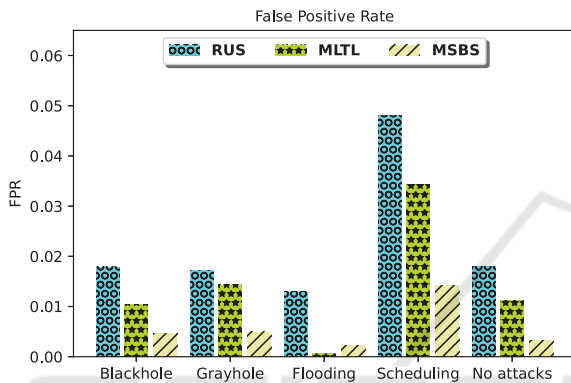


Figure 4: The False Positive Rate of all cyber-DoS attacks.

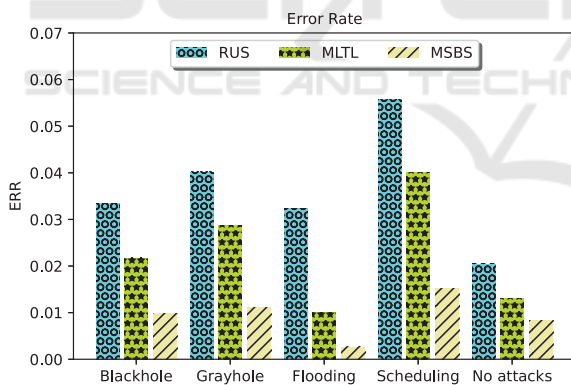


Figure 5: The Error Rate of all cyber-DoS attacks.

5 CONCLUSION AND FUTURE WORK

This paper suggests a new undersampling technique for dealing with IDSs to detect different cyber-DoS attacks specific to the IoT. In this study, we constructed the proposed technique using the WSN-DS dataset. The proposed technique solves the problem of imbalanced multiclass distribution for cyber-DoS attacks in the IoT by using the FCM clustering algorithm to group the entire dataset to find instances

with the same pattern and balance the dataset based on similarity. Experiments showed that the proposed method was more accurate (97.6

The proposed technique stood out from other proposed techniques because it was able to classify blackhole, grayhole, and scheduling attacks by getting the best TPR and FPR and the lowest EER from other techniques in the literature. Also, it classified the flooding attack more accurately than the other techniques but demonstrated a high FPR. Moreover, it was effective at identifying patterns of non-attacking behaviour. However, it was not the most effective technique among those described in the literature, despite having the lowest FPR and EER.

In the future, it is planned to investigate the low correlation between the WSN-DS features and develop the optimal method for selecting the most pertinent features to enhance the detection performance of DoS in IoTs. In addition, the proposed method will be tried on several Imbalanced multiclass datasets to see how well it works. Finally, it is planned to find the best machine learning algorithm to make an intelligent IDS, test it with a network simulator, and measure its reliability and transparency to detect cyber-DoS attacks in an IoT environment.

REFERENCES

Abiodun, O. I., Abiodun, E. O., Alawida, M., Alkhalaf, R. S., and Arshad, H. (2021). A review on the security of the internet of things: challenges and solutions. *Wireless Personal Communications*, 119(3):2603–2637.

Adat, V. and Gupta, B. B. (2018). Security in internet of things: issues, challenges, taxonomy, and architecture. *Telecommunication Systems*, 67(3):423–441.

Ahmad, H., Kasasbeh, B., Aldabaybah, B., and Rawashdeh, E. (2022). Class balancing framework for credit card fraud detection based on clustering and similarity-based selection (sbs). *International Journal of Information Technology*, pages 1–9.

Almomani, I. and Al-Kasasbeh, B. (2015). Performance analysis of leach protocol under denial of service attacks. In *2015 6th International Conference on Information and Communication Systems (ICICS)*, pages 292–297.

Almomani, I., Al-Kasasbeh, B., and Al-Akhras, M. (2016). Wsn-ds: A dataset for intrusion detection systems in wireless sensor networks. *Journal of Sensors*, 2016.

Aydilek, I. B. and Arslan, A. (2013). A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector regression and a genetic algorithm. *Information Sciences*, 233:25–35.

Behera, T. M., Samal, U. C., and Mohapatra, S. K. (2018). Energy-efficient modified leach protocol for iot application. *IET Wireless Sensor Systems*, 8(5):223–228.

- Churcher, A., Ullah, R., Ahmad, J., ur Rehman, S., Masood, F., Gogate, M., Alqahtani, F., Nour, B., and Buchanan, W. J. (2021). An experimental analysis of attack classification using machine learning in iot networks. *Sensors*, 21(2).
- Islam, U., Muhammad, A., Mansoor, R., Hossain, M. S., Ahmad, I., Eldin, E. T., Khan, J. A., Rehman, A. U., and Shafiq, M. (2022). Detection of distributed denial of service (ddos) attacks in iot based monitoring system of banking sector using machine learning models. *Sustainability*, 14(14).
- Jan, S. U., Ahmed, S., Shakhov, V., and Koo, I. (2019). Toward a lightweight intrusion detection system for the internet of things. *IEEE Access*, 7:42450–42471.
- Jiang, S., Zhao, J., and Xu, X. (2020). Slgbm: An intrusion detection mechanism for wireless sensor networks in smart environments. *IEEE Access*, 8:169548–169558.
- Kumari, A. and Mehta, A. K. (2020). A hybrid intrusion detection system based on decision tree and support vector machine. In *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, pages 396–400.
- Leevy, J. L., Khoshgoftaar, T. M., and Peterson, J. M. (2021). Mitigating class imbalance for iot network intrusion detection: A survey. In *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 143–148.
- Masengo Wa Umba, S., Abu-Mahfouz, A. M., and Ramotsoela, D. (2022). Artificial intelligence-driven intrusion detection in software-defined wireless sensor networks: Towards secure iot-enabled healthcare systems. *International Journal of Environmental Research and Public Health*, 19(9).
- Mbarek, B., Ge, M., and Pitner, T. (2020). Enhanced network intrusion detection system protocol for internet of things. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing, SAC '20*, page 1156–1163, New York, NY, USA. Association for Computing Machinery.
- Meneghello, F., Calore, M., Zucchetto, D., Polese, M., and Zanella, A. (2019). Iot: Internet of threats? a survey of practical security vulnerabilities in real iot devices. *IEEE Internet of Things Journal*, 6(5):8182–8201.
- Patel, H., Rajput, D. S., Reddy, G. T., Iwendi, C., Bashir, A. K., and Jo, O. (2020). A review on classification of imbalanced data for wireless sensor networks. *International Journal of Distributed Sensor Networks*, 16(4):1550147720916404.
- Pereira, R. M., Costa, Y. M., and Silla Jr., C. N. (2020). Mtl: A multi-label approach for the torek link under-sampling algorithm. *Neurocomputing*, 383:95–105.
- Pokharel, P., Pokharel, R., and Sigdel, S. (2020). Intrusion detection system based on hybrid classifier and user profile enhancement techniques. In *2020 International Workshop on Big Data and Information Security (IW BIS)*, pages 137–144.
- Stoyanova, M., Nikoloudakis, Y., Panagiotakis, S., Pallis, E., and Markakis, E. K. (2020). A survey on the internet of things (iot) forensics: Challenges, approaches, and open issues. *IEEE Communications Surveys & Tutorials*, 22(2):1191–1221.
- Tabbaa, H., Ifzame, S., and Hafidi, I. (2022). An online ensemble learning model for detecting attacks in wireless sensor networks.
- Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., and Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access*, 7:41525–41550.