DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

Meerkat Clan-Based Feature Selection in Random Forest Algorithm for IoT Intrusion Detection

Adil Yousef Hussein[1], Ahmed T. Sadiq[2]
[1,2]*Computer Sciences Department, University of Technology, Baghdad, Iraq*
[1]*Adil.alomran@gmail.com*, [2]*Ahmed.T.Sadiq@uotechnology*

Abstract— *Hackers can conduct more destructive cyber-attacks thanks to the rapid spread of Internet of Things (IoT) devices, posing significant security risks for users. Through a malicious process, the attacker intended to exhaust the capital of the target IoT network. Researchers and company owners are concerned about the reliability of IoT networks, which is taken into account because it has a significant impact on the delivery of facilities provided by IoT systems and the security of user groups. The intrusion prevention system ensures that the network is protected by detecting malicious activity. In this paper, the focus is on predicting attacks and distinguishing between normal network use and network exploitation for intrusion and network attack and we will use Swarm Intelligence (SI) which is one of the types of artificial intelligence (AI) that we harness to choose features to determine the task of them and specifically we will use an algorithm Meerkat Clan (MCA) for this purpose, as this research suggested a modified IDS in machine learning (ML) based IoT environments to identify features and these features will be input into Random Forest algorithm. The IoTID20 dataset is used where nominal traits are removed, so the final dataset contains 79 traits. The data set contains three categories: the label that identifies whether it is a natural use or exploitation, the category that characterizes the type of exploitation, and the subcategory that describes that exploitation more accurately. The number of trees in a random forest (RF) classifier for binary, class, and subclass will be determined by the experiment. The trained classifier is then tested and the approach achieves 100% accuracy for binary target prediction, 96.5% for category and accuracy ranges of 83.7% for sub-category target prediction. The proposed system is evaluated and compared with previous systems and its performance is shown through the use of confusion matrix and others.*

Index Terms— *IoT, IDS, IoTID20 dataset, Random Forest, Meerkat Clan Algorithm (MCA).*

I. INTRODUCTION

Today, in the world of the Internet, trends seem to be starting to shift as a result of the advent of intelligent things capable of the generation and transmission of the data through Internet in a manner which is comparable to the humans. The IoT is a collection of cutting-edge technology and applications that have a potential for revolutionizing our way of life. Internet of Things has permeated almost every aspect of the life. IoT is utilized by the governments worldwide to collect data from various sectors and to offer better programs for health, governance, development and security [1].

As IoT gets more pervasive every day, threats against it rise. At present, there are billions of Internet-connected appliances. By 2020, the estimate will increase to 20 billion [2]. An IoT will be a goal for attackers to conduct malicious behaviors and, through exponential expansion, raise the attack surface of IoT networks. This led to many of organizations suffering the loss of facilities, the results became the cyber-attacks more damaging, especially after the emergence of a new concept of data, which is Big

Received 30/September/2021; Accepted 31/December/2021

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

data, and the resulting conclusions and decisions that depend on the accuracy of that data, to recognize malicious activity in the smart infrastructure, IoT devices requires a sophisticated instrument [3, 4].

Intrusion detection systems (IDSs) can be defined as tools which may be utilized in information systems to track known threats or anomalies. IDS may be network-based or networks that are host-based. They may be utilized to safeguard a computer against networks or terminal attacks. The networking may be IoT networking, or an integrated platform like sensor devices could be an end-user device. IDS can also be both hardware and software in the form of a single software suite [5].

There are also two detection methods available:

1- Detection based on signatures,

Signature-based detection techniques are sufficient in identifying recognized assaults by analyzing network traffic or data stored in computer memory for specified patterns.

2- Detection based on abnormalities

Anomaly-based detection is utilized to discover unknown threats through the monitoring of the whole systems objects', or traffic's activities and comparing them to preset behavior considered to be normal. Any deviation from regular operations is considered a probable assault [6]. Because of the complexity of attackers and the increase in 0-day assaults, the anomaly-based IDS is seen to be well adapted to the present environment [4].

To detect abnormalities, anomaly-based intrusion detection systems rely on AI and ML. The theory behind AI and ML is making a machine able to learn on its own and differentiate between system activity that is normal or abnormal [7].

The enormous traffic for data amount with high-speed networks represents a major challenge to the efficiency of the IDS in real-time [8]. As conventional IDSs normally concentrate on improving the performance of detection and lack adequate attention to timeliness, it’s hard for them to monitor traffic data in short period. For attackers, it leaves an opening. It is also imperative to gain access to IDS solutions for the real-time traffic data processing [9].

There has been suggested five classification methods utilizing dataset to classify network assaults [10], Those approaches include J-48 decision tree, Back Propagation, Bayesian Rules, SVM and NSL-KDD.

And 3 strategies of feature selection [11], the methods of the feature discovery complies the correlation-base feature selection (CFS), information gain (IG), and decision tables.

A feature selection method which was based upon Modified Artificial Immune System that was presented there [12]. The suggested approach takes advantage of benefits of Artificial Immune System to maximize speed and feature randomization.

In this paper, IoTID20 [4] dataset is used where the nominal attributes are removed and thus the final dataset has 79 features. The dataset has three classes which are the label, category and subcategory. The paper proposed a modified IDS in IoT environments based on Meerkat Clan Algorithm (MCA) [13] for feature selection and the chosen features are input for Random Forest algorithm [14].

Received 30/September/2021; Accepted 31/December/2021

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

II. RELATED WORK

A. IOTID20

Mahmood & Ullah 2020 [4], have utilized feature similarity, feature rating, and a variety of the machine-learning algorithms for the classification for the analysis and comparison of IoT-ID20 data-set. They have utilized ML methods and data normalization techniques to normalize and analyses the IoTID20 dataset. The capabilities of the identification of ML algorithm can be harmed by the related features. For the IoT-ID20, 12 related attributes have been removed from data-set. Shapira-Wilk algorithm has been utilized for rating features in IoTID-20 data-set, testing the regularity of feature-related distributions of the occurrences. Over 70% of features that have been graded by a score > 0.5, which indicates the fact that they are of a high level of rating. The binary, sub-class, and sub-category mark data-sets have all been assessed. Models of ML have been built with the use of the ensemble, Gaussian, Naive Bayesian, SVMs, Latent Dirichlet Allocation, Logic Regression, Random Forest, and Decision Tree classifiers. For the evaluation of effectiveness of various classifiers, they utilized a variety of K-fold cross-validation tests, including three, five, and ten folds cross-validation experiments. The maximum accuracy levels have been accomplished through the use of the Ensemble, RF, and DT classifiers, whereas minimal accuracy levels have been accomplished by the SVMs and LR.

Shami & Yang in 2020 [10], have proposed LightGBM model adaptation for the analytics of the IoT data that has good accuracy when utilizing little memory and time. The suggested model will respond automatically owing to already data streams of sophisticated IoT networks incorporations of their suggested novel algorithm of drift-handling, tool for hyperparameters Particle Swarm Optimization (PSO), an ensemble machine learning algorithm (i.e. LightGBM) and Optimized Adaptive and Sliding Windowing (OASW). The suggested approach is tested and discussed using experimentations on two available data-sets of the IoT anomaly detection, IoT-ID20 [7] and NSLKDD [9]. On the basis of the IoTID20 and NSL-KDD data-sets, the method for the effective outperforms numerous advanced drift adaptation approaches in detecting with 99.92% accuracy in IoT assaults and 98.31% accuracy in responding to concept drift.

Farah in 2020 [3], had utilized 2 freely accessible data-bases of simulation, the Bot-IoT [15] and IoT-ID20 [4], for the testing and training, which have been advanced for capturing the IoT networks for a variety of the attacks like the Scanning and DoS. The models of ML which have been applied in those data-sets have been tested in every one of the data-sets prior to being assessed through the data-sets. There have been large variations in analyses that have been obtained with the use of the 2 data-sets. The supervised models of ML have been advanced and tested for the binary classifications that had differentiated between the standard cases and the anomaly attacks, in addition to the multi-class classification that had classified the attack type on IoT network. For the purpose of ensuring the fact that a model operates sufficiently, the network packet flow identifiers, like source and destination IP addresses, time-stamp, and port numbers, had to be eliminated. Due to the fact that the attackers will utilize a variety of the IP addresses and times for the initiation of the attacks on a network, in the case where a model has been trained with the use of those characteristics, it could fail at generalizing well when utilized. In addition to that, 10 more functions have been removed due to the fact they only had one value and didn’t add much value to models of ML. Those models have been trained then with the use of a total of 67 of the features. The models worked adequately and had the ability to detect the irregularities in data. On the two data-sets, DT, kNN, and ensemble scores were both higher than 0.95.

Qaddoura etal. In 2021 [16], have proposed a three-stage solution which included the clustering with oversampling, reduction, and classification by using the Single Hidden Layer Feed-Forward Neural

Received 30/September/2021; Accepted 31/December/2021

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

Network (SLFN). The innovation of this study lies in data reduction and over-sampling approaches which had been utilized in order to generate balanced and usable data of training, besides hybrid considerations of the controlled and unsupervised approaches for the detection of the activities of intrusion. Tests have been split to 4 stages and tested based on the accuracy, precision, G-mean, and recall: which measure impacts of the clustering on the data reduction, the performance of the framework's against the basic classifiers, the effects of the over-sampling, and a comparison to the fundamental classifiers. SLFN classification with the choice of the SVMs and Synthetic Minority Oversampling Technique (SVM-SMOTE) with a 0.90 ratio and a k value of 3 for the k-means++ clustering approach have given better performance, with a 98.40 % score, on chosen IoT-ID20 [4] data-set.

Qaddoura *etal*. In 2021 [17], have proposed deep multilayer classification approach, which included: a strategy oversampling with the use of SMOTE for addressing the issue of the mismatch data-sets; and a deep multi-layered classification approach. For the purpose of improving the results of the classification, 2 approaches have been suggested. SLFN technique's 1st classification level predicts the interference and routine processes. The DNN predicts the intrusion activity type in the 2nd classification stage. Tests have been conducted on the IoTI-D20 [4], which showed that the solution that has been proposed had yielded better efficiency.

B. SWARM FOR FEATURE SELECTION

Arslan and Ozturk in 2019 [18], introduce a technique whose goal is to discover relevant characteristics and minimise noisy attributes, increase classification accuracy, and reduce dataset size while selecting a subset from a larger dataset. On four distinct data sets, artificial bee colony programming (ABCP) has been presented and applied to the feature selection for classification issues. The best models have been created with the use of a sensitivity fitness function that is established based on the total number of classes in a dataset, and they have been compared to models that were created via the use of the genetic programming (GP).

Peng et al. In 2018 [19], published a method that selects a sub-set data-set from a large dimensional data-set for the reduction of the size of the data-set and selects relevant features while discarding irrelevant features. The fitness function for feature selection was created to improve the ant colony's route transfer probability technique by addressing flaws in current methods. Meanwhile, the 2-stage pheromone update method has been utilised for the purpose of adding pheromones to more routes, preventing the algorithm from prematurely slipping into local optimum. Finally, using the KDD CUP99 dataset, a simulation experiment demonstrate that the FACO method may enhance the efficiency of the classification and the classifiers' accuracy.

Lu et al. in 2017 [20], published a technique whose objective is to minimise the dimensionality of a big dataset by extracting useful features and removing irrelevant features. This work improved the accuracy of feature classifiers and was utilised in a variety of applications. They presented a hybrid feature selection method which combined the mutual information maximization (MIM) and adaptive genetic algorithms (AGA). The suggested MIMAGA-Selection technique greatly lowers the gene expression data dimension and eliminates redundancies for the classifications, according to experimental results. When compared to traditional feature selection techniques, the reduced gene expression data-set gives the best accuracy of the classification.

Yamany et al. in 2016 [21], published a proposal with the purpose of reducing the amount of data-sets, selecting the sub-set of important attributes, and selecting the optimum solution quickly. A correlation-based filter model for the reduction of the attributes was suggested as part of a system for the reduction

Received 30/September/2021; Accepted 31/December/2021

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

of the attributes. The cuckoo search (CS) optimization method has been used for searching attribute space for the attributes with the least amount of correlation. The first solutions, which are assured to have little correlation, are then candidates for additional improvements in the classification accuracy fitness function. The suggested system's performance was evaluated by implementing it using a variety of data sets. Its performance was also compared to that of other commonly used attribute reduction methods. The suggested multi-objective CS system outperforms the traditional single-objective CS optimizer, in addition to the PSO and GA optimization methods, according to the findings.

III. PROPOSED SYSTEM

Random forest is a machine learning technique that generates a collection of categorization methods based upon many decision trees.

In this paper, we will use SI which is a type of AI for feature selection in order to determine which features are important to build the forest and specifically we will use the Meerkat Clan (MCA) algorithm.

In the feature selection, artificial intelligence has the aim of obtaining a sub-set of the features for the purpose of describing the problem, which is used in a wide range of the applications for understanding data, storing data, reducing data size, and enhancing accuracy. Feature selection has been utilized in a wide range of the applications, however, the swarm intelligence effectively found sub-set of the data in comparison with the conventional approaches.

During the previous era, optimization for this methods grown in popularity. It operates in a decentralized manner that resembles the behavior of groups of swarms, groups of animals, or swarms of fishes. These strategies are stronger and more adaptable than traditional techniques. SI is a fruitful enterprise model for algorithms that meet the needs of complex problems through diversity and random selection of features.

SI which is a type of AI subject related to the development of smart multi-agent systems based on the shared activities of social insects like termites, ants, wasps, bees, as well as other animal groups such as fish and birds. For a long time, researchers were fascinated by colonies of social insects, and the mechanisms that regulate their behavior have remained a mystery. Despite the fact that these colonies' single members are uncivilized individuals, they are nonetheless capable of completing tough tasks in teams. Colony behavior is organized as a result of basic actions or relationships among colonies individuals.

Meerkat Clan Algorithm (MCA) can be defined as a set of SI algorithms inspired by meerkat behaviour in South Africa's Kalahari Desert. Meerkats live in groups of 20 to 50 male and female partners, with each group containing 20 to 50 male and female partners. Meerkats are sociable creatures living in groups of 5 to 30 people. They share both toilet and parental care responsibilities since they are social beings. There is a leading alpha male and female in every one of the mobs. Each mob has a unique territory, which they must periodically shift to if food cannot be obtained or if a tougher mob demands it. In the case where the latter occurs, the weaker group will either try to grow in another way or stay until they get stronger and recoup their losses.

A method for intrusion detection is presented by merging random forest technique with Meerkat Clan Algorithm (MCA). The method takes IOTID20 dataset, the number of required features, the classification output target, the worst care ratio (Cr) and the worst foraging ratio (Fr) as inputs and the decision tree is the output. The method begins with initializing the Meerkat clan population N for all

Received 30/September/2021; Accepted 31/December/2021

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

features and then splitting it into care and forage. It then chooses a random forage as the root of the tree and then loops on the other features

The meerkat clan algorithm used in the proposed technique to find the best sub-set of attributes which may replace the total set of attributes for intrusion detection. The MCA process is utilized for the improvement the performance of random forest by selecting a subset of trees that represent good solutions and relying on them instead of relying on all trees and characteristics. There several random forests are built and find the accuracy of classifying intrusions.

The following algorithm provides our modified MCA algorithm.

Algorithm 1: Meerkat Clan-Based Features Selection for Random Forest Algorithm

Inputs: **Training Data-set, select target, Number of Features, Cr: worst care ratio,**
Fr : worst foraging ratio,

Output: **Tree of random forest**

```
Begin:
    select Number of Trees depending on target
    Initial Meerkat clan population (N) from all features
    split the population into 2 group (foraging and care)
    For i From 1 to Number of trees
        Begin
           Choose a random foraging solution as $root_i$ for trees
           For j from 1 to Number of features in FS
           Begin:
              Split dataset depending on chosen $root$
              Choose Random feature from foraging solutions as $root_j$
           End for
           Update foraging solutions using neighbor generation technique
           Until Best features met
           Swap the worst (Fr) solutions in foraging group with the optimal ones in care group
           Drop the worst (Cr) solutions in care group and generate randomly ones
        End for
    Each tree voting of which class elective

End
```

IV. RESULTS

A. DATA DESCRIPTION (IOTID20)

The IoTID20 dataset [4] includes intrusion as well as the regular activities that have been recorded by tablets, notebooks, and smart-phones in an IoT smart home network that includes a WiFi router, SKT NGU PC, and EZVIZ camera the dataset has 83 characteristics and 625783 occurrences. The data-set contains intrusion detection mark, the label of the category, and subcategory label. Table I displays the binary, category, and sub-category labels from IoT-ID20 dataset.

Received 30/September/2021; Accepted 31/December/2021

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

TABLE I. IOT-ID20 DATA-SET

Binary	Categories	Sub-Category
Anomaly	Anomaly-Scan	Hot Port
Normally	Anomaly-Mirai	Port OS
	Anomaly-MITM ARP	ACK Flooding
	Anomaly-DOS	Host BruteForceg
	Normal	HTTP Flooding
		UDP Flooding
		MITM ARP Spoofing
		Synflooding
		Normal

Data-set record distribution between the standard and the intrusion processes. Through the addition of more damaging risks, the IoT systems increased surface of the attack. The Man-in-the-Middle (MITM), Distributed DoS (DDoS), DoS, and active scan have been considered as the most malicious actions that are injected and tracked for the production of data-set.

An attack of DoS which floods the synchronized (SYN) packets into the TCP-based connections is being studied (Transmission Control Protocol-based connections). SYN packets have been commonly used to establish TCP connections between communication parties by the reserving resources on the two sides, most notably ports and buffers. It might be deployed in order to target availability of the server and/or victim computers. In addition to that, DDoS attacks in a form of the IoT Mirai have been introduced using flooding Acknowledgement, User Datagram Protocol (UDP) and HyperText Transfer Protocol (HTTP) packets. A brute force attack had been also utilised to decode data and disclose the secrecy. MITM was deployed as well to corrupt the ARP data-base, mapping the attacker's MAC address to the IP address of the router. Which is why, the intruder will mimic a network router, interfering with network entity interactions. This attack's primary objective is sniffing or modifying the data that is sent [16].

IoT-ID20 features are ranked and reduced and as Table II shows the distribution of the dataset after it has been separated into 75% training and 25% testing.

TABLE II. IOTID20 DATASET SPLIT INTO TRAINING AND TESTING DATASET

		Binary		Category		Sub_Category					
Dataset IoTID20	**625.783**	**Anomaly**	**585.710**	**Anomaly-Scan**	**75,265**	**Hot Port**	**22,192**	**For Training 75%**	**16,644**	**For Test 25%**	**5,548**
						Port OS	**53,073**	**For Training 75%**	**39,805**	**For Test 25%**	**13,268**
				Anomaly-Mirai	**415,676**	**ACK Flooding**	**55,124**	**For Training 75%**	**41,343**	**For Test 25%**	**13,781**
						Host Brute Forceg	**121,181**	**For Training 75%**	**90,886**	**For Test 25%**	**30,295**
						HTTP Flooding	**55,818**	**For Training 75%**	**41,864**	**For Test 25%**	**13,955**

Received 30/September/2021; Accepted 31/December/2021

ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

						UDP Flooding	183,553	For Training 75%	137,665	For Test 25%	45,888
				Anomaly-MITM ARP	35,377	MITM ARP Spoofing	35,377	For Training 75%	26,533	For Test 25%	8,844
				Anomaly-DOS	59,392	Synflooding	59,391	For Training 75%	44,544	For Test 25%	14,848
		Normally	40.073	Normal	40,073	Normal	40,073	For Training 75%	30,055	For Test 25%	10,018
								Total Sample for Training	469,337	Total Sample for Testing	156,446

B. EXPERIMENTAL RESULT

The accuracy of predicting the binary target approaches 100% where the number of used trees is the random forest algorithm is -- and the number of used features is --, cr is --, fr is --. Table III lists comparison between suggested system for the prediction of binary (anomaly/ normal) target and other algorithms of ML according to time and accuracy. Accuracy has been characterized by total amount of the correct predictions that have been divided by total amount of the predictions. The suggested system's accuracy and neural networks are the highest and the time of the neural network is very high, while the test time for Decision Tree and Naïve Bayes is the lowest while their accuracy is not high compared to others. The general evaluation illustrates superiority of the suggested algorithm.

TABLE III. THE COMPARISONS BETWEEN SUGGESTED SYSTEM AND OTHER APPROACHES THAT PREDICT BINARY TARGET BASED ON THE TIME AND ACCURACY

Algorithm	Testing Accuracy	Training Time [s]
Modification System	***99.9%***	***340.7***
Neural Network	99.8%	769.52
Decision Tree	98.6%	7.6
Logistic Regression	96.6%	14.76
Naïve Bayes	94.7%	5.7
SVM	65.5%	123.47

Table IV provides comparison between suggested system for the prediction of the class (i.e. normal/ Scan, MITM ARP, Mirai, DOS) target as well as other approaches of ML based on the time and accuracy. The suggested system and Decision Trees' accuracy level are maximal and their time is not high, while the test time for Naïve Bayes is the lowest while the accuracy is very low. The overall evaluation shows the superiority of the Decision Tree and our proposed algorithm.

Received 30/September/2021; Accepted 31/December/2021

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

TABLE IV. THE COMPARISON BETWEEN SUGGESTED SYSTEM AND OTHER APPROACHES THAT PREDICT CATEGORY TARGET ACCORDING TO TIME AND ACCURACY

Algorithm	Testing Accuracy	Training Time [s]
Modification System	***96.5%***	***540.46***
Neural Network	95.7%	704.33
Decision Tree	98.7%	12.64
Logistic Regression	74.2%	30.64
Naïve Bayes	75.1%	2.43
SVM	42.9%	281.8

Table V provides comparison between suggested system for the prediction of sub-categories (Normal/ Host Port/ Port OS/ MITM ARP Spoofing/ ACK Flooding/ Synflooding/ HTTP Flooding/ Host BruteForce/ UDP Flooding) as well as other ML approaches, concerning the time and accuracy. The suggested system and Decision Trees' accuracy values have been maximal and their time is not high, while the test time for Naïve Bayes is the lowest while the accuracy is very low. The overall evaluation shows the superiority of the suggested algorithm.

TABLE V. THE COMPARISONS BETWEEN SUGGESTED SYSTEM AND THE REST OF THE APPROACHES THAT PREDICT SUBCATEGORY WITH REGARDS TO TIME AND ACCURACY

Algorithm	Testing Accuracy	Training Time [s]
Modification System	**83.7%**	**490.96**
Neural Network	73.6%	648.85
Decision Tree	**79.4%**	63.67
Logistic Regression	50.7%	38.73
Naïve Bayes	52.5%	**3.66**
SVM	15.8%	45143.91

V. CONCLUSIONS

This study proposed IDS in IoT environments on random forest hybrid feature selection techniques that can detect intrusions with high speed and accuracy. The IoTID20 dataset uses three target classes, the binary class normal or abnormal, and the binary class and subclass categories. In this paper, the purified IoTID20 dataset is used after nominal traits are removed, thus the final dataset contains 79 traits. The research suggested a modified IDS in IoT environments based on the Meerkat Clan (MCA) algorithm for variable and random feature selection to be an input to the Random Forest algorithm. The highest rated attributes are selected in the dataset and other attributes are reduced, to reduce execution time and improve accuracy, the number of trees in the binary class random forest classifier is reduced, and the number of trees in the binary class, subclass is reduced. Then the trained classifier is tested and the approach achieves 100% accuracy for binary target prediction, 96.5% for category and accuracy ranges of 83.7% for sub-category target prediction. The proposed system is evaluated and compared with previous systems.

Received 30/September/2021; Accepted 31/December/2021

DOI: https://doi.org/10.33103/uot.ijccce.22.3.2

REFERENCES

[1] Alghuried, A. (2017). A model for anomalies detection in internet of things (IoT) using inverse weight clustering and decision tree.

[2] Middleton, P., Kjeldsen, P., & Tully, J. (2013). Forecast: The internet of things, worldwide, 2013. *Gartner Research*.

[3] Press, G. (2018). Internet Of Things By The Numbers: What New Surveys Found (forbes.com). Last Accessed Dec, 1, 2020.

[4] Ullah, I., & Mahmoud, Q. H. (2020, May). A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks. In *Canadian Conference on Artificial Intelligence* (pp. 508-520). Springer, Cham.

[5] Gordon, A. (Ed.). (2015). *Official (ISC) 2 guide to the CISSP CBK*. CRC Press.

[6] Prabha, K., & Sree, S. S. (2016). A survey on ips methods and techniques. *International Journal of Computer Science Issues (IJCSI)*, *13*(2), 38.

[7] Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*, *18*(2), 1153-1176.

[8] Jamshed, M. A., Lee, J., Moon, S., Yun, I., Kim, D., Lee, S., ... & Park, K. (2012, October). Kargus: a highly-scalable software-based intrusion detection system. In *Proceedings of the 2012 ACM conference on Computer and communications security* (pp. 317-328).

[9] Jin, D., Lu, Y., Qin, J., Cheng, Z., & Mao, Z. (2020). SwiftIDS: Real-time intrusion detection system based on LightGBM and parallel intrusion detection mechanism. *Computers & Security*, *97*, 101984.

[10] Yang, L., & Shami, A. (2021). A Lightweight Concept Drift Detection and Adaptation Framework for IoT Data Streams. *IEEE Internet of Things Magazine*.

[11] Assi, J. H., & Sadiq, A. T. (2017). NSL-KDD dataset classification using five classification methods and three feature selection strategies. *Journal of Advanced Computer Science and Technology Research*, *7*(1), 15-28.

[12] Assi, J. H., & Sadiq, A. T. (2018). Modified artificial immune system as Feature Selection. *Iraqi Journal of Science*, 733-738.

[13] Al-Obaidi, A. T. S., Abdullah, H. S., & Ahmed, Z. O. (2018). Meerkat clan algorithm: A new swarm intelligence algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, *10*(1), 354-360.

[14] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.

[15] Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *Future Generation Computer Systems*, *100*, 779-796.

[16] Qaddoura, R., Al-Zoubi, A. M., Almomani, I., & Faris, H. (2021). A Multi-Stage Classification Approach for IoT Intrusion Detection Based on Clustering with Oversampling. *Applied Sciences*, *11*(7), 3022.

[17] Qaddoura, Raneem, M. Al-Zoubi, Hossam Faris, and Iman Almomani. "A Multi-Layer Classification Approach for Intrusion Detection in IoT Networks Based on Deep Learning." *Sensors* 21, no. 9 (2021): 2987.

[18] Arslan, S., & Ozturk, C. (2019). Feature Selection for Classification with Artificial Bee Colony Programming. In *Swarm Intelligence-Recent Advances, New Perspectives and Applications*. IntechOpen.

[19] Peng, H., Ying, C., Tan, S., Hu, B., & Sun, Z. (2018). An improved feature selection algorithm based on ant colony optimization. *IEEE Access*, *6*, 69203-69209.

[20] Lu, H., Chen, J., Yan, K., Jin, Q., Xue, Y., & Gao, Z. (2017). A hybrid feature selection algorithm for gene expression data classification. *Neurocomputing*, *256*, 56-62.

[21] Yamany, W., El-Bendary, N., Hassanien, A. E., & Emary, E. (2016). Multi-objective cuckoo search optimization for dimensionality reduction. *Procedia Computer Science*, *96*, 207-21

Received 30/September/2021; Accepted 31/December/2021

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352