Original Article

A Malicious Botnet Traffic Detection Using Machine Learning

M. Sakthivel[1*], S. Sivanantham[2], V. Akshaya[3], D. Sivakumar[4], H. Karthikeyan[5]

[1*]Professor, Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India. E-mail: sakthisalem@gmail.com

[2]Assistant Professor, Department of CSSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India.

Email: sivanantham.s@vidyanikethan.edu

[3]Assistant Professor, Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India. E-mail: akshaya.v@vidyanikethan.edu

[4]Assistant Professor, Department of CSSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India.

E-mail: sivakumar.d@vidyanikethan.edu

[5]Assistant Professor, School of Engineering and Technology, CHRIST (Deemed To Be University), Bengaluru, Karnataka, India.

E-mail: karthikeyana.h@gmail.com

Abstract

Detection of incorrect and malign data transfers in the Internet of Things (IoT) network is important for IoT safety to observe an eye on and prevent unwelcomed traffic flow to the network of IoT. For it, Machine Learning (ML) strategic methods are produced by several researchers to prevent malign data flows through the network of IoT. Nonetheless, because of the wrong choice of feature, a few malign Machine Learning models differentiate especially the movement of malign traffic. Still, what matters is the problem that needs to be deliberated in-depth to select the best features for better malign traffic acquisition in the network of IoT. Dealing with the challenge, a new process was proposed. 1st, the metric method of selecting a novel feature called the proposed CorrAUC, and hinged on CorrAUC, a new highlight for choosing the Corrauc algorithm name is also being developed, designed hinged on the system folding filter features precisely and select the active features of the choose ML method using AUC metric. After that, we apply a combined application Order of Preference by Similarity to Ideal Solution Using Shannon Entropy (TOPSIS) built on a bijective set which is soft to verify selected features for identification of malign 1traffic in IoT network. We test our method using data set of Bot-IoT and 4 dissimilar ML classifiers. Practical outcomeanalysis showed that our proposed approach works as well and can achieve greater than 96% results on average.

Keywords: Botnet, Machine learning, TOPSIS, Bot-IoT.

DOI: 10.47750/pnr.2022.13.04.131

INTRODUCTION

Today, Internet of Things (IoT) technologies are evolving at a great speed and every inch of time devices are establishing a connection to this tech.With the usageof this tech, our life becomes easier and is well designed. For instance, at starting, IoTtech was permitted and used in small workplaces and apartments, but today, IoTis integrating into the industry to be more reliable and more time-saving [1]. However, IoT tech inculcating an integral part of our day-to-day lives. By 2021, this tech will mature and tens of millions of IoT devices will connect, this will be a huge switch in the IoT world. Although IoT tech is developing with time, on the other sideattacks of cyber are also becoming much stronger and wilder. Therefore, many assess people in the field of IoT tech proposed many different constructs to protect the Internet, and the widespread use of Internet security has proposeda system to save their information from network strikes as well as uncertified access [2]. Recently, the security of IoT has become hotter than an article, which also received a lot of attention from IoT network security. To avoid the spike in the Internet of Things IoT attacks, knowledgeable people and scientists are trying out their best and most popular cybersecurity programs.Similarly, many network security programs on the IoT network are introduced and used to protect the importantinformation that is also protected from uncertifiedconnection to the network of IoT. In 2017, Internet of Things strikes such as denial of service (DDoS) became common and reached 72%, which generated a high look in the network IoT.

From the view of Kaspersky Lab 2019, malware attacks on IoT network websites increased from in 2017juxtaposed to IoT network malware attacks in 2013. This attack, many of youare very dangerous attacks like botnet attacks and more [3]. With Entry Access, the pioneer of the Intrusion Detection System (IDS) is Anderson in 1980. Denning introduced a new Accessprototype. Real-time interference identificationinput in 1987. Recommended to their ExpertsIntrusion Detection systems can detect hacks,hacks,Trojans,other electronictrespassing that led to computer systems being vulnerable, and more. Anyway, their model is hinged on the assumption that any security vulnerabilities can be discovered using descriptive and monitoring logs. That can detect an attack or mysterious

activity in thenetwork using user action. In today's IoT, the common, as well as dangerous strikes, are the man in the middle (MITM) through DDoS (Distributed Denial of Service). Anyway, many researchers in the community 4,400 researchers worked hard to find and implement an effective remediation program. These 4300 malign threats are common in the network environment of IoT.

For eloquent recognition of workingoutputs, type in dataset plays a crucial role in using Machine Learningclassifiers. Hence, for faultless and efficient spotting of irregularities and intrusions in the network of IoT in using Machine Learning, it is vital to pick an efficient type in feature set and take out unrequited functionality [4]. So, theattribute selection methods provide enoughcredentials and can discard the unrequired features from the data set of the givenfeatures. In their work, they aim to malign attacks in networks of IoT. The developed data set covers discrete types of malignstrikes, especiallybotnet strikes.

Much deeply, the datasetwas formed in a virtual test environment by a defined feature set, covering both normal and cyber infectdata flows. To analyze the test, a statistical analysis was doneto find out the feature which has the correct information for the identification of cyberattacks on the network of IoT. Anyway, the10 best features are selected out of thederived feature sets [5]. Next, they use the famous ML classifiersto an analysis of the performance of the selected feature set. Deeply search which featureprovides the most elegantouts of the 4 differentmetric kinds of performance analysis used.

In our prior study, the feature selection problems were investigatedand the electric shock characterization by suggesting different ways to define instant message (Instant Messaging) traffic. And a traffic-gathering attack. However, from our study, we conclude that there are more choices than Trait sets do not perform well in correctly recognizing using Machine Learningmethods and show more options than fifty feature sets can reduce the accuracy of the ML splitter becausesinks can addmathematical calculation complexity [6]. Anyway, in IoT to identify net grid attack traffic, no active MLalgorithm is currently provided. Thus, it is pivotal to cram Function selection problem and anomalous characteristics 4334 of traffic on IoT networks and come up with a new strategy 4334 to overcome this issue.

In this composition, a newefficient feature declaration method is derived for the efficient feature selection obstacle forattacks of cyber in network traffic of IoT using the dataset of BotIoT and to modify the working of Machine Learning methods [7]. Anyway, the keyaims of this treatise are:

To solve the problem of effective feature picking in identifying strikes in IoT networks. First, a new feature picking index process named CorrAUC was introduced to solve the problem of efficient feature selectionto identify cyber-attacks in the IoT network. This is the first time it is proposed to combine the correlative attribute assessment measure and the Specific Outcome AUC ofmachine learning to effectively select features in Bot- IoT attack detection.

Then, a new feature picking algorithm named Corrauc is developed and designed hinged on CorrAUC for the feature selection problem to identify malign BotIoT traffic in the IoT network, derived from wrapper method to precisely defined features and selectfeatures. The set contains ample information for the MLclassifier that chose to use AUC metrics to identify Bot strikes in IoT environments. Anyway, the produced algorithm in build2 processes for optimal feature pick. TheCAE (Correlation Attribute Evaluation) parameter and one metric which is specific to ML AUC (Area Under ROC Curve) parameter [8].

Next, we apply Shannon's built-in entropy andTechnique for Order of Preference by Similarity to Ideal Solution (TOPSIS) on a bijective set which is a soft, fusion process to checkpicked features to identify malign data flow in IoT grids [9]. It is derived from the identification of appropriateattributes that are defined characteristics to detect malignattacks more effectively in IoT networks. In addition, we contrast the results of Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Shannon Entropy hinge on the software suite 1-1 with the results obtained from the proposed method.

We then conclude and give the optimally picked feature set selected by our recommendation procedure, which transports adequatedata to detect malign BotIoT traffic in the IoT grid. Testing shows that 5optimal feature sets carry sufficientdata and have discriminatory power to detect malign attacks in networks of IoT by the usage of ML.

Related Work

In previous years, trust issues and security have become a hot topic, with many researchers attempting to tackle the problem and proposing a variety of viable models, including the wireless sensor network (WSN), the future Internet, IoVs, and IoTs[10]. However, we discussed studies on the selection of efficient features for harmful Bot-IoT in IoT networks in this part. An efficient selection of feature technique hinged on mutual information analysis technique is proposed in our fresh study, work for the flawless feature selection difficulty in Instant Messaging appsdata transfers categorization [11].

However, employing the specified feature set for Instant Messaging application traffic detection, the suggested method obtains improved performance results, according to the experimental data analysis more specifically, the research we presented is confined to feature picking for a variety of views, and reducing the computational complexity of the used Machine learning algorithms [12]. The proposed method could be used on an unbalanced data set that is staggeringly high. The results revealed that the newly presented way was capable of achieving higher performance outcomes for the classification of Instant Messaging software traffic [13].

The method of efficient picking of features, which was highly useful, can be used to increase Machine Learning performance. As we all know, feature selection is a strategy for selecting the best set of features from a large number of options and discarding characteristics that don't give adequate recognition and data (information) for the spotting are repeated in nature. In 2018, S Egealooked into mostly cited research studies which were linked to feature pickingmethod, he burned the midnight oil on the correlation coefficient method, by using that knowledge he proposed a new feature pickingmethod named "Fast Based Correlation Features algorithm". Which was used for developing the performance of the network of IoT[14], especially in the environment of industrial work. We can cut several even parts of feature space with even size which was their study's main contribution. By the usage of the producedmethod, they had produced betteroutputs of correlation Machine Learning of allworkingpoints in the network of IoT. They showedeffectivepractical results, that with accuracy and execution time as variables, the proposed approach can achieve efficient performance results, Accurate Identification is very important. Similarly, Meidan Yair et al, in 2018, To solve the attacks which were identified by the Internet of Things devices mugged the spotting of strikes in the network of IoT and introduced a new methodology and they used autoencoder for the anomalies identification in traffic of IoT. The strikes of botnet Bashlite and Miraigleaned from the IoT were the data clusters they used in their study to evaluate their proposed approach [15].The used datasets, however, are also present on many corrupted IoT devices. They demonstrated in their investigation that the proposed approach can identify strikes in network devices of IoT with huge working outputs. Shen Su investigated the highly cited attributepicking strategy and presented an attribute picking method for IoT devices, performance optimization, and anomaly detection. For their research, they first grouped IoT sensors to identify deployed sensors. They then take over the correlation difference of data for the picking of sensors for anomaly detection [16]. Added specifically, they used the curve alignment technique for sensor clustering, and the calculation window size for data is explored. Then, for feature selection, the Multi-Cluster Attribute Selection (MCFS) approach is used. They demonstrated their producedmethod is efficient and effective for working add-ons in IoT and peculiarity identification in networks of IoT, in their experimental research [17]. In-depth, a variety of IoT security techniques can be used in the IoT security environment for accurate cybersecurity purposes, such as cyber-attack detection in an effective management scheme evidence framework, and so on.

PROPOSED WORK

This header describes the producedmethod. The recommended steps for making effective decisions in an IoT network include the 4 steps. Initially, a new attributepicking metric way called CorrAUC was introduced and put in to select attributes with sufficient information, and a new attribute picking algorithm called Corrauc was built on CorrAUC. Use AUC metrics and Bot-IoT datasets to select features that work well and are effective for the ML algorithm of your choice. The producedmethodhaves Correlation Attribute Evaluation (CAE) and Area Under Roc Curve (AUC) to avoid the difficulty of effective attribute picking for Bot traffic detection in IoT using specific ML algorithms. In combination with the following metrics, we applied an integrated Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and Shannon entropy gleaned from the bijection software set to check the attributes selected for data transfer identification of Bot strikes in IoT networks. Specifically, the bijection set which is a soft and mathematical approach used for picking in different sectors. This method has very fruitful outputs regarding productive attribute picking for bots.

A. Attribute Selection Metrics

This part describes the attribute selection variables that are performed. The correlation-based metric is displayed first, followed by the AUC metric. However, see the next subsection for more information.

1) Correlation-based metrics

To throw the difficulty of productiveattribute selection, Pearson Moment Correlation Technology was introduced to find malign Bot IoT strikes in networks of IoT. This technique is used to further investigate and identify the relationship between the functionality of the independent class and the functionality of the class which is targeted. F. Galton introduces the idea of Pearson Moment Correlation which is basic in the 1880s. Likewise, In 1896, Pearson's product-moment correlation was made by K Pearson. Thismethod is used to identify relationships between distinct attributes. Although, the updated method is hinged on statistical analysis operations. The correlation coefficient formula:

For two distinct M and N

$$C_{X,Y} = \frac{Covariance(A, B)}{\sigma_x \sigma_y} \quad (1)$$

In Equation 1, the correlation coefficients are CA and B, and (A, B) indicates covariance. Likewise, the standard deviation of the A and B attributes σAσB is ab. Detailed, Equation 2 can be used for two sets of features to calculate the correlation coefficient.

$$C = \frac{\sum_{i=1}^{n}(a_i - \overline{a})(b_i - \overline{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \overline{a})^2}\sqrt{\sum_{i=1}^{n}(b_i - \overline{b})^2}} \quad (2)$$

For example, two sets of features A and B with their individual properties can be represented as a1, a2, a3,... an, and B be represented as b1, b2,... bn. Likewise, the

instancescount is indicated by n. In which values of the data be ai and bi. Similarly, a_bar and b_bar equals mean values of Equation 2, but when the value of the C coefficient + 1 and -1 is reached. In other words, if the coefficient value is +1 then the linkamong the attributes is strong, and null means that there is no link or dependency between the attributes [18]. Negative coefficient values, on the other hand, mean that the relationships amongattributes are very fragile. Pearson's correlation method is very fruitful for ranking and accurate attribute selection. Hence, correlation attribute evaluation techniques have been adopted and applied to suppress the difficulty of productive and sturdy feature selection, malign botstrikes in networks of IoT, and the effectiveness of features in multiple specific feature groups. Detects sex. Gender is ranked. The basetheory of usage of these ranking characteristics is to determine the importance of a feature set in a dataset hinged on the correlation between attributes. However, if you use machine learning to spotmalign Bot strikeson your network of IoT, the feature is useful if it has a strong relationship and no correlation with the feature. Similarly, this method can be used to calculate and analyze the effectiveness of a characteristic for accurate detection as follows:

The expression Corr shows the correlation among attributes, and kavg (corr FC) shows that we need to find the average of the correlations between features and their classes. Similarly, Avg (corrff) shows the average correlation between features, and k is the number of features. However, when the above formula is applied to the discriminatory correlation amongcharacters, the main reasons are: A strong correlation among feature sets indicates a weak correlation between feature sets and feature classes. Similarly, a fierce correlation between a feature set and a dependent class indicates a strong correlation between the feature set and the class, and a large number of attributes indicates a strong correlation between the feature and the dependent class.

2) Area Under Curve (AUC) - based Metrics

By the usage of the Corr variable, it is important to spot the toughest characteristics that contain precise information for detecting Bot strikes on IoT networks. In this scenario, a method called a wrapper is implemented hinged on the ROC curve (AUC) below the area as the metric. However, accuracy metrics are ideal for classifying network traffic through machine learning. However, here we are interested in finding key attributes for spotting bot strikes in the network environment of IoT. Therefore, the AUC variable is acrucial metric for spottingmalignstrikes in networks of IoT and is a handyvariable for classifying attributes into multiple functional groups. However, there are two clear facts about the use of the AUC variable in this experimentalwork. If the AUC variable is very high, the output of the model will be potent, and vice versa for detecting bot IoT attacks on IoT networks. In particular, AUC metrics are also very convincing for workingscrutiny and ranking attributes. Therefore, in this experimentalwork, AUC metrics are applied to rank efficaciousattributes, contain enough information to spot malign Bot strikes on networks of IoT, and attributes with very high metric scores.

3) Proposed algorithm

This section gives a detailed explanation of the Corrauc suggested algorithm. Two steps make up the projected Corrauc. The method filters the feature set in this stage by using a correlation approach to determine the correlation between the feature and the class. Next, the approach employs a particular machine-learning algorithm to filter out characteristics with high AUC metric values. The developed technique also selects helpful attributes that provide enough data for Bot detection in IoT networks. Anyway, here is how the phased phase's information is described:

As explained in the line above, the introduced Corrauc method is a compoundattributepicking method primarily hinged totally on correlation method as well as region beneath the roc graph variable, used to pick function which has sufficient statistics for detection of Bot-IoT assaults in IoT community. Anyway, the produced set of rules 1st is going in to compute the correlation among capabilities as well as pick capabilities which might be an excessive correlation relationship. In extra info, the better the brink mark, the better the introduced version fast, however, it`s now no longer powerful for the gadget to gain knowledge of a set of rules, due to the fact excessive limit values lower the identity and overall working of Machine learning classifiers. After computing the correlation and percolating with threshold values, the proposed set of rules clear out every function via way of means of the usage of the AUC metric of the precise ML set of rules. However, the set of rules filters every function one after the other via way of means of the usage of AUC metric and pick the one's capabilities which offer excessive AUC metric outputs for the spotting of Bot assaults in a community of IoT in addition to if the AUC values of a function are low then the set of rules will do away with from listing and set of rules would visit subsequent leap forward to Swapper.

B. Shannon Entropy Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)

For getting efficient features, Shannon Entropy (TOPOSIS) is applied hinged on soft-set technique to get the Bot attacks in IoT network environment. For our knowing, firstly motives and then basic definitions for efficient selection of feature and its operations are discussed. Today, making the right decision has become one of the most difficult problems in a research environment and many researchers are trying to overcome the problem of making the right decision and come up with models of making the right decision. Efficient decision-making using a soft-set for choice and making the right decision in features selection from multi-scale properties was proposed by Molodsov then followed by Gong giving the target software suite. Similarly, soft

aggregates of Type II are produced to skip the difficulty in classifying. From above, research has shown that the soft-set technique is a helpful process for efficiently getting features from certain objectives. Anyway, to get rid of the hurdle, selecting an efficient attribute picking technique is applied. Verification is more important than the proposed selection of feature technique. So, we use a conceptual way of making decision technique to choose a rigid, feature-set to know the Bot attack in the network of IoT environment. Correspondingly, Shannon applied entropy weight technology, inspired by this work in 2019; we also use the same theory to select the efficient features out of many.

4) Methodology

In this sub-section, the introductory definition and the basic functions of the software are proposed efficiently.

A) Soft-set: X is a regular set, as well as O, is Constrained then X gives P(X) and W is a subset of O, for instance, W ⊂ O. At this point, the coordinates (f, W) are set to X and Function f would be like f: W→ P (X).

B) Bijective-soft set: (f, O) is a soft-set and X is the universal set and its parameter is O, ex :(f and O) is a bisector if two conditions are given below is true:

1) $X_{BB\in O}$ f (BB) = X

2) Give two characteristics; BB_i, BB_j, BBj∈ O, BBi =BB_j, f (BB_i) ∩ f (BB_j) =∅

Input: Set of data-set features.

Output: Selection of Efficient feature set that we want.

a) Find a feature's set hinged on attacks in IoT, and traffic caused by those attacks.

b) The software assembly language is developed from the assembly language features defined from each, which is very efficient and ignores the other features. However, these concepts are effective theoretical concepts for our better understanding.

c) The feature-set values will be included and shown in Soft-set as well as Bijective-Soft set discretely to make a decision.

d) Create and then prioritize features for professional decision matrices like EPDM = [ρ_{ij}] aa × bb, where i = 1.....aa and j = 1.....bb; ρ_{ij}. m represents the number of experts, while n is the number of subjects displayed.

e) The projection as pv, entropy as ent, divergence as div, and weight as wt values of each dataset feature in the data set $YY_{ij}$are calculated respectively. $pv_{ij} = \frac{pij}{\sum_{i=0}^{aa} pij}$, ent= $-k1\sum_{i=0}^{aa} pv_{ij}\ln(pv_{ij})$, k1 is a constant represented as, κ1 = $(\ln(aa))^{-1}$, then div = 1 − (ent), $wt(YY_{ij}) = \sum_{k=1}^{n} \frac{\mathfrak{D}iv\ ij}{\mathfrak{D}iv\ k1}$

f) Get the desired request from a NER cyber security expert who can offer an informative selection of feature suggestions.

g) Shannon entropy weight is calculated as a soft-set along with the weight that we calculated selection value WV with the corresponding feature will be like; $wv_{ik} = \sum_j \mathfrak{D}iv_{ij}$, where $div_{ij} = wt(YY_{ij}) \times R_{ij}$. Here Rij is a choice concept.

h) Ideal (Id) and non-ideal (NId) solutions such as $YY_i^*$and $YY_i^{\check{}}$are calculated for each network expert using Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) as shown below.

YY_i^*= Max (wv_{ik1}); $YY_i^{\check{}}$= Min(wv_{ik1})

i) Calculate the separation (Δ_{ik1}^*,$\Delta_{ik1}^{\check{}}$) of the Id and NId using the n-dimensional Euclidean distance for each network expert using the relationship:

Δ_{ik1}^*= (YY_{ij} −YY_i^*) ^{2}, $\Delta_{ik1}^{\check{}}$= (YY_{ij} -$YY_i^{\check{}}$) ^{2}

Then the combined dissociation measure for each concept will be

(Δ_{k1}^* , $\Delta_{k1}^{\check{}}$); (Δ_{k1}^* , $\Delta_{ik1}^{\check{}}$)below; $\Delta^*_{k1} = \sqrt{\sum_{i=1}^{i=m} \Delta_{ik1}^*}$, $\Delta^{\check{}}_{k1}=\sqrt{\sum_{i=1}^{i=m} \Delta_{ik1}^{\check{}}}$

j) Calculate the asymptote of every object F Fζ κ1 to Ip as;

$\zeta_{k1}^* = \frac{\Delta_{k1}^{\check{}}}{\Delta_{k1}^{\check{}}+\Delta_{k1}^*}$.

The nearest measure would be the efficiency function.

Implementation

Shannon entropy TOPOSIS technique hinged on the soft-set method and to an applied effective selection of features as;

i) For efficient selection of feature, Bot attacks detection in IoT, the 5 different features are proposed to make a set of efficient selection of feature ES attributes as; ES = [ES_1, ES_2, ES_3, ES_4, ES_5], where these attributes are; ES_1 =mean, ES_2 = Std_dev, ES_3 = Ar_pt_Ddt I p, ES_4 = P_Sr_IP, ES_5 = P_D_I we will provide the following metrics for the efficient feature selection as;

ES_1 = {YY_{11}, YY_{12}, YY_{13}} = {L, Medium, H}
ES_2 = {YY_{21}, YY_{22}, YY_{23}} = {Poor, Good, Very Good}
ES_3 = {YY_{31}, YY_{32}, YY_{33}} = {Very Good, Acceptable, L}
ES_4 = {YY_{41}, YY_{42}, YY_{43}} = {Very Good, Acceptable, L}
ES_5 = {YY_{51}, YY_{52}} = {Min, Max}

ii) The concept of efficient features set are created to make combination from ES as per Table II set as; X = FF C_1 + FF C_2 + FF C_3 + FF C_4 + FF C_5 selection of feature sets are given as;

FF C_1 = {YY_{11}, YY_{21}, YY_{31}, YY_{42}, YY_{52}}
FF C_2 = {YY_{11}, YY_{23}, YY_{33}, YY_{43}, YY_{52}}
FF C_3 = {YY_{12}, YY_{21}, YY_{31}, YY_{43}, YY_{51}}
FF C_4 = {YY_{13}, YY_{21}, YY_{32}, YY_{42}, YY_{51}}
FF C_5 = {YY_{13}, YY_{21}, YY_{31}, YY_{41}, YY_{51}}

iii) For more in depth, the selected feature set is given as;

(G_1, ES_1) = {G_1 (YY_{11}), G_1 (YY_{12}), G_1 (YY_{13})}
(G_2, ES_2) = {G_2 (YY_{21}), G_2 (YY_{22}), G_2 (YY_{23})}
(G_3, ES_3) = {G_3 (YY_{31}), G_3 (YY_{32}), G_3 (YY_{33})}
(G_4, ES_4) = {G_4 (YY_{31}), G_4 (YY_{42}), G_4 (YY_{43})}
(G_5, ES_5) = {G_5 (YY_{41}), G_5 (YY_{32})}

Now the bijective-soft set can be further demonstrated as per 3 with details given below;

G_1 (YY_{11}) = {FF C_1, FF C_2}; G_1 (YY_{12}) = {FF C_3};G_1 (YY_{13}) = {FF C_4, FF C_5};
G_2 (YY_{21}) = {FF C_1, FF C_4, FF C_5}; G_2 (YY_{22}) = {FF C_3}; G_2 (YY_{23}) = {FF C_2};
G_3 (YY_{31}) = {FF C_5}; G_3 (YY_{32}) = {FF C_3, FF C_4}; G_3 (YY_{33}) = {FF C_1, FF C_2};
G_4 (YY_{41}) = {FF C_5}; G_4 (YY_{42}) = {FF C_{45}}; G_4 (YY_{23}) = {FF C_1, FF C_2, FF C_3};
G_5 (YY_{51}) = {FF C_4, FF C_5}; G_5 (YY_{52}) = {FF C_1, FF C_2, FF C_3};

The relations are true and must satisfy bijective-soft set, consider that (G_1, ES_1), then union soft-sets of (G_1, ES_1) sources, which is universal set X or X $_{YYij\in E\ FSi}$ G(YY_{1j}) = X.Deeper into two values (ES), YY_{11}, $YY_{12}\in$ ES_1, YY_{11}, YY_{12}, G_1 (YY_{11}) Ñ G_1 (YY_{12}) = ∅

iV) After applying the bijective-soft set, the preferred values are collect according to $_4$For Shannon weighted and display it as EPDM. The NER Cyber security Specialist assigns priority values as in Table I, where; L = 0.2; Avg = 0.5; H = 0.7; V.H= 0.9

v) Projection as PV, entropy as ent, divergence as (Div), and weight as(wt) of each characteristic value of the data set YY_{ij} is calculated by 5 and illustrated in table 2 After step number 5 required by a network security expert for effective selection of feature, we calculate the basic summary as follows; NER_1 = {YY_{13}, YY_{21}, YY_{31}, YY_{41}, YY_{51}} NER_2 = {YY_{12}, YY_{23}, YY_{31}, YY_{41}, YY_{52}} NER_3 = {YY_{11}, YY_{22}, YY_{32}, YY_{41}, YY_{52}}

vii) Privacy network Tabular representations of the Expert software assembly can be in Tables III, IV, and V respectively.

viii) The important part of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is held in this step so that NIp and Ip are calculated.

ix) Then, dis-aggregation measures are calculated for each NER from Ip and NIp according to 9 as illustrated in Table VII. While the above calculated and combined decomposition is presented in Table VIII respectively. After calculation, the proximity of FFζ is calculated as in Table XI, which is the efficient result of the selection of features.

x) From the table, it is given that FF ζ_5 is the result of the given function concept; FF C_5 = {YY_{13}, YY_{21}, YY_{31}, YY_{41}, YY_{51}} = ES is FF C = H, Poor, Very Good, Very Good, Min. Thus, it is given that for efficient selection of features, the feature concept needs to be considered to define precisely, as our proposal for the selection of feature-set effective.

5) Performance Measurements

Confusion metrics, which are hinged on performance measurements, are widely used to measure the recognition or discrimination performance of machine learning model results [19]. However, a graphical representation of the detailed performance calculation of the confusion matrix. In, the graphic representation row of the confusion matrix indicates an instance of the class, and the column identified by indicates the class instance [20] [26]. Nonetheless, the scales that are widely used to evaluate the ML model are described beneath:

- True-Positive (TrP): For attack identification, TrP indicates that class X was correctly identified as belonging to the same class X.
- True-Negative (TrN): This shows that class X was correctly identified as not belonging the class X.
- False-Positive (FaP): This Shows class X was not correctly detected as belonging to class X.
- False-Negative (FaN): This Shows class X was not correctly detected as not belonging to class X.

Although, you can use the upper metrics to create various metrics to better check your ML model. For precise detection, the ML classifier minimizes false positives for metric values. Anyway, the pinned metrics used in this paper are described in explained below.

- Accuracy: correctly identified sample traffic by total detected sample traffic. Although, when measuring working using metrics [21].

Mathematically it can be defined as:

Accuracy = (Trp + TrN)/ (Trp + TrN + FaP + FaN)

We used Equation four to evaluate the performance of the ML classifier, in our study. These metrics can be used to identify the validity of the ML classifier.

- Precision: Defined as a correctly detected sample as a Class A percentage of all samples detected in Class A. The formulas used in this research are shown below.

  Precision =Trp/ (Trp + FaP)

- Sensitivity: Correctly spotted traffic sample by the total recorded traffic. Anyway, this parameter can be used as a reminder for BotIoT detection in an environment of IoT. The following formula is used for this metric [22].

  Sensitivity =Trp/ (Trp + FaN)

- Specificity: This can be defined as the ability of a machine learning classifier to detect negative results. The formula for singularity is shown in Equation 7.

  Specificity =TrN/ (FaP + TrN)

However, the above metric was used to evaluate the performance of the proposed method.

RESULT AND ANALYSIS

This provides an analysis and results of the method proposed. This study proposes new ways to identify BotIoT wars in network environments of IoT [23]. To select an active feature, the method is proposed to select only 5 functional features that process sufficient information to identify BotIoT attacks in the network area of IoT [24][25]. During this functional feature selection process, four different ML algorithms used in the engineering working test are proposed. Decision Trees (C 4.5), Support Vector Machines (SVM), Naive Bayes, Random Forest Machine

Learning Algorithms, etc... However, all four ML algorithms applied can launch BotIoT attacks in an IoT network environment by usage of the set of functions pinned by the produced approach with appropriate precision, accuracy, specificity, and sensitivity are effective for detection. However, the Naive Bay performance score is lower than other ML classifiers by the usage of the feature set selected in terms of the accuracy as a metric to detect BotIoT attack. Likewise, the performance results of the SVM are a bit higher with the corresponding accuracy when compared with the Naïve Bayesian Machine Learning classifier. Although, the C4.5 decision tree and Random Forest Machine Learning Methods produce better workingoutputs in terms of accuracy. Anyway, the complete performance outs of Decision tree C4.5 provide efficient outputs when compared to other applied Machine Learning classifiers. Hence, the C4.5 Machine Learning algorithm produces better working by the usage of the features set selected to detect BotIoT attacks as 99.9%. This is an efficient performance output. Anyway, an explained accuracy result diagram is given in Figure 1.

```
------------------Checking Missing Values-------------------------

Unnamed: 0      0
pkSeqID         0
stime           0
flgs            0
flgs_number     0
proto           0
proto_number    0
saddr           0
sport           0
daddr           0
dtype: int64
```

Figure 1: Results of Data Preprocessing-Checking Missing Values

```
--------------------------Before Label Encoding--------------------
   Unnamed: 0  pkSeqID         stime  ... attack  category subcategory
0     1650261  1650261  1.528103e+09  ...      1      DDoS        HTTP
1     1650262  1650262  1.528103e+09  ...      1      DDoS        HTTP
2     1650263  1650263  1.528103e+09  ...      1      DDoS        HTTP
3     1650264  1650264  1.528103e+09  ...      1      DDoS        HTTP
4     1650265  1650265  1.528103e+09  ...      1      DDoS        HTTP
5     1650266  1650266  1.528103e+09  ...      1      DDoS        HTTP
6     1650267  1650267  1.528103e+09  ...      1      DDoS        HTTP
7     1650268  1650268  1.528103e+09  ...      1      DDoS        HTTP
8     1650269  1650269  1.528103e+09  ...      1      DDoS        HTTP
9     1650270  1650270  1.528103e+09  ...      1      DDoS        HTTP

[10 rows x 47 columns]

------------------------After Label Encoding----------------------

   Unnamed: 0  pkSeqID  stime  ...  attack  category  subcategory
0           0        0   5383  ...       1         0            0
1           1        1   5384  ...       1         0            0
2           2        2   5385  ...       1         0            0
3           3        3   5386  ...       1         0            0
4           4        4   5387  ...       1         0            0
5           5        5   5388  ...       1         0            0
6           6        6   5389  ...       1         0            0
7           7        7   5390  ...       1         0            0
8           8        8   5391  ...       1         0            0
9           9        9   5392  ...       1         0            0

[10 rows x 47 columns]
```

Figure 2: Results of Data Preprocessing-Before and After Labeling

Figure. 2 shows the result of examined precision output. By this fig, it is convincing that Decision tree C4.5 and Random Forest ML classifiers secure effectual working outputs in comparison with SVM algorithm and Naive Bayes. Anyway, common data transfer hijacking and keylog hijacking strikes are spottedpotently but perform worse than UDP DoS and other attacks with corresponding accuracy metrics. However, the average of all the performance results of the ML classifier is applied. It was found that only Keylogging Theft traffic was weakly spottedin comparison to other attacks and other attacks using fixedattributes with corresponding exact variables. All 4included ML algorithmsattain high working outputs with their discrete sensitivity variables. Anyway, the machine learning algorithm Random Forest and Decision C4.5 achieves biggerworking outputsby the usageof the pickedattribute set in comparison to other ML algorithms applied for BotIoT attack spotting in the environment of the IoT network. For sensitivity metrics, like accuracy and precision, the performance results of the SVM classifier and Naive Bayes ML are weaker than those of the C4.5 decision tree and the Random Forest ML algorithm. Figure 3 shows the sensitivity results with accuracy.

```
--------------------Random Forest-----------------------------
ACCURACY 99.99618233183172

              precision    recall  f1-score   support

           0       1.00      1.00      1.00       835
           1       1.00      1.00      1.00       418
           2       1.00      1.00      1.00     24841
           3       1.00      1.00      1.00       100

    accuracy                           1.00     26194
   macro avg       1.00      1.00      1.00     26194
weighted avg       1.00      1.00      1.00     26194

---------------------Decision Tree-----------------------------
ACCURACY 99.9809116591586

              precision    recall  f1-score   support

           0       1.00      1.00      1.00       835
           1       1.00      1.00      1.00       418
           2       1.00      1.00      1.00     24841
           3       0.96      1.00      0.98       100

    accuracy                           1.00     26194
   macro avg       0.99      1.00      0.99     26194
weighted avg       1.00      1.00      1.00     26194
```

Figure 3: Sensitivity Results

```
----------------Support Vector Machine--------------------------
ACCURACY 99.96564098648545

              precision    recall  f1-score   support

           0       1.00      1.00      1.00       835
           1       1.00      0.98      0.99       418
           2       1.00      1.00      1.00     24841
           3       0.98      1.00      0.99       100

    accuracy                           1.00     26194
   macro avg       1.00      0.99      0.99     26194
weighted avg       1.00      1.00      1.00     26194

---------------------Naives Bayes-----------------------------
ACCURACY 99.99236466366344

              precision    recall  f1-score   support

           0       1.00      1.00      1.00       835
           1       1.00      1.00      1.00       418
           2       1.00      1.00      1.00     24841
           3       1.00      0.99      0.99       100

    accuracy                           1.00     26194
   macro avg       1.00      1.00      1.00     26194
weighted avg       1.00      1.00      1.00     26194
```

Figure 4: Results of Classifiers

Learning classifiers in Figure 4. The detailed outputs of the included Machine Learning algorithms can be seen by the usage of our produced way, pickedattributes set for the detection of Botstrikes in the network environment of IoT. All the included Machine Learning algorithms conducting outputs are very effectual concerning specificity as Random Forest, and Decision tree C4.5 is 99.99% and 98.95% while SVM and Naive Bayes are 98.48% and 98.44%, this is very effectual working respect to specificity metric. Likewise, total strikes and general data transfers are very strongly identified by the usage of the picked attributes set. By the audit of upon results that our purposive selection of features method is more productive for the features picking for BotIoT recognition in the environment of IoT network.

Analysis and Discussion

The results of the produced method promise to detect BotIoT attacks in an IoT network environment by usage of 4 different ML algorithms selected, but with accuracy, accuracy, and sensitivity, it uses the peculiarities of the newly developed data set of BotIoT. Anyway, information that is useful obtained after the analysis of the experiment is shown beneath.

- This study clearly shows that the proposed method for selecting optimal features is effective in detecting BotIoT attacks in IoT network environments using freshly developed BotIoT datasets. Analyzing and evaluating the results of the uses accuracy, accuracy, sensitivity, and specificity metrics to tightly assess the performance of the presumed method.
- From this study, the proposed method must select the best features, including sufficient knowledge of detection information to spot and identify cyber-attacks in environments of IoT networks.
- Note that the strength of the applied Machine Learning algorithm is promising with the accuracy, precision, sensitivity, and specificity of the corresponding. However, while all attacks are detected very accurately, only the Key logging Theft attack is inadequately detected compared to the rest attacks.
- Analyzing the experimental results, the performance of the used Machine Learning algorithm is fruitful in detecting BotIoT attacks. However, if the uses the BotIoT dataset, the decision tree which is C4.5, and the Machine Learning algorithm for the Random Forest are very promising. The performance of the Support Vector Machine and Naive Bayes algorithms is also potent, but the performance of the bit weak equated to the decision tree used and random forest algorithms.

Conclusion

Detecting attacks on the IoT network is required for monitoring and blocking traffic flow that does not require IoT security. Many researchers have implemented several ML (Machine learning) technology models to block the attack traffic flow in IoT networks. However, some ML models misclassify most malign traffic streams due to improper feature selection. Nonetheless, the notable issue needs further investigation. In short, it is an effective feature selection for the accurate identification of malevolent traffic in your IoT network. A-frame model has been introduced for this purpose. CorrAUC is a new feature selection measure that is introduced first, and Corrauc is a new feature selection technique built on CorrAUC. The Bijection software package was used to validate the chosen characteristics for the detection of malicious traffic in the IoT network after we used Shannon Entropy and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). Utilizing the BotIoT dataset and four distinct ML algorithms, evaluate the suggested strategy. Analysis of the experiment's findings revealed that our developed strategy was effective because it often yielded outcomes of greater than 96 percent.

References

Pavaiyarkarasi, R., T. Manimegalai, S. Satheeshkumar, K. Dhivya, and G. Ramkumar. A Productive Feature Selection Criterion for Bot-IoT Recognition based on Random Forest Algorithm. In 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT). 2022; 539-545.

Ullah, Imtiaz, and Qusay H. Mahmoud. Design and Development of RNN-based Anomaly Detection Model for IoT Networks. IEEE Access. 2022.

Chauhan K, Gandotra E. Comparative Analysis of Machine Learning Methods for Classification of IoT Malware Attacks. InSoft Computing: Theories and Applications. 2022; 157-167.

Om Kumar, Chandra Umakantham, Jeyakumar Durairaj, Samsu Aliar Ahamed Ali, Y. Justindhas, and Suguna Marappan. Effective intrusion detection system for IoT using optimized capsule auto encoder model. Concurrency and Computation: Practice and Experience. 2022.

Selvakumar, B., B. Lakshmanan, and S. Sridhar Raj. Hybrid Framework Combining Deep Learning and Grey Wolf Optimizer for Anomaly Detection in IoT-Enabled Systems. In Soft Computing: Theories and Applications. 2022; 59-68.

Leevy, Joffrey. Machine Learning Algorithms for Predicting Botnet Attacks in IoT Networks. Ph.D. diss., Florida Atlantic University. 2022.

Araujo, Alex Medeiros, Anderson Bergamini de Neira, and Michele Nogueira. Autonomous machine learning for early bot detection in the internet of things. Digital Communications and Networks. 2022.

Majhi, Babita. An Improved Intrusion Detection System using BoT-IoT Dataset. In 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT). 2022; 488-492.

Azath, H., D. Beulah David, E. Chandra Blessie, A. Jayapradha, and S. Sheeba Rani. BoT-IoT based Denial of Service Detection with Deep Learning. In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC). 2021; 221-225.

Pavaiyarkarasi, R., T. Manimegalai, S. Satheeshkumar, K. Dhivya, and G.

Ramkumar. A Productive Feature Selection Criterion for Bot-IoT Recognition based on Random Forest Algorithm. In 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT). 2022; 539-545.

Sakthivel, M., R. Kamalraj, J. Udaykumar, Spectrum and Delay Efficient Routing Protocol for Mobile Cognitive Ad-Hoc Networks. International Journal of Recent Technology and Engineering (IJRTE). 2020; 8(5): 3529-3533.

Sakthivel, M., J. Udaykumar, and V. Saravana Kumar. "Progressive AODV: A Routing Algorithm Intended for Mobile Ad-Hoc Networks." International Journal of Engineering and Advanced Technology (IJEAT). 2019; 9(2): 70-74.

Kamalraj, R., and M. Sakthivel. A hybrid model on child security and activities monitoring system using iot. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE. 2018.

Goundar, Sam, J. Avanija, Gurram Sunitha, K. Reddy Madhavi, and S. Bharath Bhushan. Innovations in the Industrial Internet of Things (IIoT) and Smart Factory. IGI Global. 2021.

Kundu, Partha Pratim, Tram Truong-Huu, Ling Chen, Luying Zhou, and Sin G. Teo. Detection and Classification of Botnet Traffic using Deep Learning with Model Explanation. IEEE Transactions on Dependable and Secure Computing. 2022.

Araujo, Alex Medeiros, Anderson Bergamini de Neira, and Michele Nogueira. Autonomous machine learning for early bot detection in the internet of things. Digital Communications and Networks. 2022.

Shahhosseini, Mohaddeseh, Hoda Mashayekhi, and Mohsen Rezvani. A Deep Learning Approach for Botnet Detection Using Raw Network Traffic Data. Journal of Network and Systems Management. 2022; 30(3): 1-23.

Gupta, Brij Bhooshan, Akshat Gaurav, Enrique Caño Marín, and Wadee Alhalabi. Novel Graph-Based Machine Learning Technique to Secure Smart Vehicles in Intelligent Transportation Systems. IEEE Transactions on Intelligent Transportation Systems. 2022.

Akshaya, V., M. Sathyapriya, R. Ranjini Devi, and S. Sivanantham. "Detecting Credit Card Fraud Using Majority Voting-Based Machine Learning Approach." In Intelligent Systems and Sustainable Computing. 2022; 327-334.

Sivanantham, S., S. R. Dhinagar, P. Kawin, and J. Amarnath. "Hybrid approach using machine learning techniques in credit card fraud detection." In Advances in Smart System Technologies. 2021; 243-251.

Sakthivel M, Sivanantham S, Kamalraj R & Krishnamoorthy V, "An Analysis of Machine Learning Depend on Q-MIND for Defencing the Distributed Denialof Service Attack on Software Defined Network", International Journal of Early Childhood Special Education. 2022; 14(05): 3769 – 3776.

Lakshmi Haritha.M & K. Ramani. Impact of Deep Learning on Localizing and Recognizing Handwritten Text in Lecture Videos. International journal of Advanced Computer Science and Applications; 12(4), 2021.

P. Lakshmi Sagar (2022). Resolving sets and Dimension in Bidiakis Cube and Durer Graphs. International journal of Neuro Quantology; 20: 4988-4992.

Silpa C, RamPrakash Reddy Arava, K.K. Baseer. Agri Farm: Crop And F International Journal of Early Childhood Special Education ertilizer Recommendation System for High Yield Farming Using Machine Learning Algorithms; 14(5): 2022.

P. Dhanalakshmi et al. (2022). Application of Machine Learning in Multi-Directional Model to Follow Solar Energy Using Photo Sensor Matrix. International journal of Photo energy; 9:1-9.

Siva Kumar Depuru & Dr.K. Madhavi. (2019). Autoencoder Integrated Deep Neural Network for effective analysis of malware in distributed Internet of Things (IoT) Devices. The International journal of analytical and experimental modal analysis; 9(11): 226-232.