iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35622

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

RANFO: An Intelligent Anomaly Detection in IoT Edge Devices

Anuvidhya R[1], Subramaniyaswamy V[1]
[1, 2]*School of Computing, SASTRA Deemed University, Thanjavur 613401, India*

Abstract: ***As devices, applications, and communication networks become more connected and integrated, computer attacks on the Internet of Things (IoT) become more sophisticated. When attacks on IoT networks cause long-term outages, it affects the availability of critical end-user programmers, increases the number of data breaches and fraud, raises prices, and reduces revenue. In this paper we present the RANFO (IDS), prepared to protect inherently linked Iot systems. The proposed entry-level system can successfully enter real-world entrance, according to our experimental results. We'll illustrate how RANFO can identify a variety of harmful assaults, including DOS, R2L, Probe, and U2L.***
Keyword: Cyber-attack, Internet of Things, Random Forest, Intrusion Detection System.

I. INTRODUCTION

An IDS may be a internet security technology basically designed for sleuthing exploiting the vulnerable against a mark application . It's imperative to find attacks on IoT systems in actual time to sustain made security and defence. The stack is pushed up and examined in the protocol for these anomalies. The detection response ensures security as a service and it allows for the ability of multiple IoT communications network protocols. The sting Computing paradigm to find cyberthreats as shut as attainable to the corresponding information sources.

This attack is confusing which can be a recording software method, which consists of finding startup bugs using the wrong / incorrect data injection in the default way. A malware attack could be a code that takes the vulnerability of a software vulnerability or a security error. If the use used could be a criminal code to connect the network remotely and get higher privileges, or to logging in to someone who does not have remote connection to the network and get higher privileges, or deeper into the network. In some cases, exploitation is used as component of a massive attack. Data collection and in-depth knowledge of the target system is considered “Reconnaissance”. This data is a highway hacking system targeted program. It includes foot printing, counting, scanning. A computer worm can be a recurring malware Trojan horse, relying on security failures on a targeted computer to access it. Edge computing is the way conventions are applied to iot. Many methods did not guarantee to tell the process and will be a source of data and this is focused on the various networks connected along the way and what the paradigm does to find cyber routes and do those in the future .

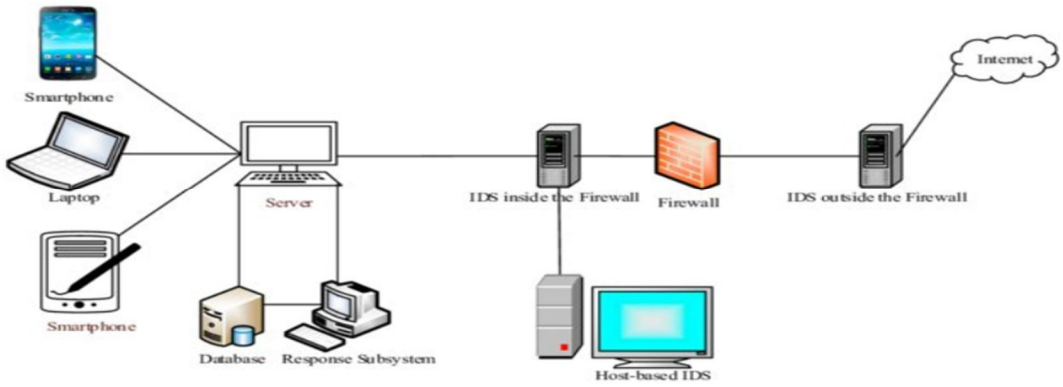


Fig:1 Architecture of IDS

A RF is fabricated into innumerable DT's that are mingled to form random forest algorithm, it makes a book of balancing the forecasts of the every elemental tree. It always has finer calculating estimate than one DT [25] . The RF Classifier, as the title suggests, creates a cover of forest with many DT's. It's a superintented classification algorithm. It's a beautiful algorithm promote to the towering fastness. The more trees in a forest, The stronger it seems. From a randomly selected section of the coaching set, the RF algorithm builds a pile of DT's . The votes from separate DT's are then combined to predispose the eventual class of the test item.

An IoT is referring to a chain of interconnected, Internet-connected elements. Random forest planners, because the name implies that a RF algorithm bring into being a forest with many deciduous trees. supervised editing algorithm [6]. Is to forge a RF and this predicts by guesstimating the predictions of entire trees of elements. It always has much better predictability than a single decision tree. Usually, when trees grow in the cover of forest, the forest looks strong. Random Forest Planning forms a tree-cutting group from a haphazardly selected set of training sets. It then includes votes from various decision-makings to regulating the mounting result of the element.

Random Forest predicts a separation algorithm that contains multiple decision trees. Random Forest (RF) creates a amount of decision-making trees in training. The answer of this value is determined as the contraction of the node pollution loaded with the atempt of contacting that node, using predictions from all trees [22] . The purport of specimens that reach the node splited by the plenary volume of specimens can be used to figure out the possibilities. The mattering much factor is higher value.

II. RELATED WOKS

Related activities are based on a baseline paper, the processes employed and test-related tasks are recognised. Recent research activities relating to intelligent IDS operating on traditional networks are abundant in the literature. In IoT environments, little has been done explicitly for IDS. History shows a limited ability to compute these devices, making it impossible to create a whole IDS, as is the technically most challenging impediment to dealing with IDS on the IoT routes. Recent literature is rich in intelligent IDS research activity on traditional networks.

Little was done for IDS in IoT environments in particular.

Yoshua Bengio et al. introduced hyper – parameter optimization using grid and manual search [19] in order to make the drawback of this experiment is grid search. The [11,15,17] work shows that random search is a natural baseline against which to judge progress in the development of adaptive (sequential) hyper-parameter optimization algorithms and different techniques are proposed

Victor G. Turrisi da Costa et al. Discovered the one class classification for an internet of things to detect botnets [14] . The successive rate of this paper is to compromise the multiple devices and perform co-ordinate the attacks. It developed a host based detection system using classification techniques. Zhiguo Ding et al. proposed a model of sliding window for the approach that is streaming data based on Isolation Forest algorithm[16] . In the concept of drift phenomena will use a anomaly detection approach .

Tsutomu Matsumoto et al. ascertained a honey pot for Internet of things of threat in revealing current called " IoTPOT " [18] . Here the telnet based attacks , honeypots , sandbox ehich are come under the Distributed denial of service attack.

Massimo Vecchio et al. distinguished smart audio sensors for anomaly detection in internet of edge things [20] . The Isolation forest and Elliptic Envelope are detecting algorithms are adopted here.

Giovanni Russello et al. achieved a security of internet of things frameworks is an new technique for the ecosystem of IoT frameworks [1] in order to make the ecosystem more reliable and flexible.

Xiaojion Wang et al. imported the intrusion detecting the IoT in botnet using machine learning technique [4] In detection , Logistic regression , svm , Random forest are experimented here to detect the anomalies in the IoT.

Nabila Faranaaz et al. established the random forest modelling in intrusion detection system for the network topology [2] taking NSL-KDD dataset for this experiment . For this methodology will detect the anomalies in the given dataset. Ishark Isam et al. proposed a model called anomaly detection in internet of things using sensors at sites using machine learning techinques.

III. METHODOLOGY

A. RANFO IDS

The RANFO IDS is the model which is calculated to encounter an IOT to a system of many decision trees. It is usually has much better predictive accuracy than a singly decision tree. The expanded of trees in the heavily booming the forest looks.

During this paper , it implies the random forest algorithm creates the forest with many decision trees [24] . Many decision trees ensemble together to procreate a random forest and it predicts by averaging the predictions of every component tree.

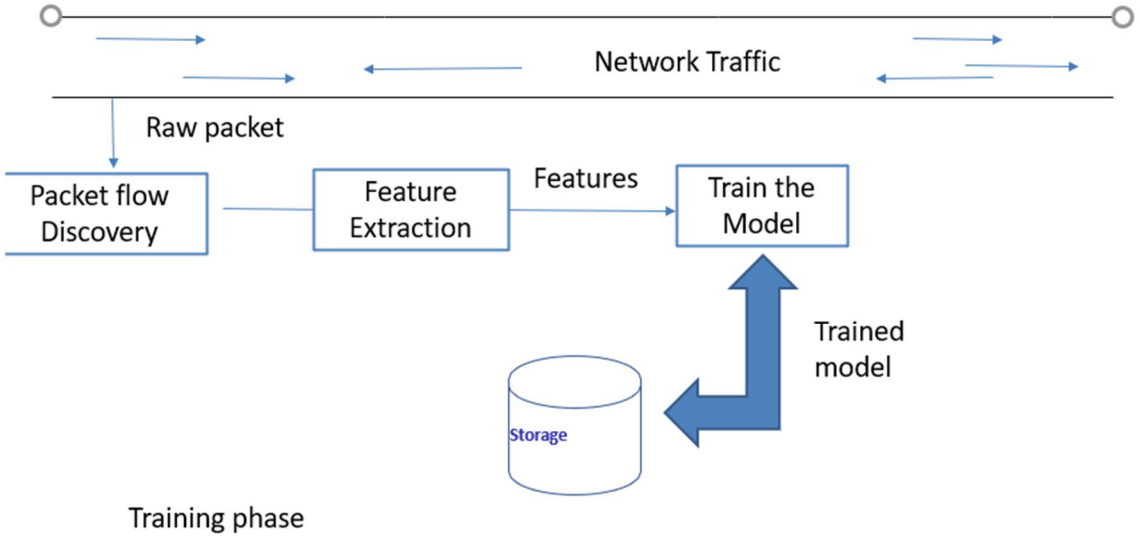


Fig : 2 Architecture Diagram of Packet capturing

Random Forest predicts a separation algorithm that contains hetrogenous decision trees. Build more decisive trees in training [3] .The forecast from all the various trees is compiled to create greater predictions. After that it includes votes on the various trees of the various decisions to make your mind set the experiment's concluding stage [7, 22] . Random Forest planning assembles a agglomeration of DT's from a asortment of randomly chosen training trees.

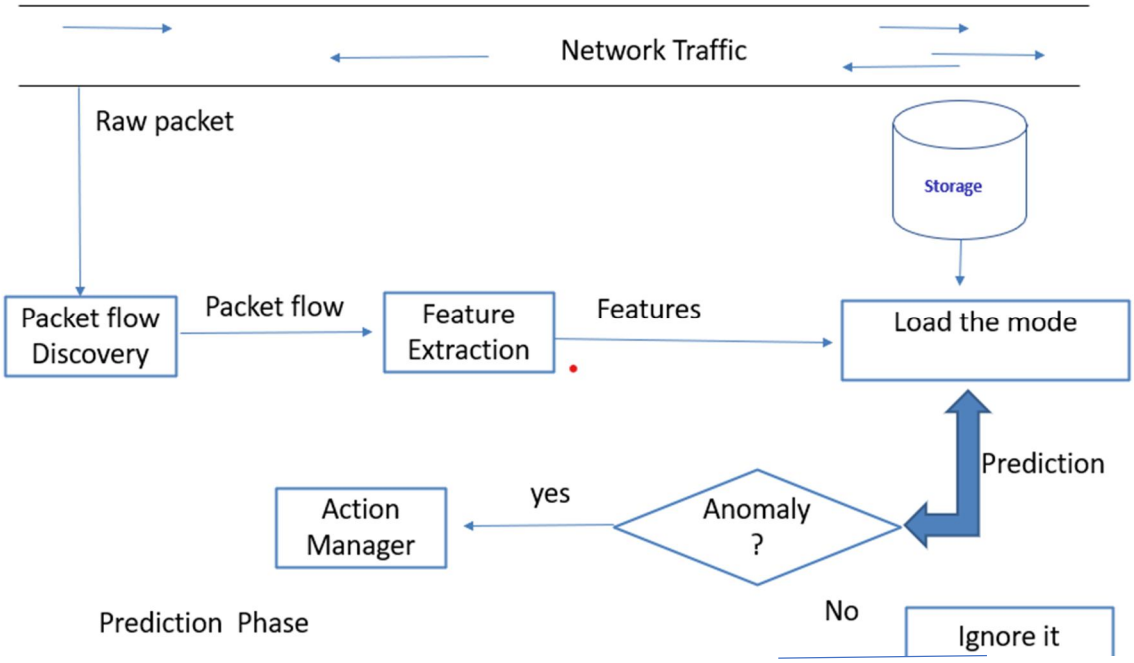


Fig : 3 Work Flow Diagram of Anomaly detection

Because the decline in node impurity is weighted by the likelihood of reaching the node, feature significance is determined. The node frequency is estimated by break down the total range of values by the amount of samples that reach the junction.

Algorithm: RANFO

```
Input: Ni: Network interface;
Output: ML: Trained Machine Learning Object;
ML = new MachineLearning ();
if (Exist trainedModel) then
        ML.loadModel(trainedModel);
end
else
        ms = new Dataset ();
        rf_data = RandomForest.process(ds) ##new line added - processing data using RF
        tf = new NetFlow ();
        for (tf in Ni.getNetFlow()) do
                tf = Ni.getNetFlow();
                rf_data.add(nf); ##new line added - Use the processed data
        end
        trainedModel = ML.train(rf_data);  ##new line added - Train the model using the RF
data
        store( trainedModel);
end
return ML;
```

The higher [8,9] the power the more important the feature. it's a supervised classification algorithm. it's a gorgeous classifier because of the high fastness.

B. Implementation

1) *Preparing Dataset (Test and Train):* Using Labelencoder and One -hot-encoding to include category features to the same 2D numpy members. Converting sector features to numbers using LabelEncoder () .One hot code coding is used here. Factors are rated to override features with large amounts that may be too heavy for the result.

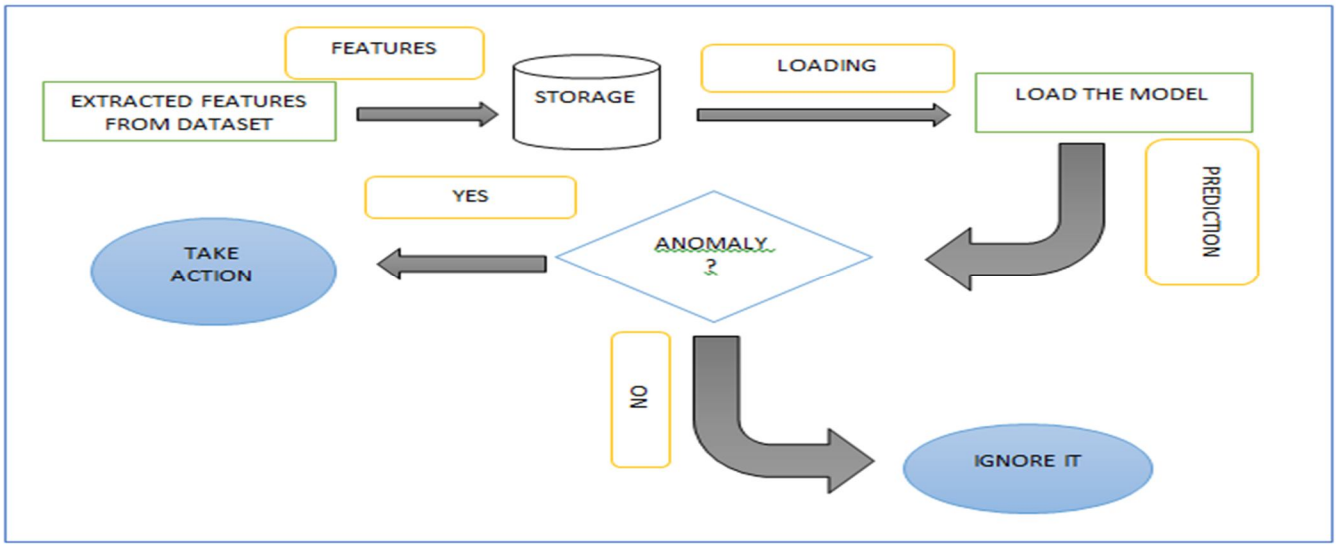


Fig :3.2 Workflow Diagram Of Ranfo Detecting System

2) *Feature Scaling:* Pre-processing is done by Transforming of all data in X and Y .
3) *Feature Selection:* Remove unwanted and inactive data by selecting the appropriate subset features that fully represent the given problem. Feature selection with Random Forest Classifier and Recursive Feature Elimination (FRE). So RFC makes feature selection based on database and algorithm. This is not a personal choice of prices[21] . This looks at individual parameters to see how strong the link between the characteristic and the labels is.
4) *Building the Model:* The model is build by the Decision tree
5) *Prediction1:* Proving a test date to make promote the calculations. Many scores are treated such accuracy score , recall , f-measure , confusion matrix. Using cross_val_score" from sklearn.

C. Attacks in IoT

There is four attacks just as Denial of service attack, Probe attack (Surveillance or other probing attack) , R2L attack (which is Unaroused access from lonesum location) and , U2R attack (which is Unaroused access to regional privileges root).There are 5 total number of classes such as Normal , Dos , Probe , R2L , U2L.

The Dos attack comprised of the following attack are Neptune attack , Back attack, Land attack, Pod attacks , Smurf attacks , Tearddrop attacks , Mailbomb attacks , Apache attacks , Processtable attacks , udpstorm attacks , worm .

The Probe attack consists of the following attacks are ipsweep attack , nmap attack , portsweep attack , satan attack , mscan , saint attack .The R2L attack are consists of ftp_wrint attack , guess_pswd attack , imap , multihop attack , spy attack , warezclient attack , warezmaster attack , sendmail attack , named attack , snmpge attack , snmp_guess attack , xlock attack , sxnoop attack , httptunnel attack.The U2R attack comprised of following attacks such as buffer overflow attack , loadmodule attack , perl attack , rootkit attack , port scanning attack , sql attack , xterm attack.

D. Advantage & Application Of Algorithm

The beauty of the random forest algorithm is that there is no problem of overrun going. For any partition problem, we can apply a random forest technique. For partitioning and regression, the same random forest approach is frequently employed. For feature engineering, the random forest algorithm is frequently used. Application land is Banking, Medicine, exchange and E-commerce.

IV. EXPERIMENTAL RESULT AND DISCUSSION

A. Dataset

The Training data and Testing data are the two division of dataset on the basis of training data the model is created [13]. The incoming test data is to test whether the model is able to detect the correct values or not. It will be based on the TP, FP, TN, FN, Recall, Precision and F1 score / F1 measure. In this dataset we have attacks that dip into four kinds such as Denial Of Service attack, R2L (guessing password attack), U2R (Buffer Overflow attack), Probing Attack (Port Scanning). Based on the feature set, the dataset will be collected. These attacks will be stimulated through the agile gateway to the packet. The feature set of the dataset .

The feature set of the dataset is collected by the features based on the attacks is given below.

Table 4.1 Feature Set Used In The Experimentation

Type	Feature ID			
DOS attack	Wrong_fragment	srv _count		
	dst_bytes	diff _srv _rate		
	Src_bytes	dst _host _serror _rate		
	num _compromised	dst _host _srv_serror _rate		
	Count	flag _SF		
Probe attack	Count	dst_host_count	dst_host_srv_count	
	Dst_bytes	dst_host_srv_count	dst_host_same_srv_rate	
	dst_host_same _port_rate	dst_host_diff_src_rate	dst_host_rerror_rate	
	Protocol_type_icmp	Protocol_type_tcp	Service_eco_i	
R2L attack	dst_bytes	Src_bytes	Hot	
	Duration	num _compromised	Logged_in	
	Is_gues_loggin	Dst_host_count	Dst_host_srv_count	
	dst_host_diff_srv_ rate	dst_host_same_srv_rate	dst_host_rerror_rate	
U2L attack	src_bytes	duration	dst_bytes	Count
	dst_host_	count	hot	root_shell
	dst_host_count	dst_host_srv_count	dst_host_same_srv_rate	dst_host_diff_srv_rate

We constructed the training data, testing data in the dataset and evaluate the metrices based on the algorithm usage.

Table 2
Dataset Summary

TRAINING DATA	TESTING DATA
❖ Total - 125973 ❖ Rows – 125973 ❖ Columns – 42	❖ Total – 22544 ❖ Rows – 22544 ❖ Columns – 42

B. Experimental Setup and Evaluation

The proposed framework is implemented using python. Sklearn, Matplotlib, Numpy and Pandas were libraries used for a variety of purposes. The test set is used here to test the data and the test matrix for accuracy, memory, accuracy and f1 rating provides greater accuracy in the use of random forest algorithm. In all of these cases, frequency, matrix provisos of the perplixity are also included to calculate Precision (P), Recall (R) and harmonic mean (e.g., F-value, alias F1 points). The calculations used to calculate this accuracy metrics are accustomed in Table 3) Finally, it is importunate to remember that FP is still a clear signal of performance in Intrusion Dectection System testing.

TABLE 3
Machine Learning Established In Evaluation Metrics Used To Check The Enforcement Of Ranfo

Performance Measure	**Formula**
Precision	P = TP / TP +FP
Recall	R = TP/ TP+FN
F-measure	F1 = 2 x (P.R) / P+R

$$\text{Precision} = \frac{\text{Number of correctly matched}}{\text{Total number of extracted}}$$

$$\text{Recall} = \frac{\text{Number of correctly matched}}{\text{Total number of assigned}}$$

$$\text{F1- score} = \frac{2 \text{ X Precision X Recall}}{\text{Precision +Recall}}$$

C. Result

The four datasets we collected and experiment is foundout. The results of our experiments for the four datasets are displayed in this Table:4. The table pageants the evaluation values of the Precision, Recall, F1 score using the technique of Random Forest algorithm will detect the anomalies in the dataset.

Table 4
Performance Evaluation Results

Attack	Technique	FP	TP	FN	TN	Precision	Recall	F1
DOS	Random Forest	28	9683	3074	4386	0.97	0.88	0.86
Probe	Random Forest	372	9339	1030	372	0.97	0.97	0.97
R2L	Random Forest	9711	0	3	2885	0.99	0.99	0.99
U2L	Random Forest	0	9711	65	2	0.97	0.88	0.86

The Correlation Matrix shows the correlation between the two columns.

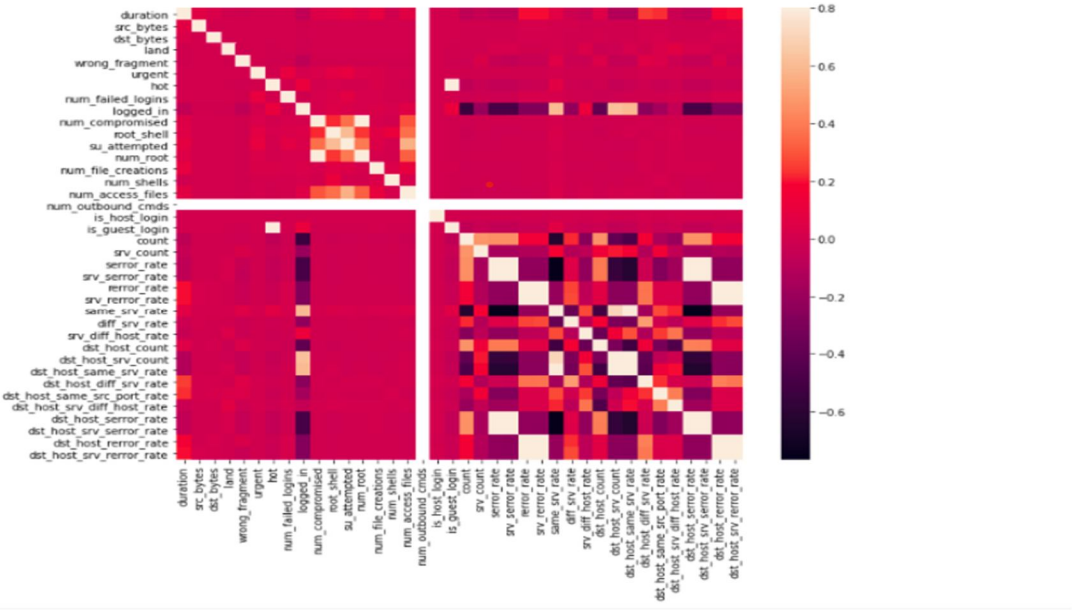


Fig :4.3 Correlation matrix of RANFO detecting Anomalies IoT Edge devices

D. Discussion

The model is observed by the random forest algorithm it will detect the anomalies in the data packet which is done with the help of feature extraction model. There are four kinds of attacks are detected in the experiment. Having five classes in this model such as Normal , Dos , Probe , R2L , U2R . Using the Random forest classifier will detect the anomalies in the given dataset. With the help of evaluation matrix (TP, FP,FN,TN) such as precision , recall , accuracy ,and F1 measure stage shows a significant aspect in this experiment. The confusion matrix and the correlation matrix shows the result of this experiment . The main motive is to minimize the False Positive rate.

The comparative effect of this model is much more accurate and more efficient than the existing model available with the Local Outlier Factor algorithm and the Isolate Forest algorithm. Here, using a random forest algorithm will detect errors or malwares or inaccuracies in the data explicitly.

V. CONCLUSION

In this work, the RANFO of the proposed novel, an anonymous intrusion detection system based on anomaly-based intrusion detection (IDS) is designed for direct handling and is done with a standard edge device. We used RANFO using and testing it with common attacks (DOS, U2R, Probe R2L). The use of the Random Forest algorithm will detect irregularities in the data. With the detection of a threat, the accuracy is very high in this model. Depending on the test matrix, (F1-score, Accuracy, Precision, Re-call) this random forest editor plays a major role in obtaining accurate results without any algorithm. A model made with high precision and memory points using a random forest algorithm and done slowly to another algorithm.

VI. FUTURE WORK

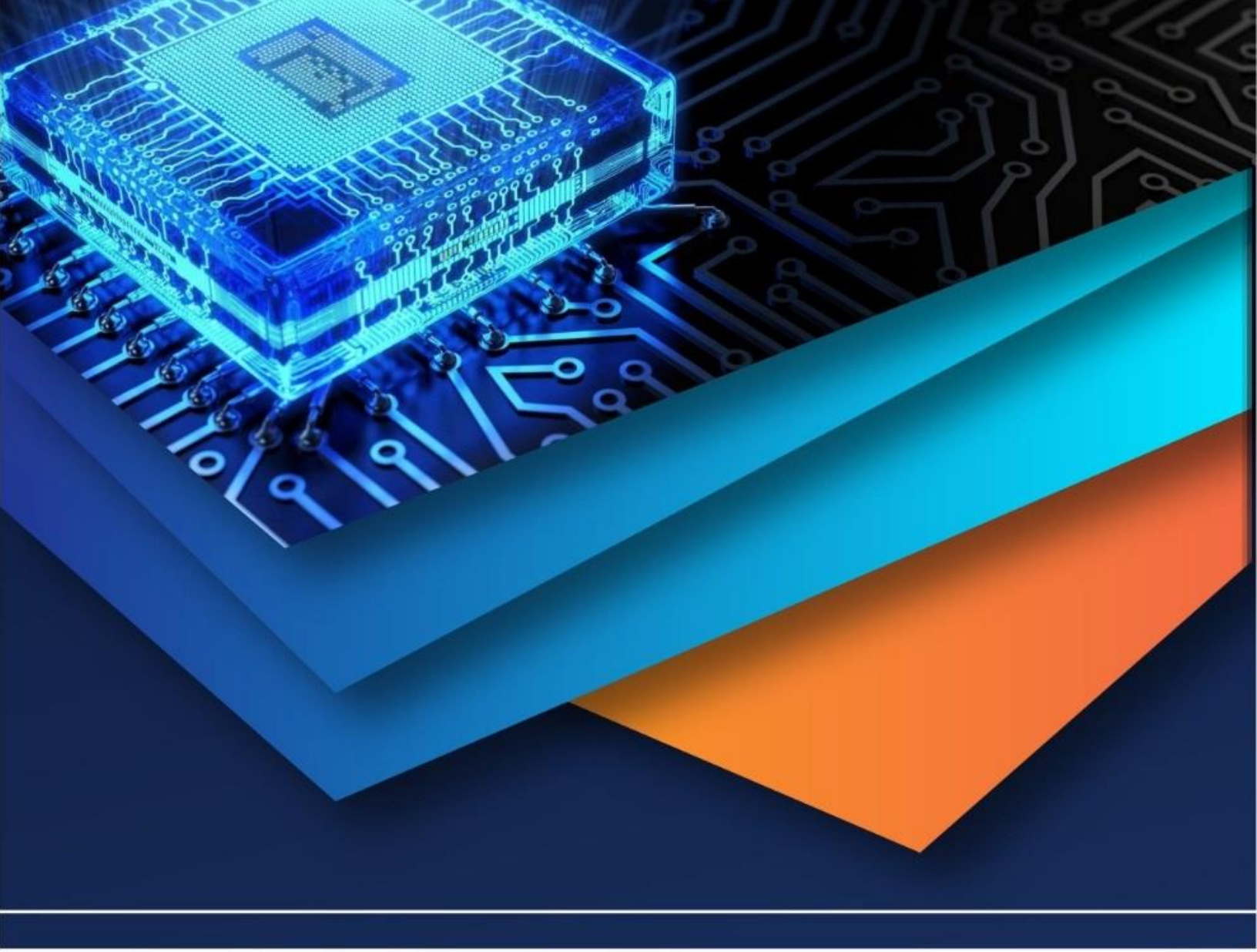
In addition, in the future we plan to use a collection of approaches in machine learning algorithms to recoganize inaccuracies of incoming packets. We will use different tools to collect the package to make it more efficient and suitable for the feature removal model**.**

VII. ACKNOWLEDGEMENT

The authors gratefully acknowledge the Science and Engineering Research Board (SERB), Department of Science &Technology, India for financial support through Mathematical Research Impact Centric Support (MATRICS) scheme (MTR/2019/000542).

REFERENCES

[1] Hasan, M., Islam, M.M., Zarif, M.I.I. and Hashem, M.M.A., 2019. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. *Internet of Things*, *7*, p.100059.
[2] Farnaaz, N. and Jabbar, M.A., 2016. Random forest modeling for network intrusion detection system. *Procedia Computer Science*, *89*, pp.213-217
[3] Lee, S., Abdullah, A., Jhanjhi, N. and Kok, S., 2021. Classification of botnet attacks in IoT smart factory using honeypot combined with machine learning. *PeerJ Computer Science*, *7*, p.e350
[4] Bagui, S., Wang, X. and Bagui, S., 2021. Machine Learning Based Intrusion Detection for IoT Botnet. *International Journal of Machine Learning and Computing*, *11*(6).
[5] Hosseini, S., Nezhad, A.E. and Seilani, H., 2021. Botnet detection using negative selection algorithm, convolution neural network and classification methods. *Evolving Systems*, pp.1-15.
[6] Hosseini, S., Nezhad, A.E. and Seilani, H., 2021. Botnet detection using negative selection algorithm, convolution neural network and classification methods. *Evolving Systems*, pp.1-15.
[7] Becker, R.A., 2015. Cyber Attack on German Steel Mill Leads to'Massive'Real World Damage. *online], PBS. Org*
[8] Eskandari, M., Janjua, Z.H., Vecchio, M. and Antonelli, F., 2020. Passban IDS: an intelligent anomaly-based intrusion detection system for IoT edge devices. *IEEE Internet of Things Journal*, *7*(8), pp.6882-6897.
[9] R. Tahir, "A study on malware and malware detection techniques," *International Journal of Education and Management Engineering*, vol. 8, no. 2, p. 20, 2018.
[10] Batra, D., 2017. Adapting agile practices for data warehousing, business intelligence, and analytics. *Journal of Database Management (JDM)*, *28*(4), pp.1-23
[11] Kasinathan, P., Costamagna, G., Khaleel, H., Pastrone, C. and Spirito, M.A., 2013, November. An IDS framework for internet of things empowered by 6LoWPAN. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 1337-1340)
[12] Becker, R.A., 2015. Cyber Attack on German Steel Mill Leads to'Massive'Real World Damage. *online], PBS. Org*
[13] Robert, L.M., Michael, J. and Tim, C., 2016. Analysis of the cyber attack on the ukrainian power grid. *USA: Electricity Information Sharing and Analysis Centre (E-ISAC)*
[14] Ding, Z. and Fei, M., 2013. An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window. *IFAC Proceedings Volumes*, *46*(20), pp.12-17.
[15] Bilge, L. and Dumitraş, T., 2012, October. Before we knew it: an empirical study of zero-day attacks in the real world. In *Proceedings of the 2012 ACM conference on Computer and communications security* (pp. 833-84
[16] Pa, Y.M.P., Suzuki, S., Yoshioka, K., Matsumoto, T., Kasama, T. and Rossow, C., 2016. IoTPOT: A novel honeypot for revealing current IoT threats. *Journal of Information Processing*, *24*(3), pp.522-533.
[17] Al-Maksousy, H.H., Weigle, M.C. and Wang, C., 2018, October. NIDS: Neural network based intrusion detection system. In *2018 IEEE International Symposium on Technologies for Homeland Security (HST)* (pp. 1-6). IEEE
[18] Zeebaree, S.R., Jacksi, K. and Zebari, R.R., 2020. Impact analysis of SYN flood DDoS attack on HAProxy and NLB cluster-based web servers. *Indones. J. Electr. Eng. Comput. Sci*, *19*(1), pp.510-517
[19] Bergstra, J. and Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2)
[20] Antonini, M., Vecchio, M., Antonelli, F., Ducange, P. and Perera, C., 2018. Smart audio sensors in the internet of things edge for anomaly detection. *IEEE Access*, *6*, pp.67594-67610.
[21] Elrawy, M.F., Awad, A.I. and Hamed, H.F., 2018. Intrusion detection systems for IoT-based smart environments: a survey. *Journal of Cloud Computing*, *7*(1), pp.1-20.
[22] Chaabouni, N., Mosbah, M., Zemmari, A., Sauvignac, C. and Faruki, P., 2019. Network intrusion detection for IoT security based on learning techniques. *IEEE Communications Surveys & Tutorials*, *21*(3), pp.2671-2701.
[23] Li, F., Shinde, A., Shi, Y., Ye, J., Li, X.Y. and Song, W., 2019. System statistics learning-based IoT security: Feasibility and suitability. *IEEE Internet of Things Journal*, *6*(4), pp.6396-6403
[24] Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D. and Elovici, Y., 2018. N-baiot—network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, *17*(3), pp.12-22
[25] Kamaraj, K., Dezfouli, B. and Liu, Y., 2019, November. Edge mining on iot devices using anomaly detection. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (pp. 33-40). IEEE



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY