ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

ANOMALY DETECTION OF IoT USING MACHINE LEARNING

K.Shanmugapriya*
Assitant Professor, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.
Mail Id: priya4sp@gmail.com

C.N.Marimuthu
Professor, Department of Electronics and Communication Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.

N.Sridhar
PG Scholar, Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamil Nadu, India.

S.Sameema Begam
Associate Professor, Department of Pharmacognosy, Nandha College of Pharmacy, Erode, Tamil Nadu, India.

Abstract

Internet of things (IoT) is a steady exchange of information between devices in an interconnected arrangement (e.g., splendid home sensors, regular sensors, vehicle and road side sensors, clinical devices, present day robots and perception contraptions). In this manner, gigantic complexity will arise to stay aware of things to future IoT establishments, which subsequently prompts troublesome shortcoming to the structure. An anomaly detection, defined as any adjustment of normal conduct, can give early caution of an issue. By using various machine learning algorithm that to identify assaults during runtime and take less handling time contrasted with different procedures. In this work, the proposed anomaly detection framework is intended to screen IoT vulnerabilities and caution the executive or the service administrations in an organization. The proposed system which uses a K-Nearest Neighbor (KNN) unsupervised machine learning algorithm and Random forest (RF) supervised machine learning algorithm for a fine tuned parameters in the distributed network. Therefore, this system maximizing the models performance without over fitting and implements a fit and a metric score using Cross Validation (CV).

Keywords: Anomaly Detection, Abnormal patterns, Protecting privacy, Machine learning, Cross Validation.

I. INTRODUCTION

Internet of Things is the administration of real things that contain equipment introduced inside the design to pass on and distinguish relationship among each other or in regards to the external environment. In the impending years, IoT-based development will offer advanced degrees of administrations and essentially change the way where people lead their regular routines. Movements in drug, power, quality medicines, agriculture, sharp metropolitan networks, and sharp homes are just a not a great large numbers of the obvious models where IoT is unequivocally settled. North of 9 billion 'Things' (genuine things) are at present connected with the Internet, now. Soon, this number is depended upon to climb to an unbelievable 20 billion. As such, the from one side of the planet to the other choice development going comparably a single key to getting this whole universe to a tiny universally related town, however IoT incorporates just two words which unequivocally depicts its definition. There are two ways of building IoT, First reason: (Real-time data) Indeed, know this as the as a matter of first importance step to start. Constant data is the unforeseen or impromptu data which is to be gathered, handled and to be conveyed quickly right away. Example: In Traffic noticing system, ceaseless information expects a fundamental part. Second reason: (Intelligent activity) if client wish to decrease the human noticing and are by and large enchanted with automating everything to make thing/organization to be a benchmark, then, at that point user can utilize IoT innovation. Think about a model: If the clients are busy with a zenith stressed work and constantly entering home at late evening. To settle this, envision the cooling framework therefore turns on before the individual who has entered the home and shows up.

The Internet of Things has been standing up to various areas like Data Technology, Healthcare, Data Analytics and Agriculture. The central spotlight is on guaranteeing protection as it is the fundamental defense behind various troubles including government investment.

An anomaly, portrayed as any change of ordinary direct, can give early reprobation of an issue. For instance,

abnormalities in an Internet of Things (IoT) sensor's time-series information can show a disappointment in an assembling unit. In any case, identifying irregularities progressively is turning out to be increasingly difficult. In this work, the proposed anomaly detection framework is intended to screen IoT vulnerabilities and caution the executive or the administrations in an organization. The proposed system which uses an K-Nearest Neighbor (KNN) unsupervised machine learning algorithm and Random Forest (RF) supervised machine learning algorithm for a fine tuned parameters in the distributed network. Therefore, this system maximizing the models performance without over fitting and implements a fit and a metric score using Cross Validation (CV).The rest of the paper proceeds as follows. The next section covers the literature survey conducted for the paper. After that, cover our proposed method and module description. Finally, concluding with result analysis on the paper.

II. LITERATURE SURVEY

J. Santos, P. Leroux, T. Wauters, B. Volckaert, and F. D. Turck [1], focuses on the Traditional quirk area approaches don't give off an impression of being legitimate for delay-delicate IoT applications since these strategies district unit broadly wedged by inaction. With the presence of 5G associations and by exploiting the upsides of continuous ideal models, like Network perform Virtualization (NFV) and edge figuring, Software-Defined Networking (SDN), low-latency, versatile quirk area becomes possible. This accomplice quirk acknowledgment objective for extraordinary town applications is offered, that has some mastery in low-power Fog Computing courses of action and evaluated among the degree of Antwerp's town of Things test-bed. An assembled immense dataset, the primary sufficient Low Power Wide space Network (LPWAN) advancements for extraordinary town use case locale unit explored.

Ibrahim Alrashdi, Ali Alqazzaz, Raed Alharthi, Esam Aloufi, Mohamed Zohdy and Hua Ming [2], presented the IoT advanced security risks in a canny city, an Anomaly Detection IoT (AD-IoT) framework is a shrewd irregularity disclosure subject to Random Forest machine learning algorithm. Their proposed plan assured to recognize IoT gadgets at conveyed haze hubs. This procedure utilized present day dataset to address the framework accuracy. The AD-IoT can effectively achieve most critical portrayal accuracy with least bogus positive rate.

Jadel Alsamiri, Khalid Alsubhi [3], surveyed diverse machine learning algorithms that can be used to quickly and reasonably perceive IoT network attacks. A new dataset, Bot-IoT, is utilized to evaluate different recognition algorithms. Some different machine learning algorithms were used, by far most of them achieved prevalent. New components were removed from the Bot-IoT dataset during the execution and differentiated and considers from the writing, and the new highlights gave better outcomes.

Md Mamunur Rashid, Joarder Kamruzzaman, Mohammad Mehedi Hassan, Tasadduq Imam and Steven Gordon [4], presents an anomaly detection strategy subject to machine learning algorithms to make preparations for and ease IoT network security risks in a smart city. Moreover examine gathering methodologies like packing, supporting and stacking to overhaul the show of the acknowledgment system. Exploratory results with the new attack dataset show that the proposed methodology can suitably recognize cyber attacks and the stacking group model outmaneuvers comparable models similar to precision, exactness, audit and F1-Score, inducing the assurance of stacking in this space.

Nanda Kumar Thanigaivelan, Ethiopia Nigussie , Seppo Virtanen, and Jouni Isoaho [5], proposes a crossover inside anomaly detection framework that shares detection errands among switch and nodes. It permits nodes to respond intuitively against the vulnerable node by authorizing impermanent correspondence restriction on it. Every node screens its own neighbors also assuming that abnormal behavior is distinguished, the node hinders the packets of the vulnerable node at interface layer and respond to its parent node. A Distress Propagation Object (DPO) and RPL control message, is formed and utilized for announcing the anomaly furthermore network exercises to the parent node and thusly to the switch. The framework's bogus positive rate examination shows lower bogus positive rate comparable to fundamental detection structure.

Xiali Wang and Xiang Lu [6], proposed a model to couple the Long Short-Term Memory (LSTM) model and the Extreme Gradient Boosting (XGBoost) together for an unusual state assessment on the IoT gadgets. The abnormal behaviors indicate by gathering of system call. The gathered structure cancel progressions are correct by the well known n-gram design, which is utilized for interruption identifications. Then, the stacking model is utilized to recognize surprising practices disguised in call progressions. Final result shows that the stacking model provides amazing execution, strength, and theory limit.

Zhongguo Yang , Irshad Ahmed Abbasi , Elfatih Elmubarak Mustafa, Sikandar Ali [7], uses extractor for consolidation of time series (Tsfresh) and an element based on genetic algorithm are applied to rapidly extricate prevailing aspects which go about as depiction for data plans. Besides, information and diverse useful algorithms were gathered chronicled data. A speedy gathering model dependent on XG-Boost is ready to collect information parts to distinguish reasonable anomaly detection system effectively at run-time. This procedure utilized to pick comfortable help and different game plan reliant upon the stream data instances. Therefore, tests are directed to evaluate the suitability of arrangements shut by genetic algorithm. The resulting analysis shows proposed procedure beats distinctive methodologies and pick help with various circumstances capably.

Milos Savic, Milan Lukic, Dragan Danilovic, Zarko Bodroski, Dragana Bajovic , Ivan Mezei , Dejan Vukobratovic , Srdjan Skrbic and Dusan Jakovetic [8],investigated 5G IoT openness, to work with a anomaly detection as a help to 3GPP versatile cell network planning. This model configuration inserts auto encoder based irregularity acknowledgment models contraptions and adaptable focus association, as such changing between the structure responsiveness and precision. Design, consolidate, show and assess a test-bed that executes the above help true sending incorporated inside the 3GPP Narrow-Band IoT (NBIoT) adaptable director association.

Ruba abu khurma, Heba al harahsheh and Ahmad sharieh[9], presents an Anomaly Detection System with clinical facility inter-network structure to recognize patient prosperity with network interferences. The structure to arrange establishment oversight and e-wellbeing checking to be further develop resources[13]. Therefore, concerning environmental elements variety are more exact. The low torpidity is ensured, an association on the edge to think about a planning close to data sources. Framing an anomaly detection system is finished and overviewed while utilizing Contiki Cooja test system depends upon a sensible informational collection assessment. Experimentation results displays inter-network interferences and information of e-well being patients accuracy.

Gregor Cerar, Halil Yetgin, Blaz Bertalani c [10], presented the exploratory IoT course of action, where the association layer recognized anomalies associates with four far off kinds. The classifiers based on machine learning (ML) and the introduction of edge to recognize subsequently on these irregularities. The overall presentation of three managed and independent systems respectively on both non-encoded and encoded (auto-encoder) feature depictions. Overall, OC-SVM beats the wide scope of different independent ML approaches coming to score for Sudden D, Sudden R, Insta D and Slow D.

III. PROPOSED METHODOLOGY

In this work, the proposed anomaly detection framework is intended to screen IoT vulnarabilities and caution the executive or the service administrations in an organization. The proposed system which uses an K-Nearest Neighbor (KNN) unsupervised machine learning algorithm and Random Forest (RF) supervised machine learning algorithm for a fine tuned parameters in the distributed network. Therefore, this system maximizing the models performance without over fitting and implements a fit and a metric score using Cross Validation (CV).

The contributions of this work are as per the following:

- Most previous experimental works for detection on IoT malicious behaviors using IDS based on signature method based on cloud center to detect only known attack. However, in this paper, anomaly detection using machine learning methods to identify attacks on the dataset including modern attack features that assumed in the IoT botnet network traffics environment [12].
- The proposed system takes the real time attack behaviors as an input.
- This system uses a UNSW-NBl5 dataset to address the issue of the lack of modern normal and attack network traffic.
- Then evaluate the proposed framework system using classification of K-Nearest Neighbor (KNN) and Random Forest (RF) algorithm predict benign or malicious data[11].
- Utilizing Scikit-Learns CV method, characterize a network of hyper-parameter ranges, and randomly sample from the grid, performing K-Fold CV with every combination of values.
- By tuning our machine learning algorithm, maximizing the models performance without over-fitting or creating too high of a variance.

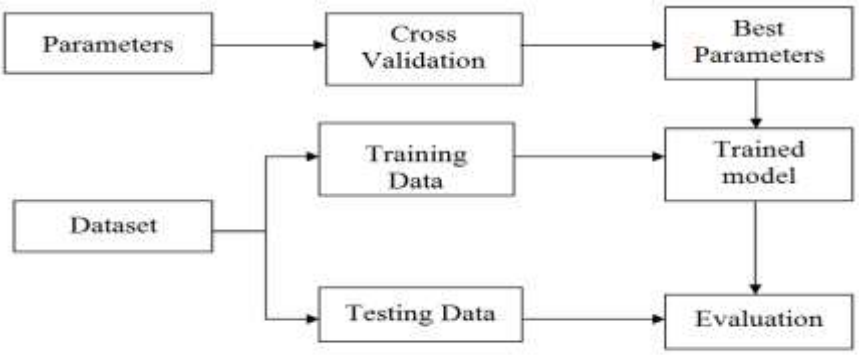

Figure 1 Evaluation of parameter tuning using machine learning algorithm

IV. MODULES AND DESCRIPTION

The framework used two machine learning models in the proposed technique of anomaly detection, which were, K-nearest neighbor and Random Forest.

This module creates a dataset using UNSW-NBl5 takes the real time attack behaviors as an input. The environment contains a testing set (82,332) and a training set (1, 75,341) for the process. This consists of various vulnerabilities: Normal, DoS, Fuzzers, Exploits, Backdoor and Reconnaissance.

In this module, in spite of the fact that KNN is a supervised machine learning algorithm, with regards to anomaly detection it adopts an unsupervised strategy[14. On the grounds that there is no real learning associated with the process and there is no pre-determined labeling of exception or not-anomaly in the dataset, all things being equal, it is completely founded on limit values. Information researchers self-assertively conclude the cutoff values past that all perceptions area units are referred to as anomalies.

Random Forest comprises of many individual decision trees operating together as a cluster. In the Random Forest, every individual tree provides a prediction of a class, and the class receiving the maximum number of votes becomes the prediction class of the model. The hyper-parameters incorporate the quantity of decision trees in the forest and the quantity of elements considered by each tree while parting a hub. The parameters are the factors and limits utilized to part every hub mastered during preparing. Scikit-

Learn carries out a bunch of reasonable default hyper-parameters for all models, however these are not destined to be ideal for an issue.

There is a proverb "Cross Validation is more trustworthy than domain knowledge" in the world of Data Science. The cross validation (CV) utilizes the specific method, K-Fold CV. Make sure to divide our data into a training and a testing set at the point when a machine learning issue approaches. The training set splits into K number of subsets, called folds. At the point of the validation of data, the K^{th} fold assessed every time the data trained K-1 folds iteratively and fits the model for K occasions. The proposed design fitting a model with K = 10. In the model the split is random and in ten approximately equal in size. In first iteration, the nine folds are used for development of trained dataset and test set evaluate on the first fold. Then rehash this process for nine more additional occasions, each time assessing on an alternate fold. Each part became available for one time validation and K-1 time development of trained data. So for total of dataset records and 10-fold, have 82,332 observations for validation and 1,75,341 training data. The benefit of this technique is that all perceptions are utilized for both training and validation, and every perception is utilized for validation precisely once.

V. RESULT ANALYSIS

The experiments were done using the Python Programming Language, and classification was done using the Scikit-Learn's Randomized Search CV method tool. Datasets were collected from the GitHub UNSW-NBl5 dataset [2]. The various vulnerabilities are considered for anomaly detection is Normal, DoS, Fuzzers, Exploits, Backdoor and Reconnaissance.

Despite the fact that KNN is a supervised machine learning algorithm, with regards to anomaly detection it adopts an unsupervised strategy. The experiment was conducted using Scikit-Learn a random split into training and test sets can be quickly registered with the train_test_split aide work. The most straight forward method for utilizing cross-validation is by calling the sklearn.metrics to import the accuracy of score. The sklearn.cluster is called for clustering of unlabeled data to perform. Then load the UNSW-NB15 data set to fit a KNN to reading the training CSV file and the testing CSV file. After that creating odd list of K for KNN with the specific range and the empty list that will hold cross validation scores. Then performing a 10 fold cross validation for number of neighbors and calculating accuracy score for neighbors.

Table 1 Cross validation score for number of neighbors using K-Nearest Neighbor algorithm.

Number of neighbors	Validation Score
1	39.0
3	39.0
5	39.3
7	39.2
9	39.7
11	39.9
13	39.9

The observations made during the experimentation for respective classifiers and the machine learning ensembles to be used in top-performing variants of the proposed method. Table 1 gives such top-performing observations regarding the performance metrics used validation score accuracy.

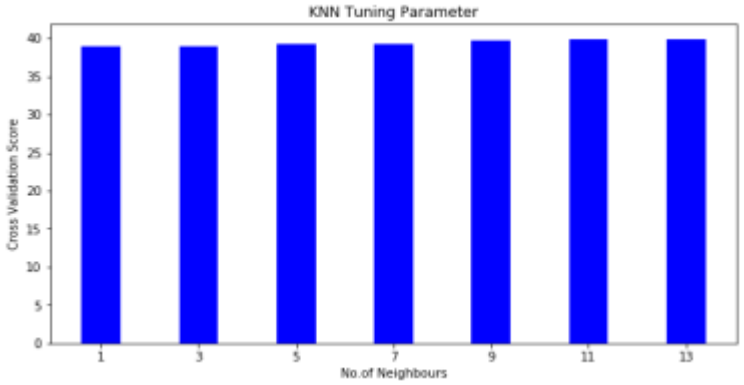

Figure 2 Cross validation score for number of neighbors using K-Nearest Neighbor algorithm.

In figure 2, the bar chart shows the Cross validation score for number of neighbors using K-Nearest Neighbor algorithm. The prediction shows the cross validation for the number of neighbours with a specific range.

Another experiment for supervised machine learning algorithm was conducted to load the UNSW-NB15 data set

to fit a Random Forest to reading the training CSV file and the testing CSV file. After that creating number of estimators list for Random Forest with the specific range and the empty list that will hold cross validation scores. Then performing a 10 fold cross validation and calculating accuracy score for number of estimators.

Table 2 Cross validation score for number of estimators using Random Forest algorithm.

Number of estimators	Validation Score
10	48.5
15	30.7
20	34.3
25	8.2
30	8.5

Another performance metric observation made during the experimentation for respective classifiers and the machine learning ensembles to be used in top-performing variants of the proposed method. Table 2 gives such top-performing observations regarding the performance metrics used validation score accuracy.

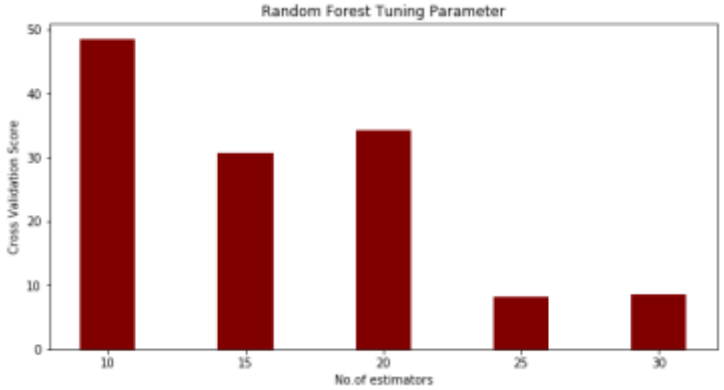

Figure 3 Cross validation score for number of estimators using Random Forest algorithm.

In figure 3, the bar chart shows the Cross validation score for number of estimators using Random Forest algorithm. The prediction shows the cross validation for the number of estimators with a specific range.

By tuning machine learning algorithm, expanding the models execution without over-fitting or making excessively high of a difference. Here, two machine learning models, and ensembles of those using cross validation as ensemble criteria, were used to classify the retrained dataset to detect anomaly. The observed results of the conducted experiments are discussed in this section.

VI. CONCLUSION

This paper proposes a mechanism for the anomaly detection of IoT using datasets of the various attacks. Extraction of various features was done using UNSW-NB l5, considered various vulnerabilities, which were Normal, DoS, Fuzzers, Exploits, Backdoor, Reconnaissance. After extracting these features, KNN for unsupervised and Random Forest for supervised machine learning models, and the ensembles of machine learning models with cross validation as their ensemble types, were trained for attack identification. Then the performance was measured using number of neighbors (KNN) and estimators (RF) validation score and accuracy. From these experimental results, found that the purpose of tuning in machine learning maximizing the models performance and also monitor the nodes without over-fitting using Scikit-Learn's with the accuracy score and was therefore comparably better for that class.

REFERENCES

[1] J. Santos, P. Leroux, T. Wauters, B. Volckaert, and F. D. Turck, "Anomaly detection for smart city applications over 5g low power wide area networks," in NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, 2018, pp. 1-9.

[2] Ibrahim Alrashdi, Ali Alqazzaz, Raed Alharthi, Esam Aloufi, Mohamed Zohdy and Hua Ming, "AD-IoT: Anomaly Detection of IoT Cyberattacks in Smart City Using Machine Learning," in Las Vegas, NV, USA IEEE International Conference on IEEE, 2019.

[3] Jadel Alsamiri1, Khalid Alsubhi2,Faculty of Computing and Information Technology King Abdulaziz University Jeddah, KSA, "Internet of Things Cyber Attacks Detection using Machine Learning," in (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 12, 2019.

[4] Md Mamunur Rashid, Joarder Kamruzzaman, Mohammad Mehedi Hassan,Tasadduq Imam and Steven Gordon , "Cyberattacks Detection in IoT-Based Smart City

Applications Using Machine Learning Techniques,“ in International Journal of Environmental Research and Public Health, 2020.
[5] Nanda Kumar Thanigaivelan, Ethiopia Nigussie , Seppo Virtanen, and Jouni Isoaho Department of Future Technologies, University of Turku, Finland, " Hybrid Internal Anomaly Detection System for IoT: Reactive Nodes with Cross-Layer Operation," in Hindawi Security and Communication Networks Volume 2018.
[6] Xiali Wang and Xiang Lu 1Institute of Information Engineering, CAS, 100093, China Schoolof Cyber Security, UCAS, 100049, China," A Host-Based Anomaly Detection Framework Using XGBoost and LSTM for IoT Devices ," Hindawi Wireless Communications and Mobile Computing Volume 2020.
[7] Zhongguo Yang , Irshad Ahmed Abbasi , Elfatih Elmubarak Mustafa, Sikandar Ali , and Mingzhu Zhang School of Information Science and Technology, Beijing, " An Anomaly Detection Algorithm Selection Service for IoT Stream Data Based on Tsfresh Tool and Genetic Algorithm ," in Hindawi Security and Communication Networks Volume 2021.
[8] Milos Savic, Milan Lukic, Dragan Danilovic, Zarko Bodroski, Dragana Bajovic Member, Ivan Mezei Senior Member, Dejan Vukobratovic Senior Member, Srdjan Skrbic and Dusan Jakovetic Member ," Deep Learning Anomaly Detection for Cellular IoT with Applications in Smart Logistics ," arXiv:2102.08936v2 [cs.NI] 2 Apr 2021.
[9] Abdel Mlak Said, Aymen Yahyaoui and Takoua Abdellatif SERCOM Lab, University of Carthage, Carthage 1054, Tunisia ," Efficient Anomaly Detection for Smart Hospital IoT Systems ," Sensors 2021, 21, 1026.
[10] Gregor Cerar, Halil Yetgin, Blaž Bertalaniˇc, and Carolina Fortuna Department of Communication Systems, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia ," Learning to Detect Anomalous Wireless Links ," in IoT Networks, arXiv:2008.05232v2 [cs.NI] 23 Nov 2020.
[11] Karthikeswaran D., Sudha V.M., Suresh V.M., Javed Sultan A., “A pattern based framework for privacy preservation through association rule mining”, *IEEE-International Conference on Advances in Engineering, Science and Management, ICAESM-2012, Pages 816-821, March 2012.*
[12] Peter K.J., Glory G.G.S., Arumugam S., Nagarajan G., Devi V.V.S., Kannan K.S., “Improving ATM security via face recognition”, ICECT 2011 - 2011 3rd International Conference on Electronics Computer Technology, Pages 373-376, 2011
[13] Deepa A., Marimuthu C.N., “Design of a high speed Vedic multiplier and square architecture based on Yavadunam Sutra”, Sadhana - Academy Proceedings in Engineering Sciences, Volume 44, Issue 9 , September 2019
[14] Deepa S., Marimuthu C.N., Dhanvanthri V., “Enhanced Q-LEACH routing protocol for wireless sensor networks”, ARPN Journal of Engineering and Applied Sciences, Volume 10, Issue 9, Pages 4036-4041, 2015