IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Overcoming Iot Security Challenges Using Machine Learning

[1]Mayank Gindodiya, [2]Sakshi Joshi, [3]Sumit Mali, [3] Tanmay Ahirrao, [3] Ashish Awate

Department of Computer Engineering,

Shri Vile Parle Kelavani Mandal’s Institute of Technology, Dhule, India

Abstract: With the rapid growth and development in Internet of Things(IoT) devices, there is an increase in cyber-attacks targeting these devices. Attackers continue to find new mechanisms and techniques for tricking systems, thereby exploiting the existing IoT-HUB for illegal purposes. Hence detection of attacks in IoT and detecting malicious traffic in early stages is very challenging due to increase in size of network traffic. Also IoT devices have low storage capacity and low processing power, hence traditional security solutions to prevent IoT systems are inappropriate. In this paper, a lightweight framework based on Machine Learning algorithms is proposed for detection of malicious network traffic. The framework uses three classification based ML algorithms, namely K-Nearest Neighbors(KNN), Support Vector Machine(SVM), Random Forest(RF) for detecting attacks.

Index Terms **- IoT, Machine Learning, Network Attacks.**

I. INTRODUCTION

The Internet of things describes physical objects that are embedded with sensors, processing ability, software, and other technologies that connect and exchange data with other devices and systems over the Internet. Iot enables physical devices to see, hear and think based on the shared information without human involvement. Global spending on IOT in 2019 reached US $745 billion, and will surpass US $1 trillion in 2022. As per above prediction the Global spending raised by 24% in 2021, led by investments in IOT software and IOT security. 130 new IOT devices are connected to the internet every second. Currently, there are 30 Billion active IOT devices running over the network. The domains where IoT is integrated include military applications and operations, healthcare, industries, telecommunications, energy productions and distributions, transportation, agriculture, natural and manmade disasters, etc.

As there is wide applicability of IoT, attackers are finding interest in exploiting the IoT network. The most common attack carried out is denial-of-service (DDOS) attack. It is characterized by an explicit attempt to prevent the legitimate use of a service. A distributed denial-of-service attack deploys multiple attacking entities to attain this goal. Hence Machine Learning is the best solution that can be used for the attack detection. All the existing models and proposed methods that suggest using machine learning for attack detection have the ability to detect only a few kinds of known attacks.

Hence we are proposing a Machine Learning architecture that can detect almost all the types of network attacks and also the unknown attacks. The architecture is basically categorized in the 4 main phases i.e. 1) Traffic Capture. 2) Packet Grouping. 3) Feature Extraction, and 4) Binary Classification for attack detection. The architecture also has the ability to identify the malicious node.

There are two main types by which Cyber-analysis can be applied in Machine Learning: Signature(misuse) based and Anomaly based. In signature based analysis, the known attacks are assigned signatures (traffic characteristics), hence this method can detect only the known attacks. The only drawback is it needs frequent update of signatures. On the other hand, the anomaly based method models the network behaviour, and anything abnormal is considered as an attack. Hence it has the ability to detect the unknown attacks as well. Hence we will be combining both of these cyber analysis techniques.

We can summarize our contributions through this research as:

- Improvement in attack detection in IoT networks as the model can detect known as well as unknown attacks.

- Use of Random Forest ensures that the model can handle larger datasets, resulting in a good fit in scalability.

- Generates a lesser number of false alarms

II. LITERATURE SURVEY

We examined and looked over a number of research papers that suggested using Machine Learning for attack detection.

In Paper 1 which is titled as “Security and Privacy in IoT Using Machine Learning and Blockchain” suggested use of both Machine Learning and Blockchain technology for network analysis. To detect malware, the authors suggested that the ML model could be trained by using three types of features including static, dynamic, and hybrid. A detailed analysis of each feature type is done using performance metrics of a dataset, features extraction technique, features selection criteria, accuracy, and detection method. Several detection methods for each feature set were analyzed, but the commonly used were RF, SVM, KNN. And also used triangle-area-based technique to speed up the feature extraction in Multivariate Correlation Analysis (MCA). But this method introduced a level of complexity and compatibility problems.

In paper 2 titled “Internet of Things Cyber Attacks Detection using Machine Learning”, the authors used seven different machine learning algorithms on the Bot-Iot dataset. Bot-Iot dataset was used because of its regular updates, wide attacks diversity, iot generated traffic and ability to generate new features from the raw dataset. Features were extracted using CICFlow Meter which is a network traffic flow generator and produces a visual documentation of features generated and also offers CSV file. Using the Random Forest regressor algorithm, features were extracted from the dataset. The authors applied machine learning algorithms in three different phases:
1. Applying on each attack in the dataset separately.
2. Applying on each attack with a set of features combining the best features in the dataset.
3. Applying on the entire dataset with the seven best features obtained in the feature selection step.

In paper 3 titled as “A Survey on Machine Learning Based Detection on DDoS Attacks for IoT Systems“. Some of the most common malicious attacks are Denial of service (DoS), and Distributed Denial of service (DDoS) attacks, which have been causing major security threats to all networks and specifically to IoT devices have been analysed and detected in this model. In this paper, Rough Set Theory (RST) and Support Vector Machine (SVM) are combined to make use of the excellent accuracy of SVM and better performance of RST which led to data reduction and better classification. For feature extraction authors looked into two classes of features: Stateless and state-full features to see how these features can help in differentiating between normal and anomalous traffic.Stateless and state-full features to see how these features can help in differentiating between normal and anomalous traffic. Stateless features can be extracted from flow-independent characteristics of each packet over the network itself thus is lightweight and realtime. Whereas stateful features are added to improve the accuracy of stateless features. The authors tested five different ML classifiers (KNN), (LSVM), Neural Network (NN), Decision tree (DT) and Random Forest(RF). The authors found that K-nearest neighbors, random forest, and neural net classifiers were the most effective.

In paper 4 titled as “DDOS attack on IoT devices”, Authors have proposed using SDN to monitor internal traffic. With the advent of Software Defined Networking (SDN), it provides a unique opportunity to effectively detect and mitigate DDoS attacks.SDN looks for increasing the number of messages, an increase in packets sent, and harmful entries that are recognized at ports to detect the attacks. When it detects suspicious traffic, it directs the packet to Honeypots(Database) that is isolated from the main server of IoT devices systems. Machine learning then plays a role to measure the size of the package and the amount of traffic, and keeping it in the records for future use in the comparison. Finally, it will prevent the attack using the network-edge preventing application. Though it enables centralized management of networking devices but increases network latency and cost due to SDN.

In paper 5 which is titled as “Attack detection in IOT using Machine Learning”, Authors have proposed a model that traces the web traffic which passes by every fog-to-things node. As fog-to-things connections resemble IoT devices, it will be more efficient to recognize these network attacks at the fog-to-things connections rather than at the cloud layer. Immediate attack detection can inform the network controllers of the IoT devices of those attacks, which will then support them to evaluate and improve their systems. Attacks of types Unauthorized to remote (R2L), Denial-of-Service (DoS), Unauthorized to root super user privileges (U2R attack) were detected by this model. The framework uses three popular classification-based malicious network traffic detection methods, namely Support Vector Machine (SVM), Gradient Boosted Decision Trees (GBDT), and Random Forest (RF).

In paper 6 titled as “Machine Learning-Based IoT-Botnet Attack Detection with Sequential Architecture”, authors mentioned that the complex cryptographic mechanisms cannot be embedded in many IoT devices. The machine learning-based detection can guarantee detection of not only the known attacks and their variances but also the unknown attacks. Authors also adopted a feature selection method to reduce the demand for processing resources for performing the detection system on resource constrained devices. The experiment results indicate that the detection accuracy of our proposed system is high enough to detect the botnet attacks. Moreover, it can support the extension for detecting the new distinct kinds of attacks by using an anomaly based approach. Mirai and Bashlite attacks were the main focus in the paper. Three different ML algorithms, including artificial neural network (ANN), J48 decision tree, and Naïve Bayes were used.

TABLE I : Summary of Algorithm accuracy and suitability

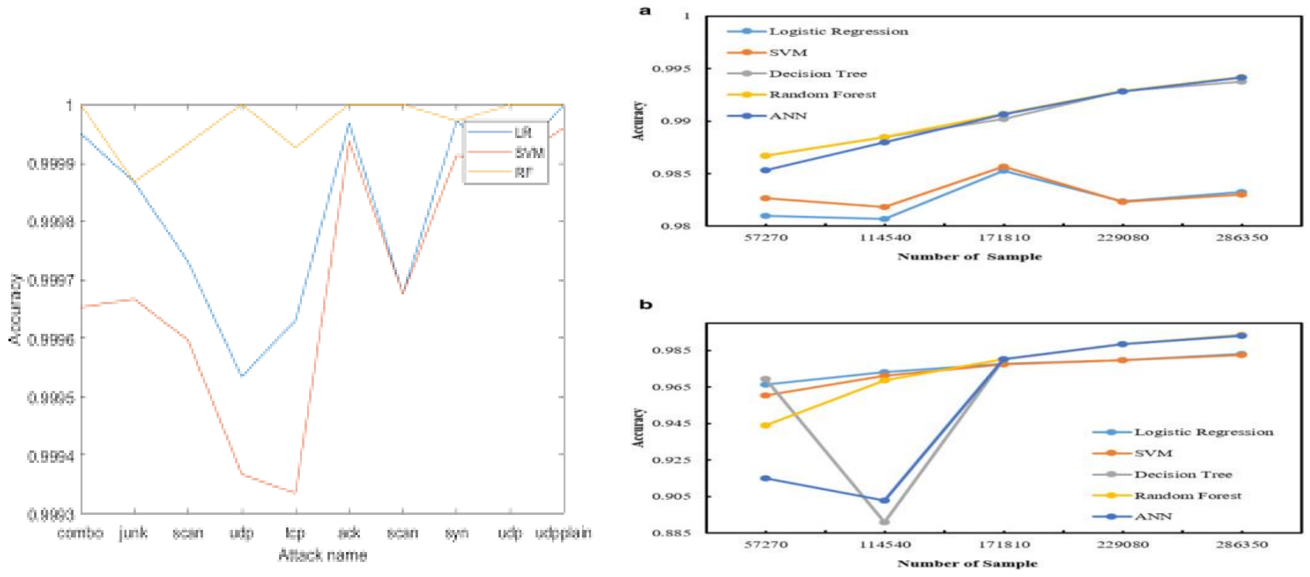
Paper no:	ML algorithm used	Accuracy	Suitability of attacks
1	Logistic regression(LR), Random Forest(RF) Support vector machine(SVM) K- nearest neighbour(KNN)	98.18 98.4 95.8 98.4	Malware Dos flooding Anomaly Malware anomaly
2	Support vector machine(SVM) K-Nearest Neighbours(KNN), Random Forest(RF) Neural network (NN), Decision tree(DT)	.932 .925 .923 .999 .972	Denial of Service(DoS) DDoS DDoS DdoS
3	Random Forest (RF), Support vector machine(SVM), gradient boosting decision tree(GDBT)	85.4 78.4 97.02	Unauthorized to remote(R2L) DoS Port scanning attack(probe)
5	Naïve bayes(NB), Random forest(RF), Multilayer perceptron classifier(MLP), Quadratic discriminant analysis(QDA), K-nearest neighbours(KNN)	0.79 0.97 0.84 0.87 0.99	DDoS_HTTP Data exfiltration Keylogging Service_scan DDoS_TCP OS_scan DDoS_UDP
6	Artificial Neural Network(ANN), J48 Decision Tree Naïve bayes(NB)	99.01 99.01 85.05	UDPflooding TCP flooding ACK SYN flooding Udp plain

As per the above table, it is quite clear that algorithms Random Forest (RF), K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) are the most suitable and efficient ones as far as IoT is concerned. Random Forest can be effectively used for feature extraction whereas K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) can be used for classification purposes.

TABLE II: Summary of Security Accuracy of different Machine Learning Algorithms.

ML Algorithm	Security Accuracy
K-Nearest Neighbor(KNN)	9.99
Support Vector Machine(SVM)	9.86
SVM combine with RF	9.94
Artificial Neural Network(ANN)	9.99
ResNet-18	9.89
Naive Bayes	9.91

FIGURE I: Graph of Security Accuracy of different Machine Learning Algorithms.



III. PROPOSED METHODOLOGY

As per the Literature Survey, it is quite clear that the existing systems that detects network attacks in IoT using Machine Learning are not efficient enough to detect known as well as unknown attacks together. Also as IoT devices are low computing and low storage devices, no other solution apart from Machine Learning can be thought of. Hence there is a need of having a Machine Learning model that can detect both the known and the unknown attacks.
Our proposed architecture will have the following 4 main phases i.e. 1) Traffic Capture. 2) Packet Grouping. 3) Feature Extraction, and 4) Binary Classification for attack detection.

The first step in our proposed architecture will be network analysis and packet collection. The system will be analysing the network traffic in real time and the packets will be captured side by side. Using a network traffic flow generator we can create a dataset of network features in either CSV or JSON format. For the same, We can make use of CICFlow Meter which is a network flow generator and can efficiently extract features from the network traffic. It also produces a visual documentation of the extracted features and offers a CSV or JSON file. Once we have the CSV file, it is very essential to clean and pre-process the data. This will make the structure of the dataset suitable for machine learning algorithms. The irrelevant or missing data can also be filtered so as to increase the accuracy during attack detection.

At this point, the dataset that is obtained contains some features that are of no use for training the model. Hence removing them to increase the accuracy and speed is utmost important. Using the Random Forest Regressor algorithm, we can select features from this dataset. Random forest is efficient and can also decrease the dimensions of the dataset. Hence it also ensures scalability as it can handle large datasets. It is lightweight when compared with other feature selection methods, And is also robust against noise and outliers.

The features can be extracted in two classes: Stateless features and Stateful features. The stateless features will be extracted in real time and are based on the characteristics of each packet which makes it flow-independent. As stateless features do not occupy any memory hence it is also lightweight. On the other hand, stateful features are maintained in the state for improving the accuracy of the model. To keep the model architecture fast and lightweight, we will majorly use the stateless class of feature extraction and will be maintaining only few in state for accuracy improvement.

As our proposed methodology will be capable of detecting known as well as unknown attacks, It is utmost important to have Cyber-Analysis. Cyber Analysis are basically of two types: Signature based cyber analysis and Anomaly based cyber analysis. Signature analysis addresses specific traffic characteristics between the attacks and thus differentiates them. These characteristics are known as signatures. Hence it has an ability to detect all known attacks efficiently based on the analysis of signatures. Signature analysis method is also convenient as it generates less number of false alarms. The drawback of this method is that it needs frequent updation of attack traffic signatures for its efficient use, And it also cannot detect unknown attacks. The second cyber analysis method is Anomaly based analysis. This method examines the network traffic flow and allows passing of normal traffic, and anything abnormal encountered is considered as an attack. Hence it has an ability to detect unknown attacks as well. The use of this method has the possibility of generating false alarms.

Hence as per our use case, We will be combining both the cyber analysis methods as a hybrid analysis technique. This resultant technique will increase the known and unknown attacks detection rate and will also decrease the false alarms rate.
After successful cyber analysis and feature selection steps, we will now train and test the model. As discussed earlier, we will be using the K-Nearest Neighbors algorithm for classification. The reason behind using this classification algorithm is that our literature review reveals that KNN algorithm produces the highest accuracy for attack detection as compared to other classification algorithms. Also, KNN is fast and easy to use and new data can be added to it without affecting its accuracy.

IV. Conclusion

The main aim of Our proposed system is it would address the security challenges for all general known attacks. Improvement in attack detection in IoT networks as the model can detect known attacks with excellent accuracy. It provides an extra security in underlying IOT applications. It also ensures computation over vast datasets resulting a good fit in scalability. It Generates lesser number of false alarms.

References

[1] NAZAR WAHEED, XIANGJIAN HE, MUHAMMAD IKRAM, MUHAMMAD USMAN, SAAD SAJID HASHMI, MUHAMMAD USMAN ‘‘Security and Privacy in IoT Using Machine Learning and Blockchain’’ACM Comput. Surv., Vol. 53, No. 3, Article 1. Publication date: April 2020.
[2] Jadel Alsamiri, Khalid Alsubhi “ Internet of Things Cyber Attacks Detection using Machine Learning” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 12, 2019.
[3] Khadijeh Wehbi, Liang Hong, Tulha Al-salah, and Adeel A Bhutta “A Survey on Machine Learning Based Detection on DDoS Attacks for IoT Systems” 2019 IEEE.
[4] Asmaa Munshi, Nouf Ayadh Alqarni, Nadia Abdullah Almalki “DDOS ATTACK ON IOT DEVICES” 2020 IEEE.
[5] Maryam Anwer, Shariq Mahmood Khan, Muhammad Umer Farooq, Waseemullah “Attack Detection in IoT using Machine Learning” Vol. 11, No. 3, 2021, 7273-7278.
[6] Yan Naung Soe, Yaokai Feng , Paulus Insap Santosa, Rudy Hartanto and Kouichi Sakurai “Machine Learning-Based IoT-Botnet Attack Detection with Sequential Architecture” Sensors 2020, 20, 4372 Published: 5 August 2020.
[7] Mohammad Dawood Momand, Mohabbat Khan Mohsin, Ihsanulhaq ‘’Machine Learning-based Multiple Attack Detection in RPL over IoT’’ 2021 International Conference on Computer Communication and Informatics (ICCCI -2021), Jan. 27-29, 2021.
[8] Fatima Hussain, Rasheed Hussain, Syed Ali Hassan, Ekram Hossain ‘Machine Learning in IoT Security: Current Solutions and Future Challenges’’ IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 22, NO. 3, THIRD QUARTER 2020.
[9] Mohammad Al-Rubaie and J Morris Chang. 2018. Privacy Preserving Machine Learning : Threats and Solutions. IEEE Security and Privacy Magazine (2018).
[10] YairMeidan, Michael Bohadana, AsafShabtai, and Juan Guarnizo, "ProfilIoT: A Machine Learning Approach for IoT Device Identification Based on Network Traffic Analysis", 2017.