Securing IoT Environment using Machine Learning Techniques

Amit Sagu, Nasib Singh Gill

Abstract – One of the most dynamic and invigorate advancement in information technology is advent of Internet of Things (IoT). IoT is territory of interrelated computational and digital devices with intelligence to transfer data. Along with swift expansion of IoT devices through the world security of things is not at expected height. As a consequence of ubiquitous nature of IoT environment most of the user do not have expertise or willingness to secure devices by themselves. Machine learning approach could be very effective to address security challenges in IoT environment. In recent related papers, the researcher have used machine learning techniques, approaches or methods for securing things in IoT environment. This paper attempts to review the related research on machine learning approaches to secure IoT devices.

Keywords - IoT, security challenges, machine learning.

I. INTRODUCTION

IoT is an umbrella term that cover all the devices over the internet which have ability to transfer the data. We specifically concern about those devices which has less or no computation ability, resource constrained, and typically small in size, for instance CCTV camera, sensor etc. There are numerous applications from smart device to smart industries. Smart home (virtual house or home automation) is popular application in which all appliances in home can be controlled and managed by smart devices for instance smart air conditioner can be switched on/off from any place, one can surveillance on their house by CCTV camera from remote location over internet and many more. There is a prediction statement by CISCO which state that there will be over 50 billion connected devices by 2020 [1]. In Health care many devices are being proposed for monitor human health, they detect condition of patient and send collected data. With the help of IoT nodes a city can be facilitate from traffic control to water distribution. Wearable devices, smart grid, industrial internet, connected car, smart farming are few more applications.

Revised Manuscript Received on February 04, 2020.

Amit Sagu, Research Scholar, Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India.
Email: saguamit98@gmail.com

Nasib Singh Gill, Professor, Department of Computer Science & Applications, Maharshi Dayanad University, Rohtak, India.
Email: nasibsgill@gmail.com

A. IoT Components

IoT components primarily include the following:
Sensor- It is physical entity which sense the environment data, e.g - temperature, air speed, humidity, movements.
Actuator – it is responsible of movement in device when it get any control signal. For instance rotate the CCTV Camera in any direction.
Network – IoT objects are tied up with networks by various wireless standards. 802.15 standard are using for wearable device, Zigbee or 802.11 used for home automation. Power efficient network standards have preferred mostly.
User – people control the object via some user interface. User interface application provide facility to people to interact with devices.

II. SECURITY CHALLANGES FOR IoT

The IoT or smart devices market is growing exponentially. Manufacturing companies are not concern about security of IoT nodes, the fact is, it is very difficult to design the node with security in mind due to device properties. They are very vulnerable to any attack. We are witnessing many attack viz. botnet, DDOS (Distributed Denial of Service).
To address the security challenge there are some computation fields, popular one is *Lightweight Cryptography*. Here lightweight mean to those algorithms which have very small footprint or having less computational space. This can be appropriate approach to secure devices and additional research are being there but it is computationally very complex and also very challenging to implement. Another computer field which can be alternate is Machine Learning (ML). We study ML in Artificial Intelligence (AI). Nowadays ML conception is using in security aspect. K. Gurulakshmi et. al used support vector machine (SVM) algorithm to classify the normal and abnormal traffic which can prevent DDOS attack [2]. Ravi kiran et. al used machine learning algorithms viz. random forest (RF), multilayer perceptron, J48 to classify the malware in android devices [3].

III. OVERVIEW OF MACHINE LEARNING

It is a conception to allow machine/computer to learn from example or experiences without using explicitly instruction. It seen as a subset of AI. We use mathematical model based on sample data which is known as training data. The longer any algorithm will run the better it will work. ML concept is not novel, the term first time coined in 1959. ML mainly divided into three category. Supervised and unsupervised are widely used categories. In supervised machine algorithm, training data has input and its corresponding output. Unsupervised machine learning, we do not have any output.

In reinforcement machine learning a software agent automatic take action to maximize the performance or award. In this review paper we are going to discuss mainly supervised machine learning approach.

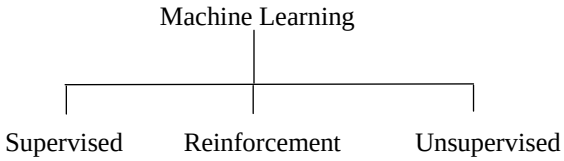

Fig. 1. Machine Learning Type

A. Supervised Machine Learning

Supervised machine learning is task of learning function from input to output. The labeled data also known as training data which has input and output label used for trained the machine, so that machine algorithm can analyze the data and can produce the infers for the new example. Supervised machine learning further divided into classification and regression machine learning. Near 80% of the time classification model is used. Classification machine learning is applicable where output variable is a category for instance identify the gender whether male or female, blood group (A+,B+,O-,B-). Classification machine learning is nothing but pattern recognition. Regression machine learning is adapted where output is some real value like weight, percentage etc.

Following are some wieldy used supervised learning algorithms:

- Support vector machine (SVM)
- K-nearest neighbor (K-nn)
- Naïve bayes
- Decision tree
- Random Forest

The most mentioned ML technique or algorithm is support vector machine. It is one of the state of art method in ML. concept in SVM is to find the hyper-plane which classify the data points. Hyper-plane is a decision boundary between two or more than two data points. Another supervised approach is K-nn, which can be used in both classification and regression.

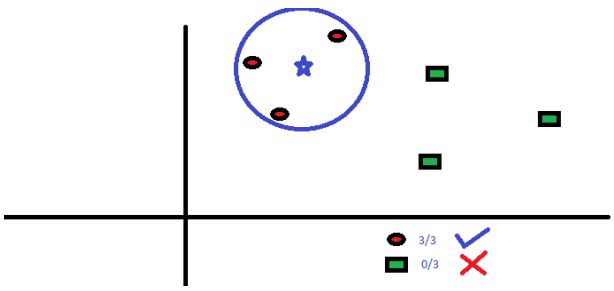

Fig. 2. K-nn

By illustrate fig 2, we can see that there are two classes red circles and blue squares. Our task is to identify the class of blue star. Here K is the nearest neighbor we want to take vote. For corresponding value of K we will draw a circle blue star as center. As we can see closest class is red circles. We infers, blue star is belong to red circles class. The efficiency of this algorithm is depend upon value of K. Naïve Bayes algorithm is based on Bayes theorem which state that a feature is presence in class is independent to other features presents in that class. It can be used for both binary class and multiclass problems. Decision tree is a simple graph with label of attribute on every node and specific edge shows a flow to some answer. Random forest is a collection of many decision tree which is more accurate than single decision tree approach.

B. Unsupervised & Reinforcement Learning

Unlike supervised machine learning we don't need to supervise or trained the model instead we allow the model to work on its own. It is work on unlabeled data and classify the data point basis on their features or pattern which can be useful for categorization

Reinforcement machine learning is about to take appropriate action and get awarded. Since there is no any training data like supervised learning it is bound to learn from their past experience.

Fig. 3. Unsupervised Learning [13]

IV. RESULT AND DISCUSSION

Diverse research has been pursued related to machine learning considering security as the focused outcome.

In [4] Mehdi et. al suggests host based Intrusion Detection System (IDS). Which can be application or device which analyze network traffic and check for abnormality. Framework specially works on smart device deployed in home environment. It investigates events based on signature, anomaly and network packet. Make use of capability of Software Defined Networking (SDN) author make framework bring to close. Sensor element capture the traffic, IDS utilize machine learning algorithm to identify the malicious activity. It is also capable to supress intruder and alert the users. Different features are identified such as number of bytes in packets, connection duration etc. feature selection play key role in machine learning based framework. Author shows their demonstration on smart light bulb named Hue [5], a Phillips brand

Supervised learning method is agreeable where attack pattern are familiar to database. While supervising attack we surely counter the attack pattern which are not familiar to our database, to handle this type of scenario we adapt unsupervised machine learning method. We have seen before that it just categorized/classify the data points without labels. [6] Parez et. al proposed a model which uses supervised and unsupervised machine learning algorithm to make hybrid machine learning techniques for intrusion detection.

Author adapt Neural Network (NN), SVM for supervised and K-mean for unsupervised learning. Hybrid models shows comparatively good result in contrast of single techniques used model. SVM is one of the most adapted classifier used in supervised machine learning.

IoT Sentinel is another form of IDS, the goal of IoT sentinel is to constraint the communication between local and remote device so that adversary would not able to exploit victim device.

Table- I: Mostly used algorithm of machine learning

Sr.	Algorithm Name
1	SVM
2	Decision Tree
3	K-nn
4	K-mean
5	Naïve Bayes
6	NN
7	Random Forest

In [7] Miettinen presents IoT sentinel which is capable to identify the device type when new device commence in environment, evaluation vulnerability of device and make communication vigorous. Here SDN is being used as security gateway where in some cases it can used for monitoring and controlling the network traffic. Author extracted 23 features as device fingerprinting for each packet including IP option, port class, ip address, packet contents. When novel device connect to network SDN evaluate its fingerprint from device behavior and send to IoT security service where with the help of machine learning algorithm isolation level are decided. Here three mark of isolation are circumscribe i.e. strict, restricted and trusted level.

Kunugi et. al[8] focused on diversified attacks which aim IoT devices viz. Distributed Denial of Service(DDOS) and mirai botnet. They also develop a system which detect abnormality by machine learning techniques. Mirai is a type of malicious program that attack IoT related components, it is a Japanese term which mean 'Future'. A bot is program which execute command from the attacker end and attack on other computers. It randomly select the ip address and use default username and password to get access into system and that system also become bot. Mirai was confirmed in 2016 August [8]. In proposed model Random forest classifier is used, which is assemblage of multiple decision tree, it give very efficient result in contrast to decision tree classifier. VizAlert tool is used which is used to visualize the alert about specific node in environment. Author also compare random forest with other technique i.e. SVM and KD tree. RF gives better results in test case.

DDOS attack is one of major threat for IoT environment. In this attack adversary continuously send requests to server resulting crash or make server unresponsive. The intention of this attack in not lucrative to any information but lay down all the IoT nodes like breakdown CCTV camera even for few second might be major security loophole. The DDOS will reach 17 Million by 2020. [9] Roopak et. al propose deep learning model to detect DDOS attack in IoT environment. Deep learning can be employed in supervised, unsupervised or semi-supervised learning. It has complex structure and it take larger time for training and mainly used for image processing, natural language processing etc. Author has implemented four diverse classification model and IDS is stationed in Fog instead of cloud as fog computing are closer to local environment. Also it resolve performance like quality of service and bandwidth. In their model they have used latest DDOS attack CICIDS2017 dataset.

Rohan Doshi et. al [10] also proposed model for DDOS detection for consumer IoT environment. Binary classification model is used to detect threat means either traffic is normal or abnormal (DDOS attack). In feature extraction they used stateless and stateful features. Stateless feature like packet size, protocol and time interval between two packets where stateful feature are bandwidth destination address novelty. Stateless feature are very lightweight since they are not changing over the time and flow, but stateful feature are overhead as they have analyzed over specific period from traffic flow. K-nn , SVM, Decision tree, random forest, Neural Network different model are proposed and tested.

Table- II: Refined Approach

Publish Year	Title of Paper	Tech. used	Refined Approach
2018	Analysis of IoT Bots against DDOS attack using machine learning algorithm	SVM, Knn	work can be extended for prevention method by applying machine learning concepts and preventing the suspicious traffic at the port entry
2018	Towards Machine Learning Based IoT Intrusion Detection Service	Random forest, Neural network	Network traffic feature closer study can provide high accuracy.
2018	Machine Learning DDOS Detection for Consumer Internet of Things Devices	Decision tree, Random forest, SVM, ANN	Their preliminary result can be used in additional research in machine learning.
2019	Machine Learning Based Network Vulnerability Analysis of Industrial Internet of Things	Random forest, SVM, ANN, Knn, Decision tree	Can be focus on utilizing a joint design of multiple algorithms to achieve better performance. The hybrid model should able to provide more accurate results compared to any of the models.

2019	Machine Learning Techniques for Recognized IoT Devices	Decision tree	Another machine learning models can be used to enhance efficiency.
2019	Use of Machine Learning in Detecting Network Security of Edge Computing System	SVM, Edge Computing	The accuracy of SVM cross validation can be enhanced.

V. CONCLUSION

In present paper we engaged in diverse research work in machine learning to secure IoT environment. As IoT network is enlarging exponentially, cyber security loophole is still a menace. Potent work has been done or being done in divergent field, machine learning is one of them which is satisfyingly emerging nowadays. ML is subset of Artificial Intelligence which allow machine/computer to act automate after being programmed. The leading characteristic of machine learning is that it enable computer to learn from data and even improve itself without explicitly programmed. Mainly three type of machine learning we mark i.e. supervised, unsupervised and reinforcement. We have peer in this review paper that mostly supervised learning is applied. When we are not apt to apprehend attack pattern unsupervised may be wholesome. We also discern hybrid approach which assimilate supervised and unsupervised learning for acquisition of better results. As a result compared analysis is done and refined approach are proposed.

REFERENCES

1. D. Evans, "The Internet of Things How the Next Evolution of the Internet is Changing Everything," CISCO, 2011.
2. K.Gurulakshmi and A.Nesarani, "Analysis of IoT Bots against DDOS attack using machine learning algorithm," in 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018), 2018.
3. K. P. S. R. Ravi Kiran, "Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms," in International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), IEEE, 2017.
4. V. a. R. Mehdi, "A Host-based Intrusion Detection and Mitigation Framework for Smart Home IoT using OpenFlow," in 11th International Conference on Availability, Reliability and Security (ARES), 2016.
5. "https://en.wikipedia .org/wiki/Phillips_Hue".
6. P. Deyben, "Intrusion detection in computer networks using hybrid machine learning techniques," in XLIII Latin American Computer Conference (CLEI), IEEE, 2017.
7. T. F. A.-R. S. S. M. N. A. I. H. S. T. Markus Miettinen, "IOT SENTINEL Demo: Automated Device-Type Identification for Security Enforcement in IoT," in IEEE 37th International Conference on Distributed Computing Systems, 2017.
8. H. S. A. K. Yuya Kunugi, "IoT Security Viewer System Using Machine Learning," Springer Nature Switzerland, p. 1071–1081, 2019.
9. G. Y. T. J. C. Monika Roopak, "Deep Learning Models for Cyber Security in IoT Networks," in IEEE 9th Annual omputing and Communication Workshop and Conference (CCWC), 2019.
10. N. A. a. N. F. Rohan Doshi, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," in IEEE Symposium on Security and Privacy Workshops, 2018.
11. B. K. Ankur Lohachab, "Critical Analysis of DDoS—An Emerging Security Threat over IoT Networks," Journal of Communications and Information Networks, Vols. Vol.3, No.3,, 2018.
12. M. A. I. E. A. K. Marwa Mamdouh, "ecuring the Internet of Things and Wireless Sensor Networks via Machine Learning: A Survey," in International Conference on Computer and Applicaitons, IEEE, 2018.
13. "guru99.com," [Online]. Available: https://www.guru99.com/unsupervised-machine-learning.html.

AUTHORS PROFILE

Mr. Amit sagu, has passed M.sc in 2016 in Computer Science and Applications from Department of Computer Science & Applications, Kurukshetra University Kurukshetra, India. He has also worked as Assistant Professor at DAV Centenary College, Faridabad, India. He is currently pursuing Ph.D. in Computer Science at M. D. University, Rohtak. His research interests include IoT, Machine Learning, Big Data Analytics and Data Mining.

Dr. Nasib Singh Gill, is at present senior most Professor of Department of Computer Science & Applications, M. D. University, Rohtak, India and is working in the Department since 1990. He earned his Doctorate in Computer Science in the year 1996 and carried out his Post-Doctoral research at Brunel University, West London during 2001-2002. He is a recipient of Commonwealth Fellowship Award of British Government for the Year 2001. Besides, he also has earned his MBA degree. He has published more than 245 research papers in reputed National & International Journals, Conference Proceedings, Bulletins, Edited Books, and Newspapers. He has authored seven books. He is a Senior Member of IACSIT as well as a fellow of several professional bodies including IETE and CSI. He has been serving as Editorial Board Member, Guest Editor, Reviewer of International/National Journals and a Member of Technical Committee of several International/National Conferences. He has guided so far 9 Ph.D. scholars as well as guiding about 7 more scholars presently in the areas – IoT, Machine Learning, Information and Network Security, Computer Networks, Measurement of Component-based Systems, Complexity of Software Systems, Decision Trees, Component-based Testing, Data mining & Data warehousing, and NLP.