Machine Learning Techniques & Internet of Things Security Issues

Urvashi Rana, Vijay Nandal
M.Tech Scholar, Assistant Professor,
Department of Electronics and Communication,
MRIEM, M D University, Rohtak, Haryana, India.

ABSTRACT: IOT define as Internet of Things where several devices are connected with internet through many technologies and make impossible things possible and gives impact on human lifestyles, IOT work on various types of sensors, layers and algorithm which helps to maintain security and perform well. IOT has become a basic part of today's industrial, agriculture, healthcare, smart city, studies and in manufacturing, transportation rising and with that it include many security on which applications works. In this paper, first section introduction of IOT explained and show how devices connected with IOT platforms, small introduction on devices used for communication purpose. In second section, various security issue, attack was discussed and explained the security requirement which based on many security attack layers like physical, network, software and encryption attack and how to solve them. Third paper, describe how machine learning provide better security in IOT devices with different techniques to improve security issue and attack.

Index Terms: IoT system; Arduino; Machine Learning; K Means algorithm.

I. INTRODUCTION

Iot is Internet of Things where things like devices connected through internet and make life easy, it reduce human effort and make life comfortable. IOT [16]was a network through which transfer data of data was easy .It was one of the most excite innovation, itwas a network of small devices where things like smart devices transfer data. In Internet of Things, things refers to smart devices like smart phones, laptops , cameras, smart watches, smart fitness trackers ,smart lock , smart security system ,automatic vehicles ,smart speakers , connected with internet through gateways and programmed with high languages to provide better security through which transfer of signals or communication process was easy and can communicate with each other without human involvement .IoT[10] was based on various process such as sensing ,networking include many technologies and protocols and standards , identifying . IoT devices help to control and monitor objects remotely.

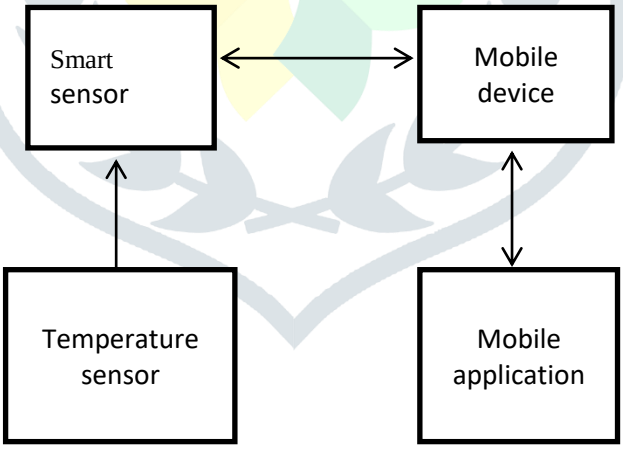


Fig I Smart AC

Above fig shows the example of smart ac which control remotely or by app install in smart phones from any place which can sense the temperature , can control power button on /off from any place and can monitor the features , which was a part of iotwhich works on hardware and software like Arduinouno[4] , Raspberry pi was the hardware and include their modified versions to control and monitor the whole process done by devices , Arduino IDE was the software through which smart devices are programmed to work better.

APPLICATIONS

In human life Internet of Things has many applications which make life easier, safe and smart some applications are as follows:

SMART HOME: Smart Home was most common and important applications and number of people using this applications increasing day by day, it include home appliances connected with internet and operate remotely and used in heating and cooling, in heating such as electric geyser, cooling like air conditioner , security system such as doorbell cam, lightning control, smart TV connected with high technology .

***SMART CITY* :** This application provide better living condition , people living in cities solve the problem like pollution , traffic jam , shortage of energy supplied , smart cities include better transportation facilities, control of traffic jams , environment monitoring , provide security , automated vehicles and provide digital life , better lightning control and have better communication devices .

SMART EDUCATION: Smart education provides smart classrooms, smart learning experience. Mostly schools, college'suses smart education system for better learning such as BYJU'S app which helps to give clear concept and can easily connect to teachers. IOT system provides lots of benefits and connect to world provide safety and security.

There are many more applications such as in industries,healthcare,military, industries, factories, agriculture, manufacturing and many more which helps to make life better and reduce human efforts. To work on these applications IOT works on many different layers which help to connect digitally and easily transfer of signals in wide area and uses security like to store personal information , protect from internal and external attack.

ARCHITECTURE

IOT architecture [5] uses sensors and works on application layer also contain devices that provide stability and security to device users. It consists of layers such as sensors and actuators, devices,gateways, protocols, cloud.Sensors and actuator layers also called perceptron layer which include sensors which senses the physical parameters can be used in face recognition , voice recognition , vision recognition and used in devices like RFID tags ,to collect information from environment and senses the other smart objects in environment , it collect the information like air quality sensors which include sensors to collect information regarding the amount of pollution in air and measure the temperature and speed , automated cars which senses the action like traffic light whether it was red or green .

Where actuators used to monitor and control the system example like heating,cooling, motors . In second stage gateways which provide[5]bridge to transfer information and uses the standards and protocols like LAN, cellular ,Bluetooth[11] ,RFID, Zigbee and many more used for connections and to transfer information to cloud and provide security from data leak and store the data , gateways are the backbone of IOT and all the process of transfer data , receiving data , storing data cloud use to control all the process and stored and perform well and secure by using high languages . In third stage protocols , through which communication is done , by using many protocols and standards such as Bluetooth , thread , z-wave , Wi-Fi and many more protocols used in network layer , application layer , session layer connect with many technologies , topologies connect with each other and communicate easily . In fourth stage cloud where all process was get controlled and monitored by using high languages such as Artificial Intelligence, Machine Learning.

II. SECURITY IN IOT

Security of devices [2] was one of the most serious challenges, security is necessary for preventing from data stealing, attacks. Sometimes, password that we was using for security was weak that attacker can easily guess through which different types of attack can easily occur such as virus, malware. Security was needed for securing personal data , security[8] in IOT devices which connected through internet include routers, webcams, home automation devices which include securities to control and monitor them which was a part of Artificial Intelligence which include Alexa and artificial robot which have securities to handle but can be hack or any physical damage can occur or can cause security issue .In recent time[12], there are large number of IOT devices increasing day by day but with that there are more chances to get attack and can include many risk .

SECURITY IMPACT ON HEALTHCARE

Things in IOT refers to devices used for treatment such as heart monitoring, pacemaker, IOT impact on Healthcare became the main priority which was related to population growth, birth rate , economic and social growth[8]which can be monitor remotely by installing app in phones which provide better security . For treatment use of symptom tracking app which gives regular update of treatment , insulin pumps which was used for store data of taking dose of insulin injections , heart rate monitor , wristband , diabetes care devices which help to improve healthcare management[1] , use of sensors for controlling each functions and can be control remotely which digitally control the whole process which helps to reduce effort but the main disadvantages was that some patient can be untested , unpatched or defective software .

SECURITY IMPACT ON SMART CITIES

Impact on smart cities to make better future , growth of smart cities[16] based on many technologies , protocols , standards , network and many layers which helps to build a new environment and provide better conversation techniques which based on many securities to work them in proper manner , it provide the network through which we can realize the dream of smart cities . While implementing[2] a smart city everything have to plan from lightning to security which make a city looks smart and can easily monitor but this process makes the system long and expensive and easily hack, work on sensors and smart city include smart transportation, smart buildings, environment , public services.

SECURITY ISSUE

In IOT devices[3] when we communicate with each other sometimes the loss of signal can occur which can cause some issue that can be based on security and privacy can cause major damage to devices such as threats,
To measure the security issue level we divide the level of issue in three parts are as follows:

- Low level security issue
- Medium level security issue
- High level security issue

Low Level Security Issue: It was the first level of security issue[3] that occurred in physical layer and data link layer in which jamming of signals that cause by radio frequency signals , low level security issue include insecure initialization , low level spoofing attack, insecure physical interface , sleep deprivation attack which caused by sensors nodes to stay sleepless .

Medium Level Security Issue: Medium level security issue caused in network and transport layers which cause authentication and secure communication, routing attack , sinkhole and wormhole attacks which can be caused by user lack of knowledge which occur by user ignorance and put everyone in risk and can attack on sensitive data which can easily hacked by using weak password which can be cure by end to end security which include several attack such as DOS attack ,to prevent from attack use of secure communication network is important.

High Level Security Issue: High level security issue include cloud based attack[1]which include insecure software, insecure windows, insecure firmware, if we use of out dated software and hardware it will attack on our devices through which devices get fully infected and data lost, regular update of software causes less harm to devices and attacker get difficulty to attack , for preventing from data lost we should have backup if data get lost.

All the security issue causes attack or threats which cause high level of attack and can occur in all security attack layers name as physical attack, software attack, network attack, encryption attack.

PHYSCIAL ATTACK

Examples of Physical Attack are:
1. Object Jamming
2. Cloning Tags
3. RF Interface on RFID
4. Hardware Torjan
5. Malicious Code Injection

SOFTWARE ATTACK

Examples of Software Attacks are:
1. Virus
2. Worms
3. DOS
4. Trojan Horse

NETWORK ATTACK

Examples of Network Attacks are:
1. Phishing Attack
2. Malware Attack
3. DDOS
4. Eavesdropping
5. Hijacking
6. Sinkhole Attack

ENCRYPTION ATTACK

Examples of Encryption Attack are:
1. Side-channel Attack
2. Cipher text Attack
3. MIM Attack

For the security requirement use of strong authentication[19,15] , key management technique , secured operating system , network connection was necessary in which use of strong password cause less number of attack where key management technique has the main concept which reduce power consumption , complexity and increase the level of security , secured operating system has less number of attack which can protect devices from damage and make window run fast and with secured network connection transfer data securely without any interference , use of end to end encryption provide better security .

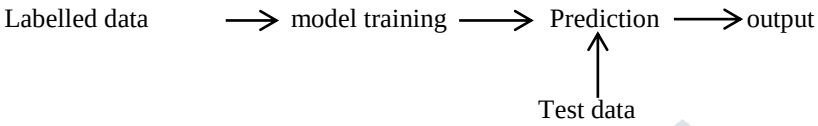
III. MACHINE LEARNING

Machine Learning [1,20]was the science of getting many computers and many programed languages to learn and act like human and improve their learning skills time to timeor by experience, when data was feed by using different algorithm like robots which was a machine fully embedded with electronic components, hardware, software, sensors and can act like human and can do many similar things and can make impossible things to possible. Machine Learning was used in many fields such as medical diagnosis, image recognition, prediction, speech recognition. In speech recognition, was used to convert audio into text form. In Medical Diagnosis, it was used in different devices that used to identify the diseases and monitor the treatment of patient.
Machine Learning consists of 4 types:

1. Supervised Learning
2. Unsupervised Learning
3. Semi-supervised Learning
4. Reinforcement Learning

SUPERVISED LERANING

Supervised Learning consist labelled data in which data was feed in the form of algorithm under the supervisor [1]who gives instructions whether the output was right or wrong, in this learning, model input was considered as training data which done on labelled datasets, in this learning number of interference level was less because it was a trained model and can do each work quickly, in this learning model it has labelled data and learns from its outputand gives feedback as compared to input. In this learning input and output was already mentioned so it takes less time to understand and to train a model.



To explain the figure let's take an example of table which has its own size, length, breadth, colour, shape when training a model under the supervisor and make sure that this values comes in input and feed to a model and model predict that value and match or test from input by its shape, size, length, breadth, colour and then gives the output that it was a table.
It further classified in two parts

- Classification – used for predicting a labelled data in which label was defined
- Regression – used for predicting a quantity values and in which labels was not defined

This learning uses many algorithms such as KNN (k-Nearest Neighbour), Random Forest[1,20,19], Decision Tree, SVM (Support Vector Machine)and many more algorithms which can be used for attack detection and mitigation and can classify the type of attack such as DDOS, algorithm like Naive Bayes which was used for intrusion detection and for malware analysis.

UNSUPERVISED LEARNING

Unsupervised learning has no labelled data, no supervisor, no training in this type of learning model learns from surroundings and find its own output, and learns from its output, it has unlabelled dataset and mostly used in investigation which was mostly used to identify the data structure, in it only input was there and has no output, it takes lot of time and the main goal of this learning was to find the structure and hidden pattern from unlabelled data and has no feedback.
This learning has no labelled data so it forms many data clusters and with that it match the same data and after make different clusters of same labelled data it recognize which type of data was and considered that as a output .

To understand easily let's have an example of different colours of balls like red, blue, green this learning first match the colour and collect different colours of balls in different clusters which was its own prediction values but can be wrong , so in above fig here colours are input data and algorithm were used to learn and to perform task after get knowledge related to input it collect that data in clusters forms which done by manual prediction and get the output and learns from their own results.
Algorithm that mostly used in unsupervised learning was K-means clustering.

K-means Clustering:

K- Means clustering algorithm [18]were mostly used in unsupervised learning, in this learning k value must be known or in integer value and forms clusters of predicted values, this algorithm was so simple in use.
For example: let considered an example of above figure,

Let number of balls = X,
X = 20

Where Red colour balls = 4
Blue colour balls = 6
Green colour balls = 10

So, model recognizes the colour of balls and put aside by matching colour of balls and forms clusters near to previous one, after matching all colours of balls it recognize balls with different colours and learn from its own predicted values or output that it was a balls with its shape and colour and with these balls we can play .

ADVANTAGES

1. It was easy to understand.
2. Its procedure for clustering making was fast.

3. It gives result when data were separated from each other.
4. K value must be known and must be in integer.

DISADVANTAGES

1. Overlapping of data can cause many problems for resolving the data and to find the number of clusters.
2. It used when k value must be known.
3. Interference or noisy data cannot be handle with this type of learning.
4. For non – linear data this algorithm cannot be used.

SEMI – SUPERVISED LEARNING

It was the combination of small labelled data or supervised data to unlabelled data or large amount of unsupervised data which was control by algorithms[14]or it was the combination of supervised or unsupervised machine learning, its cost was high and need human skills to operate, this learning uses algorithm for training labelled and unlabelled data and to train a model it first uses unsupervised learning to form clusters of different data clusters and then use the right data or labelled data to correct the unlabelled data in correct form.
For example: A student in school gaining some knowledge about any topic say maths, teacher teaches some examples of addition, subtraction, division, equations, formulas and many more and student gain that knowledge and store the concept that he learnt when lecture was going to over teacher give homework or some questions to do and a student do all the questions from the knowledge that teacher gives but it can be wrong or right. So, in this example student was a training model, teacher was a supervisor, subject that teacher teaches was labelled data or a supervised data, questions for homework was unlabelled data or unsupervised data and model learns from its own output and with that models try to experiment or gain more knowledge with unlabelled data and compare with labelled data but while training a model whole process was controlled and monitored by algorithms.

Methods that semi supervised provide like

1. ***Self – Training:*** In this method labelled data was trained by using classifier and classify the unlabelled data with their values related to labelled data and trained itself with new data and repeat the process.

2. ***CO – Training:*** It has two methods and each method was unknown to each other and performs their task without any interrupt and they are independent for labelled data.

3. ***Multi view Learning:*** This method uses algorithm such as Decision Tree, Naïve Bayes, SVM and many more algorithms to train a labelled data and but the result of both labelled and unlabelled data was necessary to be same .

RE-INFORCEMENT LEARNING

This learning was based on reward basis [11],[14], in this machine learning model was trained for better performance it get reward and for bad performance it get punishment, this learning was a reward winning learning it has three parts agent, environment, reward where agent was a trained model who learns from environment and give output on the basis that he learn and for better performance it was awarded. This learning was mostly used in gaming like chess where points were given if model reach that level. In it there was no use of supervisor. It consists of two parts on and off policy.

For example: in games there was a target to reach that point like in racing whose goal was to reach at first position and get a prize if a person was in first position, to come in that position model first watch what to do like in racing game human running in same way model start running if model win get a first prize and learn from its own prediction so for better performance it get reward and for bad performance it get punished.
Algorithm used in reinforcement learning was Q – learning but there are many more algorithm based on this learning. Q – Learning algorithm was an off policy algorithm and store data in table form where Q stands for quality which tell how good a model for winning a reward

ADVANTAGES

1. It used for solving very complex messages
2. It was time taken and prefers to take time for better result.
3. It can learn from experience.
4. It collect output from environment, on the basis of that output it perform.

DISADVANTAGES

1. Not use for simple problems.
2. Overloading can occur.
3. Need lots of data to store every step.
4. Cost was expensive.

PARAMETERS	SUPERVISED LEARNING	UNSUPERVISED LEARNING	REINFORCEMENT LEARNING
SUPERVISOR	Need	No Need	No Need
TRAINING	Need	No Need	No Need
TYPE OF DATA	Labelled	Unlabelled	Learns from environment
TYPES	Regression & Classification	Association & Clustering	Reward based
ALGORITHM	Naïve Bayes, Support Vector Machine	K – means Clustering, Neural Network	Q - Learning
DISADVANTAGES	Require large amount of data to train a model	Require less amount of data	It itself creates its own data
APPLICATIONS	Risk giving	Used in detection	Used in automated cars, games
Accuracy of result	Provide High accuracy	Provide less accuracy	Provide less accuracy

TABLE I

Above techniques were also used for security purposes such as intrusion detection [7], virus detection, they increase the level of security which helps to protect from different attacks. These algorithms are used in IOT applications which helps to prevent from data loss, hijacking and to guess the type of attack and can prevent devices from attacks.

IV. CONCLUSION

IOT in our daily life helps to build a new innovation and make us advance and create a latest technology for better performance and to provide better security to devices, for that we had discussed many techniques related to machine learning which a part of artificial intelligence and control the whole process and use highly programmed languages for better performance. In above figure shows the comparison of different machine learning techniques which tell that no one was perfect but in some areas such as for prediction Naïve Bayes algorithm were best for prediction which was a supervised algorithm and can also be used for malware analysis, for better cluster dataset K – means clustering was better which can also be used for intrusion detection, Q – Learning which was used for award winning and can be used for authentication, SVM which was used for attack detection which works on labelled or unlabelled data so all algorithm can be used in their specific task by using different high languages such as R language, python which helps to create a new or advanced level of security.

V. REFERENCES

[1] FatimaHussain , RasheedHussain , Syed Ali Hassan&EkramHossain "Machine Learning in IOT security :Current Solutions and Future Challenges," 2019 .

[2] VikasHassija, VinayChamola, VikasSaxena, Divyanshjain, PranavGoyal and BiplabSikdar " A Survey on IOT Security," 2019 .

[3]MirzaAbdurRazzaq, Muhammad Ali Qureshi , SajidHabib Gill, SaleemUllah" Security Issue in IOT," 2017 .

[4]KuldeepSingh Kaswan, Santar Pal Singh, ShrddhaSagar " Role of Arduinouno in Real world Applications," 2020

[5]Tara Salman " Networking Protocols & Standards for IOT ,"2017 .

[6]AntinoF . S Karemeta , Jose L . Hernandez – Ramos, M . Victoria Moreno " A decentralized approach for security and privacy challenges in IOT , "2017

[7]Anshuman Dash, Satyajit Pal , ChinmayHegde " Ransomware Auto – detection in IOT security by using ML," 2019.

[8] DajiQiaoBaridNait – Abdesselam , Ryan Gerdes , TIE Quu " Special Issue on security and privacy ,"2019.

[9]SandipKundu" Security and privacy of ML algorithm ," 2018.

[10] R Gurwanth, MohitAgarwal , Abhrajeet Nandi , DebarataSamanta , " An Overview : Security Issue in IOT network , "2012

[11]Liang Xiao, XiaoyueeLoav , Xiaozhu Lu , Yanyoung Zhang , Di – Wi " IOT Security Technique Based on ML, "2019

[12] Michael Negnevitsky , Artificial Intelligence " A guide to Intelligent systems," 2019.

[13]A. Monesia and N.K Jha "A Comprehensive study of security of Internet of Things ," IEEE Transactions on Emerging Topics in computing vol . 5 , pp.586602, 2017.

[14] K.A. da Costa ,J.P.Papa, C. O. Lisboa , R.Munoz and V.H.C . de Albuquerque " IOT : A survey on Machine Learning based intrusion detection technique approaches ," 2019 .

[15] Leo Louis " Working Principle of Arduino and using it as a tool for study and research ," 2016 .

[16]Hany F. Atlan and Gary B . Wills " IOT Security , privacy , safety , ethics, "2018 .

[17] JafarTanha" Semi-supervised self-training for decision tree classifiers,"2015.

[18] A.L.Buczak and E. Guven" A survey of data mining and machine learning method for security intrusion detection ,"2016 .

[19] l . Jiang and C. Li " Deep feature weighting for naviebayes and its applications to text classification," 2016.

[20] J.Caedo and A.Skjellum " Using machine learning to secure iot systems, "2016.