A Project Report

On

Adversarial Machine Learning in IoT

BY

Saksham Gupta

2018A7PS0281H

Under the supervision of

G Geetha Kumari

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF CS F376: DESIGN PROJECT

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)

HYDERABAD CAMPUS

(October 2020)

ACKNOWLEDGMENTS

The Design Project opportunity I had is a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual as I was provided with an opportunity to be a part of it. I am also grateful for having a chance to interact with so many wonderful people and professionals who led me through this project.

I am using this opportunity to express my deepest gratitude and special thanks to G Geetha Kumari of HOD of CSE Department who in spite of being extraordinarily busy with her duties in this uncertain time of pandemic, took time out to hear, guide and keep me on the correct path and allowing me to carry out my project .

I express my deepest thanks to Rashmi Sahay, Teaching Assistant for the Project who helped me in making useful decisions & giving necessary advice and guidance and arranged all facilities to make life easier. I choose this moment to acknowledge her contribution gratefully.

I perceive this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain desired career objectives. Hope to continue cooperation with all of you in the future.

Sincerely,
Saksham Gupta

Birla Institute of Technology and Science-Pilani,
Hyderabad Campus

Certificate

This is to certify that the project report entitled "**Adversarial Machine Learning in IoT**" submitted by Mr. Saksham Gupta (ID No. 2018A7PS0281H) for fulfillment of the requirements of the course CS F376, Design Project Course, embodies the work done by him under my supervision and guidance.

Date: 1, December 2020

(G Geetha Kumari)
BITS- Pilani, Hyderabad Campus

ABSTRACT

As the Internet of Things (IoT) emerges and expands over the next several years, the security risks in IoT will increase. There will be bigger rewards for successful IoT breaches and hence greater incentive and motivation for attackers to find new and novel ways to compromise IoT systems. Traditional methods and techniques for protecting against cyber threats in the traditional internet will prove inadequate in protecting against the unique security vulnerabilities that would be expected in the internet of things. Machine learning has been successfully applied in diverse areas such as image processing, natural language processing (NLP), and autonomous driving.Internet of Things (IoT) systems rely on various detection, classification, and prediction tasks, to learn from and adapt to the underlying spectrum environment characterized by heterogeneous devices, different priorities, channel, interference, and traffic effects. Machine learning has emerged as a powerful tool to perform these tasks in an automated way by learning from the complex patterns of the underlying wireless communication dynamics.Adversarial machine learning has emerged as a field to systematically analyze the security implications of machine learning in the presence of adversaries.

As the topic of the project Adversarial Machine Learning in IoT it is important to first understand the involvement of Machine Learning in Iot and the adversaries associated , also it is necessary to understand the the work done in this field therefore many research papers on the same were explored .In this report an introduction to the topic and literature survey of the twenty two papers are given.

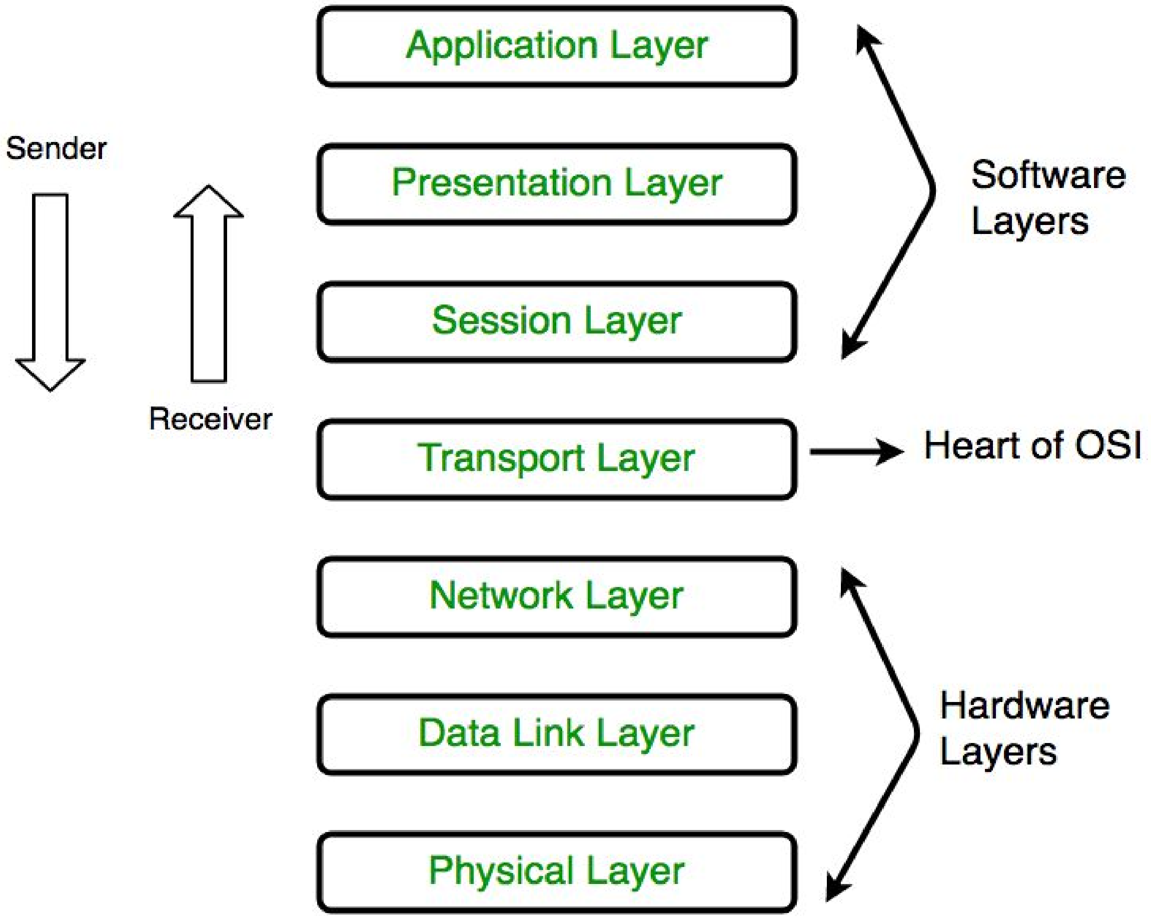
In this work we present MQTTset, a dataset focused on the MQTT protocol, widely adopted in IoT networks. We present the creation of the dataset, also validating it through the definition of a hypothetical detection system, by combining the legitimate dataset with cyber-attacks against the MQTT network. Obtained results demonstrate how MQTTset can be used to train machine learning models to implement detection systems able to protect IoT contexts. Then we show how an adversary of poisonous attack affect the accuracy of our model.

CONTENTS

1.Adversarial Machine Learning in IoT

IoT Technologies are regarded as the third industrial revolution and are defined as “the interconnection, via the Internet, of computing devices embedded in everyday objects, enabling them to send and receive data”. These devices now constitute a major portion of sectors such as personal computing, tourism, transportation, healthcare and smart cities. IoT devices have seen arise of 200% from 2 billion in 2006 to 200 billion in 2020.These devices collect and store temporal information which needs to be kept secured but just like most consumer technologies they are not designed keeping security in mind, which is a barrier in the emerging IoT network and services.This neglect in the security of the data of the IoT users can disrupt the ecosystem.

We need to have a close look on the IoT network layers in order to understand the intrusion and its detection.OSI is a 7 layered architecture on which IoT is based.Open System Interconnection model developed by ISO is a 7 layer set of architecture which defines the way any data is transmitted between two devices.



The network layer is the main layer of the architecture and also more prone to security threats, therefore this is most suited for deploying the defence mechanisms like Network Intrusion Detection Systems (NIDS).

The intrusion attacks on the IoT networks can be mainly described into four main categories:

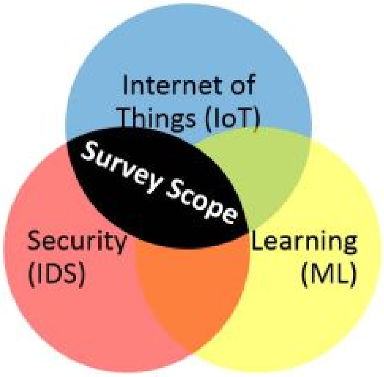
1. Probe:in this the attacker scans the network and collects the information on target.
2. DoS:(denial of service) user is denied access to a service by attackers.
3. U2R:attacker attempts to escalate a limited user's privilege to a super user.
4. R2L:attacker gains remote access to a victim machine existing local users.

User-to-Root and Remote-to-Local attacks are difficult to detect as they mimic the user's normal behaviour.

Intrusion detection are also categorised in two broad categories depending on the technique used:

- Signature based detection
- Anomaly based detection

Signature based intrusion detection system has a predefined set of activities patterns which are considered malicious and it catches any such signatures.It is however not efficient for unknown or polymorphic attacks.
Anomaly based intrusion detection system detects intrusions based on the deviations from the normal behaviour. They can easily detect unknown attacks unlike the signature based detection. Therefore it is not practical to depend on the pre defined attack patterns which are limited.
The researchers have therefore explored the fields of machine learning and artificial intelligence with greater emphasis on deep learning to deal with the intrusions and improve the security systems, and found to have improved techniques for fraud detection ,text classification and image recognition. These ML algorithms can detect the abnormalities in the IoT with ease and have significantly improved the security.In this report we have done a survey of 22 research papers discussing the improvement of IoT adversary using Machine Learning Algorithms.



2. LITERATURE SURVEY

Kunle.Ibitoye et.al(2019)[1] in their recent work on Intrusion Detection in IOT have evaluated the performance of two Deep Learning Techniques to detect and mitigate security threats against IOT. Feed-forward Neural Networks (FNN) is widely used to classify the intrusion attacks in IOT , they have compared it with Self-normalizing Neural Networks(SNN). The dataset used by them was the BoT-IoT dataset from the Cyber Range Lab of the center of UNSW Canberra Cyber.In results, FNN performs better than SNN for intrusion detection in IoT networks on the basis of multiple performance metrics such as accuracy, precision, and recall as well as multi-classification metrics such as Cohen Cappas score. However, when it was tested for adversarial robustness, SNN demonstrated better resilience against samples from the dataset, thus presenting a promising future in the quest for safer and more secure deep learning in IoT networks.

Matteo Terzi et.al(2019)[2] have talked about Adversarial Training as a procedure aiming to provide models which are robust to worst-case perturbations around predefined points.One major issue is that robustness w.r.t gradient based attackers is achieved at the cost of prediction accuracy. In their paper they have discussed a new algorithm, Wasserstein Projected Gradient Descent (WPGD), for adversarial training. It provides an easier way to achieve cost-sensitive robustness, thus giving a finer control of the robustness-accuracy trade-off. Their proposed WPGD algorithm is validated in their work on image recognition tasks with different benchmark datasets. Moreover, many real world-like datasets are unbalanced in their paper they show that when dealing with such type of datasets, the performance of training is mainly affected in term of standard accuracy.In this paper they identified the accuracy gap in adversarial training is a result of the loss of fine-grained classification capabilities in neural networks.

Yalin E. Sagduyu et.al(2019)[3] have briefly discussed the security issues of machine learning,and how its implications haven’t been considered for wireless applications as those in IoT systems which are susceptible to attacks due to the open ,broadcast nature of wireless communication. They have presented new techniques built on adversarial machine learning and applied them to three types of over the air wireless attacks, denial of service (DoS) attack, spectrum poisoning attack, and priority violation attack. They observed how the adversary starts with an exploratory attack in order to infer the channel access algorithm of an IoT transmitter by building a deep neural network classifier that predicts the transmission outcomes. Based on the prediction result, these wireless attacks continue to jam data transmission or manipulate sensing results over the air to fool the transmitter into making wrong transmit decisions in the test phase. Their results demonstrate new ways to attack the IoT systems using adversarial machine learning and designed a defensive approach based on the Stackelberg game and showed its effectiveness in improving the transmitter’s performance.

Ahmed Abusnaina et.al(2019)[4] published a research paper ‘Adversarial Learning Attacks on Graph-based IoT Malware Detection Systems’ where they discussed malware detection in IOT using control flow graph based features and deep learning networks. They investigated the robustness of these models against adversarial learning. Two approaches were designed to craft adversarial IoT software: off the-shelf methods and

Graph Embedding and Augmentation (GEA) method. In off-the-shelf method, they examined eight different adversarial learning methods to force the model to misclassify.The GEA approach aimed to preserve the functionality and practicality of the generated sample through a careful embedding of a benign sample to a malicious one. Intensive experiments were conducted to evaluate the performance of the proposed method, showing that off-the-shelf adversarial attack methods were able to achieve a misclassification rate of 100%. In addition, it was observed that the GEA approach is able to misclassify all IoT malware samples as benign. The findings of their work highlighted the essential need for more robust detection tools against adversarial learning, including features that are not easy to manipulate, unlike CFG based features. The implications of their study were quite broad, since the approach challenged in this work is widely used for other applications using graphs.

Nathalie Baracaldo et.al(2018)[5] in their recent work talks about how machine learning plays an increasingly important role in the Internet of Things, both in the Cloud and at the Edge, using trained models for applications from factory automation upto environmental sensing. They also discussed the security challenges in using machine learning in IoT environments. In particular, the training data can be manipulated by tampering with sensors' measurements. These types of attack are known as poisoning attacks and they significantly decrease the overall performance and cause targeted misclassification or bad behavior, and insert "backdoors" and "neural trojans".Taking Advantage of recently developed tamper-free provenance frameworks, they presented a methodology that used contextual information about the origin and transforms data points in the training set to easily identify poisonous data. Their approach works with or without a trusted test data set. Using their proposed approach poisoning attacks can be effectively detected and mitigated in IoT environments with reliable provenance information.

Elike Hodo et.al(2016)[6] in their paper "Threat analysis of IoT networks using artificial neural network intrusion detection system." have started by describing how Internet of things is still in its infancy and has attracted many industrial sectors including medical fields, logistics tracking,automobiles and smart cities. They described how it is susceptible to a range of intrusion threats. Their paper includes a threat analysis of the Internet of Things and uses Artificial Neural Network (ANN) to deal with such threats. A multi-level perceptron which is a type of supervised Artificial Neural Network, was trained using internet packet traces and is then assessed on its ability to thwart Distributed Denial of Service (DDoS/DoS) attacks. Their paper focuses on the classification of normal threat patterns on an IoT Network. They validated ANN procedure against a simulated IoT network and their experimental results demonstrate 99.4% accuracy and can successfully detect various DDoS/DoS attacks.

Nadia Chaabouni et.al(2019)[7] They discussed how the 2016 Dyn cyberattack exposed critical fault-lines among smart networks. Security of IoT has become a critical concern. The danger exposed by infested Internet-connected Things not only affects the safety of IoT but also threatens the entire Internet ecosystem which may possibly exploit the vulnerable Things (smart devices) deployed as botnets. Mirai malware compromised the video surveillance devices and paralyzed Internet via distributed denial of service attacks. Within the recent past, security attack vectors have evolved both ways, in terms of complexity and variety . Hence, to spot and stop or detect novel attacks, it's important to research techniques in IoT context. Their survey classifies the IoT security threats and challenges for IoT networks by evaluating existing defense techniques. Their main focus is on network intrusion detection systems (NIDSs); hence, the paper reviews existing NIDS implementation tools and datasets also as free and open-source network sniffing software. Then, it surveys, analyzes, and compares state-of-the-art NIDS proposals within the IoT context in terms of architecture, detection methodologies, validation strategies, treated threats, and algorithm deployments. The review deals with both traditional and machine learning (ML) NIDS techniques and discusses future directions. The survey provides a comprehensive review of NIDSs deploying different aspects of learning techniques for IoT, unlike other top surveys targeting the normal systems.Their paper is useful for academia and industry research, first, to spot IoT threats and challenges, second, to implement their own NIDS and eventually to propose new smart techniques in IoT context considering IoT limitations. Moreover, the survey will enable security individuals to differentiate IoT NIDS from traditional ones.

Jiaqi Li et.al(2019)[8] have discussed how Software defined Internet of Things (SD-IoT) networks take advantage of centralized management and interactive resource sharing, which reinforces the efficiency and scalability of Internet of Things applications. But with the rapid climb in services and applications, they're susceptible to possible attacks and face severe security challenges. Intrusion detection has been widely wont to ensure network security, but classical detection methods are usually signature-based or explicit-behavior-based and fail to detect unknown attacks intelligently, which makes it hard to satisfy the wants of SD-IoT networks. In this paper, they propose a man-made intelligence-based two-stage intrusion detection empowered by software defined technology. It flexibly captures network flows with a worldwide view and detects attacks intelligently. They first leverage the Bat algorithm with swarm division and binary differential mutation to pick typical features. Then, they exploit Random Forest through adaptively altering the weights of samples using the weighted voting mechanism to classify flows. Evaluation results prove that the modified intelligent algorithms select more important features and achieve superior performance in flow classification. it's also verified that our solution shows better accuracy with lower overhead compared with existing solutions.

HAMED HADDAD PAJOUH et.al(2016)[9] described the increasing reliance on Internet of Things (IoT) devices and services, and how the potential to detect intrusions and malicious activities within IoT networks becomes critical for resilience of the network infrastructure.In this paper, they present a totally unique model for intrusion detection supported two-layer dimension reduction and two-tier classification module, designed to detect malicious activities such as Remote to Local attacks and User to Root. They proposed a model using component analysis and linear discriminate analysis of dimension reduction module to spate the high dimensional dataset to a lower one with lesser features. Then they apply a two-tier classification module utilizing Naive Bayes and Certainty Factor version of K-Nearest Neighbor to identify suspicious behaviors. The experiment results using NSL-KDD dataset shows that their model outperforms previous models designed to detect U2R and R2L attacks.

Xiaolei Liu et.al(2019)[10] discuss how many Internet of Things systems run Android systems. With the continual development of machine learning algorithms, the learning-based Android malware detection system for IoT devices has gradually increased. However, these learning-based detection models are often susceptible to adversarial samples. They focus on the need of an automatic testing framework to assist these learning-based malware detection systems for IoT devices perform security analysis. The present methods of generating adversarial samples mostly require training parameters of models and most of the methods are aimed toward image data. To unravel this problem, they propose a testing framework for learning-based Android malware detection systems for IoT Devices. The key challenge is the way to construct an appropriate fitness function to get an efficient adversarial sample without affecting the features of the appliance . By introducing genetic algorithms and a few technical improvements, their test framework can generate adversarial samples for the IoT Android application with a hit rate of nearly 100% and may perform black-box testing on the system.

Zhengping Luo et.al(2020)[11] emergence of Internet of Things as the next logical stage of the web , it's become imperative to know the vulnerabilities of the IoT systems when supporting diverse applications. They have discussed the security implications of machine learning following an adversarial machine learning approach. In this paper, they proposed an adversarial machine learning based partial-model attack in the data fusion/aggregation process of IoT by only controlling a small part of the sensing devices. The numerical results demonstrate the feasibility of this attack to disrupt the choice making in data fusion with limited control of IoT devices, e.g., the attack success rate reaches 83% when the adversary tampers with 8 out of 20 devices. These results show that the machine learning engine of IoT system is very susceptible to attacks even when the adversary manipulates a little portion of IoT devices, and the outcome of those attacks severely disrupts IoT system operations.

Baracaldo et.al(2017)[12] Their predictions are involved in making decisions about healthcare, security, investments and lots of other critical applications. Given this pervasiveness, it's not surprising that adversaries have an incentive to manipulate machine learning models to their advantage. One way of manipulating a model is through a poisoning attack during which the adversary feeds carefully crafted poisonous data points into the training set. Taking advantage of recently developed tamper-free provenance frameworks, they presented a strategy that uses contextual information about the origin and transformation of knowledge points within the training set to spot poisonous data, thereby enabling online and frequently re-trained machine learning applications to consume data sources in potentially adversarial environments. To the simplest of our knowledge and it is often the primary approach to include provenance information as a part of a filtering algorithm to detect causative attacks.They presented two variations of the methodology–one tailored to partially trusted data sets and therefore the other to completely untrusted data sets. Finally, they evaluated their methodology against existing methods to detect poison data and found an improvement within the detection rate.

Nikhil Banerjee et.al(2018)[13] discussed how the advancement of Internet-of-Things devices with various sorts of sensors enables us to collect diverse information with Mobile Crowd-Sensing applications.Such a highly dynamic setting entails us to gather ubiquitous data traces, originating from sensors carried by people, introducing new information security challenges.What's needed in these settings is that the timely analysis of those large datasets to produce accurate insights on the correctness of user reports. They first modeled the cyber trustworthiness of MCS reports within the presence of colluding and intelligent adversaries. Then they rigorously assessed, using real IoT datasets, the effectiveness and accuracy of well-known processing algorithms when employed towards IoT security and privacy. By taking into consideration these spatiotemporal changes of the underlying phenomena, it was demonstrated how the concept drifts can masquerade the existence of attackers and their impact on the accuracy of both the classification and clustering processes. Their initial set of results clearly show that these unsupervised learning algorithms are in danger of adversarial infection, thus, magnifying the requirement

for further research within the sphere by leveraging a combination of advanced machine learning models and mathematical optimization techniques.
Ling Huang et.al(2011)[14] in "Adversarial machine learning” says how adversarial machine learning is similar to the effects of bringing “Byzantine” to to machine learning.In the paper they have shown how the machine learning methods like PCA and SpamBayes network anomaly detection are vulnerable to attacks .They also provided outline details on models for adversary capabilities, class limitations, the knowledge of their learning system algorithm ,data and feature space.They discussed countermeasures and defenses when an adversary attacks a learning system.Exploratory attacks on a learning system were examined and the results show that it was easy for an adversary search classifiers to find an input that is not classified as negative.Lastly the approaches and challenges for differential policy,causative privacy attacks and privacy preserving were explored.

Matthew Jagielski et.al(2018)[15] have done a systematic study of poisonous attacks and how it can be countered for linear regression models in their research paper "Manipulating machine learning: Poisoning attacks and countermeasures for regression learning.".The training data is influenced by attackers in order to manipulate the prediction results.They have proposed an optimization in poisonous attacks which requires very less information on training data.They also proposed a approach to design a robust algorithm which performs much better than the existing regression models for defending against such attacks.Their proposed defense algorithm and the attack strategy was evaluated extensively on various datasets like real estate, loan assessment and health care.They also provided a case study on health application depicting the implications of poisonous attacks and their further research is oriented to develop more secure algorithms to prevent such poisonous attacks.

Jinxin Liu et.al(2020)[16] began with describing how the security vulnerabilities in IoT devices acts as a roadblock.IoT systems have a heterogeneous nature which provides common benchmarks such as NSL KDD dataset from being used for testing and verification of the performance of different Network Intrusion Detection Systems.They have tried to bridge this gap by examining the specific attacks in the NSL KDD dataset that has a probability to impact sensor nodes in IoT network.They studied 11 machine learning algorithms to detect the attacks and reported the results, where we can see how the tree based algorithms outperform the 11 machine learning algorithms studied. 97% accuracy was seenin the supervised algorithm XGBoost, 90% in Matthews correlation coefficient and 99.6% in Area Under the Curve.It was also found that the unsupervised algorithm of Expectation Maximization performs better than the Naive Bayes classifier by 22% on detection of attacks.

Rajshekhar Das et.al(2018)[17] "A deep learning approach to IoT authentication." - low powered devices having a wireless radio attached to them compose a major portion of the IoT.The challenge for the wireless security protocols and advanced cryptographic is the secure authentication of these devices among adversaries with much higher power and computational capabilities.It is a case where chunks of signals from a low powered device is replayed by the high power software radio.A deep learning based classifier is proposed in the paper which learns the hardware imperfections of low powered devices which are challenging to emulate.LSTM framework was proposed which is sensitive to signal imperfections over longer durations.Experimental results from a testbed of 30 low-power nodes demonstrates high resilience to advanced software radio adversaries.

Zhengping Luo et.al(2020)[18] starts by discussing how IoT has become the next logical state of the internet and how important it is to deal with the vulnerabilities in IoT systems. Machine learning is used in many IoT systems hence it is necessary to study it in adversarial machine learning approach.In this paper they have proposed an adversarial machine learning method based on the partial model attack for data aggregation by controlling only a small part of the devices. Their numerical results show that the attack is feasible to disrupt the decision making with limited control of IoT devices.Success rate of attack was found to be 83% when 8 out os 20 IoT devices were tampered by the adversary, showing that the machine learning engine of IoT is vulnerable to attacks even when the adversary only manipulates a small portion of the IoT devices.

Ahmed Abusnaina et.al(2019)[19]-"Examining adversarial learning against graph-based iot malware detection systems." main goal of the study was to validate how robust are the graph based deep learning algorithms which are used for IoT malware detection against Adversarial Learning. They generated the CFGs of the IoT sample, and extracted 23 representative features from the CFGs to train the deep learning model.They have crafted two approach to adversarial IoT - first one being Off the Shelf Adversarial Attack methods and Graph Embedding and Augmentation.The GEA method preserves the practicality and functionality of adversarial sample via a careful embedding of a malicious sample vs a benign one.They found that Off the Shelf Adversarial Attack methods achieve a misclassification rate of 100% whereas the GEA method misclassify all the samples as benign.

Mahbub Khoda et.al(2019)[20] industrial IoT makes use of edge devices which are intermediary between actuators and sensors and cloud services.In such edge devices malware attacks are frequent and machine learning is used to thwart them, but these models are still vulnerable to the adversarial attacks. Small perturbations are introduced to the malware samples in order to mislead the classifier and all are classified as benign applications.Existing models work by randomly selecting the adversarial samples which thus degrades the classifier performance.They proposed two approach to retrain a classifier one being the distance from malware cluster center and other one on a probability measure derived from a kernel-based learning.In the results both the above methods outperforms the random selection method by accuracy of 6%.Their study provides path to robust security systems.

Joseph Clements et.al(2019)[21] speaks about the advancements in the field of artificial intelligence and the growing need for better defensive measures in the domain of network security and how deep learning methodologies are used in network intrusion detection systems.These algorithms allowed us to tackle conventional network attacks efficiently and enabled us to deploy practical security systems.However it is seen after adversarial machine learning that deep learning models are highly vulnerable to modification in data, this makes the deep learning modes susceptible to the weakest entry point for attack.In the paper it is shown even by modifying input features as low as 1.38 an adversary generates inputs to mislead a deep learning network intrusion detection systems and its important to check performance via a perspective of adversarial machine learning domain also along with conventional network security.

Abdur Rahman et.al(2020)[22] medical IoT devices are heavily used in healthcare sector and especially in pandemic time of COVID-19. These devices produce huge amounts of data which is further used to identify COVID-19 phenomena using deep learning techniques.CT scanning, thermal cameras and X-ray images are some of the IoT devices.Adversarial perturbation poses security a threat to the deep learning algorithms used. In the paper are examined the various methodologies based on deep learning algorithms of COVID-19 with adversarial examples.The results show that these models are vulnerable to adversarial attacks via perturbations.In detail examples are given on the adversarial generation process and how these attacks cause perturbations of deep learning formulated COVID-19 diagnosis.Their work explains the adversarial attacks and lays foundation to safeguard deep learning based healthcare systems.

3. Project Details

- We will start by introducing the MQTT Protocol (Message Queuing Telemetry Protocol) which was implemented in Cooja Simulator.

- How we extracted the information from Cooja into .pcap and its conversion to csv.
- What algorithms were implemented for Intrusion Detection System.
- How an Adversary affects our IDS.

MQTT Protocol

- MQTT stands for Message Queuing Telemetry Protocol , which is a simple messaging protocol designed for constrained devices with low-bandwidth.

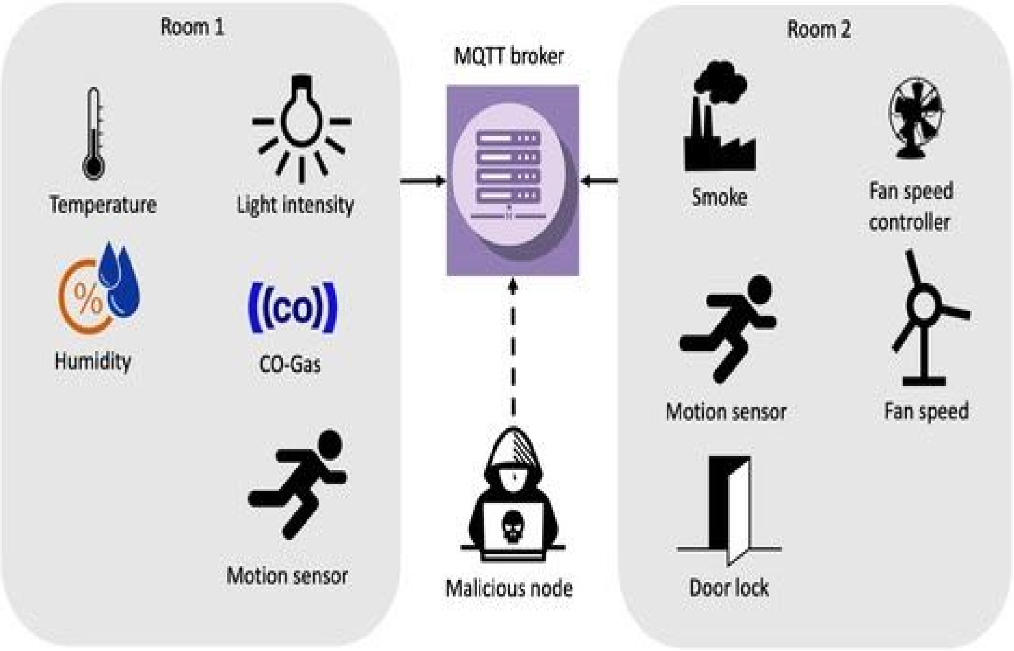
- The protocol uses a publish/subscribe communication pattern which makes it easy to send and commands to control outputs, read and publish data from sensor nodes and much more.

- The MQTT protocol is a good choice for wireless networks that experience varying levels of latency due to occasional bandwidth constraints or unreliable connections.

- The protocol is used for machine to machine communication and plays an important role in the Internet of Things (IoT).

Overview

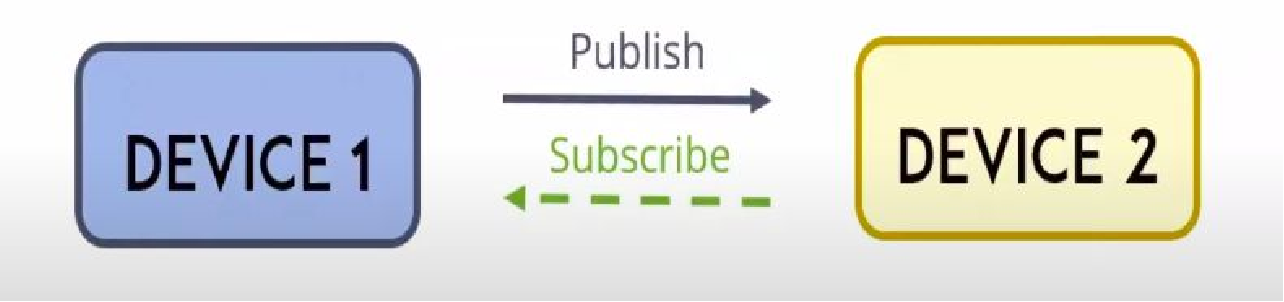
- The MQTT protocol surrounds two subjects : a client and a broker.
- An MQTT broker is a server while the clients are the connected devices.
- When a device or client sends data to a server or broker, it is called publish. When the operation is reversed it is called subscribe.



MQTT Concepts

- Publish / Subscribe

  In a publish subscribe system , a device can publish a message on a topic or it can be subscribed to a particular topic to receive messages.



- For example Device 1 publishes on a topic.
- Device 2 is subscribed to the same topic as device 1 is publishing in.
- So, device 2 receives the messages.

- MQTT Messages- these are the information that you want to exchange between your devices. Whether it's a command or data.
- MQTT Topics- topics are the way to register interest for incoming messages or how you specify where you want to publish the message.
- MQTT Broker= primary responsibility of broker is for receiving all messages , filtering the messages ,decides who is interested in

them and then publishing the messages to all subscribed clients.

- An MQTT session is divided into four stages:
    - Connection
    - Authentication
    - Communication
    - Termination

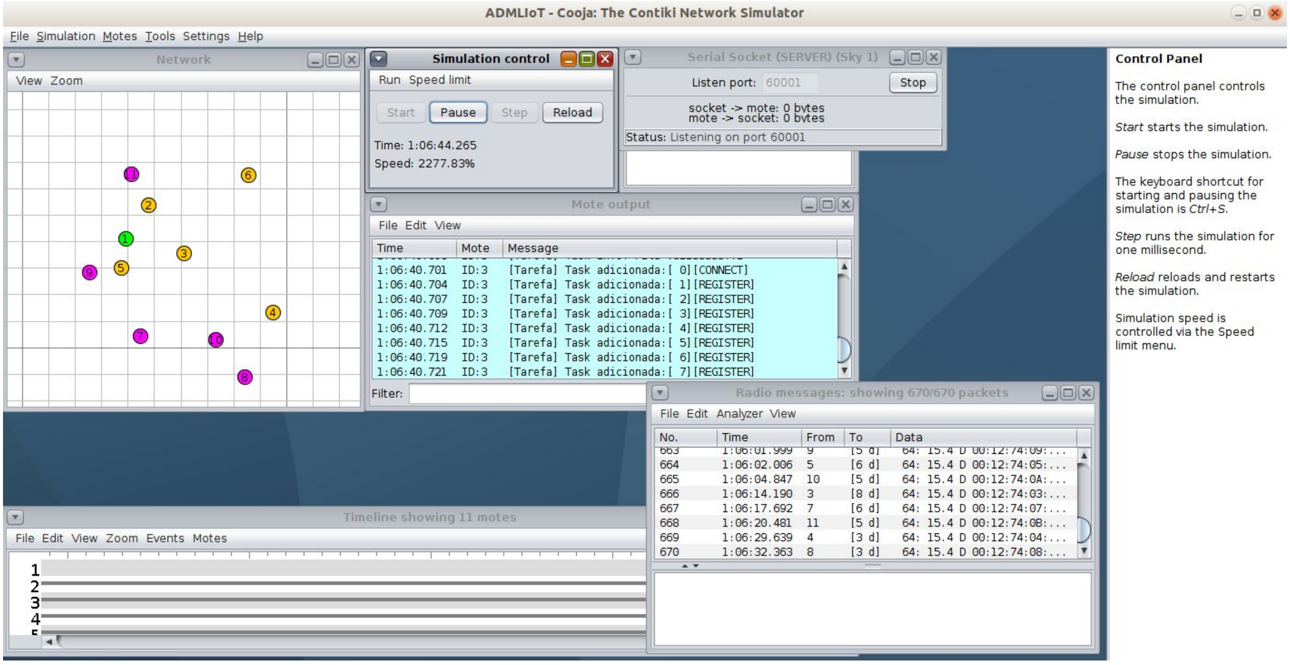
- <u>**Connection-**</u> A client starts by creating a TCP/IP connection to the broker by using either a standard port or custom port defined by the broker's operators.
- <u>**Authentication-**</u> During the SSL/TLS handshake , the client validates the server certificate and authenticates the server. The client may also provide a client certificate to the broker during the handshake.The broker can use this to authenticate the client.
- <u>**Communication-**</u> During the communication phase a client can perform publish, subscribe, unsubscribe and ping operations. The publish operation sends a binary block os data --the content-- to a topic that is defined by the publisher . Topic subscriptions are made using a SUBSCRIBE/SUBACK packet pair, and unsubscribing is similarly performed using an UNSUBSCRIBE/UNSUBACK packet pair.
- <u>**Termination-**</u> When a publisher or subscriber wants to terminate an MQTT session, it sends a DISCONNECT message to the broker and then closes the connection. THis is called a graceful shutdown because it gives the client the ability to easily re-connect by providing its client identity and resuming where it left off.

Our Setup

- We have generated data by real-life equivalent simulations using the open source Contiki/Cooja simulator , because of lack of availability of public IoT attack data sets.

- Our Simulation contains 11 nodes, of which 5 are publisher nodes , 5 subscriber nodes and 1 server node i.e border-router node.

- The Cooja simulation generates raw packet capture (PCAP) files, which are first converted into Comma Separated Values (CSV) files using the *tshark* which involves the feature extraction while converting to csv.


ADMLIoT - Cooja: The Contiki Network Simulator
File Simulation Motes Tools Settings Help
Network
View Zoom
1
2
3
4
5
6
7
8
9
10
11
Simulation control
Run Speed limit
Start
Pause
Step
Reload
Serial Socket (SERVER) (Sky 1)
Listen port: 60001
Stop
Mote output
File Edit View
Filter:
Radio messages: showing 670/670 packets
File Edit Analyzer View
Timeline showing 11 motes
File Edit View Zoom Events Motes

Time: 1:06:44.265
Speed: 2277.83%

socket -> mote: 0 bytes
mote -> socket: 0 bytes
Status: Listening on port 60001

```
Time          Mote   Message
1:06:40.701   ID:3   [Tarefa] Task adicionada:[ 0][CONNECT]
1:06:40.704   ID:3   [Tarefa] Task adicionada:[ 1][REGISTER]
1:06:40.707   ID:3   [Tarefa] Task adicionada:[ 2][REGISTER]
1:06:40.709   ID:3   [Tarefa] Task adicionada:[ 3][REGISTER]
1:06:40.712   ID:3   [Tarefa] Task adicionada:[ 4][REGISTER]
1:06:40.715   ID:3   [Tarefa] Task adicionada:[ 5][REGISTER]
1:06:40.719   ID:3   [Tarefa] Task adicionada:[ 6][REGISTER]
1:06:40.721   ID:3   [Tarefa] Task adicionada:[ 7][REGISTER]
```

No.	Time	From	To	Data
663	1:06:01.999	9	[5 d]	64: 15.4 D 00:12:74:09:...
664	1:06:02.006	5	[6 d]	64: 15.4 D 00:12:74:05:...
665	1:06:04.847	10	[5 d]	64: 15.4 D 00:12:74:0A:...
666	1:06:14.190	3	[8 d]	64: 15.4 D 00:12:74:03:...
667	1:06:17.692	7	[6 d]	64: 15.4 D 00:12:74:07:...
668	1:06:20.481	11	[5 d]	64: 15.4 D 00:12:74:0B:...
669	1:06:29.639	4	[3 d]	64: 15.4 D 00:12:74:04:...
670	1:06:32.363	8	[3 d]	64: 15.4 D 00:12:74:08:...

Control Panel

The control panel controls the simulation.

Start starts the simulation.

Pause stops the simulation.

The keyboard shortcut for starting and pausing the simulation is *Ctrl+S*.

Step runs the simulation for one millisecond.

Reload reloads and restarts the simulation.

Simulation speed is controlled via the Speed limit menu.

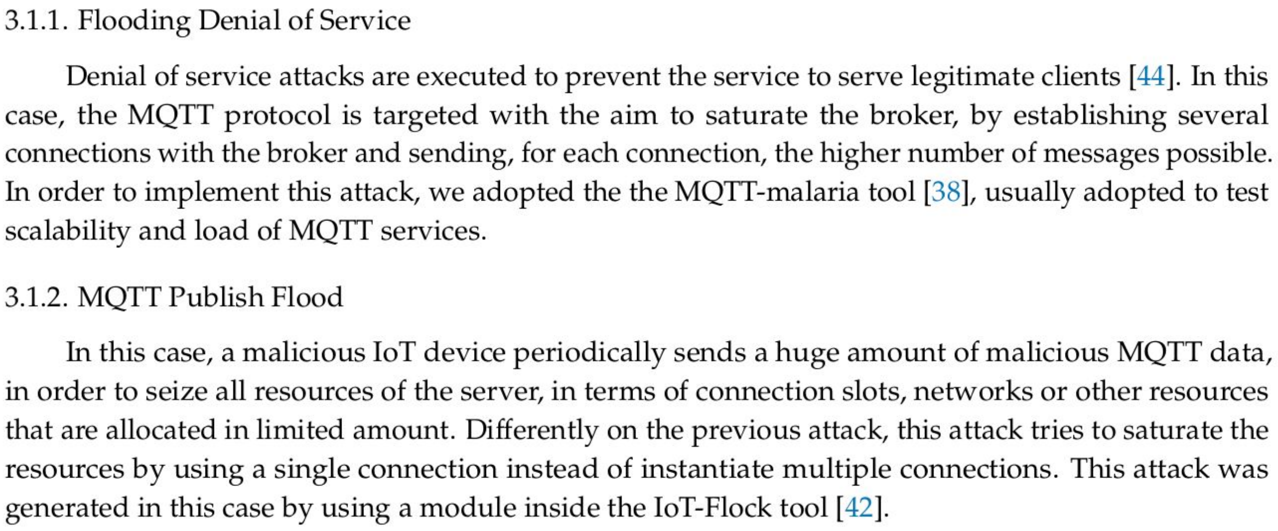
Implementation

- Previously we simulated 11 nodes that was a legitimate data without any network intrusion or any attack. The .pcap files generated were converted to csv with the help of tshark and features were extracted into csv.

- Next we simulated the same network with a malicious node causing the flood attack for network intrusion and similarly the data obtained was converted to csv.

- A number of other attacks data was also collected namely:
  - Brute-force attack
  - Malaria attack
  - Malformed attack
  - Slowite attack

Details of these attacks are given below:

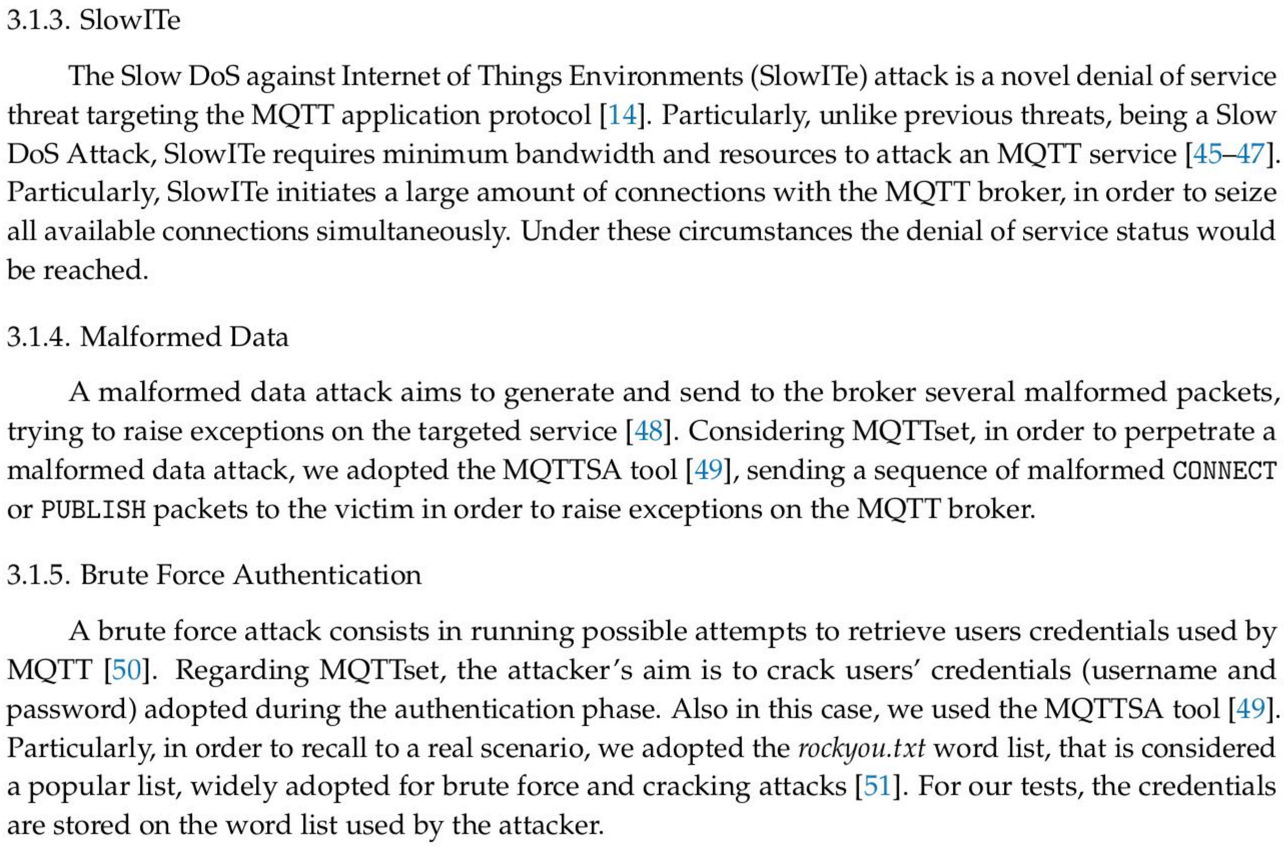


3.1.1. Flooding Denial of Service

Denial of service attacks are executed to prevent the service to serve legitimate clients [44]. In this case, the MQTT protocol is targeted with the aim to saturate the broker, by establishing several connections with the broker and sending, for each connection, the higher number of messages possible. In order to implement this attack, we adopted the the MQTT-malaria tool [38], usually adopted to test scalability and load of MQTT services.

3.1.2. MQTT Publish Flood

In this case, a malicious IoT device periodically sends a huge amount of malicious MQTT data, in order to seize all resources of the server, in terms of connection slots, networks or other resources that are allocated in limited amount. Differently on the previous attack, this attack tries to saturate the resources by using a single connection instead of instantiate multiple connections. This attack was generated in this case by using a module inside the IoT-Flock tool [42].



3.1.3. SlowITe

The Slow DoS against Internet of Things Environments (SlowITe) attack is a novel denial of service threat targeting the MQTT application protocol [14]. Particularly, unlike previous threats, being a Slow DoS Attack, SlowITe requires minimum bandwidth and resources to attack an MQTT service [45–47]. Particularly, SlowITe initiates a large amount of connections with the MQTT broker, in order to seize all available connections simultaneously. Under these circumstances the denial of service status would be reached.

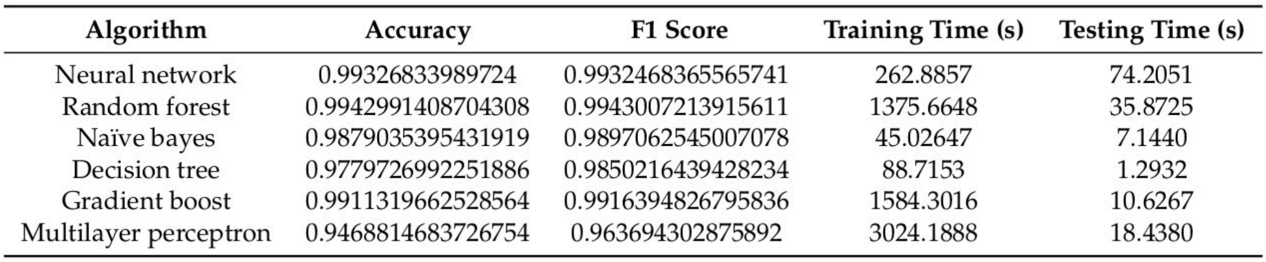
3.1.4. Malformed Data

A malformed data attack aims to generate and send to the broker several malformed packets, trying to raise exceptions on the targeted service [48]. Considering MQTTset, in order to perpetrate a malformed data attack, we adopted the MQTTSA tool [49], sending a sequence of malformed `CONNECT` or `PUBLISH` packets to the victim in order to raise exceptions on the MQTT broker.

3.1.5. Brute Force Authentication

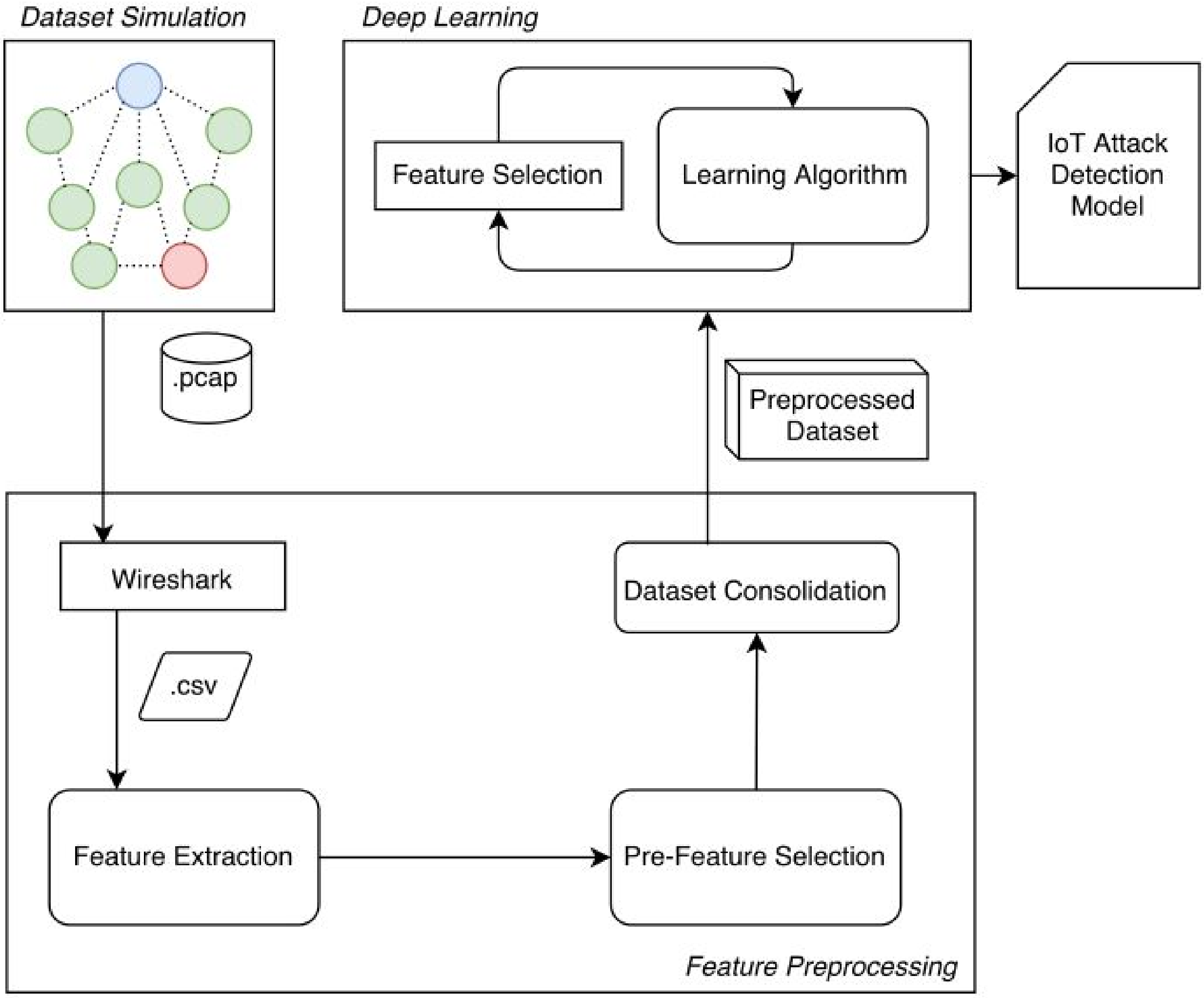
A brute force attack consists in running possible attempts to retrieve users credentials used by MQTT [50]. Regarding MQTTset, the attacker's aim is to crack users' credentials (username and password) adopted during the authentication phase. Also in this case, we used the MQTTSA tool [49]. Particularly, in order to recall to a real scenario, we adopted the *rockyou.txt* word list, that is considered a popular list, widely adopted for brute force and cracking attacks [51]. For our tests, the credentials are stored on the word list used by the attacker.

- As of now we had 6 csv data files one of which was legitimate and others were attack based.
- Next we added a attribute to each csv indicating its type i.e legitimate or attack based.
- All the csv were then combined to form a mqttdataset.csv and preprocessed where features were extracted as too many unnecessary features leads to overfitting.
- This was then split into training and test dataset in a 7:3 ratio .
- As a part of network intrusion detection we developed few machine learning (neural network, random forest, naive biased, multi level perceptron, decision tree and gradient boost) models to predict the intrusion attack correctly and compared their accuracy scores and graphs were generated for the same.



Algorithm	Accuracy	F1 Score	Training Time (s)	Testing Time (s)
Neural network	0.99326833989724	0.9932468365565741	262.8857	74.2051
Random forest	0.9942991408704308	0.9943007213915611	1375.6648	35.8725
Naïve bayes	0.9879035395431919	0.9897062545007078	45.02647	7.1440
Decision tree	0.9779726992251886	0.9850216439428234	88.7153	1.2932
Gradient boost	0.9911319662528564	0.9916394826795836	1584.3016	10.6267
Multilayer perceptron	0.9468814683726754	0.963694302875892	3024.1888	18.4380

In the Intrusion detection system that we developed we found most algorithms gave an accuracy of 0.99 however Random Forest happens to have performed the best.



Implementation- Adversarial Machine Learning

- We implemented the Poisoning attack on our Machine Learning models.

- Poisoning is adversarial contamination of training data.

- Machine learning systems can be retrained using data collected during operations.

- An attacker may poison this data by injecting malicious samples during operation that subsequently disrupt training.

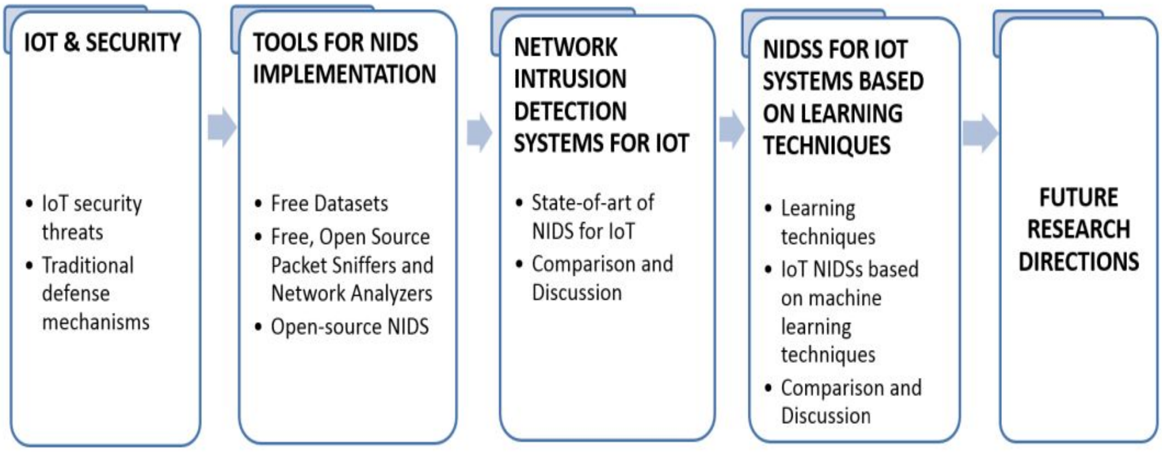
- In our Intrusion detection Machine Learning model we re-trained our model with such manipulated data and observed the accuracy of prediction to fall.

3. CONCLUSION

IoT has benefited humans in many ways and it has been evolving ever since.Iot devices are used in every domain, even the critical ones like healthcare and military.Thus it becomes highly important that these devices are secure to collect our data and there is no intrusion. Unfortunately the technologies are being developed without keeping the security perspective in mind , even one single node that is vulnerable can bring down the whole network.

NIDSs are a powerful mechanism to curb such intrusions and make systems secure.In the paper surveyed we studied many algorithms and compared their efficiencies with one another.The domain of NIDS is still open to future research and much better methodologies are yet to be followed.

In the picture given below is the workflow of our survey.


IOT & SECURITY
• IoT security threats
• Traditional defense mechanisms
TOOLS FOR NIDS IMPLEMENTATION
• Free Datasets
• Free, Open Source Packet Sniffers and Network Analyzers
• Open-source NIDS
NETWORK INTRUSION DETECTION SYSTEMS FOR IOT
• State-of-art of NIDS for IoT
• Comparison and Discussion
NIDSS FOR IOT SYSTEMS BASED ON LEARNING TECHNIQUES
• Learning techniques
• IoT NIDSs based on machine learning techniques
• Comparison and Discussion
FUTURE RESEARCH DIRECTIONS

Machine learning Algorithms are widely used today in IoT applications but as we saw in the Intrusion Detection System that we developed how an adversary can lead to false prediction of results which can cause serious damage on an Industrial level.

4. REFERENCES

[1] Ibitoye, Olakunle, Omair Shafiq, and Ashraf Matrawy. "Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks." *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019.

[2] Terzi, Matteo, Gian Antonio Susto, and Pratik Chaudhari. "Directional adversarial training for cost sensitive deep learning classification applications." *Engineering Applications of Artificial Intelligence* 91 (2020): 103550.

[3] Sagduyu, Yalin E., Yi Shi, and Tugba Erpek. "IoT network security from the perspective of adversarial deep learning." *2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2019.

[4] Abusnaina, Ahmed, et al. "Adversarial learning attacks on graph-based IoT malware detection systems." *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019.

[5] Baracaldo, Nathalie, et al. "Detecting poisoning attacks on machine learning in iot environments." *2018 IEEE International Congress on Internet of Things (ICIOT)*. IEEE, 2018.

[6] Hodo, Elike, et al. "Threat analysis of IoT networks using artificial neural network intrusion detection system." *2016 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE, 2016.

[7] Chaabouni, Nadia, et al. "Network intrusion detection for IoT security based on learning techniques." *IEEE Communications Surveys & Tutorials* 21.3 (2019): 2671-2701.

[8] Li, Jiaqi, et al. "Ai-based two-stage intrusion detection for software defined iot networks." *IEEE Internet of Things Journal* 6.2 (2018): 2093-2102.

[9] Pajouh, Hamed Haddad, et al. "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks." *IEEE Transactions on Emerging Topics in Computing* (2016).

[10] Liu, Xiaolei, et al. "Adversarial samples on android malware detection systems for IoT systems." *Sensors* 19.4 (2019): 974.

[11] Luo, Zhengping, et al. "Adversarial machine learning based partial-model attack in IoT." *Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning*. 2020.

[12] Baracaldo, Nathalie, et al. "Mitigating poisoning attacks on machine learning models: A data provenance based approach." *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*. 2017.

[13] Banerjee, Nikhil, et al. "Unsupervised learning for trustworthy iot." *2018 IEEE international conference on fuzzy systems (FUZZ-IEEE)*. IEEE, 2018.

[14] Huang, Ling, et al. "Adversarial machine learning." *Proceedings of the 4th ACM workshop on Security and artificial intelligence*. 2011.

[15] Jagielski, Matthew, et al. "Manipulating machine learning: Poisoning attacks and countermeasures for regression learning." *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018.

[16] Liu, Jinxin, Burak Kantarci, and Carlisle Adams. "Machine learning-driven intrusion detection for Contiki-NG-based IoT networks exposed to NSL-KDD dataset." *Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning*. 2020.

[17] Das, Rajshekhar, et al. "A deep learning approach to IoT authentication." *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018.

[18] Luo, Zhengping, et al. "Adversarial machine learning based partial-model attack in IoT." *Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning*. 2020.

[19] Abusnaina, Ahmed, et al. "Examining adversarial learning against graph-based iot malware detection systems." *arXiv preprint arXiv:1902.04416* (2019).

[20] Khoda, Mahbub E., et al. "Robust Malware Defense in Industrial IoT Applications using Machine Learning with Selective Adversarial Samples." *IEEE Transactions on Industry Applications* (2019).

[21] Clements, Joseph, et al. "Rallying adversarial techniques against deep learning for network security." *arXiv preprint arXiv:1903.11688* (2019).

[22] Rahman, Abdur, et al. "Adversarial Examples–Security Threats to COVID-19 Deep Learning Systems in Medical IoT Devices." *IEEE Internet of Things Journal* (2020).