Detection and Prevention of Replay Attacks in LoRaWAN Using Machine Learning

By

Abdul Samad Bin Shahid

00000203905

Supervisor

Dr. Muazzam Ali Khan Khattak

Co-Supervisor

Dr. Shahzad Younis

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters in Information Technology (MS IT)

In

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan.

(February 2020)

Approval

It is certified that the contents and form of the thesis entitled "Modeling, Simulation, and Forecasting of Wind Power Plants in Pakistan using Agent-Based Approach" submitted by Abdul Samad Bin Shahid has been found satisfactory for the requirement of the degree.

.

Advisor: Dr. Muazzam A. Khattak

Signature: _____

Date: _____

Committee Member 1:

Dr. Shahzad Younis

Signature: _____

Date: _____

Committee Member 2:

Dr. Safdar Abbas

Signature: _____

Date: _____

Committee Member 3:

Dr. Muhammad Zeeshan

Signature: _____

Date: _____

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Mr. Abdul Samad Bin Shahid, (Registration No 203905), of School of Electrical Engineering and Computer Science (SEECS) (School/College/Institute) has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: Dr. Muazzam Ali Khan Khattak

Date: _____

Signature (HOD): _____

Date: _____

Signature (Dean/Principal): _____

Date: _____

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: Abdul Samad Bin Shahid

Signature: _____

Acknowledgment

I thank Allah for giving me the strength to make this little effort reach fruition and for constant reminder that His plans are better than my dreams.

I am greatly obliged to my family for their unconditional love, continuous support, encouragement and prayers in all my endeavors. My deepest gratitude to my mother and my brother Sarmad for bringing the confidence in me and always providing the support I need to chase my dreams.

I am grateful to my Supervisor, Dr. Muazzam Ali Khan Khattak for his patient guidance, encouragement and advice to complete my thesis and for giving me numerous opportunities to learn and grow.

I am thankful to Dr. Shahzad Younis, Dr. Safdar Abbas and Dr. Muhammad Zeeshan for being on my thesis guidance and evaluation committee.

I am overwhelmed to acknowledge my depth to all those who have helped me to put these ideas into something concrete, specially Farooq Ahmad, Ammarah Irum, Komal Fayyaz, Ahmad Hasan, Muneeb Ahmad and Naveed Iqbal.

Last, but not the least, I thank my all friends for making NUST a home.

Table of Contents

List of Notations

ABP	Activation By Personalization
AES	Advanced Encryption Standard
AppEUI	Application Unique Identifier
AppKey	Application Key
AS	Application Server
EAP	Extensible Authentication Protocol
ED	End-Devices
FNwkSIntKey	Forwarding Network Session Integrity Key
IoT	Internet of Things
JoinEUI	Join Server Unique Identifier
JoinNonce	Join Server Nonce
JS	Join Server
LoRaWAN	Long Range Wide Area Networks
LPWAN	Low Power Wide Area Networks
MIC	Message Integrity Code
MITM	Man-In-The-Middle
NBIoT	Narrowband-IoT
NETID	Network Identifier
NwkSEncKey	Network Session Encryption Key
NS	Network Server
OTAA	Over The Air Activation
PKI	Public Kay Infrastructure
SNwkSIntKey	Serving Network Session Integrity Key

List of Tables

List of Figures

Abstract

LoRaWAN is a very popular globally adopted protocol for IOT networks. Despite many security measures introduced, there are still some flaws one of which is replay attacks. These attacks are a serious threat for the network traffic. Many solutions have been proposed for different scenarios of replay attacks, but till now no solution has been declared complete. In this paper we have analyzed different possible scenarios of replay attack and later compare the propose solutions. We have also proposed machine learning based solution.

Keywords: IOT, Machine Learning, Cyber Security, LORAWAN

Chapter 1

Introduction

This chapter provides the opening and general information of the research to provide a clear understanding about this thesis. It also covers the problem statement along with solution statement.

During last few years, we have seen a revolution in our lifestyles. One of driving factors of this change is the launch of sensors enable devices. This has given birth to a whole new era of research called IoT or Internet of Things. The devices equipped with sensors are called IoT devices. A prediction of 20 billion IoT devices will be in use of 3 billion users by the end of 2020 with the market share of 24.5 Billion USD by year 2021.[12][13] These IoT devices are equipped with micro-sized sensors are not only aware of its surroundings but are also capable of capturing and transmitting the data. These properties of sensors are not only helpful in improving our quality of lives but have also changed the manufacturing industry and services sectors drastically.

Since most of these sensors are comparatively smaller in size, they come with their own shortcomings like low-battery life, lower computational powers etc., and because of these limitations, the protocols designed for normal network traffic are not well enough for the communication between these sensors or devices equipped with sensors also called IOT devices. Different protocols have been proposed specifically for IOT networks so far such as, 802.15.4, 6LoWPAN, RPL, and LoRaWAN. LoRaWAN stands for long-range wide area network. It is a LPWAN protocol which means it is specifically designed for low power devices. Over the past few years it has become vastly widespread [2].

Initially released in 2015, LoRaWAN [1]is a globally adopted LPWAN protocol, popular due to its easily adaptability. It is specifically designed for devices with low power to communicate over long range and targets Internet of Things (IoT)devices requirements like two-way communication, localization services, end-to-end security, and mobility. Star-of-stars topology is used in LoRaWAN. In LoRaWAN, the end-devices can be classified in three types i.e. Class A, B and C, based on the bi-directional communication options and battery lifetime. Class A end-devices are able to enters into a low-power sleep mode for as long as their application defines: network doesn't need them to wake-up periodically. Class B devices sleep and wake-up periodically. This is done using periodic beacons. While Class C devices never sleep and are always up for communication [1]][2] [20].

1.1 Challenge

There are many noble techniques for different application domains but the problem domain is so vast and diverse that it is almost impossible to secure the end devices by using a technique solely.

There are lot of issues in LORAWAN such as its generous and continuous growth, mobility, network security, data handling and recovery. So, to develop a standard security solution to cope with the diverse problems is very difficult, but AI and Machine Learning has solved many longstanding problems lately.

Despite of many security features introduced, there are still many vulnerabilities present in LoRaWAN. Vulnerabilities of replay attack on individual devices may lead to a selective denial-of-service, recovery of data in plaintext, delivery reports

falsification, attacks on battery exhaustion and malicious message modification.[4]Most common of them is a replay attack, where uplink traffic from the gateway is sniffed via eavesdropping and then modified by hacker/intruder, resulting slowing down or total stoppage of network and also unauthorized access. These attacks can extremely vandalize the network, but can be detected and prevented based on the traffic data features. Many solutions have been proposed for different scenarios of replay attacks, but till now no solution has been declared absolute.

1.2 Motivation

The principal incentive of the Internet-of-Things (IoT) is to empower new worth cases by remotely checking and controlling distributed embedded systems, which together with their minimal effort will bring about inescapable arrangements. It is anticipated that there will be 13.5 billion associated questions being used by 2020. With the ever changing scenarios, these are always at some kind of threat. Our aim is to bring together three different domains of computer science i.e IoT, Cyber Security and Data Science, and propose a solution to a problem that can prove hazardous in future.

1.3 Problem Statement

All the existing solutions for the prevention of Replay Attacks in LORAWAN are either specific scenario based or requires major changes in the protocol, and thus are challenging to implement and adopt. Moreover, these solution fails in case if the

attacker finds a new way to carry out the attacks. There is a need for robust and generic solution for the prevention of Replay Attacks in LORAWAN.

1.4 Solution Statement

We propose to carry out different scenarios of replay attacks on the physical LoRaWAN network and generate the traffic data after successfully launching the replay attack. This data, will then be used for testing and training of different machine learning models which will help us to detect and prevent replay attacks. We proposed to apply different ML algorithms e.g. support vector machines, K Nearest Neighbors, Random Forest Tree Naïve Bayes and ANN in our model and perform binary classification on the network traffic and into normal and attack traffic. These classifiers will then be ensemble using AdaBoost and will be deployed on Gateway. The features on which the classification is performed are frequency, bandwidth, spreading-factor, device EUI, RSSI, SNR and location of the end device.

1.5 Research Impact

Replay Attacks are a serious threat for the LoRaWAN networks, as LoRaWAN being a widely adopted protocol for IoT all over the world. Our solution can provide security measure for millions of sensors and IoT devices communicating over LoRaWAN protocol around the globe.

1.6 Model of Study

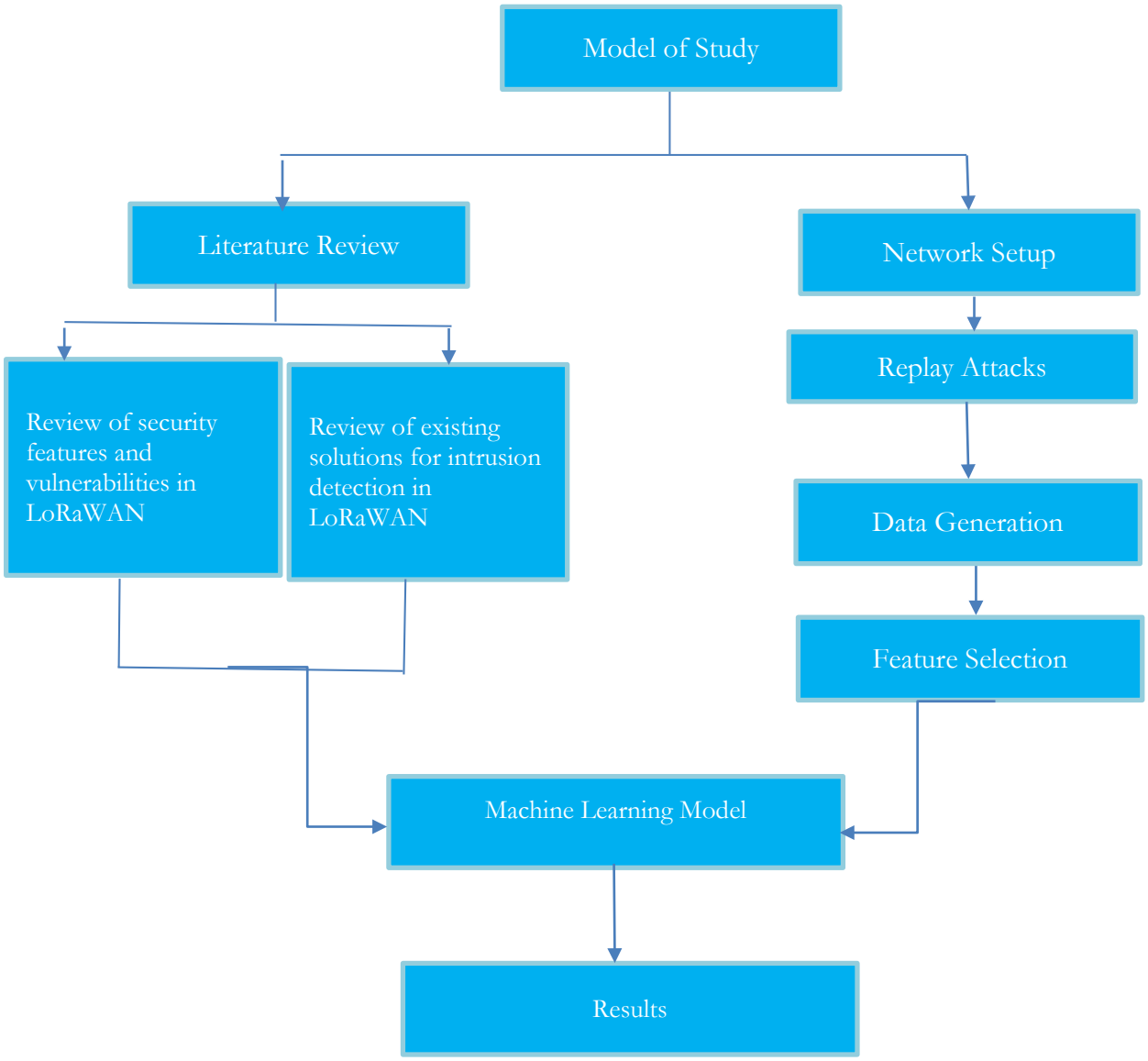

Figure 1: Model of Study

1.7 Thesis Organization

Rest of the thesis is organized in following chapters

1.7.1 Chapter 2: Background

Chapter 2 provides brief overview of LoRaWAN, its security features and vulnerabilities. This chapter also explains Replay Attacks and its different possible scenarios. Moreover, few preliminary concepts, used in methodology chapter are discussed as well.

1.7.2 Chapter 3: Literature

This chapter explains the work done so far related to Replay Attacks, existing solutions for replay attacks, and different methods of intrusion detection using machine learning. It summarizes the hierarchal approaches of model component and model composability and formulates research directions for this dissertation.

1.7.3 Chapter 4: Methodology

The chapter describes the servers, devices services and their communication in the network and the scenarios of replay attacks carried out. It also describes the collection of data and development of machine learning model.

1.7.4 Chapter 5: Verification of Results

This chapter is dedicated to demonstration of the functionality of our proposed framework taking into account the graphs. The chapter is concluded by results and their detailed discussion.

1.7.5 Chapter 6: Conclusion and Future work

Brief description of my thesis research work is presented in this section provided with tasks that can be carried out later for further research studies.

Chapter 2

Background

This chapter provides a brief overview of LoRaWAN, it's Security features and Vulnerabilities. It also includes introduction to replay attacks and its possible scenarios in LoRaWAN.

2.1 LoRaWAN

The Initial version of LORAWAN V.1.0 was released in 2015. Several improvements were made in the later version 1.1 and was released in 2017. It is a MAC Layer Protocol, specifically designed for devices with low power to communicate over long range. It works on top of LoRa which is a physical layer protocol. LoRa is a spread spectrum innovation; however, it's anything but an immediate succession spread spectrum innovation. Direct succession spread spectrum is modulating the bearer with chips/symbols to spread the transmission across more spectrum, which expands coding increase and symbol profundity. LoRa utilizes an unmodulated transporter in a FM tweet, which has similarities to M-ray FSK. So it is spread vitality over a more extensive band, yet not similarly DSSS is.

LoRaWAN basically targets key Internet of Things (IoT) requirements such as bi-directional communication, end-to-end security, mobility and localization services. LoRaWAN is based on star-of-stars topology. Low Power Wide Area

Networks (LPWAN) offers affordable connectivity of the wireless low power devices over very large areas. Many LPWAN technologies exist today, a detailed survey is presented in [1]. Among these technologies the most popular are LoRa by Semtech and NBIoT developed by IEEE in 802.15.4 standard. LoRa technology is developed for LPWAN where delay tolerant low data rate per device is required, the devices are spread over a large geographic region and typically extremely high number of devices exist in a communication cell.

In LoRaWAN, the end-devices can be classified in three types i.e. Class A, B and C, based on the bi-directional communication options and battery lifetime. Class an end-device can enter low-power rest mode for insofar as characterized by its own application for example there is no system necessity for occasional wake-ups. Class B devices are synchronized to the system utilizing intermittent beacons, and open downlink 'ping slots' at defined intervals. No encryption key or signatures are used for protection of scheduled reception windows for Class B devices. Desynchronization of reception can be performed by manipulation of timing references. This could lead to possible eavesdropping and replay attacks. The third and last type of devices are Class C devices, which are always up for communication.

One of the appealing features of LORAWAN is that it allows user to set the data rate (DR) for every device individually to communicate. The selection of the DR allows a dynamic trade-off between communication range and message duration. Less DR allows devices to communicate over a long distance and vice versa. In case we need devices over a long distance (more than 100 meter), multiple gateways need to be installed.

The devices in LORAWAN can communicate over chirp spread-spectrum (CSS PHY) modulation. Due to the spread spectrum technology, communications with different DRs do not interfere with each other and create a set of virtual 'code' channels increasing the capacity of the gateway.

To augment both battery life of the end-devices and generally speaking network limit, the LoRaWAN network server deals with the DR setting and RF output power for each end-gadget separately by methods for an Adaptive Data Rate (ADR) scheme. LoRaWAN baud rates go from 0.3 kbps to 50 kbps. [1] [2] [20] [24].

2.2 Security Features in LoRaWAN

2.2.1 Channel Confidentiality

A pair of two distinct keys is used by LoRaWAN v1.0.2 i.e. the network key NwkSKey, and the application key AppSKey. The Network key encrypt the message between end node and the network server. In case network server is further connected to some third party application, AppSKey is used to encrypt the message sent to the application server.

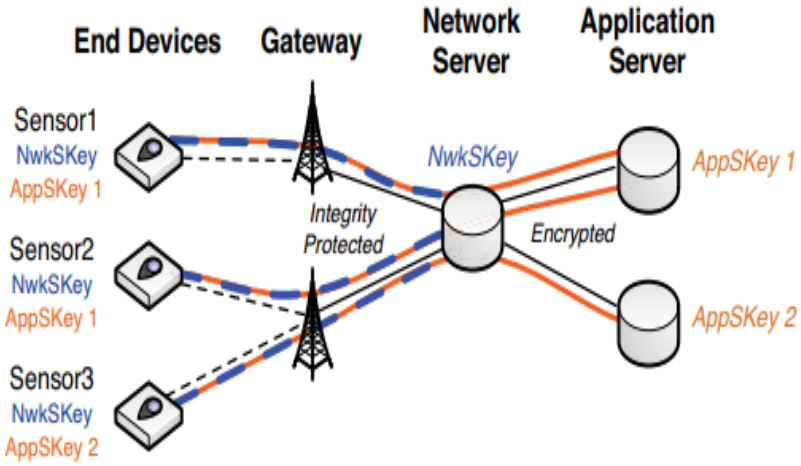

Figure 2: Encryption Keys in LoRaWAN

2.2.2 Enrollment Protocol

Following are the two methods defined by LoRaWAN to enroll a new end device into the network.

2.2.2.1 Activation by Personalization (ABP)

This method skips the exchange of join messages, as the devices are added manually by the network administrator. Unique parameters like DevAddr, NwkSKey and AppSKey are assigned to each end device. This information is then stored in the server.

End device sends messages directly to the server. The messages sent by these devices are encrypted and signed, such that only the corresponding network server can read the message.

2.2.2.2 Over-the-Air Activation (OTAA)

This module derives the network by encrypting the data using the *AppSKey*. An End device sends a join request containing a 3-byte DevNonce to the server. Network server decides if device to added in the network or not. If server doesn't respond, it means that the join-request is not accepted. Otherwise a digitally signed join accept message is sent, containing 3-byte AppNonce, which is generated by the network server. Join accept is secured using AES encryption.

Both sides use the nonce to generate the network and the application keys. The format for both keys is as follows:

AppSKey = AES*E* (AppKey, 0x02 || AppNonce || NetID || DevNonce || pad)

NwkSKey = AES*E* (AppKey, 0x01 || AppNonce || NetID || DevNonce || pad)

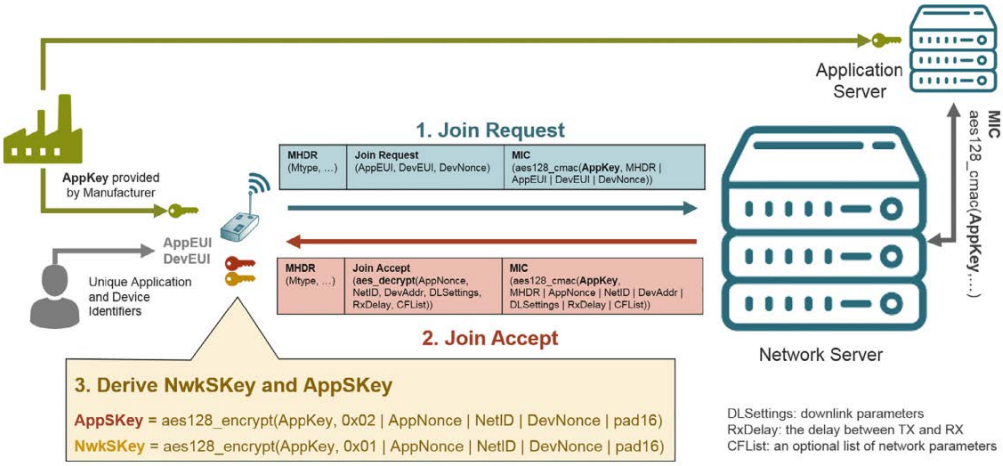

Figure 3: Activation procedure for OTAA in LORAWAN version 1.0.2

The OTAA procedure has been significantly changed in LoRaWAN v1.1. In the new architectural layout, a new server called Join Server has been introduced to manage the OTAA procedure. Furthermore, instead of a single network server, there are three network server roles introduced: home, forwarding and serving. The unique identifier of the Join Server, the *JoinEUI* and the *DevEUI* (unique identifier of the *End Device*) and both are pre-configured in the *End Device* during fabrication. The *End Device* also needs to be configured with the *NwkKey* and the *AppKey* (this can happen also during fabrication). After the End Device is deployed, it communicates with the Join Server (JS), through a gateway to initiate the OTAA Join Procedure. The session starts with the join request message sent from the End Device. The receiving Network Server checks the message and forwards the request to the JS which checks the entry for this specific End device in the Supported Devices List, matching the DevEUI of the End Device to its associated NwkSKey and AppSKey.

After a successful match, the join server responds with a Join Nonce. Then, the NS appends a NETID and also some radio and configuration parameters, along with a Message Integrity Code (MIC) to send back to the End Device in the join accept message. End Device validates the MIC and then decrypts the message to obtain the Join Nonce, NETID and parameters.

Finally, the *Join Nonce, Join EUI, DevNonce* and *NwkSKey* are used to create network session-long keys. The *FNwkSIntKey* - Forwarding Network Session Integrity Key - is used for the message integrity code (MIC) of uplink data messages. *SNwkSIntKey* - called Serving Network Session Integrity Key - is used for the message integrity code (MIC) of downlink data messages. *NwkSEncKey* and *AppSKey* keys (network and application) are used for confidentiality and integrity of the messages exchanged afterwards. *Join Nonce, Join EUI, DevNonce* and *AppSKey* are used to create application session-long key: *AppSKey*, the session key shared between the *ED* and *AS* and used to encrypt/decrypt application layer payloads.

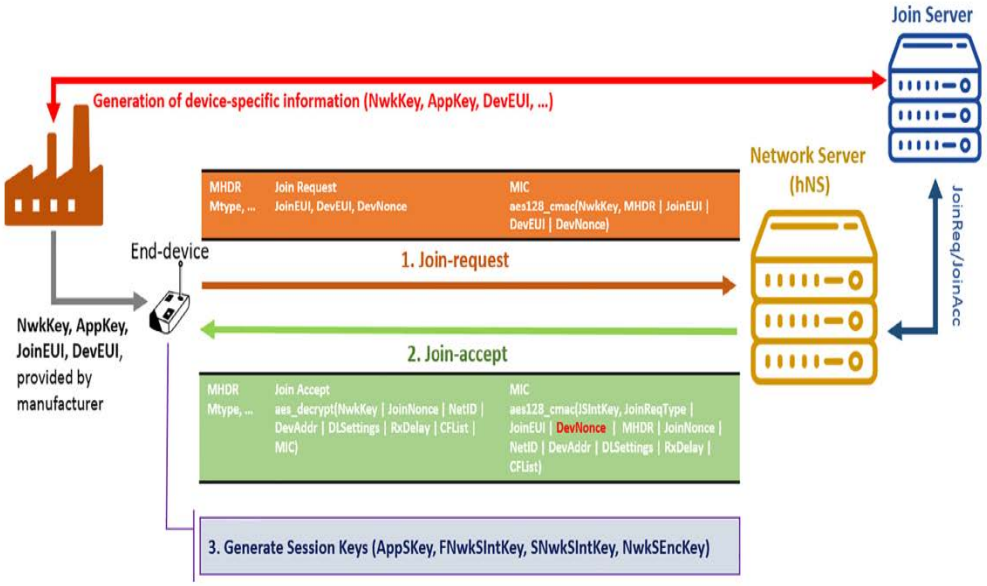

Figure 4: Activation Procedure for OTAA in LoRaWAN V 1.1

2.2.3 **Message Counters**

For each end device, there are two counters named FCntUp and FCntDown. FCntUp is including uplink messages at last device, while FCntDown is including downlink messages in the network server. So as to keep uplink and downlink messages in sync, there is a breaking point esteem MAX FCNT GAP. In the event that the distinction between number of uplink and downlink messages is bigger than MAX FCNT GAP, consequent edges will be disposed of. On the off chance that the counter overflows, it will be begun from 0 once more. These counters help in protection against replay attacks

2.2.4 Integrity and Authenticity Check

LoRaWAN uses a cryptographic message integrity code to give a trustworthiness data. The MIC for a data message is determined utilizing the NwkSKey and AES-CMAC technique. At the point when uplink messages land at the network server, the server will initially check the message respectability and, on the off chance that it passes the check, move the message to the application server keep an eye on the MAC header and payload.

2.2.5 Proprietary Handshake

This a kind of a secret question, which is asked when we forget our password and try to reset it. A secret will be shared between end device and the server, when the device will first join a network. This will act as a backup in addition to the DevNonce.

2.2.6 DevNonce Expiry

DevNonce should expire after a specific period and join process will be carried out again. Use of RSSI to keep track of location of end devices.

2.3 Replay Attacks

According to Kaspersky, "A replay attack occurs when a cybercriminal eavesdrops on a secure network communication, intercepts it, and then fraudulently delays or resends it to misdirect the receiver into doing what the hacker wants. The added danger of replay attacks is that a hacker doesn't even need advanced skills to decrypt a message after capturing it from the network. The attack could be successful simply by resending the whole thing."

Consider this true case of an assault. A staff part at an organization requests a monetary exchange by sending an encoded message to the organization's money related director. An assailant eavesdrops on this message, catches it, and is presently in a situation to resend it. Since it's a true message that has essentially been detest, the message is now accurately encoded and looks real to the money related manager. Right now, money related chairman is probably going to react to this new demand except if the person has a valid justification to be suspicious. That reaction could incorporate sending an enormous whole of cash to the aggressor's financial balance.

2.4 Scenarios for Replay Attacks

Replay attacks in LORAWAN can be carried out in various ways. Different possible scenarios are described as follows:

2.4.1 Scenario 1:

Initially, probe for sniffing the traffic is installed by the attacker within target area. The probe is use to collect join request within the target area. Once the join request is gathered, the probe identify the target end devices that sends periodic join request messages through collected message analysis. [6]

The targeted end device is ready for lunching the attack. On the basis of analytics patterns of join request messages frequency, and identify optimal time DT; is used for lunching the attack.

The attacker forwards gathered join request messages of targeted devices at a constant rate DT. As a result, the network server makes attempts to get connected with attacker device and discard the legitimate request messages received from targeted device. During the attack phase, the targeted devices will not be allowed to be part of the network till all join request messages are exhaust.

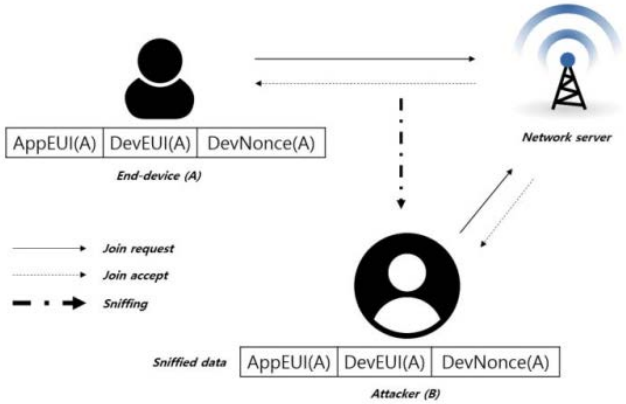

Figure 5: Replay Attack scenario 1

Vulnerability and seriousness of the threat is based on quantum of collected join request messages and methodology used of pattern analysis.

2.4.2 Scenario 2

A user joins the network by sending a join-request to the server. This join-request message which consist of user's AppEUI, DevEUI, and DevNonce. If an attacker succeeds to steal the packets of this join-request, he can use this this information to try to join the network. The server, upon receiving such request, compares it with the information already stored. Since the packets contains the same DevNonce, the server will consider it the same device and allow it through the network.

2.4.3 Scenario 3

In LORAWAN, two frame counters are kept for each end device as a security mechanism named as FCntUp and FCntDown. Former keeps count of uplink messages while other is for downlink messages. A threshold value MAX FCNT GAP is set so that both uplink and downlink remains in sync. When the difference between FCntUp and FCntDown reaches its threshold value, counter is reset to 0, while the authorization and session keys remains the same. If an attacker had grabbed the messages from the previous sessions with the greater counter values, he just needs to wait till the counter is reset. After that, he can send those packets to the network server. The server will disconnect the legitimate end device and allow the attacker, thus will become the part of the network. [4]

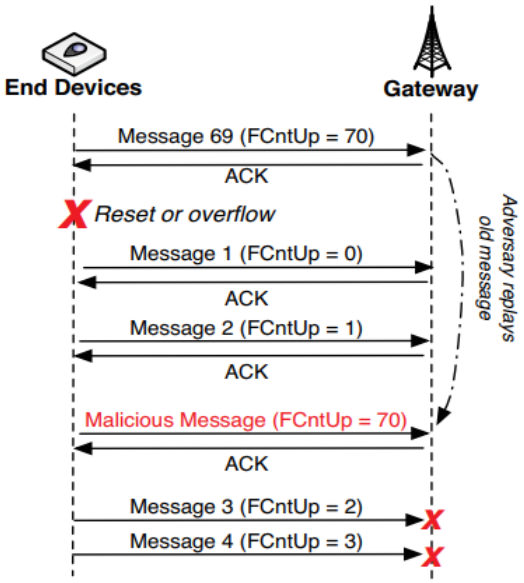

Figure 6: Replay Attack scenario 3

Assumption of value of uplink counter malicious message is FCntm, the value of uplink counter of the end device is FCntcurr, and the maximum accepted counter gap is Gap. If any message with FCntm - FCntcurr ≤ Gap is replayed to fit the running window algorithm of LoRaWAN and accordingly accepted by the network. If the counter value is selected as FCntm = Gap + FCntcurr, it becomes deadliest attack as devices will take the longest time for recovery.

2.4.4 **Scenario 4**

Replay attacks can also be carried out by selective RF Jamming for OTAA session. Attacker needs to observe join-request with DevNonce from an End Device and

Jam the corresponding Join-accept message from the Network Server for the same End Device.

The End Device will try again to join the network after timeout by sending the join-request again with the same DevNonce. Network Server will respond again but join procedure will fail again because of the selective jamming. Now as the legitimate End Device is unable to communicate with the network server, the attacker can use the stolen join-request packets join the network. The limitations for this scenario are I) the Jammer should be placed in such a way that both Network Server and the End Device are in its range, so that signal should be captured before reaching its destination II) and the packets should be stolen before it reaches the End Device from the Network Server. The first condition is sometimes impossible to achieve as it is dependent on the transmission range and the how close we are to the nearest gateway. The more gateways are there, the more difficult it is to launch the attack.

Chapter 3 Literature Review

This chapter documents the existing techniques to detect and prevent Replay Attacks provided in literature. Moreover, it contributes to understand development in the area of research.

3.1 Area of Research

A detailed literature review has been conducted to identify studies related to Security Analysis of LoRaWAN, Prevention of Replay Attacks in LoRaWAN and Intrusion Detection using Machine Learning as well. Provided Literature Review is divided into following sub-sections.

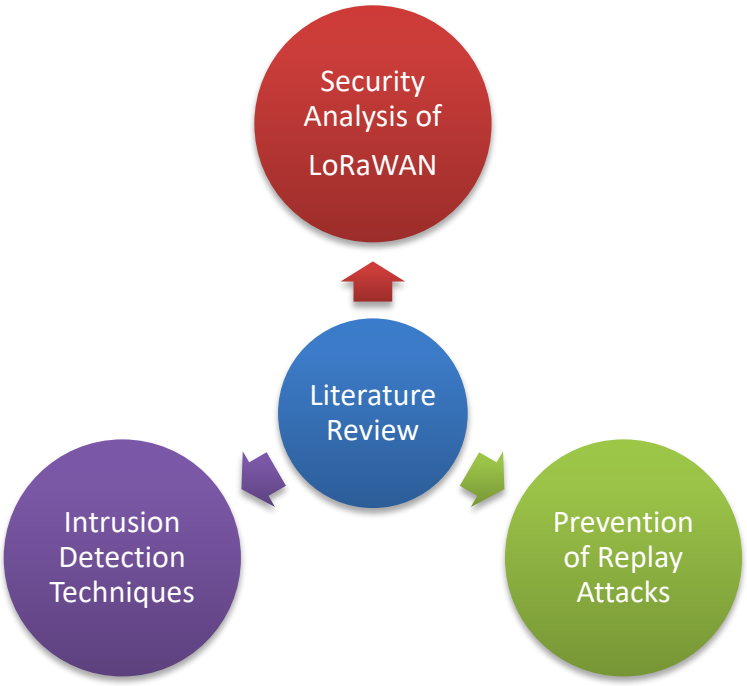

Figure 7: Areas of Research

3.2 Replay Attacks Prevention

Several techniques have been suggested to prevent the replay attacks over the years. Here is a brief overview of most common of these.

3.2.1 Application and Networking Keys

In 2018, [8] Sanchez-Ibarra et proposed a light weight EDHOC derived session keys to secure the communication in the LORAWAN networks. They named the keys NwkSKey and AppSKey. application-level packets are encapsulated into UDP datagrams which are then converted to IPv6 packet using Context Header Compression (SCHC) algorithm. This conversion is reverted once the packet reaches network server. This mechanism is explained in the figure below

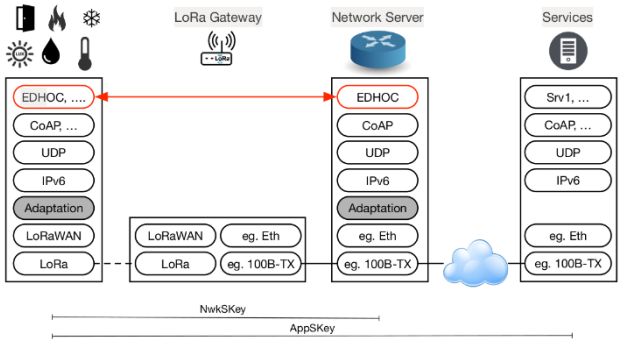

Figure 8: Certification Authority

To overcome the detention of trust by application server on Network or join server of LoRaWAN, Suman Bala et al., [5] proposed in 2019 an alternate different key

generation approach for Network Server and Application Server. In this approach the joining of End Device to Network Server and Application Server are not dependant with one another. In this methodology, two distinctive master keys are used to determine separate sessions keys and application session keys. The system master key is retained with Network Server and the application master key is retained with Application Server. This approach will permit the jobs isolation between Network Server and Application Servers. Accordingly, this methodology can keep up the start to finish security between Network Server and end devices, and between Application Server and End Device. This methodology likewise enables an End Device to join different Application Servers, and even in a solitary solicitation

3.2.2 DES and AES based Encryption

Yang et al. Proposed messages-based encryption using Data Encryption Standard (DES) and Advanced Encryption Standard (AES) algorithms. Since the message is encrypted, it is not possible for the intruder to read the encrypted cypher text without the decryption key(s). They assumed ideal conditions for cryptography with unbreakable encryption. Data Encryption Standard (DES) and Advanced Encryption Standard (AES) are classified as symmetric key encryption techniques, Scyther does not differentiate between these. Since, Scyther only take care the logical part of the security protocols, our model might fail to detect few deficiencies in (i) root keys separation in the derivation of NwkSKey and AppSKey, (ii) nonce randomization weaknesses, and (iii) any physical layer attacks like radio jamming etc.

3.2.3 AES-128 based SeLPC

AES-128 based Secured Low-Power Communication was proposed by Kun-Lin Tsai et al. [10]. The proposed methodology is aimed to accomplish security and low-power-consumption for LoRaWAN. It is proposed that the encryption key (AppSKey) and lookup table (D-Box) are periodically updated on both end devices and application server. This will enhance significant the security level of LoRaWAN communication. The periodic update can be set for both 10-round and 5 rounds per K days. This will save power consumption used for encryption and enhance the battery life of end-devices. Results showed that the SeLPC can save up to 26.2% of power consumption as well as providing security to resist eavesdropping attack, known-key attack, and replay attack.

3.2.4 Proprietary Hand Shake and RSSI-based scheme

Woo-Jin et al [7], proposed a solution to scenario 2 replay attacks. It detects if a user has been attacked based on the network features like RSS and proprietary hand-shake in addition to DevNonce, AppEUI and DevEUI. RSS can be used to detect the attacker based on their movement, while proprietary hand-shake defines some rule between the server and end node regarding join process so that even if the attacker is successful in sniffing and spoofing, they won't be able to join the network. [7]

3.2.5 Hybrid Message Authentication and Passing (MAP)

C. L. Abad [9] suggested that security from Replay attack on nodes can be provided through MAP along with identity & timestamp keys in packet transaction state. A base station with N nodes sent request REQ as first transaction message, the same is replied as REP from other end. Pernicious nodes could be distinguished by contrasting the ID & timestamp of the sent & received messages. MAP algorithm is advantageous for both multicast as well as unicast communication networks.

3.2.6 Unique Token Assignment (DevNonce)

Seung Jae et al [11], proposed countermeasure of these attacks is to assign a unique token to each of join request messages. To distinguish the duplicate join request messages at the network server end, DevNonce is used which discards the join request message if it has already been used. Regardless of whether an aggressor captures join request messages; the join messages are most certainly not legitimate to misuse on the grounds that each message is conceal with different tokens. The session key is also changed every time a join request message is retransmitted which is calculated using a AES128 encryption method. [6]

3.2.7 Rabbit-based Enhanced Key Management Scheme

Jialuo at al, [12] After analyzing how the keys are managed in LoRaWAN, Lora End Device could be classified as the most vulnerable part of the LoRaWAN that could be easily manipulated by the attackers since it saves the root keys used for deriving

session keys. Root-keys are static through there life time in LoRaWAN 1.1 that gives attackers sufficient time to manipulate the system and trace the session key derivation, currently the session key derivation method AES-ECB mode is used to reduce the risk by updating root keys periodically but it's also not suitable since its prone to pattern analysis. It is recommended that the updates of root key should be kept less frequent and range is set in a reasonable frame as it effects the life span of the resource node. Simple arithmetic based high efficiency stream cypher technique 'Rabbit' is used as two step KDF for root key update. The output is a 128-bit string from keystream generator in every round, with suitable randomness. This will comply the proprieties of LoRaWAN compatibility. Two step KDF comprises of two phases which uses the PNG of the Rabbit stream cipher for obtaining key-streams. Key Extractor also have two steps, an initialization process and keystream generation. In initialization process, it takes a root key and from a series of shared context keys; it extracts uniformly random values for session key. This will update immediately as root key and it gets updated. The number of iterations selection should balance the resulting randomness and computing requirements in second step. 128-bit keystream will be extracted once it reaches the last block of keying material. A Key Expander that takes output of extractor which is a sequence of 128bits and after doing some procedures classifies it as the new root-key.

Lora End devices and JS has the information regarding root keys which are distributed remotely in Lora end devices. Transmission of updated root keys are not essential using KDF. Request for accessing and updating the root key is initiated from network server using downlink MAC command which is verified by JS. The same process is to be adopted by end devices. Once the request is accepted, the device will send 'New Key Ready' command to JS which contains a nonce with a MIC generated using newly generated root keys. MIC is verified on JS by using its

newly generated root keys with received one. Another join- request process will start as per specified procedure defined in LoRaWAN version 1.1 after confirmation from ED. The process of new root key update is accomplished once the end device will re- join the network. Whilst updating root keys, few factors should be considered such as power consumption, lifetime of the root keys, length of transmitted messages with the valid root keys. Keys should be updated if it is used for longest times.

3.2.8 Random Key Pre-distribution

It is used in wireless sensor networks to achieve secure communication between nodes. A set of random keys is assigned to a node, and ordered to distribute these keys with the neighbours to find the matching keys. So, these keys ensure their respective privacy. The complete session implemented is hidden from the others. The main idea is the link of the key with the node to which it is assigned and validation of the key. To ensures that the network must have the ability to compute the process of validation and assigning the keys, this technique is used. The attackers will not pass the key tests so they will be unveiled. [18]

3.2.9 Auditory Inspired Spatial Differentiation for Replay Spoofing Attack Detection

Automatic Speaker Verification Systems are seriously vulnerable to various types of spoofing attacks. In which, replay attacks are very common and can be deployed easily. An auditory inspired detection technique [] based on spatial differentiation for replay spoofing attacks was proposed for replay attacks in speech detection. They

used a parallel filter bank comprising of 2nd order Infinite Impulse Response (IIR) band pass filters for experimentation. On the other hand, processing similar to spatial differentiation was engaged to attain higher order stable IIR filters, which resulted in highly selective filter banks. At front-end, the spatial differentiation used as pre-processing technique and at back end, two systems with conventional 2nd class Gaussian Mixture Model (GMM) to perform baseline were used. So, fusion of the described systems brought the error rate down with the improvement in selectivity and detection rate.

3.2.10 Detection of Fabrication, Replay and Suppression Attack in VANET- A Database Approach

Another Database based approach [26] to enhance road safety and decrease misfortunes in functioning vehicular Ad hoc Networks was proposed for various attacks including Fabrication, suppression, and especially Replay Attacks. In their proposed VANETs, each vehicle is certified at Certification Authority (CA) at root level. The vehicles are certified by providing Electronic License Plate (ELP) embedded with Encrypted Vehicle Identification Number. The Intermediate level consisted of base stations (BS), and leaf level consisted of vehicles. Whenever a vehicle enters a BS, the base station verifies its Identification and assigns a digital signature to it, so each authentic unit has an effective and valid digital signature. Each vehicle sends Beacon and service message in response of multiple events. The service messages module creates certification revocation list (CRL) by inserting Vehicle Identification Number of malicious attempters. The base stations bear the responsibility of detecting false units in their concerned area.

3.2.11 Replay Attack Detection Using Magnitude and Phase Information with Attention-Based Adaptive Filters

Another Replay Attack detection technique for Automatic Speech Verification (ASV) systems was proposed [27] for ASV spoof challenge, which was based on multi-channel feature extraction scheme with attention based adaptive filters. Their suggested framework is consisted of three fragments. First, discrimination abilities using frequency ratio techniques for full frequency band analysis. Secondly, they proposed an innovative adaptive relative phase (ARP) feature and Adaptive Frequency Cepstral Coefficient (AFCC) feature, and finally detecting replay attacks by fusing phase and magnitude info. For magnitude and phase information they used a Gaussian Mixture Model (GMM) based replay speech detector.

3.2.12 Secure Data Timestamping in Synchronization-Free LoRaWAN

Internet of things (IoT) require ubiquitous connectivity, so state of the art technologies like LoRaWAN are very crucial to achieve this. LoRaWAN has very limited bandwidth, so the key function of LoRaWAN is to gather low-rate monitoring data from distributed sensors. In distributed sensor networks, data timestamping is of critical importance. An attack aware, and low overhead synchronization-free approach [28] to timestamp the uplink data at the LoRaWAN gateway was proposed to attain milliseconds. The said approach was vulnerable to a frame delay attack implemented by maliciously jamming and delayed replay. To cope with this susceptibility, a SoftLoRa gateway design was suggested in the same work. The design consisted of a commodity LoRaWAN gateway with a low-power

software defined radio receiver to trail the essential frequency biases of LoRaWAN and devices, and a set of competent signal processing algorithm based on LoRaWAN modulation methods. Integration of both parts resulted in achieved a remarkable resolution up to 0.14 parts per million of the central frequency of channel. The resolution is sufficient for detection of attacks in case of any manipulation.

3.2.13 Transmission Line Cochlear Model Based AM-FM Features for Replay Attack Detection

Cochlea is the inner part of an ear involved in hearing. The transmission line cochlear model bears a resemblance to human ear cochlea more accurately than conventional parallel filter bank models in front end replay attacks detection models. The basilar membrane is demonstrated as a cascade of digital filters with declining resonant frequencies. A recent research proposed [29] two feature based scheme [], consisted of transmission line cochlea frequency modulation (TLCFM) and transmission line cochlea amplitude modulation (TLCAM). The modulation features of the speech are extracted from simulated membrane displacements by these two modulations. The TLC amplitude modulation output is resembled to the output of inner hair cell bending movement to accurately capture the amplitude modulation component of the speech. And on the other hand, TLC frequency modulation is extracted by originating in-phase and out-phase signals from basilar membrane movement. These modulations outpaced the parallel filter bank baseline systems. The score level fusion of TLC AM & FM proved the improvement than previous systems. TLC AM achieved higher frequency than FM.

3.2.14 Machine Learning for Attacks Detection

In 2018[2], Rohan, Noah and Nick demonstrated that specific features selection can be used for the detection of DDOS attacks in consumer IoT. The IoT traffic has repetitive features like regular traffic intervals and limited number of end points. Due to unavailability of public datasets on attacks in IoT, they simulated a network with IoT devices including router and end nodes. The router acted as an on-path device able to observe traffic between end devices, LAN and internet. It could also inspect, block or modify any data in transition. Once the network as simulated, the traffic data generated. DOS attacks were performed on the subjected virtual network by some adversarial devices. After the success implementation of attacks, the packets of traffic data were captured that included features like source IP address, source port, destination IP address, destination port, packet size, and timestamp of all IP packets sent from smart home devices. This process produced a dataset of 491,855 packets, comprised of 459,565 malicious packets and 32,290 benign packets. These packets were then categorized by devices and timestamps. Features engineering was performed on these packets and two classes of features i.e. stateless and stateful. The flow-independent features are called stateless like packet size, inter-packet interval and protocol, while stateful features changes over time during transmission e.g. bandwidth and IP Destination Address Cardinality and Novelty.

Applied machine learning algorithms and their accuracy on different features are as:

ML Technique	Stateless Features	All Features
K Nearest Neighbor	0.967	0.995
Support Vector Machine	0.92	0.921

Decision Tree	0.977	0.995
Random Forest Tree	0.981	0.998
Neural Network	0.939	0.989

3.2.15 **Self-Organizing Map (SOM)**

Self-Organizing Map is classification techniques in machine learning. Using this technique, C. Langin et al. proposed a P2P botnet version of model for distinguishing normal traffic with malicious traffic. The model is described in this way as starting with considering a bot communicates with C&C which is P2P in this case. The P2P network also includes infected nodes if firewall is not configured properly on it. After that infected nodes start communicating with other nodes to make it infected, here author illustrates that these sorts of infection initiated from over 40,000 unique source IPs. Now if firewall is properly configured and deployed, then the session initiating is denied and logged. After getting logged, these logs sent to log server for further analysis on it. Holistically, this node is not getting infected and all the incoming traffic which is denied is tagged as random. Along with that this model also assures that P2P botnet will not reach to any locally hosted node. Based on the model explained, Author proposed methodology is 7-step methodology. In Step -1, Firewall that is placed, log all the entries in a file. i.e. syslog, for some specified time interval. In step - 2, scanning the log file and calculate the net counted values of external log entries that are denied against each IP address on the network, considering the local IP would be the destination IP in log file. For matching with bot cluster, it will compare with threshold value defined for botnet. It should be kept track that the count value in file is equals to the required count value for comparison. In step - 3, Log file will scan again for the

values that meet the threshold value. In step 4, for generating summary, also save the entries that meet the threshold into a vector file. In step 5, Analyze each entry for best matching of bot cluster and label if any of listed is suspected. In Step - 6, for each suspected vector, load entries to database for further analyzing and obtaining detailed information & finally in Step 7, Analyze, evaluate & investigate manually.

3.2.16 **Clustering (K-means)**

Pragati Chandank hede et al. Proposed architecture for detection of botnet in an unsupervised way. Firstly, traffic is captured from different sources, after that data is aggregated so that the bot communicated to their bot master. Along with that evidences are also captured by keep tracking the process that is running and extracted communication and commands. These evidences are captured by Evidence Accumulation technique and for detecting hidden bots one applies subspace clustering with the application of evidence accumulation technique.

3.3 Comparative Analysis

Sr.#	Name	AKR	E	ECR	ARR	En	O
1	Application and Networking Keys	Y	Y	N	N	Y	Y
2	AES-128 based SeLPC	Y	Y	N	N	Y	Y
3	Proprietary Hand Shake and RSSI-based scheme	Y	Y	N	Y	Y	Y
4	Hybrid Message Authentication and Passing	N	S	N	Y	N	N
5	Unique Token Assignment	Y	S	N	N	N	N
6	Random Key Pre-distribution	Y	N	N	N	Y	Y
7	Secure Data Timestamping in Synchronization	Y	Y	N	Y	N	Y
8	Media Authentication and passing & Media Access Control	N	Y	N	N	N	N
9	Rabbit-based Enhanced Key Management Scheme	Y	Y	Y	N	Y	Y
10	Auditory Inspired Spatial Differentiation	N	Y	N	Y	N	Y
11	Trust-based distributed Kalman Filtering	N	Y	N	Y	N	Y
12	DES and AES based Encryption	N	S	N	N	Y	Y
13	One-time Password Authentication Scheme for WSN	Y	Y	N	Y	N	N

AKR- Additional Key Required E- Efficient ECR- External Certification Required ARR- Additional Resource Required O- Overhead En- Encryption	Y-Yes N- No S-Sometimes

Table 1: Summary of Replay Attacks Detection and Prevention Techniques

3.4 Problems in Existing Techniques

The first technique discussed is "Application and Networking Keys" which seems to be the best technique to mitigate Replay attacks but it bears some issues. Due to its infrastructure and implementation it is too expensive. It has general application domain.

The second technique "AES-128 based SeLPC" which is energy efficient but we need to manage a lookup table. Moreover, the keys also need to re-generated after a specific interval.

The third technique discussed is "Proprietary Hand Shake and RSSI-based scheme". Received signal strength Indicator only indicates the new attempts to manipulate identities. It is a costly technique in term of metrics, due to the difficulty in calculations of locations. Moreover, it also requires hardware for RSSI monitoring.

The Fourth technique, Hybrid Message Passing requires additional time before the devices and server actually starts communication.

The fifth discussed technique is "Unique Token Assignment". Comparatively this works better and requires fewer resources comparatively.

The sixth technique is Random Key Pre-distribution quite efficient in preventing the Replay attacks, but we need to assign keys before the devices and server actually starts communication. This has proven efficient but its process is quite time consuming and require computational power.

The 7th technique that has been discussed is Secure Timestamping and Synchronization. It has been proven quite efficient but is costly in terms of computational operations required.

The 8th discussed technique is "Media Authentication and passing & Media Access Control". Comparatively this works better. Although we need to keep record of the timestamps.

The 9th technique that has been discussed is Rabbit-based Enhanced Key Management Scheme. It has been proven quite efficient but is costly in terms of computational operations required.

The 10th technique is Auditory Inspired Spatial Differentiation. It only works for auditory systems. Moreover, we need to have location information of all the end nodes.

The 11th technique is Trust-based distributed Kalman Filtering. In this proposed solution, devices need to handshake first, which makes it slow.

The 12th technique discussed is "DES and AES based Encryption" which is based on Scyther, which is a tool for verification of security protocols. Data Encryption Standard (DES) and Advanced Encryption Standard (AES) are classified as symmetric key encryption techniques, Scyther does not differentiate between these. Since, Scyther only take care the logical part of the security protocols, our model

might fail to detect few deficiencies in (i) root keys separation in the derivation of NwkSKey and AppSKey, (ii) nonce randomization weaknesses, and (iii) any physical layer attacks like radio jamming etc.

The last discussed technique is One Time Authentication Scheme. This has proven efficient but its process is quite time consuming as we have to setup id and password first. I also require additional computational power.

Chapter 4

Methodology

This chapter the design and implementation of the experiment along with data acquisition and machine learning algorithms applied.

4.1 Proposed Methodology

Our idea is to simulate the LoRaWAN network implemented in SEECS for replay attack and gather the traffic data after successfully launching the replay attack. This data, will then be used for testing and training of different machine learning models which will help us to detect and prevent replay attacks using different ML algorithms. We aim to apply two different ML algorithms i.e. support vector machines and K Nearest Neighbors in our model and perform binary classification on the network traffic and into normal and attack traffic. The result produced by both techniques will also be compared to find out the which technique works best. The suggested features on which the classification will be performed are packet size, ip address, round trip time and physical payload.

This model would train against the historical dataset of both normal and intruded packets and will binary classify the traffic into normal and attack traffic. This model will be deployed on the gateway. The gateway on receiving suspicious packets, will be discard them and thus preventing the network from possible attacks.

Different machine learning algorithms will be implemented and later ensemble to improve the accuracy so that normal traffic should be prevented from discarding.

This model will be able to provide better performance with higher accuracy than the already existing techniques because not only will this handle all the scenarios of replay attacks but will also be able to evolve and update according to the changing traffic patterns

4.2 Design of Experiment

Inspired by Rohan et.al. [2], we setup a LoRaWAN network with a variety of devices. These devices were set to send and receive data packets to and from server. Some basic network parameters of this network traffic i.e. SNR, RSSI, Frequency, Location and data rate were set to be stored in a database. A few devices among these acted as attacker nodes and performed sniffing and spoofing to carry out replay attacks in the network. This experiment was carried out for week at NUST School of Electrical Engineering and Computer Science. After a week, we had enough data to perform machine learning on it. Once the Machine Learning models were successfully implemented, we took the output of these models, calculated the confidence interval from the model with the best performance, and wrote a trigger on the database table to prevent the malicious packets being saved.

4.3 Network Setup and Topology

To conduct the experiment, we took six RN-2483 microchip devices. For the sake of variety in data devices and the dataset, three devices were configured on ABP and while other three were set to work on OTA mode. Out of these six devices four were stationary and two were mobile. The mobile devices were used as attacker nodes.

These devices were linked to a Laird RG 186 gateway, which was further connected to our Lora server.

All these devices were Class C devices i.e. these were always up for the commination during the experiment. Details of these devices are mentioned in the below table.

Device Name	**Mode**	**Mobility**	**Type**
T62	ABP	Stationary	Legitimate
T63	ABP	Stationary	Legitimate
T65	ABP	Mobile	Attacker
T66	OTA	Stationary	Legitimate
T67	OTA	Stationary	Legitimate
T68	OTA	Mobile	Attacker

Table 2: Details of end devices used in experiment

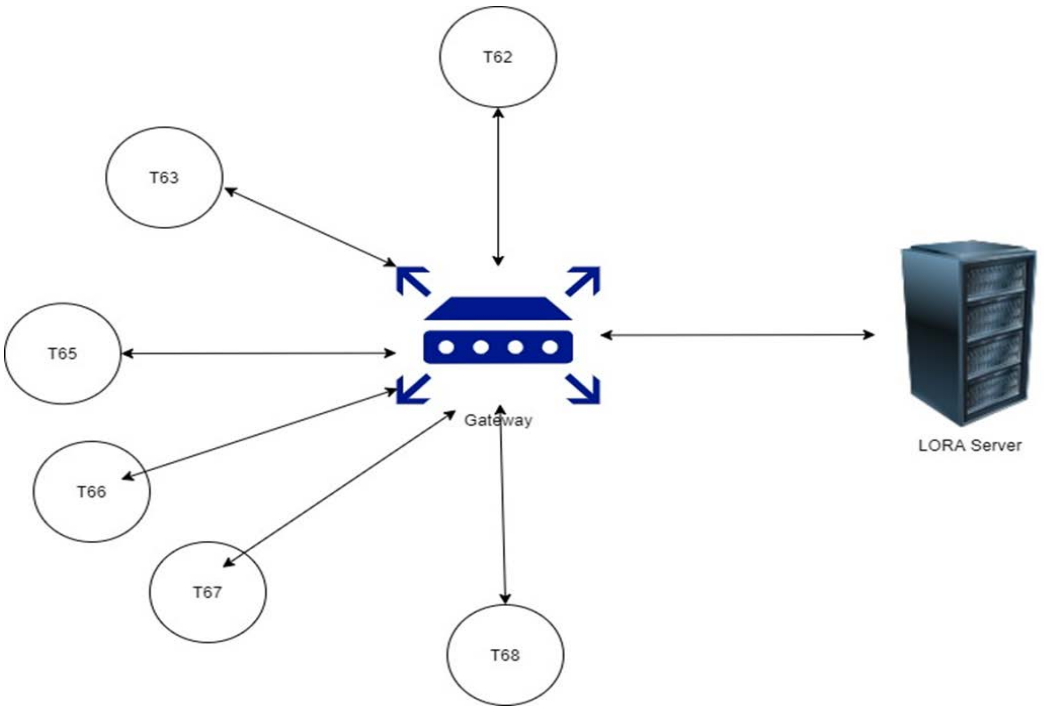

Figure 9: Network Topology Diagram

4.4 Data Generation

The devices enlisted in the above-mentioned table are deployed at different locations at NUST School of Electrical Engineering and Computer Science with the range of our gateways. These devices were set for bi-direction communication with the server in order to generate dataset for normal traffic, periodically. For three days, these devices sent and received data frames to and from the LoRa server. These data frames along with traffic features like RSSI, SNR, Longitude, Latitude, Data Rate and Frequency etc. were stored and continuously monitored at the server end. After three

days we had a data gathered a dataset of 35656 packets. These packets were labeled as normal traffic.

4.5 Data Structure

The data collected during the experiment contains the following parameters:

Parameter	**Description**	**Type**
Application Id	Unique Id Of Application	Numeric
Application Name	Defined Name Of Application	Char
Device Name	Defined Name Of Device	Char
Mode	Mode On Which The End Device Is Working	Char
DevEUI	Unique Identifier Of End Device	Char
Gateway ID	Unique Id Of Gateway	Numeric
RSSI	Radio Signal Strength Indicator	Numeric
LoRa SNR	Signal To Noise Ratio	Numeric
Latitude	Latitude	Numeric
Longitude	Longitude	Numeric
Frequency	Frequency On Which Device Is Working On	Numeric
Data Rate	Data Rate	Numeric
ADR	Describes If Adoptive Data Rate Scheme Is On Or Off	Boolean
Is_Malicious	Describes If The Traffic Is Malicious Or Not	Boolean

Table 3: Structure of Gathered Data

4.6 Data Preprocessing

At first, the collected data was in raw JSON format. Which was then cleaned and preprocessed. The null values were removed. Data types were rectified. Headers were defined. And to make classification easier, a new column was introduced and normal traffic was labeled as 0 and malicious data was labeled as 1.

4.7 Feature Selection

The parameters that were constant throughout the dataset i.e. Application ID, Application etc. were ignored during the trainings of machine learning models. The features that were used to train models were Lora SNR, RSSI, Longitude and Latitude, device name, device mode, frequency and data rate.

4.8 Machine Learning

After the preprocessing of data, different machine learning algorithms are applied on data to perform supervised classification. The list of applied algorithms include Naïve Bayes, Multiple Regression, K Nearest Neighbor, Support Vector Machine, Random Forest Tree, and Decision Tree. All these algorithms are implemented in python version 3.7 using Anaconda Spyder. The python libraries used in the process are: Pandas, Scikit-learn, tensor flow, Numpy and Matplotlib.

4.8.1 Multiple Regression

It is a statistical machine learning algorithem used for regression problems. Formally its a supervised learning method an extension of linear regression which is used to perform classification based on only one variable multiple regression is used when we want to classify a variable based on one or more variables simply put it could used many variables to classify the new variable basing them on a linear seperator.

Its most important use could be to see how the independant variables are effected by dependant variable or viceversa. More for it could be easily visulaized with how changing values of one variable effects or changes the values of others.

This algorithm was implemented using Scikit-Learn package of python. The 75% of the dataset was used to train the model, while the remaining 25% was used for testing. As this is a regression algorithm and instead of classifying a record as 0 or 1 and provides a continuous output between 0 and 1, we set a threshold at 0.9 and classify the records with greater than 90% output as malicious.

4.8.2 K Nearest Neighbor

It is a simple, no parameter machine learning algorithem used for both descriptive classification and regression problems. Formally its a supervised learning method, It works fairly simply by initializing the data with random startpoints, all it does is calculate the distances it puts the objective data point based on the minimum

distance between the number of data points near to it and thus claasifies it into the group having the most minimum distance to it.

Due to its simplicity its the most used lazy learning algorithem, it has a database with fairly seperated data based on which it classifies the new data points.

This algorithm was implemented using Scikit-Learn package of python. The 75% of the dataset was used to train the model, while the remaining 25% was used for testing. This algorithm provided results with 98.65% accuracy.

4.8.3 **Support Vector Machine**

It is a populer, non-probablistic machine learning algorithem used for both descriptive classification and regression. Formally its a supervised learning method, we input it with labelled data and it works on it by seperating it into hyperplanes (hyperplanes are basically the lines that could basically be used to seperate data into either of the classes) which could then be used to define new examples or classify them. It does all this work by employing kernels to transform given inputs and then based off of that transformation it classifies new examples on the new obtained boundrylines which provides user with the maximized most optimal seperation between the different classes.

This algorithm was implemented using Scikit-Learn package of python. The 75% of the dataset was used to train the model, while the remaining 25% was used for testing. This algorithm provided results with 97.24% accuracy.

4.8.4 **Random Forest Tree**

It is a ensemble machine learning algorithem used for both descriptive classification and regression problems. Formally its a supervised learning method that works by creating forests most commonly known as the trees of data mostly for classification. the higher the depth of the forrest goes the more accuracy it has. It classifies a new data point as it visits all the trees in ensemble which were created by randomly selected data at training time and now the new data point is classified to having the most similarity to the tree.

This algorithm was implemented using Scikit-Learn package of python. The 75% of the dataset was used to train the model, while the remaining 25% was used for testing. This algorithm provided results with 98.77% accuracy.

4.8.5 **Decision Tree**

It is a populer machine learning algorithem used for both classification and regression problems. Formally its a supervised learning method, which works by creating decion trees that use the knwledge already present in the input data and at classifiation it uses the obtained inferences from the input to make decions on the output classifications. Most importantly used for strategy making inorder to acheive the desired goals.

This algorithm was implemented using Scikit-Learn package of python. The 75% of the dataset was used to train the model, while the remaining 25% was used for testing. This algorithm provided results with 97.46% accuracy.

4.8.6 **K Means**

The k-means clustering algorithm attempts to part a given unknown informational index into a fixed number (k) of clusters.

At first k number of aimed centroids are picked. A centroid is an information point (fanciful or genuine) at the focal point of a cluster. Every centroid is a current information point in the given information informational index, picked indiscriminately, to such an extent that all centroids are novel (that is, for all centroids ci and cj, ci $\neq$ cj). These centroids are utilized to prepare a KNN classifier. The subsequent classifier is utilized to group (utilizing k = 1) the information and in this way produce an underlying randomized arrangement of clusters. Every centroid is from there on set to the number juggling mean of the cluster it characterizes. The procedure of grouping and centroid alteration is rehashed until the estimations of the centroids balance out. The last centroids will be utilized to deliver the last characterization/clustering of the information, adequately transforming the arrangement of at first mysterious information focuses into a lot of information focuses, each with a class personality.

This algorithm was implemented using Scikit-Learn package of python. The 75% of the dataset was used to train the model, while the remaining 25% was used for testing. This algorithm provided results with 99.16% accuracy.

4.9 Prevention Mechanism

After analyzing from results from our machine learning models. We discovered that 99% of attack traffic is coming from devices t65 and t68. So we wrote a pre-insert trigger on the database table where the packets data was being stored i.e. the packets coming from devices t65 and t68 will be discarded before being stored in the database. Thus, only the legitimate packets will be kept in the database.

Chapter 5

Results

This chapter describes the results and discussion related to the results of applied models.

5.1 Results

We applied the following machine learning algorithms to identify the network traffic coming from attacker nodes.

- Support Vector Machine
- Decision Tree
- Random Forest Tree
- K Nearest Neighbors
- K- Means

All of these algorithms were then assessed against the following evaluation techniques.

- Precision
- Recall
- Accuracy
- F1-Score
- Micro Average
- Macro Average
- Weighted Average
- Root Mean Square Error (RMSE)

The summary of the results we achieved from these models is described in the below table.

Technique	SVM	Decision Tree	KNN	Random Forest Tree	K means
Accuracy	0.9724	0.9746	0.9865	0.9877	0.9916
Micro Avg	0.9612	0.9768	0.9741	0.9781	0.9819
Macro Avg	0.9615	0.9863	0.9881	0.9872	0.9946
Weighted Avg	0.9724	0.9869	0.9866	0.9877	0.9912
Precision	0.9726	0.9869	0.9866	0.9875	0.9913
Recall	0.9725	0.9869	0.9865	0.9875	0.9912
F1-Score	0.9726	0.9868	0.9864	0.9875	0.9912
RMSE	0.1661	0.1592	0.116	0.111	0.0935

Table 4: Results of Machine Learning Algorithms

Charts on the next page compare all these algorithms against each performance measure individually.

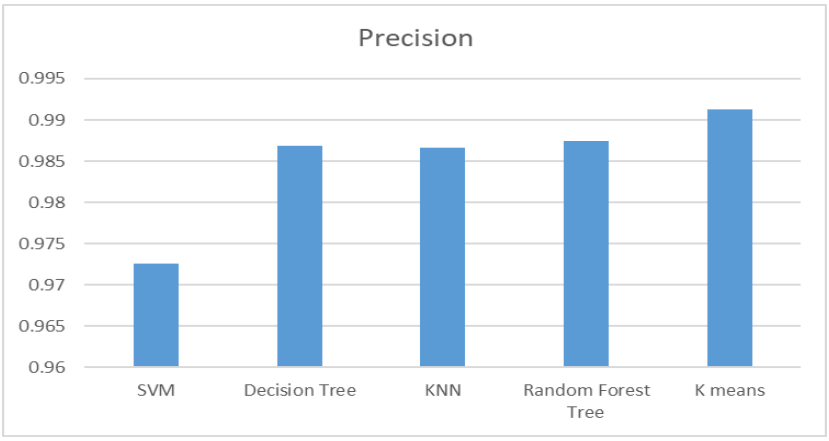

Figure 10: Precision results of applied techniques

The above chart describes, K Means provided the best results against Recall with a measure of 0.9912 followed by Random Forest Tree and Decision Tree with recall of 0.9875 and 0.9869 respectively.

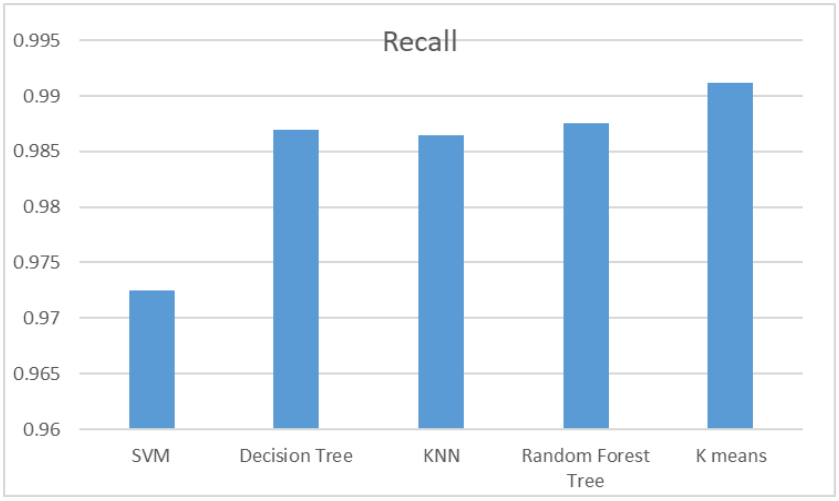

Figure 11: Recall results of applied techniques

The above chart describes, K Means provided the best results against Recall with a measure of 0.9912 followed by Random Forest Tree and Decision Tree with recall of 0.9875 and 0.9869 respectively.

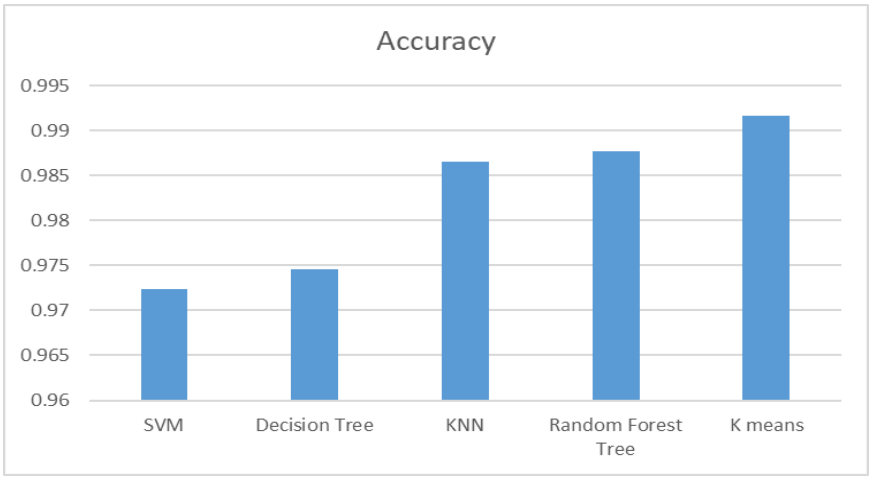

Figure 12: Accuracy results of applied techniques

The above chart describes, K Means provided the best results against Accuracy with a measure of 99.16% followed by Random Forest Tree and KNN with accuracy of 98.77% and 98.65% respectively.

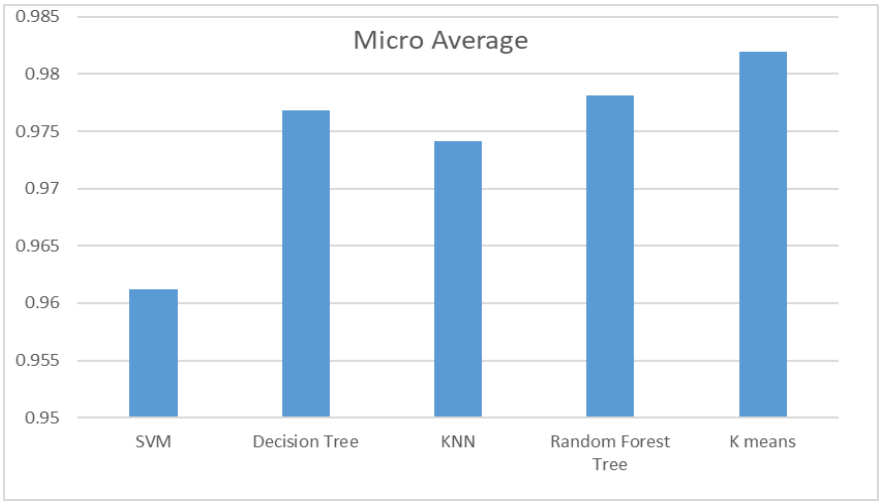

Figure 13: Micro Average of applied techniques

The above chart describes, K Means provided the best results against Micro Average with a measure of 0.9819 followed by Random Forest Tree and Decision Tree with values of 0.9781 and 0.9768 respectively.

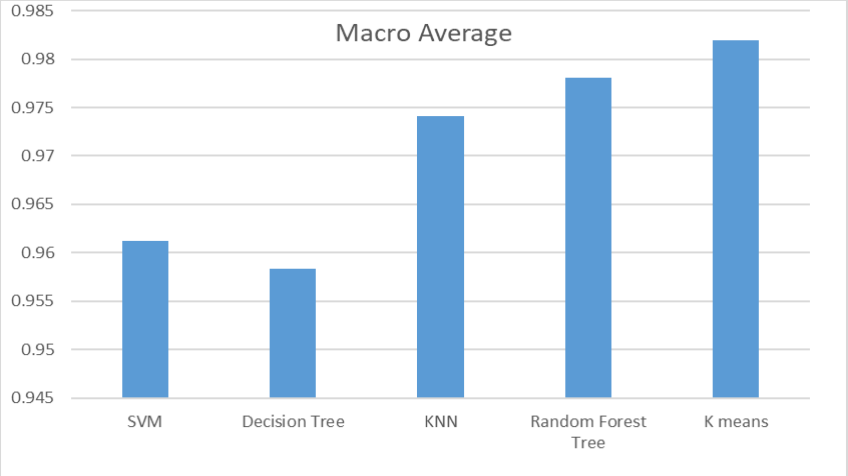

Figure 14: Macro Average of applied techniques

The above chart describes, K Means provided the best results against Macro Average with a measure of 0.9946followed by Random Forest Tree and Decision Tree with values of 0.9881 and 0.9768 respectively.

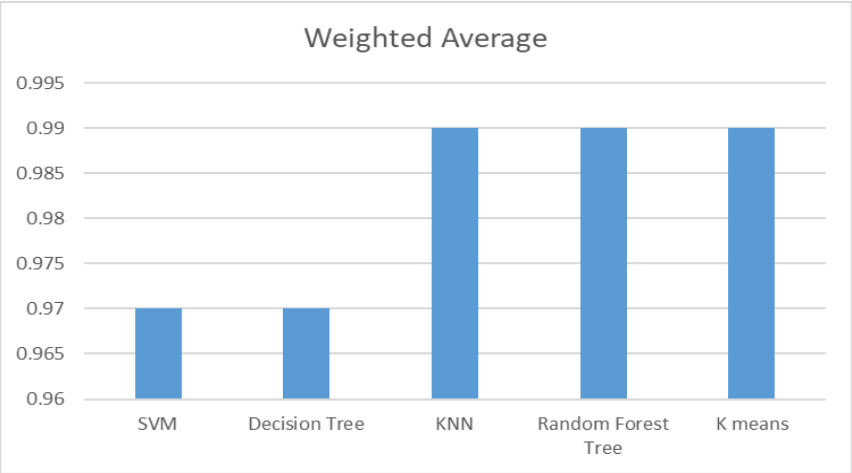

Figure 15: Weighted Average of applied techniques

The above chart describes, K Means provided the best results against Weighted Average with a measure of 0.9912 followed by Random Forest Tree and Decision Tree with values of 0.9875 and 0.9869 respectively.

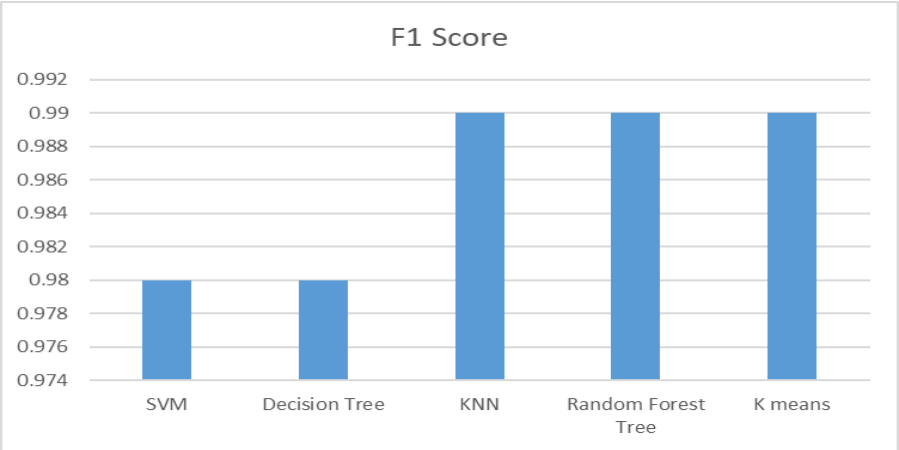

Figure 16: F1- Score of applied techniques

The above chart describes, K Means provided the best results against F1- Score with a measure of 0.9912 followed by Random Forest Tree and Decision Tree with values of 0.9875 and 0.9869 respectively.

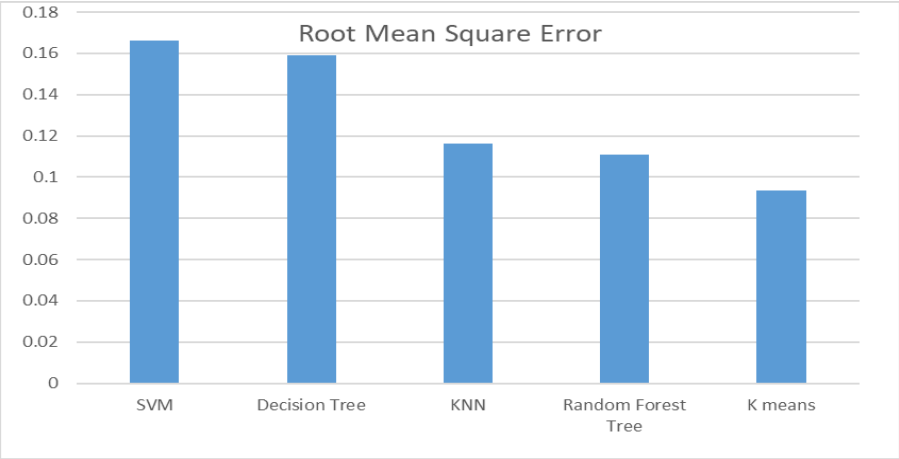

Figure 17: RMSE of applied techniques

The above chart describes, K Means provided the best results against RMSE with a measure of 0.0935 followed by Random Forest Tree and KNN with values of 0.111 and 0.116 respectively.

Chapter 6

Conclusion and Future Work

This chapter provides the discussion, conclusion and future work of the thesis.

6.1 Conclusion

In the recent years, internet of thing has attained a lot of attention. This has been raised some of new challenges, we were unfamiliar before. Challenges in IoT are quite distinctive as compared to standard internet and other wireless networks.

In this thesis, we have performed a comprehensive analysis regarding the detection and prevention techniques for replay attacks in LORAWAN proposed in the past years. All the techniques discussed have a common goal to prevent the network from the harmful activities of the attackers using as minimum resources as possible.

Over all the proposed techniques require either additional resources, external certification, or are dependent on the attackers. Our survey on the detection and prevention of Replay attacks techniques in LORAWAN reviled that it is not possible to propose such a mechanism that work best in all the scenarios and for all applications. Although many of these techniques seems promising, there are still many hurdles that need to be addressed.

We have applied a machine learning based solution that will classify the network traffic into normal malicious traffic. After detecting the traffic from intruders, the spam packets were discarded at the server's end, and thus securing the network from the attacks.

Apart from securing a LORAWAN network from replay attacks, this work has also provided a comparison of different machine learning algorithms.

6.2 Findings

- K means gave us the best results against all the performance measures.
- Overall mean for accuracy is 0.98256.
- Cumulative Standard Error against accuracy was calculative to be 0.0038.
- Overall mean for RMSE is 0.1292.
- Overall mean for F1-Score is0.9849.
- Average Precision against all models is 0.985.
- Average Recall against all models is also 0.985.

6.3 Future Work

Currently, our developed machine learning model only performs the detection of Replay attacks in LoRaWAN networks. But, as this is a machine learning based solution which learns from its past experience, so in future our work can be extended and this model can also be trained on other kinds of threats like DDOS Attacks, Battery exhaustion attack, and Jamming Attacks etc. The researcher will need to

design and perform the experiment based on the type of attack he wills to work on. Therefore, given that dataset is available for the training purpose, this idea can be applied to prevent networks from any type of attacks.

References

[1] LoRaWAN Backend Interfaces 1.0 Specification, LoRa Alliance Inc.2017

[2] Norbert Blenn and Fernando Kuipers. LoRaWAN in the wild: Measurements from the things network. arXiv preprint arXiv:1706.03086, 2017.

[3] C. Tang et al., "Comparative investigation on CSMA/CA-based opportunistic random access for Internet of Things," IEEE Internet Things J., vol. 21, no. 1, pp. 33–41, Apr. 2014.

[4] Yang, X.; Karampatzakis, E.; Doerr, C.; Kuipers, F. Security Vulnerabilities in LoRaWAN. In Proceedings of the 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), Orlando, FL, USA, 17–20 April 2018; pp. 129–140.

[5] Suman Bala, V., Barthel, D., & Gharout, S. (2019). *Separate Session Key Generation Approach for Network and Application flows in LoRaWAN* Paper session presented at 34th ACM/SIGAPP Symposium On Applied Computing (Apr 08-12: Limassol, Cyprus).

[6] SeungJae Na, DongYeop Hwang, WoonSeob Shin, Ki-Hyung Kim, "Scenario and countermeasure for replay attack using join request messages in LoRa Wan", *2017 International Conference on Information Networking (ICOIN)*, pp. 718-720, 2017.

[7] Sung, W.-J., Ahn, H.-G., Kim, J.-B., Choi, S.-G.: Protecting end-device from replay attack on LoRaWAN. In: 2018 20th International Conference on Advanced Communication Technology (ICACT), pp. 167–171. IEEE (2018).

[8] Sanchez-Iborra, R.; Sánchez-Gómez, J.; Pérez, S.; Fernández, P.J.; Santa, J.; Hernández-Ramos, J.L.; Skarmeta, A.F. Enhancing LoRaWAN Security through a Lightweight and Authenticated Key Management Approach. Sensors 2018, 18, 1833

[9] C. L. Abad and R. I. Bonilla, "An Analysis on the Schemes for Detecting and Preventing ARP Cache Poisoning Attacks," 27th International Conference on Distributed Computing Systems Workshops (ICDCSW'07), Toronto, Ont., 2007, pp. 60-60.

[10] Tsai, K.L., Huang, Y.L., Leu, F.Y., You, I., Huang, Y.L., Tsai, C.H.: AES-128 based secure low power communication for LoRaWAN IoT environments. IEEE Access **6**, 45325–45334 2018

[11] SeungJae Na, DongYeop Hwang, WoonSeob Shin and Ki-Hyung Kim, "Scenario and countermeasure for replay attack using join request messages in LoRaWAN," *2017 International Conference on Information Networking (ICOIN)*, Da Nang, 2017, pp. 718-720.

[12] Butun, I., Pereira, N., Gidlund, M.: Security risk analysis of LoRaWAN and future directions. Future Internet **11**(1), 1–22 (2019). Article 3

[13] Rohan, M. Low Power Wide Area Network Market Worth 24.46 Billion USD by 2021. Biz Journals.

[14] [SeungJae Na, DongYeop Hwang, WoonSeob Shin and Ki-Hyung Kim, "Scenario and countermeasure for replay attack using join request messages in LoRaWAN," *2017 International Conference on Information Networking (ICOIN)*, Da Nang, 2017, pp. 718-720

[15] C. L. Abad and R. I. Bonilla, "An Analysis on the Schemes for Detecting and Preventing ARP Cache Poisoning Attacks," *27th International Conference on Distributed Computing Systems Workshops (ICDCSW'07)*, Toronto, Ont., 2007, pp. 60-60.

[16] Kun-Lin TsaiEmail authorFang-Yie LeuShuo-Wen ChangJiun-Yi LinHuei-Tang Luo.; A LoRaWAN Based Energy Efficient Data Encryption Method.Advances in Intelligent Systems and Computing book series (AISC, volume 994)

[17]Kim, J., Song, J.: A dual key-based activation scheme for secure LoRaWAN. In: Wireless Communications and Mobile Computing, vol. 2017, pp. 1–12, Nov 2017. Article 6590713

[18]W. Du, J. Deng, Y. S. Han, and P. K. Varshney. A pairwise key pre-distribution scheme for wireless sensor networks. In ACM CCS 2003, pages 42–51, Oct. 2003

[19]Mohamed Eldefrawy, Ismail Butun, Nuno Pereira, Mikael Gidlund, Formal security analysis of LoRaWAN, Computer Networks, Volume 148,2019, Pages 328-339, ISSN 1389-1286

[20]E. V. Es, H. Vranken, A. Hommersom, "Denial-of-service attacks on LoRaWAN", *Proc. 13th Int. Conf. Availability Rel. Secur. (ARES)*, pp. 17, Aug. 2018.

[21]R. Doshi, N. Apthorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," *2018 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, 2018, pp. 29-35.

[22]C. Yin, Y. Zhu, J. Fei and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," in *IEEE Access*, vol. 5, pp. 21954-21961, 2017. doi: 10.1109/ACCESS.2017.2762418

[23]Taeshik Shon, Jongsub Moon.; A hybrid machine learning approach to network anomaly detection, Information Sciences, Volume 177, Issue 18,2007, Pages 3799-3821

[24]F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui and T. Watteyne, "Understanding the Limits of LoRaWAN," in IEEE Communications Magazine, vol. 55, no. 9, pp. 34-40, Sept. 2017. doi: 10.1109/MCOM.2017.1600613

[25]B. Wickramasinghe, E. Ambikairajah, J. Epps, V. Sethu and H. Li, "Auditory Inspired Spatial Differentiation for Replay Spoofing Attack Detection," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 6011-6015.

[26] Ling, Chung-Huei et al. “A Secure and Efficient One-time Password Authentication Scheme for WSN.” I. J. Network Security 19 (2017): 177-181.

[27]M. Liu, L. Wang, J. Dang, S. Nakagawa, H. Guan and X. Li, "Replay Attack Detection Using Magnitude and Phase Information with Attention-based Adaptive Filters," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 6201-6205.

[28]Gu, Chaojie, Tan, Rui, Huang, and Jun, “Secure Data Timestamping in Synchronization-Free LoRaWAN,” arXiv.org, 05-May-2019. [Online]. Available: https://arxiv.org/abs/1905.01679.

[29]T. Gunendradasan, S. Irtza, E. Ambikairajah and J. Epps, "Transmission Line Cochlear Model Based AM-FM Features for Replay Attack Detection," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 6136-6140.

[30]I. Butun, N. Pereira, M. Gidlund, Analysis of lorawan v1.1 security: Research paper, in: Proceedings of the 4th ACM MobiHoc Workshop on Experiences with the Design and Implementation of Smart Objects, in: SMARTOBJECTS ’18, ACM, New York, NY, USA, 2018, pp. 5:1–5:6, doi: 10.1145/3213299.3213304 .

[31] S. Tomasin , S. Zulian , L. Vangelista , Security analysis of lorawan join procedure for internet of things networks, in: Wireless Communications and Networking Conference Workshops (WCNCW), 2017 IEEE, IEEE, 2017, pp. 1–6 .

[32] J. Kim , J. Song , A dual key-based activation scheme for secure lorawan, Wireless Communications and Mobile Computing 2017 (2017) .

[33] B. Blanchet , et al. , An efficient cryptographic protocol verifier based on prolog rules., in: csfw, 1, 2001, pp. 82–96 .