SELINUS UNIVERSITY
OF SCIENCES AND LITERATURE

# Intrusion Detection System for Internet of Things (IDS for IoT)

By Rupendra Man Rajkarnikar

## A DISSERTATION

Presented to the Department of
Information Technology & Security
program at Selinus University

Faculty of Computer Science
in fulfillment of the requirements
for the degree of Doctor of Philosophy
in Information Technology & Security

2023

# Acknowledgement

First of all, I would like to express my sincere thanks and gratitude to my advisor, Prof. Dr. Salvatore Fava, whose invaluable advice, inspiration, support and enthusiasm have made this work possible. Prof. Dr. Salvatore Fava provided step-by-step instructions from the project development to thesis structure and grammar correction. The proponent would have been lost without him.

I would like to give special thanks to my father (Tulsi Man Rajkarnikar), mother (Karna Laxmi Rajkarnikar), my wife (Suhana K. C Rajkarnikar), daughter (Nista Rajkarnikar), Son (Rupsan Rajkarnikar), Dr. John Hummel, Mrs. Monique Beun, my friends and families as a whole for their continuous support and understanding when undertaking research and writing my thesis as well as providing a loving environment. I have gotten by so far thanks to your prayer for me.

I would like to give special thanks to all the members of Selinus University for their kind support.

Lastly, I want to express my gratitude to God for guiding me every day despite all of the challenges. All of you are the ones who let me finish my Ph. D. I'll continue to have trust in you for the future.

Rupendra Man Rajkarnikar
Regd. No: UNISE1786IT

# Index

# List of Abbreviations

| | | |
|---|---|---|
| IEEE | : | Institute of Electrical and Electronics Engineers |
| CSMA/CA | : | Carrier Sense Multiple Access/Collision Avoidance |
| LAN | : | Local Area Network |
| WAN | : | Wide Area Network |
| WLAN | : | Wireless Local Area Network |
| Wi-Fi | : | Wireless Fidelity |
| IoT | : | Internet of Things |
| IDS | : | Intrusion Detection System |
| CPS: | : | Cyber Physical System |
| RFID | : | Radio-Frequency Identification |
| Web App | : | Web Application |
| OBE | : | On-Board-Equipment |
| OLED | : | Organic Light-Emitting Diode |
| ABS | : | Anomaly Based System |
| GRC | : | Governance, Risk and Compliance |
| SIEM | : | Security Information and Events Management |
| Email | : | Electronic Email |
| TCP/IP | : | Transmission Control Protocol / Internet Protocol |
| OFDM | : | Orthogonal Frequency Division Multiplexing |
| ITS | : | Intelligent Transportation Systems |
| GPS | : | Global Positioning System |
| OOK | : | On Off Key |
| VR | : | Virtual Reality |
| AR | : | Argument Reality |
| LLC | : | Logical Link Control |
| MAC | : | Media Access Control |
| QoS | : | Quality of Service |
| DS | : | Distributed System |
| ToDS | : | To Distributed System |
| FromDS | : | From Distributed System |
| NAV | : | Network Allocation Vector |
| IBSS | : | Independent Basic Service Set |
| AP | : | Access Point |
| SA | : | Source Address |

|  |  |  |
|---|---|---|
| DA | : | Destination Address |
| RA | : | Receiver Address |
| TA | : | Transmitter Address |
| WDS | : | Wireless Distribution System |
| BSS | : | Basic Service Set |
| BSSID | : | Basic Service Set Identifier |
| AID | : | Association Identifier |
| SSID | : | Service Set Identifier |
| TBTT | : | Target Beacon Transmit Time |
| TID | : | Traffic Identifier |
| EOSP | : | End of Service Period |
| ACK | : | Acknowledgment |
| TXOP | : | Transmit Opportunity |
| PS | : | Power Save |
| ELBA-IoT | : | Ensemble Learning Model for Botnet Attack – Internet of Things |
| RF | : | Random Forest |
| SDN | : | Software Defined Network |
| NIDS | : | Network Intrusion Detection System |
| SVM | : | Support Vector Machines |
| ANN | : | Artificial Neural Networks |
| OWASP | : | Open Worldwide Application Security Project |
| DT | : | Decision Tree |
| VM | : | Virtual Machine |

# List of Figures

# List of Tables

# Abstract

Cyber-attacks are becoming more hi-tech threats now that could disable computers, steal, alter and destroy information as well as the Internet of Things (IoT). Every day the devices are increasing with a large diversity of size, shape, usage as well as complexity. Nowadays IoT devices are driving the world and changing the lifestyle of the people that provided numerous services through applications. IoT is the network of physical devices from various sectors (Like Government, Household, Transportation, Banks, Agriculture, Healthcare, Energy and so on) with hardware, software, sensors and network connectivity that allow these items to gather and send data over the internet.

Most people are vulnerable to many security flaws and attacks, including denial-of-service attacks (DDOS), vulnerabilities in Internet of Things (IoT) devices, inadequate testing and updating, subpar IoT, remote access, AI and automation, and sink holes. It has become challenging to the users because it could take down the credibility of security services such as data Confidentiality, Integrity and Availability (CIA). In this regard, an Intrusion Detection Systems for Internet of Things (IoT) will be proposed in the literate study to tackle computer threats. Data and device security is the practice of protecting digital information from unauthorized access and protection from unauthorized attacks to enable a response to that violation.

**Keyboards**: Cyber-security, Intrusion detection, Internet of Things, AI.

# Chapter 1

## Introduction

### *1.1 Internet and digital society*

The Internet's history goes back some decades by now – email has been around since the 1960s, file sharing since at least in 1970s, and TCP/IP was standardized in 1982. Over the last few decades, the Internet has grown and consolidated itself as a very powerful platform that has forever changed the way we do business and the way we communicate. The internet has become the Universal source of information for billions of people, at home, at school and at work. Internet user's distribution in the word -2021 statistics as shown below.

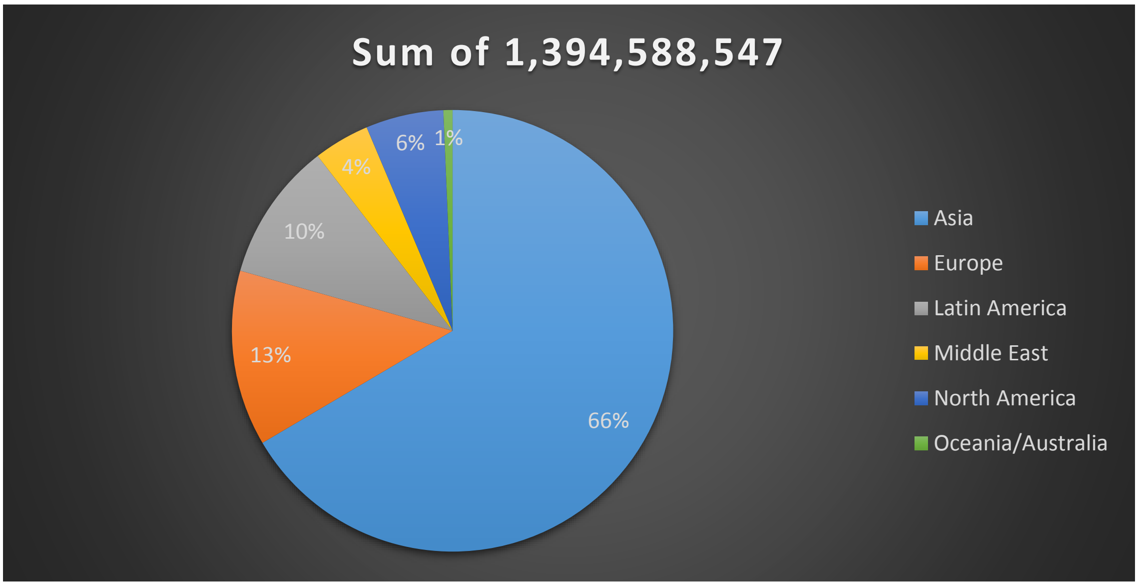


Figure 1 1: Internet user's distribution - 2021

A digital society is one that has adopted and integrated information and communication technologies into daily life, including work, play, education, and the home. New innovations which are really good to adopt and integration of advanced technologies will reshape our industry, economy, culture and society. Internet, cloud, mobile and big data technology offer big opportunities, improvement of living efficiency to many areas including transportation, education, energy, hospital, agriculture, manufacturing, retail and public administration. The emerging technologies that are responsible for developing a true Digital Society include ICT (Information and Communication Technology), information science and computing, business studies, commerce, humanities and social science. The primary focus of the digital society is on extremely sophisticated wireless and telecommunications technologies and solutions. It depends on the Digital Economy which is one of the emerging concepts of economic development with proper support from IT technologies and digital tools. It depends on technology and knowledge with digital products. Advantages of Digital Society are:

- Digital society enhances social connectivity making it easier to stay in touch with friends, family and even cam work remotely. Social media, messaging and chatting apps, smartphones keeps you updated and connected with the world and also provides ways to learn for people with disabilities.
- There are many instances where digital technology positively affects our lives, for example, learning opportunities, automation, information storage, communication speeds, editing, transportation, GPS and mapping, and much more.
- Digital technology is making machines smarter as they are automated and no longer need humans to operate. Smarter machines mean better standards of safety and a better user experience.
- Nowadays devices can be more compact, faster, lighter, and even more versatile. Therefore, it’s easy to access to the internet and then you are able to communicate on a global level and not just through text, but through videos, music, and other media.
- Work culture has changed as remote working becomes increasingly common. People can collaborate and work from different locations with the use of digital technology.

Information Security is basically the practice of preventing data from unauthorized users, disclosure, disruption, modification, inspection, defamation, recording or destruction of information. There were more than 15 million internet users and 3,11,34,363 mobile users in Nepal in January 2020. The Internet is a worldwide computer network that can be accessed through computers, mobile telephones, PDAs, games machines, digital TVs, etc. Therefore, individuals and organizations can reach any point on the internet without any regard to national or geographic boundaries or time of day.

Information can be electronic or physical. Information can be anything like personal details or your profile on social media, your data in mobile phone, ATM pins.

Therefore, a wide range of academic fields are addressed by information security, including cryptography, mobile computing, cyber forensics, online social media, etc.

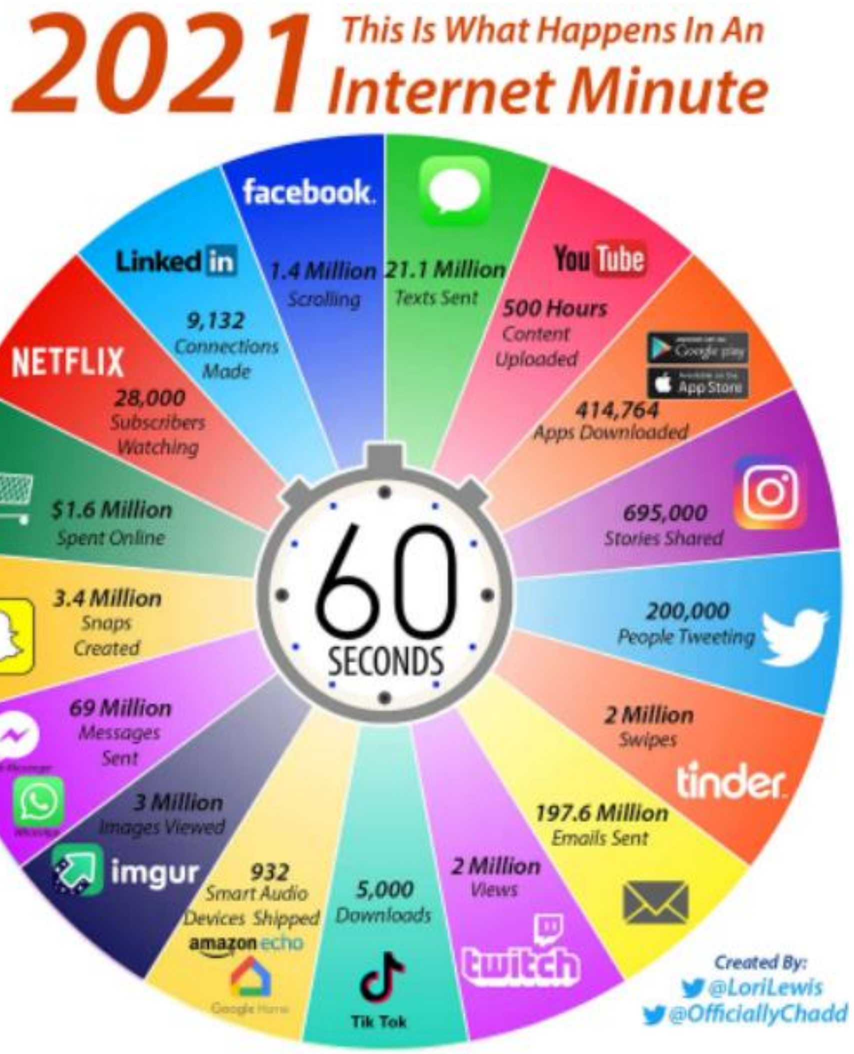


Figure 1 2: Internet Minutes - 2021

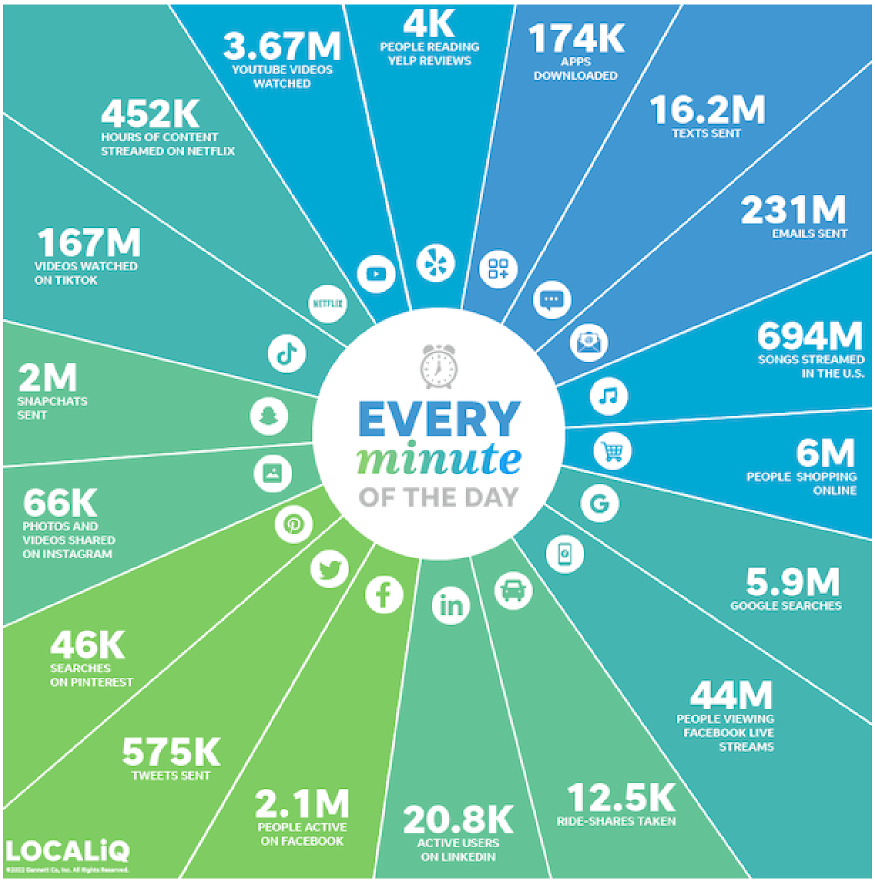


Figure 1 3: Internet Minutes - 2023

There are three basic Information Security concepts.

1. Confidentiality
2. Integrity
3. Availability

**Confidentiality:**

Confidentiality refers to when information is read and copied by someone who is not authorized to do so then it will be loss of confidentiality. For example: if I have a password for a Hotmail account but someone saw while I was typing then in that case my password has been compromised and confidentiality has been violated. Some measures to keep your information confidential are:

- Encryption
- Password
- Two-factor authentication
- Biometric

**Integrity:**

Information can be corrupted or manipulated if it’s freely available on an insecure computer or network then it will be loss of integrity. It means maintaining accuracy and completeness of data. It measures to protect information from unauthorized alteration. So, Data must not be changed in transit, and precautionary steps must be taken to ensure that data cannot be altered by unauthorized people.

Some security controls designed to maintain the integrity of information include:

- Encryption
- User access controls
- Version control
- Backup and recovery procedures
- Error detection software

**Availability:**

Information can be accessed and erased or inaccessible resulting in loss of availability. It means information must be available whenever needed. For example, if the director of a company wants to view details of an employee then it should be available to access details about him/her. Information security measures for mitigating threats to data availability include:

- Off-site backups
- Disaster recovery
- Redundancy
- Failover
- Proper monitoring
- Environmental controls
- Virtualization
- Server clustering
- Continuity of operations planning

The modern internet supports a wide variety of features and services such as cloud computing, social networking, content delivery services, blogs, online banking, e-marketing and e-shopping. Deployment of these technologies has made the internet a household commodity. Despite the development of the internet, technologies relating to sensors, wireless communications, and mobile computing have seen an unprecedented growth, which has contributed to the introduction of the new paradigm which is called Internet of Things (IoT). The Internet of Things (IoT) is the concept of connecting any device (so long as it has an on/off switch) to the Internet and to other connected devices. The Internet of Things (IoT) is a vast network of interconnected objects and people that gather and exchange data about their usage and the surroundings. Although only computers and servers were able to communicate with each other by sharing data and information in the past, any digital (computer-like) device can be enabled to share data via the Internet. Therefore, devices that can sense/ share the data from their surroundings via application for data communication with the help of the Internet are called Internet of Things (IoT). For example, a motorbike with a GPS sensor (thing) continuously transmits its location information to any user via the Internet. With the advent of hardware and software development of digital technology over the past few years, IoT has become one of the essential technologies of the 21st century. Therefore, almost all digital appliances, including kitchen appliances, cars, thermostats, baby monitors, etc. are connected to the Internet as IoT. Many devices can share the same data over the Internet thanks to cloud computing, mobile technologies, big data, analytics, and low-cost computing. Smart devices are proliferating our daily lives rapidly, wherein wearable technology (see Figure 1.4) like smart watches, smart ring, smart shoes, smart glasses, pants, belt and other devices like smartphones, smart refrigerators, smart cars, light, curtains etc. are becoming an integral part of our lives.

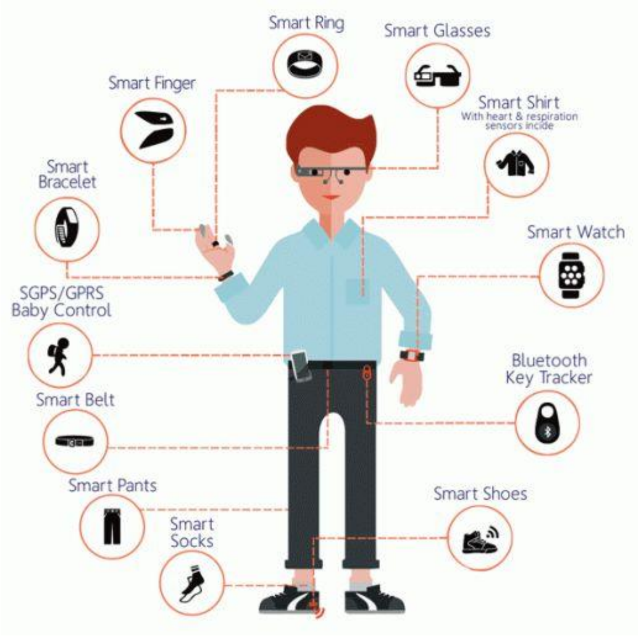


Figure 1 4: Smart wearables

The core technologies of IoT are not new where sensing technologies have been used since a few years ago on factory manufacturing floors for monitoring and tracking livestock, etc. Since machine-to-machine communication is the foundation of the

Internet, the concept is obviously not new. IoT is an evolution in the use of these technologies in terms of number of devices, the type of devices and the services provided via these devices. IoT promises, it will lead to the development of a new generation of devices that will provide personalized services; services that are tailored and modified to each user's needs and demands.

The enterprise IoT market grew 22.4% to $157.9 billion in 2021, according to the March 2022 update of IoT Analytics' Global IoT Enterprise Spending Dashboard. The market grew slightly slower than the 24% forecasted last year due to several factors, including a slower-than-anticipated overall economic recovery, a lack of chipsets, and disrupted supply chains. North America was the fastest-growing region in 2021 (+24.1%), and process manufacturing was the fastest-growing segment (+25%).

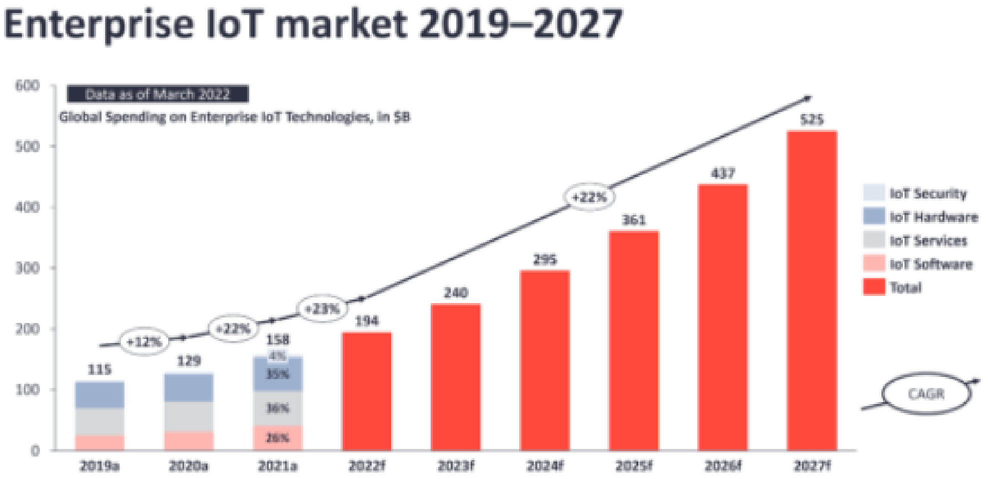


Figure 1 5: growth of IoT by 2027

At this point, IoT Analytics forecasts the IoT market size to grow at a CAGR of 22.0% to $525 billion from 2022 until 2027. Compared to last year, the five-year prediction is lower. More than expected, a number of growth headwinds have had a significant impact, including labor shortages, particularly for highly sought-after software jobs, and supply disruptions and shortages (most notably, chip shortages, which are now predicted to last well into 2024 and possibly even beyond). The Internet of Things (IoT) is still a hot technology topic, with numerous projects entering the rollout phase despite the revised growth estimates. By the end of 2022, there will be an estimated 14.5 billion connected IoT devices worldwide.

Cyber Crime is the criminal activities carried out by means of computers or the Internet. Cybercrime especially occurs through the internet which has grown in importance as the computer has become central to ecommerce, entertainment, education, social network, government as well as Internet of Things (IoT). Organizations can profit from the Internet of Things in several ways, including the following:

Monitor their overall business processes

- Improve the customer experience
- Save time and money
- Enhance employee productivity
- Integrate and adapt business models
- Make better business decisions
- Generate more revenue.

The meaning of computer crime varies based on the context, the individual, and their personal viewpoint. Various techniques are employed by cybercriminals, contingent upon their objectives and skill set. Malicious software is also known as "Malware" which poses a critical challenge to the design security system like Intrusion Detection Systems (IDS). Some well-established malware attacks such as breaches, supply chain, ransomware, DDOS, Crypto-jacking etc. Example, when ransomware attacks then it will lock your files and computer, demand money and threaten to destroy your data so, once payment is made then they will release your data. The Malware authors use different bypassing techniques for information concealing to prevent detection by an Intrusion Detection Systems and increased day per day security threats. Therefore, computer security has become one of the most essential parts of our daily lives. Hence, in every company, organizations, institutions, private and public sectors acquire Intrusion Detection Systems as part of their security systems. Common applications of the IoT include smartphones, wearable, smart city, smart grids, industrial internet, connected car, connected health, small retail, smart supply chain, smart farming.

Furthermore, the attack propagation time used to range in weeks in the 1980s to now ranges to a fraction of a second to reach a sizable number of Internet enabled devices as shown in Figure 1.6.

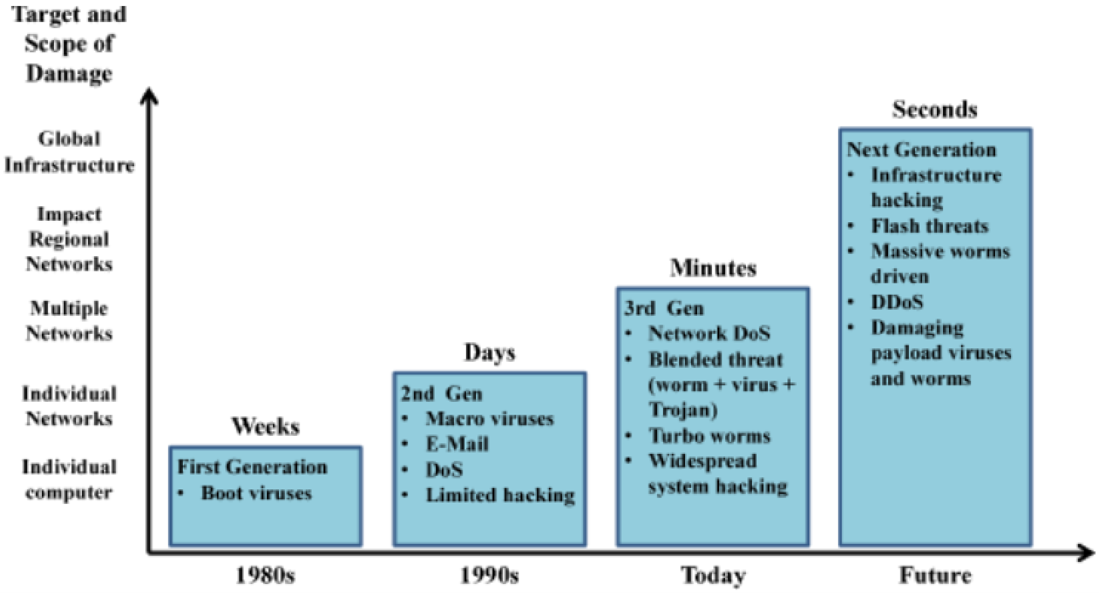
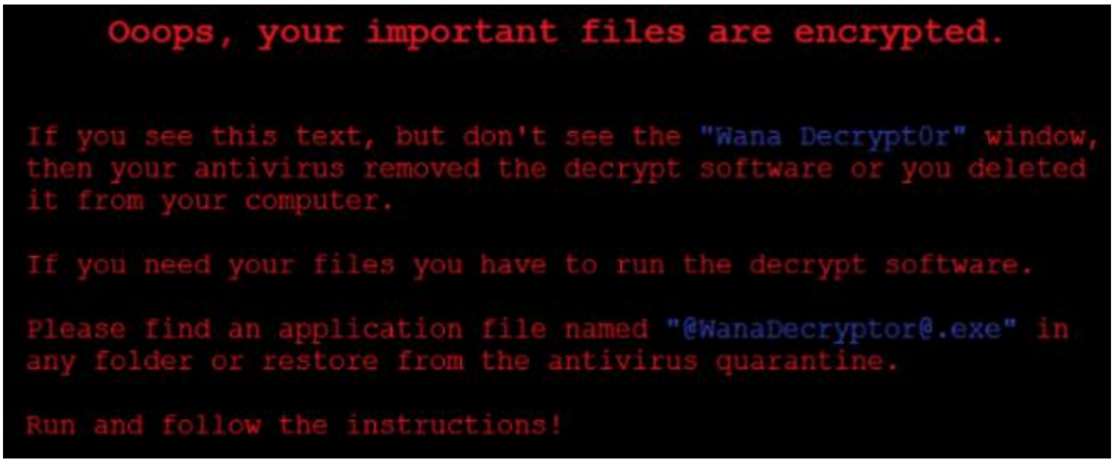


Figure 1 6: Timeline of cyber threads

As an example, in May 2017 Early on Friday morning, hospitals and businesses around Europe, Asia, and the United States were hit by a widespread ransomware attacks. According to reports it shows a warning message on the wallpaper as shown in figure 1.7.



```
Ooops, your important files are encrypted.

If you see this text, but don't see the "Wana Decrypt0r" window,
then your antivirus removed the decrypt software or you deleted
it from your computer.

If you need your files you have to run the decrypt software.

Please find an application file named "@WanaDecryptor@.exe" in
any folder or restore from the antivirus quarantine.

Run and follow the instructions!
```

Figure 1 7: Warning Message

Companies in more than 70 countries have reported incidents as of Friday afternoon. The attacks are being caused by ransomware called “WannaCry,” which quickly moves across systems to encrypt large amounts of computer data. Ransom demands seen during the current attack have requested Bitcoin amounts that equal between $300 and $600 in return for the decryption key. According to security researchers, the ransomware exploits a vulnerability in Microsoft’s Windows operating system that was disclosed in an April leak of NSA spying tools. Confirmed targets of the attack include Telefonica, Spain’s largest telecommunications provider, and the National Health System (NHS) in the United Kingdom. This attack would have impacted more systems at even a faster rate if aimed at the deployed IoT devices and sensors.

### *1.2 Statement of the problem*

Since everyone has free access to the internet and computers have connected globally. The guidelines of cyber laws are there and these privileges should not be abused since users have invested a lot to make them available well at all times. however, it is difficult to secure our data. Out data and information is at risk by misusing it in different ways such as:

- Security breach
- Supply chain
- Ransomware
- DDOS attacks
- Crypto-jacking
- Spyware, adware, virus.
- Network key policy violation
- Internet of Things (IoT)
- Black hole attack
- False data attack etc
- Hospital Radiology Med jack
- Black Energy Trojan Strikes Again
- Passenger Jets at risk of cyber attack
- Hackers attack a light bulb.
- Hackers attack emergency vehicle etc.

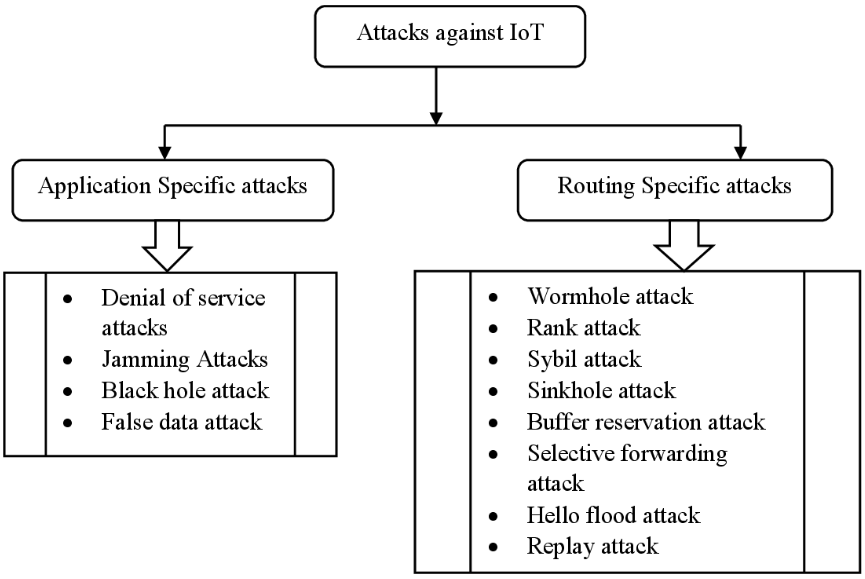


Figure 1 8: attacks against IoT

Since data is very costly and IoT is sensitive, handling it in a proper way has become a challenging job. System administrators have to be alert when they assign to track accessing of data/devices by unauthorized users and they should be able to identify their purpose to avoid its misuse. Including all these difficulties, an administrator should also be able to monitor the real-time network traffic and activity logs.

Because of this, the proponent is determined to conduct a study Intrusion Detection System (IDS) for Internet of Things (IoT) that will help all private/offices/departments to monitor attackers due to every day the device count increases with large diversity of size, shape, usage and complexity. IDS have many advantages to prevent such activities on a network. Whenever we use IDS, generally it checks and matches up to a previously detected attack signature and then it reports the activities to the console. It can notify security personnel of infections, key loggers, spyware, leakage of information, unauthorized access as well as misconfiguration of key policy and violations. IDS are widely used in the internet.

### *1.3 Problem statement*

At present, with the growth of the internet, cloud computing and IoT, smart devices have become widely used in our daily lives. These smart devices support a wide variety of services and applications, which has resulted in a plethora of networking layers and protocols.

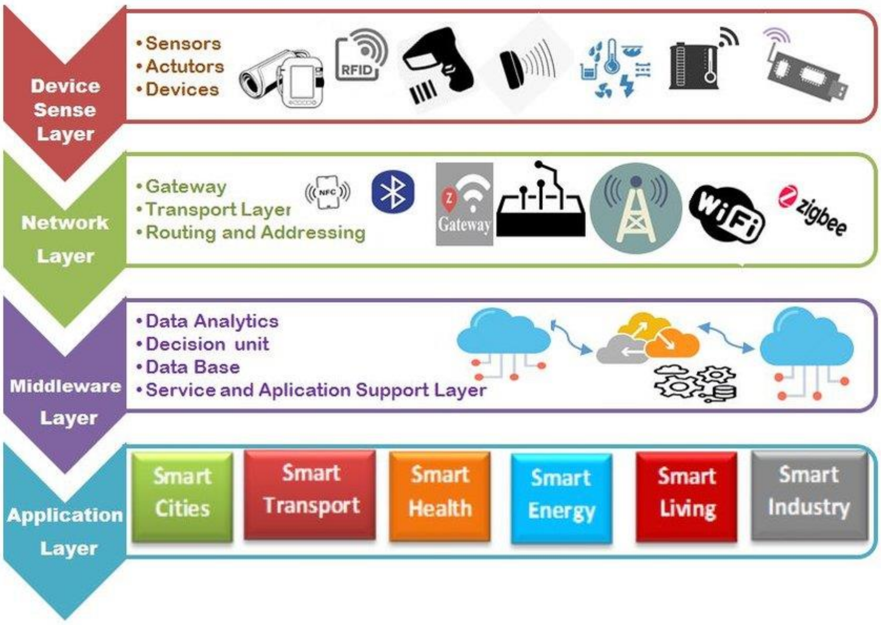


Figure 1 9: IoT network layers and protocols

Figure 1.10 shows a subset of protocols deployed on the internet. Each of these protocols are vulnerable and can be easily attacked so, need to be secured, however we can design a multi-level intrusion detection system (ML-IDS) as shown in figure 1.9 to secure the computer networks which will protect individual protocols and perform decision fusion to detect threats on the whole network devices.

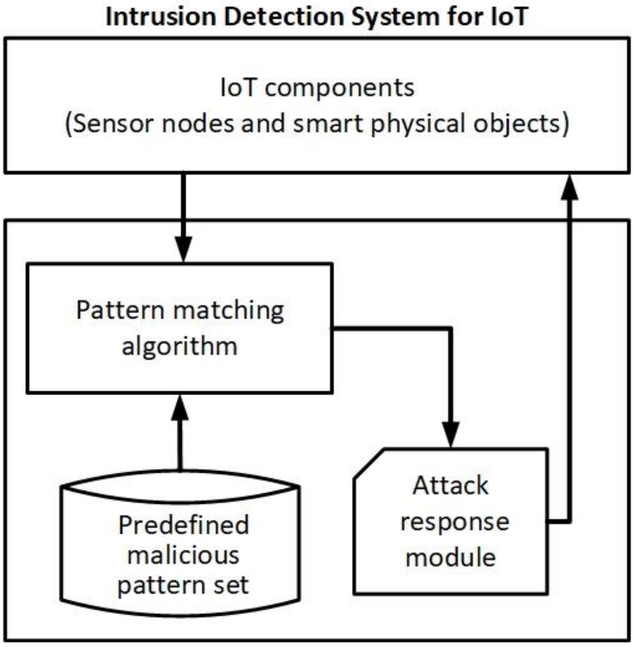

Intrusion Detection System for IoT
IoT components
(Sensor nodes and smart physical objects)
Pattern matching algorithm
Predefined malicious pattern set
Attack response module

Figure 1 10: Protecting IoT devices

This research extends the work and presents a methodology that helps in developing micro intrusion detection systems to secure each protocol used by IoT devices. The proposed methodology has 3 steps.

1. Threat modelling analysis
2. Behavior of protocols and
3. Develop machine learning models that characterize accurately the normal behavior.

Due to the growth of IoT systems, it is critically important to formalize this methodology to design IDS for IoT to secure each protocol.

### *1.4 Objective of the study*

The main objective of the study is to secure their data and devices with the help of IDS for IoT. The research is to design a general methodology that can streamline the development process of IDS for different protocols. The following highlighs are the specific goals of this research.

- Develop a methodology to design specialized IDS that can detect attacks on their protocols. This will help to detect malicious activities that trigger abnormal behavior including zero-day attacks.
- Develop innovative data structures for each protocol that will be used to accurately and efficiently characterize the normal behaviors of each protocol. These footprint data structures will lead to significant reduction of the amount of data that needs to be monitored and stored in the database in real-time. This will realize normal and abnormal operations of different networking protocols.
- Evaluate the performance of the developed IDS on the Wi-Fi, DNS and the HTML protocols. In this evaluation, different performance statistics like accuracy, false positive and false negative will be used.

The Conceptual Model

The conceptual model consists of the three major phases namely: Input, Process and Output.

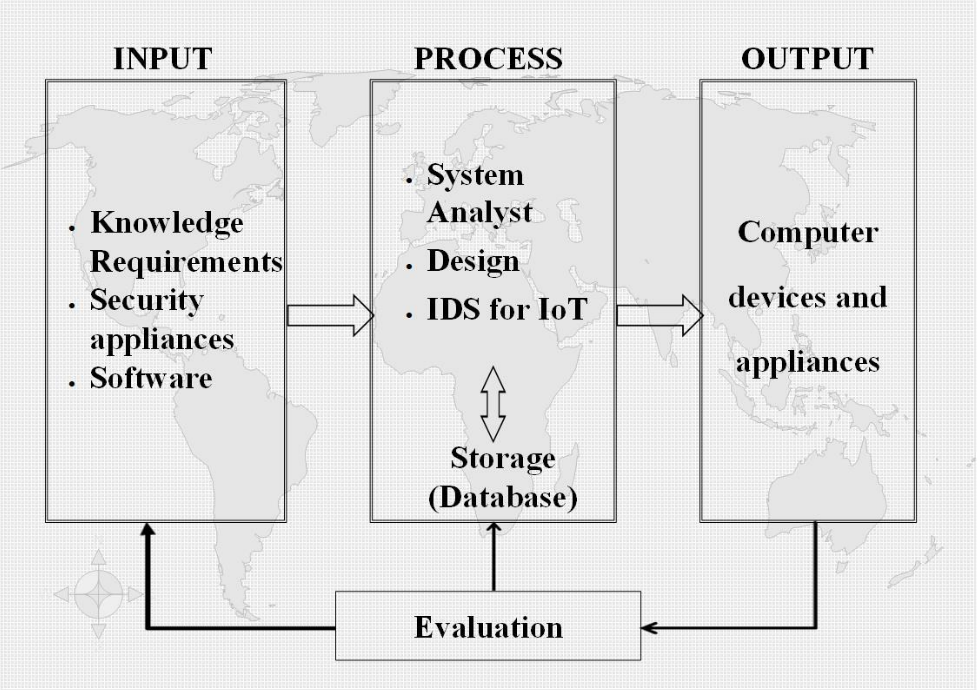


Figure 1 11: The Conceptual Model of IDS for IoT

### *1.5 Significance of the study*

This research study aims to enhance the smart devices with embedded processors, sensors and communication to enhance and collect data and send data from different environments. The devices connected to the IoT hub share the data that they collect and analyze locally. It may connect with other devices and share the information they get from each other. It works without any human interaction. Mainly it works in three ways. 1 Collect data 2. Transfer the data and 3. Analyze and take action.

### *1.6 Scope and Limitation*

Scope

- Connect devices in the various systems to the internet.
- It can be controlled from anywhere.
- It provides efficiency of data.
- It provides safety and security.

- It can be implemented in most of the things.

Limitation

- Hackers can hack the user data and private information.
- Encryption with IoT devices can be difficult with a large fleet of devices.

# Chapter 2
## Review of Related Literature and Studies

### *2.1 Introduction*

A methodology to develop anomaly based Intrusion Detection Systems (IDS) for different Networking protocols. In this chapter I present the background for this research approach and begin by discussing the IDS taxonomy presented by Axelsson. Will discuss here the types of intrusion detection systems, characteristics of intrusion detection systems and provide a brief overview of the machine learning algorithms. According to Axelsson intrusion detection systems are of three major types: Anomaly based intrusion detection systems, signature-based intrusion detection systems, and compound intrusion detection systems.

### *2.2 ANOMALY BASED INTRUSION DETECTION SYSTEMS*

According to Stefan Axelsson an, an anomaly-based intrusion detection system is designed to model the normal behavior of the system and makes an inherent assumption that any attack will lead to an anomalous behavior. That means if we could establish a normal activity profile for a system then theoretically flag all system states verifying from the established profile by statistically significant amounts as intrusion attempts. [1]

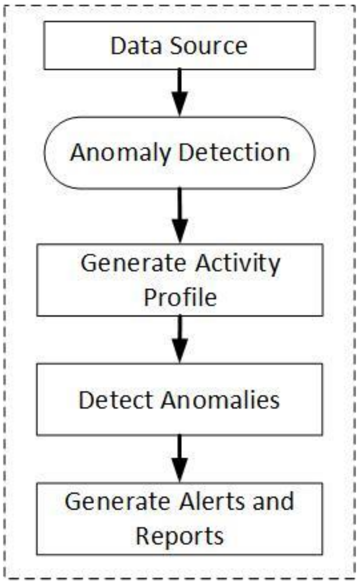


Figure 2 1: Anomaly Detection Systems

Designing and implementation of anomaly-based intrusion detection systems begins with collection of information on what constitutes normal behavior for the system that is called the “target system”, or the “target”. Anomaly based intrusion detection systems use this understanding of the normal behavior of the target to build models that allow classification of harmful anomalies from the normal behavior. Anomaly based intrusion detection systems can be categorized into two. These are Self learning and Programmed (supervised).

### *2.2.1 Self learning*

Self-learning systems refers to learning the normal behavior on their own. It observes the target at runtime and has the capabilities to judge and extract features that are characteristics of the target’s normal behavior. This system builds models that incorporate these characteristics. The systems that fall in this category may use different approaches to model the normal behavior of the system. Artificial neural networks and clustering algorithms may have been used to build such systems. For example, service providers like Microsoft, Amazon or Google that aggregates device-type specific anomaly detection models trained by all Security Gateways in the system. The systems identify its device type and retrieve the corresponding anomaly detection model for this type from IoT security service and detects abnormal communication behavior that is potentially caused by malware based on anomaly detection models that trains locally and which are aggregated by the IoT service to a global detection model.

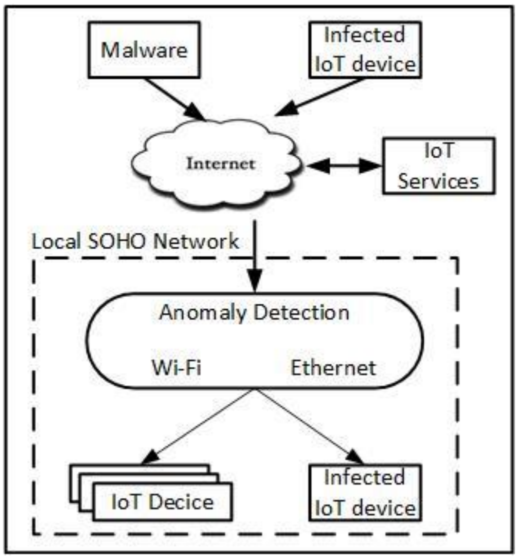


Figure 2 2: Self-Learning Systems

### *2.2.2 Programmed/Supervised Learning*

The programmed class requires someone or a third party, who teaches the system to program it and then to detect certain anomalous events. As a result, the system user develops an opinion about what defines abnormal enough for the system to alert the user to a security breach. This is generally done by feeding the system with different parameters that have statistical values that help decide if the system is operating normally or not.

## *2.3 Signature based intrusion detection systems*

Basically, signature based intrusion detection systems work on the basis of the specific patterns such as number of bytes or number of 0's or 1's in the network traffic. Additionally, it makes the detection based on the malware's known malicious instruction sequence. Thus, those detected patterns in the IDS are known as signatures. It is very easy to detect signature based IDS because it's already captured in their database and then it will give a message or alert whereas it is quite difficult to detect the new malware attacks as their pattern or signature is not known and not available in their database and then it will not take any action.

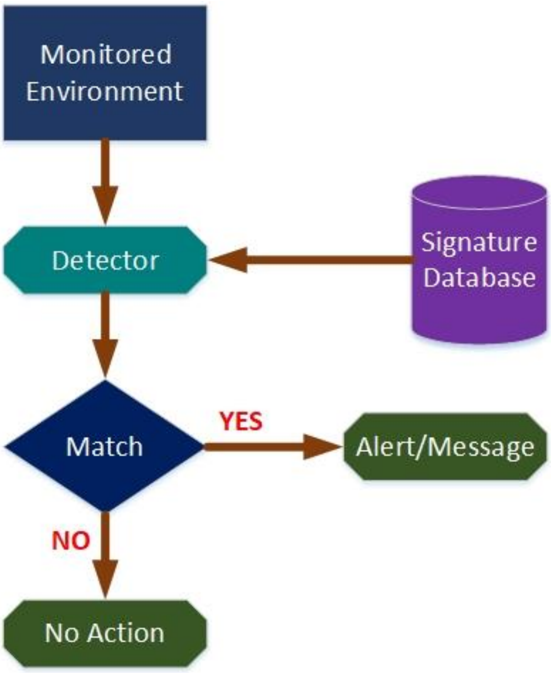


Figure 2 3: Signature based IDS

### *2.4 Compound Intrusion Detection Systems*

Generally, it uses signature-based detection on normal traffic. It is a composite of a signature-based or anomaly-based intrusion detection system.

### *2.5 SYSTEM CHARACTERISTICS OF INTRUSION DETECTION SYSTEMS*

System characteristics are independent type of detection system that performs following:

#### *2.5.1 TIME OF DETECTION*

It can be real time that checks for intrusion at runtime and responds to attacks in a timely manner or non-real time intrusion detection systems analyze network traffic offline and can run very sophisticated models to improve detection and maintain integrity.

#### *2.5.2 GRANULARITY OF DETECTION*

Granularity of detection system is the smallest unit of data that is processed by IDS can process data continuously or in small groups.

#### *2.5.3 SOURCE OF AUDIT DATA*

Source of audit data is the either network packets tapped directly from the network interface card or system logs that are maintained by the operating system.

#### *2.5.4 Active Response Systems*

It is a system that responds to detection of an intrusion on the network by creating an alarm and then taking counter measures against the detected attack and the counter measures range from closing of the network connections to even attacking the resources used by the attackers. Once response systems detect an anomalous network situation depending upon the configuration policy it responds to alerts. When someone strikes on the cards, the alarm is triggered and then the car owner is notified.

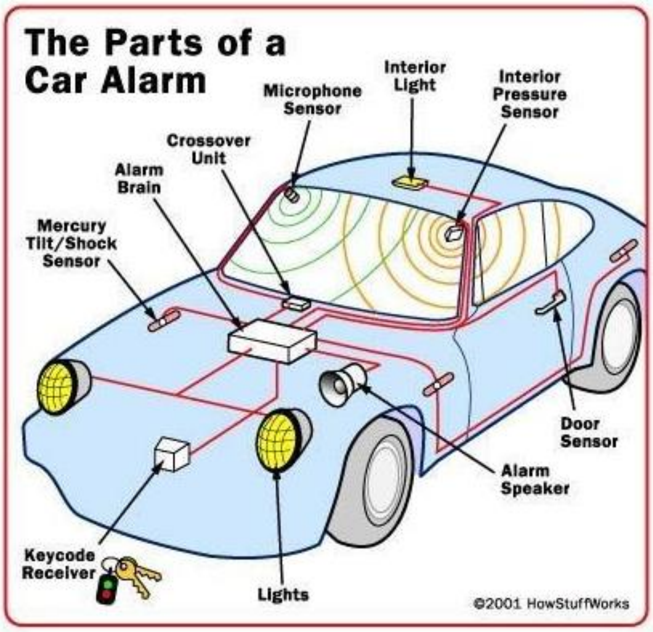


Figure 2 4: Vehicle alert system

*2.5.5 Passive Response Systems*

Passive response systems are Intrusion Detection Systems that respond to detection of an intrusion on the system by sending an alarm. They warn the user of the attack, but they do not take any preventive actions or countermeasures against the detected attacks.

*2.6 MACHINE LEARNING AND DATA MINING*

Machine learning is a type of artificial intelligence that focuses on the use of data and algorithms to imitate the way that humans learn and gradually improving its accuracy. Generally, there are three types of machine learning.

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

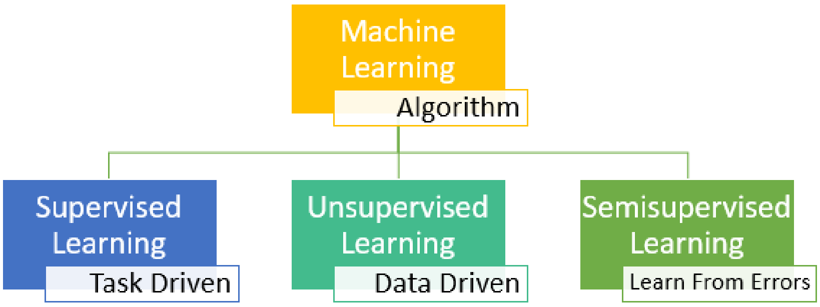


Figure 2 5: Types of machine learning

Supervised Learning

It is one of the most basic types of machine learning where algorithms are trained on labeled data. Supervised learning is extremely powerful when the data needs to be labeled accurately for this method to work in the right circumstances. In supervised learning, the algorithm is given a small training dataset to work with a smaller part of the bigger dataset and serves to give the algorithm a basic idea of the problem and solution [22]. Example for the algorithm of supervised learning as mentioned below.

Input data is called training data and has a known label such as spam and not-spam. A model is prepared through a training process in which it is required to make predictions and is corrected when those predictions are wrong. The training process continues until the model achieves a desired level of accuracy on the training data as shown on picture. [2]

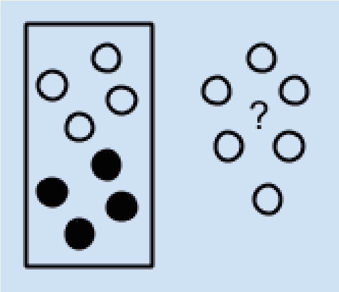


Figure 2 6: supervised learning algorithm

Unsupervised learning

Unsupervised learning involves algorithms that train on unlabeled data that scans through datasets looking for any meaningful connection. In this learning algorithm trains on as well as recommendations they output are predetermined. A model is prepared by deducing structures present in the input data may be to extract general rules and through a mathematical process to systematically reduce redundancy to organize data by similarity. [2]

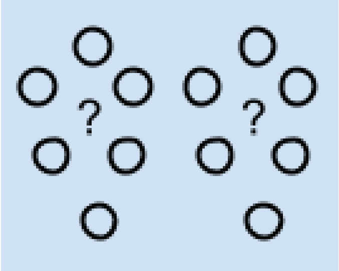


Figure 2 7: Unsupervised learning algorithm

Semi-supervised Learning

Semi-supervised learning involves a mix of the two preceding types of data that is labeled and unlabeled. There is a desired prediction problem but the model must learn the structure to organize the data as well as make predictions.

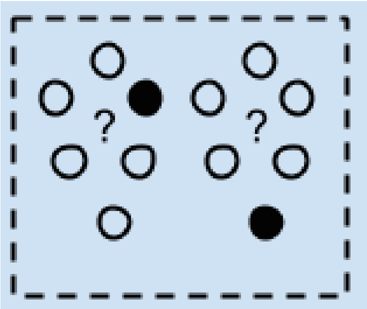


Figure 2 8: Semi-supervised learning algorithm

For the ML application on network intrusion detection systems (NIDS), we use the F1-score, precision, and recall as the assessment metrics for selecting the trained models with precision representing the exactness, and recall its completeness, then F1-score is the weighted harmonic mean between precision and recall. These metrics use the values retrieved from the traditional confusion matrix (Table 2.1), such as True-Positive (TP), True-Negative (TN), False-Positive (FP), and False-Negative (FN):

$$\text{Precision} = \frac{TP}{TP+FP}\ , \ \text{Recall or TPR} \ = \ \frac{TP}{TP+FN}\ ,$$

$$\text{F1-score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}\ .$$

Table 2 1: Confusion Matrix

| **Predicated Class** | | | |
|---|---|---|---|
| **Actual Class** | | + | - |
| | + | True Positive (TP) | False Negative (FN) |
| | - | False Positive (FP) | True Negative (TN) |

The if-else expressions used in looping are comparable to the Decision Tree Algorithm. A decision tree is a graphic depiction that, under certain circumstances, provides solutions to the difficulties we face on a daily basis.

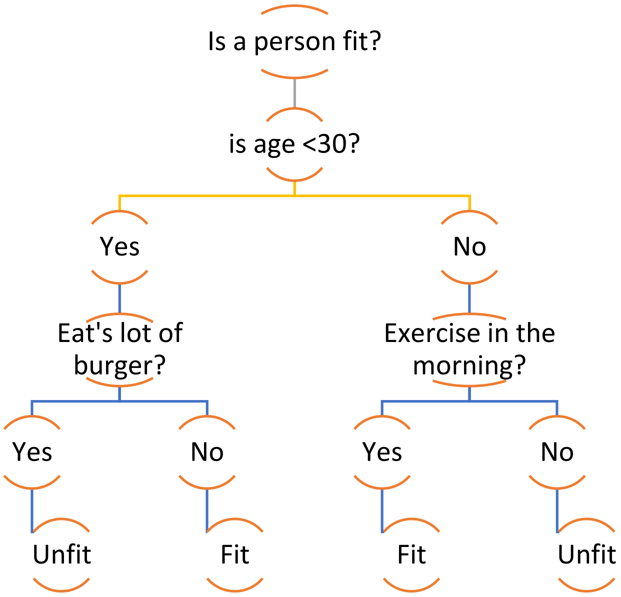


Figure 2 9: Decision tree algorithm

Classification rules as shown in the table below

Table 2 2: Classification rules

| Classification Rules | |
|---|---|
| Rule 1 | If age <30 AND not eating too much burger then you are fit; |
| Rule 2 | If age > 30 AND do exercise in the morning then you are fit; |

Li Yang and Abdallah Shami have made important advances in the field of smart technology for Internet of Things devices and systems through their presentations and analyses. IoT data analytics encounters concept drift issues because of the dynamic nature of IoT systems and shifting patterns of IoT data streams. Nevertheless, they launched IoT services and functionalities based on the analytics of IoT streaming data. They propose an adaptive IoT streaming data analytics framework for anomaly detection which use optimizes LightGBM, a concept rift adaption. [3]

Hasan Alkahtani and Theyazn H. Aldhyani have presented a presentation discussion and analysis intrusion detection system for Internet of Things using sensor devices to collect data from a smart grid environment. These data sent and stored in the cloud to serve and provide different services to different smart infrastructure such as smart homes and smart buildings. It creates a large space specially for hackers to launch destructive programs to cyberattacks. An IoTID20 dataset attack was used to develop their proposed system. [4]

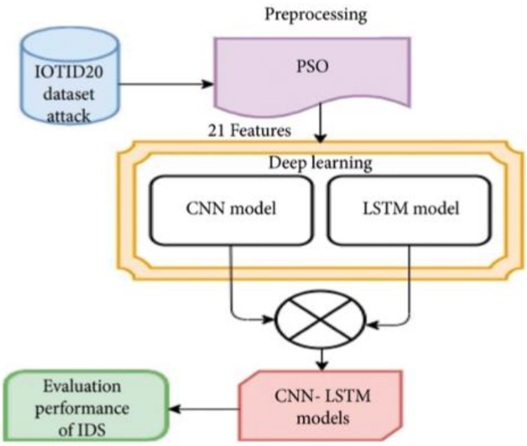


Figure 2 10: IDS for IoT

Albulayhi, K.; Abu Al-Haija, Q.; Alsuhibany, S.A.; Jillepalli, A.A. have presented Identifying and selecting relevant features in the dataset has become crucial to improving Machine Learning model performance for anomaly-based IDS. To address the challenge of improving an anomaly-based IDS in the IoT ecosystem, they have given node data attributes to identify relevant and redundant features. Redundant features affect the models' performance. They assumed that the current FS techniques do not always guarantee the best relevant features or eliminate redundant features. Therefore, they have aimed to find new strategies in dealing with the discovery of useful features and concurrently the removal of unsupportive features. Although the current approaches are practically simple to use, they can and often do use a lot of resources. [5]

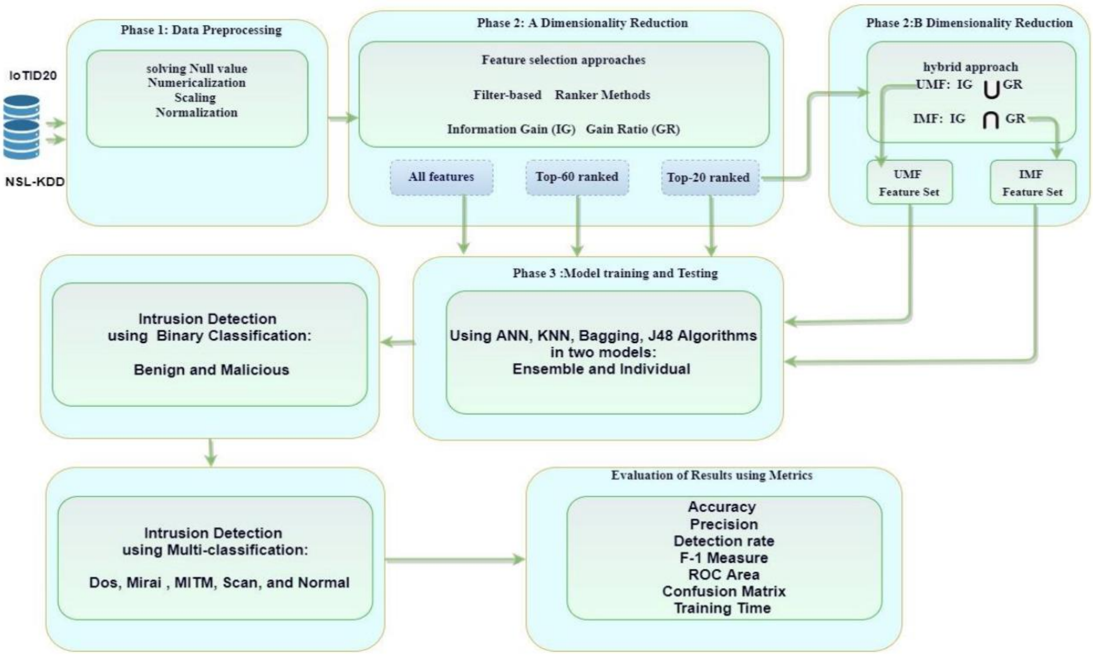

IoTID20
NSL-KDD
Phase 1: Data Preprocessing
solving Null value
Numericalization
Scaling
Normalization
Phase 2: A Dimensionality Reduction
Feature selection approaches
Filter-based Ranker Methods
Information Gain (IG) Gain Ratio (GR)
All features
Top-60 ranked
Top-20 ranked
Phase 2:B Dimensionality Reduction
hybrid approach
UMF: IG ∪ GR
IMF: IG ∩ GR
UMF Feature Set
IMF Feature Set
Phase 3 :Model training and Testing
Using ANN, KNN, Bagging, J48 Algorithms in two models:
Ensemble and Individual
Intrusion Detection using Binary Classification:
Benign and Malicious
Intrusion Detection using Multi-classification:
Dos, Mirai , MITM, Scan, and Normal
Evaluation of Results using Metrics
Accuracy
Precision
Detection rate
F-1 Measure
ROC Area
Confusion Matrix
Training Time

Figure 2 11: ML based IoT using Hybrid

Qasem Abu AI-Haija and Mu'away AI-Dala'ien has proposed, developed and evaluated lightweight, intelligent and accurate IDS for an IoT network system called ELBA-IoT. The proposed solution approach evaluated the performance of three supervised machine learning models i.e. (AdaBoosted, RUSBoosted, and bagged) in order to provide more inclusive experiments and gain more insight into the solution approach. The N-BaIoT-2021 is a comprehensive and contemporary dataset comprising real-world IoT network traffic—was used to evaluate the performance of ELBA-IoT. However, in their future work they have given one recommended extension to deploy ELBA-IoT using physical IoT gateway devices to provide a real time botnet detection using Raspberry Pi, ARM Cortex or Arduino. [6]

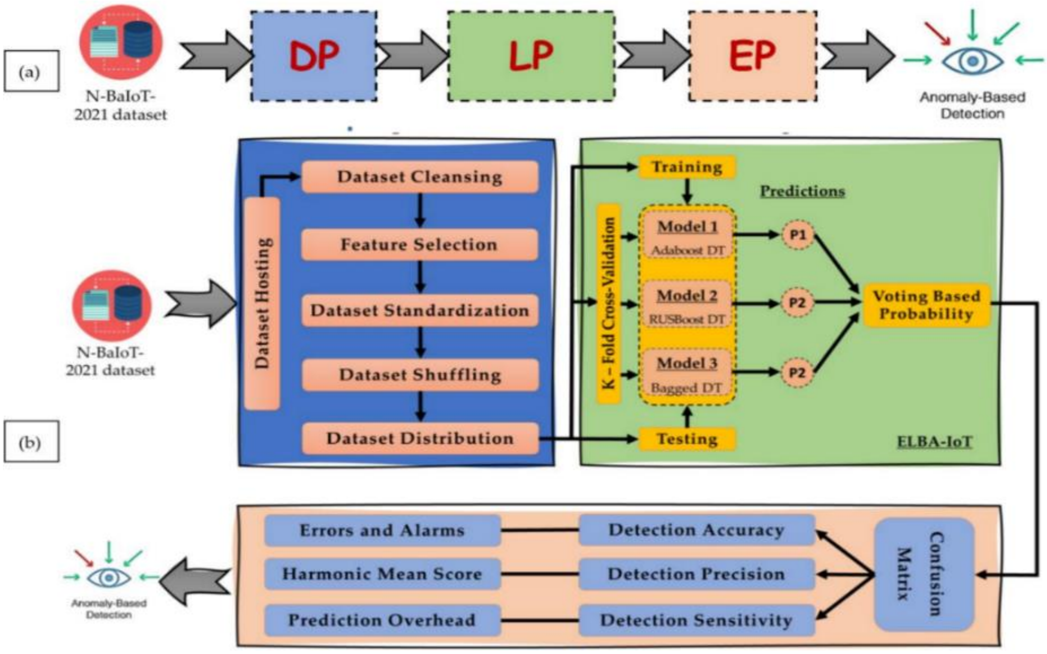


Figure 2 12: ELBA-IoT

Abu Al-Haija, Q. and Al Badawi surveyed statistical, machine learning, and data mining methods for constructing NIDSs in Software Defined Networks (SDNs). Based on their study, they recommend the suitability of machine-learning methods over other methods due to their flexibility, lightweight inference overhead, and high accuracy rates. To solve the network intrusion detection issue, the authors proposed a two-stage deep learning and machine learning approach. They reported that Random Forests (RF) achieved 0.996 F-measure and found that ANN with wrapper feature selection outperformed Support Vector Machines (SVMs) on the NSL-KDD dataset. [7]

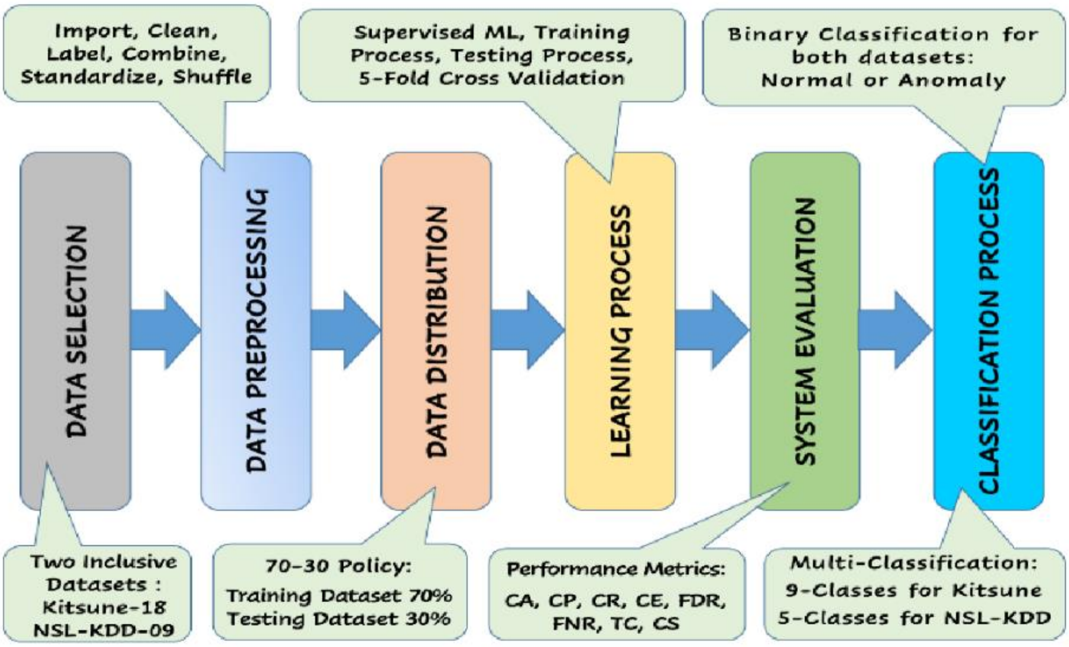

Import, Clean, Label, Combine, Standardize, Shuffle
Supervised ML, Training Process, Testing Process, 5-Fold Cross Validation
Binary Classification for both datasets: Normal or Anomaly
DATA SELECTION
DATA PREPROCESSING
DATA DISTRIBUTION
LEARNING PROCESS
SYSTEM EVALUATION
CLASSIFICATION PROCESS
Two Inclusive Datasets : Kitsune-18 NSL-KDD-09
70-30 Policy: Training Dataset 70% Testing Dataset 30%
Performance Metrics: CA, CP, CR, CE, FDR, FNR, TC, CS
Multi-Classification: 9-Classes for Kitsune 5-Classes for NSL-KDD

Figure 2 13: Workflow Diagram for attack-aware IoT network traffic routing

Stefan Axelsson has presented and analyzed anomaly detection systems with help of self-learning, programmed and supervised learning for Intrusion Detection Systems. Authors employed different principles of detection. Some systems used a two tiered model of detection as well as one lower level feeds a higher level. These systems like MIDAS, Haystack are the types of signature programmed that permits decisions on anomaly data that helps to detect anomalies from signature data. [8]

Author Ioulianou, Philokypros, Vasilakis, Vasileios, Moscholios, Ioannis, Logothetis, Michael have presented a signature based IDS for IoT networks. They presented a high level IDS architecture and their main components which involved both centralized and distributed modules for detecting IDS from external networks as well as for the internet. However, the DOS attacks may constitute some nodes unreachable and may negatively impact their power consumption. In their future work, they planned to implement and test the proposed design in Cooja that will improve the IDS performance by reducing the false positives during the attack detection process. Finally, they have imported the IDS modules to Contiki OS in order to test its performance in a real-world IoT environment. [9]

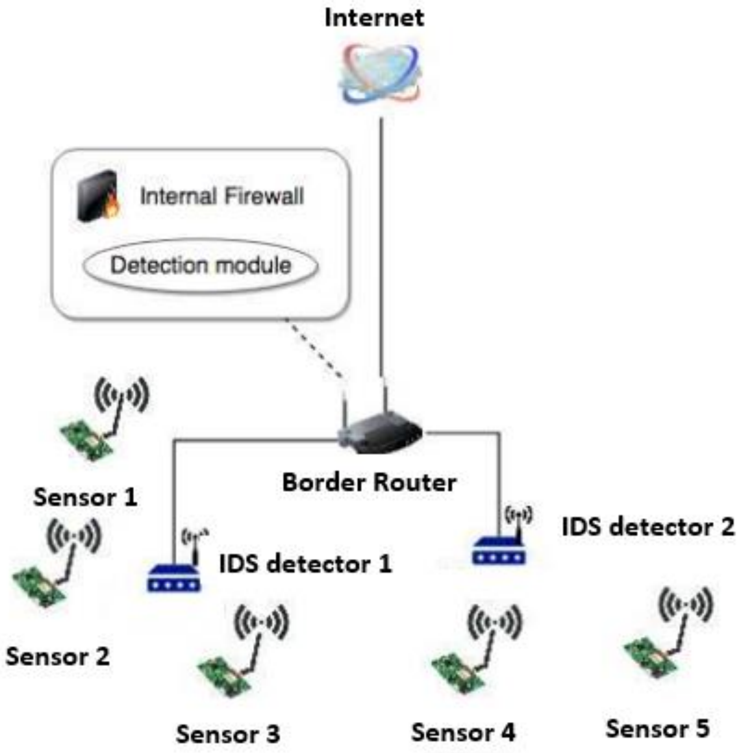


Figure 2 14: High level IDS architecture

Mahbod Tavallaee have analyzed and demonstrated that the most commonly recommended methods for evaluating and comparing intrusion detection systems are imperfect. According to MT, intrusion detection system should not be evaluated based on the areas under their receiver operating characteristics (ROC) curves or false alarm rates or distance from a goal. It need to be predicated on anticipated expenses that account for the price of a false alarm, the price of failing to detect, and the likelihood of an incursion in the past. Furthermore, the operating point of an IDS depends on the probability of a false alarm (α) and probability of missed detection ($\beta$) and should be established to minimize the expected cost. IDS's ROC curve is: Min $\{C\beta\, p, (1 - \alpha)\, (1 - p)\}$ + Min $\{C\, (1 - \beta)\, p, \alpha\, (1 - p)\}$, where $C = C_{\beta}/C_{\alpha}$ and $p$ is the prior probability of intrusion. [10]

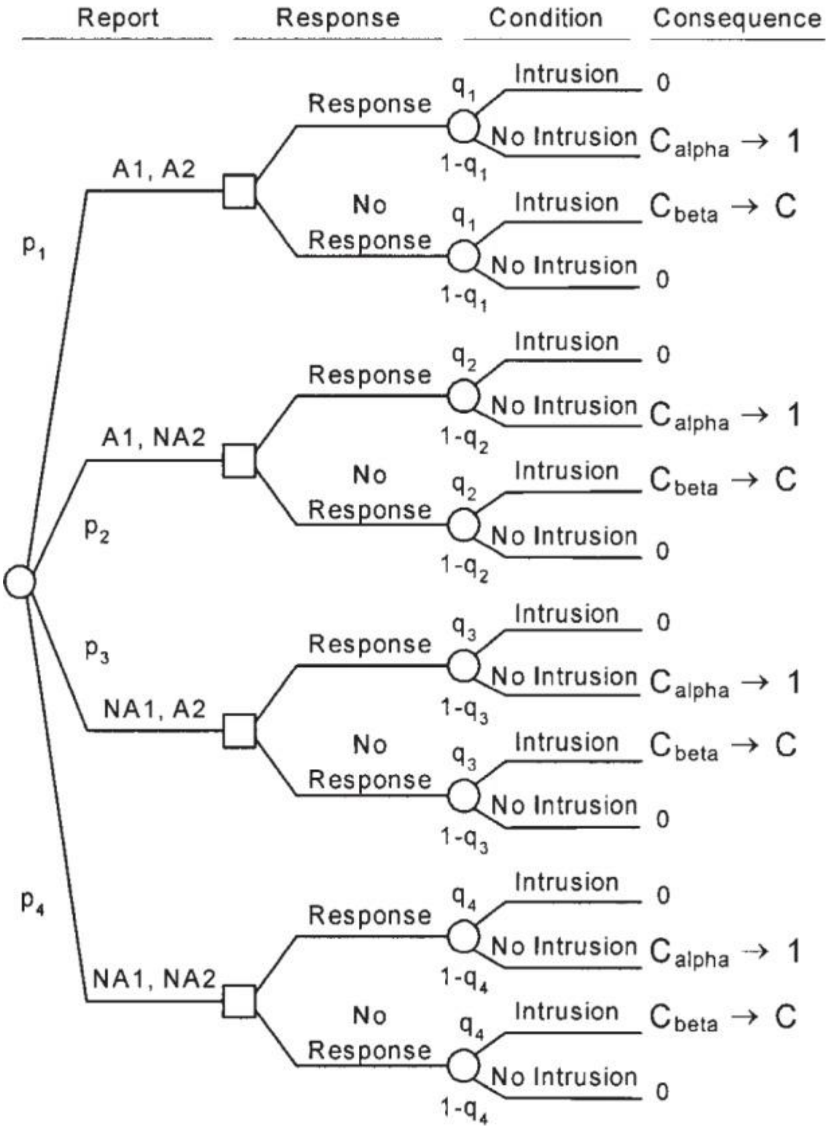


Figure 2.15: Decision tree for a compound IDS.

Canfora, Gerardo, Francesco Mercaldo, and Corrado proposed an approach which builds machine learning models on static and dynamic analysis features to detect malicious JavaScript. They extracted features like script execution time, calls to JS functions and number of function calls made by the JavaScript code via dynamic analysis features. [11]

Likarish, Peter, Eunjin Jung, and Insoon Jo demonstrated a classification-based approach to detect malicious HTML, JavaScript. The authors used a unigram and bigram-based feature extraction approach which is similar to natural language text processing. For example, to extract features like human readable characters, whitespace etc. to train models using Naïve Bayes, ADTree, SVMs and RIPPER algorithm. They were able to train a model. [12]

# Chapter 3

## Methodologies and tools

This chapter three presents a methodology for designing the Anomaly Based Intrusion Detection Systems and introduce the IoT architecture as well as methodology. It covers IoT threat modelling, methodology, anomaly behavior analysis, IDS and then data structures used to model the normal behavior of the analyzed protocols.

### *3.1 IoT Architecture*

The IoT (Internet of Things) and CPS (Cyber Physical Systems) are both complex platforms, however both of them share a common goal of technology, reliability, expanding advancement opportunity and exposing areas of untapped potential. IoT refers to a network comprised of physical objects to gather and share electronic information (smart devices) whereas Cyber Physical Systems refers to engineered systems where functionalities and salient properties emerge from the network and physical components. Therefore, in this project highly availability and flexibility is very important at any given time. System failure could cause of lose some business in the best case.

The IoT devices have no sizes or limits but it works with sensors which fits all architecture for IoT projects. IoT and CPSs have become an integral part of modern computer networks. Therefore, securing computer networks and IoT/Cyber Physical Systems leads successfully to IoT architecture. I will be presenting here IoT threat modeling as well as CPS with IoT protocols and networking protocols.

IoT architecture helps understand the behavior of the IoT/cyber physical devices to identify their vulnerabilities. IoT modelling consists of four layers as shown below in figure 3.1.1. These are End devices, communications, services and applications.

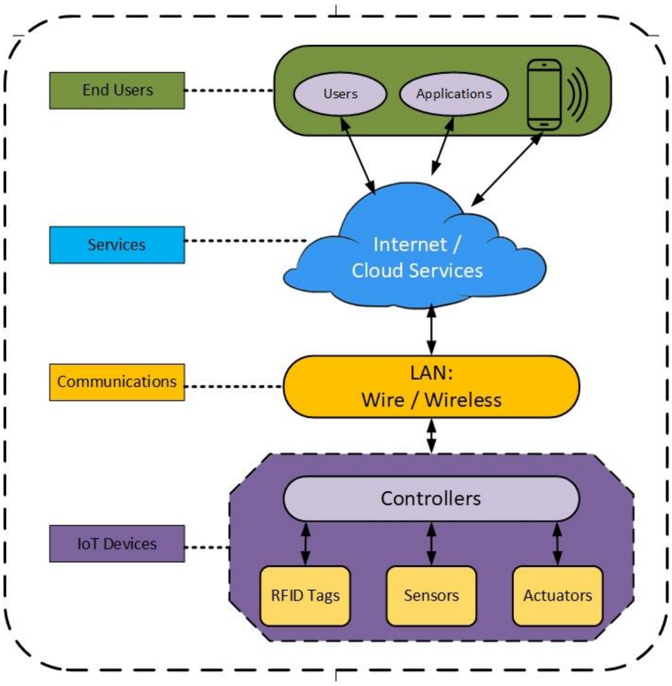


Figure 3 1: IoT Modelling

In the above figure, the bottom level is composed of IoT devices where controllers that control these devices like RFID, Sensors and Actuators interact by end users and then it collects data and passes it to the processing services that are internet (cloud services) which will connect through wire or wireless communication. The services layer provides common middleware and functions to build sophisticated IoT services in the application layer. Users access Internet of Things (IoT) devices using end-user applications, which are provided by the applications layer.

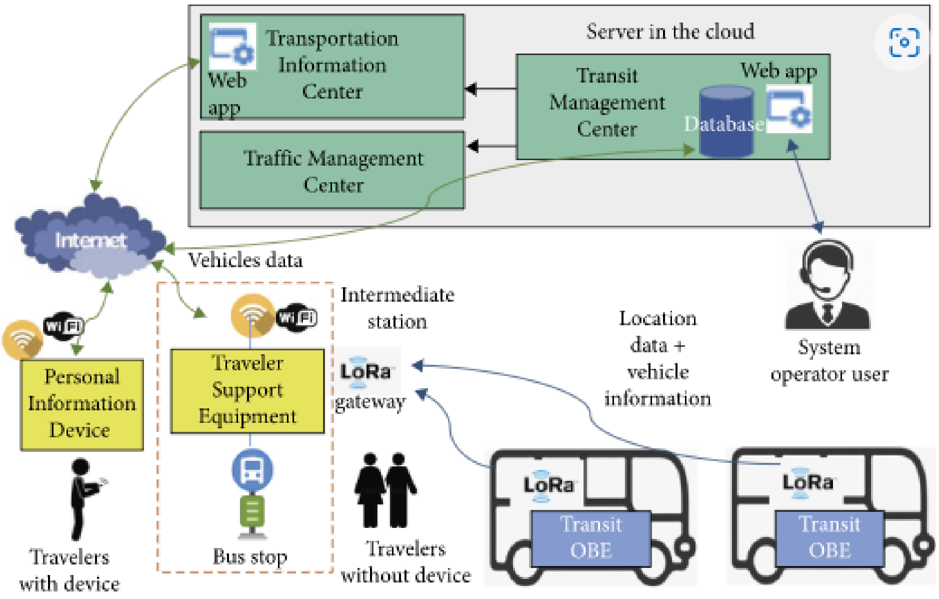


Figure 3 2- Application of IoT architecture for public vehicles

For the development of this system, we need wireless devices for the end users, and services we will be using cloud based server and web apps, transmit management center through internet and the elements in the On-Board-Equipment (OBE) in the bus, microcontroller cards and a gateway device. The selected microcontroller card was the HiLetgo ESP32 LoRa 0.96 inch OLED display, manufactured by Heltec. The selected GPS was the Ublox Neo 6M. Regarding the gateway, we had the opportunity to review two references, the RAKWireless RAK831LoRa/LoRaWAN and the Dragino gateway LG01. microfter performing some tests with the two gateways, we deemed the Dragino gateway LG01 as the best option (see Figure 3.3) because this reference did not require an additional card (such as a Raspberry Pi) to operate. In addition, the Dragino gateway was easily programmable through the Arduino IDE, which was not possible with the other gateway.

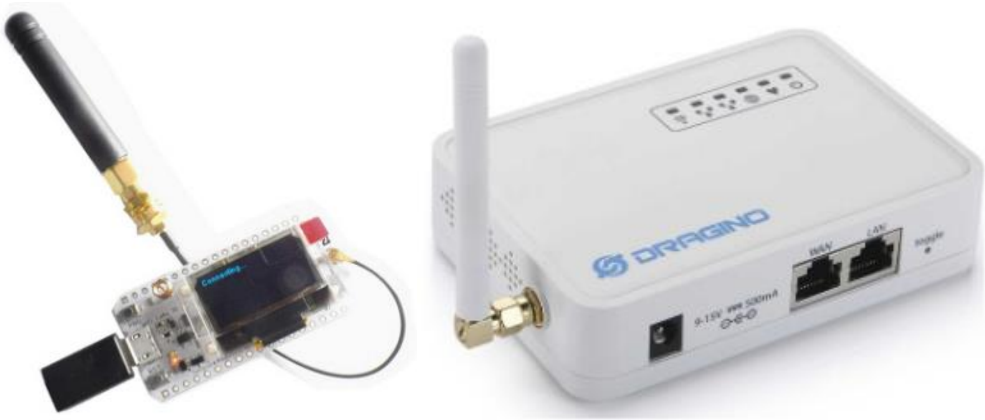


Figure 3 3: Wireless device and Gateway

### *3.2 IOT THREAT MODELING METHODOLOGY*

IoT threat modeling methodology helps analyze the vulnerabilities in cyber physical systems like IoT devices and help develop methodologies to remove the impact of these vulnerabilities if they were victimized by attackers in the cyber physical systems. IoT threat modelling consists of four layers where each layer has 5 different tasks as shown in figure 3.2.1.

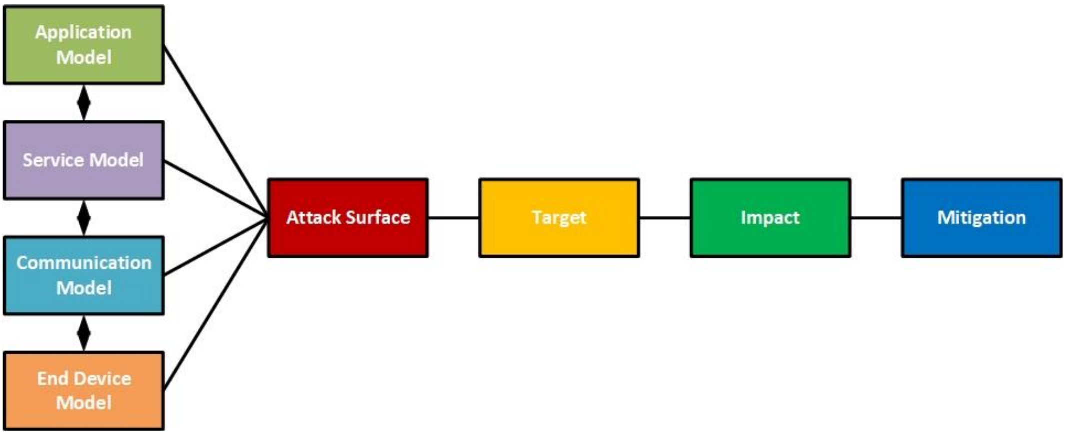


Figure 3 4: IoT threat modelling methodology

The layer model will then be analyzed to identify the attack that characterizes the entry points that can be attached by attackers to inject malicious events into that layer functions.

The IoT threat modelling consists of four layers and each layer has five different tasks as shown in the figure 3.2.1.

In these layers it discusses the functions as investigating the attack surface for the model, investing the targets of the attack surface, investigating the impact of the attack if executed and finally mitigation strategies. According to proponent, first we develop a model that captures the functions to be provided by that layer and then the layer model will be analyzed to identify the attack surface which is characterized by the entry points that can be attacked by attackers to inject malicious events into that function. Each vulnerability target layer will identify attack by attackers and then it measures the risk and impact if the attack was successful and the final step in each row of the IoT threat modelling methodology is to develop methods to mitigate or eliminate these vulnerabilities in order to achieve the necessary secure operations of functions provided by each layer.

**End Device Layer:**

It shows the attack surface, impact, mitigation, and mitigation mechanisms associated with the end-devices layer as shown below table 3.1

Table 3 1: End Device Layer

| Attack Surface | Target | Impact | Mitigation Mechanism |
|---|---|---|---|
| **Controllers** | Control, Information | Lost control, human life, safety and failure | Anomaly behavior analysis to detect abnormal control information, encryption |
| **Sensors** | Information access to the system | Lost control, human life, safety and failure | Lightweight encryption, anomaly behavior analysis, secure sensor identification and authentication. |
| **Actuators** | Controls | Lost control, human life, safety and failure | Lightweight encryption, anomaly behavior analysis and Anti-jamming |
| **Entertainment** | Access to the system | Lost control, human life, safety and failure | Encryption, moving target defense, anomaly behavior analysis. |

**Communication Layer:**

This layer shows attack surface, target, impact and mitigation mechanisms associated with communication layer as shown table below 3.2.

Table 3 2: Communication Layer

| Attack Surface | Target | Impact | Mitigation Mechanism |
|---|---|---|---|
| **Protocols** | Access control, Information | Lost control, human life, safety and failure | Authentication, Anomaly behavior analysis , moving target defense, anti-jamming. |
| **Firewalls** | Access control to the system | Lost control, human life, safety, long time, energy waste | IDS, behavior analysis, authentication. |

|  |  |  |  |
|---|---|---|---|
| **Routers** | Access information | Lost control, human life, safety and time | Anomaly behavior analysis and Anti-jamming, encryption |
| **Communication Bus** | Access information | Privacy, money, human life safety, long time and failure. | Encryption, IDS, moving target defense, anomaly behavior analysis. |

**Services Layer**

This layer shows attack surface, target, impact and mitigation mechanisms associated with services layer as shown table below 3.3

Table 3 3: Service Layer

| Attack Surface | Target | Impact | Mitigation Mechanism |
|---|---|---|---|
| **Cloud Services** | Personal and confidential information | Data lost, money, time and safety. | Encryption, Anomaly behavior analysis , moving target defense, behavior analysis, selective discloses, data distortion, big data analysis. |
| **Web Services** | Control and monitor | Control, human life, safety, money and cyber crime | Authentication, secure identity management, encryption and anomaly behavior analysis. |

**End user / application layer**

This layer shows attack surface, target, impact and mitigation mechanisms associated with services layer as shown table below 3.4.

Table 3 4: End user / application layer

| Attack Surface | Target | Impact | Mitigation Mechanism |
|---|---|---|---|
| **Smart phones (mobile devices)** | Information and control | Human life safety, personal information, money. | Authentication, access control, anomaly behavior analysis, moving target defense. |
| **Programs / applications** | Access to the system and Control, information | Time, money, safety and reputation. | Encryption, white listing, continuous Authentication and anomaly behavior analysis. |

### *3.3 Anomaly behavior analysis for IoT*

The integration of IoT with Fog and Cloud Computing enable IoT services to be distributive, cost effective and can be accessed easily at any time and from anywhere. In every IoT application and communications are crucial to deliver the required information and take actions during crisis events, however IoT components such as sensors, gateway and other nodes will introduce major security challenges as they contribute to increase the attackers and preventing the IoT from delivering accurate information to the end users.

Here I am presenting some behavior regarding ABS for IoT to detect when an IoT network nodes are being compromised. This will help to accurately detect known and unknown anomalies due to cyber-attacks with high detection rate and low false alarms. Anomaly behavior refers to even though traditional approaches need to be complemented, still **SIEM** (Security Information and Events Management) and **GRC** (Governance, Risk and Compliance) are required.

In Anomalous Behavior, GRC says what is approved the tasks you can do, the fates you can go through. Abnormal Behavior Detection says whether you should have. Extend using anomalous Behavior Detection approaches are:

Learning what is normal means the difference between approved and allowed. Identifies what is anomalous and categories the risk and finally alert so you can react before it becomes a problem. Whereas new outcomes are also possible. Example: it shows where allowed is not normal and the scope of the deviation from the norm. It detects social engineering attacks as well as network level detections. It minimizes the exposure time and loss and then finally potentially predict the leakage areas ahead of the attack. It can be applied to both GRC and non-GRC areas to build up a broader pattern of behavior. Detection of Anomalous Behavior from insight to action.

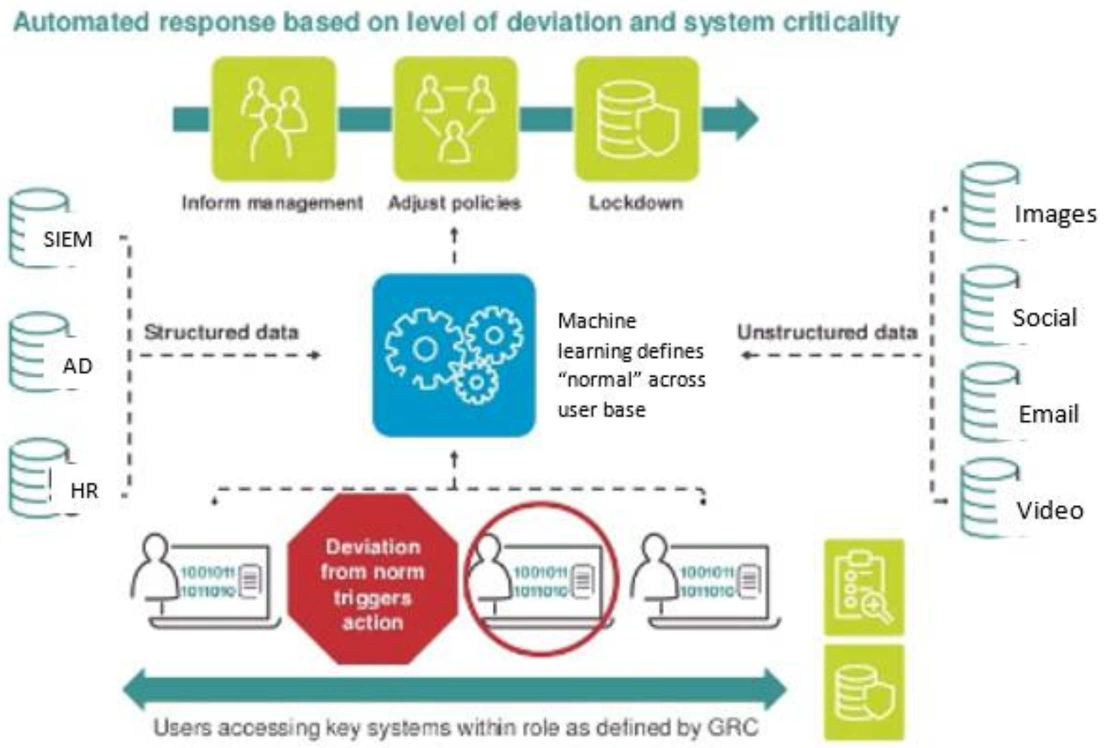

Automated response based on level of deviation and system criticality
Inform management
Adjust policies
Lockdown
SIEM
AD
HR
Structured data
Machine learning defines "normal" across user base
Unstructured data
Images
Social
Email
Video
1001011
1011010
Deviation from norm triggers action
Users accessing key systems within role as defined by GRC

Figure 3 5: Anamolous behavior

Example- SIEM, GRC and Detection of Anomalous Behavior.

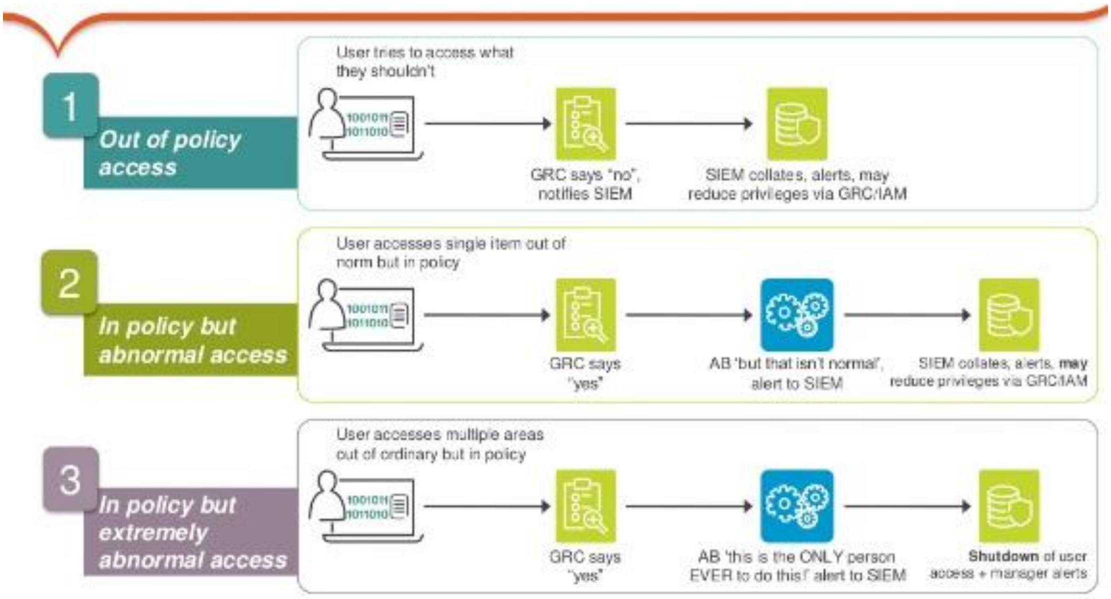

1
Out of policy access
User tries to access what they shouldn't
GRC says "no", notifies SIEM
SIEM collates, alerts, may reduce privileges via GRC/IAM
2
In policy but abnormal access
User accesses single item out of norm but in policy
GRC says "yes"
AB 'but that isn't normal', alert to SIEM
SIEM collates, alerts, may reduce privileges via GRC/IAM
3
In policy but extremely abnormal access
User accesses multiple areas out of ordinary but in policy
GRC says "yes"
AB 'this is the ONLY person EVER to do this!' alert to SIEM
Shutdown of user access + manager alerts

Figure 3 6: SIEM, GRC detection

### *3.4 Intrusion Detection System Design Methodology*

In chapter 3 section 3.3, I have shown that the designing of an Anomaly Behavior using analysis involves three layers.

1. Threat modeling analysis of the protocol
2. Features selection and protocol of foot printing to characterize the behavior of the protocol
3. Use the selection set of features to develop machine learning models which characterize the normal behavior.

In the beginning, I have applied the threat modeling methodology under section 3.2 about different types of layers including state machines and architectures that describe the normal behavior for the protocol. Because of this model, we can easily identify and impact analysis is performed on these attack surfaces and Mitigation strategies are devised to mitigate the threats from these attacks.

Figure below shows anomaly behavior methodology.

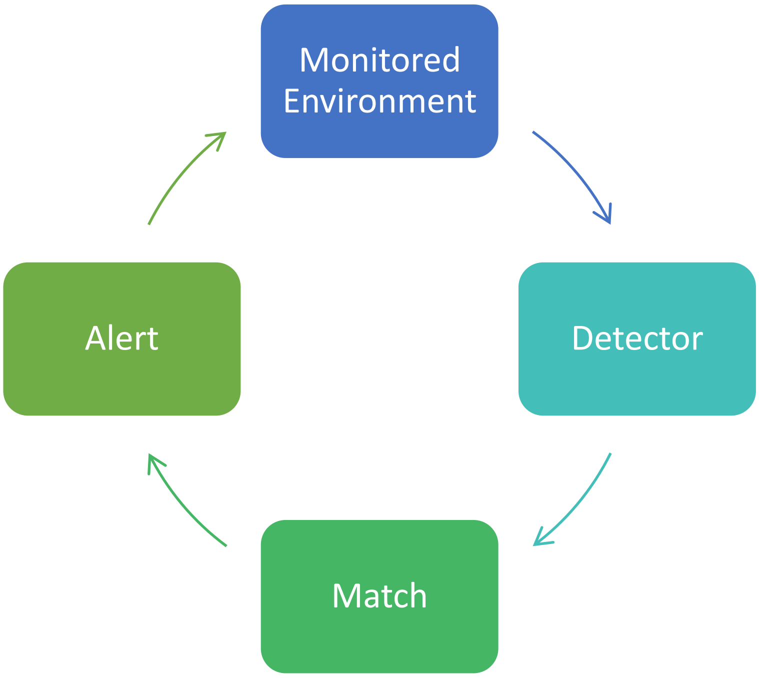


Figure 3 7: Anomaly Behavior Design Methodology

### *3.5 FINGERPRINT/FOOTPRINTING DATA STRUCTURES*

It is a deep testing technique to capture the state machine technique made by the system and protocol. Some information in fingerprint includes network topology, cluster architecture, application software, operating system platform and database. It involves scanning network traffic and outgoing packets from the target systems/network in the form of a digital signature.

Fingerprint techniques generally categorized into two as follows:

1. Active Fingerprinting

Active fingerprinting methods are riskiest as they are easier to capture with IDS and port scanning or network mapping are some of the used active fingerprinting tools to identify the types of packets and other information. Active fingerprint technique used by anonymous users by sending suspicious packets to the system and analyze their response by using TCP/IP protocol.

2. Passive Fingerprinting

It is another technique in which the hacker sniffs network traffic as scouting for creating a digital footprint of the corporate network. Usually, hackers use network scanning and other systems to penetrate testing and log file activity. Picture below shows the general architecture of fingerprints.

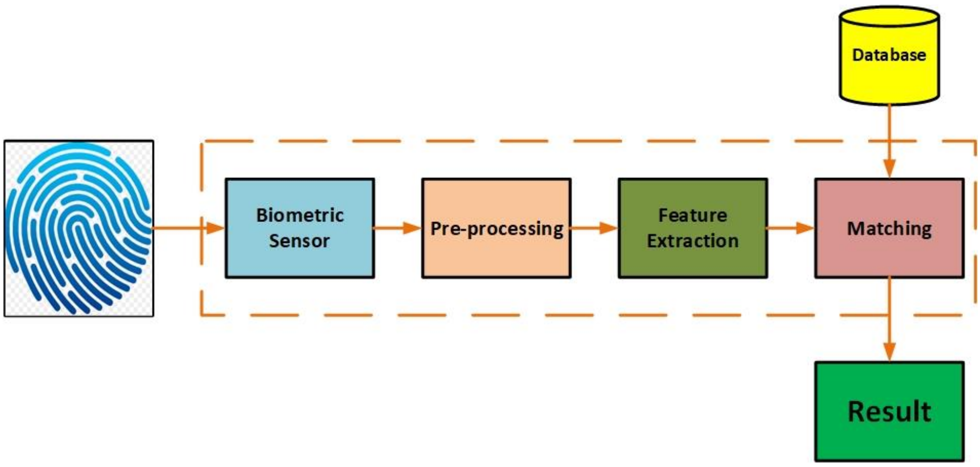


Figure 3 8: Fingerprinting Architecture

Generally, we use N-grams and Observation-flow data structure to effectively fingerprint the normal behavior using different protocols. N-grams are extensively used in text mining and natural language processing tasks. It is basically repeating symbols or tokens.

Observation flow is a continuous flow of frames or packets between a source and destination pair sampled at a specific interval of time ‘t’ which is made by the protocol.

If X=Num of words in a given sentence K, the number of n-grams for sentence K would be: $\text{Ngrams}_K$ = X-(N-1)

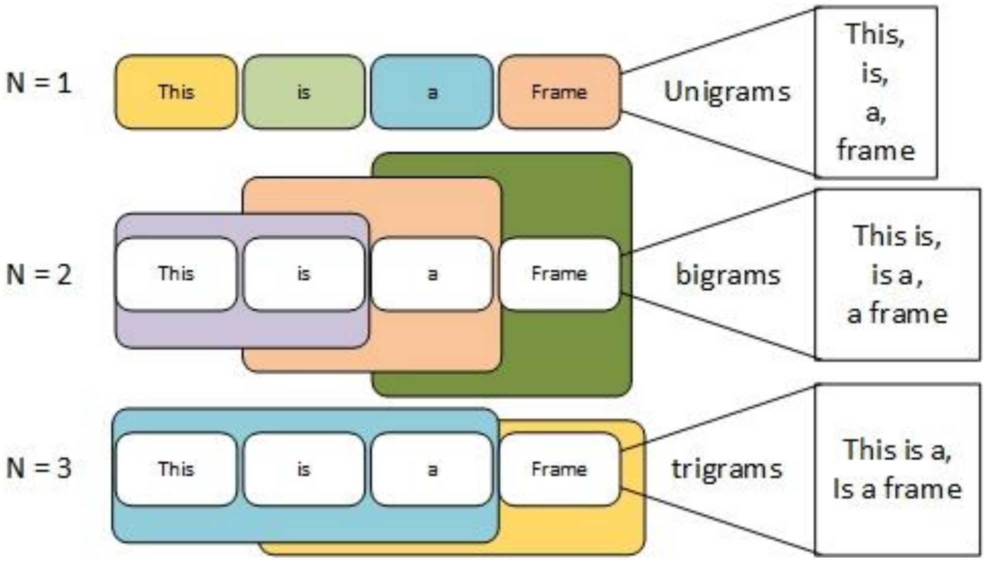


Figure 3 9: N-grams and observation flow

# Chapter 4

## Designing Anomaly Behavior IDS for the Wi-Fi protocol

The Wi-Fi protocol is also known as IEEE 802.11 which is wireless LAN protocol. Wireless network is a part of Ethernet which was first formalized in the year of 1997. Nowadays, it has been upgraded and has incurred many changes however most of these upgrades have been to enhance the data rate and the link quality of the network and also one for the security.

### *4.1 The IEEE 802.11 standard*

IEEE 802.11 standard is also known as Wi-Fi that lays down the architecture and specifications of wireless LANs (WLANs). Wi-Fi or WLAN uses high frequency radio waves for connecting the nodes. There are several standards of IEEE 802.11. WLAN can be categorized into 802.11, 802.11a, 802.11b, 802.11g, 802.11n and 802.11p. All the standards use CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) and support for both centralized based stations as well as hoc networks.

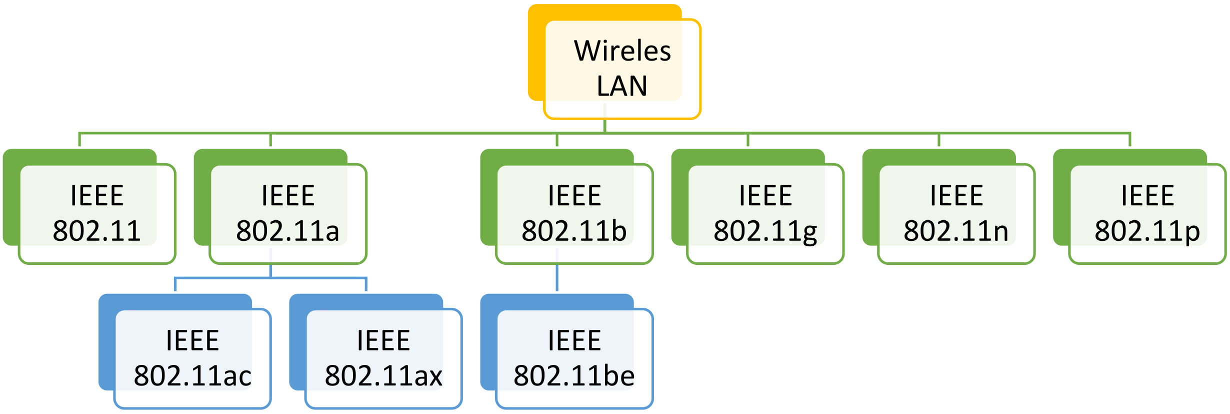


Figure 4 1: Wireless Local Area Network Standards

### *4.1.1 IEEE 802.11a*

IEEE 802.11a was one of the first Wi-Fi standards to be launched. It is a modification to 802.11 and it was published in 1999. It provide a maximum data rate of 54 Mbps operating in the 5 GHz band. It also provides error correcting code. During that time IEEE 802.11a was costlier and a little bit more difficult to implement as it operated at 5GHz rather than 2.4 GHz because of it was used less.

### *4.1.2 IEEE 802.11b*

IEEE 802.11b appeared in early 2000 A.D. which is direct extension of the original 802.11 standard. The modulation technique is same as 802.11 and it works in the 2.4GHz band. Data rate for IEEE 802.11b is 11Mbps. However, 2.4GHz band is pretty crowded even though 802.11b faces interference from other devices.

### *4.1.3 IEEE 802.11g*

IEEE 802.11g published in 2003. It operates in the 2.4GHz band however it provides average throughput of 22 Mbps. It operated Orthogonal frequency division multiplexing technique as in 802.11a. The main advantages of this protocol is that it is fully backward compatible with 802.11b and 802.11g.

### *4.1.4 IEEE 802.11n*

802.11n was approved and launched in 2009 and operates on both 2.4 GHz and the 5 GHz bands. It works at different data rate ranging from 54 Mbps to 600 Mbps.

### *4.1.5 IEEE 802.11p*

802.11p is an amendment for including wireless access in vehicle environments to support ITS (intelligent transportation systems). It is specially used for vehicles moving at high speed and environment. It has a data rate of 27 Mbps and operates in 5.9 GHz band.

### *4.1.6 IEEE 802.11ac*

802.11ac was approved in the year 2014. It allows each single link to have a throughput of 500 Mbps while the overall throughput for a multi-station wireless local area network would be at least 1 Gbps. That allows the use of wider channels and up to 8 multiple input multiple output (MIMO) spatial streams and use of 256 Quadrature Amplitude Modulation (QAM).

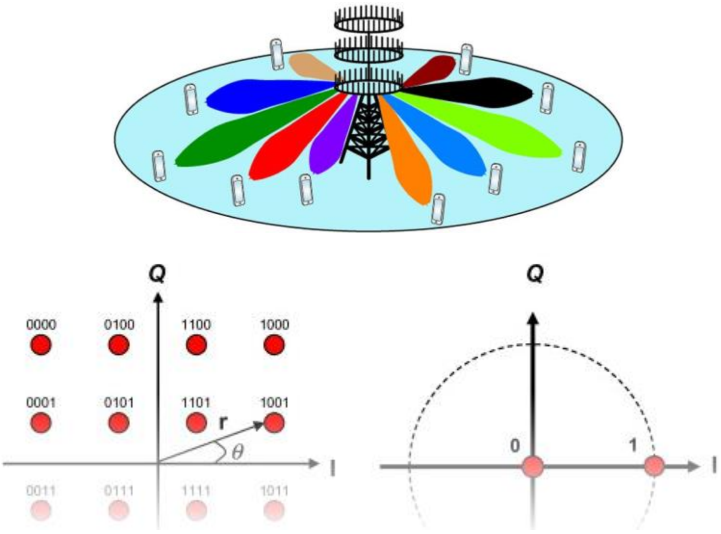


Figure 4 2:  MIMO and QAM

QAM is a modulation format that combines two carriers and their amplitudes are modulated independently with the same optical frequency and whose phases are 90° apart. It is called in-phase carriers (*I*) and quadrature-phase carriers (*Q*). The QAM can be assigned into two states by using *I* and *Q*, which is called $2^n$ QAM. Figure above shows signal processes *N* bits in a single channel, so it can realize *N* times the spectral efficiency of OOK (On-Off Key).

### *4.1.7 IEEE 802.11ax*

802.11ax also known as Wi-Fi 7 which was introduced in 2021. It is the new generation of Wi-Fi technology with a new focus on performance and efficiency. The foundation of Wi-Fi 6 technology is the improved and more effective utilization of the current radio frequency medium.

### *4.1.8 IEEE 802.11be*

802.11be standard established to implement wireless communication as much faster speeds and larger capacities then 802.11ax. It is also known as Wi-Fi 7 and bandwidth range start from 20 MHz to 320 MHz. It also uses QAM symbols and 16 spatial streams and this Wi-Fi 7 data throughput beyond 30 Gbps with low latency will be a key technology supporting new applications and services. For example: video streaming over 4K, virtual reality (VR) and argument reality (AR).

Evolution of IEEE 802.11 as shown below:

Table 4 1: IEEE 802.11 Evolution

| 802.11 | 11a | 11b | 11g | 11n | 11ac | 11ax | 11be |
|---|---|---|---|---|---|---|---|
| Transmission Vector Format | Non-HT Non-High Throughput | Non-HT Non-High Throughput | Non-HT Non-High Throughput | HT - High Throughput | VHT Very High Throughput | HE - High Efficiency | EHT Extreme High Throughput |
| Definition | 1999 | 1999 | 2003 | 2009 | 2014 | 2021 | Planned in 2024 |
| Freq 2.4 GHz | | ✓ | ✓ | ✓ | | ✓ | ✓ |
| 5 GHz | ✓ | | | ✓ | ✓ | ✓ | ✓ |
| 6 GHz | | | | | | ✓ | ✓ |
| Bandwidth [MHz] | 20 | 22 | 20 | 20 /40 | 20 /40 /80 /160 /80+80 | 20 /40 /80 /160 /80+80 | 20 /40 /80 /160 /320 |
| Maximum throughput rate [bps] | 54M | 11M | 54M | 540M | 6.93G | 9.6G | 30G |
| Modulation scheme | BPSK<br>QPSK<br>16QAM<br>64QAM | DBPSK<br>DQPSK | BPSK<br>QPSK<br>16QAM<br>64QAM | BPSK<br>QPSK<br>16QAM<br>64QAM | BPSK<br>QPSK<br>16QAM<br>64QAM<br>256QAM | BPSK<br>QPSK<br>16QAM<br>64QA | BPSK<br>QPSK<br>16QAM<br>64QAM<br>256QAM<br>1024QAM<br>4096QAM |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | M 256QAM 1024 QAM | |
| Stream | | | | 4x4 | 8x8 | 8 Steam OFDMA | 16 Stream OFDMA |

*4.2 Data Link Layer and Frame Structure*

On a network, the data link layer creates and breaks connections between two physically connected nodes. Packets are divided into frames and sent from the source to the destination. It contains two sub-layers. That is LLC and MAC layers.

- **LLC (Logical Link Control):**

Layer is It is responsible for transferring the packets to the Network layer of the receiver that is receiving. It also provides flow control.

**MAC (Media access control)**

This layer is a link between the Logical Link Control layer and the network's physical layer. It is employed in the network packet transfer process.

- The protocol also defines other types of frames that are responsible for the maintenance and management of the link. The general structure of the frame is as shown in the Figure 3.2.13

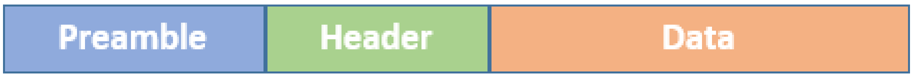


Figure 4 3: Frame Structure

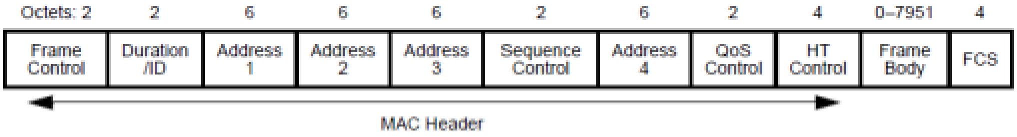
**Preamble**

Preamble is used to send and retrieve the clock(time) synchronization signal which works in a 56-bit alternate binary number that is 1010101010. The preamble marks the start of the frame. It helps the receiver extract the header and the data from the modulated signals that are sent over the channel.

**Header**

The Wi-Fi Header is made up of a two octets (two Bytes) frame control field. The Header keeps the record about where the frame is going on that includes data rate, encryption and more. It contains four address fields are source, destination, transmitter and receiver. Header has 9 major fields. These are:

Table 4 2: Mac Header



| Octets: 2 | 2 | 6 | 6 | 6 | 2 | 6 | 2 | 4 | 0–7951 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frame Control | Duration /ID | Address 1 | Address 2 | Address 3 | Sequence Control | Address 4 | QoS Control | HT Control | Frame Body | FCS |

| | | |
|---|---|---|
| 1. Frame Control | => | 2 Bytes |
| 2. Duration/ID | => | 2 Bytes |
| 3. Address 1 | => | 2 Bytes |
| 4. Address 2 | => | 2 Bytes |
| 5. Address 3 | => | 2 Bytes |
| 6. Sequence Control | => | 2 Bytes |
| 7. Address 4 | => | 2 Bytes |
| 8. QoS Control | => | 2 Bytes |
| 9. HT Control | => | 4 Bytes |

Table 4 3: Size of MAC Header

**Frame Control:**

All the bits in the Frame Control field are used according to rules described on that field. These Frame Control bits may affect the interpretation of other fields in the MAC header file and address fields which depend on the value of the ToDS and FromDS bits.

**Duration ID:**

The Duration/ID field carries the value of the NAV (Network Allocation Vector). The time is specified by the NAV to restrict access to the medium. There are four rules for the **Duration/ID field.**

1. Transmitted frames during the contention-free period set the Duration field to 32,768.
2. Frames transmitted to a broadcast or multicast destination that is the group bit set address 1 has a duration of 0. Receivers do not acknowledge such frames and they are not a part of an atomic exchange. Therefore, contention-based access to the

medium can begin after the conclusion of a broadcast or multicast data frame. In this case NAV is used to protect access to the transmission medium for a frame exchange sequence.

3. If the Frame Control field is zero, then no more fragments remain in the frame. The only final fragment need reserve the medium for its own acknowledgement. The Duration field is set to the amount of time required for one short inter frame space and the fragment acknowledgement. Figure below shows its process.

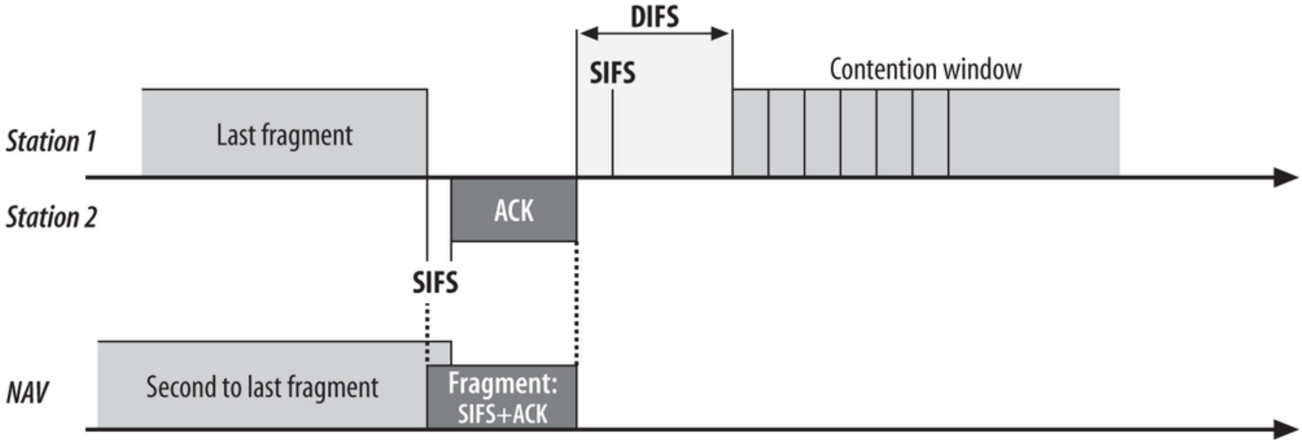


Figure 4 4: Duration setting on final fragment

More fragments will remain if the Frame Control field is set to 1, even if there are more fragments. The Duration field is set to the amount of time required for transmission of two acknowledgements, plus three short inter-frame spaces, plus the time required for the next fragment. Figure below shows the process.

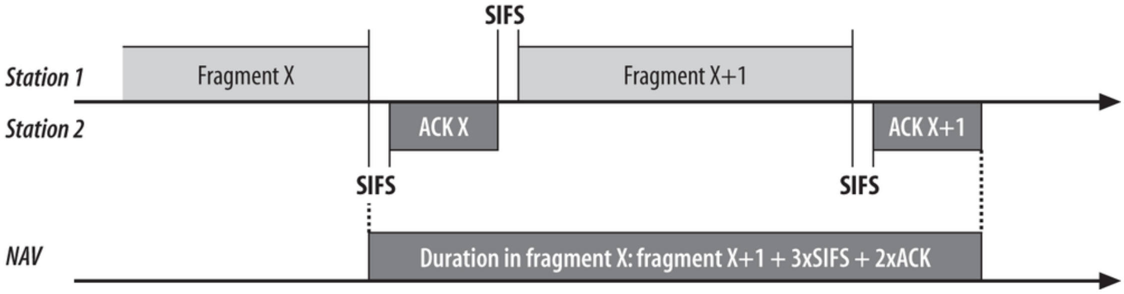


Figure 4 5: Duration settings on non-final fragment

Address 1 to Address 4:

There are altogether four addresses from address 1 to address 4. The distribution system bits that are set determine the quantity and purpose of address fields; so, the type of network being used indirectly determines how address fields 1 through 4 are used. Table below shows the use of the four address fields in the data frame.

Table 4 4: Summaries of the address fields

| Function | ToDS | FromDS | Address 1 (receiver) | Address 2 (Transmitter) | Address 3 | Address 4 |
|---|---|---|---|---|---|---|
| **IBSS** | 0 | 0 | DA | SA | BSSID | NOT USED |
| **To AP** | 1 | 0 | BSSID | SA | DA | NOT USED |
| **From AP** | 0 | 1 | DA | BSSID | SA | NOT USED |
| **WDS** | 1 | 1 | RA | TA | DA | SA |

Address 1 indicates the receiver of the frame and address 2 indicates the transmitter of the frame used to send acknowledgment. The address 3 field is used for filtering by access point and distribution system whereas address 4 sometimes used and not used depends on the function.

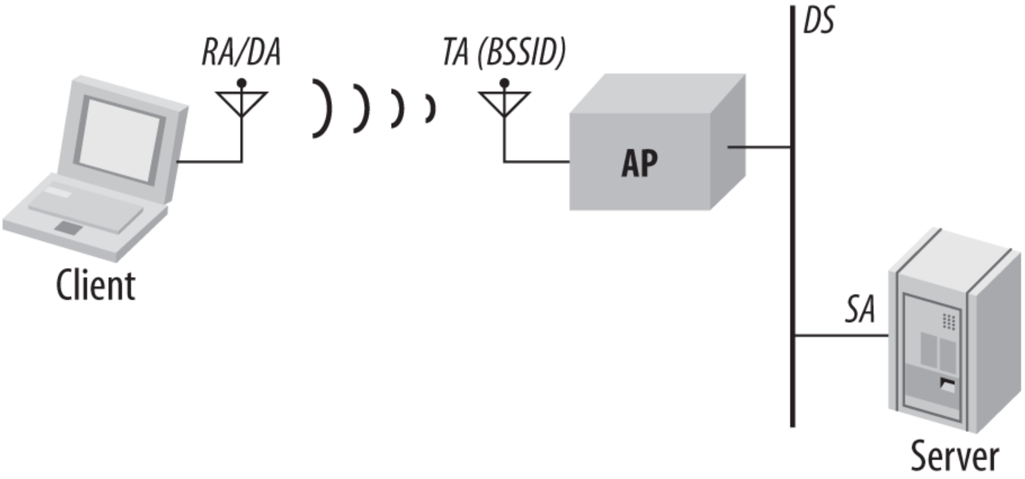


Figure 4 6: Address field usage in frames from the distribution system

**Sequence Control:**

Sequence control contains sequence number and the fragment number. It indicates the sequence number of each frame and then each frame sent of a fragment frame.

**QoS Control:**

QoS stands for Quality of Service and it consists of 16-bit field that identifies the QoS parameter of a data frame. It is comprised of five subfields as shown below.

a) Traffic Identifier (TID)

b) End of Service Period (EOSP)

c) ACK Policy

d) Reserved

e) TXOP Limit, TXOP Duration, AP PS Buffer State and Queue Size

**HT Control:**

In the IEEE 802.11 wireless LAN protocols, a MAC frame is constructed of common fields and specific fields. The Frame Control is the first two octets that a station transmits. The first three subfields within the frame control and the last field are always present in all types of 802.11 frames

**Frame Control**

The Frame Control field in the Wi-Fi header is a 2 Bytes long field that acts as a control for the frame. As demonstrated in the image below, it is separated into 11 subfields. A Wi-Fi frame can be of three type namely Management frame, Control Frame and Data Frame.

• Management Frames

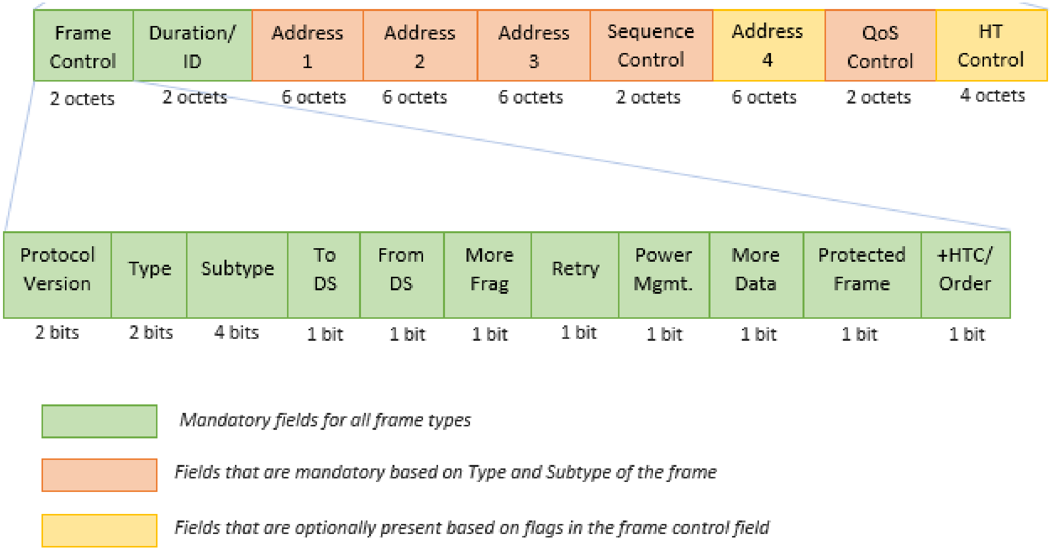


Figure 4 7: Frame control Field and Sub Fields.

### 1. Protocol Version (2-bits)

Protocol version consists of 2 bits and it is simply used to indicate protocol version of 802.11 is being used by the frame. Default value is always set to “0” as currently one version of 802.11 technology exist.

### 2. Type (2-bits)

Type field consists of 2 bits and it can be categories into 3.

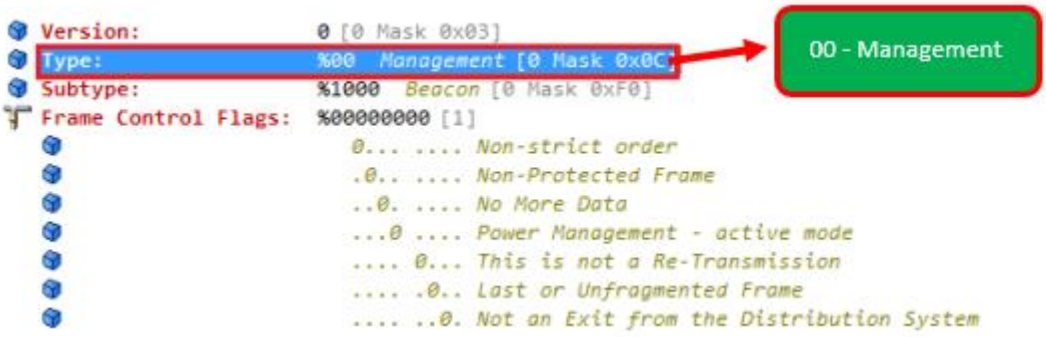
1. Management
2. Control
3. Data

Wireless frames defined in the standard. The bit value of the "Type" field with regard to each distinct type of frame is displayed below.

00– Management Frame

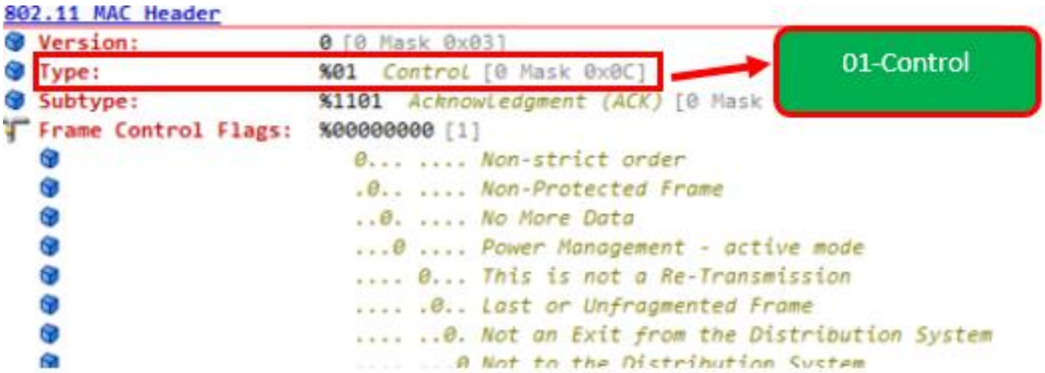
01– Control Frame

10– Data Frame



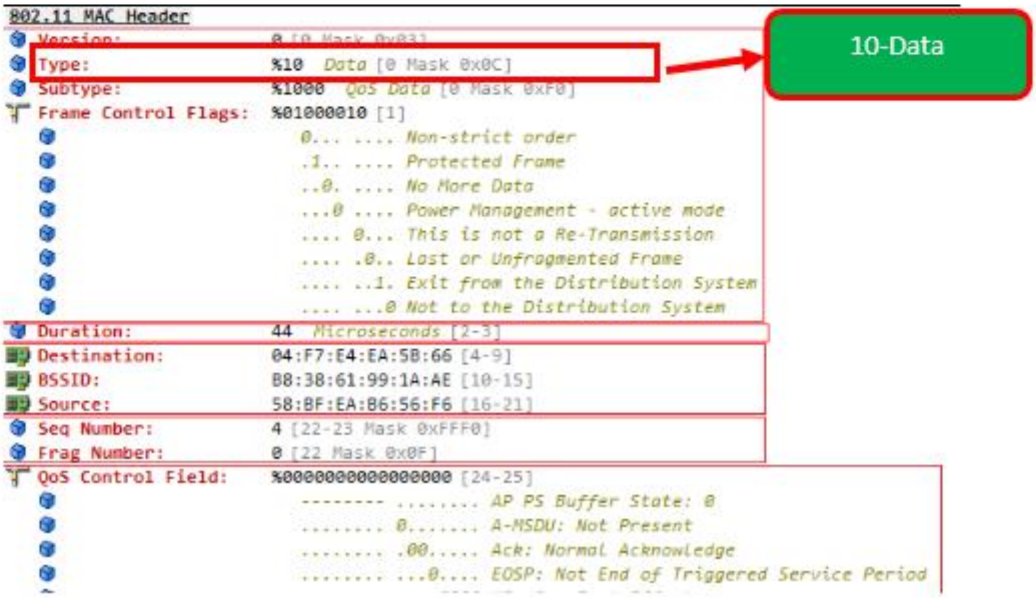
```
Version:              0 [0 Mask 0x03]
Type:                 %00  Management [0 Mask 0x0C]
Subtype:              %1000  Beacon [0 Mask 0xF0]
Frame Control Flags:  %00000000 [1]
                          0... .... Non-strict order
                          .0.. .... Non-Protected Frame
                          ..0. .... No More Data
                          ...0 .... Power Management - active mode
                          .... 0... This is not a Re-Transmission
                          .... .0.. Last or Unfragmented Frame
                          .... ..0. Not an Exit from the Distribution System
```

Figure 4 8: Management Frame



```
802.11 MAC Header
Version:              0 [0 Mask 0x03]
Type:                 %01  Control [0 Mask 0x0C]
Subtype:              %1101  Acknowledgment (ACK) [0 Mask
Frame Control Flags:  %00000000 [1]
                          0... .... Non-strict order
                          .0.. .... Non-Protected Frame
                          ..0. .... No More Data
                          ...0 .... Power Management - active mode
                          .... 0... This is not a Re-Transmission
                          .... .0.. Last or Unfragmented Frame
                          .... ..0. Not an Exit from the Distribution System
                          .... ...0 Not to the Distribution System
```

Figure 4 9: Control Frame



```
802.11 MAC Header
Version:                 0 [0 Mask 0x03]
Type:                    %10  Data [0 Mask 0x0C]
Subtype:                 %1000  QoS Data [0 Mask 0xF0]
Frame Control Flags:     %01000010 [1]
                           0... .... Non-strict order
                           .1.. .... Protected Frame
                           ..0. .... No More Data
                           ...0 .... Power Management - active mode
                           .... 0... This is not a Re-Transmission
                           .... .0.. Last or Unfragmented Frame
                           .... ..1. Exit from the Distribution System
                           .... ...0 Not to the Distribution System
Duration:                44  Microseconds [2-3]
Destination:             04:F7:E4:EA:5B:66 [4-9]
BSSID:                   B8:38:61:99:1A:AE [10-15]
Source:                  58:BF:EA:B6:56:F6 [16-21]
Seq Number:              4 [22-23 Mask 0xFFF0]
Frag Number:             0 [22 Mask 0x0F]
QoS Control Field:       %0000000000000000 [24-25]
                           -------- ........ AP PS Buffer State: 0
                           ........ 0....... A-MSDU: Not Present
                           ........ .00..... Ack: Normal Acknowledge
                           ........ ...0.... EOSP: Not End of Triggered Service Period
```

Figure 4 10: Data Frame

3. **Subtype (4-bits)**

Subtype field consists of 4-bits and it can be categories into three that is management, control and data frames. To differentiate them there are 4-bits of Subtype field is required.

Table 4 5: Subtype

| Type Value | Type Description | Subtype Value | Subtype Description |
|---|---|---|---|
| 00 | Management | 1000 | Beacon |
| 00 | Management | 1010 | Disassociation |
| 01 | Control | 1010 | Block ACK |
| 01 | Control | 1011 | RTS |
| 01 | Control | 1101 | ACK |
| 10 | Data | 0000 | Data |
| 10 | Data | 0100 | Null – No Data |
| 10 | Data | 1000 | QoS Data |
| 10 | Data | 1100 | QoS Null – No Data |

4. **To DS (1-bit)**

This field consists of 1-bit and When it set to “1” that indicate data frame is going from client station (STA) to Distribution System (DS)

**5. From DS (1-bit)**

This field is made up of one bit, and when it is set to "1," it indicates that the data frame is being sent from the Distribution System (DS) to the client station (STA). The combination of fields labeled "00, 01, 10, & 11" in the To DS & From DS fields indicates distinct scenarios.

To DS=0, From DS=0

– It can be management or control frames where it does not go to DS

– Station to Station communication in IBSS

– STSL: Station to Station Link: direct client-to-client data frame exchange.

To DS=0, From DS=1

– Streaming from an AP downstream to a client station.

To DS=1, From DS=0

– Traffic going upstream from an AP to a client station.

To DS=1, From DS=1

Data frames uses four address format. This typically happens when a wireless distribution system (WDS), such as a mesh network or wireless bridge, is in operation.

**6. More Fragments (1-bit)**

More Fragments field also consists of 1-bit. If this bit is set to “1” that indicates that frame (data or management type) have another fragment of the current MSDU or current MMPDU to follow. MAC layer fragments only those frame having unicast receiver address & never fragments broadcast or multicast frames (as those never get acknowledged)

**7. Retry (1-bit)**

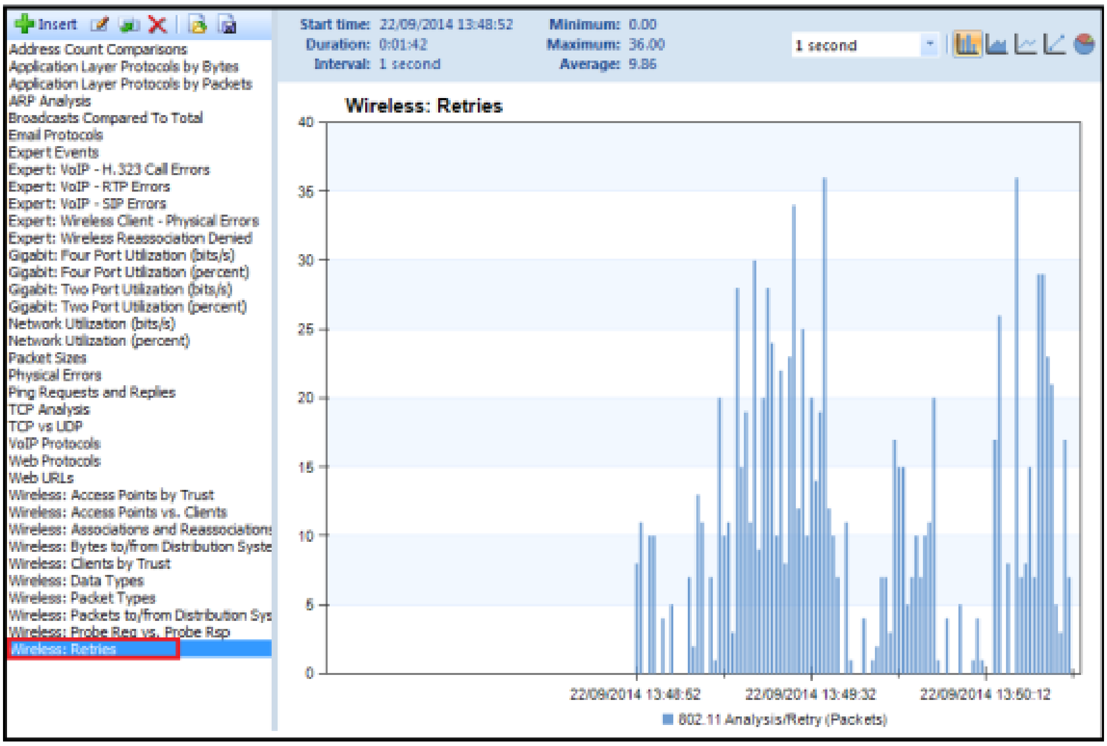
Retry field consists of 1-bit. A management frame or data frame with the Retry bit set to "1" indicates to the TX radio that the frame being sent is a "retransmission." If a TX station did not receive an ACK for a unicast frame, then frame will be retransmitted. In certain cases, where ACK is not used. For example, in RTS/CTS frame exchange, CTS server as ACK). For example, using wireless protocol analyzer here onmiPeel Analyzer has shown below.

L2 retransmissions affect WLAN performance in two ways.

▪ Increases overhead resulting decreasing throughput

- impact timely delivery of application traffic (affect voice/video services

Typically, most data applications operate in environment up to 10% retransmissions without any noticeable degradation in performance however time sensitive applications like VoIP required less than 5% retransmissions.

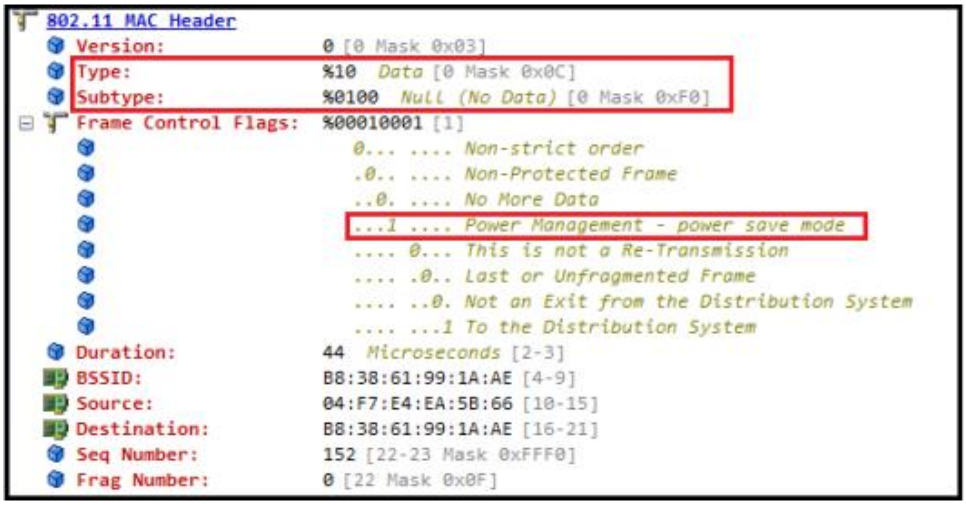

Insert
Address Count Comparisons
Application Layer Protocols by Bytes
Application Layer Protocols by Packets
ARP Analysis
Broadcasts Compared To Total
Email Protocols
Expert Events
Expert: VoIP - H.323 Call Errors
Expert: VoIP - RTP Errors
Expert: VoIP - SIP Errors
Expert: Wireless Client - Physical Errors
Expert: Wireless Reassociation Denied
Gigabit: Four Port Utilization (bits/s)
Gigabit: Four Port Utilization (percent)
Gigabit: Two Port Utilization (bits/s)
Gigabit: Two Port Utilization (percent)
Network Utilization (bits/s)
Network Utilization (percent)
Packet Sizes
Physical Errors
Ping Requests and Replies
TCP Analysis
TCP vs UDP
VoIP Protocols
Web Protocols
Web URLs
Wireless: Access Points by Trust
Wireless: Access Points vs. Clients
Wireless: Associations and Reassociations
Wireless: Bytes to/from Distribution Syste
Wireless: Clients by Trust
Wireless: Data Types
Wireless: Packet Types
Wireless: Packets to/from Distribution Sys
Wireless: Probe Req vs. Probe Rsp
Wireless: Retries
1 second
Wireless: Retries
40
35
30
25
20
15
10
5
0
22/09/2014 13:48:52
22/09/2014 13:49:32
22/09/2014 13:50:12
802.11 Analysis/Retry (Packets)

Start time: 22/09/2014 13:48:52
Duration: 0:01:42
Interval: 1 second
Minimum: 0.00
Maximum: 36.00
Average: 9.86

Figure 4 11: Retransmissions

## 8. Power Management (1-bit)

The Power Management field consists of 1-bit as well. It helps to shut-down some of the transceivers components for a period of time to converse power. By setting the Power Save mode bit to 1, the station signals that it is in Power Save mode. As we can see below “Null” data frames used to inform AP about clients in Power Save mode.



```
802.11 MAC Header
    Version:                   0 [0 Mask 0x03]
    Type:                      %10   Data [0 Mask 0x0C]
    Subtype:                   %0100  Null (No Data) [0 Mask 0xF0]
⊟   Frame Control Flags:       %00010001 [1]
                                 0... .... Non-strict order
                                 .0.. .... Non-Protected Frame
                                 ..0. .... No More Data
                                 ...1 .... Power Management - power save mode
                                 .... 0... This is not a Re-Transmission
                                 .... .0.. Last or Unfragmented Frame
                                 .... ..0. Not an Exit from the Distribution System
                                 .... ...1 To the Distribution System
    Duration:                  44  Microseconds [2-3]
    BSSID:                     B8:38:61:99:1A:AE [4-9]
    Source:                    04:F7:E4:EA:5B:66 [10-15]
    Destination:               B8:38:61:99:1A:AE [16-21]
    Seq Number:                152 [22-23 Mask 0xFFF0]
    Frag Number:               0 [22 Mask 0x0F]
```

Figure 4 12: Power Management

**9. More Data (1-bit)**

The Data field consists of 1-bit. It is used when a client associates to an access point then client services an association identifier. AP uses AID to keep track of stations associated with the AP. If AP is buffering data for a station during power save mode and when AP transmits its next beacon, the AID of the station will be seen in the field called “Traffic Indication Map (TIM). When the station receives the beacon during the awake state, it checks to see whether its AID is set in TIM or not. If so, the station will remain awake & will send a PS-Poll frame to the AP and then AP will send buffered unicast frames to the station. To indicate there are more frames AP will set the “More Data” field to 1, so the station can awake to receive all of those frames. Below diagram summarizes this process.

**FIGURE 3.13** Power Save mode

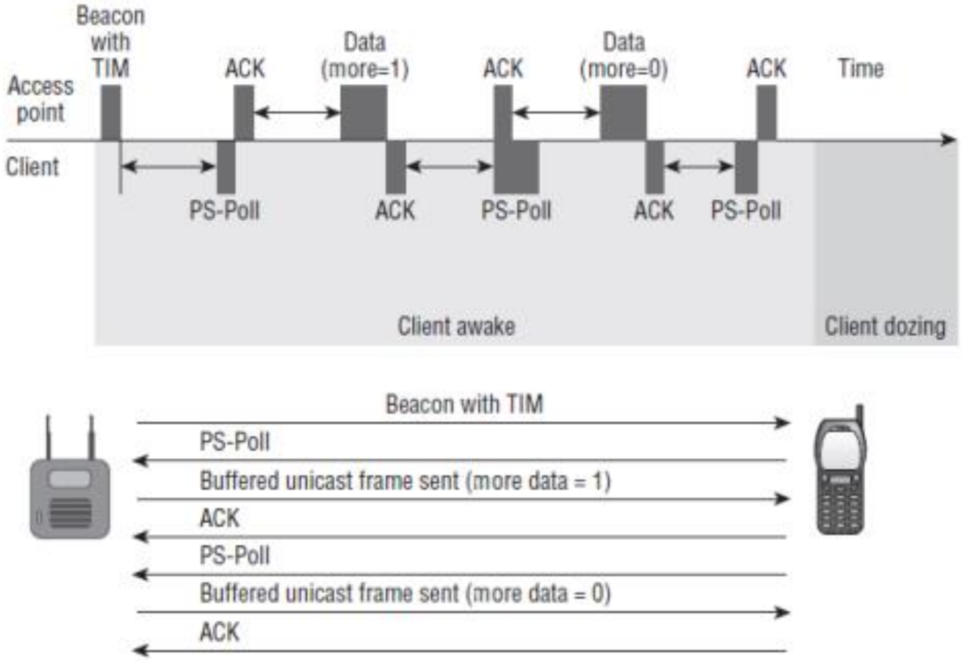
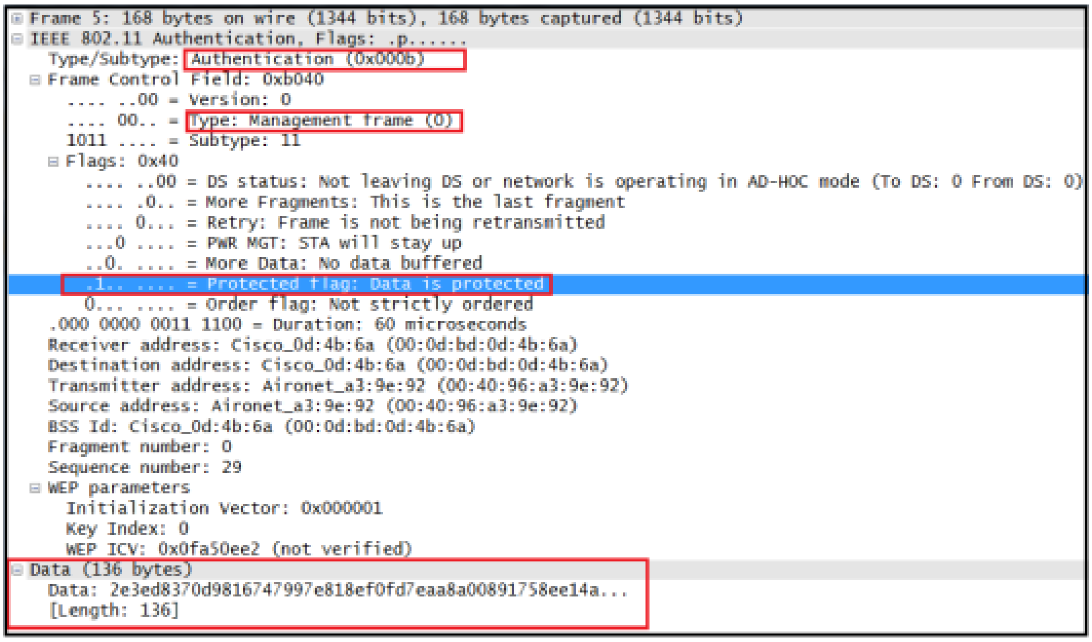


Figure 4 13: Power Save Mode

**1. Protected Frame (1-bit)**

Protected Frame also consists of 1-bit. This frame is used to check a data frame is encrypted. Below figure shows a data is protected or not by setting bit to “1”.

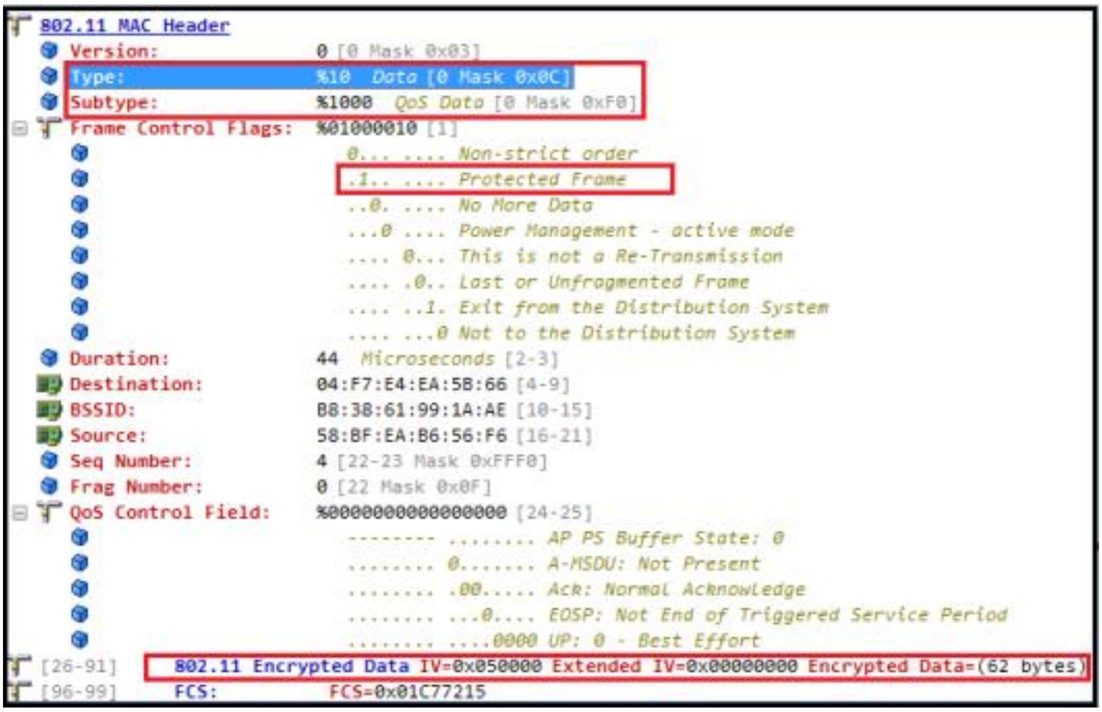


```
⊞ Frame 5: 168 bytes on wire (1344 bits), 168 bytes captured (1344 bits)
⊟ IEEE 802.11 Authentication, Flags: .p......
    Type/Subtype: Authentication (0x000b)
  ⊟ Frame Control Field: 0xb040
      .... ..00 = Version: 0
      .... 00.. = Type: Management frame (0)
      1011 .... = Subtype: 11
    ⊟ Flags: 0x40
        .... ..00 = DS status: Not leaving DS or network is operating in AD-HOC mode (To DS: 0 From DS: 0)
        .... .0.. = More Fragments: This is the last fragment
        .... 0... = Retry: Frame is not being retransmitted
        ...0 .... = PWR MGT: STA will stay up
        ..0. .... = More Data: No data buffered
        .1.. .... = Protected flag: Data is protected
        0... .... = Order flag: Not strictly ordered
    .000 0000 0011 1100 = Duration: 60 microseconds
    Receiver address: Cisco_0d:4b:6a (00:0d:bd:0d:4b:6a)
    Destination address: Cisco_0d:4b:6a (00:0d:bd:0d:4b:6a)
    Transmitter address: Aironet_a3:9e:92 (00:40:96:a3:9e:92)
    Source address: Aironet_a3:9e:92 (00:40:96:a3:9e:92)
    BSS Id: Cisco_0d:4b:6a (00:0d:bd:0d:4b:6a)
    Fragment number: 0
    Sequence number: 29
  ⊟ WEP parameters
      Initialization Vector: 0x000001
      Key Index: 0
      WEP ICV: 0x0fa50ee2 (not verified)
⊟ Data (136 bytes)
    Data: 2e3ed8370d9816747997e818ef0fd7eaa8a00891758ee14a...
    [Length: 136]
```

Figure 4 14: Protected Frame by setting bit to “1”

## 2. +HTC/Order (1-bit)

Order field consists of 1-bit and if it set to “1” in any non-QoS data frame then higher layer has requested that the data be sent using a strictly ordered class of service otherwise it is set.



```
802.11 MAC Header
  Version:                0 [0 Mask 0x03]
  Type:                   %10  Data [0 Mask 0x0C]
  Subtype:                %1000  QoS Data [0 Mask 0xF0]
⊟ Frame Control Flags:    %01000010 [1]
                            0... .... Non-strict order
                            .1.. .... Protected Frame
                            ..0. .... No More Data
                            ...0 .... Power Management - active mode
                            .... 0... This is not a Re-Transmission
                            .... .0.. Last or Unfragmented Frame
                            .... ..1. Exit from the Distribution System
                            .... ...0 Not to the Distribution System
  Duration:               44  Microseconds [2-3]
  Destination:            04:F7:E4:EA:5B:66 [4-9]
  BSSID:                  B8:38:61:99:1A:AE [10-15]
  Source:                 58:BF:EA:B6:56:F6 [16-21]
  Seq Number:             4 [22-23 Mask 0xFFF0]
  Frag Number:            0 [22 Mask 0x0F]
⊟ QoS Control Field:      %0000000000000000 [24-25]
                            -------- ........ AP PS Buffer State: 0
                            ........ 0....... A-MSDU: Not Present
                            ........ .00..... Ack: Normal Acknowledge
                            ........ ...0.... EOSP: Not End of Triggered Service Period
                            ........ ....0000 UP: 0 - Best Effort
[26-91]   802.11 Encrypted Data IV=0x050000 Extended IV=0x00000000 Encrypted Data=(62 bytes)
[96-99]   FCS:          FCS=0x01C77215
```

Figure 4 15: HTC/Order Bit

### Management Frames

Management frames that help to control of the link. These frames setup the link and tear it down once the communication is complete. Management Frames are described below:

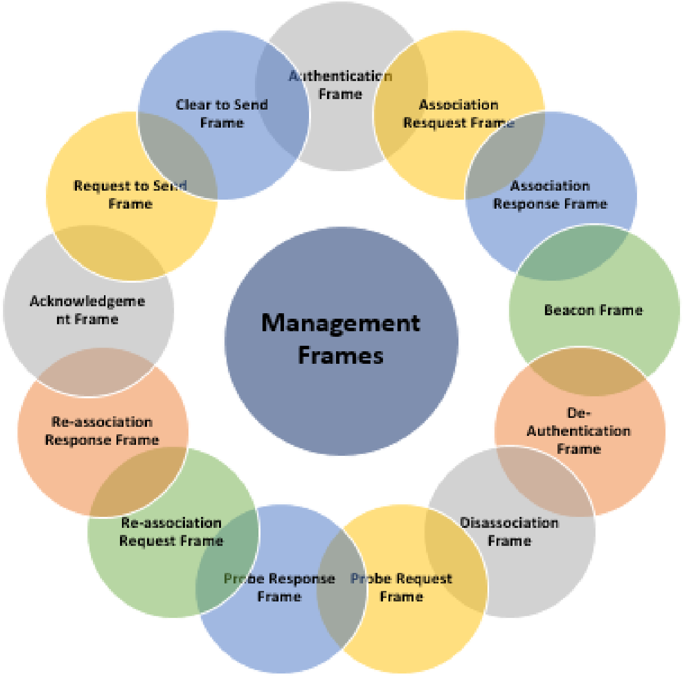
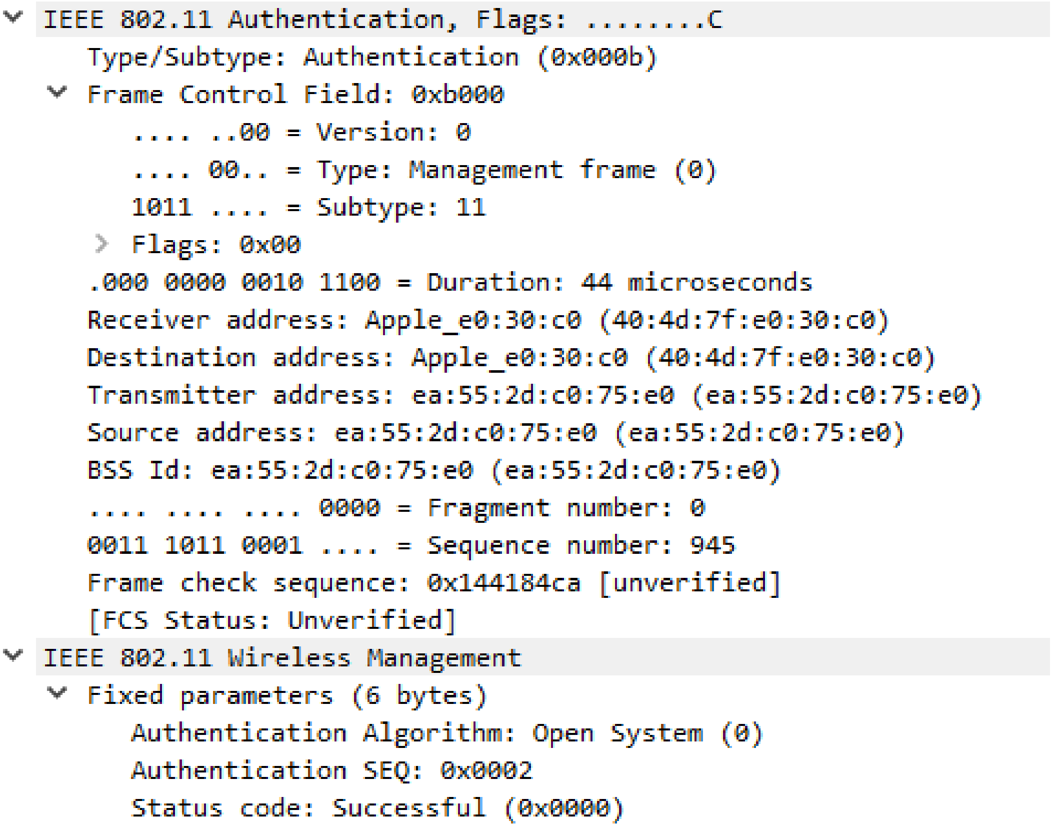


Figure 4 16: Management Frames

### a) Authentication Frame

Authentication frames are used to join the BSS (Basic Service Set) as part of the open system authentication process. It takes place with an access point when the link setup between the access point (AP) and the user device. It helps to establish a connection to the network.



```
IEEE 802.11 Authentication, Flags: ........C
    Type/Subtype: Authentication (0x000b)
    Frame Control Field: 0xb000
        .... ..00 = Version: 0
        .... 00.. = Type: Management frame (0)
        1011 .... = Subtype: 11
        Flags: 0x00
    .000 0000 0010 1100 = Duration: 44 microseconds
    Receiver address: Apple_e0:30:c0 (40:4d:7f:e0:30:c0)
    Destination address: Apple_e0:30:c0 (40:4d:7f:e0:30:c0)
    Transmitter address: ea:55:2d:c0:75:e0 (ea:55:2d:c0:75:e0)
    Source address: ea:55:2d:c0:75:e0 (ea:55:2d:c0:75:e0)
    BSS Id: ea:55:2d:c0:75:e0 (ea:55:2d:c0:75:e0)
    .... .... .... 0000 = Fragment number: 0
    0011 1011 0001 .... = Sequence number: 945
    Frame check sequence: 0x144184ca [unverified]
    [FCS Status: Unverified]
IEEE 802.11 Wireless Management
    Fixed parameters (6 bytes)
        Authentication Algorithm: Open System (0)
        Authentication SEQ: 0x0002
        Status code: Successful (0x0000)
```

Figure 4 17: Authentication Frame

BSS helps to connect wireless clients to wireless networks through an access point. A BSS in the one where we use mostly wireless networks. BSS helps that the AP is responsible for the wireless network. BSS network shown below.

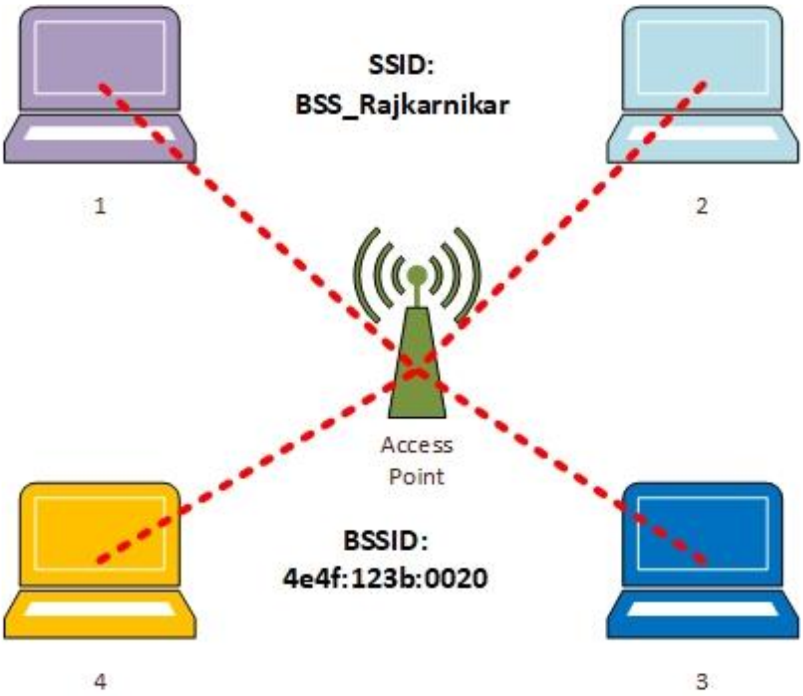


Figure 4 18: BSS Network

**b) Association Request Frame**

Association request frame makes the device ready to send data on the network and allocate resources for the device.

**c) Association Response Frame**

Association response frame helps to respond to message sent by the access point. The response frame may be a positive response or a negative response to the device.

**d) Beacon Frame**

Beacon frame allows AP to broadcast after a fixed interval of time.  To advertise the SSIDs they support, APs send beacons at a regular interval known as the target beacon transmit time (TBTT). It informs the devices that are trying to the AP of the various characteristics of the AP. For example, name, operating frequency, transfer rate and different types of encryption scheme that are used.

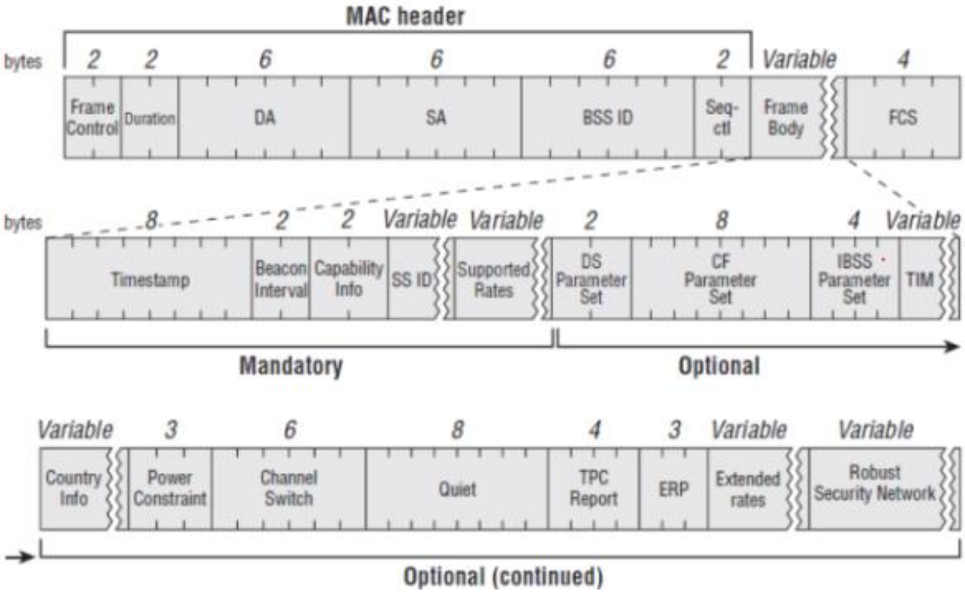


Figure 4 19: Beacon Frame Structure

**e) De-Authentication Frame**

It is a complement of authentication frame that helps to disconnect devises to the AP and network when the user wants to disconnect.

**f) Disassociation Frame**

Disassociation frame is used to reset the associated client. The authentication process takes place to associate so, if a station is de-authenticate then it is disassociated. Disassociation Frame is a type of management frame sent from an access point or any other hub. It is used to terminate the access points Disassociation Frame is a complement of the Association Frame. It informs the access point that it can de-allocate the resources that it had allocated for the device as the device no longer plans to use the network.

**g) Probe Request Frame**

Probe Request Frame is used to send from one station to another station to get information about that station. For example, to search SSID or active scanning, all the station still goes through each channel in turn, however instead of passive or listening to the signals on that frequency, stations send a request that is known as probe request management frame for asking what network is available on the channel.

**h) Probe Response Frame:**

Probe response frame is the response sent by a station for the probe request.

**i) Re-association Request Frame:**

Re-association Request Frame sends a device depending on signal strength more than the current AP and it is a frame that is sent when a device moves out of the range of one access point and moves into the range of another.

**x. Re-association Response Frame:**

Re-Association Response Frame is the response frame that is sent in response to the Re-Association Request. The response may be a positive response or a negative response depends on AP.

**xii. Acknowledgement (ACK) Frame:**

It helps to send acknowledgement from destination to the source.

**xiii. Request to Send (RTS) Frame:**

It is the request to send that acts as an optional contention control over the station or network.

**xiv. Clear to Send (CTS) Frame:**

It is the optional Clear to Send Frame that is sent in response to the Request to Send Frame on AP.

The term "web technology" refers to the various programming languages and multimedia applications that are used to create dynamic websites. It refers to the World Wide Web and focuses on the technologies used in creation, maintenance and development of web based applications. Web technology involves the use of HTML and CSS. According to White Hat security's report 2015 more than 86% of all websites have one or more critical vulnerability and the likelihood of information leakage is 56%. In this chapter we apply our AB-IDS methodology to design IDS' to detect malicious HTML files. The open vulnerabilities for the application or applications covered in this report are grouped by White Hat vulnerability class in the following table.

Table 4 6: Vulnerabilities by White Hat

| **White Hat Vulnerability Classes** | **No. of Occurrences** | **Risk Rating** | | | | |
|---|---|---|---|---|---|---|
| | | Critical | High | Medium | Low | Note |
| SQL Injection | 4 | 4 | 0 | 0 | 0 | 0 |
| Clear Text Password | 1 | 0 | 0 | 0 | 0 | 1 |
| Cross Site Scripting | 1 | 0 | 0 | 1 | 0 | 0 |
| Debug Enabled | 1 | 0 | 0 | 1 | 0 | 0 |
| DoS | 1 | 0 | 0 | 1 | 0 | 0 |
| Information Leakage | 1 | 0 | 0 | 1 | 0 | 0 |
| Path Traversal | 1 | 0 | 0 | 1 | 0 | 0 |
| URL Redirect Abuse | 1 | 0 | 0 | 0 | 0 | 1 |
| Total | 11 | 4 | 0 | 5 | 0 | 2 |

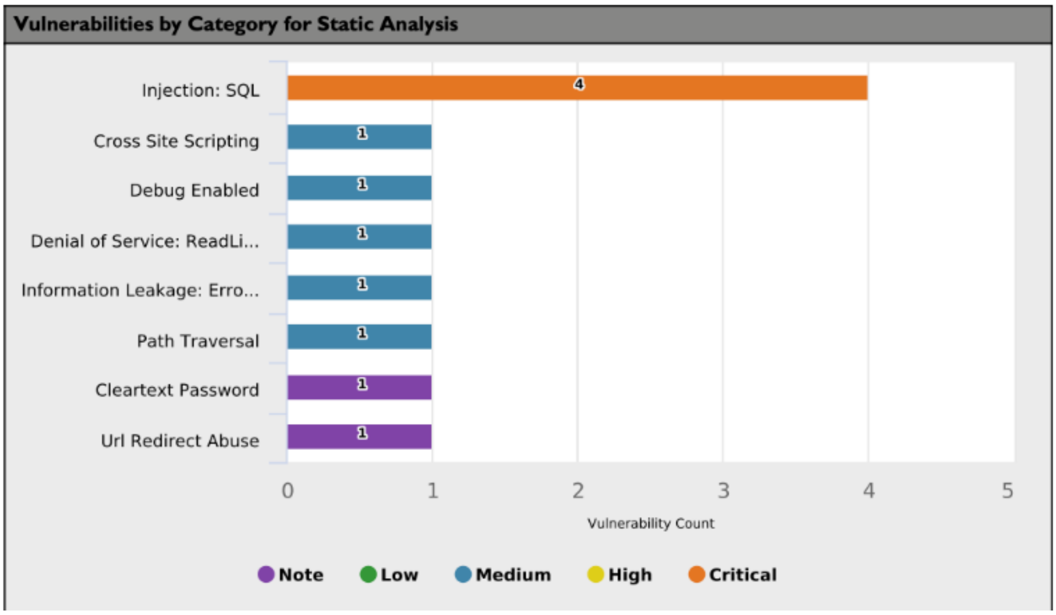


Figure 4 20: Bar Chart Summarizes by White Hat vulnerability

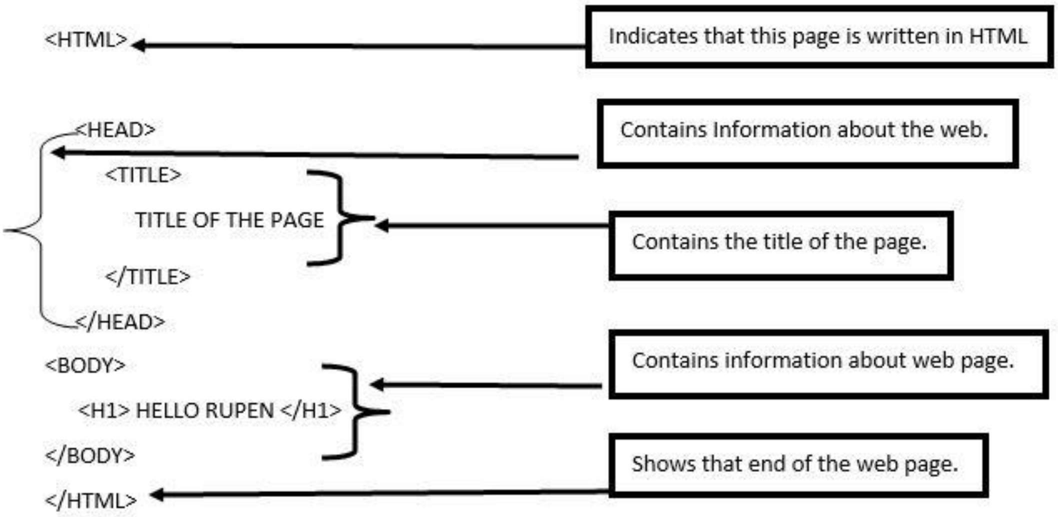
Some White Hat vulnerability class are associated with multiple Open Worldwide Application Security Project (OWASP) Top 10 Categories as shown below.

Table 4 7: OWASP

| OWASP Category | No. of Occurrences | Critical | High | Medium | Low | Note |
|---|---|---|---|---|---|---|
| **Broken Access Control** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Cryptographic Failure** | 1 | 0 | 0 | 1 | 0 | 0 |
| **Injection** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Insecure Design** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Security Misconfiguration** | 1 | 0 | 0 | 1 | 0 | 0 |
| **Vulnerable and Outdated** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Identification and Authentication** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Software and Data Integrity Failures** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Security Logging and Monitoring** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Server Side Request Forgery** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Total** | 2 | 0 | 0 | 2 | 0 | 0 |

## HTML Protocol

Every facet of our lives, including the economics, education, entertainment, and more, are now impacted by the information services available on the internet. More than a billion websites are already hosted on the internet, and even more people utilize it every day. Numerous heterogeneous devices, both mobile and fixed, use the internet for a variety of purposes. With the advent of the Internet of Things (IoT), this number is predicted to reach more than 50 billion devices. Websites, which are essentially HTML webpages, host the majority of the content that is available online. Web browsers, the software used to access the internet, ask web servers for HTML pages. The HTML file is parsed by the web browser to convert it into a document object model (DOM). The screen rendering of the content is performed on the DOM.



<HTML>
Indicates that this page is written in HTML
<HEAD>
Contains Information about the web.
<TITLE>
TITLE OF THE PAGE
</TITLE>
Contains the title of the page.
</HEAD>
<BODY>
<H1> HELLO RUPEN </H1>
</BODY>
Contains information about web page.
</HTML>
Shows that end of the web page.

Figure 4 21: Structure of HTML

## USING THE PROPOSED METHODOLOGY TO DESIGN AN AB-IDS FOR THE WI-FI PROTOCOL

Four states make up a Wi-Fi state machine when a user is connected to the network during a session.
These are:

**State 1:** The client is unauthenticated and unassociated during this state Wi-Fi is not connected to the network that means the client is not connected in any shape or form to the network. Clients in this state are passing Class 1 frames and contain the following frames:

- Control Frames
- Management Frames

- Data Frames
- Data Frames between STAs and IBSS
- Data Frames between peers using DLS

**State 2:** The client is Authenticated but Unassociated. During this state, we can pass Class 1 frames as well as introduce the ability to pass Class 2 frames which are management frames.

**State 3:** The client is Authenticated and Associated. If we are using 802.1X, it shows the wireless network has been properly connected, but RSN authentication is still pending. At this point allow Class 1, Class 2 and Class 3 frames. Class 3 frames consists of the following:

- Data Frames(ALL)
- Management Frames
- Control Frames

**State 4:** Finally, in this state we have successfully performed all the necessary steps to be connected to the wireless network.

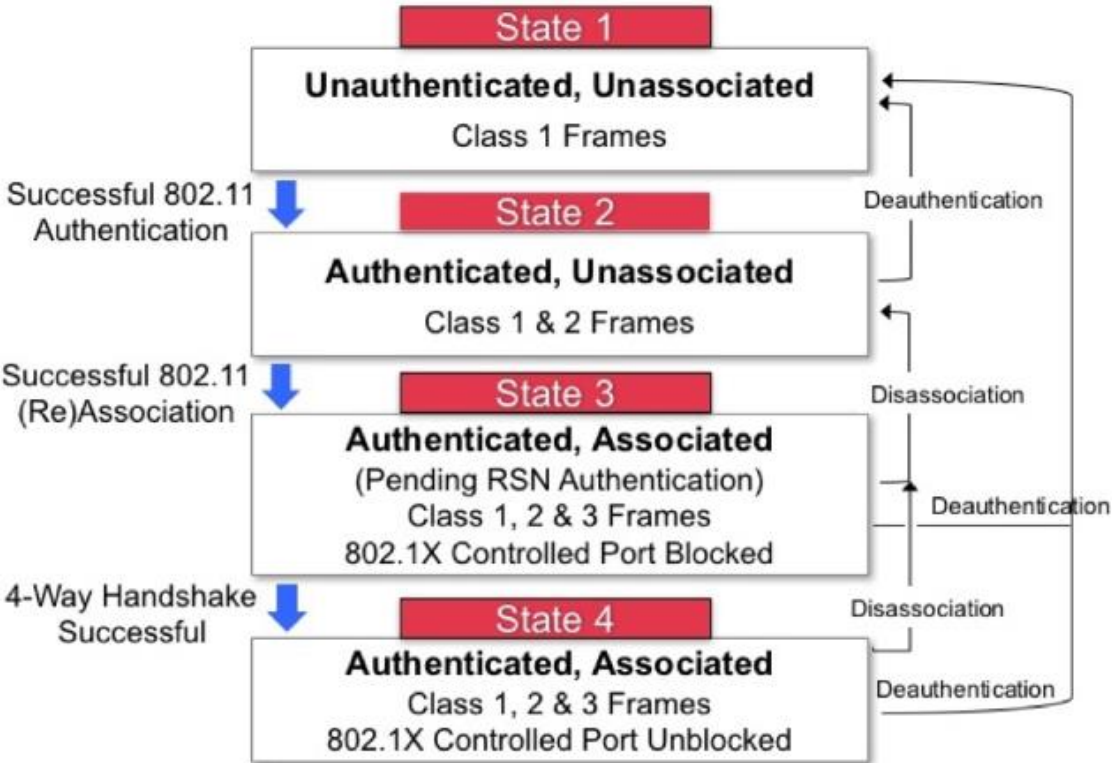

State 1
Unauthenticated, Unassociated
Class 1 Frames
Successful 802.11 Authentication
State 2
Authenticated, Unassociated
Class 1 & 2 Frames
Successful 802.11 (Re)Association
State 3
Authenticated, Associated
(Pending RSN Authentication)
Class 1, 2 & 3 Frames
802.1X Controlled Port Blocked
4-Way Handshake Successful
State 4
Authenticated, Associated
Class 1, 2 & 3 Frames
802.1X Controlled Port Unblocked
Deauthentication
Disassociation
Deauthentication
Disassociation
Deauthentication

Figure 4 22: Wi-Fi state machine diagram

Accordingly, many years ago there had been issues in handling IDS in the Internet of Things. Most of the IDSs are anomaly based systems developed by machine learning and deep learning techniques to provide decision making and intelligent cybersecurity. Well, machine learning and deep learning techniques operate using records and feature’s datasets which are used to train and test the predictive Intrusion Detection System models however not all of the features and records in a datasets are important while training and testing IDS models.

Basically Internet of Things have three-layer scheme.

1. **Application or Upper Layer:**

Application layer provides the network services to the end-users. It is used by end-user software such as web browsers and email clients. It provides protocols that enable applications to communicate with one another and give users relevant data. Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP), Post Office Protocol (POP), Simple Mail Transfer Protocol (SMTP), and Domain Name System (DNS) are few examples of application layer.

2. **Network or Middle Layer:**

It is a layer 3 that manages device addressing and tracks the location of devices on the network. The network layer's primary duty is internetworking. It provides a logical connection between different devices. The upper layer sends packets to a network layer, which transforms them into packets. This process is known as Packetizing. It is achieved by internet protocol (IP).

3. **Perception or Lower Layer:**

The main functionality of the physical layer is to transmit the individual bits from one node to another node through wire and wireless medium. In the Open System Interconnection (OSI) model, it is the lowest layer. It provides the mechanical, electrical and procedural network interface specifications. It establishes, maintains and deactivates the physical connection.

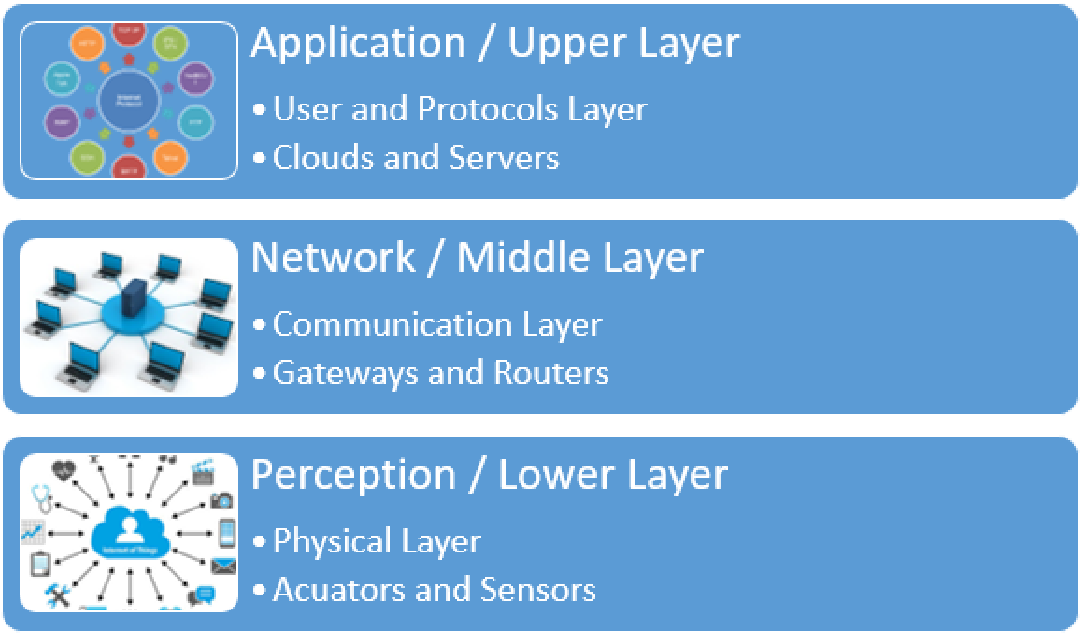


Figure 4 23: Three Layers Scheme

Data engineering and features are formulated for a core phase of machine and deep learning based intrusion detection system which played a major role in making the raw data collected from the IoT for more analysis and predictions. Therefore, it is good to use both data engineering, machine learning and deep learning approaches. Significant and successful efforts have been made to solve the security issues that have arisen for the IoT ecosystem in recent years. Many good techniques for anomaly based IDS models had been implemented for IoT security using machine and deep learning approaches. Only a few number, however, were created through research into the effects of applying various feature selection strategies to raise prediction and classification accuracy.

There are different types of wireless network attacks as shown below.

1. Sniffing

Capturing and tracking network traffic is known as sniffing. It can be done by using software to capture all data packets passing through a given network interface. Sniffing can be performed using specialized hardware or software tools, and can potentially capture sensitive information such as usernames, passwords, and credit card numbers.

There are two types of sniffing attacks.

a. Active Sniffing:

Active sniffing is a type of attack by sending different packets to more than one target on a network to retrieve sensitive data which is carried out via switch. Attackers can often bypass security measures by using special packets and it helps to inject malicious code into target systems that allows attackers to take control of their devices and steal sensitive information.

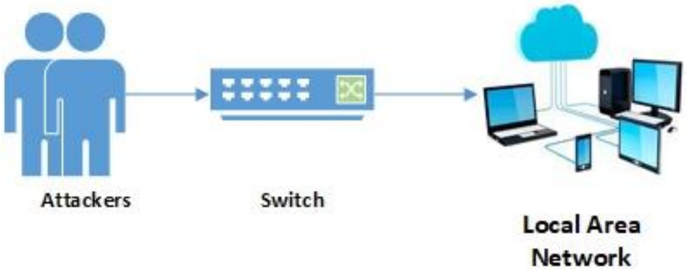


Figure 4 24: Active Sniffing

b. Passive Sniffing:

Passive sniffing is the type of attack that does not interact with the target. They just simply hook on to the network and capture sensitive data/packets through the hub.

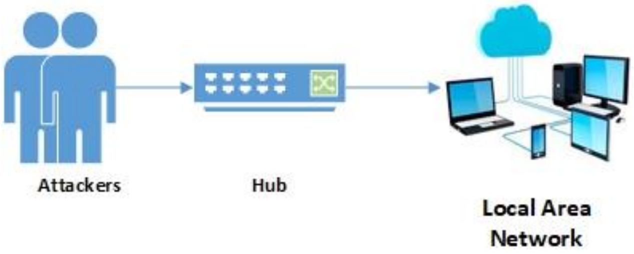


Figure 4 25: Passive Sniffing

2. Rogue access points

Rogue access points are unauthorized Wi-Fi access points that are installed on a network without the knowledge or permission of the network administrator. These access points can be used to intercept and capture network traffic, and can potentially give an attacker access to sensitive information.

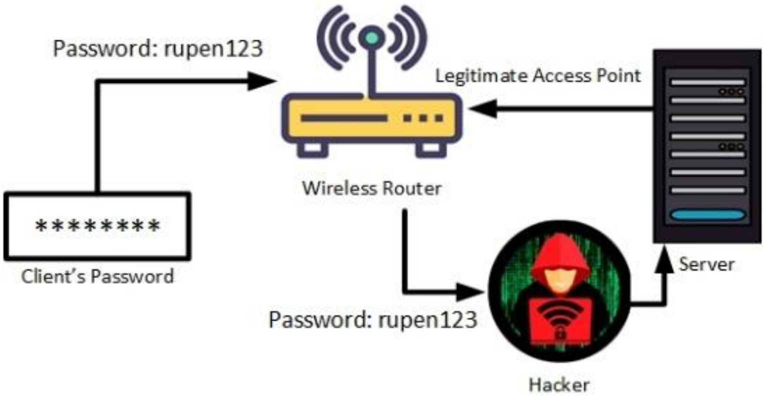


Figure 4 26: Rogue Access Point

3. Wi-Fi evil twins and phishing

Wi-Fi evil twins are rogue access points that are designed to look like legitimate Wi-Fi networks. Phishing is the act of using social engineering to trick users into providing sensitive information, such as usernames and passwords. Wi-Fi evil twins and phishing attacks are often used together to trick users into connecting to a rogue access point and providing sensitive information.

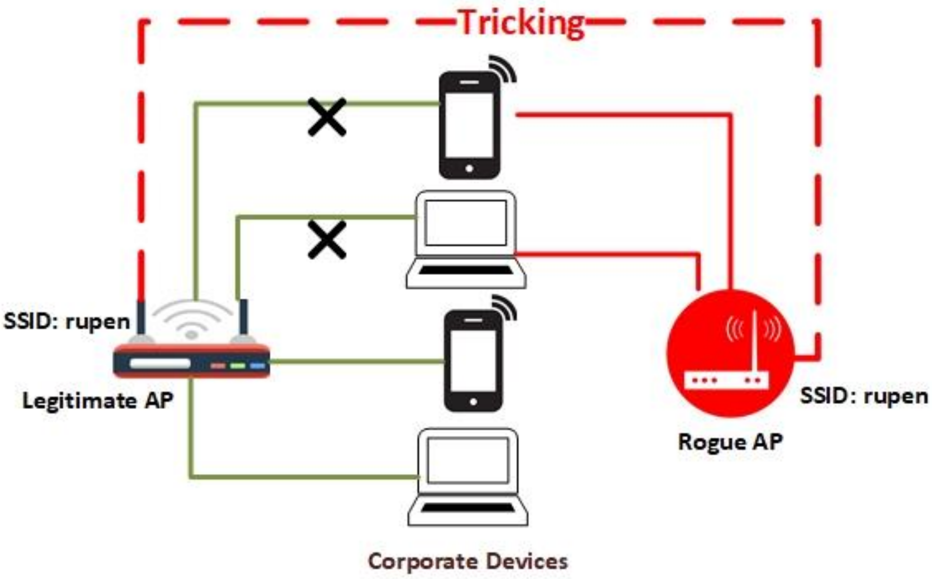


Figure 4 27: Wi-Fi Twins Phishing

4. Spoofing attacks:

Spoofing attacks involve impersonating a legitimate device or network to gain access to sensitive information. There are different types of spoofing attacks as shown below.

- MAC spoofing attack is where a hacker hunts the network for validity and then attempts to start new connections impersonating the authorized devices.

- Frame spoofing, also known as *frame injection*, occurs when attackers send malicious frames that appear to be from legitimate senders.
- IP spoofing takes place when attackers use modified IP packets to hide where the packets originate.
- Data replay occurs when attackers capture wireless data transmission, modify the transmission and resend the modified transmission to a target system.
- Authentication replay happens when attackers capture authentication exchanges between users and reuse those exchanges in attacks.

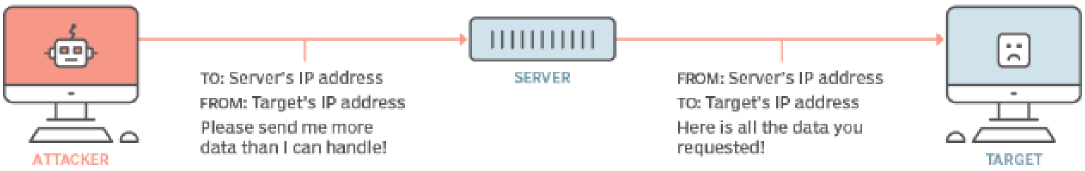

ATTACKER
TO: Server's IP address
FROM: Target's IP address
Please send me more data than I can handle!
SERVER
FROM: Server's IP address
TO: Target's IP address
Here is all the data you requested!
TARGET

Figure 4 28: Spoofing attack

5. Encryption cracking -- WEP/WPA attacks:

WEP and WPA are Wi-Fi encryption protocols that are used to secure wireless networks. However, these protocols have been found to have weaknesses that can be exploited to gain access to the network. Encryption cracking attacks involve using specialized tools to crack the encryption and gain access to the network.

6. MitM attacks:

MitM stands for Man in the Middle that attacks involve intercepting network traffic and modifying it before forwarding it to its destination. This can be used to steal sensitive information or to modify data in transit.

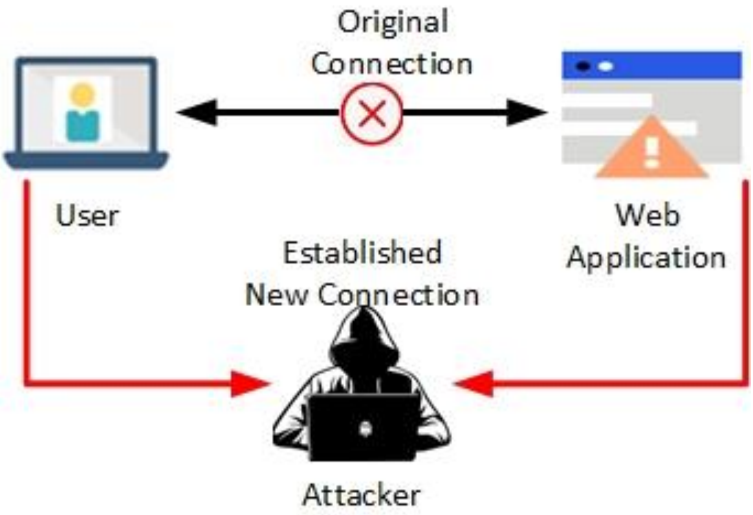


Figure 4 29: Man in the Middle attack

7. DoS attacks: Denial-of-service:

   (DoS) attacks involve overwhelming a network or server with traffic, rendering it unavailable to legitimate users. DoS attacks can be used to disrupt business operations, extort money, or as a distraction while another attack is carried out.

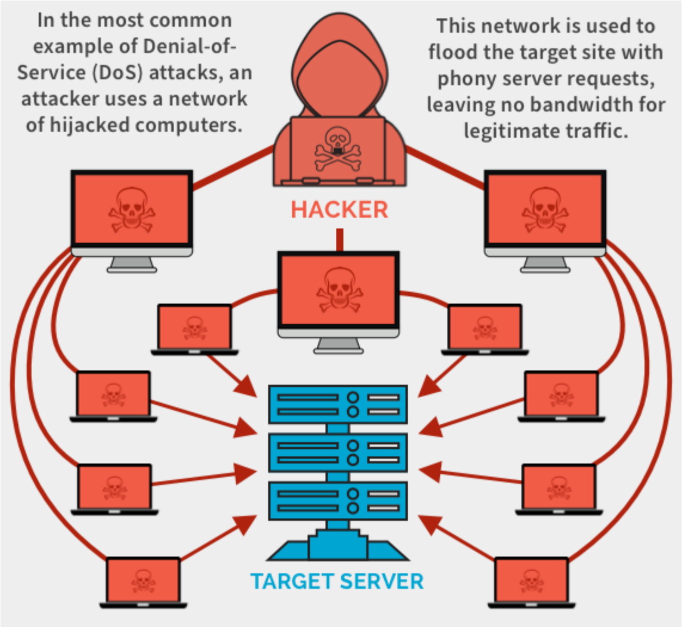

In the most common example of Denial-of-Service (DoS) attacks, an attacker uses a network of hijacked computers.
This network is used to flood the target site with phony server requests, leaving no bandwidth for legitimate traffic.
HACKER
TARGET SERVER

Figure 4 30: DoS attack

8. Wi-Fi jamming:

   Wi-Fi jamming involves flooding a wireless network with interference to disrupt communication. This can be used to disrupt the network or to prevent users from connecting to the network.

9. War driving and shipping attacks:

   Driving about in search of unprotected wireless networks is referred to as "war driving." Attackers can use war driving to find vulnerable networks and gain access to sensitive information. War shipping involves sending a package containing a small computer device that is capable of connecting to wireless networks. The device can be used to gain access to the network and steal sensitive information.

10. Theft and tampering:

    Physical theft and tampering involve stealing or manipulating network equipment, such as routers or switches. Attackers can use this to gain access to the network or to intercept network traffic.

11. Default passwords and SSIDs:

    Default passwords and SSIDs are often used by network administrators to set up wireless networks. However, if these defaults are not changed, attackers can easily gain access to the network by using well-known default passwords or SSIDs. It is important to change default passwords and SSIDs to prevent unauthorized access.

12. Attacks on HTML Files:

    Generally, there are four common ways for inserting malicious code into HTML Files as shown below”

    **Hidden Iframes:** it is an Inline Frame which is a way of loading one web page inside another, usually from a different server. For example:

    <iframe src=”http://www.virus.com” width="1" height="1"></iframe>

    **Malicious Reference:** Malicious reference is a method used to link one page to a malicious page or to download a malicious file when clicked on link. For example:

    <a href=”http://www.hacker.com”>Claim your money</a>

    **Malicious Script:** Malicious script can be written using JavaScript. Companies are hacked via malicious JavaScript code.

    **Abnormal Construction**: This category contains several different malicious activities such as, seldom used tag names, HTML structure inconsistencies, and malicious obfuscation.

    HTML IDS architecture used two methods to perform the data analytics. These are static analysis and dynamic analysis as shown below.

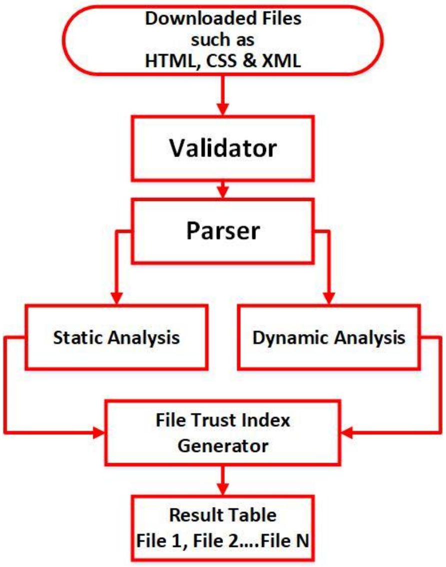


Figure 4 31: HTML IDS Architecture

**IDS for HTML**

**Validator**: The validators job is to verify the data that has been extracted from the parser block is valid and correct and has the correct format.

**Parser**: The parser receives the html file as an input. It is used to verify that the input file is an html file. The parser then parses the html file and extracts the data pertaining to the features from the html file. The extracted data is then passed to the validator block.

**Classifier:** It utilizes several classification models that are obtained as a result of machine learning on the data collected during the training phase. It classifies the output from the validator as either normal or abnormal. Details results shown below.

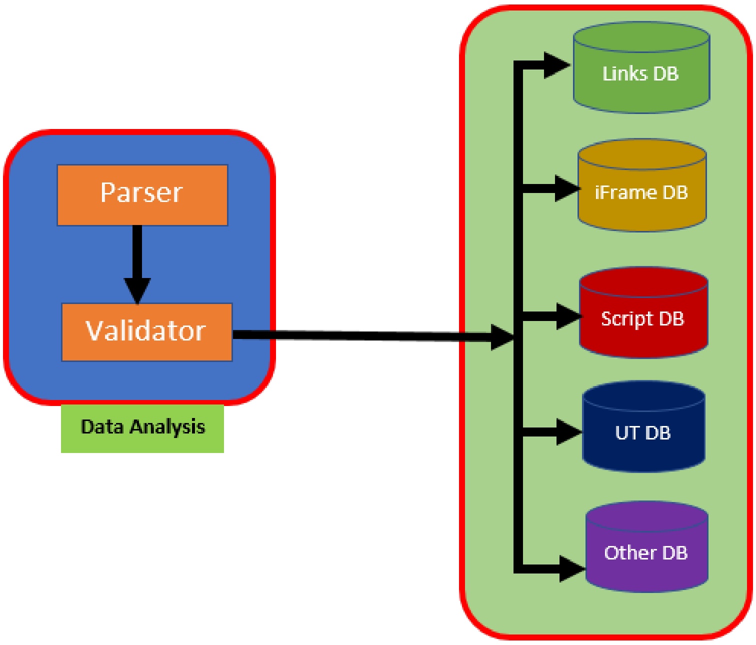


Figure 4 32 : General architecture of HTML file static analysis

Intrusion Detection System operates in two phases.

1. Training Phase:

   In this phase the html file data is used to develop models to classify normal structure of the html files. Feature extraction, data processing and model development are involved in this phase.

2. Operational Phase.

   In this phase the models that have been developed during the training phase and it is used to determine if the html file observed is normal or abnormal. In this presentation more than 10,000 html files applied for experiment as well as evaluation of this approach.

**Static**: Static analysis is the testing and evaluation of an application by examining the code without executing the application.

**Dynamic**: Dynamic analysis is the testing and evaluation of an application during runtime.

**Features Selection Methodology on Static Analysis**

The literature review and an understanding of the HTML files presented us with a large set of features that could be collected from an HTML file as shown in table 4.7.

Table 4 8: Features of Malicious Files

| Feature Name | Associated Category | Feature Type |
|---|---|---|
| Minimum Length of 'a' Tags | Malicious References | Discrete |
| Average Length of 'a' Tags | Malicious References | Discrete |
| Total External 'a' Tags | Malicious References | Discrete |
| Total 'a' Tags | Malicious References | Discrete |
| Total Harmful Links | Malicious References | Discrete |
| Total Redirects | Malicious References | Discrete |
| Length of 'script' tags | Malicious Script | Discrete |
| Length of stings in 'script' tags | Malicious Script | Discrete |
| Maximum Entropy | Malicious Script | Continuous |
| Maximum Length of 'a' Tags | Malicious Script | Discrete |
| Total 'script' Tags | Malicious Script | Discrete |
| Total External 'script' Tags | Malicious Script | Discrete |
| Total Entropy | Malicious Script | Continuous |
| Total Obfuscated HTML Tags | Malicious Script | Discrete |
| Ratio of Key Words to Words | Malicious Script | Continuous |
| Ratio of Whitespace | Malicious Script | Continuous |
| Total 'iframe' Tags | Malicious Script | Discrete |
| Total External 'iframe' Tags | Malicious Script | Discrete |
| Total Hidden iframes | Malicious Script | Discrete |
| Total 'form' Tags | Miscellaneous Features | Discrete |
| Total Event Attachments | Miscellaneous Features | Discrete |
| Total External 'form' Tags | Miscellaneous Features | Discrete |

| | | |
|---|---|---|
| Total Interaction Events Total Tags Unique Tags | Miscellaneous Features | |

Performed feature selection on this feature set to reduce the number of features to be extracted from the HTML file. The feature selection algorithm used as a part of its approach is based on the FEA (Feature Extraction Algorithm) presented by Qu et al. In this FEA, information theory is used to identify the most important and relevant features as shown below.

**Entropy**:

Entropy is a measure of uncertainty of the random variable. It is determined by the equation:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Here, the uncertainty is represented as the log to base 2 of the probability of a category (pi). The number of readily accessible categories is indicated by the index I. The equation is graphically depicted by a symmetric curve as shown below where x-axis is the probability of the event and the y-axis indicates the heterogeneity denoted by H(X).

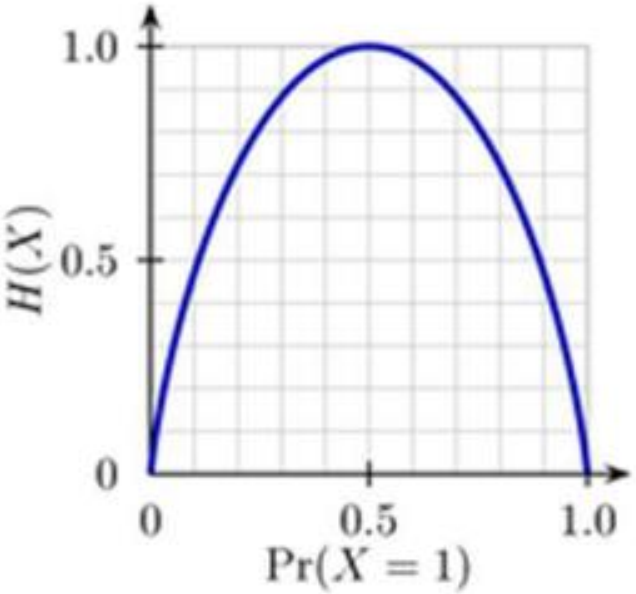


Figure 4 33: Symmetric Curve

**Conditional Entropy:**

Conditional entropy is quantified the amount of information needed to describe the outcome of a random variable Y given that the value of another random variable X. Given X, the conditional entropy of Y is defined as follows:

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

**Mutual Information:**

The amount of knowledge that one random variable knows about another random variable is called mutual information. The definition of mutual information can be described as:

$$I(X;Y) = H(X) - H(X|Y)$$

**DYNAMIC ANALYSIS OF HTML FILES**

Dynamic analysis performed by a HTML file when it is opened through a web browser that allows to perform a detailed functional analysis on the file's action. For this approach I used Sandbox, Origin pro and SPSS environment. It helps to find malicious JavaScript hyperlinks pointing to phishing websites or to perform iframe attacks as well as to generate reports. Fine grained data analysis can detect these types of sophisticated attacks by analyzing the execution traces of the HTML file whenever it is running in a web browser. Below figure shows the architecture of dynamic analysis of HTML which is carried out using sandbox tool.

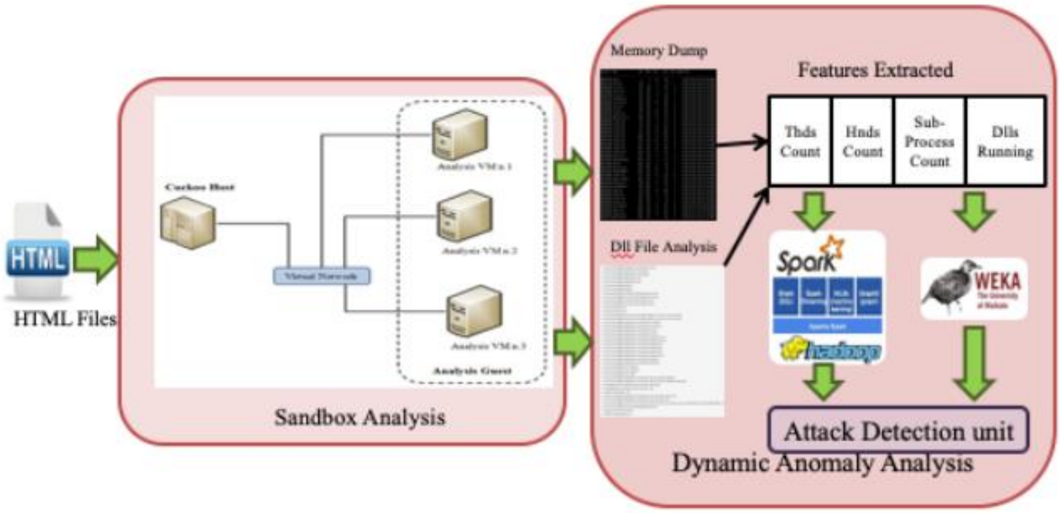


Figure 4 34: Sandbox Analysis

# Chapter 5

## Results

According to Ericsson, there will be 22 billion devices on the IoT by the end of 2022 and these devise are particularly vulnerable to network attacks such as data theft, spoofing, DDOS, phishing attacks and Sonic Wall also generated a report that there was an increase in malware attacks on IoT devices. Based on literature review and initial pre-analysis of normal and malicious files, I identified 23 features as shown on table 4.7. These features Discrete and Continuous. Discrete features are recorded as certain values and a computable value was assigned whereas Continuous features can be any value within certain range of numerical values.

The feature selection algorithm discussed under feature selection methodology is used on the complete dataset in order to reduce the number of features to be analyzed. Table 5.1 shows the mutual information of each of the features and Table 5.2 shows the reduced feature set obtained after running the Feature Selection Algorithm which can be computed as shown below.

$$Q_Y(X_i, X_j) = \frac{I(Y;X_i)+I(Y;X_j)-I(Y;X_i,X_j)}{H(Y)}$$

Table 5 1: Features with mutual information

| Number | Features | Mutual Information |
|---|---|---|
| **0** | url_total_forms | 0.01782 |
| **1** | url_external_forms | 0.04321 |
| **2** | url_total_links | 0.60971 |
| **3** | url_external_links | 0.34345 |
| **4** | url_max_length_links | 0.56431 |
| **5** | url_min_length_links | 0.36827 |
| **6** | url_ave_length_links | 0.26998 |
| **7** | url_unique_tags | 0.18472 |
| **8** | url_total_tags | 0.61241 |
| **9** | url_total_scripts | 0.23493 |

| | | |
|---|---|---|
| **10** | url_external_scripts | 0.04015 |
| **11** | url_obfuscated_html | 0.24171 |
| **12** | url_total_iframes | 0.05121 |
| **13** | url_hidden_iframes | 0.02232 |
| **14** | url_external_iframes | 0.04945 |
| **15** | url_total_objects | 0.00169 |
| **16** | url_keyword_count | 0.32192 |
| **17** | Keywords to Words Ratio | 0.64042 |
| **18** | White Space Ratio | 0.73919 |
| **19** | url_script_length | 0.8183 |
| **20** | url_total_redirects | 0.02179 |
| **21** | url_max_entropy | 0.79023 |
| **22** | url_min_entropy | 0.07363 |
| **23** | url_total_entropy | 0.65151 |

Table 5 2: Reduced Feature Set

| Number | Features | Mutual Information |
|---|---|---|
| **2** | url_total_links | 0.60971 |
| **3** | url_external_links | 0.34345 |
| **4** | url_max_length_links | 0.56431 |
| **7** | url_unique_tags | 0.18472 |
| **8** | url_total_tags | 0.61241 |
| **9** | url_total_scripts | 0.23493 |
| **10** | url_external_scripts | 0.04015 |
| **17** | Keywords to Words Ratio | 0.64042 |
| **18** | White Space Ratio | 0.73919 |
| **23** | url_total_entropy | 0.65151 |

For the signature based IDS, we set the value of MinFreq to 1 and validTime to be a negative number so that all the threats detected in the training period are to be included in the database. We have used a snort rule to create a rule called “Signature. The signature rule file, known as snort.conf, is the "rule" that the snort configuration uses to reference the signatures that were identified with the highest frequency throughout the training period. This program created complementary rule files taking out the most frequency used signatures for the secondary IDS afterward snort was restarted on both systems pointing to the new signature files. We have tested two tests of all IDS. In the first rule 3211 signatures enabled and attacked the network using Sneeze. In the second rule enabling the most frequency signatures which is shown table 5.4 and 5.5 below and then shows the results of the tests of the effects on packet drop rate. Which shows clearly the difference in the performance of the IDS using a small signature database compared to just enabling all signatures. The results of the signatures detected during the training period are shown in Table 5.3.

Table 5 3: Signature Occurrence

| Name of Signature | Number of Occurrence |
|---|---|
| **TTL Detection** | 3 |
| **UDP Port Scan** | 4 |
| **ICMP** | 3 |
| **Teardrop Attack** | 3 |
| **Gopher Proxy** | 4 |
| **TCP Port Scan** | 5 |
| **DDOS** | 14 |
| **WEB-CGI** | 16 |
| **Telnet** | 18 |
| **Double Decoding Attack** | 24 |
| **Backdoor Q Access** | 30 |
| **SNMP Request** | 48 |
| **DOS Arkiea Backup** | 53 |
| **NETBIOS attack** | 107 |
| **FTP overflow attempt** | 180 |

Table 5 4: Performance Measurements (Test 1)

| Number of packets | Scenario 1<br>All rules were enabled | Scenario 2:<br>Most frequency rules were enabled |
|---|---|---|
| **No of packets received** | 243375 | 258673 |
| **No of packets analyzed** | 197623 | 245654 |
| **No of packets dropped by Snort** | 12145 | 1426 |

Table 5 5: Performance Measurements (Test 2)

| Number of packets | Scenario 1<br>All rules were enabled | Scenario 2:<br>Most frequency rules were enabled |
|---|---|---|
| **No of packets received** | 213425 | 258513 |
| **No of packets analyzed** | 197223 | 245254 |
| **No of packets dropped by Snort** | 15145 | 3460 |
| **Percentage of packets dropped by Snort** | 8.496% | 1.807% |

Decision tree for a compound IDS results from the analysis of IDS as shown in figure 2.15. It shows the maximum difference in the value over IDS occurs at the point and no better than no IDS.

For the machine learning IDS, we use a port scanning detection system which is a strategic decision as it enables the attack interruption in its early stages, this is also known as the reconnaissance phase by security frameworks. According to Global Internet traffic by year research collected the following Petabytes in a network backbone between 2004 and 2017.

*Table 5 6: Global Internet traffic by year*

| Year | IP Traffic (Petabytes/month) | Fixed Internet traffic (Petabytes/month) | Mobile Internet traffic (Petabytes/month) |
|---|---|---|---|
| **2004** | 1,477 | 1,267 | n/a |
| **2005** | 2,426 | 2,055 | 0.9 |
| **2006** | 3,992 | 3,339 | 4 |

| | | | |
|---|---|---|---|
| **2007** | 6,430 | 5,219 | 15 |
| **2008** | 10,174 | 8,140 | 33 |
| **2009** | 14,686 | 10,942 | 91 |
| **2010** | 20,151 | 14,955 | 237 |
| **2011** | 30,734 | 23,288 | 597 |
| **2012** | 43,570 | 31,339 | 885 |
| **2013** | 51,168 | 34,952 | 1,480 |
| **2014** | 59,848 | 39,909 | 2,514 |
| **2015** | 72,521 | 49,494 | 3,685 |
| **2016** | 96,054 | 65,942 | 7,201 |
| **2017** | 122,000 | 85,000 | 12,000 |

Table 5 7: According to Wikipedia Predicted global Internet traffic from 2018 to 2022

| Year | Fixed Internet Traffic ExaByte/month | Mobile Internet Traffic ExaByte/month |
|---|---|---|
| 2018 | 107 | 19 |
| 2019 | 137 | 29 |
| 2020 | 174 | 41 |
| 2021 | 219 | 57 |
| 2022 | 273 | 77 |

For the attacks Dataset in a LAN environment, I have installed VMware virtual machine to check port scanning in a LAN environment.

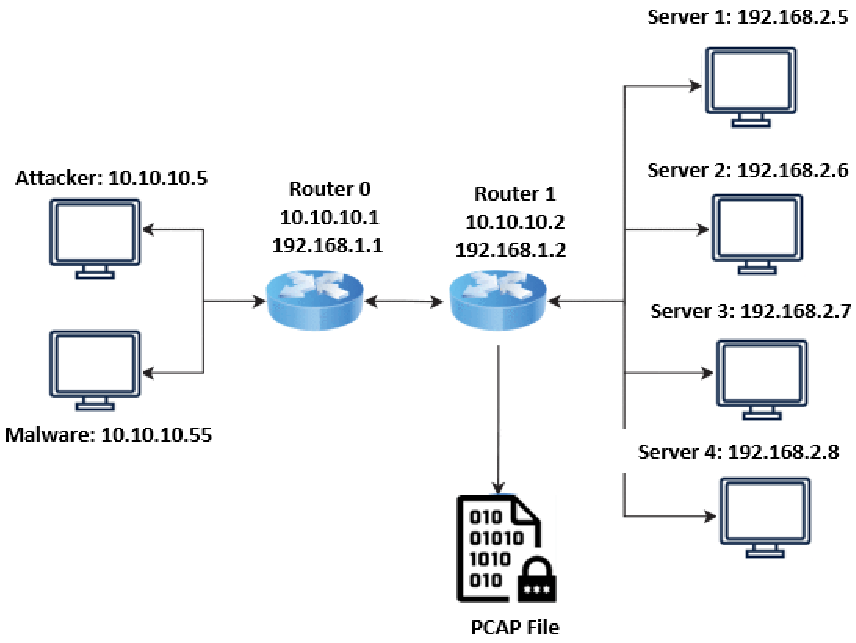


Figure 5 1: Port Scanning in a LAN

For the machine learning training models, we can use Anomaly Behavior Trap which generates dataset denoted by S. First step all attacks captured using port scanning tools for yahoo IP address on 2023 scan as shown in table 5.7.

Table 5 8: Port Scanning result for 10.137.11.164

| Port | Name | Status |
|---|---|---|
| 21 | FTP | CLOSED |
| 22 | SSH | CLOSED |
| 23 | TELNET | CLOSED |
| 25 | SMTP | CLOSED |
| 26 | SMTP | CLOSED |
| 2525 | SMTP | CLOSED |
| 587 | SMTP SSL | CLOSED |
| 43 | WHOIS | CLOSED |
| 53 | DNS | CLOSED |
| 67 | DHCP | CLOSED |
| 68 | DHCP | CLOSED |
| 69 | TFTP | CLOSED |
| 80 | HTTP | OPEN |
| 443 | HTTPS | OPEN |

| | | |
|---|---|---|
| 110 | POP3 | CLOSED |
| 995 | POP3 SSL | CLOSED |
| 143 | IMAP | CLOSED |
| 993 | IMAP SSL | CLOSED |
| 123 | NTP | CLOSED |
| 137 | NetBIOS | CLOSED |
| 138 | NetBIOS | CLOSED |
| 139 | NetBIOS | CLOSED |
| 161 | SNMP | CLOSED |
| 162 | SNMP | CLOSED |
| 389 | LDAP | CLOSED |
| 636 | LDAPS | CLOSED |
| 989 | FTP SSL | CLOSED |
| 990 | FTP SSL | CLOSED |
| 3306 | MySQL | CLOSED |
| 2095 | WEBMAIL | CLOSED |

Table 5 9: Average Feature Score

| **Machine Learning Algorithm** | | **Average Feature Score** |
|---|---|---|
| **DT** | Decision Tree | 0.96 |
| **RF** | Random Forest | 0.96 |
| **LR** | Logistic Regression | 0.70 |
| **NB** | Naïve Bayes | 0.55 |

According to the figure 5.2 shown below the feature score increases at a max value of 0.96. Therefore, this decision results in a max depth of eleven features with Gini criteria and a balance class weight as shown below.

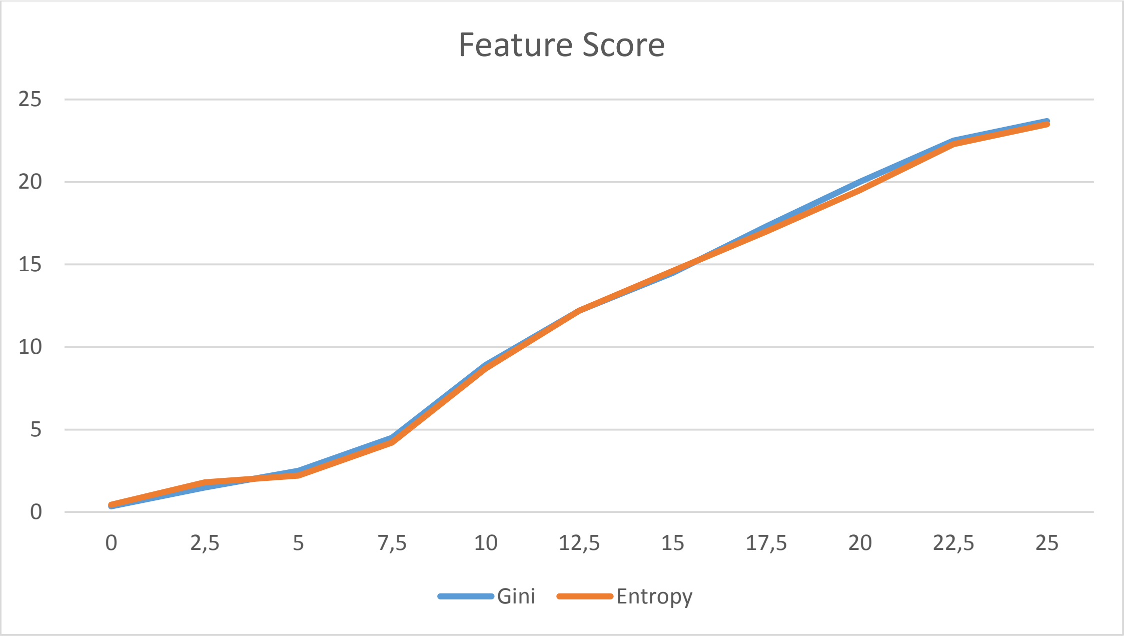


Figure 5 2: Performance of grid search

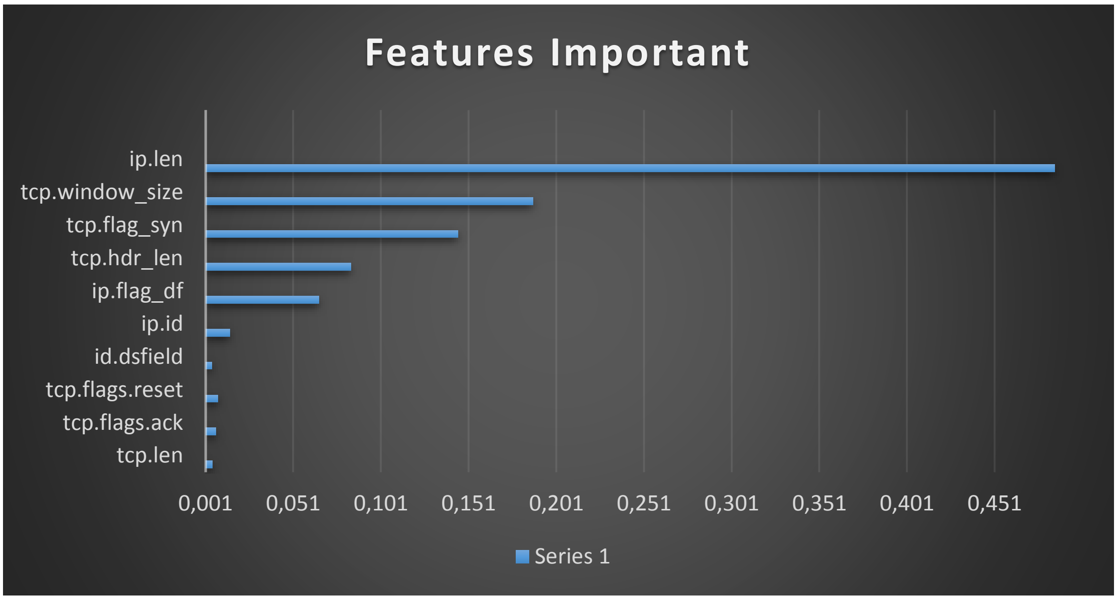


Figure 5 3: Feature importance

The Intrusion Detection System (IDS) assists in locating network anomalies and implements the required corrective actions to guarantee the secure and dependable operation of the Internet of Things (IoT). It also identifies risks and protects the network from intruders and malicious attacks. Intrusion Detection System controls most of those vulnerabilities to protect our Internet of Things. Therefore, it creates positive thinking; positive vive leads to positive life. Hence, the main objective of my study is fulfilled.

# Chapter 6

## Result

Chapter 1 described the background and the related work on IDS and data mining technique. I have discussed taxonomy for Intrusion Detection Systems like signature based IDS, anomaly based IDS and compound IDS work done by Axelsson et al in 2000. After this again, I discussed different characteristics of intrusion detection systems for example the time of detection of the IDS, granularity of IDS and then a brief introduction to machine learning, classification learning, associative learning/clustering and algorithms like isolation forest etc.

Chapter 2 presented Literature studies about Anomaly Based IDS, Self-Learning, Programmed or Supervised Learning, Signature Based IDS, Compound IDS, System Characteristic of IDS, Time of Detection, Source of Audit Data, Active/Passive Response Systems and Machine Learning IDS.

In Chapter 3, presented IDS development methodology. I have started by presenting our architecture for IoT devices which presents a threat modelling methodology designed to model IoT devices and then I apply this methodology to detect threats and design IDS for IoT. I have also discussed data structure like the observation flow and n-grams that we used to model the normal behavior of the protocol.

In Chapter 4, I have presented the behavior of IDS for the Wi-Fi protocol. I have presented different types of IEEE standards and how it works. Aside from that I have discussed different types of wireless network attacks. Aside from that presented IDS for HTML.

In Chapter 5, I described results of IDS for HTML, AB-IDS, Signature based-IDS, Compound IDS and Machine Learning IDS.

Intrusion Detection System (IDS) helps identify anomalies in the network and takes the necessary countermeasures to ensure the safe and reliable operation of IoT (Internet of Things). It also identifies risks and protects the network from intruders and malicious attacks. Intrusion Detection System controls most of those vulnerabilities to protect our IoT (Internet of Things). Therefore, it creates positive thinking; a positive environment leads to positive life. Hence, the main objective of my study is fulfilled.

# 7. References

[1] V. Tirumaladass, S. Axelsson, M. Dougherty and M. A. Rasool, "https://ieeexplore.ieee.org/abstract/document/9182814," 2020. [Online].

[2] Wikipedia, "https://en.wikipedia.org/wiki/Supervised_learning," 2022. [Online].

[3] L. Yang and A. Shami, "https://arxiv.org/abs/2007.15745," 30 July 2020. [Online].

[4] H. Alkahtani and T. . H. Aldhyani, "https://www.semanticscholar.org/author/Hasan-Alkahtani/73109651," 29 December 2020. [Online].

[5] K. Albulayhi, Q. Abu Al-Haija, S. Alsuhibany and A. Jillepalli, "https://www.mdpi.com/2076-3417/12/10/5015," 2022. [Online].

[6] Q. A. Al-Haija, "https://www.researchgate.net/publication/371913857_Cost-effective_detection_system_of_cross-site_scripting_attacks_using_hybrid_learning_approach," 2022. [Online].

[7] Q. A. Al-Haija and A. Al-Badawi, "https://www.mdpi.com/1424-8220/22/1/241," [Online].

[8] S. Axelsson, "https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7a15948bdcb530e2c1deedd8d22dd9b54788a634," 14 March 2000. [Online].

[9] P. V. V. Ioulianou, "https://eprints.whiterose.ac.uk/133312/1/ictf_2018_IoT.pdf," 11 July 2018. [Online].

[10] M. Tavallaee, "https://www.researchgate.net/publication/224138101_Toward_Credible_Evaluation_of_Anomaly-Based_Intrusion-Detection_Methods," Oct 2010. [Online].

[11] G. Canfora, F. Mercaldo and Corrado, "https://dl.acm.org/doi/abs/10.1145/2804345.2804349," 2015. [Online].

[12] Likarish, Peter, E. Jung and I. Jo, "https://www.semanticscholar.org/paper/Obfuscated-malicious-javascript-detection-using-Likarish-Jung/3f37b496268cc5eb75af2af78b26f2d17fafe0c1," 1 October 2009. [Online].

[13] S. Technology, "6 core IoT," 02 April 2022. [Online]. Available: https://www.seeedstudio.com/.

[14] V. Capatalist, "Internet Minute," 22 March 2021. [Online]. Available: https://www.visualcapitalist.com/.

[15] T. Points, 15 January 2023. [Online].

[16] D. E. DENNING, 11 Feb 2022. [Online].

[17] T. P. S. University, 11 March 2022. [Online].

[18] T. P. S. University, 06 April 2022. [Online].

[19] Nasscom, 11 April 2022. [Online].

[20] A. A. Ansam Khraisat, 20 March 2022. [Online].

[21] IEEE, 27 Feb 2022. [Online].

[22] J. McHugh, 7 March 2022. [Online].

[23] Medium, 5 April 2022. [Online].

[24] K. Nguyen, 7 December 2022. [Online].

[25] NetworkLessons, 01 January 2022. [Online].

[26] P. Satam, "https://repository.arizona.edu/handle/10150/632978," 7 March 2022. [Online].

[27] R. A. David Mudzingwa, 8 March 2022. [Online].

[28] NIST, 26 Feb 2022. [Online].

[29] Geeksforgeeks, 26 Feb 2022. [Online].

[30] Á. P. d. l. C. J. M. M. M. Ricardo Salazar-Cabrera, 16 April 2022. [Online].

[31] S. P. P. Rashmi Ravindra Chaudhari, 14 Feb 2022. [Online].

[32] Mongodb, "https://www.mongodb.com/cloud-explained/iot-architecture," 9 April 2022. [Online].

[33] P. Analytics, 17 April 2022. [Online].

[34] M. A. W. Shalaby, 7 Dec 2022. [Online].

[35] R. A. David Mudzingwa, 8 March 2022. [Online].

[36] R. K. A. B. J. K. Mohit Tiwari, 7 March 2022. [Online].

[37] D. A. B. H. W. Smys Smys, 15 March 2022. [Online].

[38] C. Yonghui, 26 Feb 2022. [Online].

[39] Researchgate, 05 April 2022. [Online].

[40] S. University, 17 March 2022. [Online].

[41] I. Analytics, 01 April 2022. [Online].

[42] T. Target, 23 March 2022. [Online].

[43] E. d. Argaez, 23 March 2022. [Online].

[44] Venerbilt, 15 April 2022. [Online].

[45] J. Interactive, 01 April 2022. [Online].

[46] BMC, 17 April 2022. [Online].

[47] nayarasi, 11 January 2023. [Online].

[48] P. Satam, "An Anomaly Behavior Analysis Intrusion Detection System for IoT," 22 November 2022. [Online].

[49] L. A. I. a. G. M. Atzori, "A Survey of Computer networks 54, no. 15 (2010): 2787-2805," 17 August 2022. [Online].

[50] A. A. A. a. L. D. X. Whitmore, "The Internet of Things, A survey of topics and trends.Information Systems Frontiers 17, no. 2 (2015): 261-274," 16 November 2022. [Online].

[51] Internetworldstats, "https://www.internetworldstats.com/stats.htm," 27 March 2022. [Online].

[52] G. V. a. R. Kemmerer, "A Network-Based Intrusion Detection Approach," 2 December 2022. [Online].

[53] G. T. O. a. K.-R. M. Rätsch, "Machine learning 42, no. 3 (2001): 287-320," 04 January 2023. [Online].

[54] S. L. W. a. T. V. Raza, "Real-time intrusion detection in the Internet of Things." Ad hoc networks 11, no. 8 (2013): 2661-2674," 07 January 2023. [Online].

[55] P. N. W. a. C. C. Sangkatsanee, "Computer Communications 34, no. 18 (2011): 2227-2235," 19 December 2022. [Online].

[56] I. S. 802.11-1997, "Wireless Lan Medium Access Control (MAC) And Physical Layer (PHY) Specifications. 1997," 11 January 2023. [Online].

[57] IEEE, "802.11a-1999 High-speed Physical Layer in the 5 GHz band," 18 December 2022. [Online].

[58] IEEE, "802.11b-1999 Higher Speed Physical Layer Extension in the 2.4 GHz band," 24 December 2022. [Online].

[59] IEEE, "IEEE 802.11g-2003: Further Higher Data Rate Extension in the 2.4 GHz Band," 22 December 2022. [Online].

[60] D. Madory, "New methods of spoof detection in 802.11 wireless networking," 9 January 2023. [Online].

[61] a. T. C. F. Guo, "Sequence number-based MAC address spoof detection," 07 January 2023. [Online].

[62] J. S. A. C. R. Gill, "Specification-Based Intrusion Detection in WLANs Computer Security Applications Conference, 2006," 29 January 2023. [Online].

# 8. Appendix

Koselee Publication, writer Rupendra Man Rajkarnikar Grade 6 to Grade 12, written about AI, Robotics, IoT, Networking, JavaScript, PHP, MySQL, Cybercrime etc.

CCTV and AI Camera seminar conducted at St. Xavier’s School, Tribhuvan University, Virinchi College.

Global Positioning System (GPS) demonstrated at St. Xavier’s School and Virinchi College in 2019.

Product demonstrated of Blue-Tooth Bulb, IoT door, light and curtain at Group Four in 2019.

VMWare Seminar conducted at e-Asia University and Ratna Laxmi Campus affiliation of Tribhuvan University.

Cyber Security and Cyber-Crime Seminar conducted at different schools and colleges (Kitini College, SXJ, Campion College, Saint Mary School, Government School and College) in 2018.

Using Arduino and mBlock, a seven-day robotics and IoT workshop was held at St. Xavier's School in Jawalakhel and Godawari. Students were able to create smart water pumps, smart cars, smart mirrors, smart dustbins, and smart irrigation water supplies.