

# Burglar detection on SCION using motion sensor by data analytics and machine learning

1

Abhijith Remesh

Email: [abhijith.remesh@st.ovgu.de](mailto:abhijith.remesh@st.ovgu.de)

Manish Ramachandran

Email: [manish.rama@st.ovgu.de](mailto:manish.rama@st.ovgu.de)

**Abstract**— There are several Internet of things (IoT) applications running on legacy networks which are not flexible in terms of path selections, network outages and security concerns. Thus, the current situation demands switching over to a new network platform which is highly scalable, secure, and isolated in terms of its architecture. Thus, a basic IoT application is being attempted which will use Scalable Control and Isolation on Next generation networks (SCION) architecture and hence, surpasses the existing concerns on the current network protocols. The chosen Internet of things (IoT) application is based on Automatic Burglar Detection, where an Intruder is identified without indulging camera setups and rather employing optimal motion sensor, mac ids and virtual switches. The implementation involves a simple client server architecture where there are two clients connected to a single server (Raspberry Pi) running a SCION Autonomous System (AS) instance and Message Queuing Telemetry Transport (MQTT) broker. The Espressif (ESP) module with Passive Infra-Red (PIR) sensor act as one of the clients publishing values to the server (Raspberry Pi) via MQTT protocol. The SCION AS enabled local machine being the other client and the user interface to the user connects to the server (Raspberry Pi) via MQTT over SCION SIG configuration. The key process includes (i) The Raspberry Pi is configured as sever with MQTT broker and a SCIONAS instance.(ii) the client ESP module integrated with PIR sensors publishes motion detection over MQTTprotocol to the SCION AS enabled Raspberry Pi server. (iii) the other client side, SCION AS enabled machine subscribes and receives the data from the server (Raspberry Pi) via MQTT protocol running over the SCION network.

## I. INTRODUCTION

We propose to develop an intruder detection application for the Smart Home use case. The main objective of our project is to detect the presence of an intruder at home without the application of cameras as installing cameras within the home is a security concern for the personnel residing at home. In order to detect the presence of a person at home, an idea was proposed to monitor three parameters (i) motion, (ii) a switch functionality made available to the user and (iii) another parameter which conveys if the user's mobile is connected to the home network. The passive infrared motion sensor has been employed to detect the motion readings at home which

will be installed at an appropriate space at home which would produce high output when a motion is detected and produces low output otherwise. An online toggle switch functionality has been developed so that the user can manually operate it whenever he arrives or leaves home. A Medium Access Control (MAC) address look-up in the Address Resolution Protocol (ARP) table has been used which would look up the MAC address of the user's mobile in the ARP table so that it would produce a high output whenever the mobile is connected to the home network and produces low output otherwise. These three parameters namely motion data which will be either a 0 or 1, switch data which will be either a 0 or 1 and mobile mac data which also will be either a 0 or 1 along with the respective timestamps are considered as the source data for the application which processes this data based on various approaches to determine the times when the person will be likely to be present/absent at home and further determines the presence of an intruder. Various analytical and machine learning approaches have been used in our application to predict the intruder presence namely 'historical data analysis', 'Long short term memory model', 'Classic Statistical methods'. The prediction results produced by these approaches are taken into account and are compared and evaluated against each other considering the merits and demerits of each of the approach. The best performing approaches are eventually specified in the conclusion.

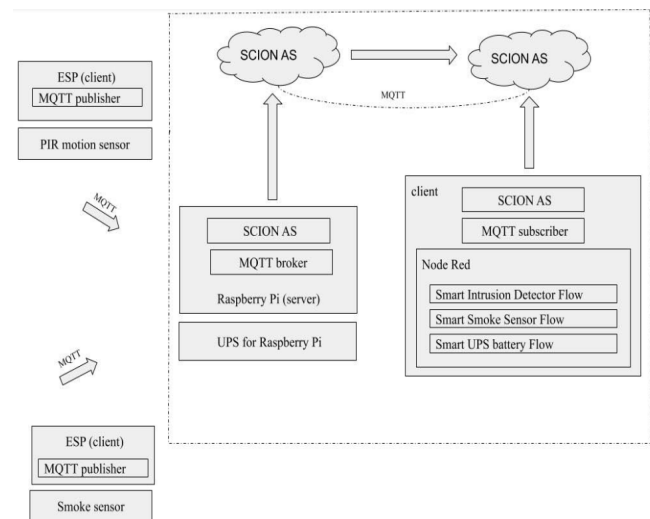


Fig1. Application schema

## II. COMMUNICATION FLOW

THIS chapter discusses the data and communication pathways in our application.

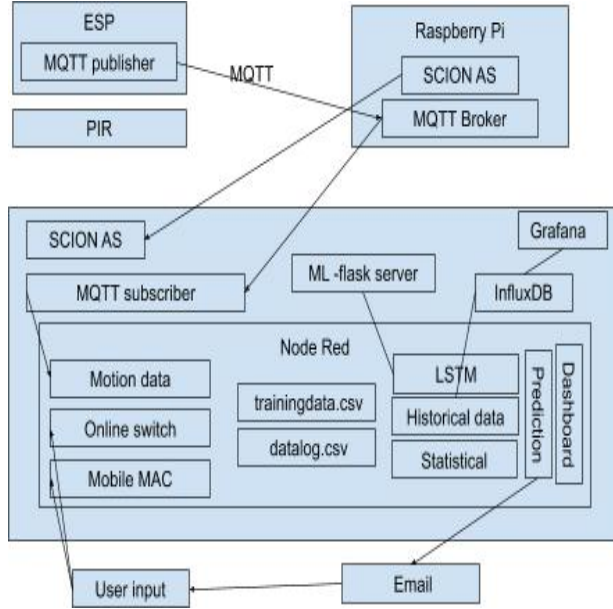


Fig2. Communication flow

- The Passive Infra-Red, PIR[1] motion sensor is coupled with ESP8266[2] module so that motion data from the PIR[1] sensor is sent over MQTT[3] protocol (WiFi) through an ESP[2] module as a MQTT[3] message "Motion/Voltage" to the MQTT[3] broker installed at the Raspberry Pi[4] end with topic "/feeds/motion". The ESP[2] module is configured to act as a MQTT[3] publisher which publishes the message "Motion/Voltage" to the MQTT[3] broker installed at the Raspberry Pi[4] end whenever a motion is being detected by the PIR[1] sensor.
- A SCION[5] AS is configured to run on the pi which provides the SCION[5] Internet Gateway (SIG) through specific defined ip addresses. The MQTT[3] broker installed at the Raspberry Pi[4] end will be always listening on the port 1883 of the pi's network interface. The ESP[2] module publishes the MQTT[3] messages to the Raspberry Pi[4] on a particular topic '/feeds/motion'. Upon receipt of this MQTT[3] message at the defined ip address and port 1883 of the Raspberry Pi[4], the MQTT[3] broker at the pi end will forward this message to the respective subscriber (local remote machine running another defined SCION[5] IP address) who is subscribed on this topic via MQTT[3] over SCION[5] network.
- This MQTT[3] message from the PIR[1] motion sensor acts as the trigger point and captures the

instantaneous mobile mac data and online switch data along with the timestamp.

- This data is logged in two different csv files namely trainingdata.csv and datalog.csv. The datalog.csv acts as the source file for historical data analysis and is uploaded weekly to the InfluxDB[6] for the same in a specific format whereas the trainingdata.csv acts as the source file for training the Long short-term memory, LSTM[7] model and the model is trained on a monthly basis. Hence, the datalog.csv is refreshed on a weekly basis and the trainingdata.csv is refreshed on a monthly basis. The training data.csv is also used for the statistical approaches and providing the weekly summary detection report.
- The motion data, mobile mac data and the online switch data is visualized using Grafana[8] and the predictions results are visualized using the Node-RED[9] dashboard. Whenever the prediction results suggests an abnormal intrusion behavior, an email notification is sent out to the user.

## III. HARDWARE DEVICES IN CONCERN

### A. Device Specifications

The PIR[1] motion sensor has been configured with different wired setups over the time and some of these setups were sidelined for their inefficient functionality and portability.

The PIR[1] sensor opted for the application is HC-SR501[10] whose specifications are as below

- View area: 110 degree like cone
- Operating voltage: 5-20 V
- Power consumption: 65mA
- PIN 1: Ground
- PIN 2 Output: 3.5V when the motion is detected.
- PIN 3: Vcc-5V

The PIR[1] motion sensor is operated in single trigger mode.

The time delay begins immediately when the motion is detected. No Continued detection.

- Jumper setting on PIR[1] motion sensor must be low
- Output becomes high when motion detected.
- Output becomes low after a certain time delay even if the object is in motion.
- After the output being low for a certain time (3s), it again becomes high.

The Time delay range for the sensor ranges from 5 seconds to 5 minutes.

- Clockwise towards the extreme right position gives a delay of 5 minutes.
- Counter-clockwise towards the extreme left position gives a delay of 5 seconds.

If the motion is detected, and if the time delay is set as 5 seconds, then the PIR[1] motion output becomes high for 5 seconds and then it becomes low for 3 seconds.

The Sensitivity ranges from 3m to 7m.

- Clockwise towards the extreme right position gives a range of 3 meter.
- Counter-clockwise towards the extreme left position gives a range of 7 meter.

Hence, the PIR[1] motion sensor is physically configured to operate at a delay of 5 seconds and at a sensitivity of 7m.

The application uses a Raspberry Pi[4] as the micro-controller unit which acts as the server with a SCION[5] AS instance and a MQTT[3] broker is running.

### B. Hardware Interfacing methodologies

There were a couple of circuits designed over time to interface the PIR[1] motion sensor and Raspberry Pi[4] in an optimal way. Some of them were dropped eventually considering the efficiency and portability factors. The methodologies that were developed and enhanced over time are described below.

**Method 1:** The conventional wired setup wherein the motion detection values and person's presence are obtained from the serially interfaced, PIR[1] sensor and hardware switch with Raspberry Pi[4].

The PIR[1] motion sensor and a hardware switch have been interfaced with Raspberry Pi[4] serially. The basic functionality of the hardware switch is to indicate the presence of person so that the user is expected to turn on the switch (Switch value 1) at his presence and turn off the switch (Switch value 0) at his absence. The script at the server side "Client\_PIR.go" fetches the motion values, the corresponding switch values and amalgamate timestamps along with it which forms the data. This data is then delivered to the client-side script at the remote system in subsequent periods over the SCION[5] network.

Because of these concerns, a new idea has been proposed to replace the existing wired setup with the ESP[2] module so that the PIR[1] motion sensor can be connected to this ESP[2] module wirelessly, thus establishing a wireless communication over Wi-Fi via MQTT[3] protocol.

Advantages	Disadvantages
Easy set up	No portability
Easy defect probing	High latency from Client local machine side because of serial communication.
Low latency from server side	Limitation of PIR[1] sensor placement due to lack of portability.
High duration of ON time.	
Easier to deploy	

Table 1. Pros and Cons of method 1

**Method 2:** ESP8266[2] [(Wemos D1 Mini[11]) and PIR[1] motion sensor

This setup employs the state of the art IoT protocol, Message Queuing Telemetry Transport (MQTT[3]). It is a wireless protocol that sends message to any other network in a device by configuring the devices as broker and client.

The ESP8266[2] is a Wi-Fi micro controller which enables it to connect to a Wi-Fi network. However, Wemos D1 Mini[11] is employed for the use case which is a mini board with 4mb flash based on ESP8266EX. Since a TTL CH340G converter module adapter is required to burn programs in ESP-01 module, the Wemos D1 Mini[11] has been used which can be connected directly to any machine to burn the programs. As the Wemos D1 Mini[11] is powered by an external battery, there exist a privileges for Battery monitoring by supplying up to 3.3 V across its analog pin which will be discussed in later part.

MQTT Brokers, Publishers and Subscribers

MQTT[3] comprises of publishers and subscribers, connected to the same broker. The broker can be acts as the server and the publishers/subscribers acts as the clients.

MQTT Topics and Messages

All clients attempting to connect via MQTT[3] protocol needs to know the address of the broker. Any client connected to a broker can publish a message on a specific topic and any client connected to the same broker and same topic can subscribe and receive the respective message.

Publisher(client) publishes a message on a topic. Topic is an intermittent channel via which any subscriber enrolled to the same broker and topic can receive the message.

Hosting an MQTT broker

A MQTT[3] broker can be hosted in a computer, Raspberry Pi[4] or any virtual machine. Similarly any program (Python or go or Arduino[12] etc) can enable MQTT[3](publish/subscribe) with the latest MQTT[3]

libraries like Paho MQTT for Python and PubSub for Arduino[12].

#### Mosquitto Installation and Service Activation

- Sudo apt install mosquitto mosquitto-clients
- Sudo systemctl enable mosquitto

Arduino[12] IDE has been used to burn the code into Wemos D1 Mini. The PubSubClient library has been used in the code for the ESP[2] to publish motion value to the Raspberry Pi[4] where the MQTT[3] broker is installed.

Topic: /feeds/motion

Value: Motion/(Person\_Value) for eg. Motion/5.2

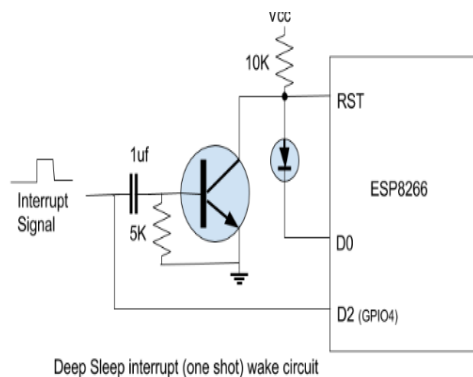
As described earlier, the Hardware switch has been used to indicate the presence of person at home. If the person is present at home, it would indicate 1 (person switches on) and otherwise 0.

Advantages	Disadvantages
Wireless Set up	Heavy battery drain
flexible to mount at remote location	
Away from the Raspberry Pi[4].	

Table 2. Pros and Cons of method 2

Because of the concern of high battery drain, a new idea has been proposed to design a PCB which would put the ESP[2] in deep sleep when not in use.

#### Method 3: ESP8266 with Sleep PCB Circuit and Physical Switch



Deep Sleep interrupt (one shot) wake circuit

Fig3. ESP Deep sleep circuit

The above circuit has been designed to put the ESP8266[2] in deep sleep mode when not in use and its corresponding PCB circuit has been developed. With this PCB circuit we were able to cut down the power usage of ESP[2] hence saving the battery life. The module wakes with an interrupt which is nothing but the detected motion which would further publish a MQTT[3] message and goes back to sleep.

#### Battery Usage Calculation

During the normal mode, the PIR[1] motion sensor and the ESP[2] consumes 3.3V and 80mA of current.

In the deep sleep mode, the estimated / calculated current consumption of ESP[2] is around 8mA (substantial drop)

We assume 80 motion in an hour which varies from person to person.

- Motion in an hour = 80
- Current consumption by ESP[2] during a motion = 80mA
- Time taken by ESP[2] to wake up, publish the message and sleep = 2.5sec (on an avg)
- In an hour =  $80 * 2.5 = 200$ seconds
- Remaining 3400 sec in hour is in deep sleep mode.
- $200 \text{ sec} * 80\text{mA} + 3400\text{sec} * 8\text{mA} = 12\text{mAh}$  in an hour
- Source = Battery capacity of our setup = 2900 mAh (4 AA batteries in series 6.5 V)
- Output efficiency of battery =  $0.80 * 2900 = 2320\text{mAh}$
- $2310 \text{ mAh} / 12\text{mAh} = 193 \text{ hours} = 8\text{days}$

The battery should approx. last 8 days if there is 80 motion in an average in an hour. Hence, it is required to find much optimal solutions.

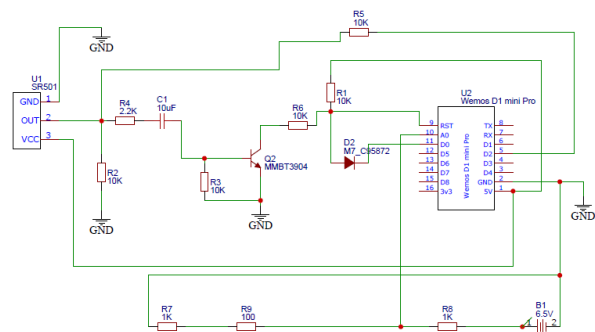


Fig4. PCB fabrication

Hence, the battery voltage needs to be stepped down from 6.5v to 3.3 v to feed it across the analog pin of ESP[2]. This has been successfully implemented with the help of voltage divider (1 and 1.1K) where voltage across 1.1 K (total 2.1k) is fed across the ESP[2] analog pin (3.3v). This is scaled back in the Arduino[12] code to detect the real voltage and sense the health of battery.

Because the MQTT[3] message from ESP[2] is "Motion/(battery voltage)", the user in the client side get continuous notification about the battery health via Node-RED[9] user interface. Alerts is also set when the battery voltage steps down below 4.2V.

The last enhancement in this setup has been the implementation of Wi-Fi manager library in ESP[2] (Wemos D1 Mini[11]).

The WiFi network is never constant and it is required for the ESP[2] to connect to different Wi-Fi networks/broker/topic dynamically as per our needs. Technically, it's a repetitive task as the program needs to be burned to ESP[2] using Arduino[12] every time, to connect to a new Wi-Fi network. This concept has been made super easy with the Wi-Fi manager library. Every time the code is burned into ESP[2] (without Wi-Fi cred) it creates a mount point (server) which can be accessed for setting all the parameters (Wi-Fi/topic/broker). Once the ESP[2] is reset, next time it connects to the parameters that were set in its configuration file earlier.

Whenever there is a need to edit any of these parameters the ESP[2] is required to create a mount point so that the parameters could be reconfigured. A small two-way physical switch has been implemented to realize this functionality. When the switch is turned on (set high) it resets all the parameters in the ESP[2] configuration file and hence creates a mount point. Once all our parameters are set, the switch needs to be turned off (set low) so as to save the parameters. This process can be repeated to wipe out all parameters and set again. Also, as an enhancement, the hardware switch (indicating person presence) has been replaced with an online switch so that the user could have an access to it via the webpage rather than expecting the user to enable and disable the switch physically. This is achieved by the "Online switch node" configured in Node-RED[9]. The switch can be turned on/off via Node-RED[9] user interface.

Apart from switch functionality, Address Resolution protocol (ARP) has been incorporated in the application which is realized using the ARP node in Node-RED[9]. The MAC id of the user's mobile is specified in the ARP node. Whenever, the user's mobile is connected to Wi-Fi at his presence, the ARP outputs a value of "1" indicating that the user is at home.

#### IV. DATA ANALYSIS APPROACHES

This chapter discusses about the establishment of client server communication over the SCION[5] network. The sequence of operation is as follows:

##### A. Historical data analysis

###### 1) Logging Data

This approach involves the periodical querying of a InfluxDB[6] database containing the historical data (timestamp data, week index data, time index data, motion

data, online switch data, mobile mac data). This historical data in the InfluxDB[6] database is updated on a weekly basis. In every five minutes, a query is passed to the database to return the number of detections for that five minute interval from the historical data which conveys the usual number of detections that normally happen on that day, for that five minute interval. This query result is then compared with the current number of detections observed in the current five minute interval. The detailed description of this approach is as follows:

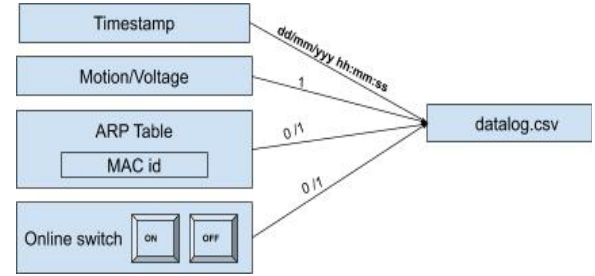


Fig5. Data components logged

When the MQTT[3] message "Motion/Voltage" is received at the server end which as the MQTT[3] subscriber, the corresponding online switch data and the mobile mac data at that instant of time is extracted. The MQTT[3] message is fetched using the "MQTT[3] in" node which is configured by providing the ip address of the Raspberry Pi[4] "\_\_\_\_\_" and the topic of the MQTT[3] message "/feeds/motion". The MQTT[3] message "Motion/Voltage" is then further mapped as 1. The online switch value is extracted using the "switch" node which is a toggle button representing "0" at OFF state and "1" at ON state. The mobile MAC data is derived from the "ARP" node where the MAC id of the user's mobile is specified so that this MAC id is checked in the ARP table of the home's network. Presence of the MAC id corresponds to "1" and absence of the same corresponds to "0". As a result, the motion data will be always "1", the online switch value will be either "0" or "1" and the mobile mac value will be either "0" or "1". All these data are then combined together along with the corresponding timestamp and is logged in a csv file, datalog.csv. An entry in the log file only happens upon a motion detection and on further receipt of the MQTT[3] message "Motion/Voltage" at the server end. The format of the datalog.csv is as shown below :

```

28/08/2019 20:17:23,1.0,1.0,1.0
28/08/2019 20:23:44,1.0,1.0,1.0
28/08/2019 20:25:31,1.0,1.0,1.0
28/08/2019 20:31:24,1.0,1.0,1.0
28/08/2019 20:42:15,1.0,1.0,1.0
28/08/2019 20:42:42,1.0,1.0,1.0
28/08/2019 20:56:21,1.0,1.0,1.0
28/08/2019 21:07:07,1.0,1.0,1.0
28/08/2019 21:11:02,1.0,1.0,1.0
28/08/2019 21:15:54,1.0,1.0,1.0
  
```

Fig6. Datalog.csv file format



## 2) Uploading logged data to InfluxDB

This datalog.csv file acts as the basis file for the historical data analysis. This datalog.csv file is then uploaded to the InfluxDB[6] database “historicaldb” on a weekly basis and hence, the

datalog.csv file is also refreshed on a weekly basis to avoid duplicate data being uploaded to the “historicaldb”. A retention policy of six months is applied on the database so that the historical data older than six months are automatically removed from the database as the time progresses and eventually, the database will only contain the recent historical data.

The datalog.csv file is uploaded to “historicaldb” in a particular format by passing the data over a series of Python scripts namely “timestamp\_formatting.py”, “timestamp\_influxconversion\_format.py” and “timestamp\_influx\_lineprotocol\_format.py”.

The input data to “timestamp\_formatting.py” is of the format:

Timestamp	motion	Online switch	Mobile mac
dd/mm/yyyy hh:mm:ss	1	0 or 1	0 or 1

Table 3. input data format

Which is then transformed into the format as below:

Timestamp	WDay	Time index	moti-on	Online switch	Mobile mac
dd/mm/yyyy hh:mm:ss	1-7	0-287	1	0 or 1	0 or 1

Table 4. input data format after first transformation

The “timestamp\_formatting.py” script aggregates the datalog.csv on a five-minute interval basis such that the timestamps are in five-minute intervals in the sequence dd/mm/yyyy hh: 00:00, dd/mm/yyyy hh:05:00, dd/mm/yyyy hh:10:00 and so on.

The script also derives the WDay and the time index from the date and time information of the timestamp respectively. The weekday, WDay is derived from the date information of the timestamp ie 1 for Sunday, 2 for Monday, 3 for Tuesday, 4 for Wednesday, 5 for Thursday, 6 for Friday and 7 for Saturday. Hence, the WDay value will be any value between 1 and 7. The time index value is also derived from the time information of the timestamp. The time index value is the incremental value for every five minutes starting from 0 to 287 such that the first five minutes of a day (00:00:00 - 00:05:00)

corresponds to 0 and the last five minutes of a day (23:55:00 - 00:00:00) corresponds to 287. Thus, the time index will have values ranging from 1 to 287 as a day contains 287 five-minute intervals.

As a result, an entry in the output file conveys the following information: five minute interval timestamp, the WDay which tells the day when the motion was detected, time index which tells five minute interval on that day, motion which tells the number of times, motion has been detected for that five minute interval on that day, the online switch and the mobile mac value for that five minute interval on that day.

The “timestamp\_influxconversion\_format.py” script converts the timestamp into the corresponding UNIX timestamp format.

Timestamp (UNIX)	WDay	Time index	moti-on	Online switch	Mobile Mac
dd/mm/yyyy hh:mm:ss	1-7	0-287	*	0 or 1	0 or 1

Table 5. input data format after second transformation

The “timestamp\_influx\_lineprotocol\_format.py” also transforms the existing format to a particular line protocol format such that the data suits to be fed into a measurement “sensor” within the “historicaldb”. The output file from this script is then fed to the InfluxDB[6] “historicaldb” using the InfluxDB[6] API POST request. The data gets updated in the “historicaldb” as follows:

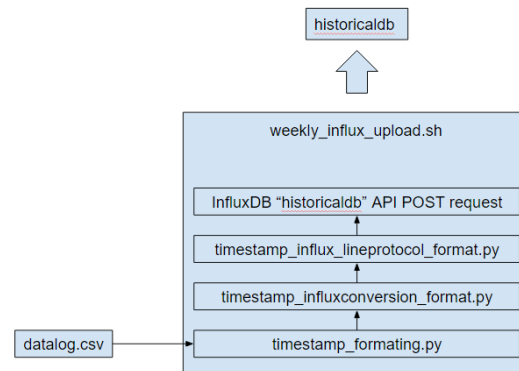


Fig7. InfluxDB upload scripts

The weekly\_influx\_upload.sh is run weekly using the crontab services which is scheduled to run on every Sunday at 23:20:00 so that the historical data update happens automatically every week. and the datalog.csv is also refreshed weekly so that every week’s datalog.csv gets uploaded to the “historicaldb” to avoid duplicate upload.

20 23 \* \* 0 /bin/sh /home/pi/weekly\_influx\_upload.sh

### 3) Historical Search logic

In every five minutes of a day, a SELECT query is passed to the “historicaldb” to return the number of detections happened for that five minute on that day. This is done using the “primaryhistoricalsearch.py” script which is embedded within the “Python Shell” node in the Node-RED[9].

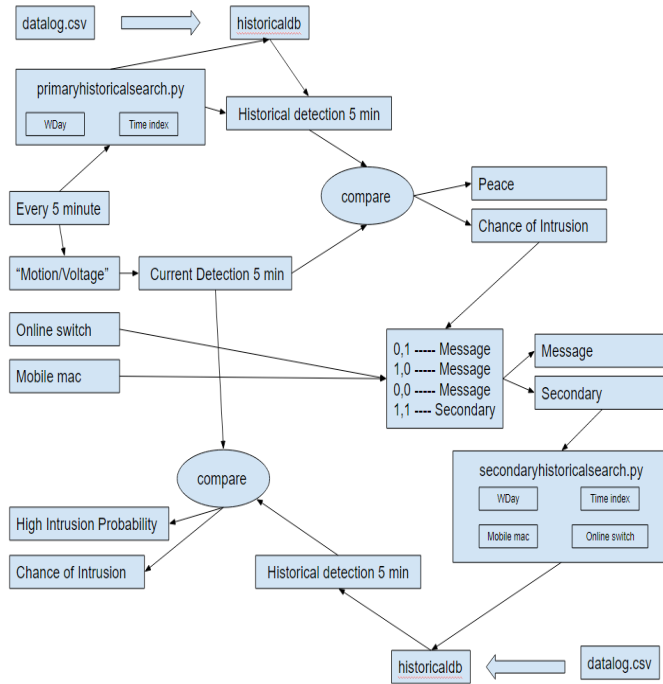


Fig8. Historical data search logic

In every five minute of a day, the instantaneous timestamp is passed as arguments to the “primaryhistoricalsearch.py” script. This script takes this timestamp and derives the respective current WDay and current time index from the timestamp information. So, the week index will be any value between 1 to 7 and the time index will be any value between 0 to 287.

The “historicaldb” already contains the historical data in the format (timestamp, WDay, time index, detections, online switch, mobile mac). The script then passes a couple of select queries on this historical data :

- "Select the “sum” of detections from “historicaldb” where time index = current time index and WDay = current WDay "
  - This will fetch out the total sum of detections that happen on this particular week day and time index. This value is nothing but the sum of detections that happen in a five minute interval on specific week day.

- "Select the “count” of WDay from “historicaldb” where time index = current time index and WDay = current WDay"
  - This will fetch out the count of specific WDays existing in the “historicaldb”

The division of the two results corresponds to the usual number of detections that happen on a five minute interval of a specific day based on the historical data. This corresponds to the historical five minute detections value.

In every five minute, the number of times, the MQTT[3] message “Motion/Voltage” being received at the “MQTT[3] in” node is counted. This value corresponds to the current five minute detections value.

This historical five minute detections value and current five minute detections value are compared with each other on a function node. If the current five minute detections value exceeds the historical five minute detections value by a nominal percentage, then the situation can be considered as abnormal situation and the situation is classified as a "Chance of Intrusion" and otherwise as “Peace”.

Condition	Situation
Current five min detection > historical five min detection	Chance of Intrusion
Current five min detection < historical five min detection	Peace

Table 6. primary historical search decisions

If an abnormal situation or "Chance of Intrusion" case arises, the current online switch value and current mobile mac value are taken into consideration and different cases are formulated based on their value as follows:

Situation	Online Switch	Mobile mac	Action
Chance of Intrusion	0	0	message
	1	0	message
	0	1	message
	1	1	Secondaryhistoricalsearch.py
Peace	nil	nil	nil

Table 7. primary historical search actions

The message corresponds to “Detections are observed or check your dashboard to see if the switch value and mobile mac value is correctly configured or not“. This message is displayed in the NodeRed dashboard.

This corresponds to the historical primary search.

If both online switch and mobile mac are in high state, a SELECT query is passed to the script “secondaryhistoricalsearch.py” embedded in the Python shell of Node-RED[9] along with the current timestamp value as arguments. The script takes this timestamp and derives the respective current WDay and current time index from the timestamp information. The script passes a couple of queries to the “historicaldb”

- "Select the “sum” of detections from “historicaldb” where time index = current time index and WDay = current WDay " and mobile mac = current mobile mac and online switch = current online switch value.
  - This will fetch out the total sum of detections that happen on this particular week day and time index if the person is present at home. This value is nothing but the sum of detections that happen in a five minute interval on specific week day at user’s presence. The result can be even empty which indicates that the passed timestamp does not exist at all at user’s presence in the historical data.
- "Select the “count” of WDay from “historicaldb” where time index = current time index and WDay = current WDay”.
  - This will fetch out the count of specific WDays existing in the “historicaldb”

The division of these two results corresponds to the usual number of detections that happen on a five minute interval of a specific day when the person is present based on the historical data. This corresponds to the historical five min detections. The number of times , the MQTT[3] message “Motion/Voltage” being received at the “MQTT[3] in” node is counted in every five minute which corresponds to the current five minute detections value. Both these values are compared against each other as follows:

Condition	Situation
Current five min detection > historical five min detection	High Intrusion probability
Current five min detection < historical five min detection	Chance of Intrusion
Historical five min detection	High Intrusion probability

Table 8. secondary historical search decisions

This corresponds to the historical secondary search.

The result status of the primary and secondary search is displayed at the Node-RED[9] dashboard for user reference which displays and refreshes the status in every five minute interval. Despite of that, whenever an abnormal situation arises, the respective message is sent out to the user via an email notification.

## B. Statistical methods

Two statistical methods are used to derive relevant information from the logged data ,trainingdata.csv.

### 1) Standard deviation method

Since the data is only logged in the trainingdata.csv file whenever a motion is detected, the timestamps logged in the dataset will be the timestamps of motion detection and the motion value logged in the dataset will be always “1” as follows :

```
28/08/2019 20:17:23,1.0,1.0,1.0
28/08/2019 20:23:44,1.0,1.0,1.0
28/08/2019 20:25:31,1.0,1.0,1.0
28/08/2019 20:31:24,1.0,1.0,1.0
28/08/2019 20:42:15,1.0,1.0,1.0
28/08/2019 20:42:42,1.0,1.0,1.0
28/08/2019 20:56:21,1.0,1.0,1.0
28/08/2019 21:07:07,1.0,1.0,1.0
28/08/2019 21:11:02,1.0,1.0,1.0
28/08/2019 21:15:54,1.0,1.0,1.0
```

Fig9. Datalog.csv file format

The script “stdprediction.py” transforms the datalog.csv into another format as follows

Timestamp	WDay	Time index	moti -on	Online switch	Mobile mac
dd/mm/yy yy:yy:yy	1-7	0-287	1	0 or 1	0 or 1

Table 9. data format after transformation

The script also filters the dataset by applying the selection condition such that online switch value = 1 , mobile mac value = 1 and motion = 1.

Timestamp	WDay	Time index	moti -on	Online switch	Mobile mac
dd/mm/yy yy:yy:yy	1-7	0-287	1	1	1

Table 10. Data after filtering

Hence, the dataset only contain the corresponding WDay and time indices when the person is present and when the motion



is detected which indicates the time periods of motion detections at users presence.

The script groups the motion detected time indices with respect to the WDay values such that the motion detected time indices of the WDay 2, Monday is stored in one list and the motion detected time indices of the WDay 3, Tuesday is stored in another list and so on. As a result, the motion detected time indices of each day is stored in each specific list or each group.

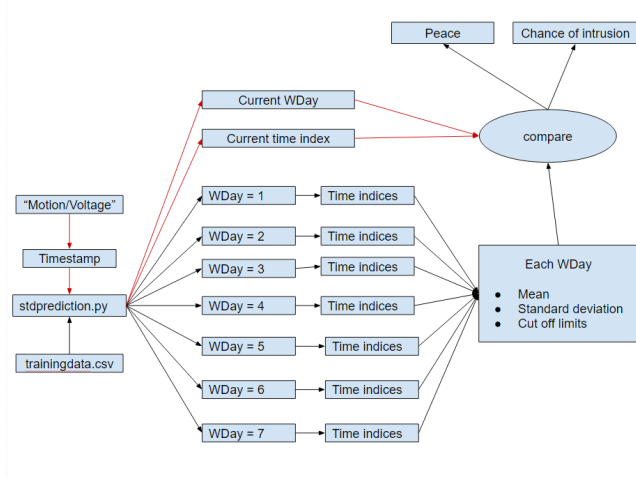


Fig10. Standard deviation logic

The mean and standard deviation of the time indices of each WDay is calculated dynamically with every new detection that happens logged in the csv file. The cut off limit is defined as the thrice times the standard deviation. Likewise, the upper and the lower cut off limit with respect to the standard deviation value is calculated for each WDay. These values gets changed dynamically when a new detection gets logged into the csv file. This logic is implemented on the script which runs on every detection that happens. On every detection, the instantaneous timestamp is passed as arguments to the script which triggers the script to run and the respective week index and time index is derived from the timestamp.

The current WDay specifies the list to be checked, that is the WDay to be checked and if the current time index exceeds the upper cut off limit or fall short of the lower cut off limit of that WDay, then the situation is regarded as an abnormal situation like chance of Intrusion otherwise is regarded as a peaceful situation.

## 2) Interquartile region method

The trainingdata.csv acts as the source file for this method.

The script “iqrprediction.py” transforms the datalog.csv into another format as follows :

Timestamp	WDay	Time index	moti-on	Online switch	Mobile mac
dd/mm/yyyy hh:mm:ss	1-7	0-287	1	0 or 1	0 or 1

Table 11. data format after transformation

The script also filters the dataset by applying the selection condition such that online switch value = 1, mobile mac value = 1 and motion = 1.

Timestamp	WDay	Time index	moti-on	Online switch	Mobile mac
dd/mm/yyyy hh:mm:ss	1-7	0-287	1	1	1

Table 12. Data after filtering

Hence, the dataset only contain the corresponding WDay and time indices when the person is present and when the motion is detected which indicates the time periods of motion detections at users presence.

The script groups the motion detected time indices with respect to the WDay values such that the motion detected time indices of the WDay 2, Monday is stored in one list and the motion detected time indices of the WDay 3, Tuesday is stored in another list and so on. As a result, the motion detected time indices of each day is stored in each specific list or each group.

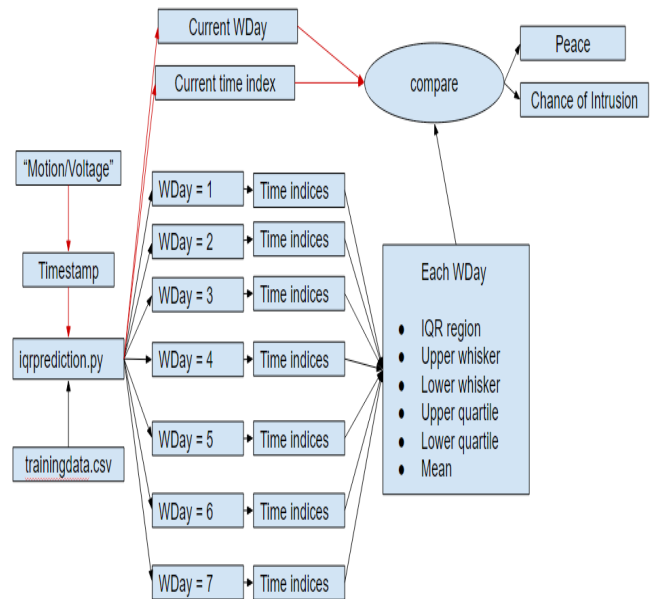


Fig11. Inter-Quartile region logic

The interquartile region, upper whisker, lower whisker, upper quartile, lower quartile and mean of the time indices of each

WDay is calculated dynamically and stored with every new detection that gets logged in the datalog.csv file. These values get changed dynamically when a new detection gets logged into the csv file. This logic is implemented on the script which runs on every detection that happens. On every detection, the instantaneous timestamp is passed as arguments to the script which triggers the script to run and the respective WDay and time index are derived from the timestamp. If the passed current time index of a particular WDay is greater than the upper whisker/ upper quartile or lesser than the lower whisker/ lower quartile or not within the interquartile region of that WDay, then that passed timestamp is regarded as an abnormal situation, that is Chance of Intrusion otherwise is regarded as a peaceful situation.

### C. Machine Learning Approach

Since the application deals with motion and time series, it was decided to choose one of the best suited Time series forecasting model for the Machine Learning functionality. The idea is to track, predict and compare the new motions instances with past occurred instances. The motion data is unstructured, huge and pattern less data (having bigger duration gap length). So, Long short-term memory, LSTM[7] with its capability to remember past data even older ones, was the perfect choice. Apart from these, the lack of many training parameters, for instance in our case, time and motion values are only the parameters that are passed to the machine learning model. LSTM[7] is preferred in such univariate time cases with plethora amount of data.

Whenever there is a motion detected, from the sensor, the data is recorded in the client side (Node-RED[9]) as **Timestamp**: mm/dd/yy HH:mm:ss and **motion value** as "1". This is the data format which is fed into our machine learning model as the input for training.

The machine learning model was built with 67 percent of motion data (timestamp aggregated as 5min duration and the corresponding motion values for the duration is added.) as training set and remaining set of test data was used to validate the same. The other parameters used to train model was

Batch Size : 1  
Epochs : 1500  
Neurons : 4  
Loss : mean\_squared\_error  
Optimizer: adam

The LSTM[7] model obtained was quite satisfactory with the RMSE score of 1.6 as per test data validation given below where the orange plot refers to the predictions.

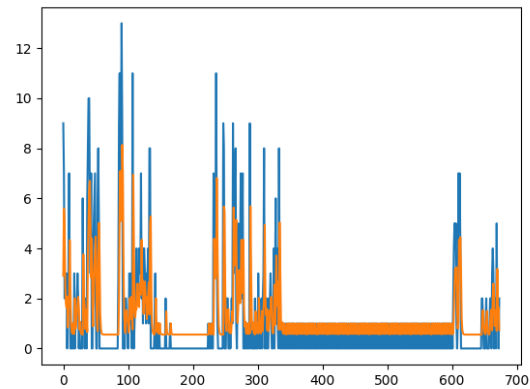


Fig12. LSTM model accuracy report

The model was saved offline using appropriate libraries and used for real time prediction by inputting real time motion values. This practical move was achieved by creating a Flask[13] application (ML server) based on Python where we would load the model once when the server is run initially. The server is good to accept API calls which in turn are motion values. These motion values obtained are passed to the model where it would predict the next motion value (for next 5 min). The output of the model which is the predicted value is fetched back to the client and compared with the real time motion value from the sensor in that specific instance. Based on the variation of the values it is decided, if it's a Intrusion or Motion caused by the user. The application server is enabled as a service "modelstart.service" which is autorun every time the machine boots. Also the service restarts monthly once the training is completed to invoke and load the newly obtained model.

#### 1) Working of LSTM prediction based on Node-RED implementation

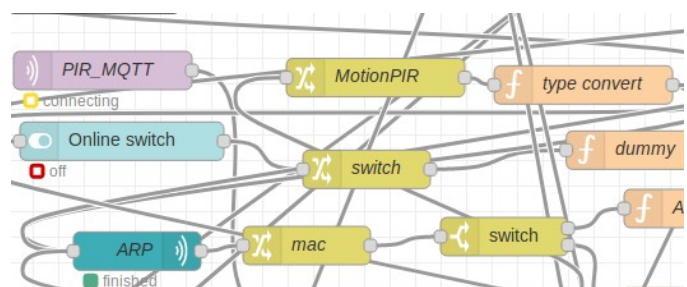


Fig13. Node-RED flow

The PIR\_MQTT[3] node in Node-RED[9] is connected to the broker installed in Raspberry Pi[4] via SCION[5] protocol. Whenever the data is sensed by the sensor, the ESP[2] publishes value "Motion/(battery voltage)" onto the mosquito broker installed in Raspberry Pi[4]. The PIR\_MQTT[3] node in Node-RED[9] (Client VM machine) subscribed to the same broker and topic in Raspberry Pi[4]. (data transfer between remote systems via SCION[5] SIG protocol) receives this data

instantly and feeds it to the member nodes. Based on this, the battery status and the motion value is processed. This motion data is saved in a CSV file “ Trainingdata.csv” which is used monthly to train the model. The training duration is up to every user and also depend on the accuracy of the model obtained.

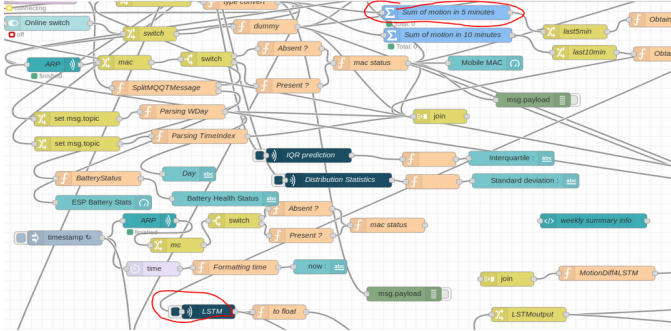


Fig14. LSTM Node-RED flow

As discussed, in case of motion, the Node-RED[9] detects the motion value as “1”. Now since the model was trained with the data aggregated for 5min interval, the model predictions will always be for future 5 mins. Hence during a motion, the real time motion values across last 5 min interval is fed into LSTM[7] to predict motion value for next 5 min interval. As discussed this is done by running a Python file in “Python shell node” in Node-RED[9]. The script would take the motion value as its input argument and make an API call to our Flask[13] application server where the model is loaded and running. This value is fed to model for the prediction. The predicted value is fed back as the output of the script. This predicted value from the Python shell is put in a register and compared with the real time motions at next 5min interval end. Based on the deviation or the overlapping of data, the final decisions are made.

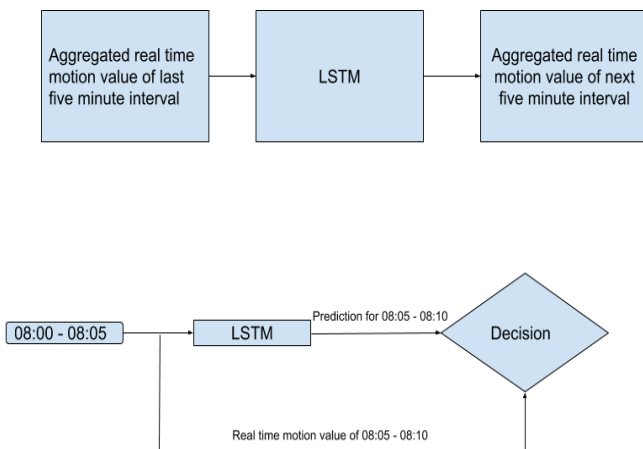


Fig15. LSTM prediction logic

After comparison of real time and predicted values, a difference of these values are calculated. If the difference is greater than 5 motions (which is a customization parameters and could vary with different applications and users), the final decision from the machine learning side as Intrusion. The output decision/response again depends on the value of ARP and online switch obtained. Again this decisive part depends on person to person who can customize the algorithm as per their choice

If the output from LSTM[7] is in a considerable range (threshold 5 motions) it is concluded as “Peace” ignoring switch and mac values.

However if the output from LSTM[7] is considerably deviated from the normal range, then the decision is based on the obtained switch and mac values

- Switch = 1 Mac = 1 message: No chance of Intrusion as Switch and mac configured rightly
- Switch = 0 Mac = 1 message: Detection are observed. Either Switch/Mobile misconfigured, else Intrusion
- Switch = 1 Mac = 0 message: Detections are observed. Either Switch/Mobile misconfigured, else Intrusion

Online switch	Mobile MAC	Message
1	1	“No chance of Intrusion”
0	1	“Detections are observed. Either Switch/Mobile misconfigured, else Intrusion”
1	0	“Detections are observed. Either Switch/Mobile misconfigured, else Intrusion”

Table 13. LSTM decision logic

If the decision turns out to be any of the above 3 other than “Peace” (from the machine learning side) the user is notified by an alert mail sent to their respective mail id. This is done using the mail node in Node-RED[9].

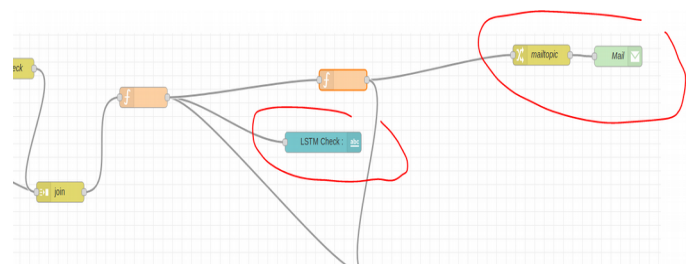


Fig16. LSTM prediction output node

## V. REPORTS

### A. Weekly Detection Summary Report

The trainingdata.csv acts as the source file for this application. The script “weeklysummary\_boxplot.py” transforms the trainingdata.csv into another format as follows:

Timestamp	WDay	Time index	motion
dd/mm/yyyy hh:mm:ss	1-7	0-287	1

Table 14. data format after transformation

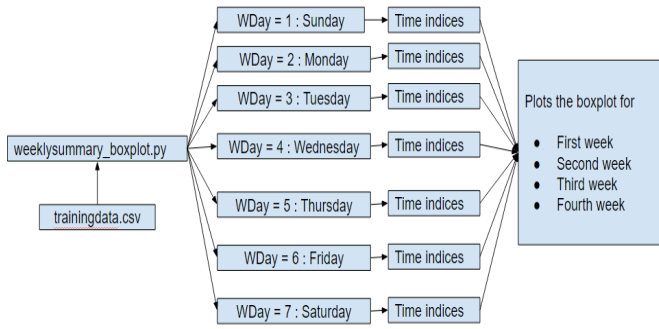


Fig17. Weekly summary report logic

The box plot maps the time indices values with respect to the WDay value. In other words, the script creates the box plot for Sunday by considering the time indices belonging to the WDay = 1 and creates the boxplot for Monday by considering the time indices belonging to the WDay = 2 and so on. This process is iterated for every week so that daily box plots of each of the week is obtained. A week’s box plot summary comprises of boxplots of Sunday, Monday, Tuesday, Wednesday, Thursday, Friday and Saturday. The Summary consists of box plot of the last four weeks. The script is scheduled to run daily at 21:00:00 using the crontab services as shown below:

```
00 21 * * * /bin/sh /home/pi/weeklysummary_boxplot.py
```

### B. LSTM Predictor model Report

After each cycle of LSTM[7] model training based on the recent updated dataset, a report on the LSTM[7] model created is generated specifying the accuracy of the model with respect to the RMSE score. The lower the RMSE value, the better is the model accuracy and performance. With this report, the user is able to access and evaluate the model performance. The blue line indicates the actual prediction values and the orange line indicates the expected prediction values. This report is available to the user via the Node-RED[9] dashboard.

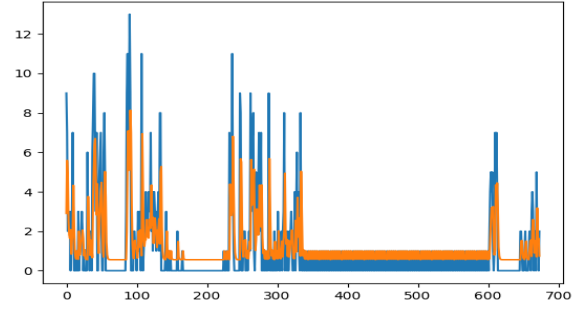


Fig17. LSTM model accuracy report

## VI. IMPLEMENTATION OF GUI

THIS chapter explains how the graphical user interface has been implemented at the client SCION[5] AS node. The flow-based, visual programming development tool, Node-RED[9] has been employed for this purpose. Several types of nodes has been used and configured accordingly for several use cases like to retrieve the MQTT[3] message over the SCION[5] network, to realize the online switch functionality, to realize the ARP based mobile mac functionality, to run the Python scripts at real time, to customize and parse the data, to provide dashboard facility for the users and so on. The most important nodes and it’s configuration and it’s requirement has been discussed below

- “MQTT In”, a network node which is configured with the Raspberry Pi[4]’s ip address/ SCION[5] instance (MQTT[3] broker) so as to receive the MQTT[3] message ‘Motion/Voltage’ whenever motions are being detected.
- “Switch”, a dashboard node which provides a switch functionality at the Node-RED[9] dashboard so that the user can either turn ON or OFF depending on his presence at home.
- “ARP” node, a network node which provides the content of the ARP table with which we can detect the ip address and the MAC address of the devices connected in the network. The node is configured by specifying the MAC address of the user’s mobile. This tells if the user’s mobile is connected to the network by checking if the MAC address of the mobile is present in the ARP table.
- “Join”, a sequence node which is used at several instances to combine the different messages together so that certain functionalities and comparison checks can be made among the different messages being received.
- Custom “Function”, a function nodes which is used to parse, format the messages being received, to create



custom context dependent functionalities and comparison checks with the messages.

- “Python shell”, an input node which has been used to run the Python scripts upon triggering of the node. The node gets triggered whenever it receives an input which can be also passed as arguments into the Python[13] script. The Python shell has been used to run various scripts which gets triggered on motion detected timestamps (weeklysummary\_boxplot.py, iqrprediction.py, stdprediction.py) or every five minute intervals(primaryhistoricalsearch.py), every five minute current detection value.

- “Email, a network node which is used to sent an email notification whenever any abnormalities in detections are observed at peculiar timings based on the implemented approaches. The node is configured by specifying the email credentials of the user.

- “Csv”, A parser node to parse the data in the csv format which has been used to log the data being received: timestamp, PIR[1] motion data, online switch data, mobile mac data in two different csv files namely datalog.csv and trainingdata.csv. The former acts as the source file for InfluxDB[6] upload and the latter acts as the source file for various already discussed approaches.

- Dashboard nodes which comprises of text nodes, gauge nodes, chart nodes has been used to serve the purpose of visualization to the user so that the user is able to know the live status of the home remotely.

This acts as a Graphical User Interface for the users so that they can monitor the live status of the home remotely. On this GUI, the user can control the online switch value based on his/her presence. He/She can turn the switch to either ON or OFF based on his presence/absence. Whenever any email alerts are sent to the user stating any abnormalities, the user is able to monitor the system from any remote location.

## VII. RESULTS

THIS chapter describes the results observed after the completion of relevant development tasks during the entire course of the application development which involves the following:

- Reception of MQTT[3] messages which are being published from the ESP[2] module on motion detection at the Raspberry Pi[4] MQTT[3] broker end.
- Generation of trainingdata.csv and datalog.csv which combines four parameters namely timestamp, PIR[1] MQTT[3] motion message, online switch value, mobile mac value.

- Successful upload of datalog.csv to the InfluxDB[6] database.

```
Connected to http://localhost:8086 version 1.7.9
InfluxDB shell version: 1.7.9
> use historicaldb
Using database historicaldb
> SELECT * from sensor
name: sensor
time                Mobile PIR Switch Time_index WDay type
----
1567022100000000000 1      8   1    239      4 motion
1567022400000000000 1      7   1    240      4 motion
1567022700000000000 1      2   1    241      4 motion
1567023000000000000 0      0   0    242      4 motion
1567023300000000000 1      1   1    243      4 motion
1567023600000000000 1      1   1    244      4 motion
```

Fig18. InfluxDB data upload result

- LSTM[7] prediction status based on the current five minute detection value. The model takes the current five minute detection count as the input and outputs the predicted five minute detection count for the next five minute interval and classifies the situation as normal or abnormal by comparing them over a custom percentage.

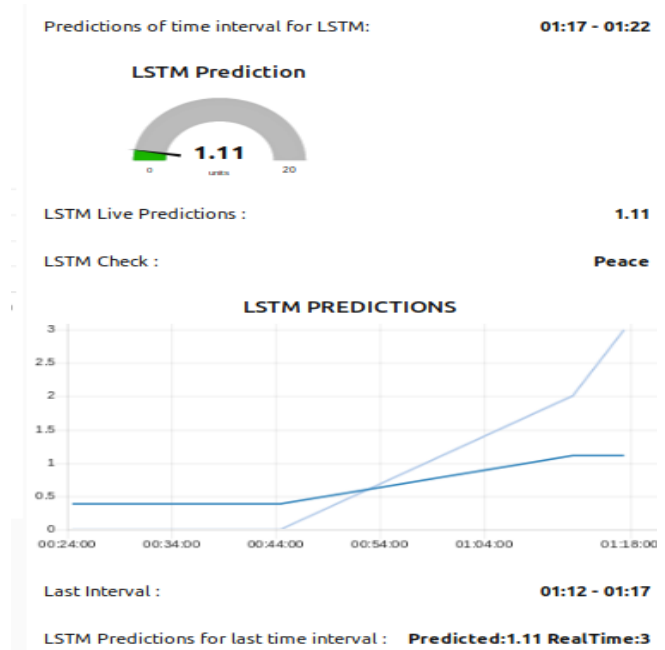


Fig19. LSTM prediction result on Node-RED dashboard

- Historical Search status in every five minute interval by comparing the current five minute detection count with the past five minute detection count of that respective day which is stored in the InfluxDB[6] “historicaldb”.

Predictions of time interval for Search: 01:17 - 01:22

Historical **Detections are observed, Either Online Switch/**  
 Primary Search **mobile wifi misconfigured, else suspects a**  
 Check : **intrusion**

Historical Secondary Search Check :

Fig20. Historical search prediction result on Node-RED dashboard



- Statistical methods prediction status in every detection which classifies the detection as normal or abnormal based on the timing of the detection and the usual pattern. It checks if the detected timestamp lie inside or outside the inter-quartile range and if the detected timestamp deviates from the usual median value.

Interquartile :	17:00:36 is At Peace
Standard deviation :	17:00:35 At Peace

Fig21. Statistical methods results on Node-RED dashboard

- Weekly summary view which creates a box plot of detected timestamps of each day for four weeks. With this, the user can get a summary of timings when he is usually present/absent at home and can view any abnormal detection (outliers) happened on any day which he never observed at first place and hence, he/she can be cautious for this regard.

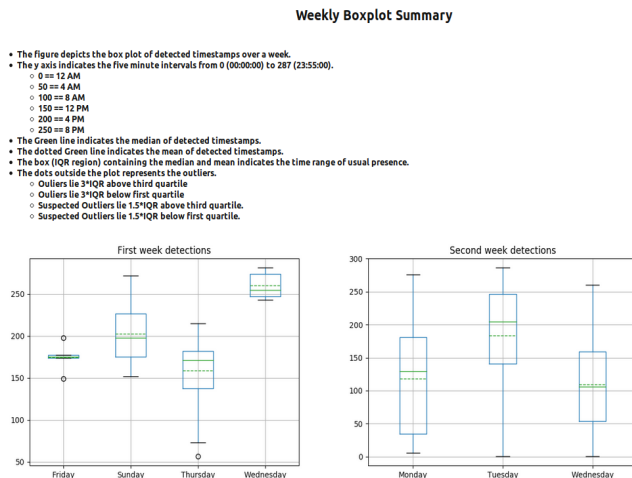


Fig22. Weekly summary report on Node-RED dashboard

- Mail notification to the user in case of abnormal situation.

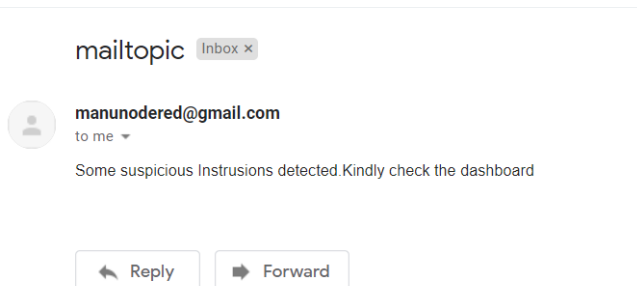


Fig23. Email notification to the user

## VIII. CONCLUSION AND FURTHER WORK

This scientific work focuses on the implementation of a Smart Home IoT use case, Burglar detection system with the use of motion sensor rather than the camera via the SCION[5] network using the SCION[5] IP gateway (SIG). The application involves a client server architecture where the SCION[5] AS instance configured, MQTT[3] broker enabled Raspberry Pi[4] acts as the SCION[5] AS server which receives the motion data from the remotely placed ESP[2] module coupled PIR[1] motion sensor over WiFi and the SCION[5] AS instance configured local machine acts as the SCION[5] AS client. This application posses wide range of development possibilities by employing more number of motion sensors so that the readings from each motion sensor can be sequenced together to derive the velocity ,direction of movement and even can suggest the approximate number of personnel present in an area. With this, it is possible to identify the intruder in a more accurate way and hence, the accuracy of the burglar detection can be improved. There is no need to employ cameras within the home to detect the burglar presence which is actually a security concern for the user as it disturbs the privacy of the user because the user might have a feeling of insecurity when he becomes conscious of the fact that a camera is being monitoring him continuously.

## REFERENCES

- PIR motion sensor - This motion sensor module uses the LHI778 Passive Infrared Sensor and the BISS0001 IC to control how motion is detected. Available: <https://www.epitran.it/ebayDrive/datasheet/44.pdf>
- ESP8266/ESP - Espressif's ESP8266EX delivers highly integrated Wi-Fi SoC solution to meet users' continuous demands for efficient power usage, compact design and reliable performance in the Internet of Things industry. Available: [https://www.espressif.com/sites/default/files/documentation/0a-esp8266ex\\_datasheet\\_en.pdf](https://www.espressif.com/sites/default/files/documentation/0a-esp8266ex_datasheet_en.pdf)
- MQTT, MQTT is a machine-to-machine (M2M)/"Internet of Things" connectivity protocol. Available: <http://mqtt.org/>
- Raspberry Pi - A small and affordable computer that you can use to learn programming. Available: <https://www.raspberrypi.org/>
- SCIONLab Coordination Service, SCION is the first clean-slate Internet architecture designed to provide route control, failure isolation, and explicit trust information for end-to-end communication. Available: [https://www.SCION\[5\]lab.org](https://www.SCION[5]lab.org)
- InfluxDB – InfluxDB, an open source time series database, purpose-built by InfluxData for monitoring metrics and events, provides real-time visibility into stacks, sensors. Available: <https://www.influxdata.com/>
- LSTM, The Long Short-Term Memory recurrent neural network has the promise of learning long sequences of

observations.. Available :<https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python/>

## APPENDIX

[8] Grafana, An open source analytics and monitoring solution for every database. Available: <https://grafana.com/>

[9] Node-RED, A programming tool for wiring together hardware devices, APIs and online services in new and interesting ways. Available: <https://nodered.org/>

[10] HC-SR501, This motion sensor module uses the LHI778 Passive Infrared Sensor and the BISS0001 IC to control how motion is detected. Available: <https://www.epitran.it/ebayDrive/datasheet/44.pdf>

[11] Wemos D1 Mini -Wemos D1 Mini is a mini Wi-Fi device based on ESP8266EX chip. Available: [https://docs.zerynth.com/latest/official/board.zerynth.wemos\\_d1\\_mini/docs/index.html](https://docs.zerynth.com/latest/official/board.zerynth.wemos_d1_mini/docs/index.html)

[12] Arduino UNO -The Arduino UNO is programmed using the Arduino Software (IDE), our Integrated Development Environment common to all our boards and running both online and offline. Available: <https://www.arduino.cc/en/Guide/ArduinoUno>

[13] Flask-Flask depends on the Jinja template engine and the Werkzeug WSGI toolkit. Available: <https://flask.palletsprojects.com/en/1.1.x/>

[14]<https://www.datacamp.com/community/tutorials/lstm-python-stock-market>

[15]<https://appcodelabs.com/introduction-to-iot-build-an-mqtt-server-using-raspberry-pi>

[16]<https://diyprojects.io/esp8266-deep-sleep-mode-test-wake-pir-motion-detector/>

[17]<https://blog.keras.io/building-a-simple-keras-deep-learning-rest-api.html>

[18]<https://code.tutsplus.com/tutorials/scheduling-tasks-with-cron-jobs--net-8800>

[19]<https://www.2daygeek.com/how-to-check-all-running-services-in-linux/>

[20] <https://learn.adafruit.com/pir-passive-infrared-proximity-motion-sensor/how-pirs-work>