PROJECT

Shubhika GARG

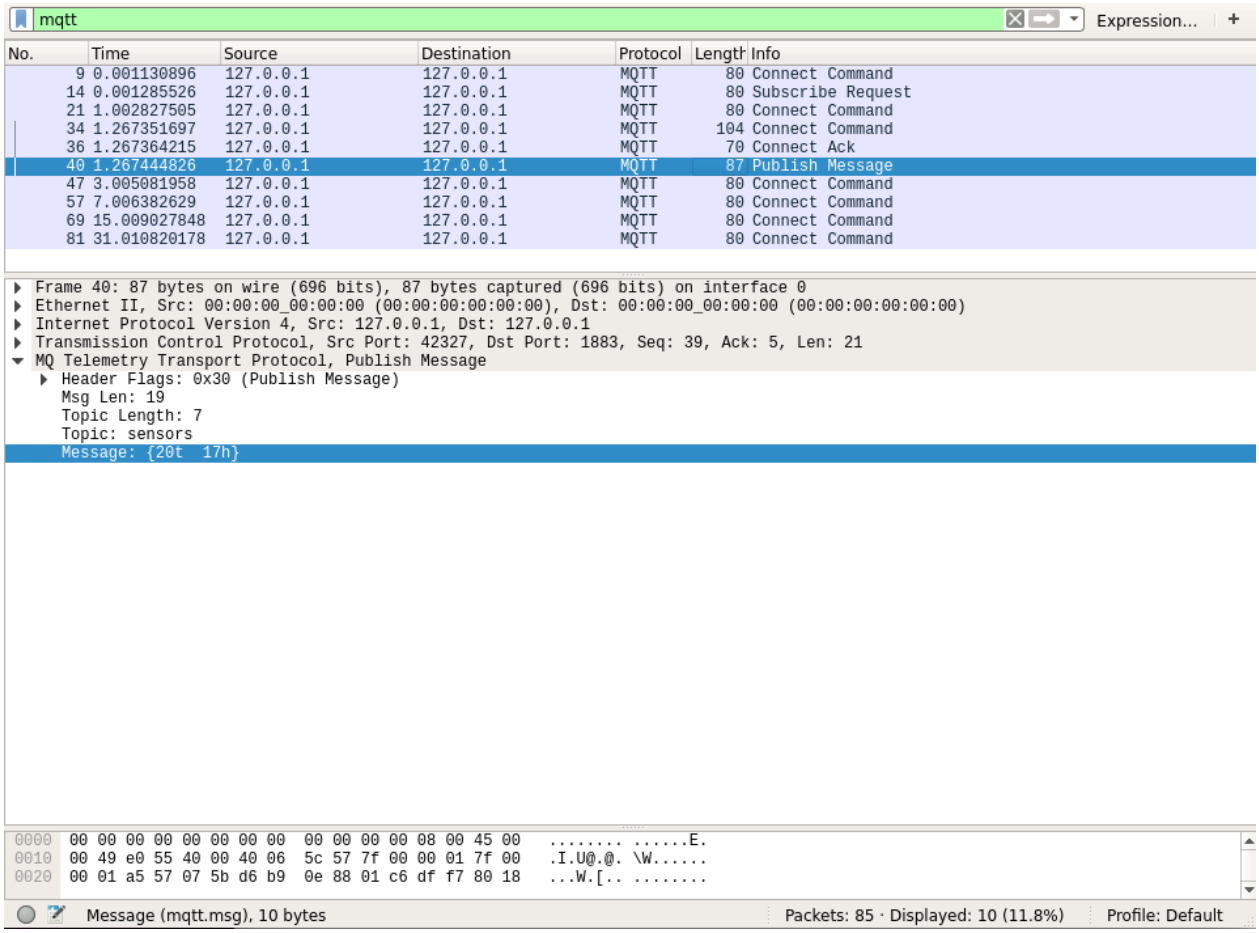
1. A company wants to deploy 10000 devices in a City. Each device (Raspberry) includes two sensors, to monitor humidity and temperature. The devices send the data to a data centre. In case the temperature and humidity are collected twice per day:

a) What is the application protocol that you advise to use? Why?

I would advise using MQTT Protocol. With MQTT, devices can easily publish their sensor data to a central broker, which can then be consumed by the data centre. MQTT also grants a flexible and extensible message format that can accommodate sensor data from multiple sources. With 10,000 devices sending data to a central data centre, scalability is an important consideration. MQTT is well-suited to handle this scale, as the broker can easily handle a large number of incoming messages, and the devices can efficiently transmit their data to the broker without requiring excessive bandwidth or processing power. MQTT supports Quality of Service (QoS) levels that allow for reliable message delivery even in the face of network disruptions or other issues. MQTT also has better support for unreliable network conditions and intermittent connectivity

b) Implement the device (device.py) and the data centre collector (datacenter.py) using Python and a library of your choice. The script "device.py" sends the data in a format of your choice (justify why you choose this format) to "datacenter.py" following the proposed protocol. The data size is 10 Bytes.

Wireshark capture:



Python Codes:

device.py:

```
import paho.mqtt.client as mqtt
import random
import time

client=mqtt.Client(protocol=mqtt.MQTTv31)
client.connect("localhost")


while True:
                temp = random.randint(10,30)
                humd = random.randint(10,30)
                payload = '{{{0:02d}t  {1:02d}h}}'.format(temp,humd)
                client.publish("sensors", payload)
                time.sleep(12*3600)
```

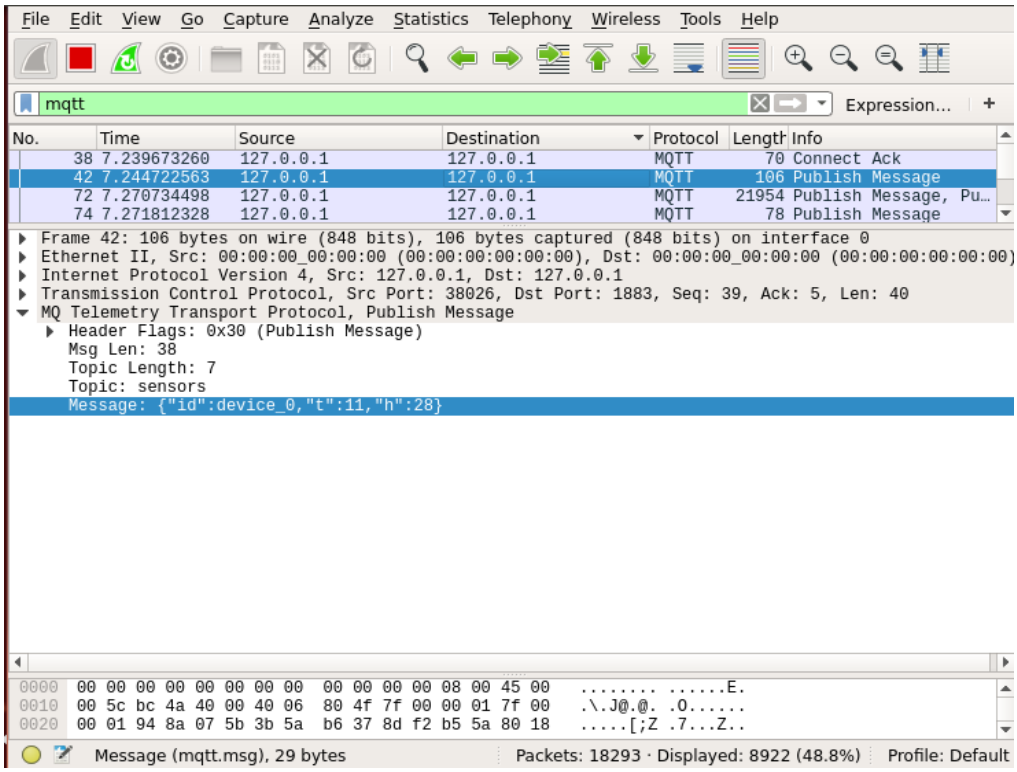
datacenter.py:

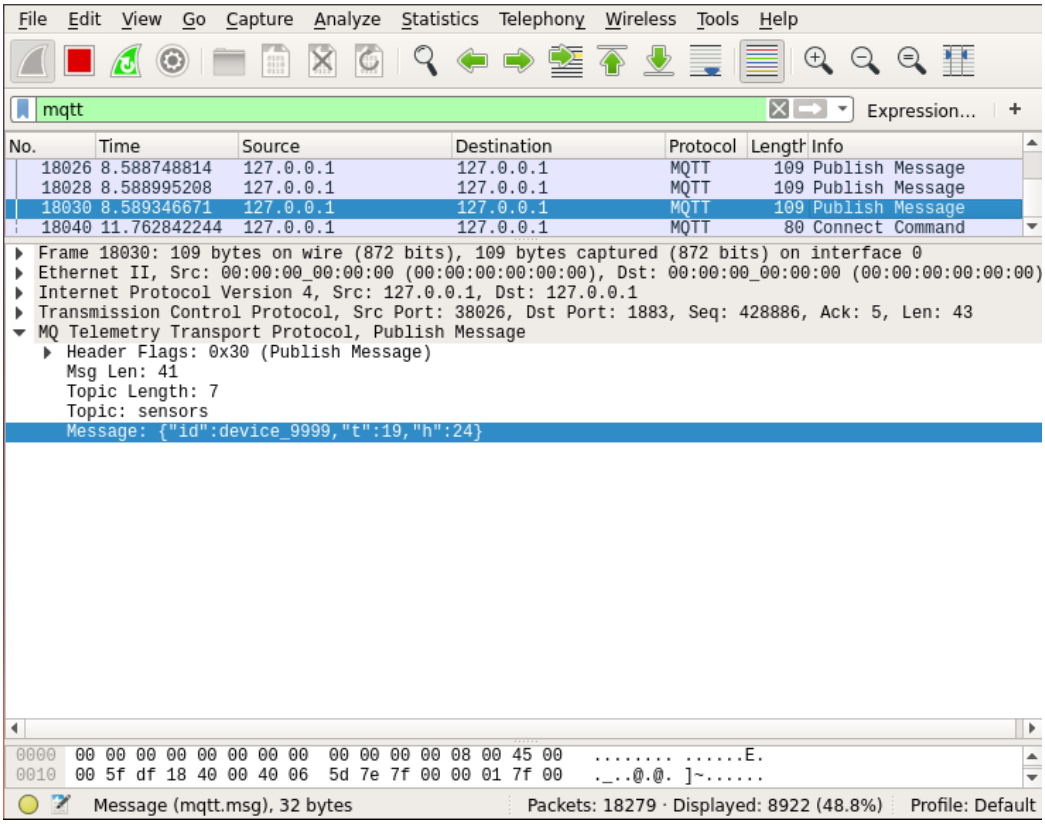
```
import paho.mqtt.client as mqtt

def on_message(client, userdata, msg):
                print("Payload:", msg.payload())
client = mqtt.Client()
client.connect("localhost",1883,60)
client.subscribe("sensors")
client.on_message=on_message
client.loop_forever()
```

Message format: ***payload = '{{{0:02d}t {1:02d}h}}'.format(temp,humd)***

This is a compact and readable format that represents temperature and humidity values as two digits each, with a leading zero if necessary to ensure a consistent format with no separators or other overhead. This can be useful when bandwidth or message size is a concern.

c) Write a script that simulates 10000 devices. Run the simulation (you can change the simulation time scale).





I chose the message format ***payload = '{{"id":{0},"t":{1},"h":{2}}}'.format(did,temp,humd)***

This format uses key-value pairs to represent the data, where the keys are strings that represent device id, temperature and humidity ("id", "t", "h") and the values are the actual data. Using this format allowed the data centre to easily parse and interpret the data received from the devices, while also allowing the devices to efficiently send only the necessary data without any unnecessary overhead.

d) How many messages are received per day ? What is the total amount of data transferred in Bytes. Hint: first compute it mathematically then confirm it in practice.

Every device gives two sensor readings per day (as temperature and humidity are collected twice per day). We have 10,000 devices.

Number of messages received per day = 2 * 10000 = 20,000 messages/day

Each message is 10 bytes.

20,000 messages/day * 10 bytes/message = 200,000 bytes/day

2. In case the temperature and humidity are sent only if a threshold is exceeded.

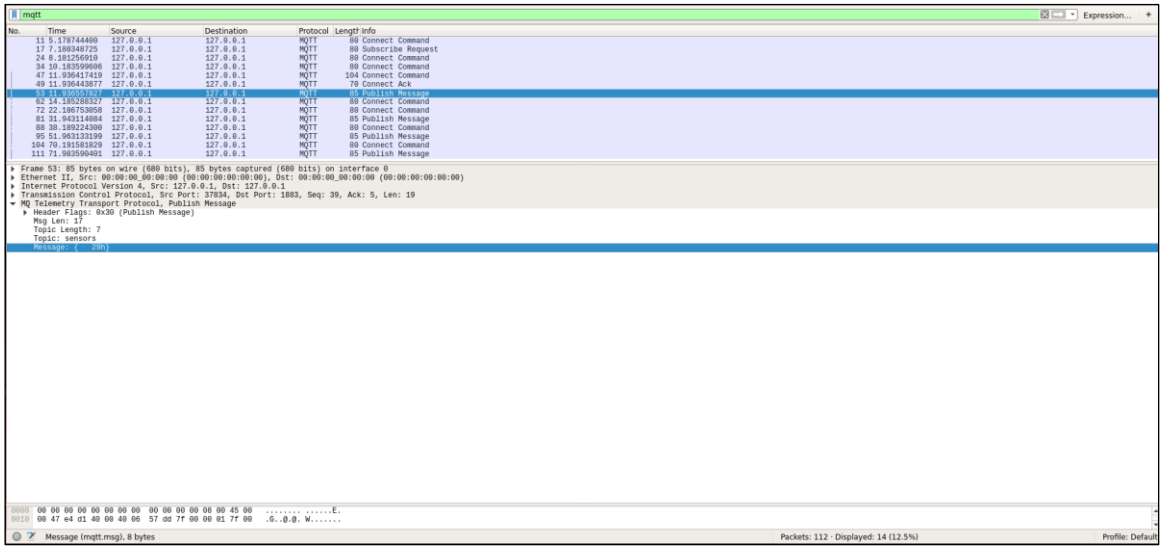
a) What is the application protocol that you advise to use? Why?

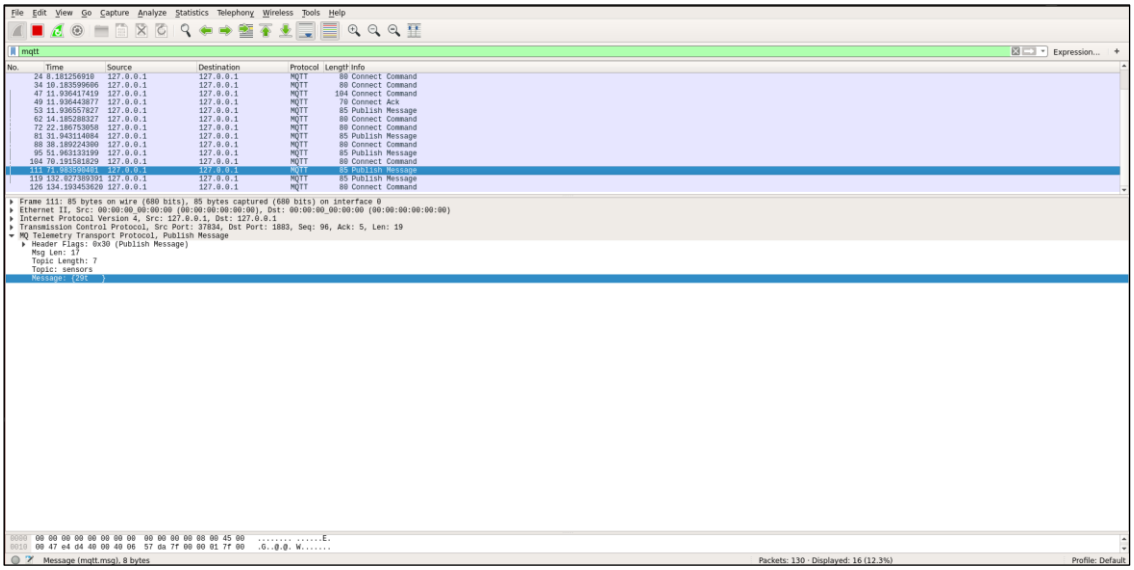
I would advise using MQTT Protocol. With MQTT, devices can publish messages to the broker only when the temperature or humidity exceeds a certain threshold, rather than sending data at fixed intervals or continuously. This selective transmission of data helps to reduce the amount of data being transmitted and processed, which can be important for large-scale IoT applications. Using MQTT in this scenario allows for efficient, reliable, and selective transmission of temperature and humidity data from the Raspberry Pi devices to the central data centre, while also minimizing the bandwidth and processing power required by the devices.

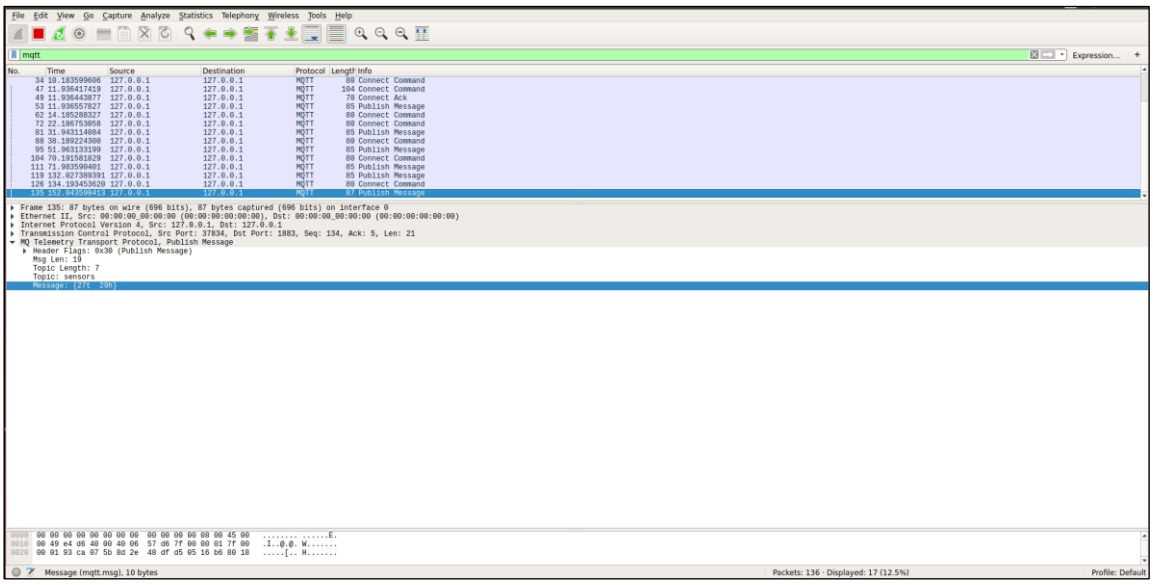
b) Implement the device (device.py) and data centre collector (datacenter.py) using Python and a library of your choice. The script "device.py" sends the data in a format of

your choice (justify why you choose this format) to "datacenter.py" following the proposed protocol. The message size is 8 Bytes if it includes humidity or temperature only or 10 Bytes if both.

Wireshark Captures:







Python codes:

device.py:

```
import paho.mqtt.client as mqtt
import json
import random
import time

client=mqtt.Client(protocol=mqtt.MQTTv31)
client.connect("localhost")

humd_threshold = 20
temp_threshold = 25

while True:
      temp = random.randint(10,30)
      humd = random.randint(10,30)

      if humd > humd_threshold or temp > temp_threshold:

            if humd > humd_threshold and temp > temp_threshold:
                  payload = '{{{0:02d}t  {1:02d}h}}'.format(temp,humd)
            elif humd > humd_threshold:
                  payload = '{{{0:02d}t   }}'.format(humd)
            else:
                  payload = '{{   {0:02d}h}}'.format(temp)

            client.publish("sensors", payload)

      time.sleep(20)
```

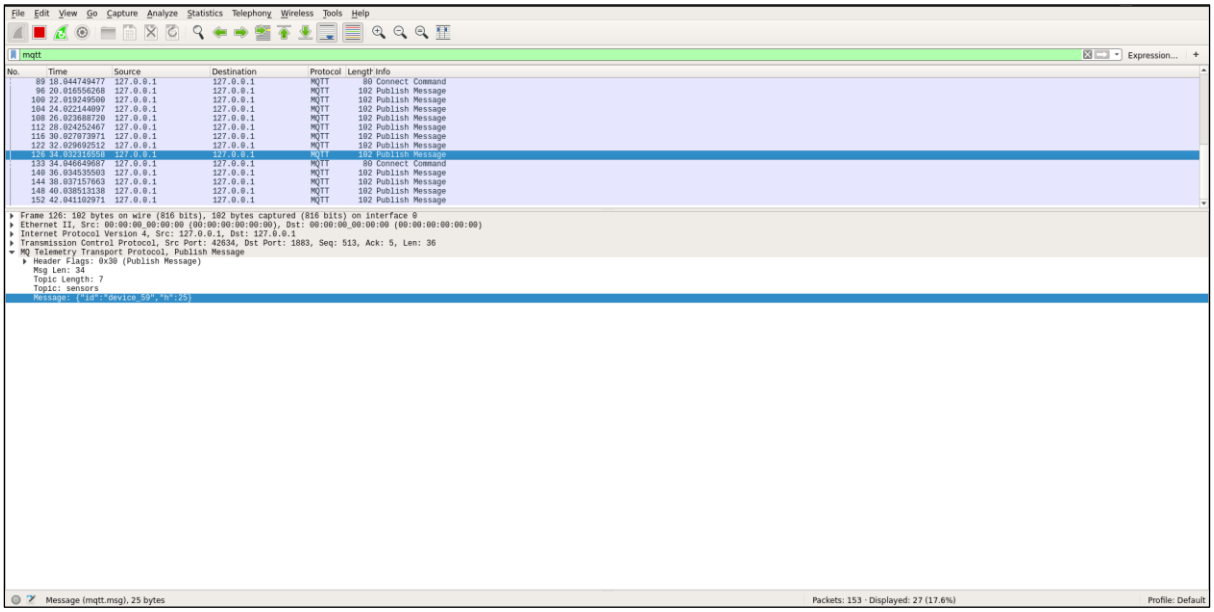
datacenter.py:

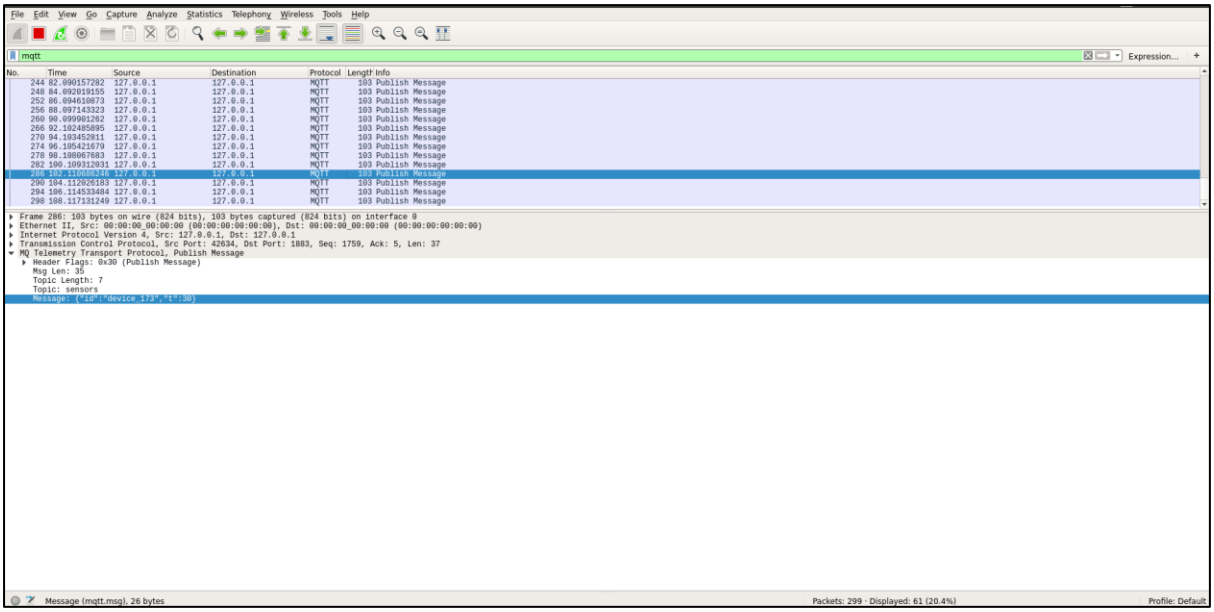
```
import paho.mqtt.client as mqtt
import time

def on_connect(client, userdata, flags, rc):
    print("Connected with result code " + str(rc))
    client.subscribe("sensors")

def on_message(client, userdata, msg):

                print("Payload:", msg.payload())

client = mqtt.Client()
client.on_connect = on_connect
client.connect("localhost",1883,60)
time.sleep(2)
client.subscribe("sensors")
client.on_message=on_message
client.loop_forever()
```

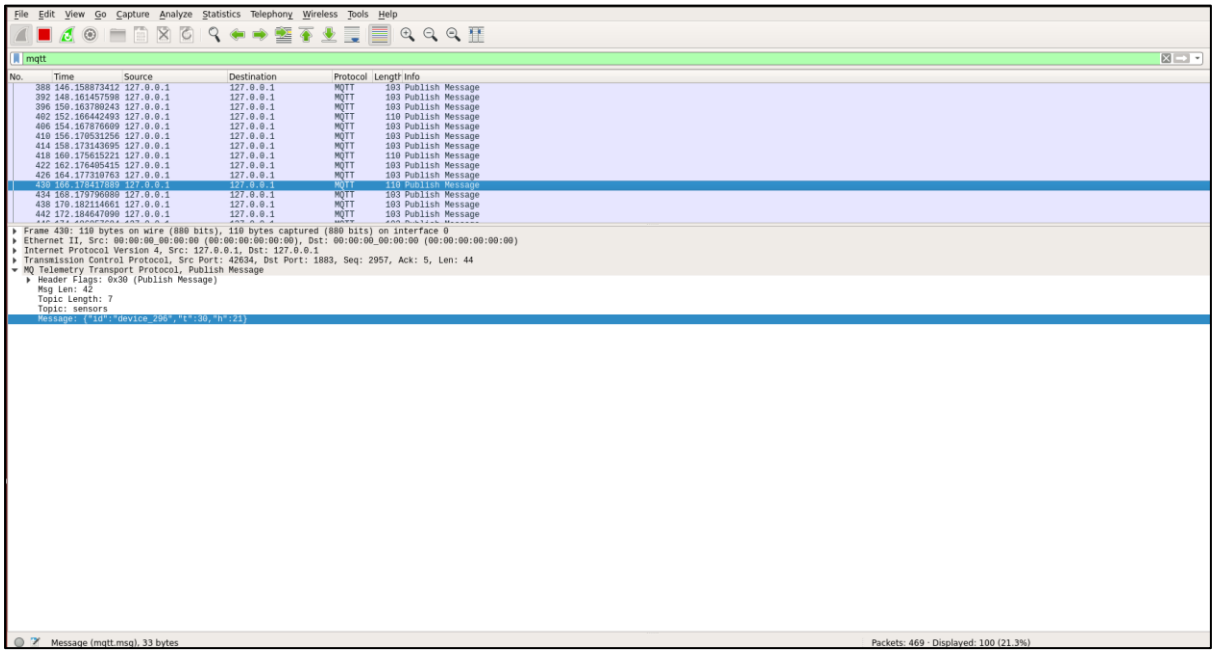
Choice of message format:

The format is compact, taking up only 10 bytes of data if both temperature and humidity are transmitted, or 8 bytes if only one value is transmitted. It is designed to use the minimum amount of data necessary to transmit the temperature and humidity values. It can be easily extended to include additional data fields or metadata (We can include a device ID field to identify the source of the data).

c) Write a script that simulates 10000 devices. Run the simulation (you can change the simulation time scale). Knowing that the probability that humidity exceeds the threshold is 0.5 and temperature exceeds the threshold is 0.2. The events are independent.







d) How many messages are received per day ? What is the total amount of data transferred in Bytes. Hint: first compute it mathematically then confirm it in practice.

P(humidity exceeds threshold) * P(temperature exceeds threshold) = 0.5 * 0.2 = 0.1

P(humidity exceeds threshold) * P(temperature does not exceed threshold) = 0.5 * 0.8 = 0.4

P(temperature exceeds threshold) * P(humidity does not exceed threshold) = 0.2 * 0.5 = 0.1

The expected number of messages per day per device is:

(0.1 * 1 + 0.4 * 1 + 0.1 * 1) = 0.6

For 10,000 devices:

The expected total number of messages per day for 10,000 devices is:

10,000 * 0.6 = 6,000

The expected total amount of data transferred per day can be calculated as:

(0.1 * 10 + 0.4 * 8 + 0.1 * 8) * 6,000 = 30,000 Bytes

The expected total amount of data transferred per day for 10,000 devices is 30,000 Bytes

Note:

To compute this, I added three counter variables for all three cases in my Python code:

```
usertp@usertp-VirtualBox:~/project2$ python devices.py
('Total messages sent for only temperature: ', 378)
('Total messages sent for only humidity: ', 2328)
('Total messages sent for both: ', 123)
usertp@usertp-VirtualBox:~/project2$
```

So, when only either temperature or humidity is sent, the message size is 8 bytes while if both are sent, the size is 10 bytes:

Therefore, (378*8)+(2328*8)+(123*10) = 22,878 Bytes for 10,000 devices

```python
import paho.mqtt.client as mqtt
import random
import time

humd_threshold = 20
temp_threshold = 25

client = mqtt.Client(protocol=mqtt.MQTTv31)
client.connect("localhost")

n = 10000
device_ids = ["device_{}".format(i) for i in range(n)]
count_temp = 0
count_humd = 0
count_both = 0

for did in device_ids:
    temp = random.randint(10, 30)
    humd = random.randint(10, 30)

    if random.random() < 0.5 and humd > humd_threshold:
        if random.random() < 0.2 and temp > temp_threshold:
            payload = '{{"id":"{0}","t":{1},"h":{2}}}'.format(did, temp, humd)
            count_both = count_both+1
        else:
            payload = '{{"id":"{0}","h":{1}}}'.format(did, humd)
            count_humd = count_humd+1
    elif random.random() < 0.2 and temp > temp_threshold:
        payload = '{{"id":"{0}","t":{1}}}'.format(did, temp)
        count_temp = count_temp+1
    else:
        continue

    client.publish("sensors", payload)

print("Total messages sent for only temperature: ", count_temp)
print("Total messages sent for only humidity: ", count_humd)
print("Total messages sent for both: ", count_both)
```