International Journal of Computing and Business Research (IJCBR)
ISSN (Online) : 2229-6166
International Manuscript ID : 22296166V12I4202203
Volume 12 Issue 4 September 2022

Real Time Data Extraction and Analytics on Internet of Things using MQTT

Gurpreet Kaur Brar
Brar Dhaliwal Services
E-mail : brardhaliwalservices@gmail.com

Abstract

Nowadays, with the advent and adoption of smart gadgets, the information is getting disseminated to huge level. A number of social media platforms are in usage and almost every person is using these platforms. This segment of huge data on social media and other platforms gives the research dimension towards real time data extraction and its analytics so that the human emotions can be extracted. Sentiment analysis also known as opinion mining or emotional AI refers to the systematically identifying, extracting, quantifying and studying affective and subjective statements using natural language processing, text analysis, computational linguistics and biometrics. The sentiment analysis applies widely to customer voices, for example reviewing and monitoring responses, online and social media, and medical supplies for applications ranging from marketing to customer service to clinical medicine. In addition, introduction of intelligent cities, intelligent workplaces, intelligent homeland robotics and several more are part of the government and business agenda. In addition to the classical applications of data extraction, the data from interconnected devices can be fetched using Internet of Things (IoT) based advanced protocols. IoT is also referred to as Ubiquitous or Pervasive Computing and these technologies are used for many real time applications. This research manuscript is focusing on the

usage of Message Queuing Telemetry Transport (MQTT) protocol in IoT for the real time data extraction and analytics patterns for assorted research applications.

Keywords: Data Extraction, IoT, IoT Communications, MQTT, MQTT and Cloud, Real Time Data Extraction

Introduction

Sentiment Data Collection is one of the key fields that academics and clinicians use extensively. There are a range of methods and software to capture live data sets, tweets, emotional features in this approach. With the aid of such tools, Twitter, Facebook, WhatsApp and many other social media portals can extract tweets and messages in real-time. This method of sentiment analytics can be used to evaluate and forecast the emotional characteristics of Internet users on social media portals. Assume that we want to determine a celebrity's total average ranking. The advanced programming scripts are used to get the live tweets from social media. Then the data gathered can be analyzed in the form of tweets or messages using a toolkit for natural language processing and a forecast is made whether this individual human, film or celebration gets famous or not [1].

Following is the statistical reports from InternetLiveStats.com and Statista.com about the real time data on social media and related web portals.

As per the research analytics, on Twitter more than 500 million users broadcast every day with 350 million tweets. In addition, there are approximately 571 new websites on the World Wide Web every minute. On smart phones there are more than 5 billion subscribers at the same time and the data needs to be evaluated.

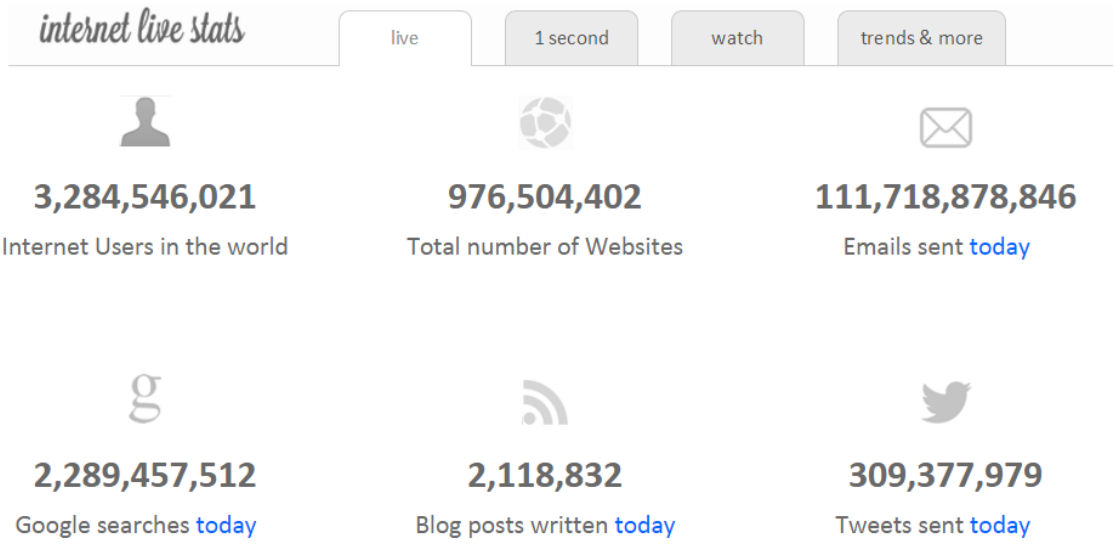

Figure 1 - Screenshot from InternetLiveStats.com

Every second, on Facebook, five new profiles are created. Any 83 trillion false profiles remain. 890 trillion daily active users upload about 300 trillion images per day. Around 320 TB data is transmitted every day with an average usage time of 21 minutes on social media and there is huge scope of research in real time data analytics [2].

Real-time data collection and processing is one of the main areas for a variety of applications, including nostalgic data analysis and criminal data investigation. This technique is often referred to as site crawling and commonly used for statistical mining and the exploration of information in real time to allow social networks understand the individual data about the single user or object.

The political parties are working out equal introduction to get their party citizens' reviews in the election with the possibility of winning. Moreover the business giants often use such methods to get input from the general public on their product [3].

Real time statistics are collected from social media and web portals, and customer ratings are gathered in order to see if users feel about their goods and services. This is further diversified into the manipulation of consumer values or emotion [4, 5].

Data Extraction and Internet of Things (IoT)

The Internet of Things (IoT) and Cloud are now used for high-performance protocol convergence. The Toll Plaza Fastag contactation is a conventional IoT-implementation, which automatically carries out a tiny chip on the nearby sensor [6].

These innovative implementations resolve scalability and security problems for effective IoT and advanced Cellular Networks in the era of conducting networking, including 4G, 5G and related technology. [2]. As per the reports, the Fifth Generation (5G) broadband networks will cross 200 million connections; according to Statista.com's research analytics, the performance system needs to be developed [7, 8].

Data Application Protocols with IoT

IoT blends smart gadgets with less resource access and low-energy concerns. There is a special protocol, Message Queuing Telmetry Transport (MQTT) [9, 10], designed for light and low-energy communications to communicate in the IoT environment, to reduce device and gadget loads [11].

MQTT is the primary IoT protocol for sophisticated networking and wireless. MQTT is used as a low-powered communication data protocol [12].

Table 1: Key Protocols in IoT communications

Protocols	**Layer**
6LowPAN, RPL, IPv4/IPv6,	Infrastructure
MQTT, AMQP, CoAP, Node, Websocket,	Data Protocols
Wifi, LPWAN, Bluetooth,	Communication Transport
Physical Web, DNS-SD, mDNS,	Discovery
OMA-DM, TR-069,	Device Management
Web Thing, JSON-LD	Semantic
URIs, EPC, IPv6, uCode,	Identification

Implementation Dimensions of MQTT Broker

The key challenge of lightweight, secured and low power networking is the speed and efficiency of communication between multiple equipment and gadgets in high frequency networks. The vast range of devices and gadgets with different configurations are made possible by MQTT to communicate [13].

Two types of objects, namely publisher and sender, are found in IoT communication. Transmission of data signals is called publication. Technically, it is called Publishing the transition of data from the machine to a different end. The device of the user is known as the subscriber for IoT communication [14]. The author is called a publisher and the recipient is called a subscriber. The MQTT broker or the IoT server that manages contact between publisher and subscriber is the key point. MQTT broker manages the publisher's data for the other subscribers as shown in Figure 1.

For example, the data must be transmitted from a temperature sensor to a handheld phone of the farmer. In this example, the temperature sensor (publisher) sends the data signal while the farmer's unit is the data receipt (subscriber) [15].

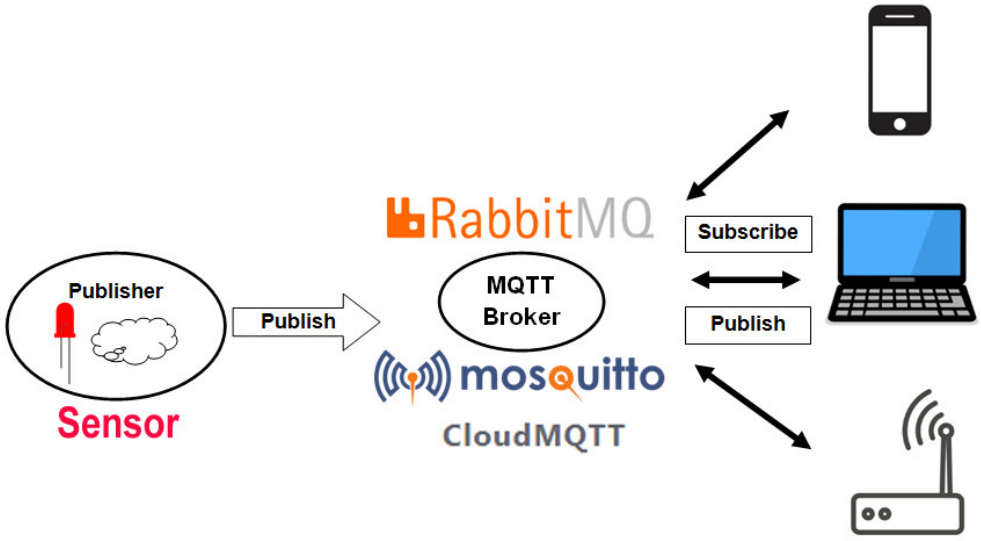

Figure 2: Key Constituents of MQTT

A number of MQTT Brokers, including Mosquitto that implements the MQTT protocol for IoT related communications, are available in free and open source distribution. It is lightweight, versatile and suited for use on any system, including computers and servers, from a low-power single-board, like Arduino, ESP8266. It can be used on on a cloud-based server which implements the Mosquitto broker, rather than using the Mosquito on a local PC, so that IoT communication is controllable on the internet [16].

MQTT is an IoT (Internet of Things) basic OASIS message protocol. It is designed as a lightweight message transport to publish/subscribe, which is suitable for linking remote devices with a minimal code footprint and low network bandwidth. Today,

MQTT is used in a wide range of industries, such as automobiles, production, telecoms, oil and gas, etc.

Two types of network entities are specified in the MQTT protocol: a message broker and many clients. A MQTT broker is a client's server which transmits all messages from its clients to the appropriate target clientes. [13] MQTT is a client of any system which runs a MQTT library and connects to a MQTT broker through the network [17]. MQTT is a network-based server which receivet the messages from clients and then routes them to the appropriate destination clients [18].

Knowledge in a hierarchy of subjects is ordered. When a publisher would have a new item of data to distribute, it sends a control message with the data to the connected broker. The broker then distributes the details to all customers who subscribe to the subject. The publisher does not have to have any information about the number or positions of subscribers and in exchange, no publishers details have to be configured [19, 20].

If a broker receives a letter on a subject on which no existing subscribers remain, the broker will discard the post unless it is a preserved message by the publisher. A retained message is a standard MQTT message with a flag that is retained. For the subject picked, the broker stores the last kept message and the respective QoS. Any user who subscribes to a subject trend resembling the subject of a retained message directly after signing up gets the retained message. This helps new users to access the latest benefit instead of waiting for the next update from a publisher. The broker only holds one retention message per theme [21].

Where a publishing client communicates with the broker for the first time, it will set a default message to the subscriber if the courier notices the unintended disconnection from the courier from the publishing client [22].

Clients communicate only with a broker, but a device may have multiple broker servers that share data depending on the topics of their established subscribers [23]. MQTT will only hold two bytes of data in a limited control message. If required, a control message can contain almost 256 megabytes of data. There are 14 defined message types used to connect a client from a broker and disconnect it to publish its data, accept data receipt, and monitor client-server connection [24].

For data transfer, MQTT depends on the TCP protocol. MQTT-SN version is used in other transportations, for example UDP and Bluetooth [25]. MQTT is a standard for transferring data from IoT to server. The de-facto IoT standard for connecting IoT devices in all sorts was originally developed in 1999 by Andy Standford-Clark and Arlen Nipper in the area of monitoring pipelines for oil and gas via satélite. Today the connectivity with MQTT is supported by all major IoT, IoT Cloud, IoT, gate and device IC.

MQTT is a lightweight publication/subscription protocol, which requires a minimum footprint and bandwidth for the IoT device connection. In contrast to the request/answer paradigm in HTTP, MQTT is event driven and allows customer messages to be pushed. This architecture unites the customers so that there is no dependency between data producers and data consumers for a highly scalable solution.

Table 2: Prominent MQTT Brokers for IoT Based Communications

URL	MQTT Applications
eclipse.org/paho	Paho MQTT
mqtt.fluux.io	FLUUX
hackage.haskell.org/package/net-mqtt	NET-MQTT
bevywise.com/mqtt-broker	Bevywise MQTTBroker
emqtt.io	EMQTTD
emqx.io	EMQ X
cloudmqtt.com	Cloud MQTT
github.com/eclipse/paho.mqtt.m2mqtt	M2MQTT
mosquitto.org	Mosquitto
hivemq.com	HiveMQ
rabbitmq.com	Rabbit MQ
thingstream.io	Thingstream
flespi.com	FLESPI
vernemq.com	VerneMQ
github.com/mcollina/mosca	Mosca
github.com/moquette-io/moquette	MOQUETTE
wolfssl.com/products/wolfmqtt	wolfMQTT

Relevant devices such as Raspberry Pi, Arduino, Notebooks, Server and other for real-world applications will instal these MQTT-broker applications [26].

Literature review

Nevertheless several works proposed solutions using different IoT protocols. An IoT Structures Security Survey was presented in a recent publication by Ammar et al. The benefits and weaknesses of different IoT Protocols are also discussed by

Fysarakis et al. today. The emphasis was on Thomann et al's 2015 encoding of MQTT. They have learned how to publish attribute-based encryption in their work. message patterns like MQTT. They contributed to a new approach to how ABE can be generalised and compared with several models already explored in combination with MQTT.

A comparative research on conventional AES operational methods was conducted by Almuhammadi et al. in 2017. This distinction was made for the time of encryption, decryption and the size of the data packets. They showed that the ECB mode is the fastest in other modes of operation. Shin et al. later introduced a security framework for MQTT.

The primary objective was to provide security for MQTT and to incorporate major over-the-counter access and computation through certificates to the proposed SSL/TLS solution. They proposed a fundamental architecture for MQTT security, i.e. AugMQTT does not require validity certificate inspections or revocation certificate inspections. Niruntasukrat et al. suggested that in addition to MQTT authentication, a mechanism be built to enable Federated Identity Management and Fine Graining Access Control. They then submitted an OAuth-based authorisation framework for specification and implementation.

In 2017, Mathur et al. studied a safe IoT strategy that has been proven to be secure against multiple attacks, including privacy abuse, denial of service, etc. Unlawful access to documentary documents.

Key Research Dimensions

• After clarifying the background details and related work it is important to remember the goals of this report.

• The primary goal is to make secure use of MQTT as an IoT protocol for the sensor and the cloud.
• The first entity in the cloud that collects IoT packets will have a low overhead connection between the sensor and the broker of the MQTT.
• A agreement between the two facets must be reached in order to assess possible alternatives.

MQTT Cloud Services

A variety of MQTT Cloud Broker services are offered in the MQTT brokers without the need of a physical MQTT broker as an Infrastructure on-demand device.

CloudMQTT

CloudMQTT is a widely used MQTT broker for IoT interfacing and collaboration with gadgets. CloudMQTT can be implemented in various situations such as Smart Toll Plaza, Smart Cities, Smart House and Smart Office Automations for IoT scenarios deployment.

The CloudMQTT contains the free Adorable Cat packet for researchers, which contains 5 users / connections and can connect ioT devices as shown in Figure 2. You will create the free instances via the current Google Account on the CloudMQTT network.

Here are Cloud MQTT Broker's core features.
• 24 / 7 MQTT Broker Availability
• Mosquitto cloud servers are operated by CloudMQTT
• Introduction of the MQ Telemetry Transport Protocol, MQTT
• Provides lightweight messaging methods by means of a message queuing model publish/subscribe

International Journal of Computing and Business Research (IJCBR)
ISSN (Online) : 2229-6166
International Manuscript ID : 22296166V12I4202203
Volume 12 Issue 4 September 2022

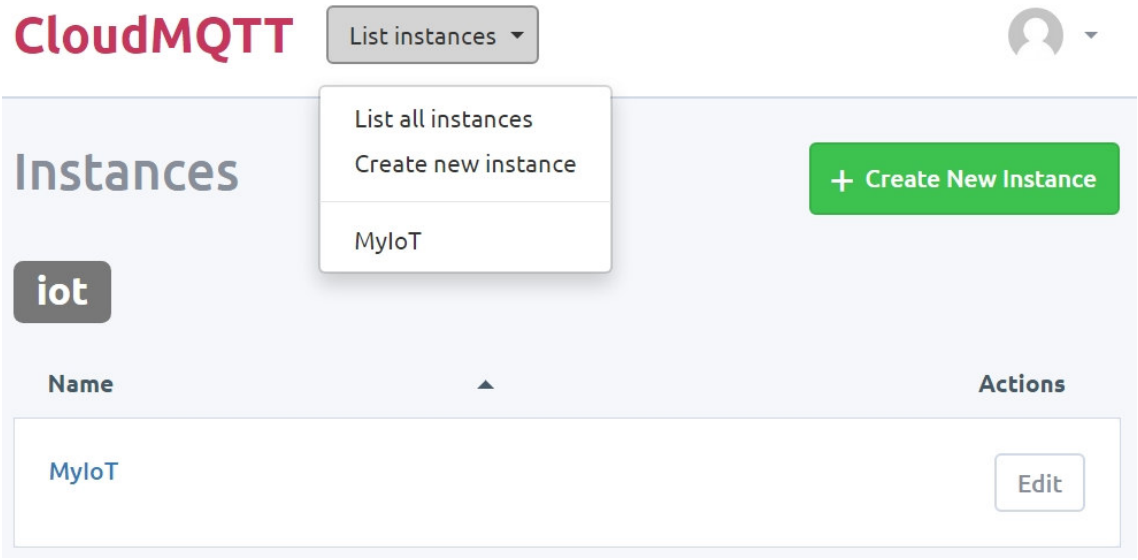

Figure 3: Creating Free Instance on CloudMQTT

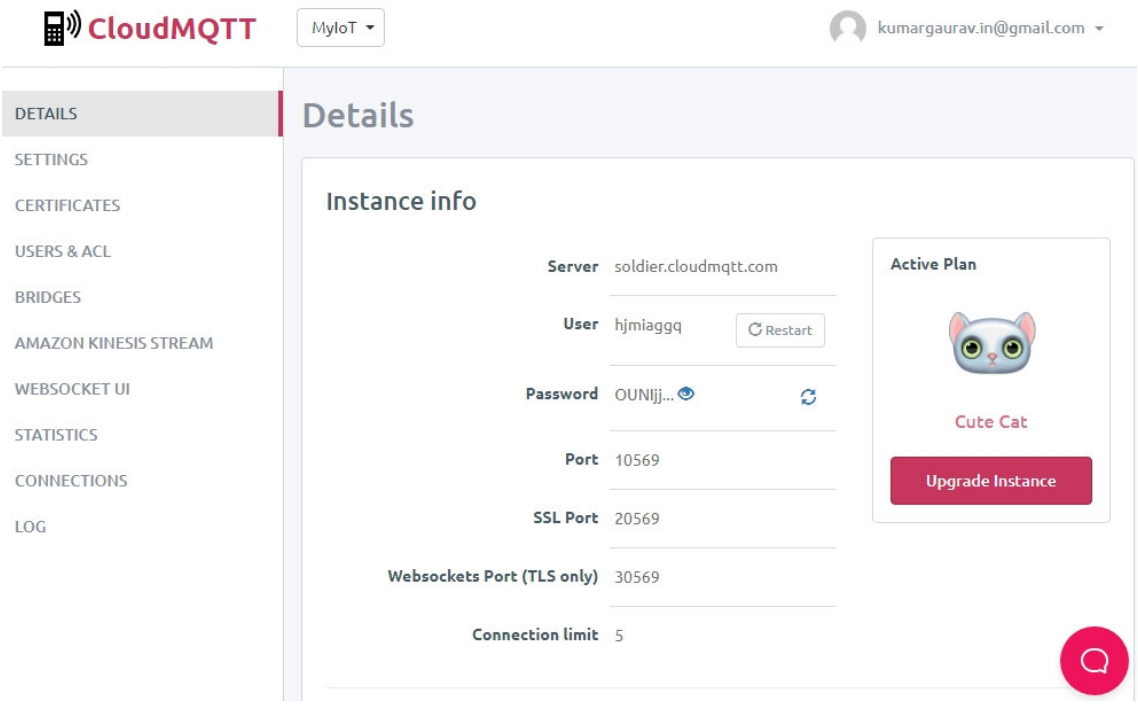

Figure 4: View Authentication in CloudMQTT for Connections

CloudMQTT authentication information can be utilised on IoT hardware or smartphones to create a real-time link between CloudMQTT and gadget, as illustrated in Figure 3.

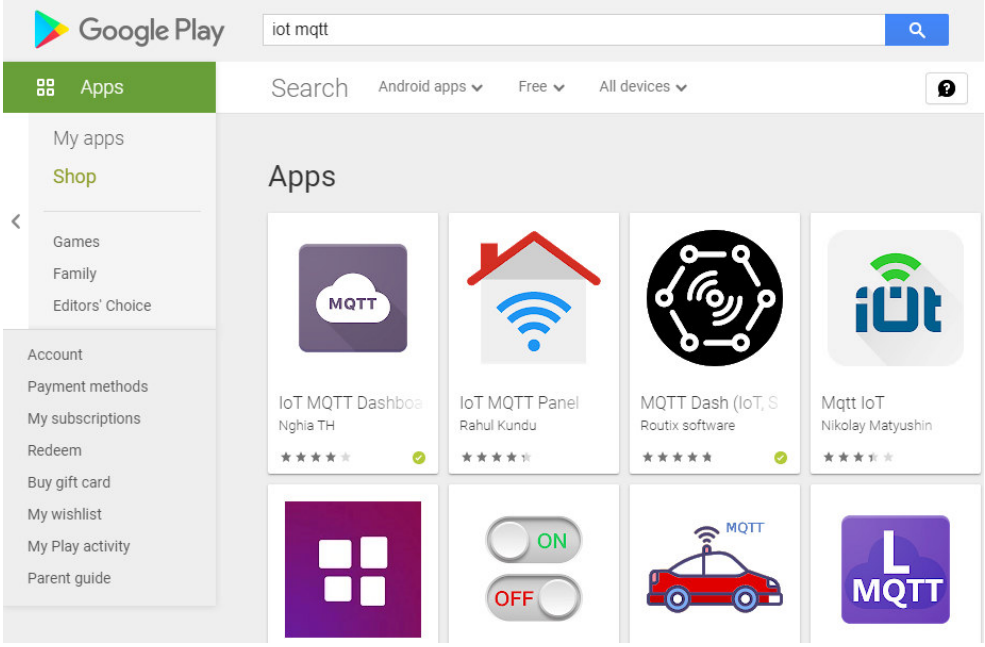

Figure 5: IoT MQTT Apps on Google Play Store

As shown in Figure 4, a number of Android apps are available from Google Play Store on the IoT MQTT Dashboard. These IoT MQTT applications interface directly with the MQTT Broker systems to enable the connectivity and interaction of IoT users, tablets and cloud brokers.

DIOTY

URL: http:/www.detectable.co

DIoTY is a free cloud-based MQTT broker programme for accessing IoT gadgets as seen in Figure 5. DIoTY provides the scripts to allow MQTT broker connections to a variety of programmes.

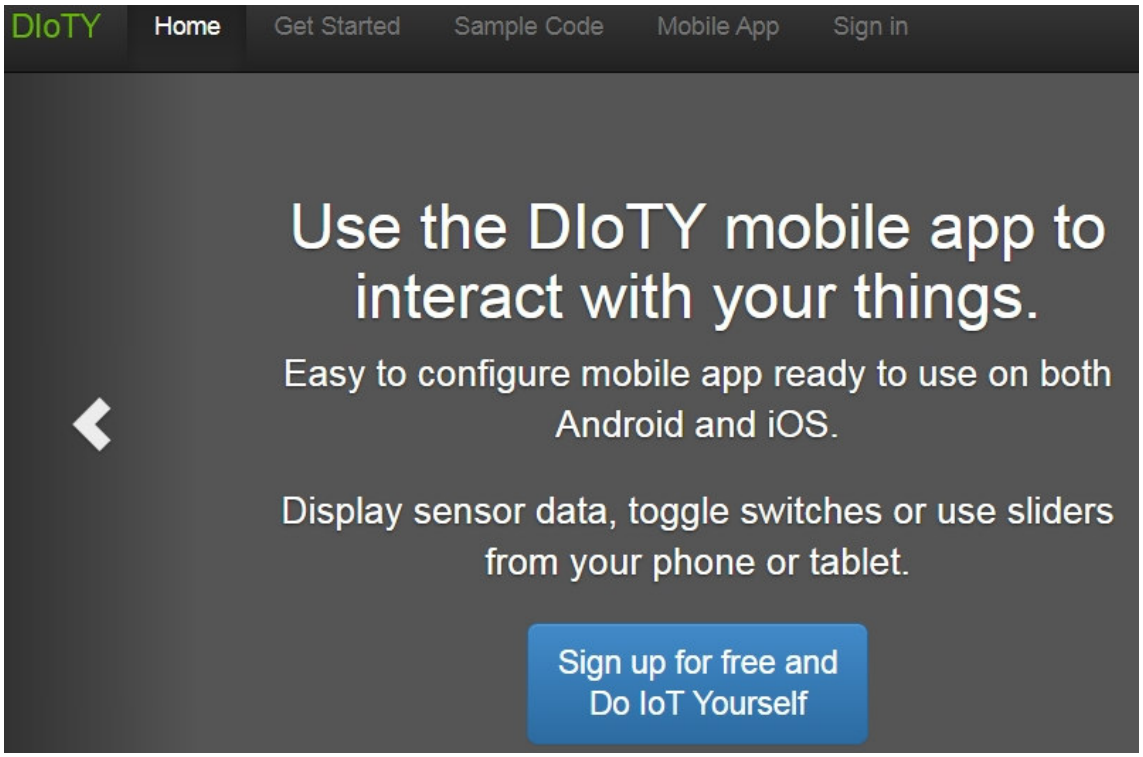

Figure 6: Dashboard of DIoTY

Following are the few of the powerful programming platforms that can be used to communicate the data and signals of the MQTT Broker with the Smart Gadget:

- C#
- Arduino
- Python
- NodeJS
- Go-long
- Java
- PHP
- MCU Node

Paho Client is one of Python's great libraries for computer interfacing and collaboration.

Then the snippet code can be implemented in computers and gadgets based on IoT:

```
# Define event callbacks
def on_connect(client, myRData, CN):
 if CN == 0:
 print("Connected successfully.")
 else:
 print("Connection failed. CN= "+str(CN))
def on_publish(client, myRData, MyDataSignal):
 print("Message "+str(MyDataSignal)+" published.")
def on_subscribe(client, myRData, MyDataSignal, granted_qos):
 print("Subscribe with MyDataSignal "+str(MyDataSignal)+" received.")
def on_message(client, myRData, msg):
 print("Message received on topic "+msg.topic+" with QoS "+str(msg.qos)+" and
payload "+msg.payload)
SecuredMQTTclient = mySECUREDMQTT.Client()
# Assign event callbacks
SecuredMQTTclients.on_connect = on_connect
SecuredMQTTclients.on_publish = on_publish
SecuredMQTTclients.on_subscribe = on_subscribe
SecuredMQTTclients.on_message = on_message
# Connect
SecuredMQTTclients.username_pw_set('mymail.in@gmail.com','7881730a')
SecuredMQTTclients.connect('mySECUREDMQTT.dioty.co', 1883)
```

```
# Subscribing
SecuredMQTTclients.subscribe('/mymail.in@gmail.com/')
# Publishing a message
x=input('Message')
SecuredMQTTclients.publish('/mymail.in@gmail.com/', x)
# Loop; exit on error
CN = 0
while CN == 0:
        CN = SecuredMQTTclients.loop()
        print("CN: " + str(CN))
```

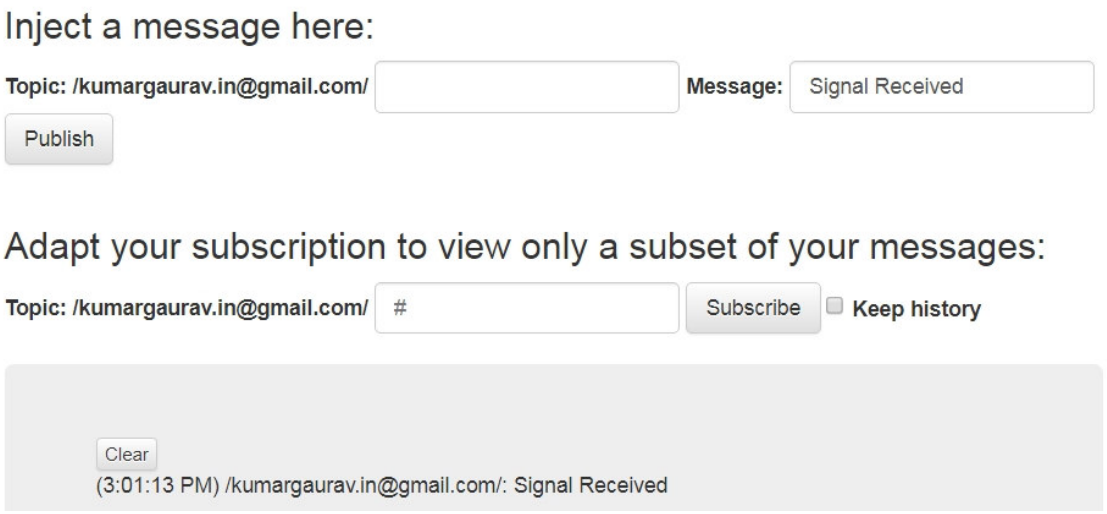

Figure 7: Receiving Data on DIoTY Dashboard using MQTT

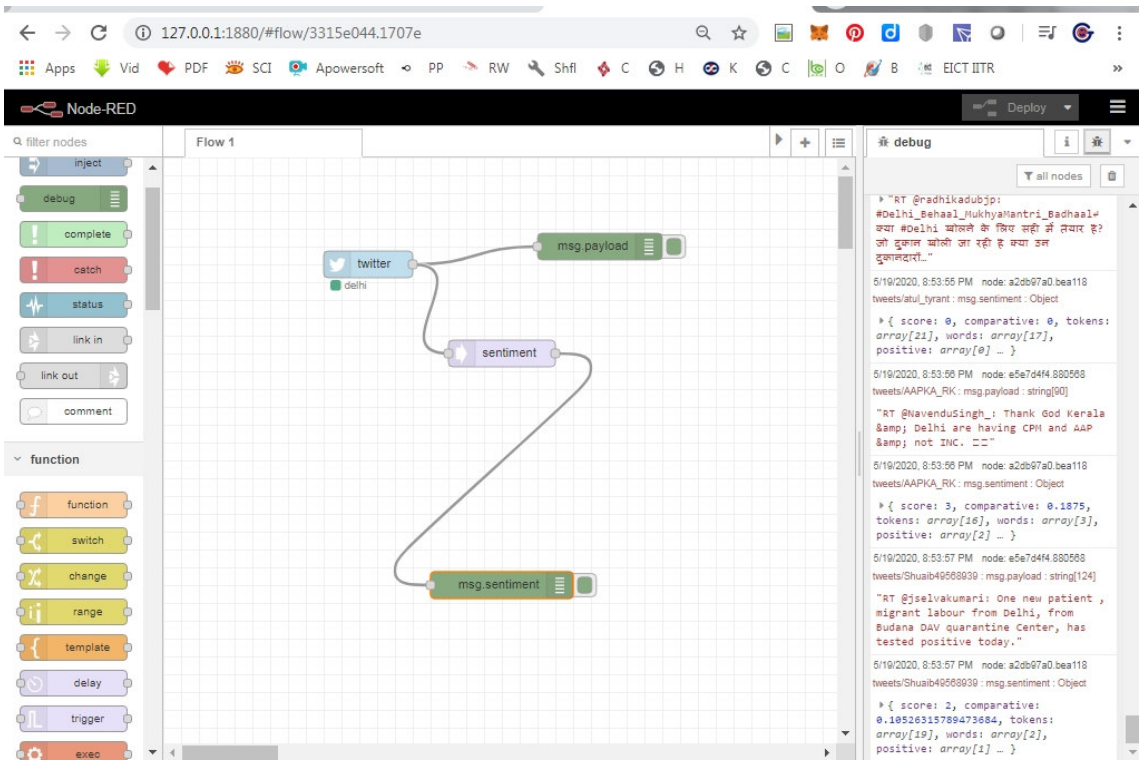

Figure 8: Fetching and Analytics of Real Time Data using MQTT Cloud

After executing the sentiment analytics patterns using MQTT broker, the data is handled and processed for the research implementations. With the implementation of these scripts and MQTT brokers, the real time data can be extracted and analyzed in real time.

Conclusion

The real time data extraction and analytics can be done effectively using MQTT broker platforms on cloud. There are a range of research formulations in data extraction and analytics using MQTT for assorted applications including telemedicine, personal safety gadgets, surveillance systems, military vehicles, smart agricultural agro-boats and several others can be applied. In many applications including crime data analysis, market analysis, citizen opinion, customer ratings, and

many more, web data extraction is needed in real-time. Market researchers and scientists can gather data for user behavioral research on relevant subjects or items from various portals for exploration of information.

References

[1] Dey, B., & Kundu, M. K. (2015). Efficient foreground extraction from HEVC compressed video for application to real-time analysis of surveillance 'big'data. IEEE Transactions on Image Processing, 24(11), 3574-3585.

[2] Elaggoune, Z., Maamri, R., & Boussebough, I. (2020). A fuzzy agent approach for smart data extraction in big data environments. Journal of King Saud University-Computer and Information Sciences, 32(4), 465-478.

[3] BanavathuMounika, S. K., Khadherbhi, R., Maddumala, V. R., & Lakshmi Patibandla, R. S. M. (2020). Data distribution method with text extraction from big data. Journal of Critical Reviews, 7(6), 376-380.

[4] Yu, J., & Couldry, N. (2020). Education as a domain of natural data extraction: analysing corporate discourse about educational tracking. Information, Communication & Society, 1-18.

[5] Semlali, B. E. B., El Amrani, C., & Ortiz, G. (2020). SAT-ETL-Integrator: an extract-transform-load software for satellite big data ingestion. Journal of Applied Remote Sensing, 14(1), 018501.

[6] Restuccia, F., & Melodia, T. (2019, April). Big data goes small: Real-time spectrum-driven embedded wireless networking through deep learning in the rf loop. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications (pp. 2152-2160). IEEE.

[7] Fenil, E., Manogaran, G., Vivekananda, G. N., Thanjaivadivel, T., Jeeva, S., & Ahilan, A. (2019). Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM. Computer Networks, 151, 191-200.

[8] Ed-daoudy, A., & Maalmi, K. (2019). A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment. Journal of Big Data, 6(1), 104.

[9] Zhang, X., & Ge, Z. (2019). Automatic Deep Extraction of Robust Dynamic Features for Industrial Big Data Modeling and Soft Sensor Application. IEEE Transactions on Industrial Informatics, 16(7), 4456-4467.

[10] Yassein, M. B., Shatnawi, M. Q., Aljwarneh, S., & Al-Hatmi, R. (2017, May). Internet of Things: Survey and open issues of MQTT protocol. In 2017 International Conference on Engineering & MIS (ICEMIS) (pp. 1-6). IEEE.

[11] Soni, D., & Makwana, A. (2017, April). A survey on mqtt: a protocol of internet of things (iot). In International Conference On Telecommunication, Power Analysis And Computing Techniques (ICTPACT-2017).

[12] Niruntasukrat, A., Issariyapat, C., Pongpaibool, P., Meesublak, K., Aiumsupucgul, P., & Panya, A. (2016, May). Authorization mechanism for MQTT-based Internet of Things. In 2016 IEEE International Conference on Communications Workshops (ICC) (pp. 290-295). IEEE.

[13] Kashyap, M., Sharma, V., & Gupta, N. (2018). Taking MQTT and NodeMcu to IOT: Communication in Internet of Things. Procedia Computer Science, 132, 1611-1618.

[14] Dhar, P., & Gupta, P. (2016, September). Intelligent parking Cloud services based on IoT using MQTT protocol. In 2016 International conference on automatic control and dynamic optimization techniques (ICACDOT) (pp. 30-34). IEEE.

[15] Hou, L., Zhao, S., Li, X., Chatzimisios, P., & Zheng, K. (2017). Design and implementation of application programming interface for Internet of things cloud. International Journal of Network Management, 27(3), e1936.

[16] Rocha, M. S., Sestito, G. S., Dias, A. L., Turcato, A. C., Brandão, D., & Ferrari, P. (2019). On the performance of OPC UA and MQTT for data

exchange between industrial plants and cloud servers. Acta IMEKO, 8(2), 80-87.

[17] Raikar, M. M., Desai, P., Kanthi, N., & Bawoor, S. (2018, September). Blend of Cloud and Internet of Things (IoT) in agriculture sector using lightweight protocol. In 2018 international conference on advances in computing, communications and informatics (ICACCI) (pp. 185-190). IEEE.

[18] Iskandar, H. R., Hermadani, H., Saputra, D. I., & Yuliana, H. (2019). Eksperimental Uji Kekeruhan Air Berbasis Internet of Things Menggunakan Sensor DFRobot SEN0189 dan MQTT Cloud Server. Prosiding Semnastek.

[19] Al-Joboury, I. M., & Al-Hemiary, E. H. (2018). Performance analysis of internet of things protocols based fog/cloud over high traffic. Journal of Fundamental and Applied Sciences, 10(6S), 176-181.

[20] Firdous, S. N., Baig, Z., Valli, C., & Ibrahim, A. (2017, June). Modelling and evaluation of malicious attacks against the iot mqtt protocol. In 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (pp. 748-755). IEEE.

[21] M. Ammar, G. Russello, and B. Crispo. “Internet of Things: A survey on the security of IoT frameworks”. In: Journal of Information Security and Applications 38 (2018), pp. 8–27.

[22] D. Thatmann et al. “Applying Attribute-based Encryption on Publish Subscribe Messaging Patterns for the Internet of Things”. In: IEEE International Conference on Data Science and Data Intensive Systems (2015).

[23] S. Almuhammadi and I. Al-Hejri. “A comparative Analysis of ES Common Modes of Operation”. In: IEEE 30th Canadian Conference on Electrical and Computer Engineering(CCECE) (2017).

[24] Vern Paxson. “Introduction to Communication Networks”. In: https://www.openssl.org/ 17 (2011), pp. 4711–4721.

[25] A. Niruntasukrat et al. “Authorization Mechanism for MQTTbased Internet of Things”. In: IEEE ICC2016-Workshops: W07 Workshop on Convergent Internet of Things (2016).

[26] E. Walid R. Muzaffar D. Gerard M. Avijit N. Thomas and T. Daniel. “A secure end-to-end IoT solution”. In: Sensors and Actuators A: Physical 263 (2017), pp. 291–299.