International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Realtime Data Traffic Analyser Locomotive of Big Data Analytics

Fatma Asad Al Jarah, Mazhar Hussain Malik

Abstract*: *Since last decade, the exponential growth of the internet users and the size of data over the internet is increasing day by day, which lead to increase the complexity of the systems by implementing policies and security to avoid attacks on systems and networks. It is very important to understand and analyses the real time data traffic of the communication systems.

The purpose of this paper to design a customized Java based application which enables analysts to capture the traffic at the bottleneck under the mean field communication environment where a large number of devices are communicating with each other. The sending data for further processing for analysis the trend to overcome vulnerabilities or to manage the effectiveness of the communication systems. The proposed application enables to capture 8 different types of protocol traffic such as HTTP, HTTPS, SMTP, UDP, TCP, ICMP and POP3. The application allows for analysis of the incoming/outgoing traffic in the visual to understand the nature of communication networks which lead to improve the performance of the networks with respect to hardware, software, data storage, security and reliability.

Keywords: Real-time Data traffic, Mean Field, UDP, TCP, Big Data.

I. INTRODUCTION

As size of data is also increasing with the passage of time, this happened as networks moved from traditional wire network to wireless and sensor-based network designing which are suitable to different type of field such as health sector, educational sector, organization and all other type of organizations which are interested to share data between the different branches or interest to connect with internet. Moreover, since the induction of sensors to capture some reading such as IoT based setups are causing challenges for network design and increase the size of the data. These factors are contributing to increase size of the data and it is very important to understand the data which enable to use it for different purposes e.g. planning for resources, predictions and categorizations of data.

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author

Fatma Asad Al Jarah*, Department of Computing, Global College of Engineering and Technology, Muscat, Oman. Email: f.aljarah@gcet.edu.om.

Mazhar Hussain Malik, Department of Computing, Global College of Engineering and Technology, Muscat, Oman.

The Real time data traffic is normally associated with enormous networks which cause to generate large amount of data. Big Data structures are characterized with four main dimensions or four V's [1, 5]:

1. Volume: Data volume is the most characteristic feature of big data. When any organization creating big data, it cannot be enrolled in the traditional way.
2. Velocity: Rapid data analysis to assist organizations in making their decisions and from which they can gather data.
3. Variety: Generated data, which come from different sources, have different type. There must be appropriate technology to handle this data.
4. Veracity: The accuracy and correctness of the data to be analysed and make decision.

The term "real- time" indicates that the accessibility of data is happening as it comes. Analysing the traffic generated from those data is very essential to understand the different aspect such as hardware, software, data storage, security and reliability.

The purpose of this paper is to provide a platform which can read real time data, identify them and build a visual analysis for the end user. The proposed application will also store the historical analysis in data warehouse for future analysis.

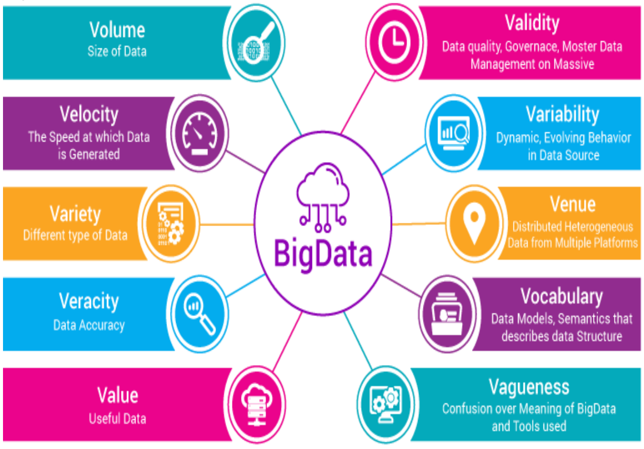

Volume
Size of Data
Validity
Data quality, Governace, Moster Data Management on Massive
Velocity
The Speed at which Data is Generated
Variability
Dynamic, Evolving Behavior in Data Source
Variety
Different type of Data
BigData
Venue
Distributed Heterogeneous Data from Multiple Platforms
Veracity
Data Accuracy
Vocabulary
Data Models, Semantics that describes data Structure
Value
Useful Data
Vagueness
Confusion over Meaning of BigData and Tools used

Figure 1: Big Data Structure [5]

II. RELATED WORK

We assessed large number of papers from some renowned journals and some of these papers are mentioned in the survey. In research paper [2] authors provide a detail analysis on the importance of big data traffic analysis, main focus was on the urban traffic congestion by addressing the issues which people communicate for their job to urban areas from their homes, this cause traffic congestion.

This paper investigates some recent technologies which are used for cater such issues. The authors investigated the MapReduce, Intelligent transfer systems and algorithms.

The authors were able to provide the different type of data structures.

Sharif et al [3] discuss the impact of Internet of Things (IoT) to include the smart traffic system (STS) along with the identification of implementation issues and challenges. The study suggests that there are challenges while capturing Realtime traffic and suggest that sensors can be deployed to capture traffic information after a specific interval such as 500 meters.

Sidorov et al [4] proposed an expert system with real time analysis for large spatial set of air traffic data. The study used a tool to analyse real time exploratory data of air traffic through use many algorithms and spatial data analysis. Consequently, will able to obtain results fast that are ideal for expletory and visual analysis of a wide range of visible data.

Based on the mentioned latest literature, it is worth to design a customised application which can capture the traffic on bottleneck and instantly visualize for further treatment and investigation.

III. PROPOSED METHODOLOGY

The overall packet capturing process of Real time traffic analyser is shown in the Figure.1. The user needs proper authentication to access the application and capturing interface will be selected and then user will choice to select the type of traffic which need to capture the application allow to select a specific type of protocol to all available protocols, in our case the application allows to capture 8 different type of protocol traffic.

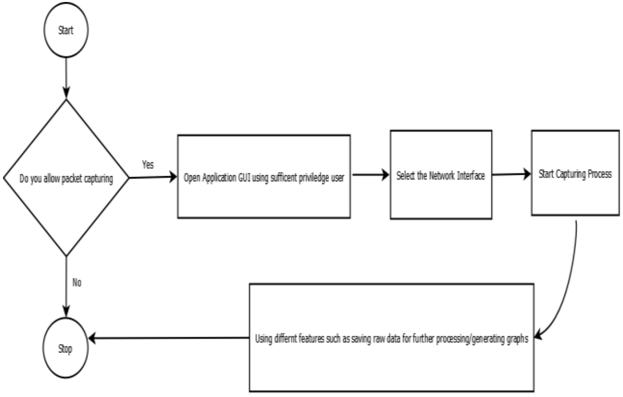

Start
Do you allow packet capturing
Yes
Open Application GUI using sufficent priviledge user
Select the Network Interface
Start Capturing Process
No
Stop
Using different features such as saving raw data for further processing/generating graphs

Figure 1: Real time packet capturing process

IV. DESIGN

The study is conducted by assumption that the size of the network is very large. There is a gateway denoted with "*G*" which is connected with large number of routers/access points (R_n). As per Figure.3, " R_1 , R_2 , R_3, ... R_n " representing the routers which are connect with different nodes " N_1 , N_2 , N_3, ... N_n " (wire/wireless) running different type of traffic such as TCP, UDP.

The gateway "*G*" is the backbone of the scenario which have polices and firewall to enable and disable incoming or outgoing data traffic.

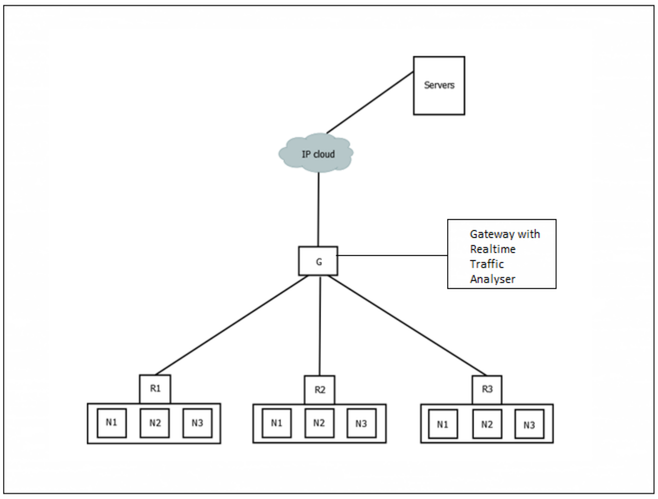


Figure.2 Overall Testing Scenario

The proposed application is tested in a network of 250+ devices in a college in Muscat (Global College of Engineering and Technology), the designed software was running at the bottleneck and was able to capture 814088 packets.

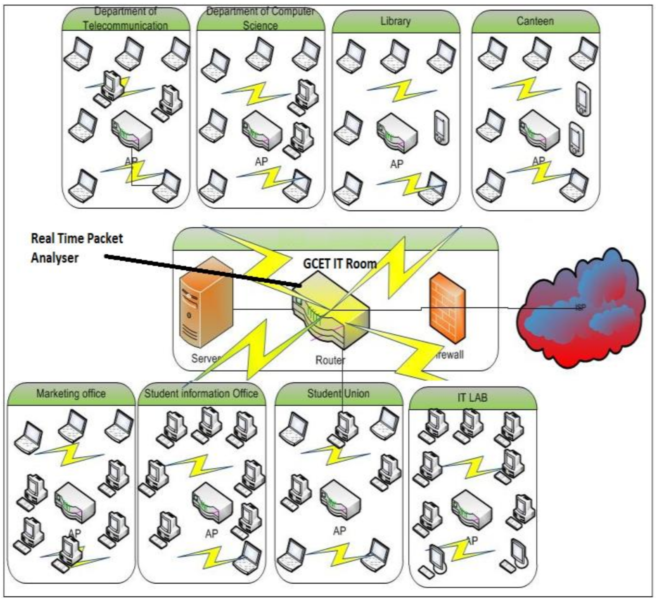

Department of Telecommunication
Department of Computer Science
Library
Canteen
AP
AP
AP
AP
Real Time Packet Analyser
GCET IT Room
Server
Router
irewall
ISP
Marketing office
Student information Office
Student Union
IT LAB
AP
AP
AP
AP

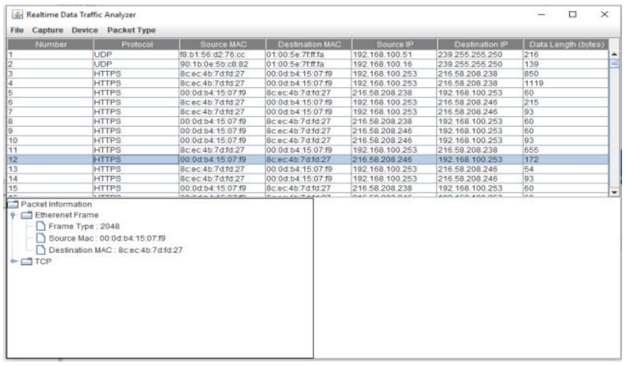
Figure.3 Overall Network Environment

The test-based environment as shown as Figure.3 consists of about 250+ devices of different type such as laptops, PC's, Mobile Phones etc. Both wire and wireless interfaces were captured at the bottleneck.

V. IMPLEMENTATION

The platform is designed in the Java Net beans by using WINPCAP to capture traffic from the interface. The Figure.3 shows the platform interface. The platform enable user to select the packet types and interfaces to capture the packets. As shown in Figure.4 the tool enable user to get information about the protocol source and destination addresses (IP's and MAC) along with Data Length in bytes.

Moreover application allows to see the detail packet information which includes frame type along with addresses and for each TCP packet the application allow to get information about the source and destination ports, sequence number, acknowledge number with flags such as Urgent Pointers (URG), Push (PSH) and Reset (RST).

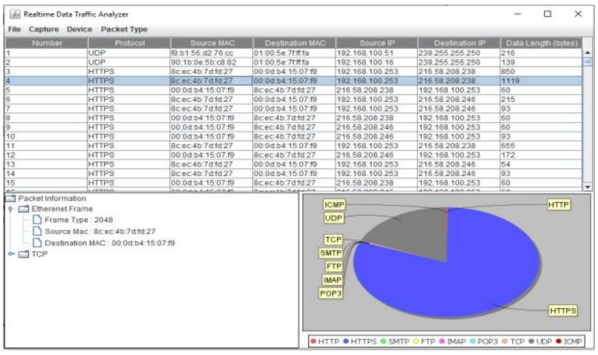


Number	Protocol	Source MAC	Destination MAC	Source IP	Destination IP	Data Length (bytes)
1	UDP	f8:b1:56:d2:76:cc	01:00:5e:7f:ff:fa	192.168.100.51	239.255.255.250	216
2	UDP	90:1b:0e:5b:c8:82	01:00:5e:7f:ff:fa	192.168.100.16	239.255.255.250	139
3	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.238	850
4	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.238	1119
5	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.238	192.168.100.253	60
6	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	215
7	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	93
8	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.238	192.168.100.253	60
9	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.246	192.168.100.253	60
10	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.246	192.168.100.253	93
11	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.238	655
12	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.246	192.168.100.253	172
13	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	54
14	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	93
15	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.238	192.168.100.253	60

Packet Information
Etherenet Frame
Frame Type : 2048
Source Mac : 00:0d:b4:15:07:f9
Destination MAC : 8c:ec:4b:7d:fd:27
TCP

Figure 4: Real time traffic Analyser

After the completion of Analysis, the analyst can produce different type of graphs such as bar chart and pie chart as shown in Figure.5.


Realtime Data Traffic Analyzer
File
Capture
Device
Packet Type
ICMP
UDP
TCP
SMTP
FTP
IMAP
POP3
HTTP
HTTPS
● HTTP ● HTTPS ● SMTP ● FTP ● IMAP ● POP3 ● TCP ● UDP ● ICMP

Number	Protocol	Source MAC	Destination MAC	Source IP	Destination IP	Data Length (bytes)
1	UDP	f8:b1:56:d2:76:cc	01:00:5e:7f:ff:fa	192.168.100.51	239.255.255.250	216
2	UDP	90:1b:0e:5b:c8:82	01:00:5e:7f:ff:fa	192.168.100.16	239.255.255.250	139
3	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.238	850
4	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.238	1119
5	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.238	192.168.100.253	60
6	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	215
7	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	93
8	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.238	192.168.100.253	60
9	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.246	192.168.100.253	60
10	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.246	192.168.100.253	93
11	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.238	655
12	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.246	192.168.100.253	172
13	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	54
14	HTTPS	8c:ec:4b:7d:fd:27	00:0d:b4:15:07:f9	192.168.100.253	216.58.208.246	93
15	HTTPS	00:0d:b4:15:07:f9	8c:ec:4b:7d:fd:27	216.58.208.238	192.168.100.253	60

Packet Information
Etherenet Frame
Frame Type : 2048
Source Mac : 8c:ec:4b:7d:fd:27
Destination MAC : 00:0d:b4:15:07:f9
TCP

Figure 5: Real time traffic Analyser Result GUI

VI. RESULT AND DISCUSSION

Figures.6-8 showing the sum and average of traffic at the bottleneck. Results shows that majority of traffic is generated by the HTTPS which contributes to 91% of total traffic. Overall data size of the HTTP is 392396822 bytes with average size of 722.5 bytes per packet, while about 9% traffic is generated by the UDP protocol, overall 40843215 bytes and average size of 154 bytes per packet. ICMP was producing total data of 1020 bytes with average size of 92.7 per packet.

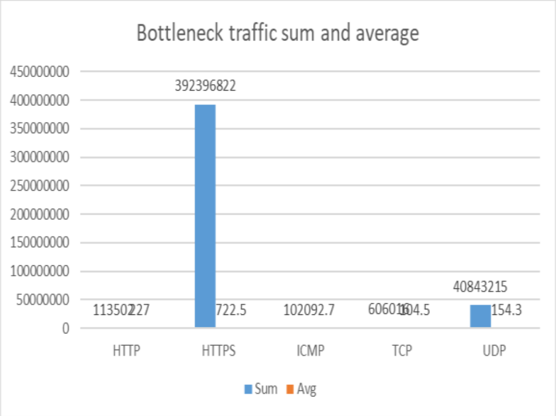


Figure 6: Bottleneck traffic sum and average

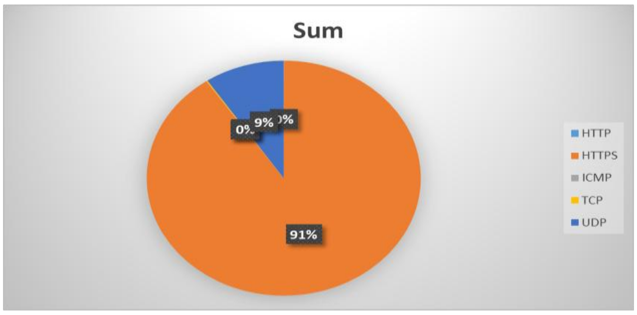


Figure 7: Bottleneck traffic Sum (bytes)

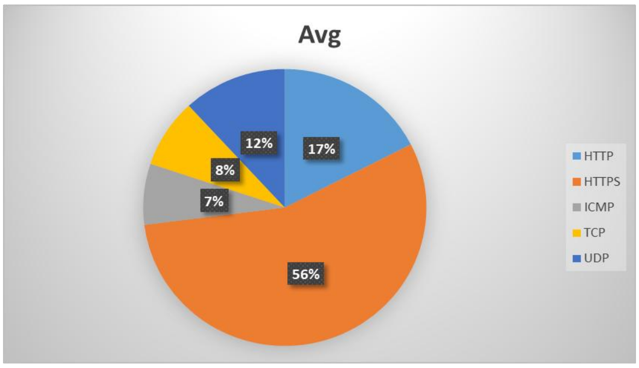


Figure 8: Bottleneck traffic Average Packet Size (bytes)

Figures.9-10 shows the number of packets for each protocol and trend of the protocol traffic. Figure.9 showing that HTTPS generating 543094 packets, which is contributing 67% of total generated packets. In the case of ICMP only 11 packets were generated, TCP was generating 5794 packets which is 1% of total generated traffic, HTTP was generating 499 packets which is less than 1% and UDP is generating 264690 packets which contributes to 32% of total packet generated, the traffic trend in Figure.10.

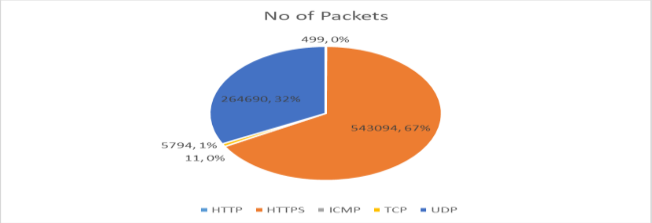


Figure 9: Protocols with No of Packets (incoming/outgoing)

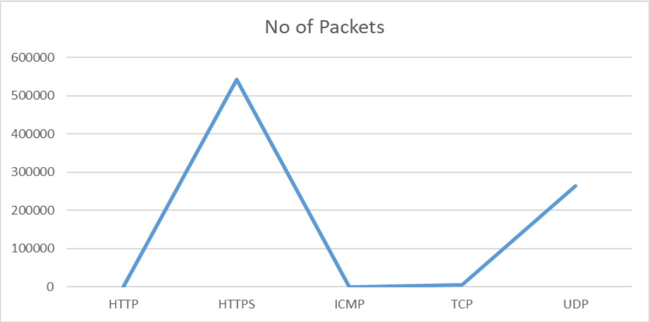


Figure 10: Protocols with No of Packets (incoming/outgoing) trend

Retrieval Number: F9809038620/2020©BEIESP
DOI:10.35940/ijrte.F9809.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VII. CONCLUSION

The study shows that majority of traffic which is generated from the studied scenario is belonging to HTTPS and UDP is second highest in the traffic category. It is also worth to mention that TCP traffic ratio is less than 1% which is quite low. The overall findings show that majority of WebPages which are running in the network are using HTTPS and UDP traffic (i.e. Audio/Video or gamming traffic) contributing as second largest incoming/outgoing traffic. The study justify that designed application allows decision makers to identify the requirements in efficient way and align required resources efficiently.

The application can be further extended by taking capture data to the any data analysis tool such as WEKA/SAS for further processing.

REFERENCES

1. He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data. Industrial management & data systems, 115(9), 1622.
2. Vidya, K. T., & Ashwini, B. P. (2018, May). Survey on Real Time Traffic Analysis on Big Data. In 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 852-858). IEEE.
3. Sharif, A., Li, J., Khalil, M., Kumar, R., Sharif, M. I., & Sharif, A. (2017, December). Internet of things—smart traffic management system for smart cities using big data analytics. In 2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) (pp. 281-284). IEEE.
4. Sidorov, V., Ng, W. K., & Salleh, M. F. B. M. (2018, December). A Practical Expert System with (Near) Real-Time Analysis of Large Spatial Sets of Air Traffic Data. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 3478-3484). IEEE.
5. Rumsfeld, J. S., Joynt, K. E., & Maddox, T. M. (2016). Big data analytics to improve cardiovascular care: promise and challenges. Nature Reviews Cardiology, 13(6), 350.

AUTHORS PROFILE

Fatma Al Jarah is final year student of B.Eng (hons) Software Engineering at Global College of Engineering and Technology, Muscat Oman. Currently, she is working on data mining techniques to analysis real time network traffic behavior.

Mazhar H. Malik has PhD in Computer Science with specialization in Computer Networks and since September 2017, working as Head of Department/Senior Lecturer in Global College of Engineering and Technology, Muscat, Oman. Prior to this post, he worked in academic and different industrial positions for various universities and companies including Institute of Southern Punjab, Multan Pakistan, SBE electronics, UK, NCR Corporation, USA, The University of Lahore, Islamabad, Pakistan.He is editorial board member of peer-reviewed journals and reviewer of IEEE Access, International Journal of Advanced Computer Science and Applications (IJACSA), International Journal of Computer Science and Information Security (IJCSIS). His research interest includes, Data Science, Wireless Communications, and Cloud Computing, Sensors networks, Network security and forensics, Artificial intelligence and Machine Learning.