

# BIG DATA ANALYTICS IN MOBILE CELLULAR NETWORKS

<sup>1</sup>Anup Raut, <sup>2</sup>Abhijeet Ghorpade, <sup>3</sup>Ketan Jadhav, <sup>4</sup>Ajit Ghanekar

<sup>1</sup> Professor, Department of Information Technology, Parvatibai Genba Moze College of Engineering Wagholi, Pune  
<sup>2,3,4</sup> BE, Department of Information Technology, Parvatibai Genba Moze College of Engineering Wagholi, Pune

**Abstract:** In Big data analytic system use in the mobile cellular network. For transfer the big data over the network. Mobile cellular systems have turned out to be both the generators and carries of massive data. Big data analytic can enhance the execution of versatile cell systems and furthermore boost the income of administrators. In this paper, we present a brought together data show in light of the irregular grid hypothesis and machine learning. At that point, we exhibit an engineering structure for applying the big data investigation in the portable cell systems. Also, we portray a few illustrative cases, including big signaling data, big traffic data, big location data, big radio waveform data, and big heterogeneous data, in versatile cell systems. In this cellular network classified the data and, send the data from sender to receiver. In this paper we utilize hadoop concept. In that utilization the mapper and reducer concepts are used.

**Keywords:** Big Data Analysis, Mobile Cellular Network, hadoop, mapper, reducer.

## I. INTRODUCTION

In big data analytic system sender send the data to the recipient over the system. At the time of data sending, data get categorized like in signal, location, traffic and waveform [1]. Big Data is an expression used to mean a gigantic volume of both organized and unstructured data that is so big it is hard to process utilizing customary database and programming procedures [2]. In most endeavor situations the volume of data is too enormous or it moves too quickly or it surpasses current handling limit. Big Data can possibly help organizations enhance operations and make speedier, cleverer choices. This data, when caught, organized, controlled, put away, and broke down can help an organization to increase valuable knowledge to build incomes, get or hold clients, and enhance operations[4]. Enormous data is a term for detail indexes that are so expansive or complex that customary data handling applications are deficient to manage them. Challenges incorporate analytic, capture data, data curation, search, sharing, stockpiling, exchange, representation, questioning, and refreshing and data security. The expression "Big data" regularly alludes essentially to

the utilization of prescient investigation, client conduct examination, or certain other propelled data investigation strategies that concentrate an incentive from data, and rarely to a specific size of detail collection. Researchers, business officials, specialists of solution, promoting and governments alike consistently meet troubles with big detail indexes in regions including Internet look, back, urban informatics, and business informatics. Researchers experience restrictions in e-Science work, including meteorology, genomics, complex material science recreations, science and ecological research [5].

Big data is the accumulation of data sets so extensive and complex that it winds up noticeably hard to process utilizing customary data preparing applications [4]. With late advances of remote innovations and expanding in the portable applications, versatile cell systems have turned out to be both generators and transporters of big data. Generally data examination manages the organized data that is data contained in social databases and spreadsheets. Enormous data investigation is fit for gathering the scattered data, for understanding the client practices from the numerous points of view [3]. It incorporate the endorsers' living propensities and the timetable can be for the most part deduced from the utilization of movement cover distinctive eras of a day, surfing propensities and their much of the time went by spots or the scope of exercises can be around gotten from home area enlist [6]. Enormous data investigation, administrators can screen their framework progressively, and settle on self-ruling and element choices. Mobile Service Providers (MSPs) are procedures enormous measures of client created call records every day. Breaking down this big data can help in settling the absolute most basic issues in MSPs. With the dangerous development of big data and high level of activity in the portable cell systems are taken care of by the hadoop structure and Map Reduce programming model can be proposed and give the security to high movement data[7]. For investigating and limiting the system activity, an expansive scale structure in light of hadoop is being utilized [3].

## II. PROPOSED SYSTEM WORK

In our proposed system there is a sender mobile from which we exchange the data [6]. This information exchange over the network, subsequently the information get arranged in signal, traffic, waveform, location data in cellular network. We are utilizing the hadoop concept for exchanging the information. In that, the extensive information goes to the mapper [1]. Mapper is use to change over these enormous information into little parts and afterward prepared on that. At that point this information is exchange to the reducer. Reducer joins this information and gives result [4].

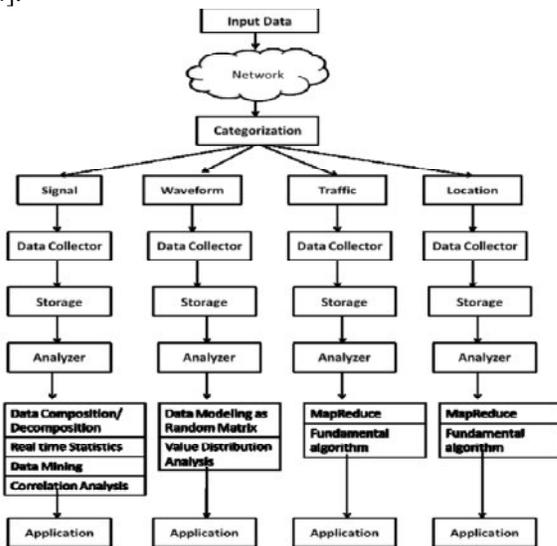


Figure 1: Big Data Analytic Using Mobile Cellular Network Architecture

In the proposed framework, we take an information document from the mobile or the system as info and pass it to the network [8]. Here, the data information will be classified into various parts to be signal log, location log, waveform, traffic log. After this classification, there is an information authority which gathers the information and stores them. Analyzer is utilized to break down the information [5]. The flag information analyzer checks the information arrangement and decay, continuous measurements, information mining and connection systematic. The waveform information analyzer checks the information displaying as arbitrary network and esteem appropriation investigative [7]. The movement information analyzer plays out the Map Reduce capacity and Fundamental calculation. The area information analyzer likewise plays out the Map

Reduce capacity and Fundamental calculation [10]. In the wake of playing out all these handling, the information is then sent to the application. Thus information is exchange from legitimate sender to beneficiary [12].

## III. RELATED WORK

Suzhi Bi, Rui Zhang, Zhi Ding, and Shuguang Cui proposed the Wireless Communications in the Era of Big Data [1]. The quickly developing flood of remote information administration is pushing against the limit of our correspondence system's preparing power. The inescapable and exponentially expanding information movement introduces inevitable difficulties to every one of the parts of the remote framework outline, for example, range proficiency, processing abilities and front haul/backhaul connect limit. In this article, they examined the difficulties and openings in the outline of adaptable remote frameworks to grasp such a "big data" period. On one hand, they audit the cutting edge organizing structures and flag handling methods versatile for dealing with the big data movement in remote systems.

Manhal Abdel Kader, Ejder Bastug, Mehdi Bennis, Engin Zeydan, Alper Karatepe, Ahmet Salih Er, Mérouane Debbah propose the Leveraging Big Data Analytics for Cache-Enabled Wireless Networks[2]. In this article, they approached this issue from a proactive reserving point of view where increases of store empowered base stations in 5G remote are contemplated. Specifically, colossal measure of genuine information from a telecom administrator in Turkey is gathered/ handled on a major information stage, and an examination is done for substance fame estimation for reserving, intending to enhance clients' involvement as far as demand fulfillments and offloading the backhaul. Along these lines, with this versatile movement information gathered from many base stations inside a few hours of time interim and the estimation of substance prevalence through machine learning instruments, we examine the additions of proactive reserving by means of numerical reproductions. The outcomes demonstrate that proactive reserving satisfies 100% of client demand fulfillment and offloads 98% of the backhaul, in a setting of 16 base stations with 15.4 Gbyte of capacity size (87% of the aggregate index size) and 10% of substance appraisals. File Terms— proactive storing, content notoriety estimation, huge information, machine learning, 5G cell systems.

He, Fei Richard Yu, Nan Zhao, Hongxi Yin, Haipeng Yao, Robert C. Qiu propose the Big Data Analytics in

Mobile Cellular Networks [3]. The term Big Data goes well past the information itself; it is additionally frequently used to allude to another procedure to approach our issues and arrangements. As pointed out in [2], our logical advances fall in various stages, or ideal models, as humankind pushes ahead. The primary worldview is known as the exact stage, which happened when logical revelation was for the most part determined by recording experimental perceptions through apparatuses, for example, telescopes. The second stage was when hypotheses were acquainted with outline the perceptions and make forecasts. Researchers, for example, Newton utilized arithmetic and physical laws to manufacture models to clarify the exact perceptions. The third worldview came accordingly of the entry of computerized PCs, when huge scale reproductions were utilized to emulate the progression of nature. With the landing of the Big Data, we are toward the start of the fourth worldview of logical disclosure, when learning revelation is done through theory testing driven by the accessibility of the huge advanced information. In this fourth-worldview method for logical considering, information turns into a top of the line subject, bringing forth the specific routine of learning revelation known as Data Science.

Navya H Ajjar, Nagalambika Swami proposes the Big Data Issues in Mobile Networking what's more, Computing Data. Enormous information is the term for informational collections so huge and confounded that it ends up noticeably hard to process utilizing customary information administration apparatuses or handling applications [4]. Portable systems administration is turning into a more vital partner of customary Internet and huge information. The portable systems administration is getting to be noticeably bigger and bigger because of discharging of a huge number of PDAs and cushions.

Mohammad Abu Alsheikh, Dusit Niyato, Shaowei Lin, Hwee-Pink Tan, and Zhu Han proposed the Mobile Big Data Analytics Using Deep Learning and Apache Spark. — The expansion of cell phones, for example, cell phones and Internet of Things (IoT) devices, brings about the current portable huge information (MBD) time[5]. Gathering MBD is unbeneficial unless appropriate examination and learning techniques are used for extricating important data and concealed examples from information. This article exhibits a diagram and brief instructional exercise of profound learning in MBD investigation and talks about an adaptable learning structure over Apache Spark. In particular, a circulated profound learning is executed as

an iterative Map Reduce processing on many Spark laborers.

#### **IV. CASE STUDIES OF BIG DATA ANALYTICS IN MOBILE CELLULAR NETWORKS**

##### ***A) BIG SIGNALING DATA***

In mobile cellular network, the transmission of voice and information is joined by control messages, which are named as flagging [3]. The flagging works as per the predefined conventions and guarantee the correspondence's security, unwavering quality, consistency and proficiency. Flagging observing assumes a vital part in proper designation of system assets, enhancing the nature of system administrations [9], ongoing recognizing system issues, and so on. With the fast improvement of different versatile cell arranges, the volume of flagging information develops enormously and the customary flagging observing frameworks have excessively numerous issues to manage [10].

##### ***B) BIG TRAFFIC DATA***

With the broad use of versatile Internet, the volume of activity information increments at a phenomenal rate. Going about as a bearer of the activity information, cell administrators need to deal with the system asset properly to adjust arrange stack and advance system use [11]. Activity checking and breaking down is a rudimentary yet basic part for system administration, empowering execution examination and forecast, disappointment location, security administration, and so forth [13]. Conventional ways to deal with screen and break down the movement information appear to be, be that as it may, direct and lacking with regards to enormous activity information, as represented in Fig. 3. In [1], the interrelationship between enormous information and software-characterized organizing (SDN) [12] has been contemplated transportation framework developments, statistic patterns, hazard notices for crowed individuals, fast rise reactions, wrongdoing problem areas investigation, and so on[11]. It can likewise increase astounding business bits of knowledge, for example, portable publicizing and showcasing. A conclusion to-end Hadoop-based framework was created with various utilitarian calculations worked accessible as needs be record points of interest (CRDs) [10]. With the data about endorsers' propensities and interests, it is equipped for giving precious data about when, where and how a

classification of people (e.g., sports fans, music significant other, et.) move [8].

### **C) BIG LOCATION DATA**

Human exercises depend on areas, and area information investigation is useful [5]. As represented in, the location based enormous information emerging from GPS sensors, WiFi, blue-tooth through cell phones, have turned out to be valuable vital assets [6]. These assets would offer help for government organization, for example, open office arranging.

### **D) BIG RADIO WAVEFORMS DATA**

Zhang and Qiu [8] utilized huge arbitrary grids as building pieces to display the enormous information emerging from a 5G gigantic MIMO framework that is executed utilizing programming characterized radios, as showed in They abused the way that all information preparing is done at CPU so all the balanced waveforms are put away at the RAMS or at the hard drives [2]. Then again, enormous information investigation in light of the arbitrary lattice hypothesis is connected to the gathered information from their tried, where a versatile client speaks with the gigantic MIMO base station while moving. The exploratory outcomes can evaluate the client's moving pace, regardless of whether unmoving, at an about consistent speed, at a moderate speed or at a higher speed [4]. These applications approve the way that the monstrous MIMO framework is a correspondence framework, as well as a huge information stage which can brings huge values through huge information investigation [7].

## **V. HADOOP**

Hadoop is an entire eco-arrangement of open source extends that give us the structure to manage huge information [6]. Hadoop works in a comparable configuration. On the base we have machines orchestrated in parallel [7]. These machines closely resemble singular patron in our relationship. Each machine has an information hub and an errand tracker. Information hub is otherwise called HDFS (Hadoop Distributed File System) and Task tracker is otherwise called delineate. Information hub contains the whole arrangement of information and Task tracker does every one of the operations[6]. You can envision assignment tracker as your arms and leg, which empowers you to do an errand and information hub as your cerebrum, which contains all the data which you need to prepare. These machines are working in storehouses and it is extremely fundamental to facilitate

them. The Task trackers (Project administrator in our similarity) in various machines are composed by a Job Tracker [5]. Work Tracker ensures that every operation is finished and if there is a procedure disappointment at any hub, it needs to dole out a copy assignment to some errand tracker. Work tracker additionally disperses the whole assignment to all the machines. A name hub then again arranges every one of the information hubs [4]. It administers the appropriation of information heading off to each machine. It likewise checks for any sort of cleansing which have occurred on any machine. In the event that such cleansing happens, it finds the copy information which was sent to other information hub and copies it once more [7]. You can think about this name hub as the general population chief in our relationship which is concerned more about the maintenance of the whole dataset [10].

## **VI. MAPREDUCE**

Map Reduce is a preparing strategy and a program display for conveyed processing in view of java. The Map Reduce calculation contains two essential assignments, specifically Map and Reduce [7]. Delineate an arrangement of information and proselytes it into another arrangement of information, where singular components are separated into tuples (key/esteem sets). Furthermore, lessen undertaking, which takes the yield from a guide as info and consolidates those information tuples into a littler arrangement of tuples. As the arrangement of the name Map Reduce suggests, the decrease assignment is constantly performed after the guide work [9].

The significant preferred standpoint of Map Reduce is that it is anything but difficult to scale information handling over various processing hubs. Under the Map Reduce display, the information handling primitives are called mapper and reducers [6]. Breaking down an information preparing application into mapper and reducers is at times nontrivial. In any case, once we compose an application in the Map Reduce frame, scaling the application to keep running more than hundreds, thousands, or even a huge number of machines in a group is just an arrangement change [4]. This straightforward adaptability is the thing that has pulled in numerous software engineers to utilize the Map Reduce demonstrates [8].

### *The Algorithm*

- Generally Map Reduce worldview depends on sending the PC to where the information lives!

•MapReduce program executes in three phases, specifically delineate, rearrange arrange, and lessen organize.

**Map arrange:** The guide or mapper's occupation is to prepare the information. For the most part the information is as record or registry and is put away in the Hadoop document framework (HDFS). The info record is passed to the mapper work line by line. The mapper forms the information and makes a few little lumps of information [8].

**Reduce organize:** This stage is the blend of the Shuffle arranges and the Reduce arrange. The Reducer's occupation is to handle the information that originates from the mapper. Subsequent to handling, it delivers another arrangement of yield, which will be put away in the HDFS [8].

Map arrange: The guide or mapper's employment is to prepare the info information. For the most part the information is as record or registry and is put away in the Hadoop document framework (HDFS). The information document is passed to the mapper work line by line. The mapper forms the information and makes a few little lumps of information.

Reduce organize: This stage is the blend of the Shuffle arranges and the Reduce arrange. The Reducer's occupation is to handle the information that originates from the mapper. In the wake of handling, it delivers another arrangement of yield, which will be put away in the HDFS.

•During a MapReduce work, Hadoop sends the Map and Reduce errands to the fitting servers in the bunch.

•The structure deals with every one of the points of interest of information passing, for example, issuing errands, checking assignment finish, and replicating information around the bunch between the hubs.

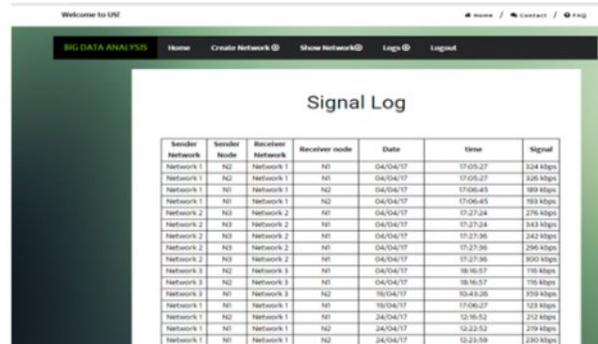
•Most of the registering happens on hubs with information on nearby circles that diminishes the system activity.

•After culmination of the given errands, the group gathers and diminishes the information to shape a suitable outcome, and sends it back to the Hadoop server.

## VII RESULT

In this paper we show the result. In that we calculate the signal log, waveform log, traffic log, location log.

Here we transfer the data from sender to receiver over the network.



Sender Network	Sender Node	Receiver Network	Receiver node	Date	Time	Signal
Network 1	N2	Network 1	N0	04/04/17	17:05:27	324 Kbps
Network 1	N2	Network 1	N0	04/04/17	17:05:27	326 Kbps
Network 1	N0	Network 1	N2	04/04/17	17:05:45	189 Kbps
Network 1	N0	Network 1	N2	04/04/17	17:06:45	193 Kbps
Network 2	N3	Network 2	N1	04/04/17	17:27:24	276 Kbps
Network 2	N3	Network 2	N1	04/04/17	17:27:24	343 Kbps
Network 2	N3	Network 2	N1	04/04/17	17:27:36	243 Kbps
Network 2	N3	Network 2	N1	04/04/17	17:27:36	256 Kbps
Network 2	N3	Network 2	N1	04/04/17	17:27:36	300 Kbps
Network 2	N2	Network 3	N1	04/04/17	18:30:57	188 Kbps
Network 3	N2	Network 3	N1	04/04/17	18:30:57	176 Kbps
Network 3	N1	Network 3	N2	18/04/17	10:43:26	359 Kbps
Network 1	N0	Network 1	N0	18/04/17	17:04:27	123 Kbps
Network 1	N2	Network 1	N0	04/04/17	12:22:52	213 Kbps
Network 1	N0	Network 1	N2	24/04/17	12:22:52	279 Kbps
Network 1	N1	Network 1	N2	24/04/17	12:23:59	230 Kbps

Figure 2: List of signals used from sender to receiver



Network	Node	IP Address	Date	Time	Remark	Required Time	Space/MB
Network 1	N1	192.168.1.23	04/04/17	17:09:25	Assign ip successfully	34	334 mb
Network 1	N2	192.168.1.23	04/04/17	17:03:28	Assign ip successfully	69	338 mb
Network 2	N1	192.168.1.23	04/04/17	17:03:40	Assign ip successfully	379	285 mb
Network 2	N2	192.168.1.23	04/04/17	17:03:43	Assign ip successfully	129	285 mb
Network 2	N3	192.168.1.23	04/04/17	17:03:46	Assign ip successfully	103	286 mb
Network 3	N1	192.168.1.23	04/04/17	18:14:33	Assign ip successfully	597	120 mb
Network 3	N2	192.168.1.194	04/04/17	18:14:45	Assign ip successfully	41	125 mb
Network 7	N1	192.168.1.23	08/04/17	15:17:18	Assign ip successfully	201	139 mb
Network 7	N2	192.168.1.23	08/04/17	15:17:22	Assign ip successfully	104	144 mb

Figure 3: IP Log describing network name, date, time and remark



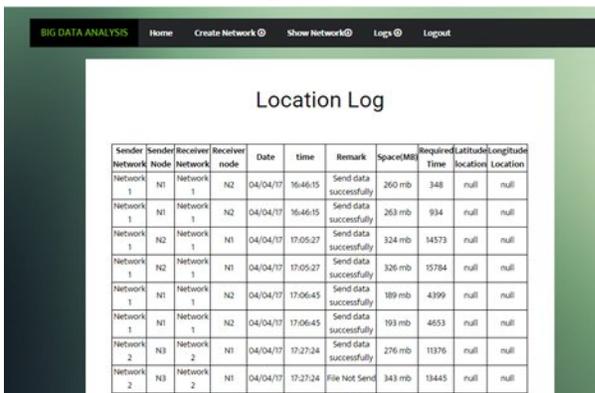
Figure 4: Time Graph showing failure and success nodes

## VIII. CONCLUSION

We proposed the big information expository in portable cell arrange. Enormous Data Analytics is utilized for diagnostic the issue in the versatile cell systems. Hadoop, remote system virtualization, delineate are the instruments for control the activity administration and this apparatuses helps in enhancing the execution in the cell systems.

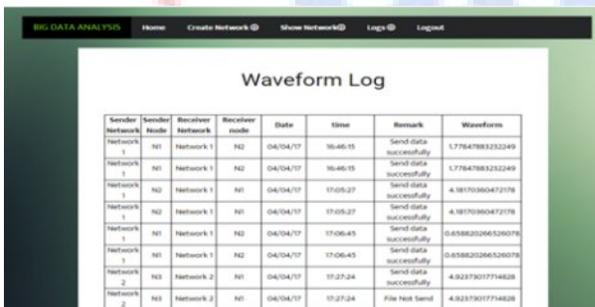


Figure 5: Space Graph showing failure and success nodes



Sender Network	Sender Node	Receiver Network	Receiver Node	Date	Time	Remark	Space/MB	Required Time	Latitude Location	Longitude Location
Network 1	N1	Network 1	N2	04/04/17	16:46:15	Send data successfully	260 mb	348	null	null
Network 1	N1	Network 1	N2	04/04/17	16:46:15	Send data successfully	263 mb	934	null	null
Network 1	N2	Network 1	N1	04/04/17	17:05:27	Send data successfully	324 mb	14573	null	null
Network 1	N2	Network 1	N1	04/04/17	17:05:27	Send data successfully	326 mb	15784	null	null
Network 1	N1	Network 1	N2	04/04/17	17:06:45	Send data successfully	189 mb	4399	null	null
Network 1	N1	Network 1	N2	04/04/17	17:06:45	Send data successfully	193 mb	4653	null	null
Network 2	N3	Network 2	N1	04/04/17	17:27:24	Send data successfully	276 mb	11376	null	null
Network 2	N3	Network 2	N1	04/04/17	17:27:24	File Not Send	343 mb	13445	null	null

Figure 6: Location Log showing receiver and sender details



Sender Network	Sender Node	Receiver Network	Receiver Node	Date	Time	Remark	Waveform
Network 1	N1	Network 1	N2	04/04/17	16:46:15	Send data successfully	1.7784788323249
Network 1	N1	Network 1	N2	04/04/17	16:46:15	Send data successfully	1.7784788323249
Network 1	N2	Network 1	N1	04/04/17	17:05:27	Send data successfully	4.18170390473719
Network 1	N2	Network 1	N1	04/04/17	17:05:27	Send data successfully	4.18170390473719
Network 1	N2	Network 1	N2	04/04/17	17:06:45	Send data successfully	0.8588202066526076
Network 1	N1	Network 1	N2	04/04/17	17:06:45	Send data successfully	0.8588202066526076
Network 2	N3	Network 2	N1	04/04/17	17:27:24	Send data successfully	4.92373017714828
Network 2	N3	Network 2	N1	04/04/17	17:27:24	File Not Send	4.92373017714828

Figure 7: Waveform Log

This overview paper helps in perceiving diverse procedures utilized as a part of the versatile cell systems. In this paper Big information investigation will be an imperative piece of the versatile cell administrators' thought of system operation, business sending, and even the plan of the cutting edge portable cell organize structures. In this paper, the association between enormous information analytics and portable cell systems has been deliberately investigated. We gave a wide diagram of huge information investigation in light of random lattice hypothesis. Next, a building

structure for the uses of enormous information investigation in cell systems was introduced. In addition, a few illustrative cases were given. At long last, we examined some examination challenges and enormous information investigation's prospects for cutting edge cell systems. Future work is in advance to address these difficulties.

## REFERENCES

- [1] Suzhi Bi, Rui Zhang, Zhi Ding, Shuguang Cui, "Wireless Communications in the Era of Big Data", University of California, August 27, 2015, pp. 1-19.
- [2] Manhal Abdel Kader, Ejder Bastug, Mehdi Bennis, Engin Zeydan, Alper Karatepe, Ahmet Salih Er, M erouane Debbah, "Leveraging Big Data Analytics for Cache-Enabled Wireless Networks", Advanced Mathematical Tools for Complex Network Engineering, pp. 1-6.
- [3] Ying He, Fei Richard Yu, Nan Zhao, Hongxi Yin, Haipeng Yao, Robert C. Qiu, "Big Data Analytics in Mobile Cellular Networks", Special Section On Theoretical Foundations For Big Data Applications, Volume 4, 2016, pp. 1985-1996.
- [4] Qiang Yang, Fellow, "Introduction to the IEEE Transactions on Big Data", IEEE Transactions on Big Data, year 2015, pp. 2-15.
- [5] Navya H Ajjar, Nagalambika Swami, "Handling Big Data Issues in Mobile Networking and computing data", year 2015, pp. 780-782.
- [6] Mohammad Abu Alsheikh, Dusit Niyato, Shaowei Lin, Hwee-Pink Tan, and Zhu Han, "Mobile Big data analysis using deep learning and apache spark", year 2016, pp. 1-8.
- [7] Jun Liu, Feng Liu, and Nirwan Ansari, "Monitoring and Analyzing Big Traffic Data of a Large-Scale cellular network with hadoop", year 2014, pp. 32-39.
- [8] Juha K. Laurila, Daniel Gatica-Perez, Imad Aad, Jan Blom, Olivier Bornet, Trinh-Minh-Tri Do, Olivier Dousse, Julien Eberle, Markus Miettinen, "The Mobile Data Challenge: Big Data for Mobile computing research", year 2014, pp. 1-8.
- [9] Ying He, F. Richard Yu, Hongxi Yin, Haipeng Yao, and Robert C. Qiu, "Big Data Analytics in mobile cellular network", year 2013, pp. 1-12.
- [10] Samiddha Mukherjee, Ravi Shaw, "Big Data – Concepts, Applications, Challenges and Future scope", year 2016, pp. 66-74.
- [11] Y. Polyanskiy, H. V. Poor, and S. Verd , "Channel coding rate in the finite blocklength regime," IEEE Trans. Inf. Theory, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [12] M. Capitaine and C. Donati-Martin, "Strong asymptotic freeness for Wigner and Wishart matrices," Indiana Univ. Math. J., vol. 56, no. 2, pp. 295–309, 2007
- [13] N. Guo, J. Aribido, and R. Qiu, "Estimation of interferer location and its transmit power using mobile user packet drop rates," IEEE Trans. Veh. Technol.