Pakistan Journal of Engineering Technology and Science (PJETS)
Volume 6, No 2, December 2016

Impact of Big Data over Telecom Industry

Muhammad Waqar Khan[1], *Muhammad Asghar Khan[1], Muhammad Alam[1, 2], Wajahat Ali[1]
[1]Institute of Business and Management (IoBM), Korangi Creek, Karachi-75190, Pakistan
[2]Institute of Post Graduate Studies,Universiti Kuala Lumpur, Jalan Sultan Ismail,10250,Kuala Lumpur, Malaysia

***Abstract*- During past few years, data is growing exponentially attracting researchers to work o a popular term, the Big Data. Big Data is observed in various fields, such as information technology, telecommunication, theoretical computing, mathematics, data mining and data warehousing. Data science is frequently referred with Big Data as it uses methods to scale down the Big Data. Currently more than 3.2 billion of the world population is connected to internet out of which 46% are connected via smart phones. Over 5.5 billion people are using cell phones. As technology is rapidly shifting from ordinary cell phones towards smart phones, therefore proportion of using internet is also growing. There is a forecast that by 2020 around 7 billion people at the globe will be using internet out of which 52% will be using their smart phones to connect. In year 2050 that figure will be touching 95% of world population. Every device connect to internet generates data. As majority of the devices are using smart phones to generate this data by using applications such as Instagram, WhatsApp, Apple, Google, Google+, Twitter, Flickr etc., therefore this huge amount of data is becoming a big threat for telecom sector. This paper is giving a comparison of amount of Big Data generated by telecom industry. Based on the collected data we use forecasting tools to predict the amount of Big Data will be generated in future and also identify threats that telecom industry will be facing from that huge amount of Big Data.**

Keywords: Big Data, Data Science, Telecommunication.

I. Introduction

Big data analytics is not just a passing trend; it is becoming an important part in every aspect of a communication service [1]. Whenever we talk about the electronic communication it means we are talking about the production of data by using wired or wireless medium. The quick growth of internet and the availability of technology everywhere, internet users are increasing day by day. On the other hand, the boom in smartphone industry makes it easier for users to access network on the move. This ease of access is becoming great threat for telecom industry. People are diverting from wired to wireless medium especially on GSM to access the internet [2]. Mobile applications such as Instagram, WhatsApp, Apple, Google, Google+, Twitter, Flickr etc., providing ease to connect, at the same time are generating huge amount of data to tackle with.

*muhammad.asghar@iobm.edu.pk

Figure 1 is representing the internet traffic trends from year 2015 to 2020 [3]. Figure 2 is elaborating it more in terms of internet users against world population in percentage, growth in using internet by using smartphones, average speed of internet and average speed of per Capita per month [2].

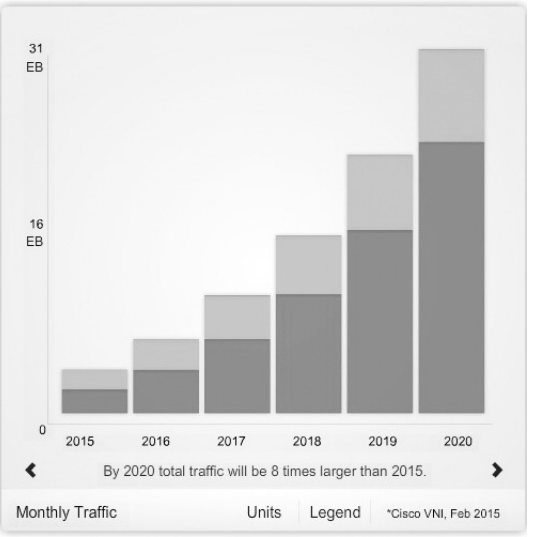

Figure 1: Internet traffic from 2015 to 2020 [3]

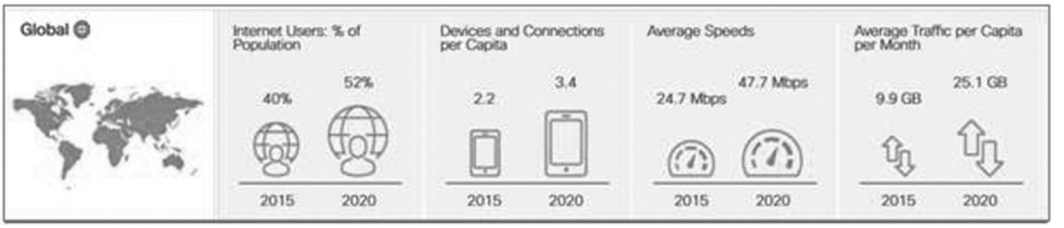

Figure 2: Internet users against world population in percentage, growth in using internet by using smartphones, average speed of internet and average speed of per Capita per month [2]

This paper is further divided into six sections; section 2 is presenting the results of previous research in Big Data and its impact over telecom industry. Section 3 is addressing the problem in the context of Big Data handling in telecom sector. Section 4 is presenting the forecasting and prediction of Big Data usage in telecom sector in next 5, 10 and 30 years. Last section concludes this paper with future directions.

II. Literature Review

This section is divided into two subsections, which are

A. Big Data

B. Big Data in Telecom Sector

A. Big Data

In [4], authors talk about the technologies, opportunities and challenges of Big Data. According to them Big Data management and analysis can be done by considering five (5) common issues which are volume, variety, velocity, value and complexity [4, 5, 6]. Volume of Big Data is normally huge. However, huge amount of data in petabytes or zettabytes is not required. In order to cope with volume problem online additional storage can be managed. Variety is another issue in Big Data management. Different type of data is produced by different devices it need to be fault tolerant. Researchers used different data mining algorithms to deal with variety. Velocity means speedup data. Normalize data provides the velocity. Authors also suggested a life cycle for Big Data by using terminologies and technologies used for Big Data. Stages they include for the life cycles are collection, filtering, analysis, storage, publication, retrieval and discovery [4].

One of the biggest problems in the Big Data is the unstructured data, which is well known as "human information". Financial records, scientific computation and simulation, videos, still images, geospatial images, weather records, call centers data generated by Facebook, tweeter, email and WhatsApp etc. all fall under this category [7]. Unstructured data makes up to 80% of all of the Big Data [7, 8]. Out of all Big Data, 84% of the data is unstructured not modeled and random. It is very difficult to analyze this huge amount of unstructured data [7].

Volume of the Big Data is becoming an alarming threat over the last past few years. Around 5 billion people over the globe call, text, tweet, send emails and browse on mobile devices [9]. In year 2012 around 89 billion emails were received and sent daily and it is forecasted that it will increase at the rate of 13% and by 2020, this amount will be touching 143 billion [10]. In year 2012, 34% of all emails were sent through smart phones [10], and this number is also increasing rapidly with the growth of and usage of smart phones [4, 11-18].

B. Big Data in Telecom Sector

Approximately 40% of the world population has internet connection today. The number of internet users has increased dramatically from 0.4 Billion in 2000 to 3.4 Billion in 2016 [19]. According to 2016 edition of Mary Meeker's annual internet trend report [20], China has the biggest internet population followed by India and USA. Total internet traffic has increased enormously in

the last two decades. In 1992, internet transferred around 100 GB of data every day, whereas these figures shoot up to 100 gigabytes per second (GBps) in 2002. Furthermore, total internet traffic reached above 20,000 GBps in 2015 [21].

Mobile computing has changed the landscape of workplace dramatically and brought many advantages to internet users such as wireless connectivity, increased productivity and easy mobility. Google Consumer Barometer [22] shows positive trends towards acceptance of mobile devices both in developed and developing countries. For example, in China and USA, the percentage of people who use a smart phone for internet has increased from 33% and 44% in 2012 to 79% and 72% in 2016 respectively. Smart phones and tablets sales have already surpassed the Workstation/PC sale. Internet live statistics [23] counter shows that more than 1.6 Billion smart phones sold in the first eight months of 2016 compared to 142 Million computers.

Applications of big data are countless and organizations of all sizes are looking to improve their performance by using big data analytics tools [24]. Yin and Kaynak [25] identified many challenges faced by enterprises when dealing with big data. One of the key challenges for the IT professionals and researcher is to deal with the application that continuously generates large quantity of data at very high speed. Enormous growth has been observed in the data production during the last two years, in fact, 90% of the world's current data has been generated in the last two years alone [26]. Massive amount of data is continuously generated and transferred through network by various internet entities, such as, video on demand, news streaming and e-commerce websites. In the era of social networking, websites are producing data at enormously high speed. For example, Twitter generates over 350,000 tweets per minute and 500 million tweets per day on average [27]. Similarly, Google processes around 40,000 search queries in every second on average [28]. Likewise, astronomy effort, such as Palomar Transient Factory in Southern California, is looking for new phenomena in the sky and roughly captures 30 terabytes of data every night [29].

Mobile devices produce huge quantity of data at enormously high speed. Cisco has developed different forecasting tools [30], based on their Visual Networking Index(VNI), that predict that "Global mobile traffic will grow three times faster than Global fixed IP traffic from 2015 to 2020". Furthermore, "Global mobile data traffic will grow 8-fold from 2015 to 2020, a compound annual growth rate of 53%". This massive shift of IP traffic from fixed to wireless network will introduce number of challenges for Network architects. Telecom companies are sitting on a gold mine, as they have plenty of data. But what they require is a proper analysis of both structured and unstructured data to get deeper insights into customer behavior, their service usage and interests real-time [31].

Since, half of the data is now being transferred via mobile and in 2012, according to study 90% of the data will be transferred through mobile and tablet devices (wireless connections). It could be a threat for telecom industry as they have already invested millions of dollars to process and store the big data via wire line connections. Now the technology is getting changed and majority of the data

is being transferred on wireless connections [32]. But just to sustain in the market, telecom industry needs to reinforce their strategies based on the current scenario. They also have an opportunity to target this untapped market. The telecom industry has an advantage over other industries due to the maximum bandwidth and depth of data it collects in the course of normal business. For example, an operator serving 8 million prepaid mobile subscribers generates around 30 million Call Data Records daily, equaling 11 billion records annually. If the same operator also provides postpaid and fixed lines services, then there is even more volume and variety of data at the ready. Big Data helps telecom operators improve their marketing effectiveness. Relevant actions can be taken based on real-time information without the need to wait for data extraction or manual data mining. Big Data can help gather real-time customer satisfaction information through social media listening or Voice of the Customer analysis. It can also improve customer experience by identifying the most valuable customers who would benefit from dedicated treatment and better services [32].

Different subscription fees can be charged to customer according to their usage on monthly basis. Cloud based services could help telecommunication industries to manage and secure customer's data. Telecom Industry can create new products just to utilize the big data with wireless connections. Top players like Google, Skype, and Netflix may be interested in behavioral data to target consumers with specific content.

III. Problem Statement

Technology is rapidly shifting from wired to wireless networks and every node is busy to generate data. This technology shift in data growth is predicting that we will face serious infrastructural challenges, which include but not limited to, bandwidth problem, network maintenance and support issues, computational bottleneck and MSC's and BSC's congestions. Currently organizations that are dealing with Big Data are increasing resources for load balancing. However, increasing the number of IT resources might solve the problem for the short run but this approach is not scalable for the future growth. IT companies, such as Google, Facebook and BT are looking for some improved protocols and algorithms to improve data transmission and computing. Forecasting the future data growth/transmission in short and long run will help IT researchers, Telecom companies, Governments and other stakeholders to understand where they currently stand and where and in which direction they need to move to solve above mentioned problem. This research paper is using Time Series forecasting to understand the data needs in 2020, 2030, 2040 and 2050.

IV. Proposed Methodology

We have collected data for past 16 years from year 2000 to 2015 of world population and compare it with internet and non-internet users. Based on the available data by using regression we predict for 2020, 2030, 2040 and year 2050 as shown in table 1. Rapid growth of internet users is showing that the amount of data will be immense in year 2050. Trends shows that most of the data will be generated by cellphones/smartphones, because of their handy and available every time nature. This will overburden telecom industry because it will be dealing a huge amount of Big Data. Current research shows that more than 52% of users are connected via smartphones as compare to 41% via desktop [36].

Table 1
Year wise distribution of world population, internet and non-internet users

Year	Internet Users	World Population	Non Users
2000	414,794,957	6,126,622,121	5,711,827,164
2001	502,292,245	6,204,310,739	5,702,018,494
2002	665,065,014	6,282,301,767	5,617,236,753
2003	781,435,983	6,360,764,684	5,579,328,701
2004	913,327,771	6,439,842,408	5,526,514,637
2005	1,030,101,289	6,519,635,850	5,489,534,561
2006	1,162,916,818	6,600,220,247	5,437,303,429
2007	1,373,226,988	6,681,607,320	5,308,380,332
2008	1,575,067,520	6,763,732,879	5,188,665,359
2009	1,766,403,814	6,846,479,521	5,080,075,707
2010	2,023,202,974	6,929,725,043	4,906,522,069
2011	2,231,957,359	7,013,427,052	4,781,469,693
2012	2,494,736,248	7,097,500,453	4,602,764,205
2013	2,728,428,107	7,181,715,139	4,453,287,032
2014	2,956,385,569	7,265,785,946	4,309,400,377
2015	3,185,996,155	7,349,472,099	4,163,475,944
2016	**3,424,971,237**	**7,432,663,275**	**4,007,692,038**
2017	3,448,735,286	7,507,238,701	4,058,503,415
2018	3,640,880,317	7,589,116,161	3,948,235,844
2019	3,833,025,348	7,670,993,621	3,837,968,273
2020	**4,025,170,380**	**7,752,871,081**	**3,727,700,702**
2021	4,217,315,411	7,834,748,541	3,617,433,130

2022	4,409,460,443	7,916,626,001	3,507,165,558
2023	4,601,605,474	7,998,503,461	3,396,897,987
2024	4,793,750,505	8,080,380,921	3,286,630,416
2025	4,985,895,537	8,162,258,381	3,176,362,844
2026	5,178,040,568	8,244,135,841	3,066,095,273
2027	5,370,185,600	8,326,013,301	2,955,827,701
2028	5,562,330,631	8,407,890,761	2,845,560,130
2029	5,754,475,662	8,489,768,221	2,735,292,559
2030	**5,946,620,694**	**8,571,645,681**	**2,625,024,987**
2031	6,138,765,725	8,653,523,141	2,514,757,416
2032	6,330,910,757	8,735,400,601	2,404,489,844
2033	6,523,055,788	8,817,278,061	2,294,222,273
2034	6,715,200,820	8,899,155,521	2,183,954,701
2035	6,907,345,851	8,981,032,981	2,073,687,130
2036	7,099,490,882	9,062,910,441	1,963,419,559
2037	7,291,635,914	9,144,787,901	1,853,151,987
2038	7,483,780,945	9,226,665,361	1,742,884,416
2039	7,675,925,977	9,308,542,821	1,632,616,844
2040	**7,868,071,008**	**9,390,420,281**	**1,522,349,273**
2041	8,060,216,039	9,472,297,741	1,412,081,702
2042	8,252,361,071	9,554,175,201	1,301,814,130
2043	8,444,506,102	9,636,052,661	1,191,546,559
2044	8,636,651,134	9,717,930,121	1,081,278,987
2045	8,828,796,165	9,799,807,581	971,011,416
2046	9,020,941,196	9,881,685,041	860,743,845
2047	9,213,086,228	9,963,562,501	750,476,273
2048	9,405,231,259	10,045,439,961	640,208,702
2049	9,597,376,291	10,127,317,421	529,941,130
2050	**9,789,521,322**	**10,209,194,881**	**419,673,559**

A. Regression Analysis of time series analysis using SPSS:

Linear trend is found between time and Internet users. Time is considered as independent variable (x) and Internet users as dependent variable. The equation describe as [33-35]:

Y (internet users) = - 3.841 x 1011 +1.92x108 x (time)

According to regression line estimated internet user would be raised 5.95 billion in 2030 and 9.79 billion in 2050; this is 69.4% and 95.5% of the total world population respectively. Following graph describe Linear regression support trend line and highly significant. The correlation r=0.992 define highly positive relationship between time and internet users. Figure 3 and 4 are representing trends of internet users till year 2050. Figure 5 is presenting the user inclination towards cellphones. 2015 data shows that 51% of users are using smartphones as compare to desktop users which were only 42% and dropping [36].

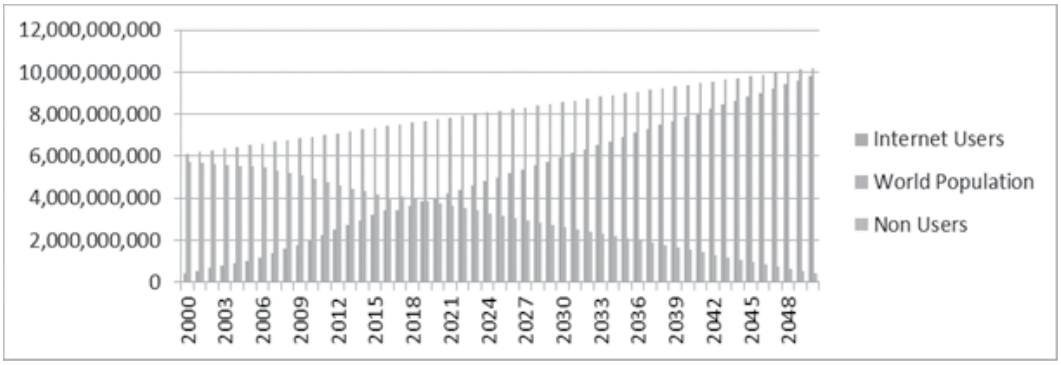

Figure 3: Bar chart representation of world population, internet users and non-internet against time.

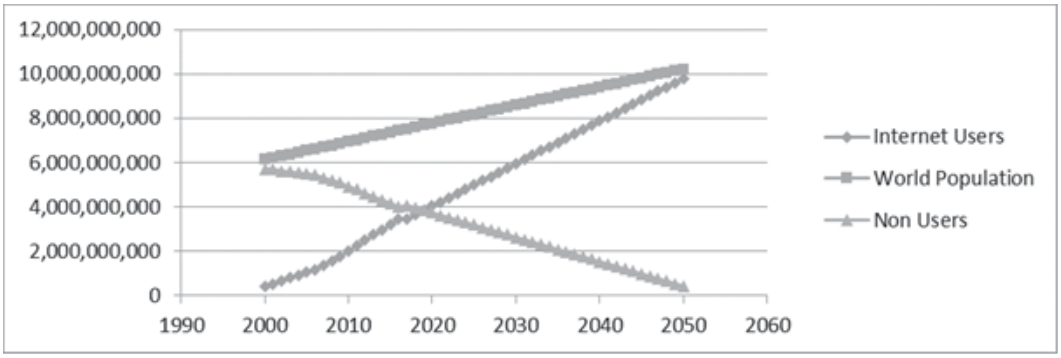

Figure 4: Trend line of world population, internet users and non-internet against time.

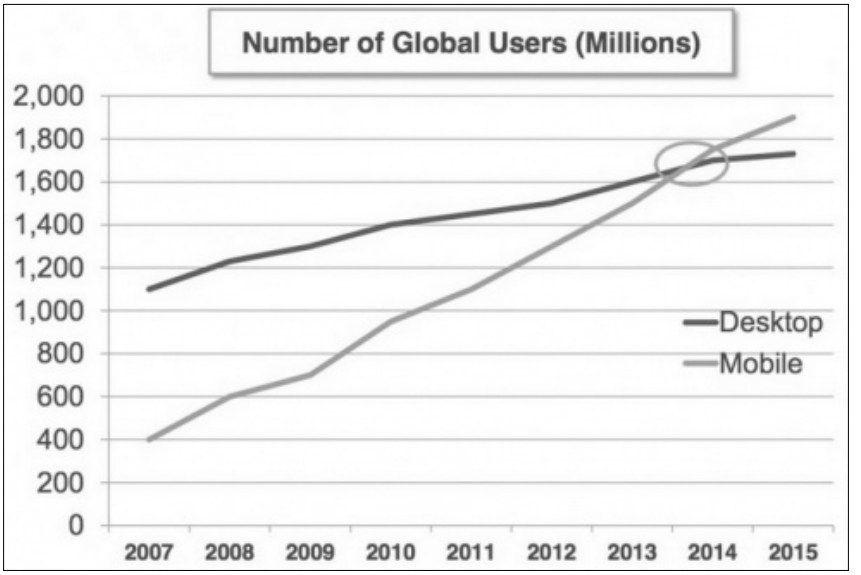

Figure 5: Figure by Dave Chaffey, showing the increase in use of mobile phones over desktop over the years [36]

V. Conclusion, Limitation and Future Directions

Forecasting by using regression shows that in year 2050 95% of world population will be connected via internet and most of it will be using wireless mode to connect especially GSM or future telecom architecture. It will generate huge amount of data for telecom industry for which the industry is not ready to deal with. Most of this data will be redundant therefore proper data mining tools and techniques are required to dig out for the required data and dump the redundant data.

References

[1] Big data analytic is just not a passing trend. http://www.analysysmason.com/Research/Content/Reports/Big-data-analytics-Jul2014/ (last visited: September 16th, 2016)

[2] http://www.cisco.com/c/m/en_us/solutions/service-provider/vni-forecast-highlights.html (last visited: September 16th, 2016)

[3] http://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-widget/forecast-widget/advanced.html (last visited: September 16th, 2016)

[4] Khan, N., Yaqoob, I., Hashem, I.A.T., Inayat, Z., Mahmoud Ali, W.K., Alam, M., Shiraz, M. and Gani, A., 2014. Big data: survey, technologies, opportunities, and challenges. The Scientific World Journal, 2014.

[5] S.Kaisler, F.Armour, J.A.Espinosa, W.Money, "Big Data: issues and challenges moving forward", Proceedings of the IEEE 46th Annual Hawaii International Conference on System Science (HICSS'13), pp.995-1004, January 2013.

[6] S.Sagiroglu, D.Sinanc, " Big Data: A review" Proceedings of the international onference on Collaboration Technologies and Systems (CTS'13), pp 42-47, IEEE, San Diego, Calif, USA. May 2013.

[7] K.Douglas, "Infographic: Big Data brings marketing baig numbers" 2002, http://www.marketingtechblog.com/ibm-

big-data-marketing/. (Last visited: August 20, 2016)

[8] A.Holzinger, c.Stocker, B.Ofner, G.Prohaska, A.Brabenetz and R.Hofmann-Wellenhof, "Combining HCI, natural language processing and knowledge discovery- potential of IBM content analytics as an assistive technology in the biomedical field", in Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured Big Data, vol. 7947 of Lecture Notes in Computer Science, pp. 13-24, Springer, Berlin, Germany, 2013.

[9] P.Russom, "Big Data analytics", TDWI Best Practices Report, Fourth Quarter 2011.

[10] S.Radicati and Q.Hoang, " Email Statistics Report", 2012-2016. The Radicati Group, London, UK, 2012.

[11] Worldmeters, "Real time world statistics", 2014 http://www.worldmeters.info/world-population/. (Last visited: August 22, 2016).

[12] M.Chen, S.Mao and Y.Liu, "Big Data: A Survey", Mobile Networks and Applications, Vol. 19, No. 2, pp. 177-209, 2014.

[13] S.Kaisler, F.Armour, J.A.Espinosa and W.Money, "Big Data: Issues and Challenges moving forward", in Proceedings of the IEEE 46th Annual Hawaii international Conference on System Sciences (HICSS'13), pp.995-1004, January 2013.

[14] A Muhammad, "Power Management of Portable Devices by Using Clutter Based Information" IJCSNS, International Journal of Computer Science and Network Security, VOL. 9 No. 4, April 2009, tel-00815919, version 1-19 Apr 2013 pp 237-244.

[15] A Muhammad, MS Mazliham, M Shahrulniza, M Amir, "Posterior Probabilities based Location Estimation (P2LE) Algorithm for Locating a Mobile Node in a Disaster Area M" MULTICONF p 13-16, 2009.

[16] Muhammad Alam, Mazliham Muhammad Suud, Patrice Boursier, Shahrulniza Musa, Jawahir Che Mustapha Yusuf, "Predicted and corrected location estimation of mobile nodes based on the combination of kalman filter and the bayesian decision theory" , International Conference on Mobile Wireless Middleware, Operating Systems, and Applications, Pages 313-325, Publisher Springer Berlin Heidelberg

[17] A Muhammad, MS Mazliham, Patrice Boursier, M Shahrulniza, Jawahir Che Mustapha Yusuf, "Terrain/clutter based location prediction by using multi-condition Bayesian decision theory" Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication. P. 110-117, 2012.

[18] Muhammad Alam, "Error corrected location determination in an outdoor wireless environment by using estimation, filtering, prediction and fusion techniques: A Wi-Fi application by using terrain based knowledge" , HAL, 2013 pages 313.

[19] http://www.internetlivestats.com (last visited: October 19th, 2016)

[20] http://www.kpcb.com/blog/2016-internet-trends-report (last visited: October 19th, 2016)

[21] The Zettabyte Era—Trends and Analysis, Cisco, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html (last visited: September 27th, 2016)

[22] https://www.consumerbarometer.com/en/trending/?countryCode=CN&category=TRN-NOFILTER-ALL (last visited: September 27th, 2016)

[23] http://www.internetlivestats.com (last visited: September 27th, 2016)

[24] BUILDING A SCALABLEBIG DATA INFRASTRUCTURE FOR DYNAMIC WORKFLOWS,

[25] http://www.emc.com/collateral/solution-overview/h11761-building-scalable-big-data-infra-so.pdf (last visited: September 27th, 2016)

[26] YIN, BS, Kaynak, O (2015),Big Data for Modern Industry: Challenges and Trends, Proceedings of the IEEE, vol. 103, No. 2.

[27] Mukerjee, S and Shaw, R 2016, Big Data – Concepts, Applications, Challenges and Future Scope, International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 2.

[28] http://www.internetlivestats.com/twitter-statistics/ (last visited: September 27th, 2016)

[29] http://www.internetlivestats.com/google-search-statistics/ (last visited: September 27th, 2016)

[30] http://www.nytimes.com/2015/08/01/science/research-scientists-to-use-network-much-faster-than-internet.html (last visited: September 27th, 2016)

[31] VNI Complete Forecast Highlights Tool (2016), http://www.cisco.com/c/m/en_us/solutions/service-provider/vni-

forecast-highlights.html (last visited: September 27th, 2016)

[32] Telecom companies are sitting on a gold mine. http://bigdata-madesimple.com/11-interesting-big-data-case-studies-in-telecom/ (last visited: September 27th, 2016)

[33] Big Data helps telecom operators improve their marketing effectiveness. http://www.teletech.com/resources/articles/numbers-three-big-data-opportunities-telecoms#.V82-gvmqqko (last visited: September 27th, 2016)

[34] Rob J Hyndman and George Athana¬sopou¬los Journal of Forecasting J. Fort. 20, 315-328(2001) DOI: 10.1002/for.795

[35] Joseph Nguyen by Regression Basics for Business Analysis

[36] Statistics for Management 7th Edition by Richard Levin

[37] Dave Chaffey, "Statistics on consumer mobile usage and adoption to inform your mobile marketing strategy mobile site design and app development", Mobile Marketing Statistics compilation: Mobile Market Analytics, April 27, 2016.