

Control-plane Traffic Analysis of UMTS Network

Yi Zhang^{1, a}, Fang Liu¹, Qiujuan Lv^{1, b} and Huan Wang^{2, c}

¹Beijing University of Posts and Telecommunications, Beijing 100876, China

²Haohandata Technology Co., Ltd, Beijing 100876, China

^azhangyi419@bupt.edu.cn, ^blvqiujuan@bupt.edu.cn, ^cwanghuan@haohandata.com.cn

Keywords: 3G, control plane, traffic monitoring and analysis, user behavior.

Abstract. In 3G networks, when user applications send or receive data, control-plane signaling messages will be triggered to initiate or release radio resources. However, massive signaling messages will increase the processing and transmission overheads of the 3G cellular network infrastructure, and this in turns degrades the performance experience of mobile subscribers. Thus, traffic monitoring and analysis are emerging as an attractive solution for network management and optimization. In this paper, we present a traffic monitoring and analysis system called UMAS for 3G network. The system can collect and process the control-plane packets in UMTS core network efficiently. Moreover, we conduct an in-depth analysis of user's radio resources usage and online preferences based on signaling records produced by UMAS. The results from the analysis illustrate the characteristics of radio resources overheads and also identify power-law and heavy-tailed properties in user online preferences. Our study can provide insights for network operators to resource allocation for control planes of 3G network.

Introduction

3G cellular data networks have recently witnessed rapid growth, especially due to the emergence of smartphones and mobile applications. As evidence of this popularity, in May 2013, all three of China's mobile telcos have just released their May reports. It shows that the huge market has hit another significant milestone - it has now surpassed 300 million 3G users in China [1]. Nevertheless, the exponential growth of mobile users imposes new challenges. One potential challenge is the control-plane signaling overhead. According to existing 3G standards, signaling messages will be triggered when the user applications initiate or release radio resources for data transfer [2]. Therefore, it becomes imperative and urgent to analyze the performance of 3G networks so that having thoroughly understanding of the packet traffic to effectively optimize and plan the networks.

In this paper, we implement a packet monitoring and analysis system named UMAS. With our system, we conduct analysis of control plane signaling overheads based on the real-world packets captured from a UMTS network of a famous province in southern China. Then, we identify heavy-tailed workload characteristics and power-law distributions in several aspects of user online preferences on the control plane. All the knowledge and insights that we gained from our study provide useful information for the ISP (Internet Service Provider).

The structure of the paper is as follows. First, we provide some related works in section II, then introduce a brief background on the control plane of UMTS in section III. In the next section, the detail illustration of UMAS is given. Section V specifies our data sets and the network traffic analysis results based on our data sets are illustrated in Section VI. Finally, conclusions are drawn in Section VII.

Related Work

There have been various monitoring and analysis studies on cellular networks. Existing studies mainly analyze the data-plane performance of cellular networks. Examples of analysis include: similarities with IP wireline traffic [3], performance of TCP flows [4], user browsing behavior [5], and network resource usage [6]. Also, some recent studies analyzed the control-plane performance of

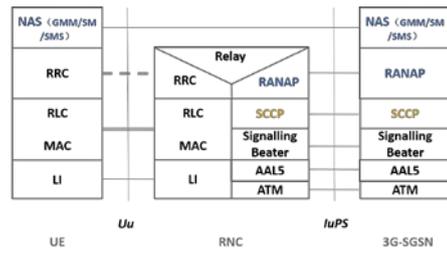
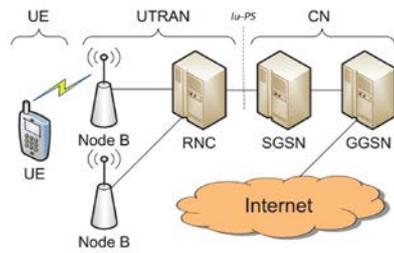


Fig.1. A simplified architectural view of a 3G UMTS network Fig.2.Control plane UE-SGSN

3G networks. Lee et al. [7] studied the signaling overhead from the security perspective. Qian et al. [8] inferred and analyzed the RRC state transitions of user sessions using data-plane traces, and mainly focus on power consumptions of mobiles. In [9], the data/control plane performance of different device types using massive traces collected from the core of a 3G UMTS city-wide network is analyzed.

We distinguish our work from the prior work as follows. First, our work focuses on the Iu-PS interface between RNC (Radio Network Controllers) and SGSN (serving GPRS support node) which is closer to the core network. To a certain extent, it reflects the signaling load of core network better. Also, we study the performance of SCCP (Signal Connection Control Protocol) connections other than RRC (radio resource control) connections. Since the setups of SCCP connections are the basis of user data/signal transmission. It enables us to infer the load of cellular networks accurately. Above all, we use the real data of a province in China. The analysis results are more general and comprehensive than simulation experiment results.

Background of UMTS Control Plane

The Simplified Structure of UMTS network. As is illustrated in Fig. 1, the UMTS network consists of three subsystems: User Equipment (UE), UMTS Terrestrial Radio Access Network (UTRAN), and the Core Network (CN). UEs are essentially mobile handsets carried by end users. The UTRAN allows connectivity between a UE and the CN. It consists of two components: base stations called Node-Bs, and RNC, which controls multiple Node-Bs. Most UTRAN features such as packet scheduling, radio resource control, and handover control are implemented at the RNC. The centralized CN is the backbone of the cellular network [8].

The Control Plane Protocols of Iu-PS Interface. Since the high overhead of protocol puts high pressure on 3G networks especially in Iu-PS interface, it's important and necessary to measure the Iu-PS interface which connects the RNC and SGSN. As is shown in Fig. 2, there are the protocols for controlling the radio access bearers and the connection between the UE and the network from different aspects (including requesting the service, controlling different transmission resources, handover and streamlining etc.). Also a mechanism for transparent transfer of NAS (Non Access Stratum) messages is included [10]. For transport of RANAP (Radio Access Network Application Part) messages over Iu-PS, an SCCP protocol shall be used. The SCCP protocol shall fully comply with ITU-T white book [11].

The Significant Signaling Messages in Control Plane. In the PS domain, the concept of 'session' is used. For each session, a PDP Context is created, which defines the characteristics of the session. The managements of PDP context, which include the procedures necessary to activate, modify and deactivate a PDP context between the UE, SGSN and GGSN (Gateway GPRS Support Node), come under the Session Management(SM) functionality [12].Fig. 3 depicts the signaling message flow of the SCCP duration per user session. When the UE wants to send or receive data, the RNC first to makes a SCCP connection request with the CN, which replies a CC (Connection Confirm) or CREF (Connection Refused) message. After the data transfer, RNC requests to release the temporary signaling connection between two users of the SCCP. Two messages are required to initiate and complete connection release: RLSD (Released) and RLC (Release complete).

During connection establishment, both a source and destination local reference (SLR/DLR) number are assigned independently to a connection section. SLR and DLR are assigned at connection section setup for a permanent connection section. Each node will select the local reference that will

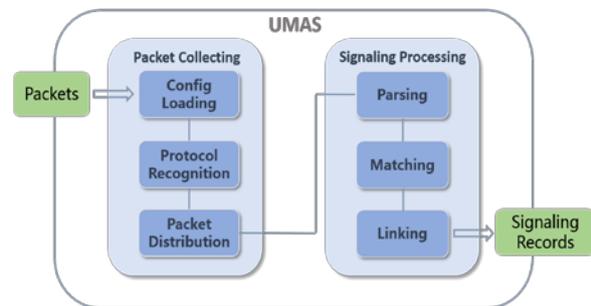
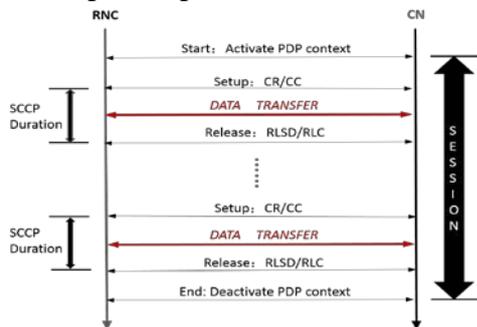


Fig.3. Signaling message flow of the SCCP Duration per user session Fig.4. System overview

Be used by the remote node as the destination local reference number field on a connection section for data transfer. Once the destination reference number is known, it is a mandatory field for all messages transferred on that connection section [13]. So SLR and DLR can identify essential information from different signaling messages to the same connection.

Software Design

This section presents the detailed design of UMAS. UMAS is deployed at the Iu-PS interface for packet capturing and signaling processing between RNC and SGSN. After signaling processing, the network traffic is stored in the format of signaling records, which include the fields of user identifier, connection duration and DATA and so on. The general view of UMAS is illustrated in Fig. 4. UMAS consists of two major modules, packet collecting module and signaling processing module.

Packet Collecting Module. The packet collecting module depends on a hardware-based probe, which is in charge of most part of packet capturing and classification. It consists of three major components: config loader, protocol recognizer and packet distributor. The config loader loads configurations on custom protocol stack for the convenience of parsing underlying protocol flexibly. Then packets will be classified on the basis of upper-layer protocol in protocol recognition component and distributed among different signaling processing modules.

Signaling Processing Module. The signaling processing module is responsible for matching and associating different signaling messages of same users or same connections. This is the core part of our system and it consists of four major components - packet parsing component, matching component, linking component. The work flow is as follows:

The protocol engine starts the signaling processing module as an entry point. The input of this module is the preprocessed packet data from the packet collecting module. The packet parsing component is in charge of dissecting the packets according to the protocol specifications thoroughly. Afterwards, the signaling matcher turns corresponding request and response message into signaling records. Finally, the linking component adds significant information such as IMSI (a unique user identifier), time of SCCP connections setup and release, and the number of SCCP connections per session to the signaling records. These records enables us to do a lot of statistics and data analysis among the control plane of UMTS network.

Data Set

As mentioned above, the dataset used in this paper was collected by UMAS which was collected in a province capital of southern China. The duration of the measurements is all set as a whole day. In order to ensure that the results we get can reflect the real situation of subscribers, we pick two days respectively April 12th and May 12th in 2013. April 12th is a weekday while May 12th is a weekend. The summary statistics of the collected data is shown in Table 1. About 1000G packet data of the two days were analyzed, which ensures the reliability of our analysis. Note that the data presented has been preprocessed. In the preprocess procedure, we split the illegal signaling records. As the field of

records should be in correct format, the value of the field should be in a normal range, and the number of fields in a record should be the same. After preprocessing, the signaling records can be used to do data analysis and statistics.

Table 1 Data Description

Collection Data	Duration	Total Number of SCCP Connections	Users
20130412(weekday)	24 hours	23245171	198522
20130512(weekend)	24 hours	25548431	311017

Data Analysis with UMAS

In the following data analysis, we analyze the characteristics of SCCP connection and user session in section B and C respectively, and then present the relationship between them in section D. Note that each SCCP connection setup/release triggers a number of signaling messages. The frequency of SCCP connection establishment per session reflects the level of radio resources overheads, and the characteristics of user session classify several different categories of users.

Overview. First, we give a graphical overview of our captured data on May 12th. Fig. 5 plots the number of SCCP connection setups over time in 24-hour traces. The horizontal axis of the graph represents the time of the day, and the vertical axis represents the number of SCCP connection setups per 30 minutes. We can observe that the curve almost stay 500,000 below from 0:00 AM to 16:00 PM, then it reaches its peak around 18:00 PM when most people end the day’s social activities. The high level of the number last to around 20:00 PM. Then the curve naturally reaches a low value since most people may return home to use Wi-Fi instead of cellular networks or go to sleep. Hence, there is strong correlation between the frequency of SCCP connection setups and user daily activities.

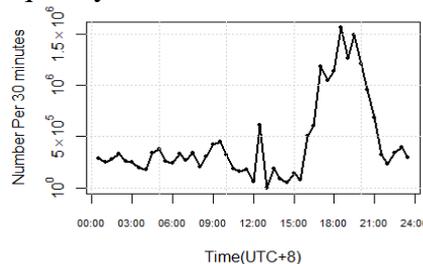


Fig.5. 24-hour time series for the number(per 30 minutes)of SCCP connection setups on 12 May 2013

The Distribution of SCCP Connection Duration. Fig. 6 provides a graphical overview of the SCCP connection duration distribution profile on a log-log scale. Fig. 6a and Fig. 6b are the distributions of April 12th and May 12th respectively. The horizontal axis of the graph represents the duration of SCCP connection, and the vertical axis represents the number of connections. We define the duration as the interval time from receiving the CC message to the reception of RLC message. The system uses the LR (including SLR and DLR) to confirm that these messages are in the same SCCP connection. As is shown in Fig. 6, distributions of the two days are basically the same. We also observe that the distributions fit the red curves approximately and the distributions exhibits heavy-tailed workload characteristics. A distribution is said to be heavy-tailed if the asymptotic shape of distribution is hyperbolic [14]. So the distributions of SCCP connection duration can be modeled well with a heavy-tailed distribution.

On April 12th, 99.33% of connections are equal or less than 100 second, while the proportion on May 12th is 99.31%. The average of SCCP connection durations on April 12th is 18.17 seconds while the value on May 12th is 19.18 seconds. However, Few SCCP connections have a long duration, while many durations of SCCP connections are very short. As different types of user applications (e.g., bulk-data transfer versus interactive applications) have different data transfer characteristics. There are a few long-lived data transfer applications that keep a SCCP connection active for a long time, which will occupy radio resource for a long time.

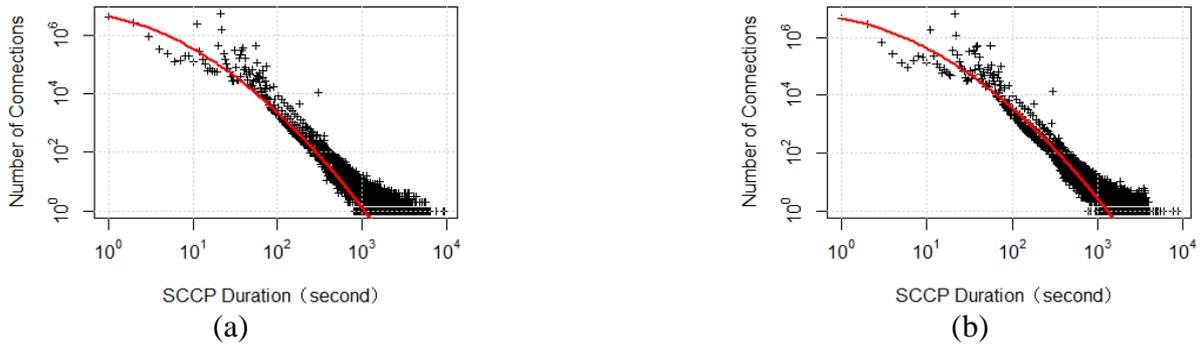


Fig.6. The distribution of SCCP connection duration :(a)2013/4/12,(b)2013/5/12

The Distribution of Session Duration. According to the definition of session, duration of session is the time from PDP activation to PDP deactivation. In this part, we mainly illustrate the cumulative distribution function (CDF) of session duration in UMTS network. The CDF describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to x .

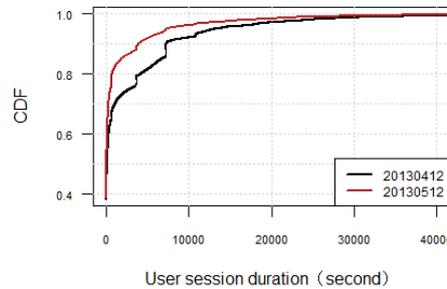


Fig.7. The CDF of User session duration

In our analysis, there are totally 200 thousand sessions on April 12th and 300 thousand sessions on May 12th. Since April 12th is weekday and May 12th is a weekend, the session number of April 12th is less than the number of May 12th. The results of the analysis are illustrated in Fig. 7. The horizontal axis of the graph represents the value of session duration in second, and the vertical axis represents the cumulative probability. There are some differences between the shapes of two curves in Fig. 7. Firstly, the user session duration per user is highly skewed. In particular, 69.7% of the user session durations are smaller than 1000 seconds on April 12th while the proportion on May 12th reaches to 82.3%. Secondly, the average of session durations on April 12th is 2741 seconds (46 minutes), and the average of session durations on May 12th is 1452 seconds (24 minutes). Moreover, based on data of April 12th, we find that 2.8% of the observed users are heavy users (users staying online for relatively long time) in the network, contributing to about 30% of the session duration. On May 12th, the proportion of heavy users is 1.4%. We can observe that users tend to stay at net longer on weekday, which increases the proportion of heavy users. However, there are more online users on weekend. The result shows that more users on weekend prefer to access the network but these users tend to stay at the net shorter. We speculate the reason of the difference is timing difference. Mobile users might prefer to join outdoor activities instead of surfing the Internet for a long time by mobile phone on weekend. We will do further study to find the deep reasons. As the users are discriminate from each other, the ISP might should pay more attention to those heavy users.

Relationship between SCCP Connections and User Session. In this section, we studies the relationship between the duration of session and the number of established SCCP connections during this session to investigate the relationship between the radio resources overhead and duration of user session. To explore the relationship, we first count the number of SCCP connections established during a user session. Fig.8a and Fig. 8b presents the distribution of number of SCCP connections per user session on April 12th and May 12th respectively. The horizontal axis of the graph shows the number of SCCP connections per session, and the vertical axis shows the number of user sessions.

A main conclusion can be drawn from Fig. 8. The established number of SCCP connections during a user session is generally quite small. The expectation of SCCP connection number is 1.16 and 1.31 respectively, but a few users introduce many SCCP connections during a session. Since the signaling

messages consume radio resources, excessive signaling messages can increase both the processing and transmission overheads, and hence degrade performance experience of mobile subscribers. So these users who established a large number of connections during a session should be paid more attention by ISP.

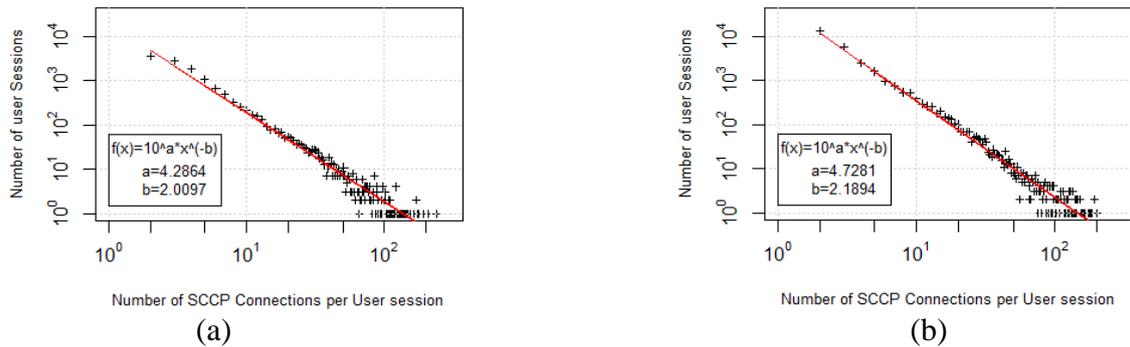


Fig.8. The distribution of SCCP connection number per user session :(a)2013/4/12,(b)2013/5/12

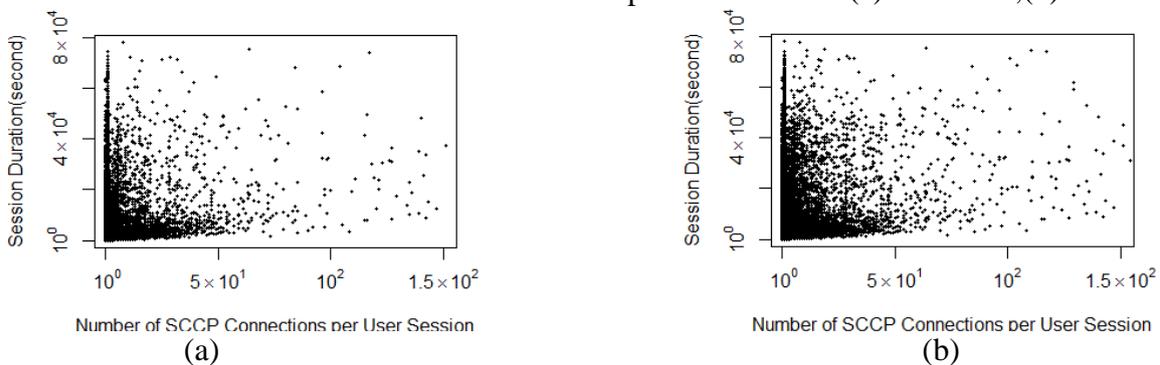


Fig.9. The distribution of SCCP connection number per user session :(a)2013/4/12,(b)2013/5/12

As is shown in Fig. 8, we observe that these distributions above are both highly skewed about 40% of the number of SCCP connections is contributed by only 6.5% of sessions. This observation naturally leads us to formulate a power law model to describe the distributions. The distributions are similar and both fit precisely the following formula:

$$f(x) = 10^a x^{-b} \tag{1}$$

Equation (1) is referred as power-law distribution which is one of the most important characteristics in complex system. The constant b is power law exponent, it has universality. And in log-log scale, the power law distribution presents as a straight line with a slope of negative power law exponent [15]. These unknown parameters (e.g. a , b) are derived by R language through modeling the data [16]. The parameter a represents the coefficient of the regression line and b represents the absolute value of the slope of the regression line. The parameter values can fairly explain that due to the timing difference, user sessions on weekday tend to have less SCCP connections per session than weekends. However, as is shown in Fig. 7, we observe that users tend to stay at net longer on weekday. The emergence of the result doesn't accord with our expectations, so it is necessary for us to investigate the relationship between the number of SCCP connections per session and the session duration.

Fig. 9a and Fig. 9b are the distributions of April 12th and May 12th. The horizontal axis of the graph shows the number of SCCP connections per session, and the vertical axis shows the user session duration. Each point in the graphs represents one user. From the graphs, the longer duration of session will not necessarily lead to more number of SCCP connections established during the session. So we can easily find that despite the time difference, the number of SCCP connections per session and the session duration are weakly correlated so that they can be modeled as separately.

Conclusion

In this paper, we present a traffic monitoring and analysis system called UMAS for UMTS network, which could collect and process the control-plane traffic in UMTS core network efficiently. In

addition, based on the packets captured from 3G control plane, we focus on connection between the UE and the network. Our results illustrate that the durations of SCCP connection and user session are both highly skewed. We also identify pow-law properties in the analysis of number of SCCP connection number per user session and heavy-tailed distribution of SCCP connection durations. Moreover, we investigate the relationships between radio resources overheads and the session duration. It shows that there are some high level of radio resources overheads in a short period of time which would increase the processing and transmission overheads of the 3G cellular network infrastructure and degrade the performance experience of mobile subscribers. All the knowledge and insights that we have gained from our study provide useful information for the ISP.

In the future, we will do the further study to research deep reasons that there are more online users on weekend, but majority of them have a short access to the network. Then, we might do the research to explore causes of high level of radio resources overheads in a short period. These further studies can be useful to improve the network performance and maximize network efficiency for the ISP.

Acknowledgment

This work is supported in part by the Fundamental Research Funds for the Central Universities (2013RC0114) and 111 Project of China (B08004).

References

- [1] S. Millward, "China now has over 300 million 3g subscriber," vol. 2014, 2013.
- [2] T. R. W. Group, "25.931 v1.1.1 utran functions, examples on signalling procedures," 2002.
- [3] J. Ridoux, A. Nucci, and D. Veitch, "Seeing the difference in ip traffic: Wireless versus wireline." IEEE, Apr 2006, pp. 1 – 12.
- [4] J. Kilpi and P. Lassila, "Micro- and macroscopic analysis of rtt variability in gprs and umts networks," ser. Lecture Notes in Computer Science, vol. 3976. Springer Berlin Heidelberg, 2006, pp. 1176 – 1181.
- [5] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao, "Profiling users in a 3g network using hourglass co-clustering." ACM, Sep 2010, pp. 341 – 352.
- [6] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks." IEEE, Apr 2011, pp. 882 – 890.
- [7] P. P. C. Lee, T. Bu, and T. Woo, "On the detection of signaling dos attacks on 3g/wimax wireless networks," Computer Networks, vol. 53, no. 15, pp. 2601 – 2616, 2009.
- [8] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Characterizing radio resource allocation for 3g networks." ACM, Nov 2010, pp. 137 – 150.
- [9] P. P. C. L. X. He, "A panoramic view of 3g data/control-plane traffic: Mobile device perspective," in ifip-tc6 networking conference, 2012, pp. 318–330.
- [10] T. R. W. Group, "Utran overall description," 1999.
- [11] "Itu-t signalling connection control part formats and codes."
- [12] N. N. Kasera S, 3G Networks. Tata McGraw-Hill Education, 2004.
- [13] "Itu-t rec. q.714: Signalling system no. 7-sccp procedures: Signalling connection control part."
- [14] C. Williamson, C. Williamson, E. Halepovic, E. Halepovic, H. Sun, and Y. Wu, "Characterization of cdma2000 cellular data network traffic," pp. Z000 – 719, Nov 2005.
- [15] L. Y. Wu Zhenyu and W. Yuying, "Prediction of social tag frequency's power law distribution with rgf model," pp. 1884 – 1889, Oct 2012.
- [16] T. R. C, "R: A language and environment for statistical computing," 2012