

User Behavior Analysis Based On Gn Interface Of GPRS Network

Yong Chen^{1, a}, Wenli Zhou^{1, b}

¹School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China

^am15117939579@163.com, ^bzwl@bupt.edu.cn

Keywords: GPRS; Gn Interface; GTP; Data Analysis; User Behavior

Abstract. The General Packet Radio Service (GPRS) has become a standard to extend the services provided by the Global System for Mobile Communications (GSM). Through the introduction of packet switching technology, GPRS supports the use of packet data transmission, allowing a more efficient usage of the radio resources with a consequent improvement in the quality of service of subscribers. This paper presents some observations of the real operational GPRS network in a certain city by analyzing the characteristics of the data which was captured by monitoring system deployed in the Gn interface of GPRS core network. The distributions of the activated users and on-offline users were analyzed in this paper, as well as the online duration characteristic of mobile users. At last, the distribution of users and sessions was discussed, and finally some valuable conclusions were concluded for the mobile network operators.

Introduction

GPRS, based on packet switching technology, has been developed to enhance the GSM system. As the transition technology from GSM network to the 3rd generation mobile communication network, GPRS has significant advantages in many ways, such as high transmission speed and low cost [1].

With wide spread use of smart mobile terminals, mobile network has become more and more important in people's life. The operators are also putting much more attention to the quality of network. It can be foreseen that the mobile internet traffic will increase explosively in the near future. So it is necessary to identify the mobile user's behavior and its impact on the network, which may help operators significantly in network maintenance and finally improve mobile subscriber's using experience.

Analyses on the expected behavior of network have been performed in previous researches, which are essentially focus on measures like throughput or the transmission delay of a frame from the source to the destination. The results were normally established on simulation, for example, [2] presented experiments conducted at a GPRS testing platform, focusing on the performance evaluation of the GPRS signal operations. In [3] a GTP state machine was introduced, while the authors in [4] proposed a methodology for characterizing and identifying user behavior in online social networks.

The initial idea of this paper was to capture packet-level traces from the GPRS network and use them to distil synthetic user behavior distribution. Our main contribution is that we present a sample set of findings and results from the exploration of real traces from the operational network with our monitoring system. And our insight based on large amount of data is valuable for network operators.

The rest of the paper is organized as follow. At first, the deployment of system which is used to capture the data in GPRS network is introduced, and then the distribution of traffic generated by mobile users from GPRS network is analyzed and some user behavior characteristics are presented. Finally the conclusion is drawn in the last part of this paper.

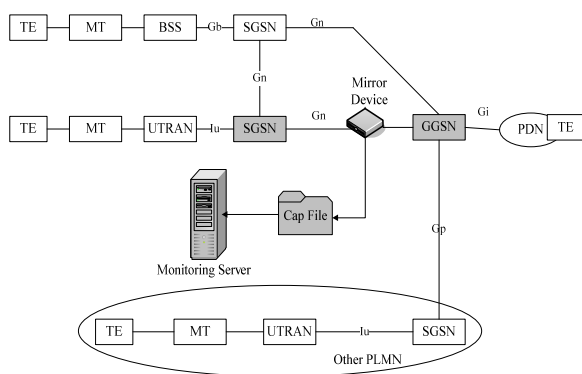
Data Collection

GPRS Overview. To introduce GPRS in the existing GSM infrastructure, additional elements are needed to provide support for packet switching: Service GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). SGSN controls the communications and mobility management

between the mobile stations (MS) and the GPRS network. GGSN acts as an interface between the GPRS network and external packet switching networks such as Internet or GPRS networks of different operators [5].

Between GPRS Support Nodes (i.e., SGSN and GGSN), GPRS Tunnel Protocol (GTP) is invited, providing with GPRS network access services for mobile users [6,7]. GTP includes GTP control plane protocol (GTP-C) and data plane protocol (GTP-U). GTP-C is used to create, modify and release the tunnel between GSNs. Each tunnel is assigned with a unique identifier associated with a particular mobile user. While GTP-U uses the GTP-C tunnel to provide user data packet transmission [8].

System Deployment. The dataset used in this paper is collected from Gn interface in GPRS network. Mirror device is deployed in front-end GGSN to collect GTP messages between SGSN and GGSN, and then transfer the captured file to monitoring system. The key components of GPRS network and the deployment of monitoring system are illustrated in Fig.1.



<i>Date</i>	<i>Log Volume</i>	<i>Records Numbers</i>
Day 1	62.66GB	1,482,432,311
Day 2	62.17GB	1,472,557,724
Day 3	62.46GB	1,496,522,080
Day 4	63.76GB	1,500,545,227
Day 5	63.20GB	1,472,557,724
Day 6	63.44GB	1,480,047,747
Day 7	62.16GB	1,464,257,031

Fig.1 Deployment of monitoring server

Table 1 Daily log volume and sessions numbers

The system collects the data running through the Gn interface and classifies network traffic employing DPI [9] and DFI [10] methods. The traffic data, which contained each user's records including login time, logout time, cell phone number, service option and so on, are stored in a log database. In this paper, we select data collected in one week on March of 2012, and the volume and numbers of records are shown in Table 1.

User Behavior Analysis

Count of Activated Users over Days of Week. We select the data of one week from 05/03/2012 to 12/03/2012 to find out the distribution of activated users. Fig.2 shows the number of activated users that begin in each 5 minutes interval throughout the whole week.

It can be seen that the distribution trends of each day of the week are basically similar, activities of both weekdays and weekends follow the same diurnal cycle. However, the count of activated users on weekday is greater than that on weekend and it fluctuates more intensely than weekend does.

Taking into account the difference between weekday and weekend, we select the average value of four working days and two rest days, Fig.3 presents the count of activated users that throughout 24 hour periods.

We can know how the work schedules of daily life affect the way people use the mobile internet from figure 3. The daily activity follows an expected diurnal cycle. There are much less activities between 1:00 and 7:00 when most people are in sleep. And a quick increase begin at 7:00 while people get start to work in a day. The 12:15 peek seems reasonable for work customers who may get on GPRS during lunch break, while the 19:30 peak may represent people access the internet on the way home after work.

Besides having a higher valley value, the weekend distribution trend follows the same diurnal cycle as the weekday, it can be concluded that most people in this area have normal schedule.

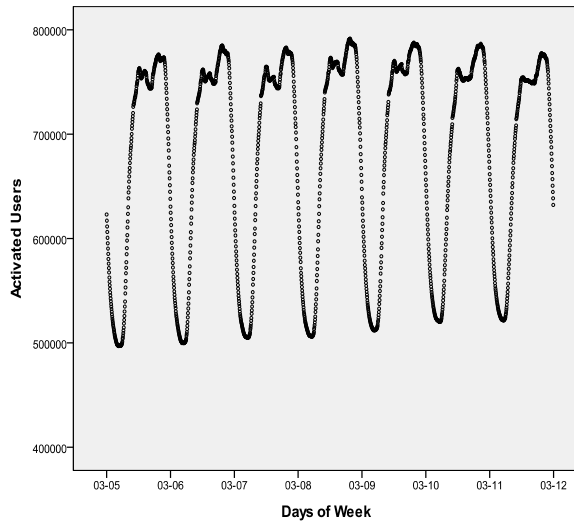


Fig.2 Activated users over one week

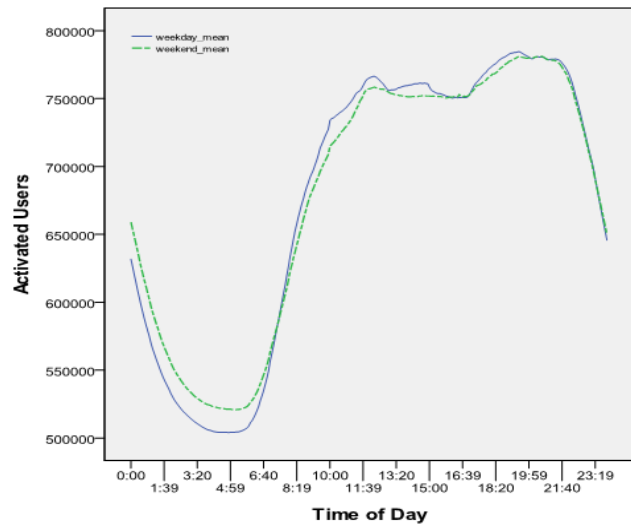


Fig.3 Activated users on weekday and weekend

Online Duration Characteristic. According to the GTP protocol, we defined user online duration as: the interval from the time user gets online as start time, to the time user gets offline as end time, accurate to seconds. The data of the same day was selected and granularity of 5 minutes was used to aggregate each user's online duration.

As it shows in Fig.4, the abscissa indicates online duration and ordinate is the count of sessions. Because of the large amount of short online duration sessions and few numbers of long duration sessions, both the abscissa and ordinate are taken as the bottom 10 of the logarithm.

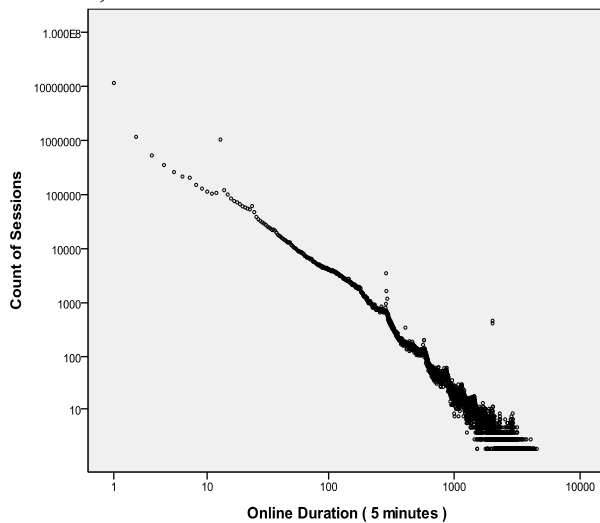


Fig.4 Distribution of the sessions with different online duration (within 5 minutes)

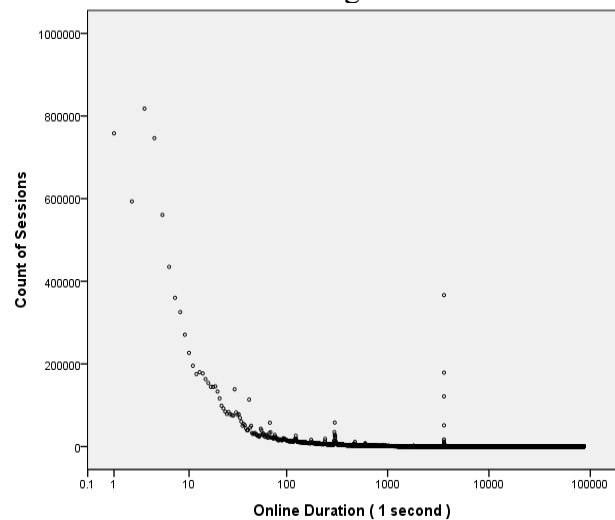


Fig.5 Distribution of the sessions with online duration less than 24 hours

It can be found that the distribution of the count of sessions and online duration is heavy-tailed distribution. With the increase of online time duration, the sessions of users decline rapidly.

The sessions with less than 5 minutes online duration occupy a high proportion of 63.99%, and above 99% of sessions just have durations less than 24 hours. It can be concluded that the short online duration is the characteristic of GPRS users, which indicates that people may be immediately disconnect the GPRS once they have obtain certain information, or they might always reconnect to the network and cause continuous short online time duration.

As the 99% of sessions are less than 24 hours online duration, we need to pay more attention to this major part of sessions and the detailed distribution is shown as Fig.5. We can notice the characteristic of short time duration of mobile users: the number of people who just access the GPRS for 10 seconds is nearly 5 million, which accounts for 28.4% of total users. It also can be found that the large quantities of users whose online duration is around 3600 seconds, by tracing these users' online records, a general characteristic of them is: they get on line immediately after getting offline or after a

very short interval, at the same time, all getting offline requests recorded are proposed by users, it means there is no possibility of a timing mechanism in GGSN side. This phenomenon of online duration maintaining at entire hours is a result of certain bias users. These users' terminals may be used as servers or collectors with a timing mechanism.

By data analysis and fitting technology, the curve is fitted by different models, and ultimately we find the power function model has the best targets. The function model coefficient of determination is 0.916, and F test statistic gets the observation value of 924891.738. Following formula (1) is the power function of proportion of sessions with different online duration.

$$f(t) = b_0 t^{b_1}, t \geq 1. \quad (1)$$

And $b_0 = 131306355$, $b_1 = -0.592$, power function model not only reflects the presence of the universal law when users surfing in GPRS network, but also provides a theoretical basis for user behavior analysis and helps operators to introduce new value-added services.

Distribution of Users and Sessions. In order to have further research on behavior of mobile users, we also analyzed the counts of sessions and numbers of users that once have online or offline activity in a day. After counting all the users through their cell phone numbers, we calculate numbers of users that have same counts of sessions, and present the distribution of them as Fig.6 shows.

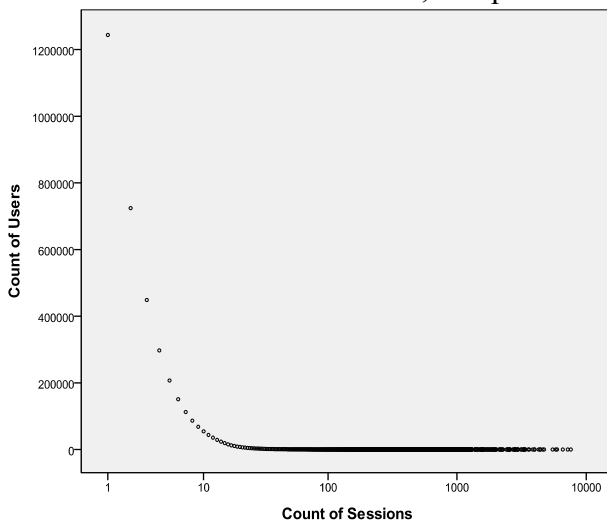


Fig.6 Distribution of counts of sessions and users

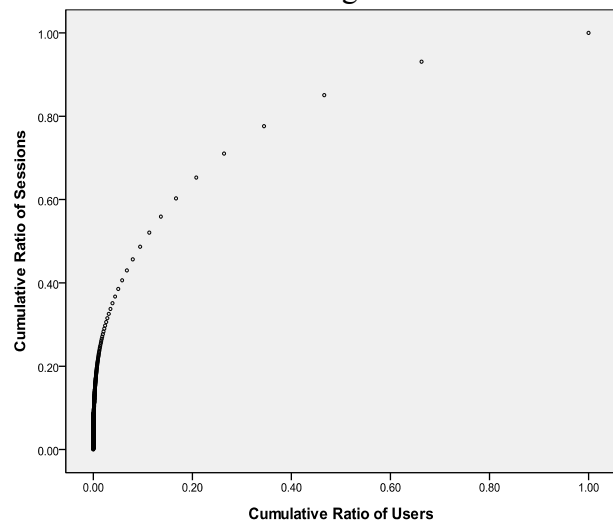


Fig.7 The Pareto Principle in GPRS network

The abscissa, which indicates the number of sessions, is taken as the bottom 10 of the logarithm in order to facilitate the analysis. The number of users who have less than 10 sessions a day has a large proportion of total users, the ratio is nearly 92%, while the user only just have one session a day contributes a proportion of 33.7%. This may indicate that the most of people in this area have a few activities within a day. It would be benefit for the network operators in this region if the business promotion is strengthen to develop the huge market potential.

Meanwhile, we can find some users have a large amount of sessions in a day, and the highest number is 7609, which means the user access to the GPRS network in every 11.35 seconds. We can imagine that the abnormal user behavior certainly will push much pressure on the network system, and it is very important for operators to control this phenomenon.

Now we sort the users by descending order according to the numbers of their sessions, and calculate the cumulative proportion of users and sessions, finally get the distribution as the Fig.7 shows. It can be found that the 20% users have created 65% of total sessions, and the top 10% users even have 50% of sessions.

We can conclude that in GPRS network, the most connections are created by few mobile users, which meet the Pareto Principle put forward by the Italian statistician Pareto. The Pareto Principle states that roughly 80% of the effects come from 20% of the causes, so long we need to handle the few important factors which may cause the most effects. As in this paper, we suggest the operators should monitor or control the users who generate a lot of sessions.

Conclusion

In this paper, the results of analysis are presented from GPRS users' activity records in certain city over a significant period.

It is obvious to know the characteristics of mobile user behavior through above analysis. The mobile users surfing habits are closely related with the people's daily life. The distributions of activated users are basically the same between workday and weekend. And a reasonable online time duration model, the power function model, has been established, which will play an important role in user behavior analysis in GPRS network. And the Pareto Principle in mobile network is found, which indicates that the most network connections are created by few mobile users. All these conclusions are very helpful for operators to maintain and optimize the service of network.

Acknowledgements

This work is supported by National Natural Science Foundation of China (61072061) and 111 Project of China under Grant No. B08004.

References

- [1] Information on <http://www.3gpp.org/3GPP>
- [2] N. Stylianides, G. Stephanopoulos, G. Tselikis and M. Dopher. Signaling Performance Trials and Evaluation Results on a GPRS Platform, *Information Systems Frontiers*, 7(2), pp. 129-140, (2005)
- [3] Z. Liu, W. Li and Y. Lai. Application of Bloom Filter for GTP Stateful Inspection in Network Processor, *IAS*, pp. 589-592, (2009)
- [4] M. Maia, J. Almeida and V. Almeida. Identifying User Behavior in Online Social Networks, *European Conference on workshop on social network systems*, pp. 1-6, (2008)
- [5] Information on http://en.wikipedia.org/wiki/GPRS_Core_Network
- [6] 3GPP, 3rd Generation Partnership Project, Technical Specification Group Core Network and Terminals; General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface (Release 6) Technical Specification 29.060, V6.19.0, (2008)
- [7] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Numbering, addressing and identification (Release 9) Technical Specification 23.003, V9.0.0, (2009)
- [8] F. Tataranni, S. Porcarelli, F. Giandomenico, A. Bondavalli, L. Simoncini. Modeling and Analysis of the Behavior of GPRS Systems, *IEEE*, pp. 51-58, (2001)
- [9] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss. An architecture for differentiated service, *RFC 2475*, (1998)
- [10] C. Wang, X. Zhou, F. You, H. Chen. Design of P2P traffic identification based on DPI and DFI, *International Symposium on CNMT*, pp. 1-4, (2009)