

Modeling and Analysis of User Behavior of GPRS Network

Cheng Wei ¹⁺, Wang Chen ², Lei Zhenming ¹, Chen Luying ¹ and Zhang Bo ¹

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

² Network Management Center, China Mobile Group Zhejiang Co., Ltd, Zhejiang, China

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 function entity, GPRS supports the use of packet data transmission, allowing a more efficient usage of the radio resources with a consequent improvement in the QoS of subscribers. In this paper we present some observations of the GPRS, by describing user behavior at Gn interface in GPRS core network. By disposing monitoring and analysis system at the Gn interface of real operational GPRS network, we analyzed the distribution of mobile user's online duration and factors affect it, such as user's business types and terminal types. At last we obtained the model of the user online duration distribution. And some methodological issues are involved in performing this type of analysis, for instance the potential bias introduced by some users.

Keywords: GPRS, Gn, GTP, user behavior, online duration.

1. Introduction

GPRS has been developed to enhance the GSM system with the introduction of services based on a packet switching technique[8]. As the transition technology from the existing GSM network to third generation mobile communication, GPRS network in many ways has significant advantages, such as high transmission speed, low-cost. With GPRS as technical support, many businesses can be achieved, such as e-mail, e-commerce, mobile office, online chat, WAP-based information browsing, PDA terminal access, integrated positioning technology and other businesses. An important goal of the technology is to make it possible for GSM license holders to share physical resources on a dynamic, flexible basis between packet data services and other GSM services.

Work on GPRS has started in 1994, and a standardization of the GPRS specification has been performed by ETSI (European Telecommunications Standard Institute). With the GPRS technology and business matured and widely used, GPRS network will gradually become the mainstream bearer for mobile data services.

1.1. Related works

Analyses of the GPRS expected behavior have been performed, essentially focusing on measures like throughput, delay for the transmission of a data frame from the source to the destination, Their results are basically established on simulation, cannot simulating the environment in the real operational network. For instance [5] presented experiments conducted at a GPRS testing platform focusing on the performance evaluation of the GPRS signaling operations related to the establishment and release of user data channels

⁺ Corresponding author. Tel.: + 62283741.
E-mail address: chengweibupt@gmail.com.

through the Gn and Gi interfaces. [6] introduced GTP state machine, [7] reported some observations about TCP Round Trip Time in the live traffic of an operational GPRS /UMTS network.

The dataset used in this work is derived by packet traces collected at a Gn interface in real operational GPRS network. To the best of our knowledge there were no previous papers presenting an analysis of the distribution of online duration.

The rest of the paper is organized as follows. Section II introduces the basic structure of GPRS network and the deployment of our monitoring equipment. Section III provides a detailed description of several measurement results and explanations. In Section IV we conclude, and give some hints for further research.

2. 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). The SGSN controls the communications and mobility management between the mobile stations (MS) and the GPRS network. The GGSN acts as an interface between the GPRS network and external packet switching networks such as Internet, or GPRS networks of different operators.

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

2.1. Equipment deployment

In this work the dataset used was collected from Gn interface in GPRS network[7]. Network monitoring devices were deployed in front-end GGSN to collect GTP messages between SGSN and GGSN, and transferred flow to analysis system.

The system collected data, analyzed the distribution of mobile users' online duration, in-depth proposed the model of the user online duration distribution. By further distinguishing features, such as users' business types and terminal types, a reasonable explanation for the model was given.

3. Data Analysis

Before data analysis, basic definitions are given.

Packet Data Protocol (PDP): a network protocol used by packet switching external networks to communicate with GPRS networks. To efficiently deliver the packets in the core network, The PDP data structure (so called PDP-context) is present on both the SGSN and the GGSN that contains the MS's session information during an active session[7]. When a GPRS terminal becomes active it establishes a PDP-context with the network that is maintained during the whole activity period. During this period the terminal is assigned a dynamic IP address by the GGSN. The details of the contents in a PDP-context can be found in [1, 3].

“PDP online duration” is introduced as: the interval from the time MS gets online as start time, to the time MS gets offline as end time, accurate to seconds. Taking into account the existence of packet loss, retransmission and so on in the real operational network, in order to get the full PDP-context record, the analysis system will remove the following conditions:

- a) Cannot create the PDP-context, including the loss of the GTP create request/response packet, GGSN refused to create a context;
- b) Repeat to create the PDP-context, including retransmission, the conflict context packets;
- c) Cannot remove the PDP-context, including the loss of GTP delete request / response packet.

This system analyzed billions of PDP on-offline records, and summed up the general law of distribution of PDP-context with online duration. Finally the distribution model of the proportion of the number of PDP-context with online duration was obtained.

3.1. PDP online duration model

By mirroring user data from the real operational GPRS network, the user on-offline state machine was maintained. Online duration was clustered in accordance with particle size of 10 seconds. The longest online duration is more than 7 days, only holding a very small proportion. To highlight the features, only PDP online proportion distribution with online duration less than 10 minutes was shown in Figure 1: online duration of mobile user in GPRS network is very short. Combined with the original statistical data analysis, online duration concentrated in 1 minute, accounting for 55%. Online duration of less than 5 minutes accounts for 72%. As the online duration grows, curve decline gently, showing a general uniform distribution. The average of PDP online duration is 1422 seconds. This phenomenon has relevant with user behavior which the network service for.

By data analysis and fitting technology, the above results were fitted by different models, and ultimately got the power function model, which gets the best targets. The model coefficient of determination is 0.931, F test statistic got the observation value of 495003.7, and F statistic probability p is 0.000. The follow is the power function of proportion of the number of PDP-context with different online duration.

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

and $b_0 = 66.229, b_1 = -2.166$, Figure 2 shows the fitting results of PDP online proportion distribution with power function model accounting for 1 hour.

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. The distribution of the number of PDP-context in GPRS network meets the power function model, meaning that the residence time for Internet access changes in different GPRS networks. The user's behavior is affected by services provided, network conditions, and the impact of tariff packages and other factors. To introduce a new package or promote new business certainly needs to take user behavior into account in order to achieve the desired results. Length of online duration for users in different stages, especially which is very short and ultra-long need to be analyzed to get underlying causes, such as summarizing distribution of business types, analyzing offline causes, surveying the frequency of getting online. On this basis, it can design businesses and web sites to meet the user preferences, or even business packages which can change the behavior of users in GPRS network.

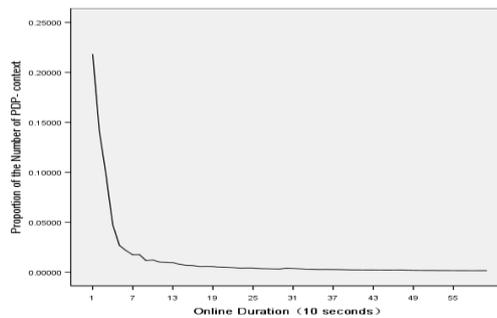


Fig. 1: Distribution of the Proportion of the Number of PDP-context with Different Online Duration.

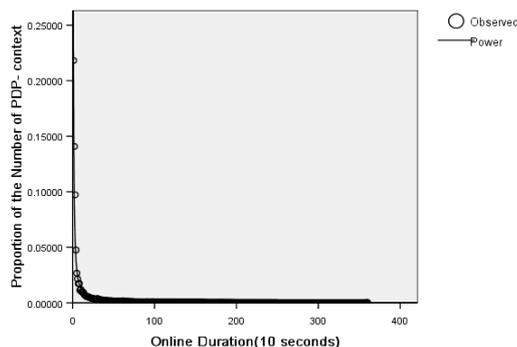


Fig. 2: Power Function Model of the Proportion of the Number of PDP-context with Different Online Duration (within

1 hour).

3.2. PDP online duration model analysis

In order to verify the PDP Online Duration Model, offline records were classified with different APNs (Access Point Name) to test the model. CMNET and CMWAP are two different APNs when using surfing in China Mobile GPRS network. CMWAP only provides WAP proxy and HTTP proxy, users can only access WAP site. CMNET provides NAT service. Mobile using the service can access the internet directly. Therefore CMWAP is mainly used for mobile phone with WAP browser. In the addition to mobile phones, CMNET is also used for PC, notebook computer, PDA and other devices to access Internet. The service objects they oriented result in diversity of online duration for the two groups. Whether CMWAP, CMNET or other kinds of access types, most of PDP online durations concentrate in a very short time. Within 10 minutes, PDP-context with CMWAP accounted for 63.1% of total and CMNET 56.8%. The average of PDP online duration with CMWAP is 1125 seconds, CMNET 1992 seconds. Total PDP online duration gets 1422 seconds. CMWAP are dominant in the current condition of GPRS network.

The distribution of the proportion of the number of PDP-context with different online duration based on CMWAP, CMNET were fitted with power function model. The following function is power function of the proportion of the number of PDP-context with different online duration based on CMWAP.

$$f_{cmwap}(t) = 140.402t^{-2.294} \quad (2)$$

The model coefficient of determination is 0.929, F test statistic got the observation value of 303650.8, and F statistic probability p is 0.000.

Power function of the proportion of the number of PDP-context with different online duration based on CMNET:

$$f_{cmnet}(t) = 140.588t^{-1.919} \quad (3)$$

The model coefficient of determination is 0.916, F test statistic got the observation value of 311297.8, and F statistic probability p is 0.000.

It can be seen from (1), when independent variable t gets small value closing to 0, constant b_0 has greater impact than b_1 , the proportions of the number of PDP-context based on CMWAP and CMNET are roughly equal. However, due to the involvement of other types of APNs, constant b_1 of power function for all APNs is smaller than that of each (CMWAP, CMNET), resulting in a smaller proportion in short duration for all APNs. When independent variable t increasing, constant b_1 has greater impact than b_0 . The bigger t is, the greater the gap between the proportions of each type of APN.

$$\text{if } (2) < (3) \text{ then } 140.402t^{-2.294} < 140.588t^{-1.919}$$

that is $t > 1$. It means, after 10 seconds, the proportion of the number of PDP-context based on CMNET is larger than CMWAP. However, the power function curve does not coincide with distribution of the proportion when independent variable t closes to 1. We can make this conclusion: PDP-context with CMNET has a longer online duration than CMWAP. It coincides with our common sense: the online duration with CMWAP which is limited due to access restriction, terminal mobility, has shorter time.

No matter which APN mobile use to accessing GPRS network, the distribution of the proportion of the number of PDP-context will be a huge peak in the vicinity of full hours. This reflects that user online duration in GPRS network has a tendency to maintain full hours. This phenomenon coincides with the distribution of PDP-context online duration with all APNs. But such model would be influenced by the population composition of terminal types (laptop, handset). It is evident that usage patterns are inherently different for laptop and handset. Since the relative ratio between the two populations of terminals has not stabilized yet, we can expect macroscopic shifts in the network-wide traffic composition in the near future, which makes the

case for the model of performances at the scale of the whole network[7]. Also the distribution of user online duration is impacted by user behavior, such as the use of new applications or new charge mode.

3.3. Shortcomings of PDP online duration model

Power function model reflects the distribution features with online duration in GPRS network to a large extent. But there are still shortcomings. There are two points in the following.

Fitting degree in power function model is not good when independent variable t closes to 1 (real online duration is 10 seconds) on the range of the distribution as shown in Figure 3: the power function value $f(1)$ is much larger than the actual proportion of the number of PDP-context when duration around 1.

To further study the reason that online duration focusing on the short-time range, types of businesses that users used in GPRS networks were classified and statistically analyzed. Figure 4 shows the distribution of the number of PDP-context with online duration for 12 types of businesses within 100 seconds: the main business used is browsing. This phenomenon conforms to the habit of mobile users in current network. In this business records, each uplink traffic flow has an average of 1845 bytes, an average of 6832 bytes downlink traffic flow. Other categories, such as downloading, e-commerce have not widely used in mobile users.

Among these types of businesses with PDP online duration in 1~3 seconds, the largest proportion of business is browsing, accounting for 32.5%, unknown type of TCP, UDP services accounted for 35%. Table I is the average proportion of various businesses, uplink and downlink packets and bytes with online duration 1~3 seconds.

With Further analysis, uplink of unknown TCP and UDP services only have the average number of 1 packet, 141.5 bytes, suspected to be notice news for certain business.

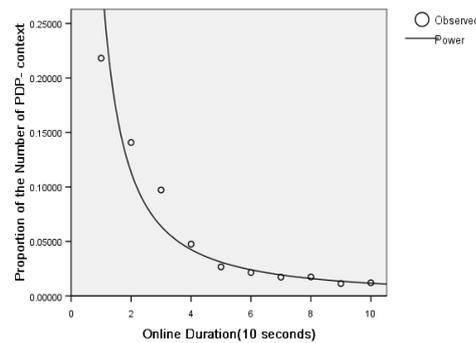


Fig. 3: Power Function Model of the Proportion of the Number of PDP-context with Different Online Duration (within 100 seconds).

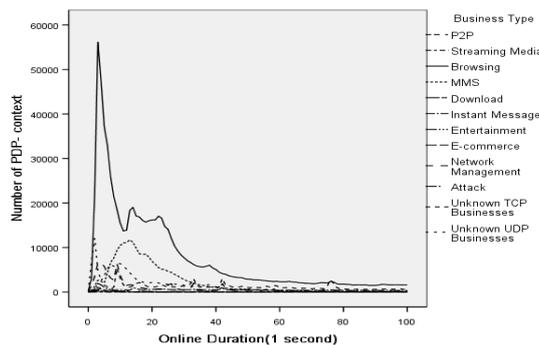


Fig. 4: Distribution of the Number of PDP-context with Different Online Duration for 12 Types of Business.

Tab. 1: Average Proportion of 12 Businesses (within 3 seconds)

<i>Business Type</i>	<i>Ratio</i>	<i>Uplink Packets</i>	<i>Downlink Packets</i>	<i>Uplink Bytes</i>	<i>Downlink Bytes</i>
<i>P2P</i>	0.506 %	3.43	3.33	765.97	1430.97
<i>Stream Media</i>	0.082%	5.89	7.05	1220.32	3500.04
<i>Browsing</i>	24.678 %	4.90	4.99	1018.46	2608.55
<i>MMS</i>	0.296 %	5.56	4.57	1153.41	1417.84

<i>Download</i>	0.168 %	6.92	8.95	1073.60	5669.58
<i>Instant Message</i>	32.534 %	1.88	1.53	230.67	316.38
<i>Entertainment</i>	1.236%	2.02	1.19	296.17	216.62
<i>E-commerce</i>	0.257%	2.30	2.00	291.72	423.84
<i>Network Management</i>	3.870%	1.47	1.44	176.89	461.65
<i>Attack</i>	0.840%	1.28	0.36	147.05	48.63
<i>Unknown TCP</i>	24.746 %	0.83	0.44	120.59	58.39
<i>Unknown UDP</i>	10.789%	1.15	0.34	189.27	54.77

The fitting degree in power function model is not good when independent variable t closes to $360*n$ (real online duration is n hours) on the range of the distribution.

The ratio(the proportion of PDP-context divided by the power function value $f(t)$) in Table II shows that, power function model does not reflect the real situation of GPRS networks when the independent variable t (units of 10 seconds) is in the vicinity of hours. Distribution of the proportion of the number of PDP-context in curve is listed below with online duration in 50~70, 110~130, 170~190, 230~250 minutes.

Tab. 2: Comparison of value for Power Function Model and Distribution of the Proportion of the Number of PDP-context with Certain Online Duration).

Online Duration (seconds)	t(10)	Online Explanation	Duration	$F(x)^a$	Proportion in certain Online Duration	Ratio b
361		1 hour 10 seconds		0.00019	0.00311	16.4
724		2 hour 40 seconds		0.00004	0.00051	12.8
1082		3 hour 20 seconds		0.00002	0.00007	3.5
1444		4 hour 40 seconds		0.00001	0.00005	5.0

a. $F(x) = 66.229x^{-2.166}$

b. Ratio = Proportion/ $F(x)$

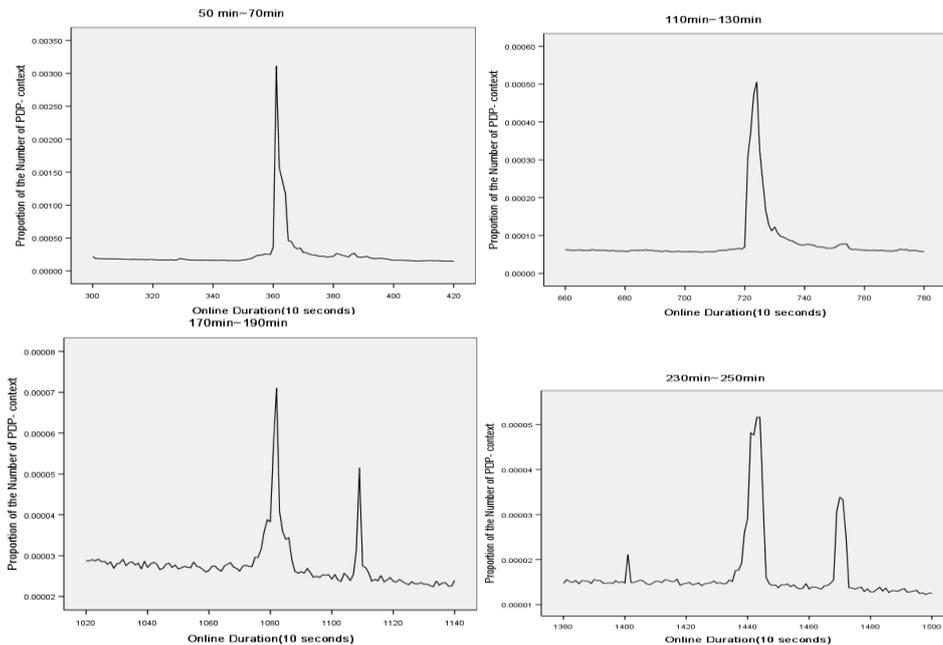


Fig. 5: Distribution of the Proportion of the Number of PDP-context with Different Online Duration (50~70, 110~130, 170~190, 230~250 minutes)

Figure 5 reveals the unique features for distribution of the proportion of the number of PDP-context with different online duration in GPRS network: distribution of the proportion of the number of PDP-context will be a huge peak in the vicinity of hours, meaning that online duration concentrated around hours for GPRS users has higher probability than surrounding area.

To further validate this conclusion, we got online records which have online duration of 50~70, 110~130, 170~190, 230~250 minutes and so on. Then duplicate records for the same user were filtered out. Distributions of the proportion of the number of users with different online duration are shown as in Figure 6: there is such a group of users, whose online duration concentrated on hours (1 hour, 2 hours, etc.). By tracking

users' online records, a general characteristic of these users 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, that is, 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. Such users do not comply with the general behavior of mobile users, and these users usually have several PDP-contexts at the same time. Such users' terminals may be used as servers or collectors, etc., with a timing mechanism.

Through sampling analysis on these users whose online duration concentrating on hours, they can be divided into 2 categories: the first group uses terminals for web browser or mobile QQ chatting. They make a large amount flow records, each record contain a small amount of packets during online. Another typical user group is different. They have a lot of flow records with regular time interval (15 seconds, 30 seconds to 1 minute); each uplink and downlink flow has 1 or 2 packets with a fixed number of bytes.

From the above analysis we can see: In GPRS network, certain types of business in order to maintain its connection, status notification message will be sent regularly. Under certain conditions, the offline trigger is caused. The existence of such business with timing mechanism results in the phenomenon that mobile users' online duration bursting on hours.

4. Conclusions and Future Work

In this work, through analyzing the PDP-context online duration in GPRS network, a reasonable PDP-context online duration model has been established, that is power function model, which will play an important role in user behavior analysis in GPRS network. Through further analysis, the model is confirmed correct. It also has distortion in some cases (online duration around full hours). The model parameters are different in different GPRS network environment, differentiated and meticulous management can be implemented effectively. First, with the model, network operators can effectively evaluate the status of current network, whether there is a lot of congestion and attacks. Second, combined with the online duration and preferences of user groups of GPRS network, effective value-added services can be achieved.

This model is given under the precondition of normal on-offline records from GPRS network. But there is a variety of unavoidable exceptions when getting online and offline. In our future work we plan to focus on these exceptions to portray a more realistic user online duration model for GPRS network environment.

5. Acknowledgment

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

6. References

- [1] GSM 09.60. Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across the Gn and Gp Interface. ETSI EN 301 347 V7.4.1 (2000-11)
- [2] 3GPP, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS); Service description; Stage 2(Release 6) Technical Specification 23.060 V6.11.0 (2005-12)
- [3] 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 (2008-09)
- [4] 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 (2009-09)
- [5] Nikolas V. Stylianides, George V. Stephanopoulos, George S. Tselikis, Michael Dopher "Signaling Performance Trials and Evaluation Results on a GPRS Platform"
- [6] Xiaoli Zhang, Yue Lai, Shengli Xie "A GTP Stateful Inspection Method Based on Network Processor" ICCS 2008
- [7] Francesco Vacirca, Fabio Ricciato, Rene' Pilz "Large-Scale RTT Measurements from an Operational UMTS/GPRS Network" WICON'05

[8] F. Tataranni, S. Porcarelli, F. Di Giandomenico, A. Bondavalli, L. Simoncini “Modeling and Analysis of the Behavior of GPRS Systems” IEEE 2001 pp. 51-58

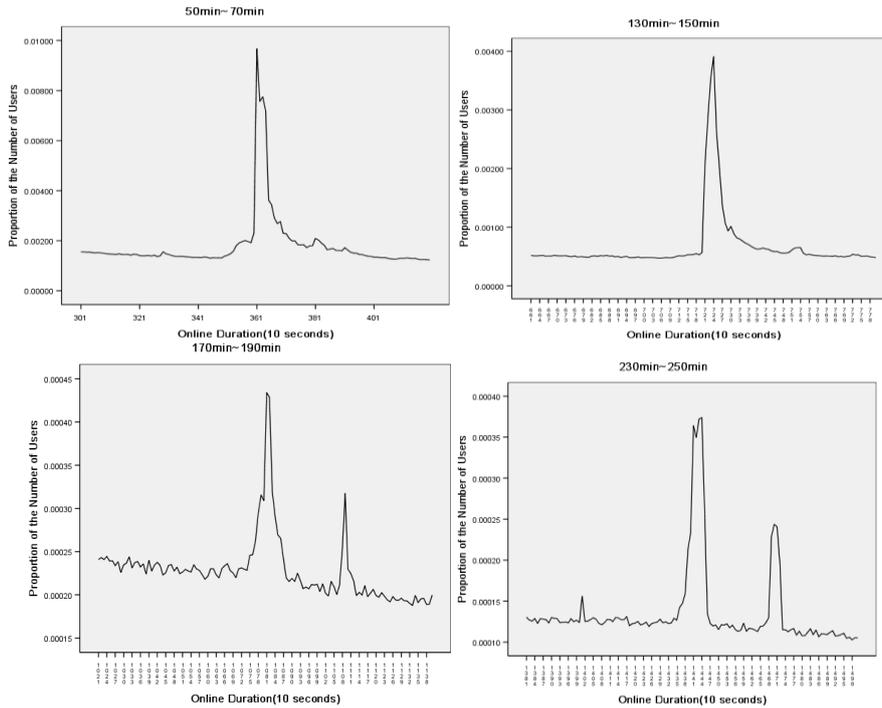


Fig. 6: Distribution of the Proportion of the Number of Users with Different Online Duration (50~70, 110~130, 170~190, 230~250 minutes)