doi:10.3772/j.issn.1006-6748.2015.04.010

Analysis on traffic characteristics and transmission efficiency of instant messager services①

Li Ke (李 克)②, Li Wenfa, Su Limin, Chen Xiaodan
(School of Information Technology, Beijing Union University, Beijing 100101, P. R. China)

Abstract

Traffic characteristics of several typical instant messager services under certain scenarios are firstly analyzed, based on real-time data collected in the commercial mobile network. Then criteria for the evaluation of the efficiency of the mobile network for the transmission of packet services are proposed in both transport layer and physical layer over air interface. The transmission efficiency of IM services is evaluated and compared under the proposed criteria. Furthermore, a so-called smart resource adaptation algorithm is verified in the effectiveness of improving the wireless transmission efficiency. Finally, improvements to the smart resource adaptation are proposed to further improve the wireless transmission efficiency, and its effectiveness is verified by the calculations.

Key words: transmission efficiency, traffic model, instant messager, keep-alive mechanism, wireless resource, TCP packet

0 Introduction

For a mobile communication system, network, terminal and service are the three key parties which interact with each other closely and have significant impact on user service perception. In recent years, with the fast growing of smart phone and commercial deployment of 3G/4G network, mobile Internet services also develop very quickly. According to the reports of CNNIC[1,2], by Jun. 2014, smart phone became the top device of internet surfing in China. That is, 83.4% of the users access the Internet through handset, which is 2.5% more than that by PC. Instead of the traditional circuit switched (CS) services like voice and short message service (SMS), packet switched (PS) services, especially a variety of mobile Internet services, are already the dominant services for the mobile subscribers. User behavior of the PS services has also changed a lot.

One major problem that is incurred by this changing of service type and service behavior is the impact of new services on the capability and transmission efficiency of mobile network. It is known that for almost all of the networks including 2G, 3G and 4G, the network is targeted to be optimized for carrying voice traffic, not mobile Internet services. The major characteristics of instant messager (IM) services are highly active, low traffic per usage and always-online. This will inevitably generate huge amount of small size packets during the service, and then results in the decreasing of transmission efficiency of the network.

Much work has been carried out to measure and investigate the traffic model of Internet services[3-6]. Pries, et al. proposed a method of service type identification and user behavior of Internet surfing via wideband radio access by passive monitoring[3]. In Ref. [4], traffic properties of Internet access over WiFi link were investigated. Since 2011, 3GPP, the well-known telecommunication standardization body, has carried out a research project called enhancements for diverse data applications (eDDA)[7-10], to investigate the traffic model of new services and its impact on the signaling burden of the mobile network. Some infrastructure vendors and operators also proposed some solutions to elevate the transmission efficiency of mobile network in carrying small packet services like IM[11,12], in which the smart resource adaptation (SRA) proposed in Ref. [11] is proved to be effective in improving the efficiency of GSM/GPRS network.

In view of this, further researches have been carried out in this paper. Firstly, the packet transmission

① Supported by the National Natural Science Foundation of China (No. 61301103), the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (No. CIT&TCD201504039), and Funding Project for Academic Human Resources Development in Beijing Union University (No. BPHR2014A03, Rk100201510).
② To whom correspondence should be addressed. E-mail: like@buu.edu.cn
Received on Mar. 17, 2015

information of IM services under typical scenarios is collected at the terminal side through active measurement. Then the key transmission properties of the IM services are achieved based on the analysis of the data collected. The definition of transmission efficiency of TCP layer and physical layer are given. By associating the data collected from terminal side and network side under typical scenarios, the transmission efficiency are then evaluated. The performance of SRA algorithm has been verified and some improvements are proposed. The proposed improvements are also proved to be able to improve the transmission efficiency through calculations.

The remainder of the paper is as follows: Section 1 introduces the method of measurement and data preprocessing in this paper; Section 2 gives the analysis of transmission properties of TCP packets of several IM services; In Sections 3 and 4, the TCP transmission efficiency and wireless transmission efficiency are investigated. Some improvements to the SRA algorithm are proposed and verified. Finally, the conclusions are drawn in Section 5.

1 Measurement methodology

In this section, description of the measurement scenario and how the data is collected during the measurement are presented, for the investigation of traffic properties and TCP/wireless transmission efficiency.

1.1 Measurement scenario

In general, there are two ways of traffic and network measurement: passive monitoring and active testing. The active testing under a fully controlled and transparent environment is employed in our study, in order to build a more accurate mapping relationship between the user behavior and relevant network parameters.

The measurement is conducted in cooperation with a mobile network operator of China. In our investigation, WeChat (the so-called Weixin within China) and QQ are selected as the target IM applications for our measurement, because of their dominant position in IM application market in China. According to the statistics of CNNIC in 2014[2], the penetration rate of QQ and WeChat are ranking top 2 among all the IM applications (i. e. 77.8% and 65%, respectively). And the usage of QQ and WeChat is very much proactive, from the viewpoints of average start times per day, average duration per usage to traffic volume.

As a supplementary to our study, Fetion, an IM service operated by the abovementioned operator of China, is also selected for the measurement, since it also has huge amount of users among the subscribers of that operator.

Fig. 1 gives the description of the measurement scenario.

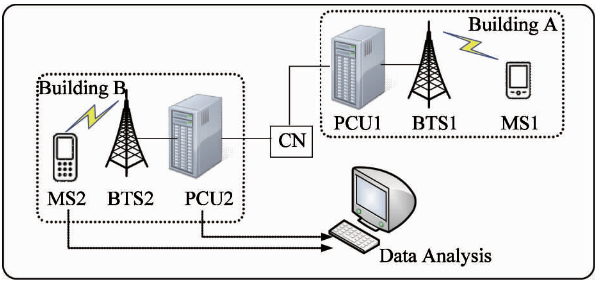


Fig. 1 Measurement setup

One test engineer in Building A sends messages or files via the target IM service of mobile station (MS) 1 to MS 2 held by another test engineer in Building B, by simulating the real behavior of general subscribers. The network under measurement is EDGE, and the MS is smart phone with Android OS. The measurements are conducted in four cases, that is,

Case 1: Short message chatting

Test engineer in Building A sends text messages continuously via the IM APP on MS 1, to MS 2 held by test engineer in Building B. Each message consists of 10 Chinese characters (i. e. 20 Bytes). Each test shall last no less than 1 hour.

Case 2: Long message chatting

The only difference to Case 1 is that the message consists of 100 Chinese characters (i. e. 200 Bytes) instead.

Case 3: Large file downloading

Test engineer sends a very large file from IM of MS 1 to MS 2. Each test shall be no less than 1 hour.

Case 4: Idle status

Test engineer keeps running the IM APP on MS 2, without any operation, for at least 2 hours.

The information collected during the tests includes:

(1) TCP packet information

The information is collected by packet sniffer software running on MS 2 during the test, by monitoring all the TCP packets passing through the network interface card. Here the TCPdump based on Linux (one of the most popular sniffers) is employed in our test. The information includes packet arrival time, packet size, type of protocol, source IP, target IP, type of content, etc. The collected data is saved on the SD card of the handset, as a .pcap file.

(2) Allocation of air interface resources

The knowledge of the air interface resources assignment is detected from network side. As we know, in a GSM/EDGE network, the packet control unit (PCU) is responsible for the assignment of air resources for all the users in packet service domain. Generally, PCU is co-located together with the base station controller (BSC). During the measurement, we can log into PCU 2 with the operation and maintenance software, to collect the downlink/uplink (DL/UL) packet data channel (PDCH) assignment information for all the users that has a temporary block flow (TBF) configuration in the target cell of base transceiver station (BTS) 2. The assignment information includes BTS ID, TRX ID, Direction (UL/DL), IMSI (subscriber ID), Date/time, TBF ID, PDCH/timeslot assigned, coding scheme and Traffic class, etc. Here a TBF refers to a physical connection between MS and BTS, which is setup for the transmission of packet data over air interface. The TBF is released once the packet transmission is completed.

Because of the multiple-process mechanism deployed in the Android OS, there may have some other applications running at the background during the measurement of IM service. In this case, additional traffic demand may be generated by these background applications and thus impact on the accuracy of the assignment information collected for the IM service. To mitigate this influence, all the processes of the 3rd party applications before the measurement will be removed.

1.2 Measurement duration

The measurement lasts from Oct. 2013 to Mar. 2014. To collect samples with sufficient statistical, each measurement was carried out for more than 1 hour and repeated in different time (working day, non-working day, day time and night). For each test case of each IM service, totally more than 10 hours of data are collected. The total number of packets captured during all the measurements is 499690, and 35600 air interface assignment information are captured in PCU. With this, sufficient number of samples are ready for the analysis hereafter.

1.3 Data pre-processing

For the raw data collected in the measurements, a pre-processing is carried out firstly, which includes identification of the target users and packets specific for the IM services, and the association of the assignment information and packet information based on the time stamps.

Wireshark, a well-known packet sniffer software, is employed in our analysis of traffic properties. And then, Matlab is utilized to analyze the transmission resource assignment and transmission efficiency.

(1) Identification of target users

The information acquired in PCU 2 includes the air resource assignment information of all the users in the target cell, in which only the information of the target user (i. e. MS 2) is of our interests. Then data of the target user are screened out by its IMSI.

(2) Identification and classification of IM packets

The service packets have to be classified after the raw data is collected. Currently there are two techniques[3] to identify and classify packets: (1) payload-based classification, (2) Host behavior classification.

The payload-based technique is to identify the keywords of the packet payload based on a deep packet inspection (DPI)-like methodology. One main drawback of this technique is the granularity of the service classification. In some cases, only the service type can be identified (e. g. HTTP), not the exact application (e. g. UCweb, or Opera browser). In addition, the character strings of the payload are not always available or the payload may be encrypted.

Due to the limitation of the payload-based classification, Karagiannis et al.[13] proposed a host behavior classification method called BLINC. The focus is shifted from classifying flows to associating hosts with applications. The flows are then classified accordingly.

In this paper, the second technique is employed. Firstly, the pcap file is opened with Wireshark. Direction of the packets can then be identified by the destination IP address. That is, for the UL (DL) packets, the source (destination) IP in the head shall be private address like 127. 0. 0. 1. The group of servers for a certain IM service provider in a certain area is always static. The application that the packet belongs to can then be identified, by analyzing the source/destination IP address. Packets with the similar quintuple can then be grouped together for the analysis hereafter.

(3) Association of data from different sources based on the aligned time stamps

The data collected from the handset and PCU can be associated by aligning the time stamps within the data. By this, we can easily set up the relationship between the operations on the handset, the packets sent and received, and the wireless resources assigned for the packet transmission.

2 Analysis of TCP packet transmission properties

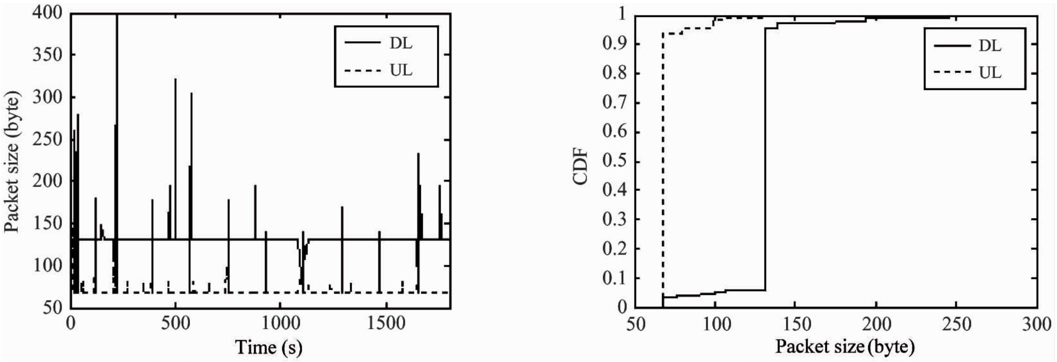
Instant messager traffic comprises a mix of user plane packets conveying the text or multimedia data, along with application layer and/or transport layer protocol signaling to verify message delivery status. Some instant messager applications may also have keep-alive behaviors (pulse-like packets) and may generate other background traffic specific to the application.

For the packets transmission in all the scenarios, our analysis is mainly on the DL packets received at the terminal and the UL packets conveying the feedback information to the server. Key factors considered in the research include packet size, inter-arrival time, and keep-alive mechanism in idle state.

2.1 QQ

Case 1: Short message chatting

Firslty, the distribution of the DL/UL packet size is analyzed and presented in Fig. 2, in which Fig. 2(a) is the time series curves of packet size, and Fig. 2(b) is the cumulative density function (CDF) of DL/UL packets, respectively.



(a) Time series of packet size (b) CDF of packet size

Fig. 2 Distribution of packet size (QQ, Case 1)

From Fig. 2(a) it can be seen that, at the beginning of the chatting, there is a short period of frequent interactions between the client and the server. After that, there is only the quasi-static transmission of packets in DL (generally 131 bytes of data for each packet) and UL (68 bytes of acknowledgement to the server). It is more clear from the CDF in Fig. 2(b).

In addition, an interesting phenominon found in Fig. 2(a) is that, for every 180s there is always a pair of packets in DL (68 bytes) and UL (99 bytes) transmitted. Together with the analysis in Case 4 mentioned below, it can be seen that acturally they are keep-alive packets with fixed period of 180s. That means, the keep-alive mechanism exists not only in idle state, but in transmission state as well.

The CDF of inter-arrival time of packets is illustrated in Fig. 3 below. It can be seen that, the distribution is the same for both DL and UL, mostly in the scope of 1 ~ 5 seconds, which is identical to the interval of operations during the test.

Case 2: Long message chatting

Fig. 4 gives the distribution of packet size in this case. It can be seen in the figure that, most of the DL packets is of 307 bytes which is correspondent to the 200 bytes chatting message, along with a 68 byte of acknowledgement in the UL. Again, the typical keep-alive packets is observed with the same characteristics. That is, every 180s, a 99 bytes UL packet and a 68 bytes DL packet are sent to update the status of both sides.

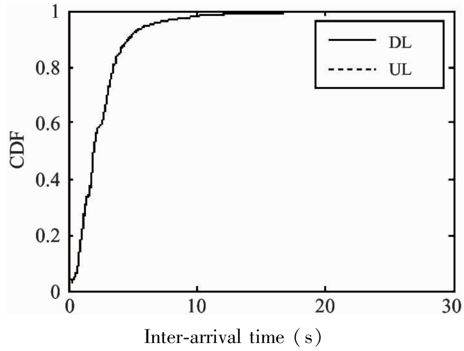


Fig. 3 Inter-arrival time of DL/UL packets (QQ, Case 1)

Case 3: Download of large file

The properties of packet size and inter-arrival time are given in Fig. 5.

It can be seen that, most of the TCP packets in DL is of 1450 bytes during the download. This is because the large packets from the application layer is segmented in TCP layer to be adapted to the tranmission over IP. Here 1450 byts corresponds to the upper bound of maximum transmission unit (MTU) set in TCP. In the UL, there is always a 68 bytes packet as an acknowlegement of the reception. Besides, Fig. 5(b) shows that the inter-arrival time is very small during the downloading.

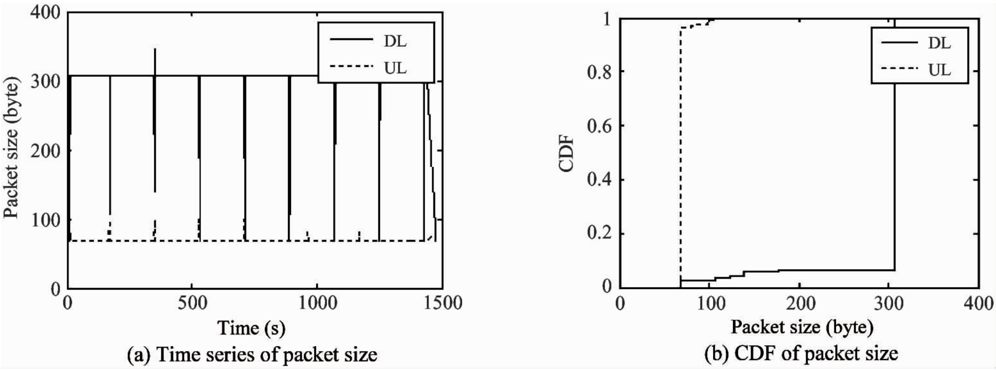


Fig. 4 Distribution of packet size (QQ, Case 2)

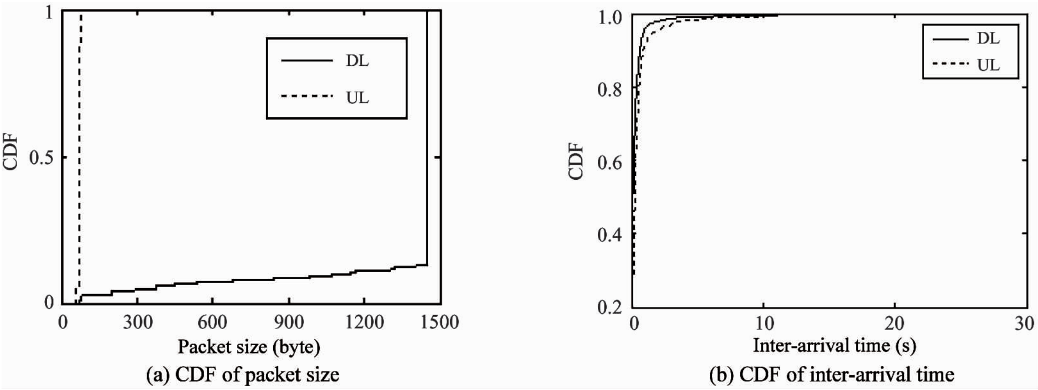


Fig. 5 CDF of packet size and inter-arrival time (QQ, Case 3)

Case 4: Idle state

There is no packets of application layer transmitted between the users in idle state. To keep the connection available for a fast response to the transmission demand and update the status of the users, a keep-alive mechanism is always undertaken, which transmits small packets from time to time between the users and the server. Fig. 6 presents the distribution of the packet size in this case.

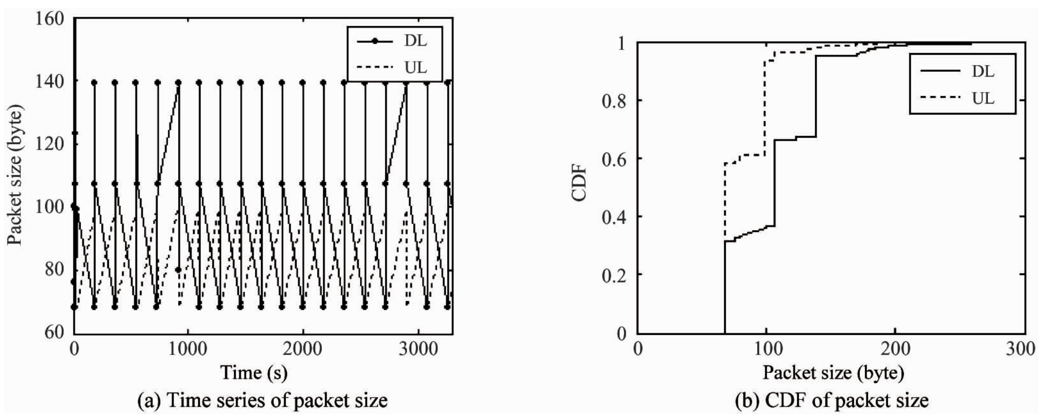


Fig. 6 Distribution of packet size (QQ, Case 4)

A typical keep-alive mechanism can be observed in Fig. 6. That is, every 180s, the terminal sends a group of packets (generally 3 packets in one group, with 99, 68, 68 bytes each) to the server. As the response to that, another group of packets are sent back in the DL (either 3 packets of 68, 139 and 107 bytes, or 2 packets of 139 and 107 bytes). The property of periodic keep-alive mechnism is more clear from the CDF of inter-arrival time in Fig. 7.

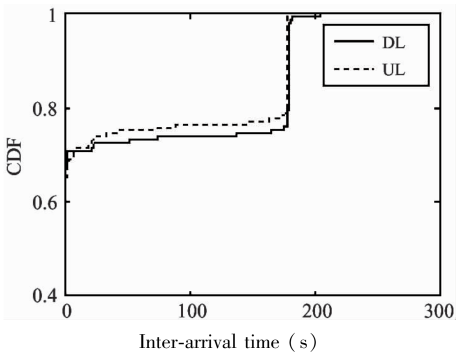


Fig. 7 CDF of the inter-arrival time (QQ, Case4)

2.2 WeChat

The measurement and analysis of WeChat has been conducted for Case 1, 2 and 4.

Fig. 8 shows the distribution of packet size in Case 1. In spite of the small signaling packets exchanged in UL/DL, most of the DL packets are the encapsulated packets for the short message. For each message, 2 TCP packets of 315 and 56 bytes are sent in DL, and correspondently two packets (56 and 258 bytes) are sent in UL.

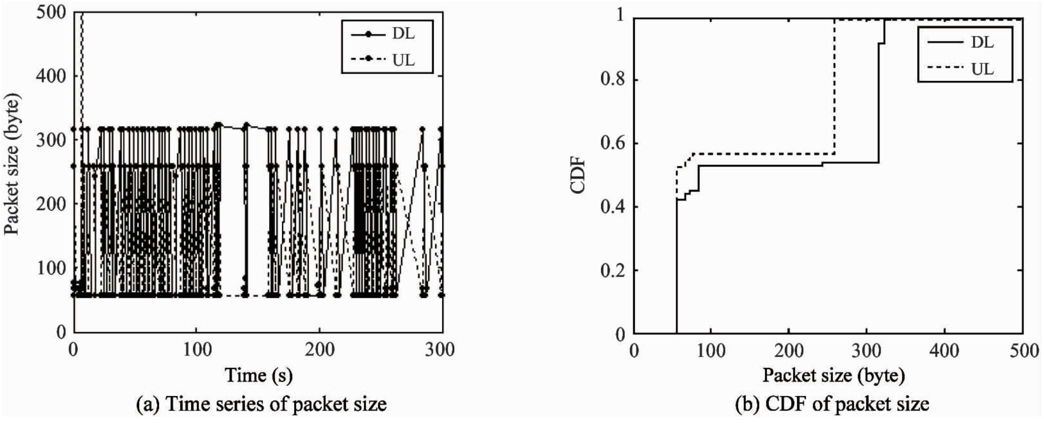


Fig. 8 Distribution of packet size (WeChat, Case 1)

The CDF of packet size for Case 2 is shown in Fig. 9. It is seen that a message of 200 bytes in the application layer is corresponding to two DL data packets in the TCP layer (331 and 56 bytes) and two acknowledgement packets in UL (56 and 258 bytes), respectively.

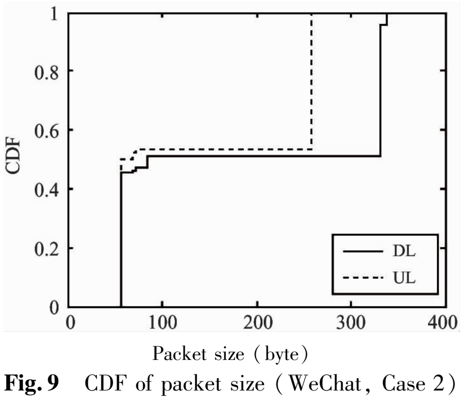


Fig. 9 CDF of packet size (WeChat, Case 2)

For Case 4 (idle state), the CDF of packet size and inter-arrival time is shown in Fig. 10. In this case, most of the DL packets are of 68 bytes and most of the UL packets are of 56 or 76 bytes. Unlike QQ, the CDF of inter - arrival time here shows no obvious periodical keep-alive mechnism. Most of the packets are of very short interval, to keep the connection alive and update the status.

2.3 Fetion

Fig. 11 gives the CDF of packet size in Case 1 and 2. It shows that, the reception of a 20 byte short message results in two DL TCP packets (68 and 340 bytes) and one UL TCP packet of 95 or 68 bytes, while the reception of a 200 byte short message results in two DL TCP packets (68 and 370 bytes) and one UL TCP packet of 95 or 68 bytes.

The statistics of packet size and inter-arrival time of TCP packets in the idle state is illustrated in Fig. 12.

From Fig. 12(b) it is obvious that many packets with large size are transmitted in this case. The size are mainly within 400 ~ 600 bytes in UL, and 700 ~ 1000 bytes in DL. A sophisticated but stable keep-alive pattern can be observed in Fig. 12(a) & (c), where the inter-arrival time of the DL keep-alive packets is around 70, 130 and 320 seconds. On the contrary, the inter-arrival time of the UL keep-alive packets are quite scattering.

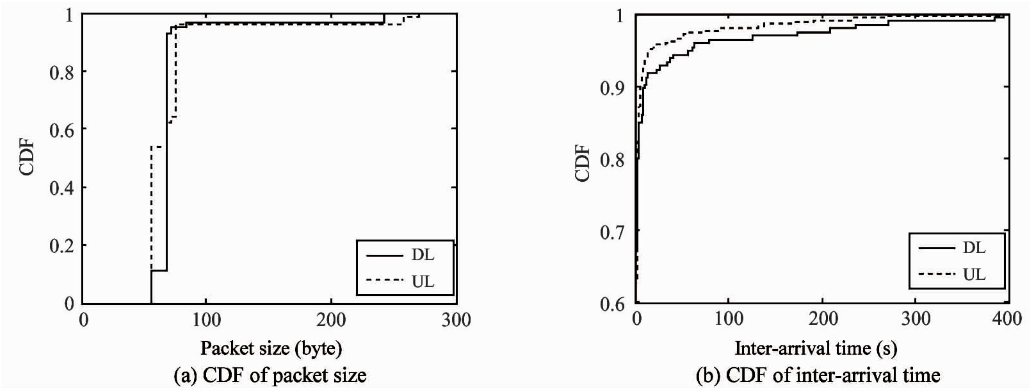


Fig. 10 CDF of packet size and inter-arrival time (WeChat, Case 4)

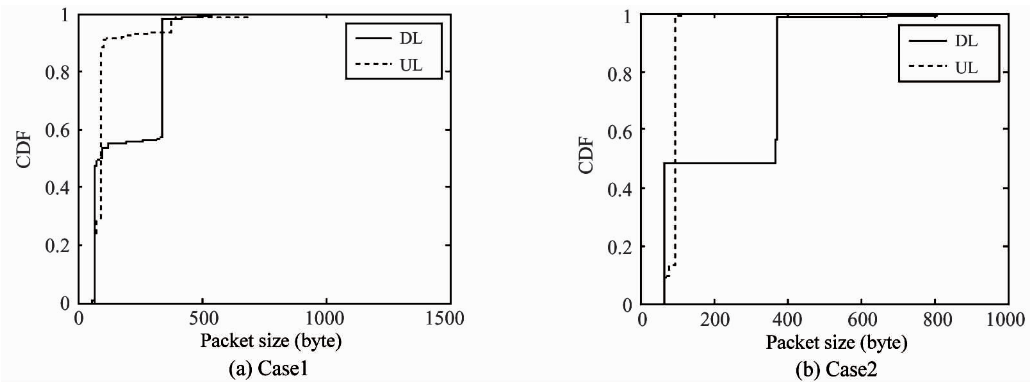


Fig. 11 CDF of packet size (Fetion, Case 1&2)

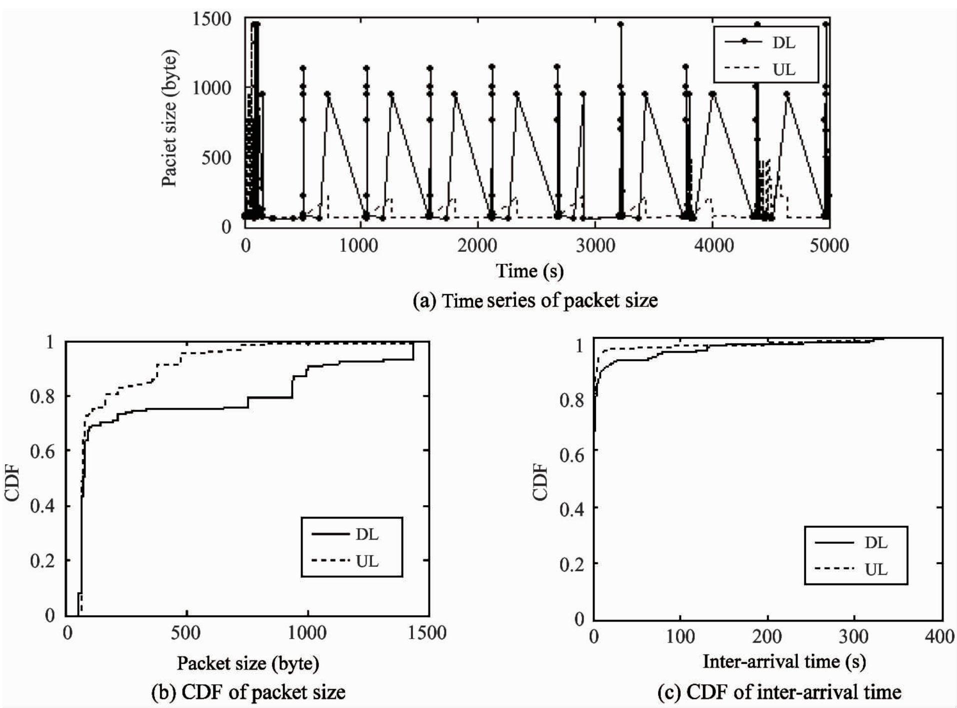


Fig. 12 CDF of packet size and inter-arrival time (Fetion, Case 4)

3 Analysis of TCP transmission efficiency

To evaluate the transmission efficiency of IM payload in the TCP layer, a so-called "DL TCP transmission efficiency" is defined as the total size of DL payload transmitted in the application layer at the expense of a unitary amount of TCP packets transmitted in DL, that is,

$$\eta_D = \frac{T_{AD}}{T_{TD}} \quad (1)$$

in which T_{AD} is the total amount of bytes received by the terminal in the application layer, and T_{TD} is the total amount of bytes transmitted in the DL of TCP layer.

By considering the fact that, during the reception of message in the DL, additional traffic is always generated in the UL as the acknowledgement to the message reception. Therefore, another index, "total TCP transmission efficiency", is defined below, as a more detailed reflection of the resource consumptions in the TCP layer,

$$\eta_A = \frac{T_{AD}}{T_{TD} + T_{TU}} \quad (2)$$

in which T_{TU} is the total amount of bytes transmitted in the UL of TCP layer during the message reception in DL.

By aligning the time stamps of the data collected, one can associate every TCP packet with the corresponding message in the application layer. Based on the analysis of Section 2, the following summary of TCP packet characterics and the estimation of TCP transmission efficiency are got in Table 1.

Table 1 Estimation of TCP transmission efficiency

Scenario	APP message size(B)	DL TCP packet size(B)	UL TCP packet size(B)	$\tilde{\eta}_D$ (%)	$\tilde{\eta}_A$ (%)
QQ Case1	20	131	68	15.3	10.1
QQ Case2	200	307	68	65.1	53.3
WeChat Case1	20	371	314	5.4	2.9
WeChat Case2	200	387	314	51.7	28.5
Fetion Case1	20	408	95	4.9	4.0
Fetion Case2	200	438	95	45.7	37.5

Based on the real data collected in the measurements, the following calculation results of TCP transmission efficiency are presented as illustrated in Table 2.

Table 2 Measurement of TCP transmission efficiency

Scenario	No. of messages received	T_{TD} (kb)	T_{TU} (kb)	η_D (%)	η_A (%)
QQ Case1	14888	2113	1139	13.8	8.9
QQ Case2	11952	3667	865	63.7	51.5
QQ Case3	–	41230	1444	95.5	92.3
WeChat Case1	5460	2108	1862	5.1	2.7
WeChat Case2	7155	2748	2224	50.9	28.1
Fetion Case1	14080	6515	2978	4.2	2.9
Fetion Case2	13638	6218	1587	42.8	34.1

By comparing Tables 1 and 2, it can be seen that, the measurement of TCP transmission efficiency is only slightly lower to the estimation. This is mainly because of the additional signalling packets to be transmitted before and during the message tranmission.

For the short message chatting, the efficiency of all the three services are much lower than that of the long message chatting. While for both Case 1 and Case 2, the TCP transmission efficiency of WeChat and Fetion is significantly lower than that of the QQ service. Therefore, the implementation of WeChat and Fetion needs to be further optimized to improve their transmission efficiency in the TCP layer.

4 Analysis of wireless transmission efficiency

4.1 Optimization of wireless transmission efficiency

As we know, that for the GPRS/EDGE system, the assignment of DL air interface resources in the packet domain is in the granularity of PDCH (i. e. timeslot). Generally, the physical layer resources over the air interface are assigned based on the capability of the terminal (most of the GSM handsets available on the market support the simutaneous transmission over 4 timeslots). Even if there is only tens of bytes to be transmitted, 4 timeslots are assigned to the terminal. And the TBF link set up for the packet data transmission will only be released after a certain time delay, even though the transmission is completed. This inevitably leads to the TBF blocking and unnecessary occupation of wireless resources, and finally the low efficiency of air resource utilization.

With the fast increasing of IM services usage among smart phone users, the situation is getting much worse, since the IM services generate huge amount of small size packets. In addition, the keep-alive mechnism popularly adopted by IM services demands a lot more small size packets transmission when the IM service is in idle state.

To improve the low transmission efficiency resulting from the small size packets, some infrastructure vendors propose optimization algorithms. Of those, the Smart Resrouce Adaption (SRA) proposed by an European vendor[11] is quite interesting. The main idea of this algorithm, is to allocate the timeslots adaptively according to the packet size. That is, for packets smaller than the pre-defined threshold (mainly conveying IM payloads and TCP control signalling), only 1 timeslot is allocated to the TBF link. In case the packet size is above the threshold (mainly for services like HTTP or other TCP payload), timeslots are allocated according to the capability of the terminal (i. e. 4 timeslots). Here the threshold is 250 bytes.

With this, the block rate of TBF setup can be reduced significantly and thus the transmission efficiency is improved to the same extent. An on-site measurement shows that, the PDCH transmission efficiency can be improved around 9% [12]. Here the PDCH transmission efficiency is defined as: PDCH transmission efficiency = PDCH throughput at busy hour / (average number of PDCH allocated in busy hour × 3600s) × 100%. PDCH throughput refers to the traffic in RLC sublayer of the air interface. It is a good measure to the transmission capability of traffic channel in busy hours. According to the data in Ref. [12], the average PDCH transmission efficiency of GPRS network nation-wide is 4.69kbps.

As an approximation to the PDCH transmission efficiency, here a new index is defined, 'DL wireless transmission efficiency', as

$$\eta_{WD} = \frac{T_{TD}}{N_{PD} \cdot t} \tag{3}$$

in which N_{PD} is the average DL PDCH allocated, t is the total time of transmission. It shall be noted that, the throughput here is in TCP layer, which is different from the definition of PDCH transmission efficiency. Furtherly, by counting the UL PDCH resources consumption during the reception of messages in DL, the total wireless transmission efficiency is defined as

$$\eta_{WA} = \frac{T_{TD}}{(N_{PD} + N_{PU}) \cdot t} \tag{4}$$

in which N_{PU} is the average UL PDCH allocated.

4.2 Measurements and further improvements

To verify the above-mentioned SRA algorithm, some measurements are conducted in three cases (Case 1, 2, and 3) for the QQ service. By aligning the time stamp of the packet information and the allocation of wireless resources, then the knowledge of the number of DL timeslots allocated exactly for each TCP packet is got.

Fig. 13 presents the mapping of timeslot allocation and the TCP packet size in time domain. It can be seen clearly that, SRA algorithm takes effects in all the cases. For Case 1, mostly 1 timeslot is assigned since the packet size is always 131 bytes. While for Case 2 and 3, mostly 4 timeslots are assigned. In contrast to the conventional mechnism of resource allocation, SRA is proved to be able to improve the wireless transmission efficiency.

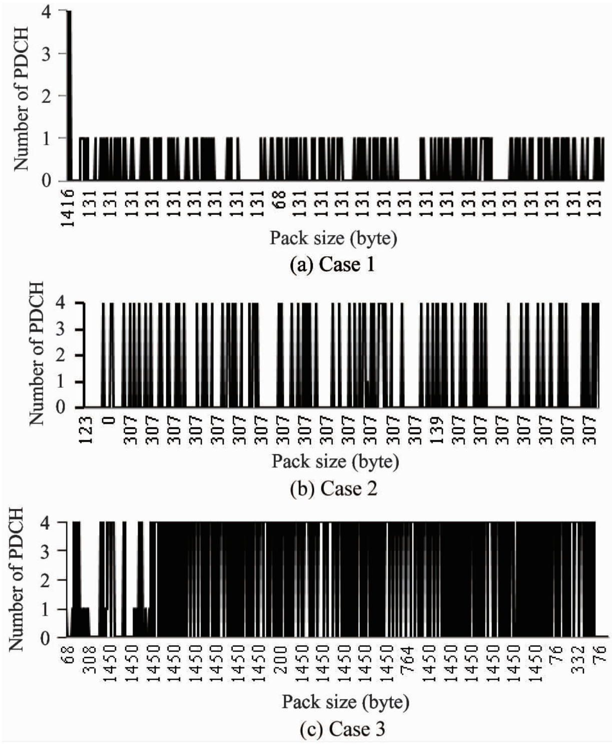


Fig. 13 Time series of timeslot allocation and TCP packet size (QQ)

The calculated wireless transmission efficiency (unit: kbps) for these cases are listed in Table 3.

Table 3 Comparison of wireless transmission effficiency (QQ)

	T_{TD} (kb)	N_{PD}	N_{PU}	t (s)	$\tilde{\eta}_{WD}$ (kbps)	η_{WD} (kbps)	η_{WA} (kbps)
Case1	16888	1.05	0.27	9164	0.46	1.76	1.40
Case2	29334	3.95	0.34	6936	1.06	1.07	0.99
Case3	329844	3.98	0.29	17200	4.79	4.82	4.49

In the table, $\tilde{\eta}_{WD}$ refers to the DL transmission efficiency in case SRA is not deployed (suppose N_{PD} = 4). It can be seen that, by applying the SRA algorithm, the efficiency for Case 1 is improved significantly (i. e. from 0.46Kbps to 1.76Kbps), whilst there is no big change for Case2 and Case3 since most of the packets in these cases are above the threshold. Especially, the transmission efficiency of Case 2 with SRA applied (0.99kbps) is much less than that of Case 1 (1.40kbps), since the packet size in Case 1 is always below the threshold thus requests much less PDCH resources than in Case 2.

Further more, since the terminal is assigned 4 timeslots for the transmission of packet with 1450 bytes, then approximately 1 timeslot is sufficient for the transmission of a 362 byte packet. By considering this, the following improvements to the SRA algorithm is proposed:

(1) adjust the threshold of SRA, from the original 250 bytes to 320 bytes for one timeslot;

(2) change 2-interval judgement into 4-interval judgement, i. e., using 3 threshold (320, 640 and 960 bytes) for the judgement. Suppose the TCP packet size is x, then the number of timeslots n to be allocated is

$$n = \min\left(\left\lceil \frac{x}{320} \right\rceil, 4\right) \tag{5}$$

where $\lceil\ \rceil$ denotes rounding towards plus infinity.

To verify the effectiveness of the proposed improvements, the transmission efficiency is re-calculated for Case 2 of QQ, based on the proposed algorithm and the raw data collected from the PCU. The results are compared with that of the SRA and the original method (Table 4). The table shows that, the wireless transmission efficiency can be improved significantly by applying the proposed improvements to the SRA algorithm, that is, from 0.99 to 2.47kbps. Therefore, in addition to small size and large size packet cases (Case 1&3), the proposed improvement of SRA is proved to be applicable to middle size packet cases (Case 2 and simular cases) as well.

Table 4 Comparison of wireless transmission efficiency (QQ, Case 2)

	T_{TD} (kb)	N_{PD}	N_{PU}	t (s)	η_{WD} (kbps)	η_{WA} (kbps)
Original	29334	4	1	6936	1.06	0.85
SRA	29334	3.95	0.34	6936	1.07	0.99
Proposed improvement	29334	1.37	0.34	6936	3.09	2.47

5 Conclusions and suggestions

In this paper, the transmission properties of TCP layer packets of IM services have been analyzed in several typical scenarios. The research is conducted for three IM services which are widely used by smart phone users in China, namely, QQ, WeChat and Fetion.

From the research, it is found that QQ deploys a regular keep-alive mechanism during the idle state which is able to guarantee a satisfactory user perception at the expense of very small consumption of transmission resources. This mechanism takes effect not only in idle state, but also in the transmission state. That is, every 180 seconds, a group of 2 ~ 3 packets are transmitted in both UL and DL. During the transmission of payload, the distribution of packet size for UL acknowledgement is nearly the same for all the scenarios in the measurements, while the packet size of the DL payload is somewhat proportional with respect to the size of the message in the application layer.

There is no periodical keep-alive mechanism observed for WeChat in the idle state. Instead, many small size packets are exchanged frequently between handset and the server, to keep the TBF alive and update the user status, which then consumes lots of air interface resources. During the transmission of payload, the distribution of packet size in UL and DL is quite similar. Even during the reception of messages in DL, around 50% of the packets are signaling instead of payloads, which consumes a lot more resources than QQ.

For the Fetion service in idle state, there are frequent interactions of packets between handset and the server. A quite sophisticated but stable keep-alive pattern can be observed. For the transmission of payload, the distribution of packet size is similar to QQ in UL, while similar to WeChat in DL.

Then the transmission efficiency of TCP layer is analyzed, by both estimation and measurement based on real data collected in the GPRS/EDGE network. It shows that, the transmission efficiency of the three services in the short message chatting case is significantly lower than that of long message chatting case. The efficiency of WeChat and Fetion is much lower than that of QQ for both cases. Therefore, the WeChat and Fetion need further optimization to improve their TCP transmission efficiency.

Taking QQ as an example, the efficiency of transmission over the GPRS/EDGE air interface is analyzed further. It can be seen that, the efficiency in the short message chatting case is much lower than other cases, mainly because of the waste of resources for large amount of small packets in this case. By applying SRA in PCU, it is proved that the wireless transmission efficiency of short message chatting case can be increased largely by more reasonable allocation of timeslots — even larger than the long message chatting case.

Finally, further improvements to the SRA algorithm is proposed, considering that SRA is still not so effective for long message chatting cases. The calculation results prove the effectiveness of the proposed method.

In the future, the proposed method will be further verified in the real network by cooperating with local operator. And the transmission efficiency of IM services in the state-of-the-are mobile network, say, LTE, is to be analyzed and optimized.

References

[1] China Internet Network Information Center. Report on the Investigation and Research of Mobile Internet of China. Beijing: CNNIC, 2014

[2] China Internet Network Information Center. Investigation Report on the User Behavior of Mobile Messaging Applications in China, Beijing: CNNIC, 2014

[3] Pries R, Wamser F, Staehle D. Traffic measurement and analysis of a broadband wireless internet access, In: Proceedings of IEEE Video Technology Conference, Barcelona, Spain, 2009. 1-5

[4] Gebert S, Pries R, Schlosser D. Internet access traffic measurement and analysis, In: Proceedings of the Traffic Monitoring and Analysis workshop, Vienna, Austria, 2012. 29-42

[5] Lu K, Chen W, Yang D. A novel approach for efficient and effective mining of mobile user behaviors, In: Proceedings of the 4th IEEE International Conference on Multimedia and Ubiquitous Engineering, San Diego, USA, 2010. 1-6

[6] Fukuda K, Cho K, Esake H. The impact of residential broadband traffic on Japanese ISP backbones, *ACM SIGCOMM Computer Communication Review*, 2005, 35(1): 15-21

[7] 3GPP. RP-120256: work item description: LTE RAN enhancements for diverse data applications (eDDA), In: Proceedings of 3GPP RAN Meeting #55, Xiamen, China, 2012. 1-6

[8] 3GPP. TR 36.822. LTE RAN Enhancements for Diverse Data Applications, Valbonne, France: 3GPP, 2012

[9] Huawei, HiSilicon, R2-115239: Analysis on instant message traffic. In: Proceedings of 3GPP RAN2 Meeting #75bis, Zhuhai, China, 2011. 1-4

[10] Samsung, R2-115432: Analysis of cause of frequent keep-alive messages, In: Proceedings of 3GPP RAN2 Meeting #75bis, Zhuhai, China. 2011. 1-3

[11] Nokia Siemens Network, Smart GSM networks require smarter packet resource allocation. http://networks.nokia.com/print/9561, Nokia: 2011

[12] Wang L, Li Y. Impact of GSM network PDCH carrying efficiency on user's perception. *Telecom Engineering Technics and Standardization*, 2013(3):14-16

[13] Karagiannis T, Papagiannaki K, Faloutsos M. BLINC: multilevel traffic classification in the dark, In: Proceedings of 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, New York, USA, 2005. 229-240

Li Ke, born in 1972. He received his Ph. D. degree in Radio Engineering Department of Southeast University in 2000. He also received his B. S. and M. S. degrees from Changchun College of Geology in 1993 and 1996 respectively. His research interests include mobile Internet service analysis, QoS and QoE modelling, antenna array processing, etc.