2015 The 5th International Workshop on Computer Science and Engineering
doi: 10.18178/wcse.2015.04.084

Understanding Latency in Cellular-based Mobile Internet

Guofeng Zhao[1,a], Chunyan Cao[1], Dan Li[1], Chuan Xu[1]

[1.]Key Laboratory of Optical Communication and Networks, CQUPT, China

[a.]Zhaogf@cqupt.edu.cn

Keywords: Cellular network, mobile Internet, gateway, latency.

Abstract. In this paper, we focus on investigating latency in internal parts of the mobile Internet ecosystem based on real data sets. We reveal and characterize three kinds of latencies: gateway processing delay, round-trip delay, and inter-ISP delay. We found that the gateway-processing delay fits well with the lognormal distribution. Besides, to decrease inter-ISP delay, we suggest mobile carriers should improve the capacity and performance of their infrastructure, enhance the level of cooperation among them and apply some specific policy such as caching strategy.

Introduction

In recent years, we have witnessed an explosive growth of mobile Internet which will continue to grow as technology and application availabilities further improve. To best serve their customers, mobile carriers and application/content providers need to understand the performance of cellular-based mobile Internet in that they hunt for better resource planning, bandwidth allocation, network design and new application development.

There have been a few works that have studied network performance in cellular data network[1-3]. Basically, latency or delay is a key QoS performance parameter. Most existing techniques, no matter active probing[4] or passive analyses[5] can infer latency in network. However, active probing may only obtain the round-trip delay, which is the sum up of latencies in the internal parts of a whole system. Besides, using passive analyses may have the limitations that the measurement covers only limited scale areas.

Unlike the wired Internet, mobile Internet has different infrastructure, as shown in figure 1(a). On behalf of better understanding what is going to happen when a subscriber surfs Internet by mobile device, we abstract the architecture of cellular-based mobile Internet and show the communication process in figure 1(b). Before a mobile device connects to Internet, in advance, it has to connect to the GGSN gateway through cellular network. The gateway is responsible for DNS resolution, IP address assignment, QoS control, etc. Therefore, it can be seen as a bridge which connecting the cellular network and the wired Internet.

Accordingly, from the viewpoint of mobile device, the total round-trip latency includes three parts: cellular network delay, gateway-processing delay and round-trip delay over the wired Internet. Thus, the way of active probing or some special passive measurement techniques could derive round-trip delay; however, it is hard to reversely infer latencies in different parts of the ecosystem.

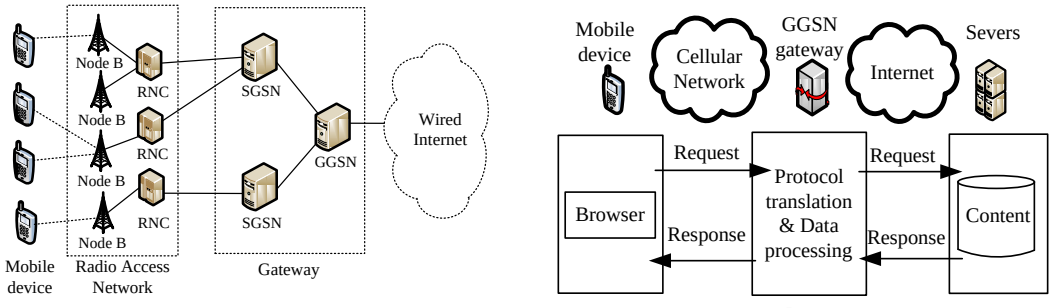


Fig.1: Architecture of Cellular-based Mobile Internet Ecosystem

In this paper, we aim to reveal and understand latencies in different parts of the mobile Internet ecosystem. They are gateway-processing delay, round-trip delay over wired Internet, and inter-ISP delay. Our study is based on real data sets obtained from a GGSN gateway that belongs to a major mobile carrier with more than five million subscribers. Based on the in-depth analysis of the data sets, we have some interesting findings and make the following contributions.

1. We found that the gateway-processing delay fits the lognormal distribution very well. This finding will benefit mobile operators and researchers who study cache policy or schedule strategy at the gateway. It is also very helpful for content providers and new application developers; that means they should concern such delay carefully in their system optimization or new applications design.
2. The Inter-ISP delay means the delay caused when mobile users belong to an ISP visiting another ISP. We analyze the Inter-ISP delay between five local ISPs, and found that the inter-ISP delay varies great from each other. We think that the level of cooperation between ISPs matters great.
3. Previous works show that the average round-trip latency in wired Internet is often tens milliseconds[6-7]; the median round-trip latency in 3G cellular network is with hundreds milliseconds[8-9]. However, interestingly, our results show that the average round-trip latency in the wired part of mobile Internet is also about hundreds milliseconds and even thousands milliseconds when visiting runs across ISPs.

The rest of the paper proceeds as follows. In section 2, we describe the data sets used in this study, and from the perspective of the GGSN gateway, we explain the data processing methodology for our analysis. In section 3, we present and analyze the results, including the overview of four delaysT1-T4 (viewed at the gateway), inter-ISP delay (delay between different ISPs). Finally, we conclude the paper in section 4.

Data Description

Data Source.

Our study is based on real transaction logs collected from a GGSN gateway, which belongs to a major mobile carrier who has more than five million subscribers in Chongqing province, China. One of the data sets covers mobile users' Internet access activities during a week (5th to 11th) in April 2010, and the other spans a week (4th to 10th) in April 2011 accordingly. each line of record in the logs is corresponding to a mobile user's request or response from web servers, containing fields of Calling Number, Client IP Address assigned, URL, Domain, Incoming Request Time, Outgoing Request Time, Incoming Response Time and Outgoing Response Time, etc. Table 1 shows the overview of the two data sets.

Table 1: Overview of the data sets

Data Sets	D1(2010)	D2(2011)	D2/D1
Period	Apr.5-Apr. 11	Apr. 4-Apr. 10	
Num. of Users	80,690	274,808	3.4 times
Num. of Requests	14,160,663	112,240,849	7.9 times
Data Size	16.1GB	122GB	7.5 times

Methodology.

In our datasets, each line of record in the log files includes four time-stamps which observed at the gateway. We model the gateway as shown in figure 2, and we have four kinds of delays - T1, T2, T3 and T4, described in table 2.

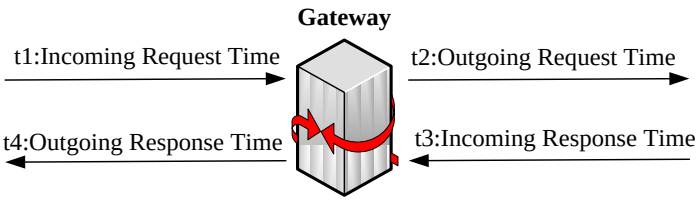


Fig.2: Time-stamps logged at the gateway

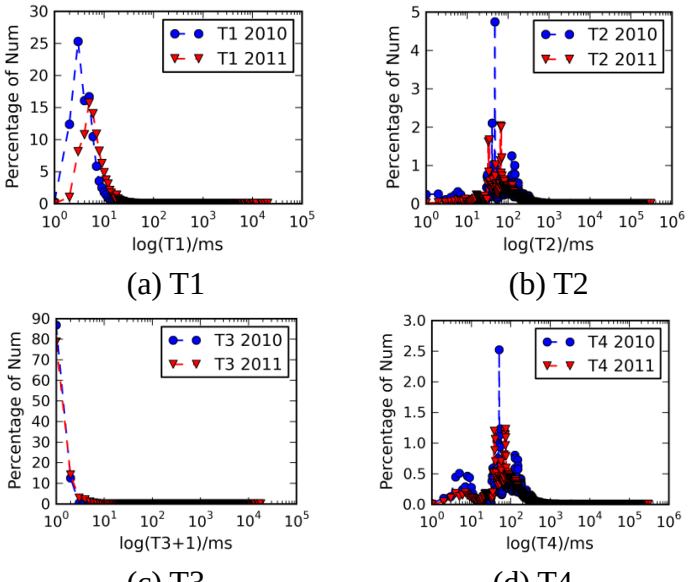
Table 2: Delays we studied at the gateway

T1=t2-t1	DNS resolution and IP binding time
T2=t3-t2	Round-trip delay in wired Internet
T3=t4-t3	Response-packet processing time at gateway
T4=t4-t1	Sum of T1,T2 and T3

A mobile device has to communicate with the gateway before it gets access to Internet. When a request arrives at the gateway, its Incoming Request Time t1 is logged. When the request leaves, the Outgoing Request Time t2 is logged. Therefore, T1 can be deemed as the time spent at the gateway. If a server produces a series of responses, each response will be logged respectively and marked as Incoming Response Time t3. Thus, we can regard T2 as the round-trip delay spent in the wired Internet. The responses need to be further processed at the gateway, such as protocol translation, encoding and decoding. It happens at the gateway in period of T3. Consequently, T4 shows the round-trip delay viewed by the cellular network. It differs from T2 in that T2 only involves the wired Internet, but T4 includes gateway-processing time as well.

Latencies Analysis

Overview of four delays T1, T2, T3 and T4.



(a) T1 (b) T2

(c) T3 (d) T4

Fig.3: Percentage of the num. of T1-T4

First, we investigate T1. Represented as red triangles in figure 3(a) in 2011, the mean value of T1 is a bit larger compared with it in 2010. Besides, there are more large values in 2011 though most of T1 is within 10ms for both. As shown in table 1, mobile Internet users in 2011 increase about 3.41 times and the number of requests is near 8 times as those in 2010. Therefore, the explosive growth

of requests will produce much burden on the gateway, and larger latency may arise when it is in heavy load. As we can see, some delays are even beyond ten seconds.

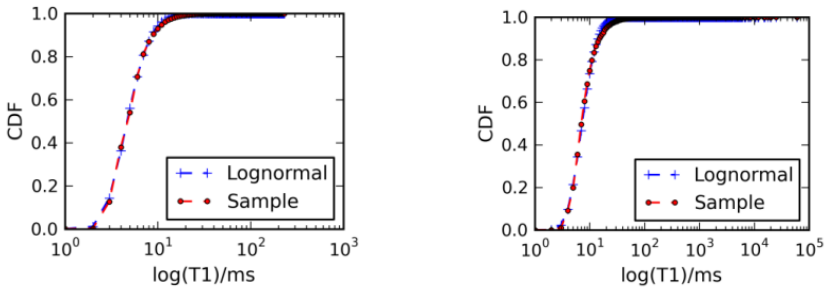
As shown in figure 3(b), the distribution curve of T2 in 2010 is more concentrative while it tends to become flat (that means it has a heavier tail) in the next year. Furthermore, the mean value of T2 in 2011 is about 12.9% higher than that in 2010. It means that the round-trip delay over the wired Internet is undergoing big changes. Some previous studies show that the average round-trip delay in common wired Internet is often tens milliseconds. However, the mean delay of T2 we got is much bigger - the average round-trip delay over the wired Internet is about hundreds milliseconds.

Despite the fact that the gateway is quite busy in 2011 for the large amount of requests, it is interesting to find that T3 changes very little in both years. So, we realize that the response packet processing raise little burden on the gateway, which contradicts much from the situation in T1. Moreover, it is easy to find that most values of T3 are very small.

As mentioned above, T4 covers the round-trip delay, and is the sum of T1, T2 and T3. In 2011, the mean value of T4 is about 12.8% higher than that in 2010. Figure 3(d) shows that the distribution of T4 in 2011 becomes more flat, shorter delays are becoming less but larger ones tend to be more. This results means the round-trip delay in mobile Internet ecosystem is becoming larger, and longer delays come more.

We further study the mathematical distribution of T1 by K-S test based on the statistics and give the results in figure 4. It shows that T1 fits the lognormal distribution very well with the probability density function described in equation(1). The parameters areμ=1.30, σ=0.57 in 2010 andμ=1.84, σ=0.57 in 2011.

$$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right) \tag{1}$$

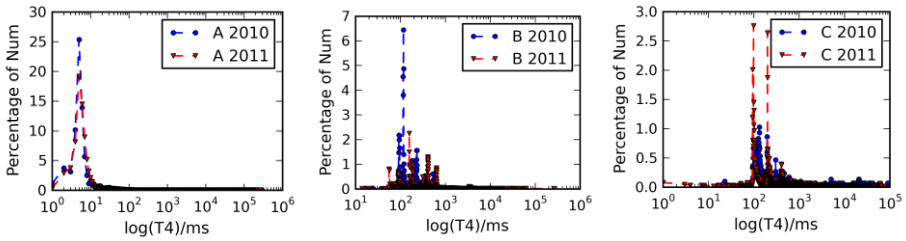


(a) 2010 (b) 2011

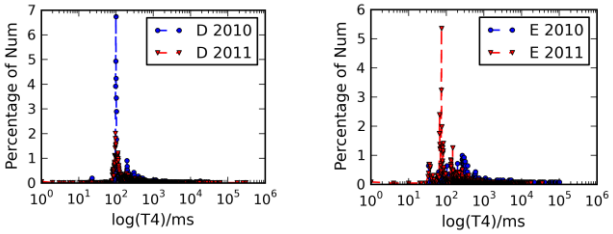
Fig.4: K-S test results of T1

Inter-ISP Delay.

Inter-ISP delay refers to the delay caused when severs and mobile users respectively belong to different ISPs. For anonymous consideration, we denote the five mobile carriers as ISP-A to ISP-E. Particularly, ISP-A is the biggest ISP which provides us the data sets. The distribution of inter-ISP delays are shown in figure 5(b)-(e). For comparison, we present the intra-ISP delay (users and servers are both belong to ISP-A) in figure 5(a). The horizontal axis is logarithm of delays in ms, and vertical axis is the percentage of frequency of each T4 value.



(a) ISP-A to ISP-A (b) ISP-A to ISP-B (c) ISP-A to ISP-C



(d) ISP-A to ISP-D (e) ISP-A to ISP-E

Fig.5: Percentage of the num. of inter-ISP T4 delay

For intra-ISP delay, figure 5(a) shows that it has almost the same distribution in 2010 and 2011. When surveying figure 5(b)-(e), some interesting findings come out. (1) The average delay of ISP-A to ISP-E is the second largest in 2010 but sharply reduces to its one fourth and even less in 2011. (2) In 2010, the mean value of ISP-A to ISP-B is almost the same as that of ISP-A to ISP-D, but the former increases nearly 78% while the latter decreases more than 35% in 2011. (3) In 2010, the largest mean delay is ISP-A to ISP-C, however, it not only still stays at the top but also grows 71% in 2011.

How should we explain above phenomenon? We think that the level of cooperation between ISPs matters great for the inter-ISP delay. When faced with its hottest rival, ISP-A may perform a limit policy to ISP-C's network. Therefore, the averages of inter-ISP latency in the two years are both far beyond the other four. In *Chongqing*, both ISP-D and ISP-E rent ISP-A's infrastructure; the cooperation between ISP-A and ISP-D, ISP-A and ISP-E are both enhanced in 2011. That is why both of the mean delays decreased though the number of requests increased substantially in 2011. Without knowing about the cooperation level between ISP-A and ISP-B, we can only see longer delays.

Conclusion

In this paper, we focus on investigating latency in internal parts of the mobile Internet ecosystem. Based on two real data logs collected from a gateway of a major mobile carrier, we have interesting and instructive findings. First, the gateway-processing delay can be well modeled by a lognormal distribution. Second, compared with round-trip delay over wired Internet in 2010, that in 2011 tends to be larger and has heavier tail. Finally, we found the inter-ISP delay has no fixed changing rules - we think it depends on the cooperation level between two ISPs. We hope our work can be helpful for better understanding both the whole and parts of latency in cellular-based mobile Internet.

Acknowledgements

This work is supported by National Basic Research Program of China (2012CB315806), Natural Science Foundation of Chongqing (CSTC2012JJA40060) and Jiangsu Future Networks Innovation Institute Prospective Research Project on Future Networks (BY2013095-5-07, BY2013095-2-03).

References

[1]U. Paul, A. P. Subramanian &M. M.Buddhikotand, Understanding traffic dynamics in cellular data networks.In Proceedings of IEEE INFOCOM'11, Shanghai, China.pp. 882-890, 2011.

[2]J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen&O. Spatscheck, A close examination of performanceand power characteristics of 4G LTE networks. In Proceedings of the 10th international conference on Mobile systems, applications and services(MobiSys), ACM New York, NY, USA. pp. 225-238, 2012.

[3]Joel Sommers&Paul Barford, Cell vs. WiFi: on the performance of metro area mobile connections. In Proceedings of the 12th ACM SIGCOMM conference on Internet measurement conference, New York, USA. pp. 301-314, 2012.

[4]M.Luckie, a scalable and extensible packet prober for active measurement of the Internet. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement conference, New York, USA. pp. 239-245, 2010.

[5]M.Lee, N.Duffield&R.R. Kompella, ascalable architecture for maintaining packet latency measurements. In Proceedings of the 12th ACM SIGCOMM conference on Internet measurement conference, New York, USA. pp. 115-122, 2012.

[6]G. Maier, A. Feldmann, V. Paxson & M. Allman, On dominant characteristicsof residential broadband internet traffic. In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, Chicago,USA. pp90-102, 2009.

[7]S.Sundaresan, W.Donato, N.Feamster, R.Teixeira, S.Crawford &A.Pescapè, Broadband Internet Performance: A View From the Gateway. In Proceedings of the ACM SIGCOMM'11, New York, USA. pp. 134-145,2011.

[8]A.Gember, A.Akella, J.Pang, A.Varshavsky&R.Caceres, Obtaining in-context measurements of cellular network performance. In Proceedings of the 12th ACM SIGCOMM conference on Internet measurement conference, New York, USA. pp. 287-300,2012.

[9]P. Romirer-Maierhofer, F. Ricciato, A. D'Alconzo,R. Franzan& W. Karner, Network-widemeasurements of TCP RTT in 3G. In InternationalWorkshop on Traffic Monitoring and Analysis (TMA),Berlin, Heidelberg.pp. 17–25, 2009.