A First Look at Cellular Network Latency in China

Xinheng Wang(✉), Chuan Xu, Wenqiang Jin, and Guofeng Zhao

School of Communication and Information Engineering,
Chongqing University of Posts and Telecommunications, Chongqing, China
xhwangcqupt@gmail.com, jinwenqiang6668@gmail.com,
{xuchuan, zhaogf}@cqupt.edu.cn

Abstract. The cellular networks in China are growing rapidly, thus various radio complex technologies, from 2G to 4G are deployed to support billions of mobile users. In this paper, we take a first attempt to analyze cellular network latency in China. A mobile app NetSense (The application is available at: https://github.com/lovesick/NetSense/) was scattered up on several hundred mobile terminals to measure the latency at large. Based on the measured results of 4 months, we find out that: (i) China's cellular network adopts hierarchical structure, with more gateways deployed in capital or major cities for each province; (ii) the latency varies largely across geographical locations, ISPs and is also dependent on the deployed cellular technologies; (iii) Considering end-to-end latency, the last mile delay contributes over 70 percentage for the whole transmission latency.

Keywords: Cellular network · Latency · Wireless delay · Last mile delay

1 Introduction

With the evolution of cellular technology, the wide deployment of 4G technologies and the rapid adoption of new smartphone devices, the cellular networks are growing explosively in China. The universal coverage of 4G networks enable the mobile devices such as phones and tablets have improved our ability of information accessing, sharing and processing. According to the report, more than 760 million users access the Internet via the cellular networks, nearly three-quarters of all the Internet users in China [1].

Although China is with the largest number of cellular users and the fastest growing market of mobile application [2], however, China's cellular networks have received relatively little attention in the measurement community. Perhaps, in China, researchers lack the infrastructure and resources for large-scale cellular network measurement studies such as those carried out in MobiPerf [3], and large data sharing from cellular carriers such as AT&T, which are essential for comprehensive and in-depth studies of the cellular network performance [4–6].

The popular applications, such as IM, social and E-commerce, have attracted more than 200 million users to use. Users share information through exchanging a large number of small packets via short http connections, the performance of information

Q. Chen et al. (Eds.): ChinaCom 2016, Part I, LNICST 209, pp. 339–348, 2018.
DOI: 10.1007/978-3-319-66625-9_33

exchange would depend on the cellular network latency. Therefore, carriers in China have widely adopted tree topology to construct their cellular networks, with more gateways deployed to provide more access points to Internet. Furthermore, for serving numerous types of mobile devices, each carrier needs to operate a variety of network technologies from 2G to 4G, for example, China Mobile provides GPRS, EDGE, TD-SCDMA and TD-LTE. These make cellular networks in China more complex, leading to very different network performances from others in US or Europe.

In this paper, we take a first look at latency of cellular networks in China. Our goal is to understand the latency characteristics of cellular network in China. To the best of our knowledge, our study is the first to investigate the characteristics of China's cellular network latency. Key contributions of this paper are presented below.

- We conduct a measurement on cellular network latency through our tool NetSense, which deployed on wide range mobile devices within commercial cellular networks in China.
- We show the topologies of cellular core networks, and found that the tree topology being widely adopted in three main cellular carriers, more gateways deployed in capital or major cities of each province, and 3 to 6 hops in the cellular core network.
- We compare the latencies when crossing different carriers, different areas, different technologies (WiFi, 3G, EDGE, GPRS), particularly the latency produced between different 3G technologies (EVDOA, HSDPA, HSPA); even the latency by the same technology crossing different carriers (TD-HSDPA and HSDPA).
- We provide a comprehensive comparison of our results to previous studies, and found that the latency in China is superior to that in US and in Norwegian. We think it is owing to more gateways being deployed in cellular core network in China; that provides more access points to Internet and shortens the latency.

2 Measurements and Data

2.1 Cellular Network Architecture

There are three major cellular carriers in China, China Mobile, China Unicom and China Telecom. Despite the difference among cellular technologies, a cellular network is usually divided into two parts, the Radio Access Network (RAN) and the Core Network (CN). The simplified architecture of cellular network is shown in Fig. 1. UEs are essentially mobile handsets carried by end users. The RAN allows connectivity between a UE and the CN. They consist of multiple base stations called BSC of 2G technologies (e.g., GPRS, EDGE, 1xRTT, etc.), Node B of 3G technologies (e.g., EVDOA, HSDPA, HSPA, etc.), and Evolved Node B of 4G technologies (e.g., LTE TDD, LTE FDD, etc.). The centralized CN is the backbone of cellular network, which consist a number of mobility management devices and gateways called SGSN and GGSN in China Mobile and China Unicom, PCF and PDSN in China Telecom, MME and S/P-GW of 4G, but the structure of the CN does not differentiate each other.

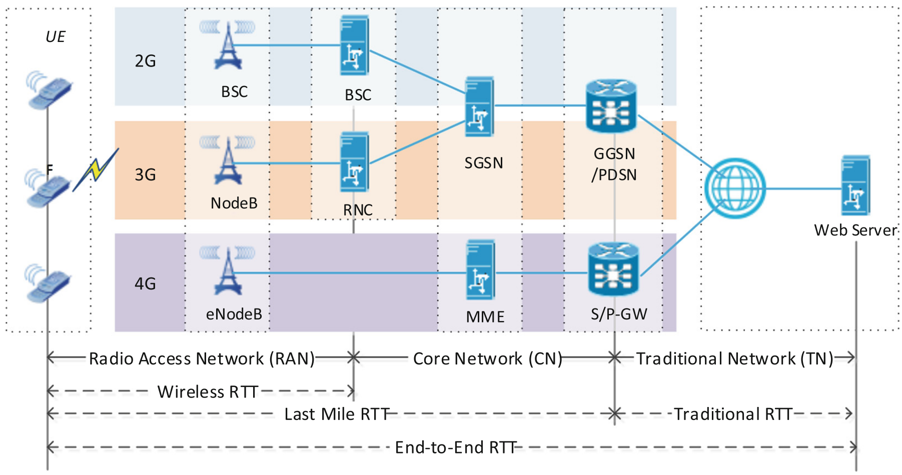


Fig. 1. Simplified cellular network structure in China.

2.2 Methodology and Data Set

In this study, NetSense, a professional measurement tool was designed and deployed on hundreds smartphones to measure the performance of cellular networks. By doing the tasks embedded in the NetSense, the network performance information can be collected directly from the end users every day, including network type, carriers, location, latency, throughput, trace, HTTP and DNS performance.

We publicly issued the Android version of NetSense application1 in December, 2015, distributed via the major application platform 360's App Store, Baidu's Mobile App Store and 91's Mobile Application Store in China. Till March, 2016 more than 400 users from across the country have installed and run our tool, and more than 300K test records have been collected. Among all records, more than 98% records include complete latency data through ping and trace tools, about 95% of records have GPS readings. According to location information, users from more than 20 provinces were observed.

As shown in Fig. 1, we break down end-to-end RTT into two components: wireless RTT which means the latency between a UE and the NodeB, last mile RTT which is the latency between a UE and the gateway (GGSN/PDSN). The wireless RTT is the latency in the RAN, and last mile RTT is the latency in the cellular core network. We use NetSense to trace route the end to end delay, the partition method is as follows. Since the IP addresses in the cellular network are mostly private address, we use the RTT of first hop with private IP address as the estimation of the wireless RTT in RAN, and RTT of the first hop but with public IP address as the estimation of the last mile RTT in core network. The number of users and records from three carriers are listed in Table 1.

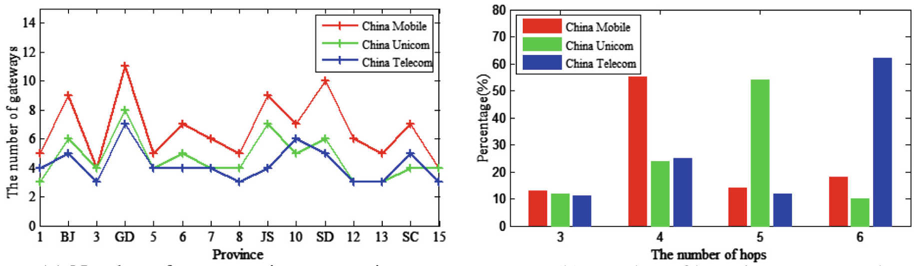
Table 1. The data set

Data	Carriers		
	China Mobile	China Unicom	China Telecom
Users	126	97	180
	From 2015-12-01 to 2016-03-30		
Records	Ping	Trace route	
	180K	150K	

3 Overview of Cellular Core Network

Gateways (i.e. GGSN/PDSN) are of extreme importance for cellular core network. They provide access points to the Internet for mobile users, make the cellular infrastructures transparent from the external network, and also plan an important role in geolocation mapping for mobile devices [13]. Similar to the broadband internet, the cellular core networks in China also adopt the tree topology. Several gateways are deployed in the capital city or major cities of each province. The packets from a UE are forwarded first to the NodeB, then to aggregation node, and finally to the Internet via one of the gateway.

Figure 2(a) plots the number of gateways for the three ISPs in each province where there are mobile terminals use our NetSense app. Notably, there are at least 3 gateways for each ISP in any province. Note that since the vantage points in our measurement do not cover all regions in China, it is impossible to identify all the gateways in every province. China Mobile deploys more gateways than other two ISPs in each province, since it serves more than 500 million mobile Internet users. In some provinces, there are as many as 11 gateways. Generally, we find more gateways are deployed in more developed provinces, such as Beijing, Guangdong, Shandong and Jiangsu. Compared to the results found in US [13], where there are only 4–6 gateways for each of the four carriers, ISPs in China deploys much more gateways. One possibility is that they need to support much more mobile users than those in US. Another possibility is that these ISPs compete with each other to provide better connectivity to Internet.



(a) Number of gateways in one province (b) Number of hops in core network

Fig. 2. The core network of different carrier.

Figure 2(b) illustrates the number of hops from mobile users to Internet gateways. We can clearly see the difference among three ISPs. For China Mobile, 70% of provincial gateways can be reached within 4 hops, while this percentage is only 40% and for China Unicom and China Telecom. In fact, the number of hops is concentrated around 5 and 6 hops for China Unicom and China Telecom, respectively. This can be explained by the number of gateways as shown in Fig. 2(a). Having more gateways in capital city as well as other big cities in each province could provide users with the close one for connection.

4 Cellular Network Latency Characterization

In this section, we first examine characteristics of cellular network latency using. We then proceed to compare our results with measurement studies of cellular network in western countries.

4.1 RTT in 4G Network

We focus on comparing cellular network latency performance among the three cellular carriers. For each carrier, given that the evolution of cellular technologies makes it provide several cellular access technologies simultaneously. We further analyze the latency of different cellular network technologies.

Since, 4G networks are widely deployed in Chinese commercial cellular network, and the most data traffics are forwarded by 4G network in mobile wireless network. We use the RTT to measure the network performance from the aspect of latency.

Figure 3(a) plots the CDF of RTT for three carriers. The latency for the 3 carriers exhibit short-tail distributions. Nevertheless, the latency for China Unicom shows a more concentrated effect around 52.9 ms. Indeed, China Telecom has the good performance in terms of latency, followed by China Telecom with average RTT of 63.7 ms and China Mobile with average RTT of 95 ms. Furthermore, Fig. 3(b) shows the transmission latency in terms of jitter, obviously, China Mobile network shows a

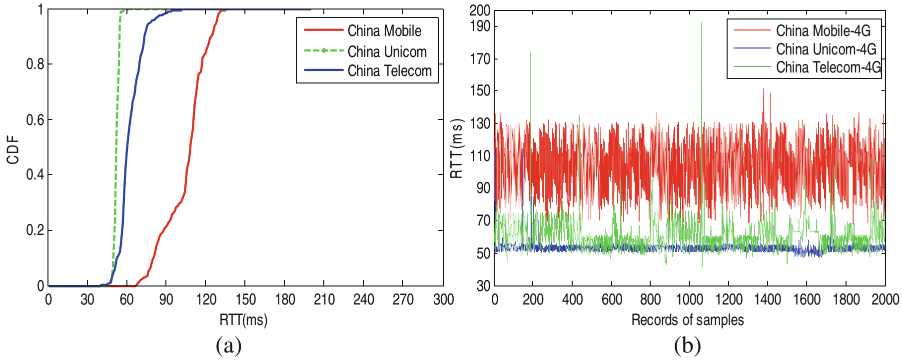


Fig. 3. RTT comparison of major carriers.

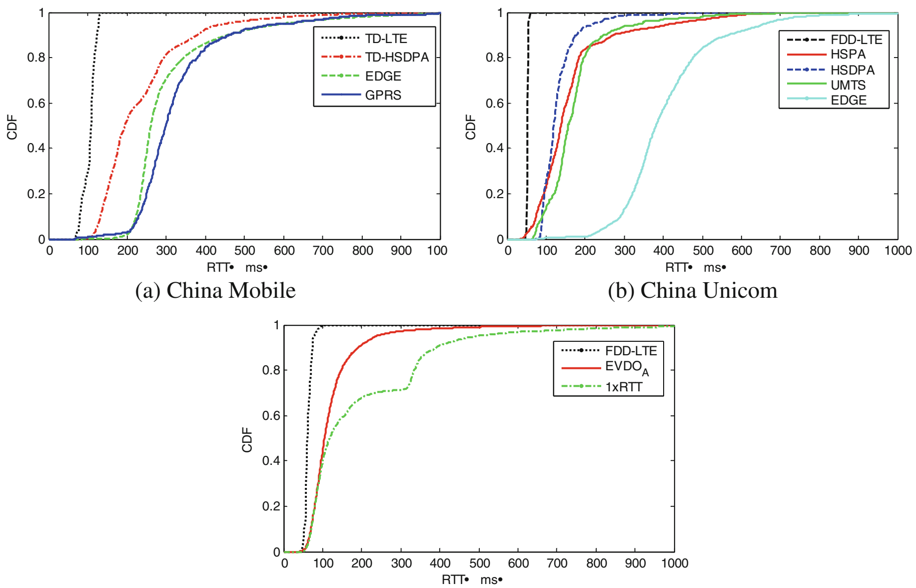
relative poor performance in transferring data packets stably with RTT values unevenly distributed in the wide range of 80 ms to 148 ms. And the RTT values of the most recorded samples in China Telecom concentrated between 50 ms and 75 ms. However, China Unicom shows an extremely stable data transmission in term of jitter with RTT values covered constantly in the range of 48 ms to 55 ms.

4.2 RTT Among Different Types of Accessing Technology

In Fig. 4, we further compare the RTT among different types of cellular networks. As shown in Fig. 4(a), when using 2.75G access technologies, such as GPRS, EDGE, the RTTs are very close to each other, with median RTT of 299.5 ms and 261.7 ms, respectively. Moreover, when considering the 3G access technology TD-HSDPA, the RTT is lower, with median RTT of 195.6 ms. However, the performance of 4G access technology is really astonishing, for TD-LTE, the median RTT is 99 ms.

It is notable that the GPRS networks in China have much better performance in terms of latency than that in US, where the median RTT is 1,000 ms in US [3]. Since GPRS is the main cellular network type of China Mobile, which has been optimized for many years. Its performance is relatively stable and comparable to 3G technology TD-HSDPA.

Figure 4(b) shows latency distribution for the 3G networks used in China Unicom, i.e. UMTS, HSPA and HSDPA. We can see HSDPA has the smallest median RTT of 120.3 ms, and UMTS has the largest median RTT of 156.8 ms. Compared to 3G



(a) China Mobile (b) China Unicom

(c) China Telecom

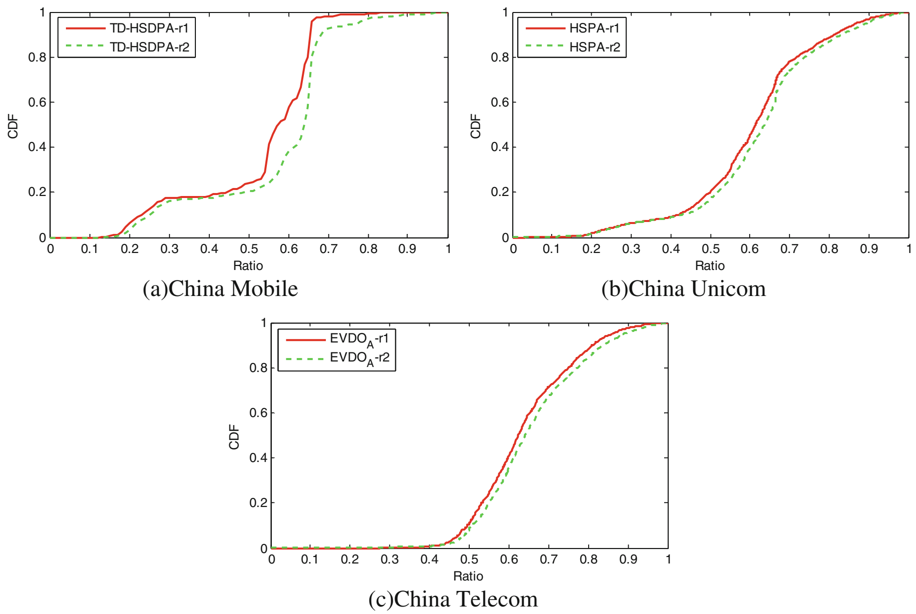
Fig. 4. RTT comparison among different network types of major carriers.

networks, the performance of EDGE is much poor, with the median RTT of 382.2 ms. And the FDD-LTE accessing technology shows an even smaller RTT with 52.9 ms as its median value in China Unicom cellar network. In Fig. 4(c), we observe that EVDOA of China Telecom has better performance than that of 1xRTT, with the median RTT of 105.4 ms compared to 118.8 ms in 1xRTT, which is the earliest CDMA 2.75G access technology. What's more, the 4G FDD-LTE accessing technology applied in China Telecom network shows a better performance as predicted which presents its median RTT value is 60.8 ms.

4.3 End-to-End Delay Analysis

We then proceed to analyze the end-to-end delay. Usually the RAN latency dominates the end-to-end delay [12]. In order to quantify the contribution of end-to-end delay, we consider the ratio (r1) of the wireless RTT to the end-to-end RTT, ratio (r2) of the last mile RTT to the end-to-end RTT.

Figure 5 illustrates the CDF of r1 and r2 among main technologies of three carriers. As shown in Fig. 5(a) for China Mobile, the wireless RTT accounts for more than 50% of the end-to-end RTT for 80% of TD-HSDPA measures. We also observe that the differences between r1 and r2 are consistent, and this gap is actually the delay in core network, which is smaller than 5%. We have the same observation for China Unicom and China Telecom, but the difference between r1 and r2 is only 2%.



(a)China Mobile (b)China Unicom

(c)China Telecom

Fig. 5. Wireless and last mile delay characteristics.

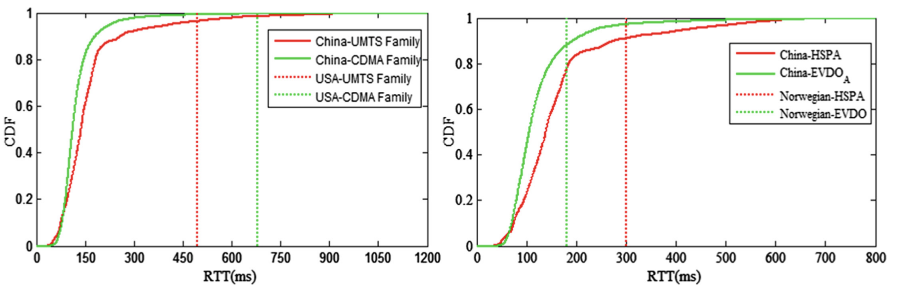
Our findings suggest that the RAN latency dominates the overall end-to-end delay, independent of ISPs and 3G technologies. We also observe that the median values of r1 are 0.64, 0.63 and 0.62 for China Mobile, China Unicom and China Telecom, respectively. However, with the evolution of cellular technology, the ratio of RAN latency in end-to-end delay becomes constantly smaller, but which is widely distributed.

4.4 Comparison with US and EU Cellular Network

We finally compare the latency characteristics of China's cellular networks with those in US and EU. In order to have a fair comparison, we focus on the delay metric with the same network technology. The 3GTest study [3] deployed an app that measures end-to-end delay from the user's handsets to the test server with many 3G tests of four large U.S. cellular carriers in late 2009. After that, they compared network performance among technology types which cover most 3G network types [14]. The recent study [6] conducted an in-depth of 4G LTE network performance using a combination of active and passive measurements. The 3G network study [10] presented a long-term delay measurement from data connections in 3 Norwegian 3G networks.

Figure 6(a) shows the network comparison results of 3G technologies with the measurement data collected by MobiPerf in U.S. Obviously, the latency of UMTS and CDMA 3G technologies deployed in China carriers are much better than those in U.S. carriers, the median RTT of UMTS family and CDMA family are 137.3 ms and 105.4 ms, comparing to 490 ms and 680 ms in U.S. The detail results are shown in Fig. 5, in UMTS 3G technologies, the UMTS, HSDPA and HSPA's median RTT of China Unicom are 156.8 ms, 120.3 ms and 138.7 ms, which are smaller than that of U. S. carriers with median RTT of nearly 500 ms. Similarity in CDMA 3G technologies, the 1xRTT and EVDOA's median RTT of China Telecom are 118.8 ms and 105.4 ms, which are also lower than that of U.S. carriers, with median RTT of 1,000 ms and 680 ms.

Figure 6(b) shows the network latency performance comparison results with the measurement data collected by wireless mobile mini-card and USB modem in Norwegian. Obviously, the latency of the same HSPA and EVDOA 3G technologies



(a) Comparing with US. Carriers (b) Comparing with Norwegian Carriers

Fig. 6. Comparing with previous measurement results.

deployed in China carriers are better than those in Norwegian carriers, with the median RTT of 138.7 ms and 105.4 ms, comparing to 300 ms and 180 ms.

Since carriers in China have to serve huge number of mobile users, they deploy much more gateways in every province than in other counties, which would be the main reason for the better latency performance in China.

5 Related Work

The characteristics of cellular networks have been studied for few years. Some existing works have studied the performance on smartphones, which compared 3G and WiFi performance on smartphones [7], compared cellular network performance on different smartphones among carriers [3], profiled diverse usage behaviors of smartphone applications [8] and studied the correlation among IP address, location and network latency for smartphones [7]. Some studies also performed large-scale measurement of cellular networks. Seetharam et al. focused on measuring cellular network parameters such as throughput, latency and packet loss by performing TCP downloads, uploads and others experiments within different wireless services providers' networks. [5] Rusan et al. analyzed the impact of backhaul packet delay on the LTE S1-U interface, which provides user plane transport between the Core Network and the Evolved NodeBs [9]. Elmokashfi et al. studied the cellular network delay characteristics from data connections in 3 Norwegian 3G networks [10]. Arlos et al. studied the influence of the packet size on the delay in 3G networks [11]. Qian et al. performed network-wide measurement studies of cellular periodic transfers [12]. Cellular network infrastructures also have been studied. Xu et al. characterized 3G data network infrastructure, found that the routing of cellular data traffic is quite restricted as traffic traverse through a small number of gateway nodes [13]. Wang et al. unveiled cellular carriers' NAT and firewall policies [4].

6 Conclusion

In this paper, we made the efforts to characterize the infrastructure of cellular core network, and the latency of three major cellular carriers in China. Our study was based on four months data collected from widely deployed measurement application Net-Sense. Compared the latency performance under situation of different carriers, technologies and areas, We found that in 2G technologies, GRPS and EDGE of China Mobile has the best performance, but in 3G technologies, the latency with HSPA of China Unicom and EVDOA of China Telecom are much better than TD-HSDPA of China Mobile, which makes China Mobile first deployed the 4G technology TD-LTE in December 2013. Furthermore, the 4G technologies provide a predictable small transmission latency in all three carriers' network, probably because more base stations are installed for functioning 4G mobile network compared with previous networking technologies.

For furthering this research, we are exploring other factors, such as network type, packet size and mobile device, which might have impact on the cellular network latency would be considered. On the other hand, we are exploring the cellular network performance more detailed from the perspective of throughput, DNS lookup and TCP flow of 3G and 4G technology.

Acknowledgments. This work is supported by the National Science Foundation of China (NSFC) under grant 61402065, Chongqing key science and technology project (No. cstc2014jcsf70001) and Young Backbone Teacher Funding Program of Chongqing Municipal Education Commission.

References

1. CNNIC: Statistical Report on Internet Development in China. http://www.cnnic.net.cn/hlwfzyj/
2. Tian, Y., Dey, R., Liu, Y., et al.: Topology mapping and geolocating for China's internet. IEEE Trans. Parallel Distrib. Syst. **24**(9), 1908–1917 (2013)
3. Huang, J., Xu, Q., Tiwana, B., Mao, Z.M., Zhang, M., Bahl, P.: Anatomizing application performance differences on smartphones. In: Proceedings of Mobisys. ACM (2010)
4. Wang, Z., Qian, Z., Xu, Q., Mao, Z.M., Zhang, M.: An untold story of middleboxes in cellular networks. In: Proceedings of the ACM SIGCOMM (2011)
5. Seetharam, A., Walker, P.: An empirical characterization of cellular network performance. In: 2016 International Conference on Computing, Networking and Communications (ICNC). IEEE (2016)
6. Huang, J., et al.: An in-depth study of LTE: effect of network protocol and application behavior on performance. In: Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM. ACM (2013)
7. Gass, R., Diot, C.: An experimental performance comparison of 3G and WiFi. In: Proceedings of PAM (2010)
8. Xu, Q., Erman, J., Gerber, A., Mao, Z.M., Pang, J., Venkataraman, S.: Identifying diverse usage behaviors of smartphone apps. In: IMC (2011)
9. Rusan, A., Vasiu, R.: Assessment of packet latency on the 4G LTE S1-U interface: impact on end-user throughput. In: 2015 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM). IEEE (2015)
10. Elmokashfi, A., Kvalbein, A., Xiang, J., et al.: Characterizing delays in norwegian 3G networks. In: Proceedings of PAM (2012)
11. Arlos, P., Fiedler, M.: Influence of the packet size on the one-way delay in 3G networks. In: Proceedings of PAM (2010)
12. Qian, F., Wang, Z., Gao, Y., Huang, J., Gerber, A., Mao, Z.M., Sen, S., Spatscheck, O.: Periodic transfers in mobile applications: network-wide origin, impact, and optimization. In: Proceedings of the 21st International Conference on World Wide Web (2012)
13. Xu, Q., Huang, J., Wang, Z., Qian, F., Gerber, A., Mao, Z.M.: Cellular data network infrastructure characterization and implication on mobile content placement. In: Proceedings of ACM SIGMETRICS, San Jose, CA (2011)
14. Balakrishnan, M., Mohomed, I., Ramasubramanian, V.: Where's that phone?: geolocating IP addresses on 3G networks. In: Proceedings of IMC (2009)