Smartphone data usage: downlink and uplink asymmetry

Bowei Yang, Weisi Guo, Yanliang Jin and Siyi Wang✉

Mobile phone usage has changed significantly over the past few years and smartphone data usage is still not well understood on a statistically significant scale. This Letter analyses 2.1 million smartphone usage data values and explore the current wireless downlink–uplink demand asymmetry for different time periods and across different radio access networks. The current data demand over 2G networks remains largely *symmetric* with strong temporal variations, whereas the demand over 3G networks is *asymmetric* with surprisingly weak temporal variations is shown here.

Background and Relevance: To better deliver telecommunication services, one must come to understand the consumer demand through real data. Mobile phone usage data represents a tremendous opportunity for telecommunication engineers to better understand the behaviour of both individuals and of demographics. For the past few years, mobile phones, and in particular smartphones have increased their penetration into our lives. Over 4 billion people on the planet are active mobile phone users and the number of new mobile users is growing four-fold faster than the population [1]. Smartphone penetration has exceeded 60% for many developed nations and people often carry multiple smart devices, i.e., smartphones, tablets, and smart-wearables. In this Letter, we focus mainly on smartphone data usage as we generate most of our wireless network data via smartphones (68%) [1].

Significance: Mobile phone data can be challenging to collect, especially on a statistically significant scale. Prior to the growing usage of smartphones, extracting personal mobile phone data was difficult. Historically, call-data-records (CDR) based research [2, 3], has yielded useful statistical models [4]: i.e., user arrival (~Pois(.)) and data demand per session ($\sim \text{Log} - \mathcal{N}(.)$). However, the main drawback is that the data is typically aggregated across multiple devices, time, and a large area of space, causing the loss of granularity that can prove to be important. CDR based research also requires active industrial collaboration and that often limits the CDR to be those of a single network operator. Cross-operator techniques such as spectrum sensing requires mass deployment of sensors [5], as well as the translation between the measured power density and traffic demand [6]. Alternatively, using survey based methods can also achieve cross-operator reach and avoid hardware investment, but have the disadvantage of relying on customer recollection and the difficulty of reaching a wide audience in a cost-effective manner [7].

Smartphones data mining allow us to access the data directly via purpose built applications (Apps). Most academic studies in this domain have been small scale, with typical data sample sizes varying between 10 to 100 s of users, many of which will move within a single location area (i.e. a single city) [8]. Small scale testing will often lead to skewing the data in favour of certain technological and social demographics, i.e., data-hungry post-graduate students. Large scale smartphone studies that involve 1000 s or more users over several months are very limited, for the simple reason that large scale data collection is either expensive or requires strong collaboration from industry as well as the enduring participation of private citizens.

In this Letter, we leverage on data collected from a popular commercial App. In the interests of commercial sensitivity, we cannot reveal the App in this Letter. To protect the users' *privacy*, we hash the identity of individual data sources and remove the specific time and location of where the data is generated. We randomly sample over 2.1 million data values from 10,000 unique smartphone users over a period of several weeks in 2014. All smartphone data is selected from users that predominantly reside in a single developed metropolitan area and this data approximately represents 0.17% of the smartphone users in that geographical area. This to the best of our knowledge represents one of the larger studies of smartphone data consumption. Fig. 1 shows an example of the data sampled from a user along with common data rates required for different audio and video services. In this Letter we only provide statistical results to show the DL/UL data demand ratio at different time periods over the 2G and the 3G radio access technology (RAT). Due to the scarcity of 4G service usage in our current data set, we restrict this study to 2G and 3G services.

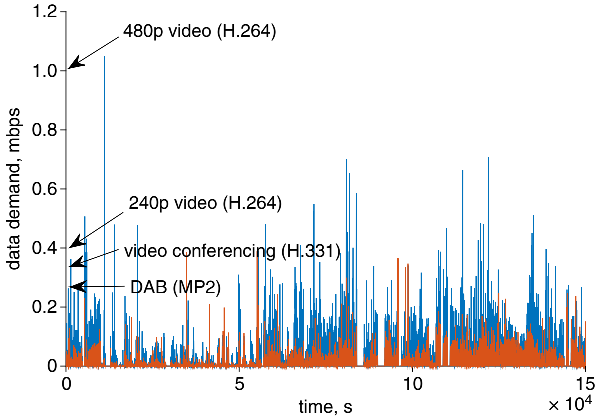

Fig. 1 *Example of smartphone data usage (DL and UL) with common data rates required for different audio and video services*

Impact on network design: Whilst voice calls were downlink (DL) and uplink (UL) symmetric, the dominance of packet-switched data services has led to rethinking in cellular network spectrum management and operations. Most existing wireless carriers use frequency-division duplex (FDD) with separate DL and UL channels operating on different frequency bands (*paired spectrum*). This is efficient for symmetric or fixed DL/UL demand ratio scenarios. As new generations of cellular networks are being rolled out (e.g. 4G LTE), most will be FDD configured with paired spectrum fixed at a ratio of 1:1 between DL and UL channels [9].

Over the past 5 years, data demand has grown stochastic and one of the most data-hungry application areas in smartphone usage is video streaming, further increasing the asymmetry of data demand. Estimates of the ratio of DL/UL ratio in video is approximately 8–11:1 [10]. Whilst FD-LTE is backwards compatible with legacy networks, TD-LTE (time-division) can be more efficient in transmitting asymmetric data since they provide the flexibility to adjust the UL/DL ratio dynamically to better match the demand. In this Letter, our goal is to use the cellular DL and UL data mined from smartphones to examine whether there is a case for future TD-LTE implementation and whether the existing FDD provisioning on 3G is sufficient.

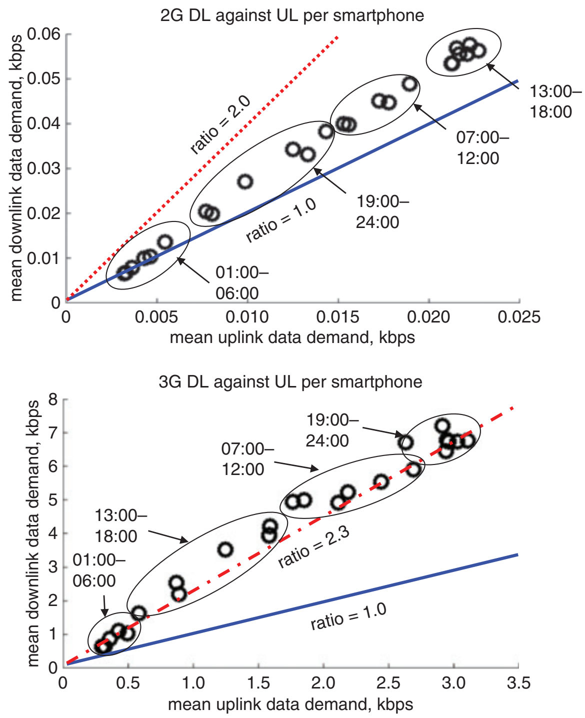

Fig. 2 *Smartphone usage over 2G and 3G RATs: plot of DL against UL data demand for different times of the day*

DL and UL asymmetry results: Fig. 2 shows the plot of smartphone mean DL against mean UL data demand over 2G (top) and 3G (bottom) RATs. The results show two important trends: (i) the DL and UL mean demand scales monotonically, and (ii) each hour can be clustered into groups of 6 continuous hours, based on the demand value. Based on this clustering, we examine the DL against UL ratio's statistical distribution. We define the instantaneous ratio between DL data demand $R_{n,\mathrm{DL}}$ and UL data demand $R_{n,\mathrm{UL}}$ as ζ_n:

$$\zeta_{n,i} = \frac{R_{n,\mathrm{DL}}}{R_{n,\mathrm{UL}}}, \tag{1}$$

where n denotes the cellular network affix and i denotes the cumulative distribution value. The cumulative distribution function (CDF) of ζ will reveal the statistics for instantaneous demand. We divide the data into clear time periods to help distinguish demand differences.

The CDF plot of $\zeta_{2\mathrm{G}}$ for smartphones transferring data over different 2G RATs is shown in Fig. 3 for different time periods. The results show a strong temporal variation between night and day time. For the late night to early morning period (01:00 to 06:00): the mean ratio is $\zeta_{2\mathrm{G},0.5} = 1.0$. The greatest asymmetry is reached in the afternoon to evening period (13:00–18:00), where the mean ratio $\zeta_{2\mathrm{G},0.5} = 2.0$. The combined results show that current usage of 2G network is largely symmetric during the night, with strong temporal variations during the day, but does not generally exceed 2.0. In fact, very few occasions of smartphone usage (<1%) achieve the asymmetry reported when using video streaming applications ($\zeta_{2\mathrm{G}} \geqslant 11$).

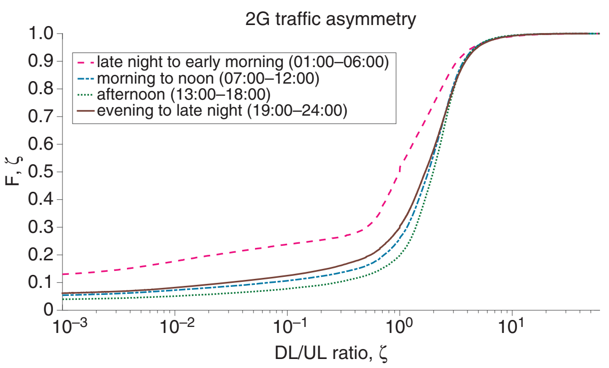

Fig. 3 *Smartphone usage over 2G RAT: CDF plot of DL/UL data demand ratio at different time periods of the day*

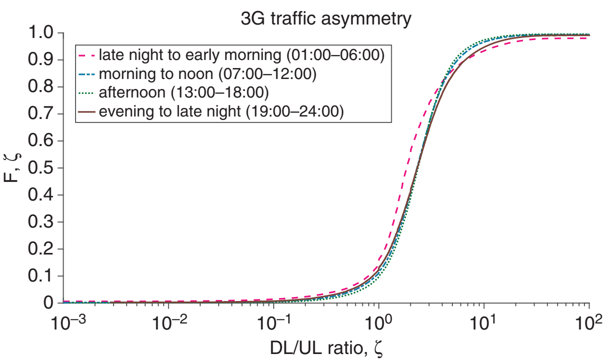

Fig. 4 *Smartphone usage over 3G RAT: CDF plot of DL/UL data demand ratio at different time periods of the day*

The CDF plot of $\zeta_{3\mathrm{G}}$ for smartphones transferring data over different 3G RATs is shown in Fig. 4. The results show a weaker temporal variation between night time and day time. For the same late night to early morning period (01:00–06:00): the mean ratio is $\zeta_{3\mathrm{G},0.5} = 1.8$. For the other day periods, they all share a common mean ratio $\zeta_{3\mathrm{G},0.5} = 2.3$. The combined results show that current usage of 3G network has largely asymmetric ($\approx 2 - 2.3$) with weak temporal variations. Compared with 2G, a few more occasions of smartphone usage ($\approx 3\%$) achieve the asymmetry reported when using video streaming applications ($\zeta_{3\mathrm{G}} \geqslant 11$).

Implications and conclusions: Our results show that current usage of 2G network remains close to being symmetric with strong DL/UL ratio variations between night time and day time. For data demand over the 3G network, there are only small variations between different times of the day and the DL/UL asymmetry is stable with a value of $\zeta_{3\mathrm{G},0.5} \approx 2 - 2.3$. That is to say, whilst we have come to expect that the 2G network experiences largely symmetric demands, we have discovered that there is a stronger temporal variation than 3G, which is surprising. On the other hand, whilst we have come to expect that the 3G network experiences a more asymmetric demand, we have discovered that the magnitude is not as large as feared [10] and the temporal variations are weak. The underlying reason is perhaps most of the heavily asymmetric video demand is being carried over the Wi-Fi network.

Network implication wise, at least based on the smartphone usage data analysed in this Letter, the need for future TD-LTE is not as strong as many researchers have come to expect. The vast majority of the data demand records are displaying a reasonably symmetric behaviour in both downloading and uploading content, with 97% of demands below the asymmetric levels of video ($\zeta = 8 - 11$). In the short term, we advice that operators either observe how the demand pattern changes and whether users are increasingly offloading video demand to Wi-Fi networks. Based on the data, we do advice that a 2-fold bandwidth is given to the DL channels to address the stable asymmetry ($\zeta_{3\mathrm{G},0.5} \approx 2 - 2.3$) found in the data. Future work will focus on using the smartphone data to examine when users choose to offload data to Wi-Fi.

Submitted: *19 September 2015* E-first: *10 December 2015*
doi: 10.1049/el.2015.3249
One or more of the Figures in this Letter are available in colour online.

Bowei Yang (*School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, People's Republic of China*)

Weisi Guo (*School of Engineering, University of Warwick, Coventry, Warwickshire CV4 7AL, United Kingdom*)

Yanliang Jin (*School of Communication and Information Engineering, Shanghai University, Shanghai, People's Republic of China*)

Siyi Wang (*Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangshu 215123, People's Republic of China*)

✉ E-mail: siyi.wang@xjtlu.edu.cn

Weisi Guo: Also with the School of Communication and Information Engineering, Shanghai University, Shanghai, People's Republic of China

References

1 Pepper, R.: 'Cisco visual networking index: global mobile data traffic forecast update'. Technical Report, Cisco, 2013
2 Na, C., and Rappaport, T.S.: 'Measured wireless LAN public hotspot traffic statistics', *Electron. Lett.*, 2004, **40**, (19), pp. 1202–1203
3 Lee, D., Zhou, S., Zhong, X., Niu, Z., Zhou, X., and Zhang, H.: 'Spatial modeling of the traffic density in cellular networks', *IEEE Wirel. Commun.*, 2014, **21**, (1), pp. 80–88
4 Laner, M., Svoboda, P., Schwarz, S., and Rupp, M.: 'Users in cells: a data raffic analysis'. IEEE Wireless Communications and Networking Conf. (WCNC), Shanghai, China, 2012, pp. 3063–3068
5 Teguig, D., Nir, V.L., and Scheers, B.: 'Spectrum sensing method based on likelihood ratio goodness-of-fit test', *Electron. Lett.*, 2015, **51**, pp. 253–255
6 Guo, W., and Wang, S.: 'Mobile crowd-sensing wireless activity with measured interference power', *IEEE Wirel. Commun. Lett.*, 2013, **2**, (5), pp. 539–542
7 Butt, S., and Phillips, J.: 'Personality and self reported mobile phone use', *Comput. Hum. Behav.*, 2008, **24**, (2), pp. 346–360
8 Yu, Q., Erman, J., Gerber, A., Mao, Z., Pang, J., and Venkataraman, S.: 'Identifying diverse usage behaviors of smartphone Apps'. ACM Special Interest Group on Data Communication (SIGCOMM), Toronto, Canada, 2011, pp. 329–344
9 3GPP: 'TS 25.2XX – UMTS-FDD: Physical Layer'. Technical Report, 3GPP, 2009
10 Erman, J., Gerber, A., Ramakrishnan, K., Sen, S., and Spatscheck, O.: 'Over the top video: The gorilla in cellular networks'. Proc. of the ACM SIGCOMM Conference on Internet Measurement Conf., Berlin, Germany, 2011, pp. 127–136