Characterizing the Phenomenon of Traffic Tide for Large-scale Mobile Cellular Data Networks

Jingtao Ding, Yong Li, Depeng Jin
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
Email: liyong07@tsinghua.edu.cn

***Abstract*—With tremendous growth in demand for broadband data, we undergo severe traffic congestion problems in mobile data networks. Researchers found out that there exists spatio-temporal inhomogeneity in data traffic, which lacks a measurement-driven study. In this work, by carrying out a thorough analysis of data traffic obtained from 380,000 base stations in a large-scale metropolis of Shanghai, we quantitatively characterize the "tide" phenomenon in traffic usage. In different days, its distribution shows an exponential decay, while in different places it follows a power law distribution. These findings would provide important insights on how to deal with traffic congestion problems.**

I. Introduction

The demand for data is surging rapidly every year with the Cisco Visual Networking Index for 2014 forecasting a nearly 11-fold increase in global mobile data traffic from 2013 to 2018 [1]. Cloud services and M2M applications, capacity-hungry applications, and bandwidth-hungry devices are contributing to the high mobile data congestion [4]. However, researchers found out that the Erlang based traffic usages in Berlin are inhomogeneous both spatially and temporally [3]. With the deployment of high-speed mobile network, users tend to frequently use data service of mobile network. Therefore, it is similarity that there also exists spatio-temporal inhomogeneity in the distribution of mobile data traffic.

This inhomogeneity in traffic distribution is closely tied with mobile traffic congestion. However, a large dataset is required in order to obtain credible results. Consequently, until now, there exists no work on characterizing spatio-temporal inhomogeneity for large-scale mobile data networks. Most works are focused on the solutions to alleviate mobile network congestion. One way is to use network pricing as a congestion management tool, for example, time-dependent pricing (TDP), where the service providers make discounts during off-peak period to incentivize the time-shifting of data [4]. We believe that quantitatively characterizing spatio-temporal inhomogeneity of data traffic can provide important hints on how much the service providers would cost when using TDP.

In this work, we focus on characterizing this inhomogeneity based on large-scale traffic trace in *Shanghai*, which can be described as a stable "tide" when considering overall traffic. We define the metrics of peak-to-average ratio (PAR) and peak-to-valley ratio (PVR) to measure temporal traffic variation. On basis of quantitative analysis, we find out that the distributions of both PAR and PVR show an exponential decay in different days at the same place, and follow a power law distribution in different space within the same time period. These results indicate the spatial differentiation of this traffic "tide" is more obvious than that of temporal. To the best of our knowledge, our work is the first trial to present a measurement-driven characterization of spatio-temporal inhomogeneity in data traffic usage.

II. Data Collection and Processing

In order to carry out a measurement-driven traffic analysis, we use an anonymous large-scale data traffic records from over 380,000 cellular base stations deployed in *Shanghai* by one of the major operators in China, within an interval of 4 weeks in Augest 2014. The specific information of each log includes: the device's ID (anonymized), timestamps of traffic start and end time, base station (BS) ID, BS location and traffic volume.

Based on the massive dataset, a preprocessing procedure is conducted to sort the traffic records by time and BS ID, and then compute data traffic according to the usage during a certain period (1 minute in our work) within same BS. Further, we measure the peak traffic during 0:00 to 11:59 and 12:00 to 23:59, which are marked as p_1 and p_2. Denoting the valley traffic between p_1 and p_2 as v, we define the peak traffic of whole day as $p = max\{p_1, p_2\}$ and the average traffic as p_a. Finally, we are able to easily compute PAR (i.e. $R_a = p/p_a$) and PVR (i.e. $R_v = p/v$) for each day and BS.

III. Main Result

To obtain the characteristic of temporal traffic variation, we average over 60 samples of traffic usage to get the overall usage $\overline{u}$ during one hour. Fig. 1(a) depicts the overall temporal traffic usage in one week, where we can observe obvious stable periodic variation ("tide") from peak time to valley time. We also find out that this "tide" shows spatio-temporal differentiation from the other three figures, which show the traffic of three typical BSs (subway station, central business district (CBD) and residential zone). For example, traffic trace near subway station tend to reach peak level in rush hours, i.e. 9AM and 18PM; while those near CBD and residential zone tend to have only one peak each day. We quantitatively show the average and variance of PAR and PVR of Fig. 1 in Table 1, where the average denoted by $Avg(R)$ indicates spatial differentiation while variance denoted by $Var(R)$ implies that "tide" changes each day. In conclusion, based on Fig. 1 and Table 1, we show that this inhomogeneity is stable in

978-1-4673-7131-5/15/$31.00 ©2015 IEEE

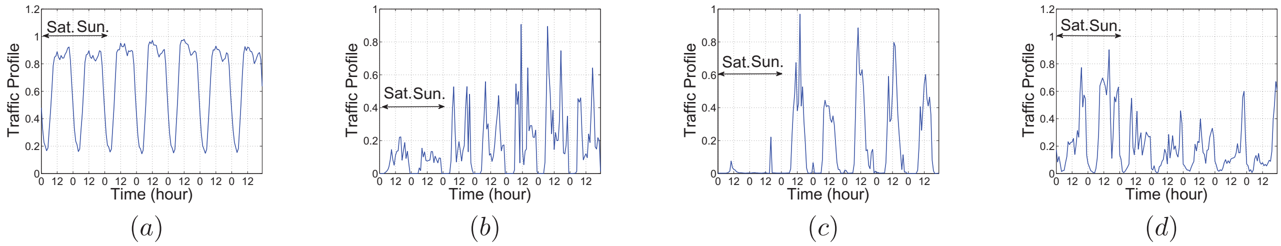

Fig. 1. (a) Normalized traffic profile during one week from data traffic information (overall traffic), (b) traffic profile of a base station near subway station, (c) traffic profile of a base station near CBD, (d) traffic profile of a base station near residential zone.

overall traffic, but exists a spatio-temporal differentiation when considering the local traffic variation.

TABLE I
THE AVERAGE AND VARIANCE OF PAR AND PVR IN FIG. 1

Base Station	Avg(R_a)	Var(R_a)	Avg(R_v)	Var(R_v)
Overall	1.41	2.92×10^{-4}	1.05	0.14×10^{-2}
Subway Station	3.20	0.57	5.65	13.70
CBD	6.14	26.56	15.96	1.34×10^{3}
Residential Zone	3.23	0.33	7.43	21.54

To further characterize the temporal inhomogeneity, we examine the distribution of PAR and PVR for the whole month at different urban area with its CCDFs, i.e., $P(R > r)$, plotted in Fig. 2. By exponential fittings we find an exponential decay for these temporal distribution. This result corresponds to the patterns of human activity. For example, since most people only stay at home during nights and go to work in the day, probability of high PVR are higher in BS near residential zone.

In terms of spatial inhomogeneity, we randomly select 100 BSs and compute their PAR and PVR value for one week. Their aggregated empirical CCDFs are plotted using log-log scale in Fig. 3, where we divide the data by two groups of weekday and weekend. Well approximated by $R^{-\alpha}$, their distributions follow a power law spatially. This is in sharp contrast to the log-normal and Weibull distribution of BS traffic density [2].

IV. CONCLUSION

In this poster, we presented our ongoing work on quantitatively characterizing the spatio-temporal inhomogeneity of data traffic in a large scale mobile cellular data network. We show a stable "tide" characteristic existed in overall traffic usage. This traffic "tide" decays exponentially in temporal and follows a power law distribution in spatial. Our future work will focus on what are the key factors that influence the spatio-temporal distribution of traffic inhomogeneity as well as how to utilize this modelled behavior of traffic "tide" to aid the management of mobile traffic congestion.

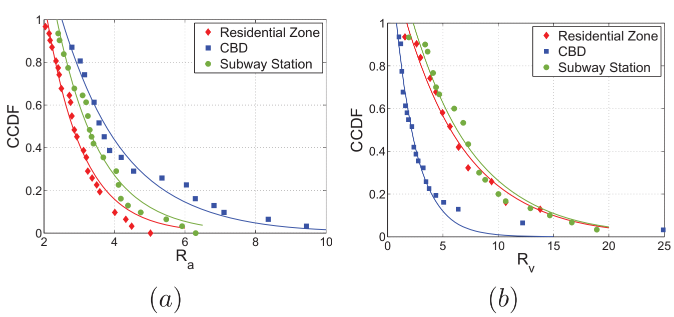

Fig. 2. (a) and (b) are the CCDFs of BS traffic PAR and PVR, respectively, for the whole month at different urban area. The solid exponential lines represent the analytical predictions of exponential distribution.

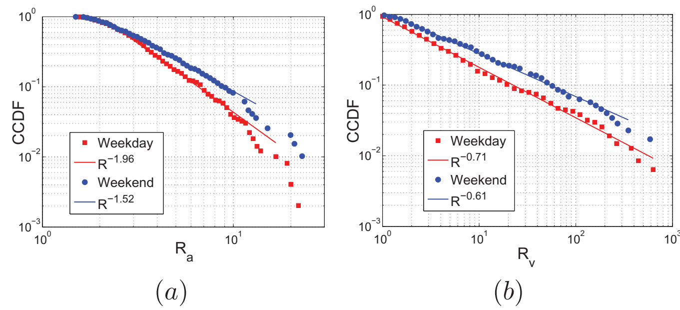

Fig. 3. (a) and (b) are the CCDFs of traffic PAR and PVR, respectively, for 100 BSs in log-log plot, at different periods of a week. The straight lines represent the analytical predictions by power law distribution

ACKNOWLEDGEMENT

This work is supported by the National Basic Research Program of China (973 Program) (No. 2013CB329001), National Nature Science Foundation of China (No. 61301080, No. 91338203, and No. 91338102), National High Technology Research and Development Program (No. 2013AA013501 and No. 2013AA013505), Chinese National Major Scientific and Technological Specialized Project (No. 2013ZX03002001), and Chinas Next Generation Internet (No. CNGI-12- 03-007).

REFERENCES

[1] Cisco Visual Networking Index, White Paper, Feb. 2014. available online at http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf, 2014. Accessed 2 January 2015.

[2] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang. Spatial modeling of the traffic density in cellular networks. *Wireless Communications, IEEE*, 21(1):80–88, 2014.

[3] M. Michalopoulou, J. Riihijarvi, and P. Mahonen. Towards characterizing primary usage in cellular networks: A traffic-based study. In *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on*, pages 652–655. IEEE, 2011.

[4] S. Sen, C. Joe-Wong, S. Ha, and M. Chiang. Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access. *Communications Magazine, IEEE*, 50(11):91–99, 2012.