An In-depth Analysis of 3G Traffic and Performance

Zhenxian Hu[+], Yi-Chao Chen[*], Lili Qiu[*], Guangtao Xue[+], Hongzi Zhu[+]
Nicholas Zhang[$], Cheng He[$], Lujia Pan[$], Caifeng He[$]
Shanghai Jiao Tong University[+], The University of Texas at Austin[*], Huawei[$]

ABSTRACT

As the popularity of cellular network grows explosively, it is increasingly important to develop an in-depth understanding of characteristics of cellular traffic and performance. In this paper, we analyze cellular network traces of different generations from three major cities in China during 2010 and 2013 and a city in Southeast Asian country in 2013. We analyze (i) cellular traffic, (ii) throughput, round-trip time (RTT), and loss rate, and (iii) how they are affected by cellular modes, time, and geographic locations. We find the TCP performance in our traces is much worse than those reported in previous 3G measurement studies from North America and Europe likely due to more expensive cellular data plan and more limited cellular resources. We further use machine learning to diagnose reasons behind high RTT and losses, and identify major factors that limit TCP throughput. Our analysis shed light on important characteristics of cellular traffic and performance in large-scale cellular networks.

1. INTRODUCTION

Motivation: The importance of cellular network performance has attracted significant work. In particular, considerable measurements have been conducted to characterize cellular traffic and performance in terms of traffic volume [5] diversity [5, 17], network performance [11, 12, 14, 15], temporal and spatial variation [16, 13].

Despite significant existing work, several open issues remain. First, the existing works focus on cellular networks in US and Europe, and our knowledge about the cellular networks in the rest of the world is much more limited even though they contain larger population. Second, most measurement works analyze the general characteristics of cellular traffic and their performance. In comparison, much fewer works try to understand factors accounting for the performance problems. Diagnosing cellular performance problems is challenging because there are a wide variety of factors that may lead to performance problems in cellular networks, such as the nature of mobile applications, specific characteristics of protocols at different network layers, and network connectivity.

Our approach: By April 2015, there are more than 467.66 million 3G subscriptions and 143.08 million 4G subscriptions in China. There are 1.91 billion 3G subscriptions in the world [6]. 3G is still widely used, so it is important to understand its traffic and performance.

In this paper, we analyze massive traces from large cellular data networks in several major cities in China and a Southeast country. The largest trace under study contains 13 TB, 27.6 billion packets, 383 million flows, 65K mobile devices, and 168 million Radio Resource Control (RRC) records. To our knowledge, this is the first large-scale study of 3G network measurement in Asia.

We analyze aggregate traffic, throughput, RTT, and loss rate, and how they are affected by cellular modes, applications, time, and geographic locations. Moreover, we further diagnose performance issues based on metrics across the network stack. We apply machine learning to diagnose reasons behind high RTT and loss rates. In addition, we examine major factors that limit the TCP throughput.

Our analyses reveal the following important characteristics of 3G traffic and performance: (i) TCP throughput in our traces is much lower than reported from the existing studies of North American and Europe traces likely due to more expensive cellular data plans in Asia. (ii) Most flows in cellular networks are short. They are shorter than reported in measurement studies of North American and Europe traces, likely due to more expensive data plans. (iii) 3G modes have significant impact on throughput and RTT, but not much on loss rate due to effectiveness of MAC-layer retransmission and TCP congestion control. (iv) TCP retransmission rate is low, but there is significant packet reordering in the downlink. So it is important to distinguish between congestion loss and packet reordering. (v) Traffic, RTT, and loss rates exhibit temporal stability and diurnal patterns. Traffic also exhibits Zipf-distribution in the granularity of individual users and sectors. Popular sectors/users are more likely popular in the next hour or the same hour next day. (vi) Traffic is moderately correlated between the sectors

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AllThingsCellular '15, August 17-21, 2015, London, UK

© 2015 ACM. ISBN 978-1-4503-3538-6/15/08…$15.00
DOI: http://dx.doi.org/10.1145/2785971.2785981

belonging to the same base station, but has little correlation between sectors belonging to different base stations. (vii) The major factors that contribute to high RTT and loss rates include poor wireless channel quality (*e.g.*, low SNR and high interference), high traffic load (*e.g.*, high load level at a base station and a large number of downlink/uplink users), low user categories (*e.g.*, UMTS as opposed to HSPA), and many soft handoffs. (viii) Over 90% traffic is limited by applications or opportunities, and above 4% traffic is limited by losses. Most of such traffic are generated by web browsing. Our findings have significant implications on the design and optimization of cellular networks and mobile applications.

2. BACKGROUND

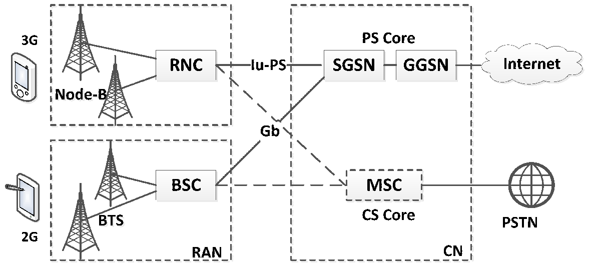


Figure 1: cellular network.

As Figure 1 shows, a typical UMTS network consists of a Radio Access Network (RAN) and a core network. The network elements are organized in a tree topology. Radio Network Controllers (RNC) controls multiple Base Stations (NodeBs), which connects with multiple User Equipments (UEs). Each NodeB is typically configured with multiple sectors (commonly 3 and up to 6 sectors) in different directions. The core network consists of Serving GPRS Support Nodes (SGSN) and Gateway GPRS Support Nodes (GGSN) to perform data access and charging functionality.

HSDPA increases the downlink rate over UMTS by using higher modulation and MAC layer retransmission. HSUPA improves the uplink by using a dedicated channel. The combination of HSDPA and HSUPA is called High Speed Packet Access (HSPA). TD-SCDMA is another type of air interface used in UMTS networks, mostly used in China by China Mobile, which serves the largest number of users in the world. TD-SCDMA is better at supporting asymmetric traffic by dynamically allocating time for uplink and downlink traffic.

3. DATASET AND PREPROCESSING

Datasets: Table 1 summarizes our traces. For confidentiality, we anonymize the city names to a, b, c. These cities are among the largest cities in China. In 2010, the cities had 8.7 million, 7 million, and 12 million population, respectively.

The first trace is most extensive in terms of types of data, duration, and volume. It contains one week traffic captured by the gateway integrated interface (Iu-PS) that connects 20 RNCs and the core network. The traces include the following information: (i) IP packet traces containing around 13TB traffic and 28 billion packets including both user plane packets and control plane packets using RANAP protocol, (ii) Radio Access Network Application Part (RANAP) elements: RANAP is the UMTS signaling protocol for the radio connection between a user equipment (UE) and the core network. The elements include all the signaling messages related to radio connection management, such as the messages of connection setup and release. (iii) PCHR logs: Maintained by RNC facility. They contain life cycle information of every Radio Resource Control (RRC) protocol functionalities, including information regarding service classes and uplink/downlink channel types, call types and serving cell and RNC ID, handover types and times during each RRC cycle. The other traces span shorter time scales, but offer diversity in terms of locations, time, and types of networks.

Network	Location	Time	Capture points	Volume	#Users
WCDMA	city a	11/25/10 - 12/01/10	lu-PS, 20RNCs	13 TB	65K
WCDMA	Southeast country	03/11/13	lu-Ps	897 GB	35K
TD-SCDMA	city b	11/01/13 - 11/05/13	lu-PS, 2RNCs	518 GB	15K
GPRS	city c	06/03/13	Gb	16 GB	2K

Table 1: Trace summary.

Data processing: To make our data processing scale to a large volume of data, we run our process on Hadoop [7] to parallelize the computation. After preprocessing, we keep three data tables in a Hive warehouse. The whole project builds up a 5-server platform, each with 24 CPU cores and 50GB memory.

4. TRAFFIC CHARACTERISTICS

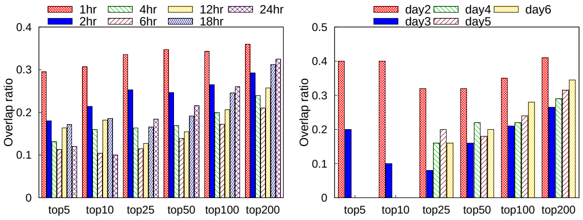
In this section, we analyze user traffic. Unless otherwise specified, we report results from WCDMA 2010 trace since this is the most comprehensive trace. For comparison, many of our analyses are also performed on the other traces.

Flow sizes: All traces have lots of short flows. For example, 90% of UMTS flows have fewer than 17KB and shorter than 39.5 seconds. In comparison, HSDPA and HSPA flows have more bytes and shorter in length due to higher capacity: 90% of their flows have within 36KB and 32KB, respectively, and last within 26.2 and 34.3 seconds, respectively. So it is important to optimize short flows in 3G networks. Moreover, WCDMA flows have similar duration as TD-SCDMA flows, but shorter than GPRS flows due to lighter traffic load and higher bandwidth in WCMDA.

Temporal characteristics: There is strong diurnal pattern. For example, the traffic volume is low during sleeping hours, rises significantly in the morning, and decreases during the late night. Predicting traffic using the same time on the previous day yields a prediction error of 13.3%.

Spatial characteristics: The WCDMA 2010 trace contains around 9500 sectors in total. We find the traffic across the top 100 loaded sectors in a decreasing order of volume on a log-log scale fits well with a straight line, which indicates a Zipf-like distribution (*i.e.*, the top i-th sector has C/i^{α} traffic), where α is 0.58 using least square fit. Similar results are observed on different days.

We further find the traffic across users in a decreasing order of volume. Again we observe a close fit with a straight line when plotting on a log-log scale, which also indicates Zipf-like distribution with $\alpha = 0.47$.



(a) across different hours (b) across different days

Figure 2: Overlap in the top N sectors.

Figure 2(a) shows the amount of overlap in the top N sectors between the first hour and the subsequent few hours for varying N. We observe the overlap is around or above 30% between two adjacent hours and decreases as the time gap increases since crowded base stations change across different time of day. Then it increases after the time gap increases to 12 hours and continues to increase further till the time gap reaches 24 hours due to diurnal pattern. Figure 2(b) further shows the overlap between different days. As we would expect, the overlap is largest on two consecutive days. Interestingly, it does not monotonically decrease with the number of days in between.

Next we try to understand the spatial correlation between the traffic across different sectors. We compute correlation coefficient of sectors at different distances using 30-minute time bins during 7am-12pm. We observe that sectors belonging to the same base station (*i.e.*, the distance between sectors is 0 km) have moderate correlation coefficient 0.48. The correlation coefficient of the other sectors drop to $0.16 - 0.33$. The correlation is even weaker in less busy periods. This information is useful for cellular provision (*e.g.*, designing load balancing algorithm).

5. TCP PERFORMANCE

In this section, we analyze TCP performance in terms of throughput, RTT, retransmission since 83% traffic uses TCP.

Throughput: Figure 3(a) and (b) plot CDF of uplink and downlink throughput in the WCDMA 2010 trace, respectively. In the uplink (*i.e.*, from a mobile), the average throughput of UMTS, HSDPA, and HSPA are 5.3 Kbps, 4.3 Kbps, and 5.6 Kbps, respectively. In the downlink (*i.e.*, towards a mobile), their average throughput are 13.2 Kbps, 44.3 Kbps, and 33.6 Kbps, respectively. The throughput does not monotonically increase with the 3G mode because most flows are limited by the applications instead of the network as we will show in Section 7. Moreover, the throughput in our traces is 1-2 orders of magnitude lower than the 3G measurement studies from US and Europe (*e.g.*, [10] reports 556-970 Kbps median downlink throughput and 207-331 Kbps median uplink throughput, and [21] reports 300 Kbps - over 1 Mbps). This is likely due to expensive data plans in China during the time of the trace collection. Most flows contain few packets (as shown in Section 7) and small flows usually do not saturate the link. For example, a 5 GB data plan costs around \$20/month in US, whereas a similar plan costs around \$150/month in China [1]. Since the average monthly income in China is below \$1000/month, only a small fraction of people can afford such expensive data rate plans.

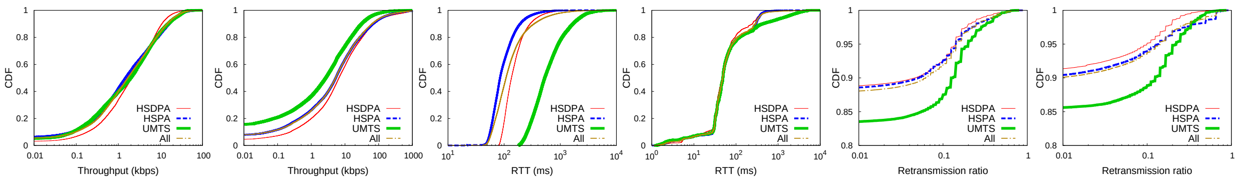
We observe WCDMA 2010 has the highest throughput, followed by TD-SCDMA 2013, WCDMA 2013, and then GPRS. GPRS experiences the worst performance as expected.

RTT: As Figure 1 shows, the data path can be viewed as: mobile device $-$ base station $-$ RNC $-$ GGSN $-$ server. We capture traces at the RNC. By taking the difference between the time of sending data from RNC to a mobile device and the time of receiving the corresponding ACK from the mobile to the RNC, we get the cellular delay. Figure 3(c) plots the cellular delay for all traffic, UMTS traffic, HSDPA traffic, and HSPA traffic. As we would expect, UMTS $>$ HSDPA $>$ HSPA. For example, the median RTT in UMTS, HSDPA, and HSPA are 527.87 ms, 134.80 ms, 84.52 ms, respectively. Figure 3(d) plots the Internet delay (*i.e.*, delay from RNC to the server and back to the RNC). The Internet delay is similar across different cellular modes. Their median RTTs are all around 46-49 ms, although a few UMTS flows experience noticeably higher upstream RTT. Both the cellular and Internet delay vary significantly over time.

Retransmission rate: We estimate the TCP loss rate based on the number of retransmissions. This is an approximation since TCP may sometimes pre-maturely retransmit (*e.g.*, due to a too short TCP timeout). But such pre-mature retransmissions are likely to be low due to the effectiveness of TCP retransmission mechanism. Figure 3(e) and (f) show the retransmission rates from RNC to the server and from RNC to the client, respectively. The former approximates the uplink loss rate, and the latter approximates the downlink loss rate. The actual loss rates can be higher since some retransmissions may be lost before reaching RNC. The average uplink retransmission rates of UMTS, HSDPA, and HSPA traffic are 3.00%, 1.71%, and 1.95%, respectively. The corresponding numbers for the downlink are 2.90%, 1.18%, and 2.18%, respectively. Both retransmission rates are low due to the effectiveness of wireless MAC layer retransmission and TCP congestion control. While the absolute loss rates are low, they are higher than reported in US 3G cellular studies (*e.g.*, [21] reports 0.1-0.3% loss rates from AT&T, Sprint, Verizon in MA, US). The root cause is analyzed in Section 6.

We also compare the retransmission rates of different traces. The lightly loaded WCDMA has the lowest retransmission rate, followed by TD-SCDMA and GPRS, as expected.

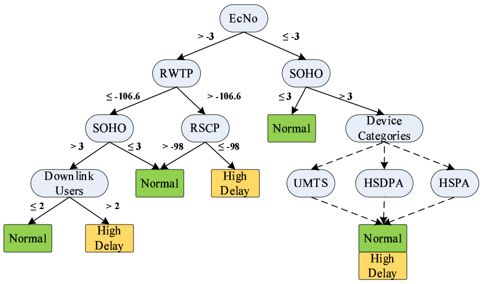
Packet reordering: The downlink packet reordering rate is higher than the uplink in all traces. GPRS and TD-SCDMA see reordering in less than 1% uplink traffic, whereas WCDMA see a nearly zero uplink reordering rate likely due to lighter load in the WCDMA traces. In comparison, 6 - 30% downlink traffic see reordering. TCP treats significant packet reorders as congestion losses and responds by cutting the sending rate. In order to avoid TCP reducing sending rate, the



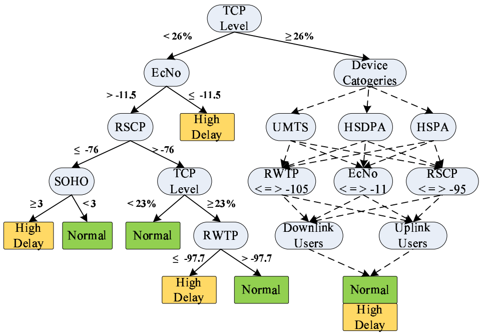
(a) Uplink throughput (b) Downlink throughput (c) Cellular delay (d) Internet delay (e) Uplink loss (f) Downlink loss

Figure 3: CDF of performance for different cellular modes in 2010 WCDMA trace.

receiver should differentiate between packet reordering and congestion losses, and send feedback to the sender.

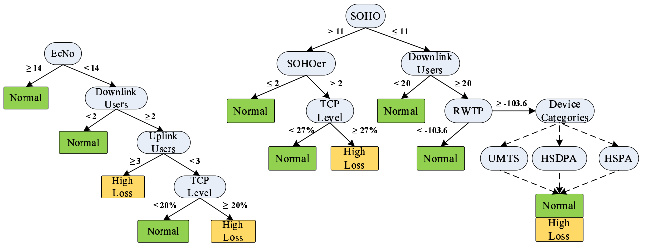


(a) 2010 WCDMA



(b) 2013 WCDMA

Figure 4: Decision trees to diagnose high RTT. Note that there are different versions of UMTS, HSDPA, and HSPA, and each of these version involve different tests. For simplicity, we abstract them to only three nodes, corresponding to UMTS, HSDPA, and HSPA and use dotted lines to connect to the other testing conditions.



(a) 2010 WCDMA (b) 2013 WCDMA

Figure 5: Decision trees to diagnose high losses.

6. DELAY AND LOSS DIAGNOSIS

In this section, we diagnose reasons for high delay and losses. We compile a comprehensive list of features from the information in the capturing Iu interface and PCHR data sets. The list includes 15 features. Then we use machine learning to identify the most important features that contribute to the high delay and loss rates and how they impact the performance. We use J48 pruned decision tree [2], which is a widely used decision tree and uses pruning to avoid overfitting. We also use support vector machine (SVM). Decision tree is from Weka [20], a popular suite of machine learning software. SVM is from LibSVM [4]. We use 10-fold cross validation, which partitions the traces into 10 pieces and uses 9 of them as training and the remaining as testing. We repeat this process 10 times and each time using a different piece for testing. The accuracy is defined as the fraction of RRC connections that are correctly labeled as high delay/losses in the testing traces. We report the average accuracy over the 10 runs.

We compute RTT and loss rates for each RRC. We consider the RTT that exceeds the 85th-percentile as high RTT, whereas the other as normal RTT. We consider loss rates above 1% as high losses in 2010 WCDMA traces, and above 4% as high losses in 2013 WCDMA traces. These loss thresholds correspond to 80-th percentile loss rates in the traces. Based on these thresholds, each RRC connection is labeled as either high RTT/losses or normal RTT/losses. We feed the features of each RRC along with its label in the training traces to the decision tree or SVM, and apply the learned tree/SVM to the testing traces.

The accuracy of both decision tree and SVM is high for RTT and losses in 2010 and 2013 WCMDA traces: between 0.85 and 0.90. Below we look into the decision tree results since it is more intuitive.

Figure 4 and Figure 5 show the output decision trees for RTT and losses, respectively. The features used in the decision tree are as follow:

- Wireless channel quality: The following three metrics are related to channel quality: (i) *Energy per chip of the pilot channel over the noise power density (EcNo)* reflects how strong a signal is above the noise in a cell (sector) and measured in dB. It is also known as signal-to-noise ratio (SNR) per bit. EcNo is measured on the pilot channel and thus may be different from the SNR of the data traffic channel. A higher EcNo indicates a better channel. (ii) *Received Signal Code Power (RSCP)* is the downlink power received by the UE receiver on the pilot channel and measured in dBm. A higher RSCP indicates a better channel. (iii) *Received Total Wideband Power (RTWP):* It measures the total level of noise within the UMTS frequency band of any cell and captures uplink interference. A lower RTWP indicates a better channel. An unloaded network has RTWP around -105 dBm, RTWP of -95 dBm indicates some interference, and RTWP of -85 dBm indicates strong interference.

- Numbers of soft handovers (SOHO) and softer handovers (SOHOer). A soft handover occurs when a UE is connected to a new cell where it establishes the new link before the removal of old one. A SOHOer handover is a soft handover between the neighboring cells (sectors) in one base station.
- Device categories: There are over 10 device categories. Here we simplify them according to the main edition of their code in the PCHR log and group them into three generations: UMTS, HSDPA, HSPA. Each main generation contains several smaller generations that differ in device capability. As we would expect, they have significant impact on the performance. A device with a lower category (*e.g.*, UMTS) is more likely to incur performance issues.
- Numbers of downlink and uplink users in a cell.
- TCP level per cell reflects the load and interference level of a cell. Note that TCP here does not refer to the transport layer protocol. A higher TCP indicates more congestion.

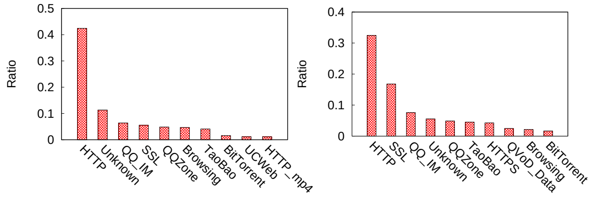
As shown in Figure 4(a), the high delay in the 2010 traces is mostly caused by poor wireless channel quality (*e.g.*, low EcNo, low RSCP, or high RTWP) and many handoffs. In comparison, the decision tree for the 2013 traces, shown in Figure 4(b), involves additional metrics of the traffic load, such as TCP level and the number of uplink users. More load-related metrics are involved in the 2013 decision tree because the 2013 traces see higher load and have high RTT due to congestion. When the TCP level is below 26%, RTT is mainly influenced by many wireless channel features, like EcNo, RSCP, RTWP (in the left subtree); otherwise, the difference of device categories starts to matter, as we would expect. For the lowest category, whenever the TCP level is above 26%, it incurs high delay. For the other categories, delay can be normal when TCP is above 26% (*e.g.*, under good wireless channel quality and small numbers of uplink and downlink users).

Next we diagnose high loss rates. Figure 5 shows the resulting decision trees. The 2010 tree is relatively simple, and reflects that loss rate is closely related with SNR and traffic load. When EcNo is larger than -14 dB, the probability of high loss rate ($> 1\%$) is smaller than 10%. When EcNo is lower than -14 dB, losses arise when the traffic load (measured in terms of the number of downlink and uplink users and TCP level) is high. In comparison, the 2013 tree is more complicated. It includes additional metrics, such as SOHO, SOHOer, RTWP, and Device Categories. High loss arises when the number of soft handovers are large and TCP level is high or when the number of downlink users is large and interference is strong. In the latter case, the loss also depends on the device categories, and the lower device categories are more susceptible to losses.

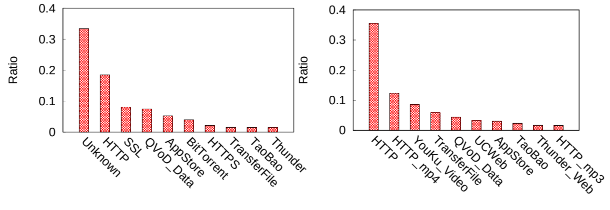
To summarize, we observe that wireless channel quality is very important in all cases, as we would expect. Second, TCP level has a significant impact on network performance. The 2010 WCDMA network was a trial system at the time and was a lightly loaded, so TCP is less important. For more loaded network in the 2013 traces, the impact of TCP increases. Third, device categories matter. More advanced categories are more resistant to many bad environmental factors and more likely to enjoy better performance.

7. THROUGHPUT DIAGNOSIS

We leverage T-RAT tool [22] to analyze TCP flows to determine the factors that limit TCP throughput. The tool is tailored to wireline networks, so we extend it to support wireless networks. Specifically, based on the tool with our modifications, we classify the flows into one of the followings: (i) *Opportunity limited:* A flow shorter than 13 MSS (*i.e.*, maximum segment size) or staying at slow-start. [22] uses 13 MSS for thresholding since it is hard to tell whether a flow is in slow start or congestion avoidance if it has fewer than 13 MSS. (ii) *Application limited:* A packet smaller than MSS was sent followed by an idle interval larger than the RTT. In this case, the application does not generate enough data packets to fully utilize the network. (iii) *Bandwidth limited:* The rate before a loss occurs is consistently close to the 3G data rate after accounting for the protocol overhead including header overhead. (iv) *Loss limited:* A flow that incurs a loss. (v) *Receiver window limited:* A flow has 3 consecutive flights with flight sizes $S_i \times MSS > awnd_{max} - 3 \times MSS$, where $awnd_{max}$ is the largest receiver advertised window size and $3 \times MSS$ aims to account for variation due to delayed ACKs. (vi) *Sender window limited:* A flow (1) is not sender or receiver window limited, congestion limited, or bandwidth limited, (2) have $S_{F_{80}} < S_{F_{med}} + 3$, and (3) have four consecutive flights with flight sizes between $S_{F_{80}} - 2$ and $S_{F_{80}} + 1$, where $S_{F_{med}}$ and $S_{F_{80}}$ denote the median and 80th percentile of flight size, respectively. (vii) *Transport limited:* A flow enters congestion avoidance, has no loss, and its flight size continues to increase. (viii) *Host window limited:* A flow passes the above sender window limited condition but its ACKs are not present. It is either sender or receiver window limited, but cannot be determined which one.



(a) Opportunity limited (b) Application limited



(c) Congestion limited (d) Transport limited

Figure 6: Traffic mix of 2010 WCDMA trace.

We analyze the 2010 WCDMA trace during 10am to 12am.

We observe application-limited and opportunity-limited are the dominant factors that limit the throughput. They each account for 26.19% and 65.88% traffic, respectively. In addition, congestion-limited and transport-limited account for 4.12% and 1.21%, respectively. While the percentages are low, they are still significant and affect over 1 million flows during 10-10:30am. Figure 6 further shows the distributions of applications that account for the application- and opportunity- limited flows. Web browsing, streaming, IM account for 40%, 15%, and 2% traffic, respectively. The results are similar in the other traces.

8. RELATED WORK

Characterizing cellular traffic: [5] examines application usage patterns and interaction time of 255 smartphone users and finds aggregate and device-specification application volume follows a Zipf-like distribution. [17] develops an iPhone 3G application to measure 25 smartphone users' usage.

Characterizing cellular performance: The authors of [10] deploy 3GTest to collect and analyze performance measurements. [8] extends the tool to analyze LTE users. The DARWIN+ group collects IP packets at a GPRS/UMTS network, and analyzes various issues, such as TCP performance and traffic anomalies [14]. [3] analyzes the impact of 3G wireless link latency and bandwidth on TCP performance based on the traces collected from the user devices. [19] uses TCP and video experiments to learn the capacity of a 3G link and reports that it is highly unpredictable due to customized engineering at each cell. Paul *et al.* [13] analyze data packet headers and various signaling messages in a national 3G network, and study their temporal and spatial variations. [15] reports significant performance degradation in two high-profile crowded events in 2012 and suggests more aggressive release of radio resources to reduce radio resource wastage. [9] studies the interactions between applications, transport protocol, and the radio layer in a large LTE network in US. It reports 52.6% of TCP flows are throttled by TCP receive window.

Our study is the largest-scale 3G measurement in Asia, whereas the existing works (*e.g.*, [10, 18, 9]) focus on US and Europe. We also go beyond the characterization of traffic and performance, and diagnose reasons that contribute to high delay, high loss, and low throughput.

9. CONCLUSION

In this paper, we analyze cellular traces from several major cities in China and a Southeast country. We report a range of important characteristics of the cellular traffic and performance. These observations have significant implications on the protocol and application design. In particular, we can accurately identify high RTT/losses using a combination of metrics, such as wireless channel quality, traffic load, and the number of soft handovers. Moreover, most of traffic are limited by applications or opportunities. Techniques such as persistent HTTP, multiple TCP connections, and prioritization, are likely to be helpful in improving network utilization and reduce perceived delay. In addition, there is considerable TCP reordering in 3G downlinks, and it is important to distinguish TCP reordering from congestion losses and avoid unnecessary TCP window reduction. Our work complements the existing work by providing valuable data points about cellular network traffic and performance in Asia and performing a fine-grained diagnosis to pinpoint the root cause for performance issues.

Acknowledgement

This work is supported in part by NSF grant CNS-1343383.

10. REFERENCES

[1] Cellular rate plan. http://wenku.baidu.com/link?url=NH6s6E_LuUcQWUI2dTZIJI9VtXwJWhLhPYRULLd%n2a1OA_q3yjxyWsc3fMw0Qphdjxiv0_yov56sQZwK_w2UHOKx08IA3UEx59MGlp9pKzi.

[2] N. Bhargava, G. Sharma, R. Bhargava, and M. Mathuria. Decision tree analysis on j48 algorithm for data mining. *Int. Journal of Advanced Research in Computer Science and Software Engineering*, June 2013.

[3] M. C. Chan and R. Ramjee. TCP/IP performance over 3G wireless links with rate and delay variation. In *Proc. of ACM MobiCom*, 2002.

[4] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011. Software at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[5] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, and R. G. D. Estrin. Diversity in smartphone usage. In *Proc. of MobiSys*, 2010.

[6] GSA 3G stats. http://www.gsacom.com/news/statistics.php4.

[7] Hadoop. http://hadoop.apache.org/.

[8] J. Huang, F. Qian, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck. A close examination of performance and power characteristics of 4G LTE networks. In *Proc. of MobiSys*, pages 225–238, 2012.

[9] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck. An in-depth study of LTE: effect of network protocol and application behavior on performance. In *Proc. of SIGCOMM*, 2013.

[10] J. Huang, Q. Xu, B. Tiwana, Z. M. Mao, M. Zhang, and P. Bahl. Anatomizing application performance differences on smartphones. In *Proc. of MobiSys*, 2010.

[11] J. Kilpi and P. Lassila. Micro- and macroscopic analysis of RTT variability in GPRS and UMTS networks. In *Proc. of Networking*, pages 1176–1181, 2006.

[12] K. Mattar, A. Sridharan, H. Zang, I. Matta, and A. Bestavros. TCP over CDMA2000 networks: a cross-layer measurement study. In *Proc. of PAM*, 2007.

[13] U. Paul, A. Subramanian, M. Buddhikot, and S. Das. Understanding traffic dynamics in cellular data networks. In *Proc. of INFOCOM*, 2011.

[14] P. Romirer-Maierhofer, A. Coluccia, and T. Witek. On the use of TCP passive measurements for anomaly detection: a case study from an operational 3G network. In *Proc. of TMA*, 2010.

[15] M. Z. Shafiq, L. Ji, A. Liu, J. Pang, S. Venkataraman, and J. Wang. A first look at cellular network performance during crowded event. In *Proc. of ACM SIGMETRICS*, 2013.

[16] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang. Characterizing geospatial dynamics of application usage in a 3G cellular data network. In *Proc. of IEEE INFOCOM*, 2012.

[17] C. Shepard, A. Rahmati, C. Tossell, L. Zhong, and P. Kortum. Livelab: Measuring wireless networks and smartphone users in the field. In *Proc. of HotMetrics*, 2010.

[18] J. Sommers and P. Barford. Cell vs. WiFi: On the performance of metro area mobile connections. In *Proc. of IMC*, 2012.

[19] W. L. Tan, F. Lam, and W. C. Lau. An empirical study on the capacity and performance of 3G networks. *IEEE Tran. on Mobile Computing*, 2008.

[20] Weka. http://www.cs.waikato.ac.nz/ml/weka/.

[21] E. M. N. Yung-Chih Chen, R. J. Gibbens, D. Towsley, and Y. sup Lim. Characterizing 4G and 3G networks: Supporting mobility with multi-path TCP. *UMass Tech. Report: UM-CS-2012-022*, 2012.

[22] Y. Zhang, L. Breslau, V. Paxson, and S. Shenker. On the characteristics and origins of internet flow rates. In *Proc. of ACM SIGCOMM*, Aug. 2002.