Characterizing Smart Phone Traffic: A Mobile Carrier's Perspective

Mahmoud Taghizadeh* Haiyong Xie†‡ Cheng He‡ Gong Zhang‡

*ECE Dept, Michigan State University, Michigan, MI
Email: taghizad@msu.edu
†Suzhou Institute for Advanced Study, University of Science and Technology of China, China
Email: haiyong.xie@ustc.edu
‡Corporate Research, Huawei Technologies, China
Email: {hecheng, nicholas.zhang}@huawei.com

***Abstract*—In this paper we study the smart phone traffic and analyze its characterization from a mobile network operator's perspective. We collected and base out study on packet-level traces captured in a tier-1 cellular network serving a populated region in China. Our results, largely consistent with existing studies from end users' perspectives, show that web browsing contributes more than half of the traffic. Our detailed analysis, however, reveals some unique user browsing behaviors that have not been shown in other studies.**

I. Introduction

IP traffic in cellular networks has been growing rapidly and significantly faster than regular Internet traffic in recent years. A key reason that contributed to this trend is that the recently emerging smart mobile devices and wireless-enabled data applications have fostered new content dissemination models in today's cellular networks. In addition, it has been predicted that mobile traffic will grow 10 times faster than regular Internet traffic (see, *e.g.*, [1], [2]). Not surprisingly, it is likely that most of such mobile traffic is generated by smart phones, which are becoming more popular and prevalent than desktop PCs.

There are mainly two factors contributing the most to the rapid growth of cellular IP traffic. The first factor is the recent advances in mobile devices (*e.g.*, iPhone, iPad, Android-based smart phones), and the proliferation of such devices. In particular, the significant improvements on mobile devices' capability of graphics, storage and computation as well as improvements on availability of wireless bandwidth (*e.g.*, GPRS, 3G, 4G) have become a key factor.

The second factor is the emergence and proliferation of mobile data applications. Typical mobile applications include but are not limited to web browsers, e-mail clients, weather and stock applications, and mobile games. The level of proliferation of mobile applications is exemplified by the fact that as of August 2011, Apple's App store offered over 100,000 applications that can be downloaded by iPhone/iPad users.

The rapid growth of IP traffic in cellular networks motivates us to study mobile traffic in more detail. There have been only a few efforts to reveal the characteristics of mobile traffic, for instance, characterization of network traffic generated by web applications in a metropolitan 3G network (see, *e.g.*, [3]) and network traffic generated by smart phones (see, *e.g.*, [4]). Most of these efforts characterize network traffic from the perspectives of applications or end users.

Different from these studies, we are more interested in understanding the properties of IP traffic in cellular networks from a mobile network carrier's perspective. Specifically, we distinguish our work from others' in the following ways. Firstly, our work is one of the very few studies of mobile traffic in a non-US based carrier. This carrier is one of the largest mobile carriers in the world. Secondly, our trace contains users with two different access technologies which helps us to provide an insight into the high-level composition of the global traffic in both technologies. Thirdly, although duration of our trace is relatively limited, it contains the detailed traffic of a large number of mobile users, which help us to statistically characterize the user-generated traffic. Last but not least, in addition to IP-level traffic analysis, we conduct a thorough analysis on web traffic which is the most prevalent part of the traffic.

Studies on mobile network traffic from the perspective of mobile network carriers could be beneficial to provision mobile networks and may even shed light on the directions towards which today's mobile networks evolve. For instance, analyses of mobile network traffic are important for carriers to understand the traffic mix, identify and fix critical bottlenecks in their networks. The result of our studies can also be used for capacity planning, network troubleshooting or even build models of bandwidth cost saving. On the other hand, by comparing the characterizations of the traffic generated by two different access technologies, we may be able to identify the trend that both the mobile networks and their traffic evolve towards.

The remainder of this paper is organized as follows. In Section II we describe the data set used in our study. In Section III we present the results of IP-level traffic analysis. In Section IV we present our analysis on web traffic. In Section V we discuss the implications of our findings on radio power management. In Section VI we present the related works. We conclude the paper with future works in Section VII.

II. Data Set

We first describe the data set used in our studies.

978-3-901882-50-0

A. Background

We illustrate in Fig. 1 the main components and the interfaces between components in the UMTS cellular network in a layered structure. The first layer comprises user equipments (UEs). Radio Access Network (RAN) is the second layer which consist of base stations (node-B) and radio network controllers (RNCs). The RAN controllers connect to core layer through Serving GPRS Support Nodes (SGSNs). In the core layer, the SGSN converts the mobile data into IP packets and send them to the Gateway GPRS Support Nodes (GGSN) through the GPRS Tunneling Protocol (GTP). The GGSN serves as the gateway between the cellular core network and the Internet. This means every IP packet sent to a UE has to go through the GGSN.

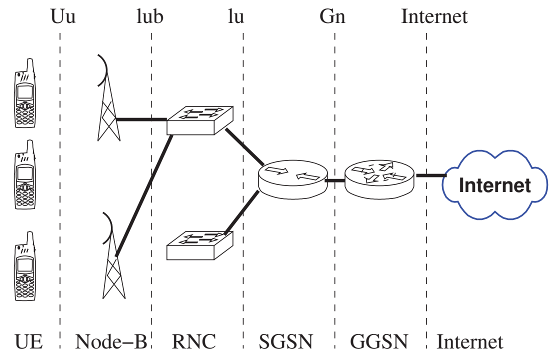

Fig. 1. The UMTS network architecture

B. Data Set

We capture the packet-level IP traffic data from the Gn interface (*i.e.*, the interface between SGSN and GGSN) in the core network of a tier-1 regional cellular network in China. The regional network is a part of a national cellular network and serves mobile users in a highly populated region in China. Both the GPRS and 3G technologies are deployed in this network. The data capturing process lasts seven days. The packet-level trace allows us to characterize both the IP and application traffic pattern of mobile devices.

1) Purify Data Set: In the early stage of our study, we notice that some IP addresses have exchanged a significantly larger amount of traffic than most others. Through a more detailed analysis we find that the data set contains the traffic generated for some internal servers and tethered desktop workstations in the carrier network, rather than for mobile devices. Since we are mainly concerned with mobile traffic generated for mobile devices, we have to distinguish the noise traffic (*i.e.*, the non-mobile traffic).

We distinguish mobile traffic from other captured non-mobile traffic in multiple attempts. In the first attempt, we try to use the Time-To-Live (TTL) value to distinguish IP addresses of mobile devices and internal servers. The rationale behind this attempt is that the default TTL values of IP packets generated by some popular mobile devices (*e.g.*, Nokia's handsets) significantly differ from those generated by desktops/laptops. However, many hand-held devices use the same default TTL value of regular operating systems such as Linux and Windows. Therefore, we cannot rely only on this approach.

In the second attempt, we try to use the average Round Trip Time (RTT) values of packets from and to the clients[1] to differentiate IP addresses of mobile devices and regular workstations. The rationale is that the average RTT value of packets generated by internal servers and workstations must be much smaller that that generated by mobile devices, since packets generated by the latter have to go through 3G or 2G wireless links which typically incur high delays to packets. Unfortunately the actual measurement based on our data set reveals that RTT is varying significantly and is not a reliable factor to classify traffic.

Finally, in the third attempt, we use the User-Agent header in HTTP requests to distinguish mobile and regular traffic. Many of mobile device manufacturers (*e.g.*, Nokia, Motorola, and Apple) use their brand names in the User-Agent header. Therefore, we can easily identify those mobile devices and their IP addresses using the User-Agent header. However, it remains a challenge to differentiate mobile devices and regular workstation, both of which may use Mozilla as their user agents. Apparently the user agent information may not be sufficient to distinguish mobile devices from regular workstations. Thus, we also take into account the information collected from users' web access traces. For instance, many operating systems (in particular Windows operating systems) as well as many Anti-virus software for desktop workstations typically automatically download OS/software patches and updates from their Internet servers; however, mobile devices typically do not exhibit such behaviors. By investigating the clients' web access traces, we are able to identify and distinguish a majority of desktop workstations from mobile devices.

2) Mobile Users: The mobile devices have two classes of IP addresses: more than 70% users are assigned IP addresses from the network prefix `10.0.0.0/8` and the remaining 30% of users assigned addresses from the network prefix `172.0.0.0/8`. The `10.0.0.0/8` and the `172.0.0.0/8` IP addresses are assigned to mobile devices adopting GPRS and 3G technologies, respectively. Throughout this paper, we refer to these two networks and their corresponding users as *net2g+* and *net3g* respectively. Thus, our data set consists of traffic from/to both networks, which enables us to compare the characterizations of traffic and applications in both networks.

TABLE I. USER AGENT SUMMARY

Device Type	% Percentage	
	net2g+	net3g
Unknown	7	9
WAP	30	0
Nokia	36	54
Other brands	27	20
iPhone	0	13
Windows Mobile	0	4

Table I summarizes the percentage of mobile users in both networks based on the observed User-Agent header in the HTTP requests. We categorize the users based on the mobile device brands (*e.g.*, Nokia and Sony-Ericsson) or browsers (*e.g.*, WAP). 7% of user agents in net2g+ and 9% of user agents in net3g are unknown.

[1]The terms "users" and "clients" are used interchangeably throughout this paper.

During our analysis, we notice that many of users (especially those in net3g) send and receive only a few packets; for instance, over 72% of net3g users send/receive less than 5 packets. With a careful analysis, we find that almost all of such tiny-volume traffic is generated mainly by a malware software, which scanned the network to find new victims. The presence of such unwanted traffic is not unexpected, since nowadays laptops with 3G datacards – often equipped with popular operating systems – coexist with handsets and smart phones in 3G networks, and it is well known that the unwanted traffic is a steady component of the traffic in the wired networks. As this traffic is unwanted and does not express user preference, we filter it out from the trace prior to our analysis.

III. Mobile Traffic Composition

We next statistically characterize the smart phone traffic in this section. Specifically, we first analyze the traffic mix, then investigate the amount of traffic being sent and received by different clients and Internet servers, and finally analyze the traffic mix from the network carrier's perspective.

A. Traffic Mix

In our traces, we observe that approximately 90% of packets are TCP packets, which contribute 87% of the total traffic in net2g+, while 91% of packets are TCP packets contributing 94% of the total traffic in net3g. This implies that TCP is dominant in mobile traffic. In addition, net2g+ and net3g have similar percentages of TCP traffic (in terms of both byte count and packet count).

We also investigate the traffic mix for different applications using TCP/UDP ports.[2] Table II and III report the TCP/UDP ports which contribute to at least 1% of packets and/or traffic. In both networks, around 50% of packets are HTTP packets. These packets contribute to 78% and 69% of the total TCP traffic in net2g+ and net3g respectively.

TABLE II. TCP PORTS USED IN IP PACKETS

TCP Port	% of (Packets,Bytes)	
	net2g+	net3g
80	(55,78)	(48,69)
14000	(38,18)	(26,7)
443	(1,1)	(2,1)
110	(1,1)	(6,10)

TABLE III. UDP PORTS USED IN IP PACKETS

UDP Port	% of (Packets,Bytes)	
	net2g+	net3g
9201	(40,12)	(0,0)
53	(1,1)	(19,8)
137	(1,1)	(0,0)
5000	(0,0)	(40,78)

It is clear that HTTP traffic is the dominant traffic in both networks. This should not come as a surprise, as Internet browsing is one of the most popular uses of smart phones and HTTP has become the workhorse of many other applications including audio/video streaming. However, the percentages of HTTP traffic reported in this paper is significantly smaller than 80-97% reported in [6]. The main difference comes from the share of TCP traffic on port 14000 which is used by QQ, a very popular instant messaging application in China. Additionally, the percentage of HTTPS traffic (on TCP port 443) is significantly lower than 37% reported in [4], [6].

Email traffic on TCP port 110 contributes only 1% of traffic in net2g+. The main reason is that mobile devices in net2g+ are largely old smart phones which use the WAP protocol for Internet access, while devices in net3g are mostly more powerful (*e.g.*, iPhone and Windows smart phones) and capable of running different applications including Email. As a result, we observe that around 10% of TCP traffic of users in net3g is generated by Mail application.

In net3g, 19% of UDP packets are DNS (on UDP port 53) packets which contribute 8% of all UDP traffic. However, DNS traffic is extremely low in net2g+. The reason is that net2g+ users mainly rely on a single proxy for Internet access and therefore domain name resolution is done by the proxy. 40% of UDP packets in net2g+ also belong to UDP port 9201 which is used by the same proxy to provide WAP service. 40% of UDP packets in net3g are generated on port 5000 which belongs to the aforementioned QQ instant messaging application.

B. Client Perspective

Fig. 2 and Fig. 3 plot the cumulative distribution of traffic received by users in both networks.

We observe that 80% of users in both networks receive less than 100KB of traffic, as shown in Fig. 2. The amount of exchanged traffic by these users are also very similar. The percentage of active users (those who send and receive more than 1MB) in net3g is slightly higher than that in net2g+. Almost 5% of users in net3g receive more than 1MB whereas this number in net2g+ is only 0.5%. This can be further observed in Fig. 3 which shows the actual amount of downstream traffic ranked by users.

Three factors may explain this difference. First, mobile devices in net2g+ are mostly older smart phones. Second, a small percentage of users in net3g use desktop or laptop (with 3G data cards) to access the 3G network, whereas we do not find any laptop or desktop in net2g+.[3] Third, a majority of net2g+ users lack regular web browsers and rely on WAP to download content objects form the web. Although not reported in this paper, we observed a very similar trend for upstream traffic of clients.

C. Server Perspective

We next report the contribution of different servers on the total downstream traffic, as shown in Fig. 4. Interestingly the trends for both networks are very similar. Almost 70% of servers send less than 100KB of traffic, and 10% of servers send more than 1MB. However, more than 90% of the total downstream traffic in both networks is generated by these top 10% servers.

We further analyze the correlation between the downstream traffic and the servers sorted in the descending order of their

[2] Although identifying applications using port is not very accurate in some cases, it is a simple indicator that works well in most of cases (see, *e.g.*,[5]).

[3] Although we separated mobile and non-mobile devices using user agent information and access behaviors, very few laptop and windows desktop may still exit in the trace.

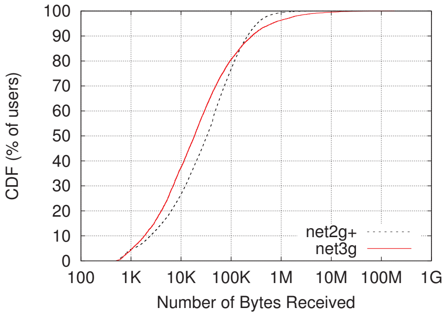

Fig. 2. Client perspective: traffic per user

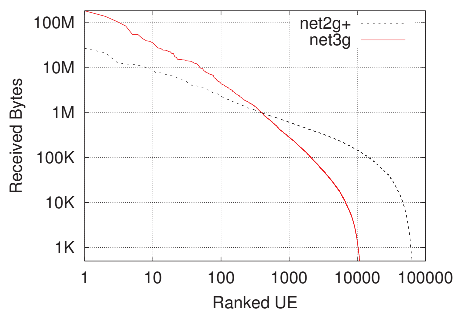

Fig. 3. Client perspective: ranked users

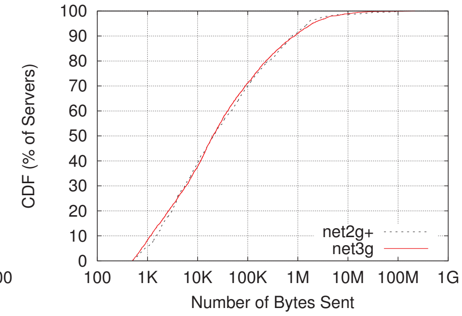

Fig. 4. Server perspective: traffic per server

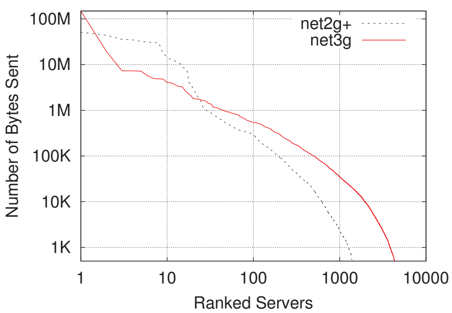

Fig. 5. Server perspective: ranked servers

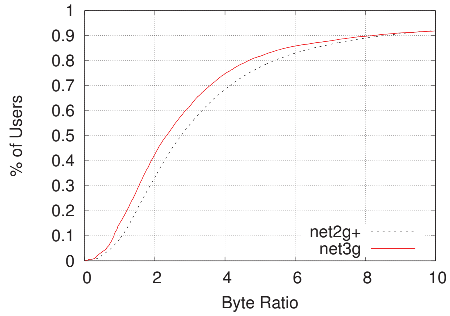

Fig. 6. Network perspective: byte ratio

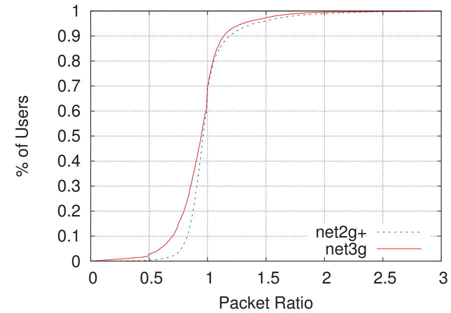

Fig. 7. Network perspective: packet ratio

outbound traffic volumes, as shown in Fig. 5. Note that the graph is plotted in the log-log scale.

We make the following observations. Firstly, the number of servers accessed by net3g users is much larger that that by net2g+. This means that the diversity of servers visited by net2g+ users is significantly less than that by net3g users. A key reason is that both the access technology and mobile devices in net3g are more advanced, allowing many applications (*e.g.*, video streaming over HTTP and instant messaging) to proliferate; while users in net2g+ have to use WAP to access the Internet, which significantly limits the proliferation of mobile applications.

Secondly, the top 10 servers have a completely different trend than the remaining servers in terms of outbound traffic volumes. Manual analysis reveals that these servers mainly belong to a very popular portal in china (*i.e.*, `qq.com`) which are mainly used for Web browsing and instant messaging. Moreover, a further investigation reveals that the top 12 servers accessed by net2g+ and net3g users are responsible for 63% and 40% of total downstream traffic respectively (note that proxies are excluded in our analysis). This result suggests that the amount of traffic originating from servers to clients due to user downloading behaviors is highly skewed; therefore, caching or replication of those servers (or the content objects they serve) can significantly reduce the broadband traffic of the network carrier.

Lastly, after excluding the top 12 servers, the outbound traffic volumes contributed by the remaining servers approximately follows the Zipf distribution. This is most likely resulted by the fact that the pattern that users request for content objects follows the Zipf distribution.

D. Network Perspective

We now report the traffic statistics from the network perspective.

1) Downstream to upstream traffic ratio: We summarize in Fig. 6 and Fig. 7 the ratio of downstream to upstream traffic in terms of both the number of packets and bytes. We observe that some users are significantly more active and download large files (probably videos and applications), as shown in Fig. 6. Packet ratio, however, does not show that much of diversity, as shown in Fig. 7. Almost 80% of users send the same number of packets they receive.

These results suggest that the average size of downstream packets is much larger than upstream. Despite differences in total traffic exchanged, the ratio in net2g+ and net3g are very similar. Due to a small percentage of highly active users in net3g, the ratio for this network is slightly higher compared to that of net2g+.

2) TCP Throughput: We next report the statistics of TCP throughput for both networks, summarized in Fig. 8. For each TCP session, its TCP throughput is measured by dividing the number of downstream bytes by the length of the TCP session. The length of a session is the time difference between the first packet (*i.e.*, SYN) and the last packet (*i.e.*, FIN or RST) in the given TCP session.

We make two observations. Firstly, the average throughput for net2g+ is lower than net3g. This is expected as users in net3g not only use more advanced phone but also they enjoy a wireless link with lower RTT. Secondly, the average TCP throughput is very low. We note that it is well known that TCP performance on wireless links with high packet loss rate is bad;

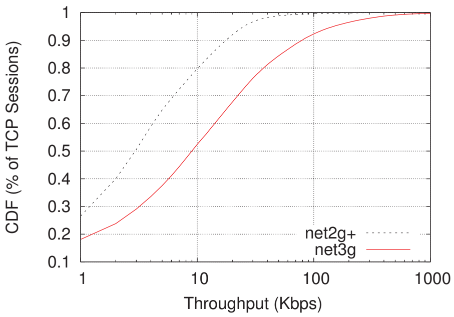

Fig. 8. TCP session throughput

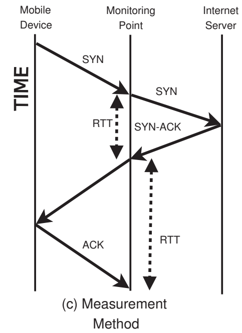

Fig. 9. RTT measurement method

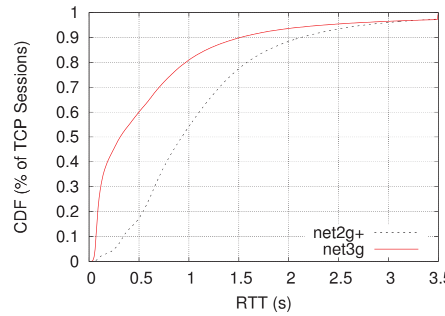

Fig. 10. Round Trip Time Statistics

however, this does not justify the extremely low throughput reported in Fig. 8.

In order to understand the causes for the low throughput, we carefully analyze the TCP sessions and find two main reasons. Firstly, a majority of sessions carry very small amount of data. Later in Section IV we shall see that over 50% of HTTP responses are smaller than 1KB. Small amount of data along with high value of RTT typically result in low throughput. Secondly, a problem especially for persistent HTTP connections is long idle time between the requests. In summary, the reported throughput does not show the actual capacity of the link but instead it shows the effective throughput over time.

3) Round Trip Time: We also measure RTT for TCP packets in both networks as shown in Fig. 9. Usually RTT is measured as the difference between the SYN and SYN-ACK packet pairs. However, the time difference between the SYN and SYN-ACK packet pairs in this case is RTT between the provider network and Internet[4], which excludes the latency of wireless access links. Therefore we also measure the difference of arrival times between SYN-ACK and ACK packet. By doing so, the wireless access link latency is included in the measured RTT.

Fig. 10 reports the measured RTT for TCP packets in both networks. We make the following observations. Firstly, RTT in net2g+ is significantly higher than RTT in net3g (a majority of users in net2g+ experience approximately 100% higher RTT than those in net3g), as the latter one use the more advanced communication technology. Secondly, RTT is highly variable. Large RTT values can be resulted by link layer retransmissions, network congestion and packet buffering in the routers. To measure the effect of buffering, we run a controlled experiment and observe extremely high delays when we flood the provider with continuous ping packets. Some packets arrived at the destination after a time period as long as 120 seconds.

IV. Web Traffic in Mobile Networks

As HTTP is the most prevalent traffic in both mobile networks net2g+ and net3g, we next examine the Web traffic more closely.

[4]Recall that we capture the traffic on the Gn interface, thus the time difference between a SYN and SYN-ACK packet pair is the time in which the SYN packet travels to the server and the server responds with a corresponding SYN-ACK packet.

A. Web Content Mix

We categorize the HTTP responses (*i.e.*, content returned by servers) into 8 different types based on the Content-Type header in HTTP response packets. All HTML, Javascript, XML and plain text contents are labeled as `text` and all WAP related contents (*e.g.*, WML, WMLC, WAP-mms-message) are labeled as `WAP`. For each content type and each network, we report the percentage of requests and the percentage of web traffic volume, as shown in Fig. 11 and Fig. 12.

We make the following observations. Firstly, 43% of HTTP responses are `WAP`, suggesting that a majority of users in net2g+ use WAP to for Internet access. In addition, `image`, `text` and `form` contents contribute to 45% of all HTTP requests in net2g+ whereas a very few requests are generated for `media` (*i.e.*, multimedia) and `flash` contents. This is reasonable as devices in net2g+ are mostly old low-end smart phones which most likely are not able to deal with such content. In comparison, 60% of responses in net3g are `text` and `image` contents, which can be treated as the regular web browsing behavior for users in net3g. The percentage of `media` responses in net3g is slightly higher than that in net2g+.

Secondly, more than 60% of web traffic incurred by HTTP responses (*i.e.*, downstream HTTP traffic) in net2g+ is generated by `WAP` and `app` (*i.e.*, application) objects. However, in net3g the `media` and `app` contents are responsible for 45% and 25% of all web traffic.

Thirdly, `image` and `text` contents contribute to only 20% of web traffic while they make more than 60% of requests in net3g, suggesting that a majority of web objects are very small.

B. Content Length and Type Distribution

We plot the distribution of content length and content type in Fig. 14.

We observe that in both networks, 50% of requests are smaller than 1KB and 90% of requests are smaller than 10KB. In particular, most of the `text`, `form` and `image` contents are very small (*e.g.*, less than 1KB) in net3g. Surprisingly, 60% of `media` requests are also reported as small contents. We investigate this more closely and find that these requests are audio/midi requests which contain attributes of music data rather than music data itself. Most of these objects are downloaded from some Nokia music store.

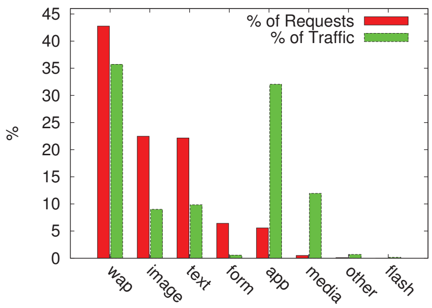

Fig. 11. Application mix in net2g+

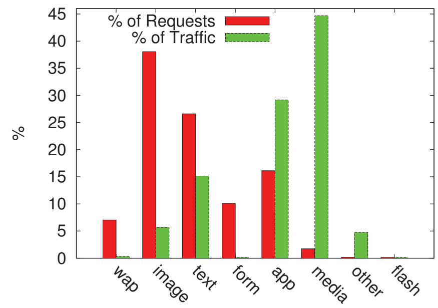

Fig. 12. Application mix in net3g

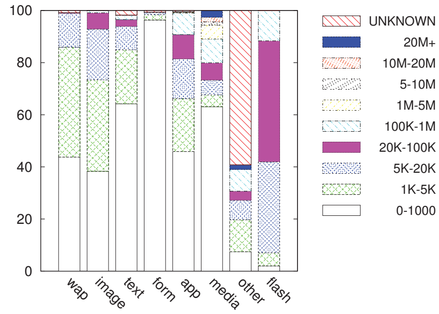

Fig. 13. Content length distribution

The collection of these results, if putting altogether, have multiple implications.

Firstly, these results suggest that for mobile web browsing, link delay is a more important factor compared to link throughput. To achieve a good quality of service, 3G network providers should favor providing links with minimum latency over links with maximum throughput.

Secondly, these results also suggest that 50% of contents can be delivered in a single packet. If without fine tuning TCP for mobile networks, content delivery using TCP as the transport protocol may suffer from unnecessary delays of 3-way handshakes or small congestion windows when accessing these small objects. A more light-weight protocol (*e.g.*, DCCP [7]) might be a better choice in this scenario (note that reliability could be provided by the application layer).

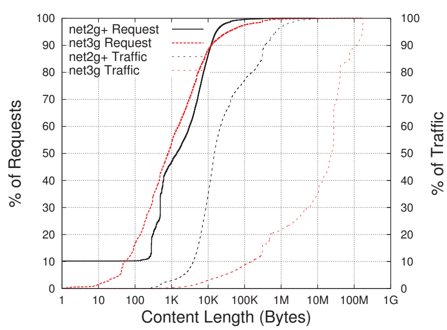

Fig. 14. Content-Type vs. Content-Length

We also observe that 90% of web requests in net2g+ contribute to less than 30% of web traffic, while in net3g this number is less than 5%, mainly because users in net3g tend to download large files. In net2g+ where users mainly use the WAP protocol, almost 80% of web traffic are generated by content objects less that 100KB.

C. Request/Response Distribution

We also investigate the distribution of HTTP requests and responses. We summarize the results in Fig. 15–17.

1) Number of requests generated by users: We plot in Fig. 15 the number of requests generated by users in both networks in a log-log scale (users are ranked in the descending order of the number generated requests). We observe that the top 10 users have completely different trends of generating web requests in net2g+ and net3g. We manually analyze the trace and find that a majority of these web requests in net2g+ are generated by a (maybe buggy or carelessly implemented) software which checks the up-to-date version by sending web requests to the `www.blovestorm.com` website once every 100ms. However, we do not find any abnormal activity for those users in net3g. One possibility is that they are laptops equipped with 3G network cards.

We also observe that for the remaining users, the number of requests generated by clients is highly skewed and follows the power law. We measure the cumulative number of requests for net2g+ and find that 75% of web requests are generated by only 20% of users. In addition, the skewness of request distribution in net3g is higher than that in net2g+. Based on our measurements, 20% of users in net3g are responsible for more than 80% of the total web requests. Surprisingly, we observe that in both networks, over 45% of users generate less than 5 web requests.

2) Number of requests for servers: We plot in Fig. 16 the number of requests for each server (*i.e.*, website). We extract the host headers from HTTP requests and compute the percentage of requests for the same host. Earlier studies (*e.g.*, [8]) have shown that in Internet, the number of requests for web pages follows the Zipf distribution. Fig. 16 shows that this in fact is valid for mobile traffic as well.

However, we observe that there are some unique patterns in this figure. Firstly, the most popular websites visited by net2g+ users include the proxy server, the `www.blovestorm.com` website, and a few web servers in the domain of `qq.com` (a popular portal in China). In net2g+ and net3g, 90% of all requests are generated for only 7% and 25% of visited websites, respectively. A further investigation reveals that over 50% of all requests in net2g+ and 24% of requests in net3g belong to different sub-domains of `qq.com`, suggesting that the significant popularity of this domain and its sub-domains in China.

Secondly, the differences between net2g+ and net3g also contribute to these unique patterns. Users in net3g have more powerful devices and higher 3G bandwidth, therefore they tend to browse many different websites. On the contrary, users in net2g+ have to use some proxy servers and due to limited bandwidth, their web browsing is limited to fewer popular web sites.

Thirdly, the total number of visited websites in net2g+ is larger than net3g. This is because the number of Internet-browsing users in net2g+is almost an order of magnitude more than that in net3g; due to the significantly larger number of

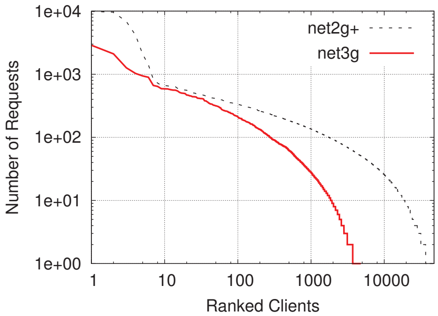

Fig. 15. Requests sent by ranked clients

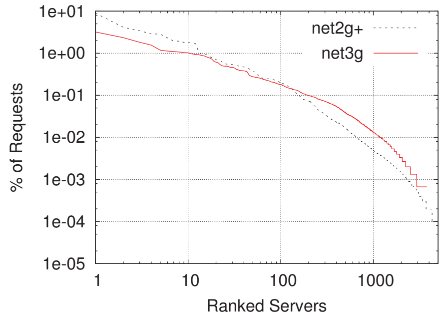

Fig. 16. Requests received by ranked servers

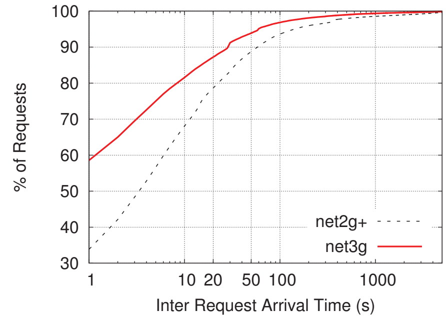

Fig. 17. Web requests' inter-arrival time

active users in net2g+ diversity of visited websites in net2g+ is higher.

3) Inter-arrival patterns: We plot in Fig. 17 the average inter-request arrival time for HTTP requests. We observe that more than 80% of requests in net3g and only 70% of requests in net2g+, respectively, are generated within 10 seconds from previous requests.

Requests generated within a short period of time are likely correlated, as browsers typically send multiple subsequent requests to retrieve the embedded contents (*e.g.*, images) in a Web page. Requests are generated faster if browsers render pages faster. Browsers of mobile devices in net2g+ are likely to have higher rendering time due to less powerful devices or dealing with WAP, which may explain the difference in inter-request arrival times in the two networks.

V. Discussions

In this section, we discuss the network and radio power management issues based on our finds.

A. Implication on Network Management

Providing a detailed statistical characteristics of mobile data traffic, our findings can help mobile network administrators to manage their networks more efficiently; for instance, our findings reveal the significance of national/regional portal sites and most HTTP responses being small, thus replication of contents or portal servers and lower the latency inside the mobile networks would significantly improve the user experiences quality and reduce the potential congestion in the mobile networks.

B. Implication on Radio Power Management

One of the major power consuming components in mobile devices is radio [9]. Several studies report that mobile devices maintain 3 different power levels for radio. A radio can be in the Idle, the low power or the Active mode. A radio in the Idle mode consumes no power. An Active radio use the highest power level and consume the maximum energy. A radio in the low power mode consumes smaller amount of energy compared to the Active mode. During transferring or receiving data, a radio is in the Active mode. When a radio has no packet to send or receive for a pre-determined amount of time (which depends on both carrier and mobile device manufacturer), it goes to the low power mode, and if it remains idle, it eventually goes to the Idle mode. An Idle radio takes about a few seconds (approximately 1.5 seconds) to switch to the Active mode. To avoid this wake-up delay, usually a radio is kept in the low power mode for a long time (between 12 and 15 seconds). Switching from the low power mode to the Active mode is much faster than switching from the Idle mode to the Active mode. However, keeping radio in the low power mode consumes and wastes the energy if no packet is being sent after this period.

Recently, finding the best values of power management timers has been the topic of several studies (see, *e.g.*, [9], [10]). There is a clear trade off in choosing the timer that control the low power mode. If we select a small timer, a radio may unnecessarily go to the Idle mode which results in long wake-up delay for the next transmission. On the other hand, selecting a large value can waste the energy when timer expires and no transmission occurs.

Some studies looked at the inter-packet arrival time and simply suggest to lower the long-tail period for saving the energy. According to [4], 95% of the packets are received or transmitted within 4.5 seconds of the previous packet. Based on this result they suggested that a 4.5-second long tail before going to sleep can cover 95% of the packets; longer tails will have diminishing returns with respect to covering more packets while wasting more energy.

Fig. 17 however shows that the inter-arrival time of HTTP requests has different pattern. Only 50% of web requests in net2g+ and 70% of requests in net3g are generated within 5 second after the last request. Simply reducing the long-tail period may reduce the energy but it will increase the delay for the next web request. This additional delay will happen for 50% of web requests in net2g+ and over 30% of web requests in net3g. We believe that using a single fixed long tail period is not appropriate in general and that it should be adjusted dynamically based on individual users' activity profiles.

Ongoing work on this topic includes detailed analysis of adjusting long-tail period and its impacts on energy saving. We also plan to study diurnal characteristic of user's activity by looking at longer traces.

VI. Related Work

We divide the related work into the following categories based on the experimental methodologies. The first category is active measurements using synthetic workloads (see, *e.g.*, [11], [12], [13], [14]). The drawback of this methodology is that it hardly provides a realistic view of what the network and users experience in reality.

The second category is active measurements using realistic workloads (see, *e.g.*, [15], [4], [16], [17], [18]). Specifically, Falaki *et al.* characterized diversity in smart phone activities of 255 mobile users [15]. They further looked at the generated network traffic and found that web browsing contributes over half of the traffic [4]. Mao *et al.* conduct a series of comprehensive studies which shed valuable light on different aspects of smart phone usage, traffic and power consumption [16], [17], [18].

The third category is offline analysis of captured traffic from the infrastructure (see, *e.g.*, [3], [6], [19], [20], [21]). Typical analyses include traffic characteristics such as application/protocol breakdown, application or end-to-end performance (delay, packet losses, jitter, response time), TCP retransmissions, packet-arrival and usage patterns, and user behaviors. In particular, in [3], Trestian *et al.* analyzed temporal dynamics of user mobility and network traffic generated by web applications in a 3G network within a metropolitan area. Shafiq *et al.* proposed a Zipf-like model and a Markov model to capture the distribution and dynamics of traffic volumes, respectively, based upon a week-long flow-level traffic data collected from a major cellular operator's core network [20].

Our study falls in the third category and is complementary to existing studies. We collect the data set from a non-US tier-1 mobile carrier. However, the data sets used in most of studies in the literature are captured from US-based carriers. Our data set also contains traces for networks based on two different technologies (3G and GPRS), which allow us to conduct comparative studies on these two networks.

VII. Conclusion

In this paper we take a carrier's perspective to analyze the mobile smart phone traffic based on packet-level traces collected from a tier-1 national cellular network in a highly populated region in China. Our results, largely consistent with existing studies from end users' perspectives, show that web browsing contributes more than half of the total traffic. Our detailed analysis, however, reveals some unique user browsing behaviors that have not been shown in other studies.

There are multiple avenues for future works. First, 3G/4G data cards are increasingly popular; therefore, we are interested in ways differentiating devices with data cards and normal smart phones as well as characterizing traffic generated by such devices. Second, we are interested in fine-tuning or customizing TCP to improve the effectiveness of content delivery in cellular networks. Last but not the least, we plan to study the quality of cellular user experiences using the collected traces.

Acknowledgement

We would like to thank the anonymous reviewers for their insightful and constructive comments. Haiyong Xie is the corresponding author and is supported in part by the National Natural Science Foundation of China under Grant No. 61073192, by the Grand Fundamental Research Program of China (973 Program) under Grant No. 2011CB302905, by the New Century Excellent Talents Program, and by the Fundamental Research Funds for Central Universities under Grant No. WK0110000014.

References

[1] Cisco, "Cisco visual networking index: Forecast and methodology, 2009–2014," White Paper, 2010.

[2] ——, "Hyperconnectivity and the approaching zettabyte era," White Paper, 2010.

[3] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: connecting people, locations and interests in a mobile 3g network," in *Proc. IMC '09*, Chicago, Illinois, USA, 2009, pp. 267–279.

[4] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A first look at traffic on smartphones," in *Proc. IMC '10*, Melbourne, Australia, 2010, pp. 281–287.

[5] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband internet traffic," in *Proc. IMC '09*, Chicago, Illinois, USA, 2009, pp. 90–102.

[6] G. Maier, F. Schneider, and A. Feldmann, "A first look at mobile hand-held device traffic," in *Proc. PAM '10*, Zurich, Switzerland, 2010, pp. 161–170.

[7] E. Kohler, M. Handley, and S. Floyd, "Designing dccp: congestion control without reliability," in *Proc. ACM SIGCOMM '06*, Pisa, Italy, 2006, pp. 27–38.

[8] L. Breslau, P. Cue, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," in *Proc. INFOCOM*, 1999, pp. 126–134.

[9] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications," in *Proc. IMC '09*, Chicago, Illinois, USA, 2009, pp. 280–293.

[10] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Top: Tail optimization protocol for cellular radio resource allocation," in *Proc. ICNP '10*, oct. 2010, pp. 285 –294.

[11] R. Chakravorty, S. Banerjee, P. Rodriguez, J. Chesterfield, and I. Pratt, "Performance optimizations for wireless wide-area networks: comparative study and experimental evaluation," in *Proc. ACM MobiCom '04*, Philadelphia, PA, USA, 2004, pp. 159–173.

[12] J. Chesterfield, R. Chakravorty, J. Crowcroft, P. Rodriguez, and S. Banerjee, "Experiences with multimedia streaming over 2.5g and 3g networks," in *Proc. BroadWiM '04*, San Jose, CA, Oct. 2004.

[13] J. Huang, Q. Xu, B. Tiwana, Z. M. Mao, M. Zhang, and P. Bahl, "Anatomizing application performance differences on smartphones," in *Proc. MobiSys '10*, San Francisco, CA, 2010, pp. 165–178.

[14] K. Mattar, A. Sridharan, H. Zang, I. Matta, and A. Bestavros, "Tcp over cdma2000 networks: a cross-layer measurement study," in *Proc. PAM '07*, Louvain-la-Neuve, Belgium, 2007, pp. 94–104.

[15] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *Proc. MobiSys '10*, San Francisco, CA, USA, 2010.

[16] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Profiling resource usage for mobile applications: a cross-layer approach," in *Proc. MobiSys '11*, Bethesda, Maryland, USA, 2011, pp. 321–334.

[17] Q. Xu, J. Huang, Z. Wang, F. Qian, A. Gerber, and Z. M. Mao, "Cellular data network infrastructure characterization and implication on mobile content placement," in *Proc. SIGMETRICS '11*, San Jose, California, USA, 2011, pp. 317–328.

[18] W. Zhaoguang, Q. Zhiyun, X. Qiang, M. Zhuoqing Morley, and Z. Ming, "An untold story of middleboxes in cellular networks," in *Proc. ACM SIGCOMM '11*, Toronto, ON, CA, 2011.

[19] P. Svoboda, F. Ricciato, E. Hasenleithner, and R. Pilz, "Composition of GPRS/UMTS traffic: Snapshots from a live network," in *Proc. IPS MoMe '06*, Salzburg, Austria, Feb 2006.

[20] M. Z. Shafiq, L. Ji, A. X. Liu, and J. Wang, "Characterizing and modeling internet traffic dynamics of cellular devices," in *Proc. ACM SIGMETRICS '11*, San Jose, California, USA, 2011, pp. 265–276.

[21] R. Keralapura, A. Nucci, Z.-L. Zhang, and L. Gao, "Profiling users in a 3g network using hourglass co-clustering," in *Proc. MobiCom '10*, Chicago, Illinois, USA, 2010, pp. 341–352.