Mobile Network Performance from User Devices: A Longitudinal, Multidimensional Analysis

Ashkan Nikravesh[1], David R. Choffnes[2], Ethan Katz-Bassett[3], Z. Morley Mao[1], and Matt Welsh[4]

[1] University of Michigan   [2] Northeastern University
[3] University of Southern California   [4] Google Inc.

Abstract. In the cellular environment, operators, researchers and end users have poor visibility into network performance for devices. Improving visibility is challenging because this performance depends factors that include carrier, access technology, signal strength, geographic location and time. Addressing this requires longitudinal, continuous and large-scale measurements from a diverse set of mobile devices and networks.

This paper takes a first look at cellular network performance from this perspective, using 17 months of data collected from devices located throughout the world. We show that (i) there is significant variance in key performance metrics both within and across carriers; (ii) this variance is at best only partially explained by regional and time-of-day patterns; (iii) the stability of network performance varies substantially among carriers. Further, we use the dataset to diagnose the causes behind observed performance problems and identify additional measurements that will improve our ability to reason about mobile network behavior.

1   Introduction

Cellular networks are the fastest growing, most popular and least understood Internet systems. A particularly difficult challenge in this environment is capturing a view of network performance that is representative of conditions at end user devices. A number of factors frustrate our ability to capture this view. For instance, carriers enforce different policies depending on the traffic types or geographic/social characteristics of different locations such as population [1, 2], causing user perceived performance to differ from advertised performance for access technologies. Other environmental factors have a significant impact on performance, including device model [3], mobility [4], network load [2], packet size [5, 6] and MAC-layer scheduling [4].

To account for various factors impacting Internet performance in mobile networks, we need pervasive network monitoring that samples a variety of devices across carriers, access technologies, locations and over time. This work takes a first look at such a view using data collected from controlled measurement experiments in 144 carriers during 17 months, comprising 11 cellular network technologies. We use this data to identify the patterns, trends, anomalies, and evolution of cellular networks' performance.

This study demonstrates that characterizing and understanding the performance in today's cellular networks is far from trivial. We find that all carriers exhibit significant variance in end-to-end performance in terms of latency and throughput. To explain this

variance, we investigate geographic and temporal properties of network performance. While we find that these properties account for some differences in performance, importantly we observe that performance is inherently unstable, with some carriers providing relatively more or less predictable performance. Last, we identify alternative sources of variance in performance that include routing and signal strength. An important open question is how to design a measurement platform that allows us to understand reasons behind most observed performance differences.

This paper differs from previous related work in that our study is longitudinal, continuous, pervasive and gathered from mobile devices using controlled experiments. In contrast, some related work [7–9] passively collected network traffic from cellular network infrastructure, using one month of data or less. These studies tend to be limited to a single carrier, hampering our ability to conduct meaningful comparisons across carriers. Other work collected network performance data at mobile devices [10, 1, 11], but did not use controlled experiments to capture a continuous view of performance.

Roadmap. We describe our methodology and dataset in §2, then present our findings regarding network performance across different network technologies, carriers, locations, and times in §3.1, §3.2, and §3.3 respectively. Then we study the root causes for performance degradation in §3.4. We discuss related work in §4 and conclude in §5.

2 Methodology and Dataset

This paper studies cellular network performance using a broad longitudinal view of network behavior impacting user-perceived performance. To this end, we consider HTTP GET throughput, round trip time latency from ping, and DNS lookup time as end-to-end performance metrics. In addition to gathering raw performance data, we annotate our measurements with path information gathered from traceroute, the identify of the device's carrier, its cellular network technology, signal strength, location and timestamp.

We focus on performance from mobile devices to Google, a large, popular content provider. We argue that Google is an ideal target for network measurements because it is highly available and well provisioned, making it easier to isolate network performance to cell networks vs. Google's network. Focusing on these measurements, we identify the performance impact of carrier, network technology, location and time. To reason about the root cause behind performance changes, we use path information, DNS mappings and signal strength readings.

Our data is collected by two Android apps using a nearly identical codebase, Speedometer and Mobiperf.[1] Speedometer is an internal Android app developed by Google and deployed on hundreds of volunteer devices, mainly owned by Google employees. As such, the bulk of our dataset[2] is biased toward locations where Google employees live and work. Speedometer collected the following measurements from 2011-10 to 2013-2 (17 months): 6.6M ping RTTs to `www.google.com` (each sample consists of 10 consecutive probes), 1.7M HTTP GETs to measure TCP throughput using a 224KB file hosted on a Google server, 0.4M UDP burst samples for measuring packet loss rate, 0.8M DNS resolutions of `google.com`, and 0.8M traceroute (without hop RTTs) from

---

[1] http://www.mobiperf.com/

[2] This dataset is publicly available at https://storage.cloud.google.com/speedometer

Table 1: Number of Measurement and Carriers for the Network Technologies

	HSPA	HSDPA	UMTS	EDGE	GPRS	LTE	EVDO	eHRPD	1xRTT
# of Measurements	439K	2326K	563K	506K	58K	1460K	2183K	301K	68K
# of Carriers	50	111	96	85	48	7	8	2	3

144 carriers and 9 network technologies. The dataset includes $\approx$ 4-5 measurements per minute. Each measurement is annotated with device model, coarse-grained location information (k-anonymized latitude and longitude), timestamp, carrier, and network type.[3] All users consented to participate in the measurement study; the anonymization process is explained in the dataset's README file. Because of anonymization, the number of users who participated in data collection is unknown.

We augment the Speedometer dataset with 11 months of data collected by Mobiperf. Mobiperf conducts a superset of measurements in Speedometer, and notably adds signal strength information. The number of measurements collected by Mobiperf for each task ranges from 17K (HTTP GET) to 58K (ping RTT test) from 71 carriers. We use Mobiperf data to study the impact of signal strength on measurement results. Table 1 shows the number of measurements collected from the most frequently seen 9 network technologies (ordered by peak speed) for both GSM and CDMA technologies in the combined datasets.

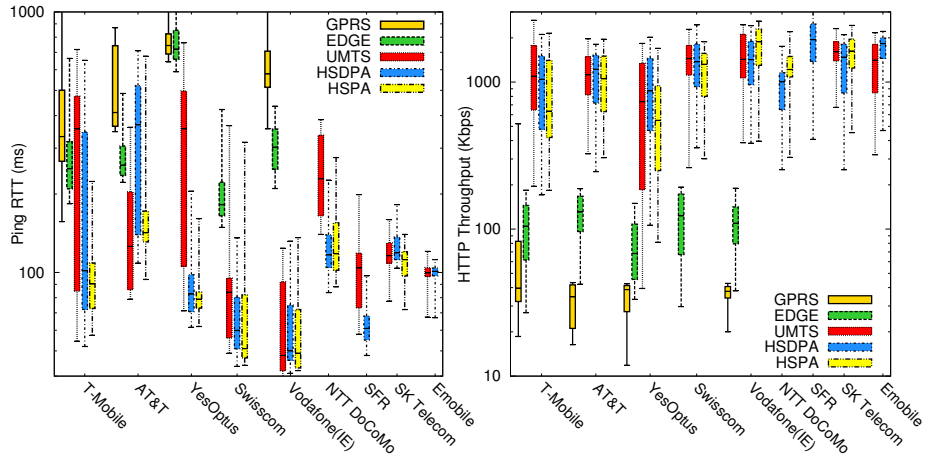## 3 Data Analysis

3.1 Performance across carriers

This section investigates the performance of five access technologies for each of several carriers. Our goal is to understand how observed performance matches with expectations across access technologies, and how variable this performance is across carriers. In Fig. 1, we plot percentile distributions (P5, P25, median, P75, and P95) of the latency and throughput of 9 carriers from Asia, America, Europe, and Australia. We select these carriers based on their geographic locations and relatively large data sample sizes. One of the key observations is that performance varies significantly across carriers and access technologies; further, the range of values is also relatively large.

For carriers that have high latency, we use traceroute data to investigate if the cause is inefficient routes to Google [12]. However, approximately half of the carriers such as SFR (French Carrier) and Swisscom have direct peering points with Google, making this unlikely to be the cause for high latency.

For carriers such as AT&T, T-Mobile US, and Airtel (India), we observe high variability in latency. In the following subsections, we investigate whether this is explained by regional differences, time-of-day effects and/or other factors.

Surprisingly, we do not observe significant latency differences across access technologies for some carriers. For example, the latency of UMTS, HSDPA, and HSPA in Emobile (Ireland), SK Telecom (Korea), and Swisscom are almost equal. Users in these networks may not see noticeable differences in performance for delay-sensitive applications when upgrading to newer technologies.

---

[3] https://github.com/Mobiperf/Speedometer

(a) Ping RTT

(b) HTTP GET throughput for downloading a 224KB file

Fig. 1: Throughput and latency across access technology and carriers.

In Fig. 1b, we plot HTTP throughput for downloading a 224KB file from a Google domain. Compared to ping RTT, the difference between the throughput of carriers is relatively smaller, indicating that the high variability in ping RTTs is often amortized over the duration of a transfer.

Note that the throughput for UMTS, HSDPA, and HSPA are almost identical. This occurs because the flow size is not sufficiently large to saturate the link for high-capacity technologies. This indicates a need for better low-cost techniques to estimate available capacity in such networks [13]. However, the figure shows significant performance difference between GPRS/EDGE and other access technologies.

We observe that lower latency is generally correlated with higher HTTP GET throughput, but this depends on the carrier. We quantify this using the correlation coefficient between HTTP throughput and ping RTT for specific carrier and network type. The strongest correlation coefficient observed was for Verizon LTE users with -0.53 and lowest was -0.01 for T-Mobile HSDPA users, using one-hour buckets.

Having observed significant differences in performance within and between carriers, we now investigate some of the potential factors behind this variability.

3.2 Performance across different locations

We now investigate the impact of geography on network performance. We focus on four major US carriers in three US regions where our dataset is densest (New York, Seattle, and Bay Area). Each of these carriers exhibits different topologies (Internet egress, Google ingress and ASes between) in different regions, potentially leading to performance differences in each region.

Despite the variety in network topologies, we surprisingly find that for AT&T, T-Mobile, and Sprint, both of the latency and throughput were similar in these three locations. However, for Verizon, we observe different LTE performance in New York,
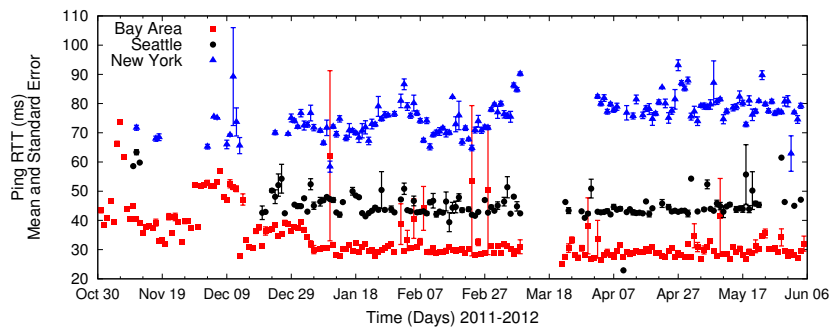
Fig. 2: Verizon LTE Ping RTT in Different Locations

Seattle, and Bay Area. Fig. 2 plots these latencies over time, and clearly show that the RTT latency for the Bay Area is lower than New York and Seattle areas. HTTP throughput in these regions exhibit similar patterns.

We use DNS data in the Seattle area and observe that 97% of DNS requests for google.com resolve to an IP for a server in the Los Angeles area instead of Seattle, in part explaining the gap in latency between the two regions. For the NY area, our measurements did not provide enough geographic information to understand whether increased latency was due to path inefficiencies.

The key takeaway from this section is that geography alone doesn't explain the variance in performance observed in the previous section; however, for one carrier (Verizon), it explains some of it. Further, we observe that each region experiences changes in performance independently – the correlation of performance across regions for each carrier is negligibly small. Last, when correlating ping RTT and HTTP GET throughput within each region, we find higher correlations than carrier-wide correlations presented in the previous section. This further suggests that performance is affected by location.

3.3 Performance over time

We now analyze how performance depends on time – both in terms of time-of-day effects and the stability of measurement performance over time. These properties allow us to identify when to measure the network (*e.g.,* during known busy hours) and when *not* to measure (*e.g.,* at ten minute intervals), thus allowing us to efficiently allocate the limited measurement resources that users provide.

Time-of-day and long-term trends. Fig. 3 plots HTTP throughput for four major carriers in the US. As expected, throughput decreases (and variance tends to increase) during the busy hours for mobile usage (8AM to 7PM), likely due to higher load on the network. Interestingly, different carriers experience minimum throughput at different times. T-Mobile and AT&T reach their minimum throughput at 1PM and 5PM, respectively; Sprint experiences minimum performance at 9PM and Verizon, two troughs occur at 8AM and 9PM. Last, these carriers experience different relative variations in performance during busy hours: AT&T and Sprint throughput drops by approximately a third during busy hours while Verizon drops by 25%, and T-Mobile by 16%.
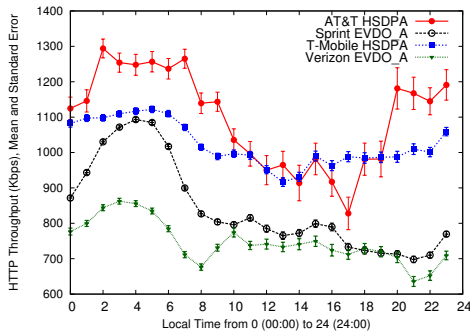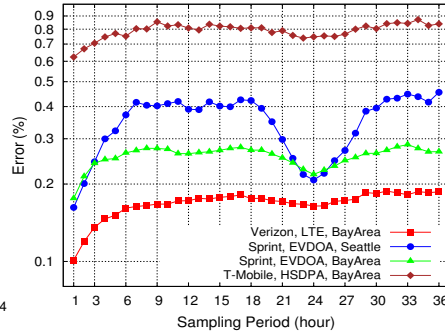
Fig. 3: Time of day pattern of HTTP throughput

Fig. 4: Weighted Moving Average Error (Median Ping RTT, $W = 2$)

Next, we investigate the long-term performance trends over the duration of our study, allowing us to tell if new cellular technologies and infrastructure are keeping pace with increased mobile Internet usage. Specifically, we look at the change in throughput and latency of carriers through time over consecutive days for each network technology they support in different areas. We did not observe improvement; despite technology upgrades, performance is highly variable over time and there is no statistically significant change during the observation period.

Stability of performance. The predictability and stability of network performance are important not only for users, who are often frustrated more by variations in performance than the average value, but also for determining how and when to conduct measurements for future experiments. In this section, we compute stability using a weighted moving average and autocorrelation.

First, we group the data into 1-hour buckets (to obtain a sufficiently large sample size). Then for each bucket, we use either the median or 5th percentile latency. We compute the moving average error for different window sizes and sampling periods.

We compute the moving average error as follows: for a window size W, we predict the next data point on that series by computing moving average for the previous consecutive W points. For each W and sampling period (*e.g.,* every N hours for $N = 1, 2, 3, \ldots$), we compute the average over different offsets.

Fig. 4 plots the average error for all data points with windows size of 2 and different sampling periods for median ping RTT (results with larger window sizes of 3, 4, and 5 are similar). We observe that prediction accuracy varies significantly by carrier, with Verizon and Sprint in the Bay Area being relatively predictable, and T-Mobile and Sprint in Seattle being relatively unpredictable. Also, for all of these carriers, prediction accuracy is best when looking at the most recent data (one hour sampling period) and error tends to increase with longer durations, with the exception of 24hr (day) and 168hrs (week) sampling periods, which are local minima. The results from autocorrelation are similar.

These predictability results indicate that despite the large overall variance in cellular network performance, there are regions and time scales over which performance is relatively predictable, depending on the carrier. Importantly, we can use this information to
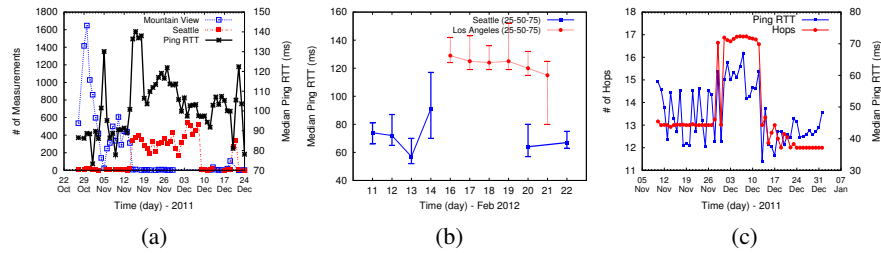
Fig. 5: Performance Degradation in: (a)T-Mobile HSDPA network in Bay Area due to server selection flapping from Bay Area to Seattle (b)T-Mobile HSDPA network in Seattle due to change in ingress point of transit AS between T-Mobile and Google (c) Verizon LTE network in Bay Area

inform the design of measurement system that uses prediction to minimize probes that would provide redundant results. For instance, if we subsample every other value (*i.e.,* 50% sampling rate) in the Verizon LTE ping data in the Bay Area (which has the lowest error in the full sample), the distribution of latencies is nearly identical.

3.4 Performance Degradation: Root Causes

We now use our measurements to identify the reasons for persistent performance degradation observed in consecutive days. We focus on cases where the issue affects both ping RTT and HTTP throughput.

Inefficient paths. A reason for performance degradation is inefficient paths. Zarifis et al [14] provide a detailed taxonomy and analysis of path inflation in mobile networks; here we focus on their time evolution and constrain our analysis to only those cases where both latency and throughput were impacted.

For example, we observe an increase in ping RTT in T-Mobile's Bay Area HSDPA network from Nov 12, 2011 to Dec 10, 2011. Using DNS lookups, we find that clients previously sent to Mountain View were being sent to Seattle, with the additional delay explained by path inflation (Fig. 5a). After Dec 10, clients are again directed toward Mountain View.

We also observed a high-latency event for T-Mobile's Seattle HSDPA network in Seattle (Fig. 5b). Prior to the event, traceroutes indicate that traffic from T-Mobile ingresses into Level 3 in Seattle, then enters Google's network. After Feb 15, traffic from these subscribers ingressed into Level 3 at a peering point in Los Angeles before entering Google's network. After Feb 20, routing returns back to its previous state (ingress and egress point in Seattle area) and the median RTT decreases to its previous value, strongly implying that the change in performance was due to the topology change.

In Fig. 5c, we observe that ping RTT and the number of traceroute hops increases for Verizon LTE users in the Bay Area. Previously, clients were sent to a Google frontend in the Bay Area; after the change clients are sent to the same Google ingress point, but then traffic is sent to a frontend in Seattle (leading to $\approx 30\%$ higher latency).

In this section we show that fixed-line inefficiencies can significantly impact the performance of LTE and HSDPA networks. For these newer technologies, since the

(a) Ping RTT and packet loss   (b) HTTP throughput (file size is 100KB)
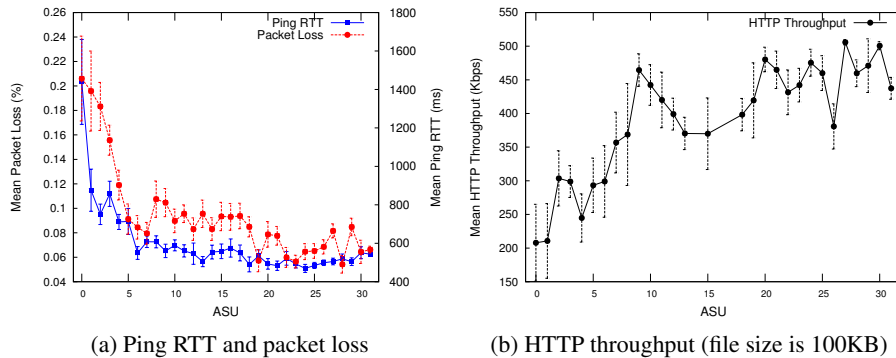
Fig. 6: Impact of signal strength on latency, packet loss, and throughput.

RTT is lower, the impact of inefficient routes is even relatively higher (around 80% increase in the RTT of T-Mobile HSDPA in Seattle).

Signal strength. It is well known that weak signal strength reduces channel efficiency for wireless communication; therefore, it is important to account for this when interpreting measurements. Using Mobiperf clients, we gather network measurements annotated with the signal strength, in Arbitrary Strength Units (ASUs),[4] reported during the probes and determine the impact of signal strength on performance.

Fig. 6 shows how three performance metrics vary with ASU values for AT&T HSDPA users in Seattle. The figures indicate high packet loss, latency and low throughput for ASU values between 0 and 8 (confirming the results in [15]); at larger ASU values that increase in signal strength has less impact on performance. These results indicate that accounting for signal strength is critically important for properly interpreting measurement results. For example, when measuring a carrier's capacity, it is important to do such tests in regions with high signal strength.

4   Related Work

Many previous studies attempt to improve our visibility into and understanding of mobile network performance. We can broadly characterize them according to *what* type of network performance they measured, *where* they conducted measurements and *how* they performed measurements. In this work, we are the first to use controlled, active measurement experiments to continuously monitor end-to-end network performance seen from mobile devices, across more than 100 carriers during a period of 17 months. Previous work differs as follows.

Passive measurements, infrastructure, single carrier. Several studies focus on passive measurements from inside mobile carriers [7–9]. While important for debugging the infrastructure components of latency, the view from such locations does not necessarily indicate the performance on mobile devices.

---

[4] Android shows zero signal bar for the ASU values between 0 and 2 and full signal bars when ASU value is more than 12.

Active measurements, end devices, single carrier. Several projects use active measurements from end devices, but focus on a single carrier for a limited duration of time, often doing a fine-grained and low-level analysis of performance. In [5], authors measured goodput, delay, and jitter of HSDPA and WCDMA networks from an operator in Finland using active measurements from a laptop. In [6], the authors compare LTE and HSPA networks by conducting high precision latency measurements for an operator in Austria. In [16, 4, 17, 18], authors studied the TCP performance in CDMA2000 networks. In[16], the authors investigate the steady-state TCP performance over CDMA 1x EV-DO downlink/uplink with the active measurement of long-lived TCP connections at the end-points for a Korean operator. [18, 4] conducted a cross-layer measurement of transport, physical and MAC layer parameters. [18] characterizes the wireless scheduler in a commercial CDMA2000 network and its impact on TCP performance by performing end-to-end experiments and sending UDP and TCP packets.

Active measurements, end devices, several carriers.: Similar to the previous examples, several studies also include comparisons across multiple carriers. In [19], by investigating the performance of three Norwegian operator and conducting active measurements from end-to-end devices, they studied the impact of the packet size on the minimal one-way delay for the uplink in 3G mobile networks. In [11], by performing active measurements for more than 6 months from 90 voting locations and by measuring the round trip delay of three network operator in Norway, they found the operator-specific network design and configurations as the most important factor for delays. In [2], by measuring data throughput, latency, and video and voice calls handling capacities, they compared the 3G performance of three carriers in Hong Kong under saturated conditions by conducting measurements at 170 sites in four months.

Active measurements, end devices, pervasive. Most closely related to our work is [1] and [3]. Both projects gather active measurements from apps running on mobile devices; however, they all rely on user-generated tests. In contrast our work uses controlled experiments to schedule measurements independent of user activity. This enables a more continuous view of performance in mobile networks.

5 Conclusion

This paper took a first look at end-to-end performance as seen from mobile devices, using a dataset of scheduled network measurements spanning more than 100 carriers over 17 months. We find that there are significant performance differences across carriers, access technologies, geographic regions and over time; however, we emphasize that these variations themselves are not uniform, making network performance difficult to diagnose. Using supplemental measurements such as DNS lookups and traceroutes, we identified the reasons behind persistent performance problems. Further, we examined the stability of network performance, which can help inform efficient scheduling of future network measurements. Overall, we find that performance in cell networks is not improving on average, suggesting the need for more monitoring and diagnosis. As part of our future work, we are investigating how to automatically detect persistent performance problems in real time, gather additional network measurements to explain them and provide this information to carriers and end users automatically.

References

1. Sommers, J., Barford, P.: Cell vs. WiFi: on the performance of metro area mobile connections. In: Proc. ACM SIGCOMM IMC. (2012)
2. Tan, W.L., Lam, F., Lau, W.C.: An Empirical Study on 3G Network Capacity and Performance. In: Proc. IEEE INFOCOM. (2007)
3. Huang, J., Xu, Q., Tiwana, B., Mao, Z.M., Zhang, M., Bahl, P.: Anatomizing application performance differences on smartphones. In: Proc. ACM MOBISYS. (2010)
4. Liu, X., Sridharan, A., Machiraju, S., Seshadri, M., Zang, H.: Experiences in a 3G network: interplay between the wireless channel and applications. In: Proc. ACM MOBICOM. (2008)
5. Jurvansuu, M., Prokkola, J., Hanski, M., Perala, P.: HSDPA Performance in Live Networks. In: IEEE ICC. (2007)
6. Laner, M., Svoboda, P., Romirer-Maierhofer, P., Nikaein, N., Ricciato, F., Rupp, M.: A comparison between one-way delays in operating HSPA and LTE networks. In: Proc. WINMEE. (2012)
7. Vacirca, F., Ricciato, F., Pilz, R.: Large-Scale RTT Measurements from an Operational UMTS/GPRS Network. In: WICON. (2005)
8. Laner, M., Svoboda, P., Hasenleithner, E., Rupp, M.: Dissecting 3G Uplink Delay by Measuring in an Operational HSPA Network. In: Proc. PAM. (2011)
9. Romirer-Maierhofer, P., Ricciato, F., D'Alconzo, A., Franzan, R., Karner, W.: Network-Wide Measurements of TCP RTT in 3G. In: TMA. (2009)
10. Deshpande, P., Hou, X., Das, S.R.: Performance Comparison of 3G and Metro-Scale WiFi for Vehicular Network Access. In: Proc. ACM SIGCOMM IMC. (2010)
11. Elmokashfi, A., Kvalbein, A., Xiang, J., Evensen, K.R.: Characterizing delays in norwegian 3g networks. In: Proc. PAM. (2012)
12. Zheng, H., Lua, E.K., Pias, M., Griffin, T.G.: Internet routing policies and round-trip-times. In: Proc. PAM. (2005)
13. Huang, J., Qian, F., Guo, Y., Zhou, Y., Xu, Q., Mao, Z.M., Sen, S., Spatscheck, O.: An in-depth study of lte: Effect of network protocol and application behavior on performance. In: Proc. ACM SIGCOMM. (2013)
14. Zarifis, K., Flach, T., Nori, S., Choffnes, D., Govindan, R., Katz-Bassett, E., Mao, Z.M., Welsh, M.: Diagnosing Path Inflation of Mobile Client Traffic. In: Proc. PAM. (2014)
15. Schulman, A., Navday, V., Ramjeey, R., Spring, N., Deshpandez, P., Grunewald, C., Padmanabhany, K.J.V.N.: Bartendr: A practical approach to energy-aware cellular data scheduling. In: Proc. ACM MOBICOM. (2010)
16. Lee, Y.: Measured TCP Performance in CDMA 1x EV-DO Network. In: Proc. PAM. (2006)
17. Claypool, M., Kinicki, R., Lee, W., Li, M., Ratner, G.: Characterization by Measurement of a CDMA 1x EVDO Network. In: Proc. WICON. (2006)
18. Mattar, K., Sridharan, A., Zang, H., Matta, I., Bestavros, A.: TCP over CDMA2000 networks: a cross-layer measurement study. In: Proc. PAM. (2007)
19. Arlos, P., Fiedler, M.: Influence of the Packet Size on the One-Way Delay in 3G Networks. In: Proc. PAM. (2010)