Characterizing delays in Norwegian 3G networks

Ahmed Elmokashfi, Amund Kvalbein, Jie Xiang, and Kristian R. Evensen

Simula Research Laboratory

Abstract. This paper presents a first look at long-term delay measurements from data connections in 3 Norwegian 3G Networks. We have performed active measurements for more than 6 months from 90 voting locations used in a trial with electronic voting during this fall's regional elections. Our monitors are geographically spread across all of Norway, and give an unprecedented view of the performance and stability of the total 3G infrastructure of a country. In this paper, we focus on delay characteristics. We find large differences in delay between different monitors. More interestingly, we observe that the delay characteristics of the different operators are very different, pointing to operator-specific network design and configurations as the most important factor for delays.

1 Introduction

We are witnessing a revolution in the way people access and use the Internet. The advent of mobile devices such as smartphones and tablets, combined with the almost universal coverage of 3G networks, has radically changed how we access, share and process information. A stable and resilient 3G network connection has become a necessity for the daily operations of individuals and organizations. Yet, we have little knowledge of the long-term stability and performance of 3G data networks, beyond the coverage maps provided by network operators. This gives a very limited basis for comparing and evaluating the quality of the offered services. To alleviate this, there is a need for long-term measurements of the stability, availability and quality experienced by users in each network.

This paper presents a first look at long-term measurements of mobile broadband (MBB) data connections from 3 different network operators in Norway, with an emphasis on delay characteristics. The measurements are carried out over a period of more than 6 months from 90 locations in 10 municipalities spread across Norway. The measurements are performed using *ping* and *traceroute* from our monitor nodes to servers placed at two different locations. These measurements were collected in connection with a trial of electronic voting during the Norwegian regional elections in fall 2011. Hence, all monitors are placed in voting locations. The number of voting locations in each municipality varies between 4 and 15. Voting locations are geographically spread according to habitation patterns in the participating municipalities, which vary in size and population density.

Our measurements have a unique combination of features:

- They are taken from a large number of geographically diverse measurement points, giving a representative view of the quality of MBB data connections experienced by customers across Norway.
- They are measured over a long period of over 6 months, giving a good basis for capturing both short-term and long-term variations in the experienced performance.
- They are performed simultaneously in 3 different cellular networks, giving a unique possibility to directly compare and correlate the performance of different networks.

In this paper, we present the measurement setup, and use the data to take a first look at an important performance metric: delay. More specifically, we focus on RTTs measured by *ping*. We characterize delay along several axis, and compare the delays experienced in different networks and at different locations. We find that there are large differences between operators with respect to both absolute delays and variations, and that each operator has its own "signature" in the delay characteristics. Interestingly, we also find that the delay characteristics are mainly network-dependent rather than monitor-dependent, indicating the key role played by network design decisions in deciding delay characteristics.

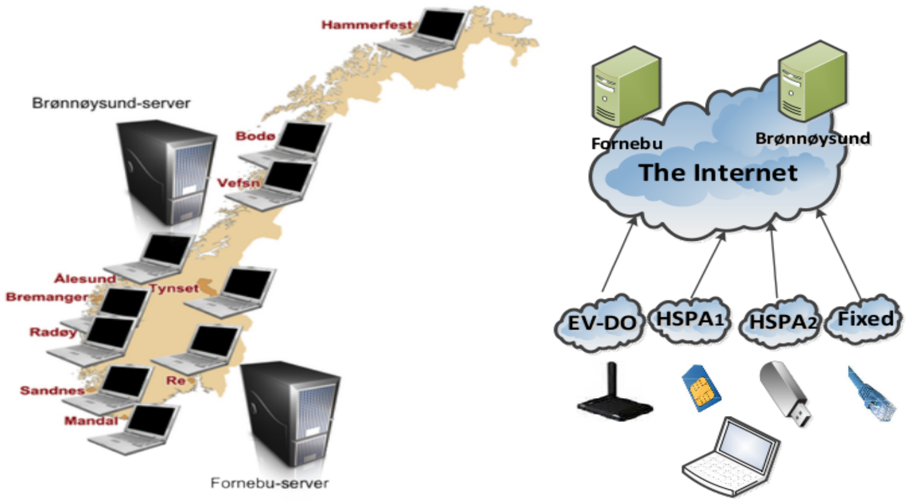
2 Measurement setup and data

We have built a measurement infrastructure consisting of 90 measurement hosts in 10 municipalities across Norway as shown in Fig. 1a. Our measurement nodes are hosted in separate locations within each municipality; the average distance between two monitors in a municipality is 7.7 km. The infrastructure also includes two servers, one is located in the middle of Norway (Brønnøysund) and the other one is located in the south east of Norway (Fornebu)[1].

Our measurement node is a Dell Latitude E6510 laptop running Ubuntu 10.04. As shown in Fig. 1b, each node is multi-homed to four ISPs, three of them are MBB providers. The fourth operator is which ever fixed broadband provider that is available on-site. This connection will have varying quality, from high-speed fiber connection in some locations to nothing at all in other. In this paper, we use fixed broadband measurements as a reference point for comparing the performance of the MBB providers. Operators 1 and 2 offer a High Speed Packet Access (HSPA) based data service, an evolution of Wide-band Code Division Multiple Access (WCDMA). In locations where the HSPA service is not available, the connection reverts to EDGE/GPRS. In the following, we refer to these operators as $HSPA_1$ and $HSPA_2$. Operator 3 offers a CDMA2000 1xEV-DO (Evolution-Data Optimized) based data service, we refer to this operator as *EV-DO*. Our measurement node connects to these 3G operators through the following devices. Dell built-in wireless 5540 HSPA mobile broadband mini-card ($HSPA_1$), ZTE MF636 USB modem ($HSPA_2$), and Huawei EC506 wireless

---

[1] For more information about our measurement setup please refer to http://nevada.simula.no/



(a) Measurement infrastructure (b) A measurement node

Fig. 1: Measurement setup

router (*EV-DO*). We discuss the impact of the different modems on the measured delays in the next section.

Each node periodically runs *ping* and *traceroute* measurements through each of its four interfaces to the two servers indicated above. Ping measurements are performed every second through the fixed connection and every 5 seconds through the wireless networks. Traceroute measurements are performed every 10 minutes. We use a modified version of Paris traceroute [2], where we have added support for specifying which interface to use for each run. We also use AT commands every minute to measure the received signal strength. Our measurements cover the period from February to August 2011, but in this paper we often use only a subset of the data collected as long as this does not influence the results. Most of our analysis is based on data collected during July 2011.

The scale and complexity of our infrastructure poses several challenges regarding its management and operation. To minimize the administration overhead (e.g. traveling to remote sites), we have designed our monitors to be as self-administered as possible. Each host maintains a reverse SSH session with our Fornebu server, to be used by the host for uploading its measurement data, and by the server for pushing new configurations and remote management when needed. Further, each node stores measurement data locally and uploads it every day to the server at around 3 AM. A monitor periodically checks the status of the SSH session and all four network interfaces and automatically tries to restore any failing session or interface. IT personnel at remote municipalities help when on-site intervention is needed on a voluntary basis. Thus, long response times are expected when a node is permanently down. Another challenge that we have faced is the instability of $HSPA_2$'s 3G USB modems; the majority of them require frequent physical removal and re-plugging. Due to these challenges we use measurements from around 60 hosts out of 90 in this study, and only 17 $HSPA_2$ monitors.

3 Delay characteristics of Norwegian 3G networks

In this study, we use the IP-layer tools *ping* and *traceroute* to measure the end-to-end delay between the measurement nodes and our servers. This means that we are not able to dissect the contribution of the different components in the 3G access networks (such as the base station and the Radio Network Controller) to the total delay. Using *traceroute*, we are still able to compare the RTT in the first IP-hop to that of the end-to-end path. The first IP hop in 3G networks will typically be the Gateway GPRS Service Node (GGSN).

In this section, we present our findings regarding delay characteristics in the measured MBB networks.

There are large differences in delay between operators. The left panel in Fig. 2 illustrates a typical CDF of RTTs measured at one of our monitoring points during July 2011. All MBB networks exhibit roughly an order of magnitude higher delay than the fixed network. Delay varies significantly between networks; we note that $HSPA_1$'s delay is higher than that of *EV-DO* and $HSPA_2$, and varies in a wider range between 200ms and 600ms.

The right plot in in Fig. 2 shows the 5th percentile, median, and 95th percentile of RTTs measured in July 2011 between each monitor and the Fornebu server[2]. This figure shows that there are large and consistent differences in delay between operators. $HSPA_1$ shows the highest delay (median RTT $\sim$ 300ms across all monitors). Then follows *EV-DO* (median RTT $\sim$ 180ms), before $HSPA_2$ (median RTT $\sim$ 104ms). Note that, as explained in Sec. 2 we have fewer monitors of type $HSPA_2$. The fixed line RTTs are significantly smaller (median RTT $\sim$ 16ms) than all MBB operators.

We also record large variations between monitors in the same operator and even within a single connection. $HSPA_1$'s RTTs in a single connection shows large variations reaching up to two orders of magnitude. In some cases, the round trip delay can reach several seconds, even tens of seconds. Across monitors, *EV-DO*'s RTTs are more stable than those of $HSPA_1$ and $HSPA_2$. It's median RTT varies between 162ms and 297ms across monitors. The same metric varies between 82.5ms and 1691ms in $HSPA_1$; and between 71.2ms and 740ms in $HSPA_2$.

The observed differences cannot be explained by different modems alone. As described in Sec. 2, we use different modems to connect to the different operators. It is therefore natural to ask whether the choice of modem can explain the observed differences. To investigate this, we have run controlled experiments with different modems for each operator. Table 1 shows the median delay recorded over a 24 hour period using different modems[3]. The measurements for each operator are taken in parallel during the same 24 hour period. All modems are USB sticks, except the internal modem and Huawei EC506 (which is a standalone wireless router). The values marked with a star represent

---

[2] Measurements to the other server show similar results.

[3] Due to the different technologies and provider locks, we are not able to test all modems across all operators.

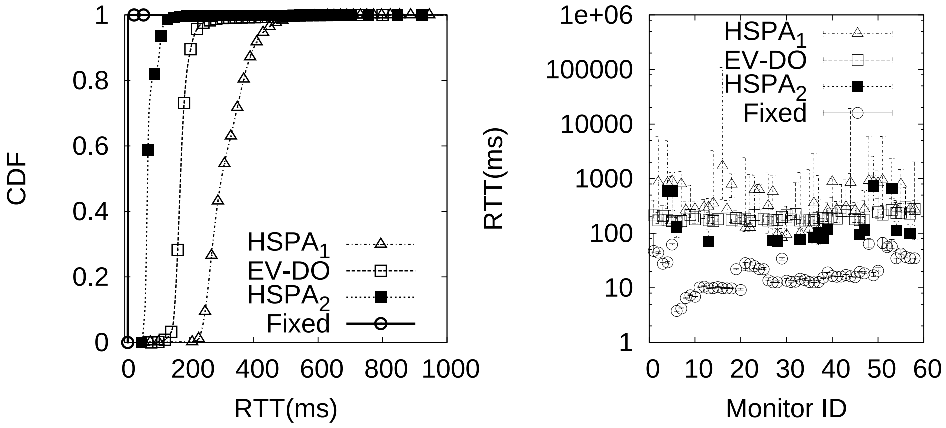


Fig. 2: Example of a typical RTT CDF (left), RTTs statistics (right)

the modem that was used in the long-term measurements. We observe that the choice of modem has a marked influence on delay, but that it is far from the dominant factor. We plan to do more systematic evaluations of the role of the modem in future studies.

Operator	Internal	ZTE MF636	Huawei E1752	Huawei EC506	C-motech D50
$HSPA_1$	282 ms*		368 ms		
$HSPA_2$	57 ms	72 ms*	64 ms		
$EV\text{-}DO$				164 ms*	81 ms

Table 1: Comparing Modems

While there are sometimes large differences between monitors of the same operator, they mainly belong to the same population. Our previous observations sometimes show large variations in delay between monitors of the same operator, thus it is interesting to check whether these differences are inherent in MBB networks or just reflect local effects near an affected monitor (e.g. poor wireless coverage). To answer this we investigate differences between delay distributions of monitors that belong to the same operator.

To compare two different delay samples as to whether they belong to the same population, we need to pick an appropriate statistical test that suits our data. First, it is reasonable to avoid parametric tests (e.g. t-test), since we cannot make assumptions about the underlying probability distribution of the RTT data. One possibility is to apply the two-sample Kolmogorov-Smirnov test [5] for comparing continuous, one dimensional distributions. But, RTT distributions are not continuous, thus we decide to employ the Kullback-Leibler (K-L) divergence test instead [7]. The K-L divergence is a measure for the closeness between two samples P and Q in terms of extra information bits required to encode a message based on P instead of Q. Note that the K-L divergence in general is not symmetric.

K-L divergence by itself cannot determine whether the two tested delay samples are drawn from the same population at a certain confidence level. Hence, we construct a hypothesis test that is inspired by the approach used in [10]. In

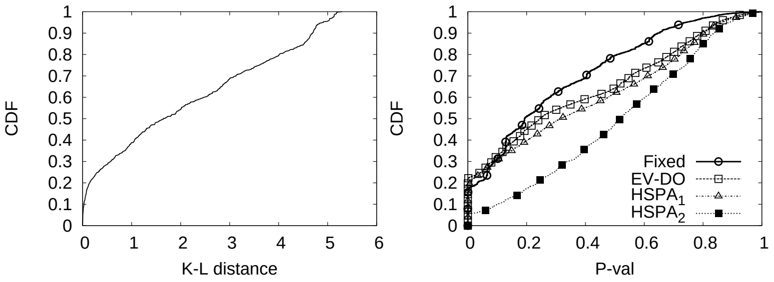


Fig. 3: KL distance distribution (left), Resulting P-values distribution (right)

the following, we present this hypothesis test; our null hypothesis H_0 is that the tested samples have identical underlying distribution.

For each interface and monitor, we draw 30 random mutually exclusive samples of equal sizes from RTT measurements in July'11. We then calculate the K-L divergence for each pair, that results in 870 values. These values are then used to estimate the corresponding empirical CDF of K-L divergence. The left panel in Fig. 3 presents an example of such CDF estimated for one of our *EV-DO* monitoring interfaces, in the following we call this interface X. If we want to compare the distribution of RTTs measured at another *EV-DO* interface Y to that of X, we first measure the K-L divergence between Y and X. Let's assume that $K\text{-}L(Y, X) = 2$. We use the empirical CDF of K-L divergence values at X to find $F(K\text{-}L \text{ distance} = 2)$. This value represents the probability that such divergence can occur between two samples drawn from the RTT population of interface X. The P-value of our test is then calculated as $1 - F(K\text{-}L$ distance $= 2)$, we accept H_0 if $P\text{-}value > 0.05$, i.e. the probability that such divergence occurs between two samples from the same population is at least 5%. In our example, the P-value is 0.46 thus we accept H_0.

Using our constructed hypothesis test we compare all pairs of distributions from the same operator. The right plot in Fig. 3 shows the CDF of the calculated P-values. We observe that a large fraction of pairs in all interfaces is characterized by a P-value larger than 0.05, meaning that the majority of RTT distributions come from the same population. Our results show that at least 75% of all monitor pairs from the same operator belong to the same population. We also compare RTT distributions across operators and find that a significant fraction of pairs do not belong to the same population. For example, only 49% of all pairs are of the same population, when comparing *EV-DO* to $HSPA_1$. This is in agreement with our earlier observations in Fig. 2.

The fact that delay distribution in most monitors of the same operators mainly come from the same population is quite interesting. First, recalling the large differences between operators, it seems that each operator has its own "signature" in the delay characteristics. Second, it shows that the delay characteristics of a connection is mainly network-dependent rather than monitor-dependent.

3G access network plays a central role in deciding delay characteristics. The 3G access network seems to play a central role for the delay characteristics. We cannot directly measure this (since there are no IP hops in this network), so we investigate this by looking at delay correlations between monitors at different geographical distances. We first calculate the temporal correlation between all pairs of RTT time series from the same operator. To construct these time series, we use one month of delay measurements per monitor, divide it to five-minute bins, and calculate the average RTT in each bin. Second, we examine how temporal correlation between two time series varies in relation to the geographical distance between the respective monitors. To estimate correlations between monitors, we use the non-parametric Kendall's τ rank correlation coefficient [5]. τ takes value between -1 and 1, and it represents the difference between the probability that the observed data are in the same order in both samples and the probability that they are not.

The left panel in Fig. 4 depicts τ's CDFs for monitors that are at most 100km apart for all operators. The MBB operators demonstrate stronger correlation than the fixed network. In the middle panel, we plot the τ's CDFs corresponding to our fixed line monitors. Each curve represents correlations between monitors that are within a specific distance range from each other. We observe that the temporal correlation between the fixed line monitors is generally low ($\tau \leq 0.3$ in almost 80% of the cases). Furthermore, *distance between monitors has a negligible impact on their correlations.*

Interestingly, we observe a quite different behavior in the MBB networks. *Monitors that are up to 300km apart are strongly correlated. Beyond that the correlation properties are similar to those of fixed line monitors.* The right panel in Fig. 4 illustrates this for $HSPA_1$. Monitors in *EV-DO* and $HSPA_2$ behave similarly. It is natural to relate this behavior to the architecture of 3G networks, where geographically close base stations share the same Radio Network Controller (RNC). This result shows that the 3G access network is an important contributor to the overall delay characteristics, and indicates that queuing at the Base Station Controller (BSC) level plays an important role.

The access network is a decisive factor for delay, but is not responsible for outliers. The common wisdom is that last mile delay constitutes a large fraction of end-to-end delay in wireless networks. The last mile includes the part of the 3G network between an end device and the first IP hop in the respective provider's cloud (i.e. the GGSN). In a wired network, the last mile corresponds to all physical infra-structure that lies between a customer's access device (e.g. ADSL modem) and the first gateway in her ISP's network.

In order to quantify the contribution of the access network to observed RTTs, we consider the ratio (r) of the last mile RTT to the end to end RTT. We employ our traceroute measurements to estimate the last mile latency (i.e. by extracting the RTT to the first IP hop in the respective provider's network). The left panel in Fig 5 illustrates r's CDF, each curve is estimated by combining r values from all monitors of the corresponding operator. As expected, we observe a clear difference between the fixed network and the three wireless interfaces. In the fixed

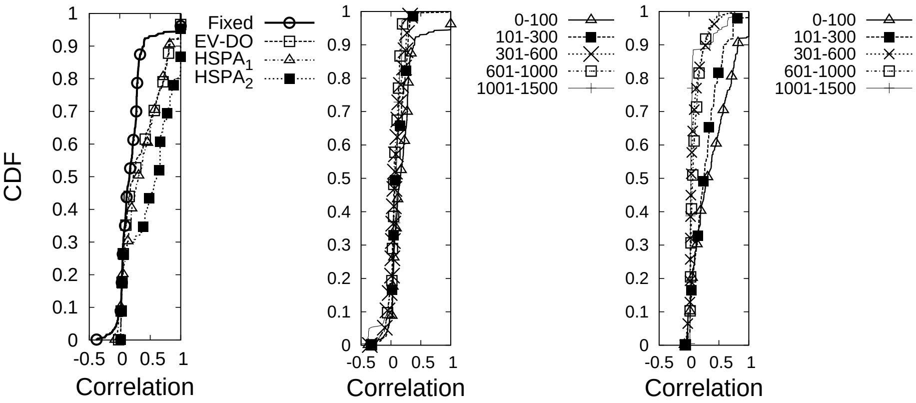


Fig. 4: Delay correlation between monitors in relation to their geographical distance. Monitors that are up to 100 km apart (left), Fixed (middle), $HSPA_1$(right)

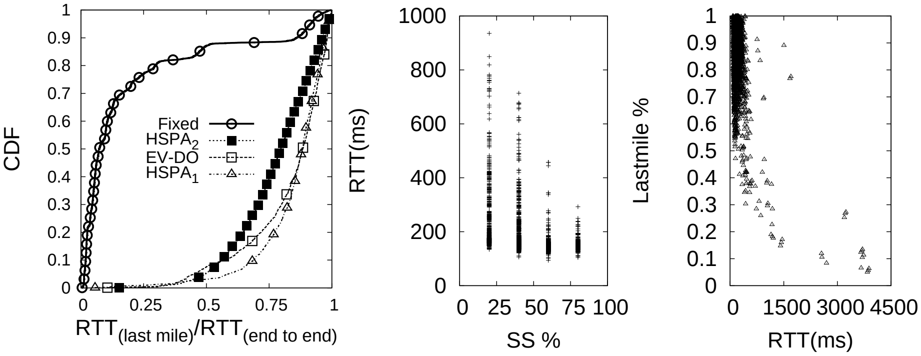


Fig. 5: Last mile delay characteristics

network, r is less than 10% in 50% of the cases, but in the wireless networks it is more than 50% in 90% of the cases. We also note that the contribution of the access network to observed RTTs is higher in $HSPA_1$ than in *EV-DO* than in $HSPA_2$. The last observation is in accordance with our earlier findings.

Several factors contribute to the last mile latency including modem performance, signal quality, queuing in the access network beyond the first wireless hop, and the impact of different components in the access network (e.g. RNC). The middle panel in Fig. 5 shows the relation between last mile RTTs, measured in all *EV-DO* monitors, and the received signal strength (SS) represented as a percentage of the best attainable signal quality. Note that, the SS we measure only covers the received signal. We observe that *last mile RTTs increase as SS deteriorates*. In our future work, we plan to investigate the impact of other factors (e.g. queuing in the access network beyond the first wireless hop) on the last mile delay.

Finally, we ask whether the access network is responsible for the very high delay values that we sometimes experience. The right panel in Fig. 5 depicts the

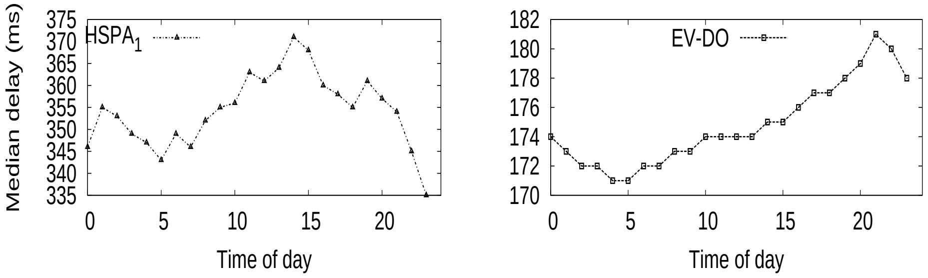


Fig. 6: Delay variation during the day

relation between r and RTTs aggregated across all $HSPA_1$ monitors. Surprisingly, *we observe that* r *decreases as RTT increases*, suggesting that such large RTTs are caused by performance degradation beyond the access network, i.e., in the GGSN or at the IP path from the GGSN to the measurement servers. *EV-DO* and $HSPA_2$ show a similar behavior.

3G delays exhibit clear diurnal patterns. To indirectly measure coarse-grained traffic patterns, we explore how delay varies throughout the day. The plots in Fig. 6 shows the median delay of $HSPA_1$ and *EV-DO* as a function in the time of the day. To calculate for operator X at hour H, we collect all RTTs from all measurements from X that are recorded at hour H throughout May'11 and then find their median[4].

Not surprisingly, we find clear diurnal patterns in delay. This has earlier been reported in many studies from fixed networks [3]. Interestingly, delay peaks at different hours in our measured networks. While $HSPA_1$ delays are highest during business hours, *EV-DO* shows a different pattern with higher delays in the evening. We have been in contact with *EV-DO*, and they confirm that they see more traffic in their network at these hours, probably because they have a large number of home and recreational users.

4 Related work

Laner et. al [8] measured 3G uplink delay in an operational HSPA network and showed that the average delay is strongly dependent on the packet size. Further, they found that last mile delay constitutes a large fraction of measured delays. The authors in [4] analyzed packet delay in UMTS networks and identified ARQ loss recovery mechanisms as the main cause behind the high variability in packet delay. Arlos and Fiedler [1] measured the influence of the packet size on the one-way delay (OWD) in 3G networks in three different operators in Sweden. They showed that choosing an optimal packet size significantly reduces OWD. In contrast with previous work that measured delay in 3G networks, we provide a more complete study that involves two different 3G technologies, three operators and about 60 monitoring points.

Other papers (e.g. [6, 9]) measured different set of performance metrics in 3G networks; including TCP and UDP performance, throughput, and network resource allocation.

---

[4] July data show similar patterns.

5 Conclusions

This work presents a first look on long-term measurements of MBB data connections from 3 different network operators in Norway. More specifically, in this paper, we investigate the characteristics of round trip delays with a focus on the role of the 3G access network. We observe large differences between operators with respect to both absolute delays and variations. Access network latency constitutes a significant part of the total delay. However, its share drops at large RTTs. We also observe that delays in 3G networks exhibit clear diurnal patterns that peak at different hours during the day depending on the operator.

Interestingly, we find that the delay characteristics in different 3G networks are mainly network-dependent rather than monitor-dependent, and that each operator has its own "signature" in the delay characteristics. These findings indicate that differences between MBB operators are mainly dictated by the way their access networks are designed and configured. The important role played by the 3G access network in deciding delay characteristic is further confirmed through analyzing correlations between monitors of the same operator. The identified strong correlation between geographically close 3G monitors indicates the presence of significant infrastructure aggregation within each operator access network.

References

1. P. Arlos and M. Fiedler. Influence of the Packet Size on the One-Way Delay in 3G Networks. In *PAM*, 2010.
2. B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira. Avoiding traceroute anomalies with Paris traceroute. In *IMC*, 2006.
3. N. Brownlee and k. claffy. Understanding Internet Traffic Streams: Dragonflies and Tortoises. *IEEE Communications*, pages 110–117, Jul 2002.
4. J. Cano-Garcia, E. Gonzalez-Parada, and E. Casilari. Experimental analysis and characterization of packet delay in UMTS networks. In *NEW2AN*, 2006.
5. M. Hollander and D. A. Wolfe. *Nonparametric statistical methods.* Wiley, second edition, 1999.
6. M. Jurvansuu, J. Prokkola, M. Hanski, and P. Perala. HSDPA performance in live networks. In *ICC*, 2007.
7. S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.
8. M. Laner, P. Svoboda, E. Hasenleithner, and M. Rupp. Dissecting 3G uplink delay by measuring in an operational HSPA network. In *PAM*, 2011.
9. W. L. Tan, F. Lam, and W. C. Lau. An Empirical Study on the Capacity and Performance of 3G Networks. *IEEE Transactions on Mobile Computing*, 7(6), 2008.
10. M. M. B. Tariq, A. Dhamdhere, C. Dovrolis, and M. Ammar. Poisson versus periodic path probing (or, does PASTA matter?). In *IMC*, 2005.