Behind the NAT – A Measurement Based Evaluation of Cellular Service Quality

Fabian Kaup[1], Foivos Michelinakis[2,3], Nicola Bui[2,3], Joerg Widmer[2], Katarzyna Wac[4,5], David Hausheer[1]

[1]Peer-to-Peer Systems Engineering Lab, Technische Universität Darmstadt, [2]IMDEA Networks Institute,
[3]Universidad Carlos III de Madrid, [4]Université de Genève, [5]University of Copenhagen
Email: {fabian.kaup|hausheer}@ps.tu-darmstadt.de, Katarzyna.Wac@unige.ch,
{foivos.michelinakis|nicola.bui|joerg.widmer}@imdea.org

Abstract—**Mobile applications such as VoIP, (live) gaming, or video streaming have diverse QoS requirements ranging from low delay to high throughput. The optimization of the network quality experienced by end-users requires detailed knowledge of the expected network performance. Also, the achieved service quality is affected by a number of factors, including network operator and available technologies. However, most studies focusing on measuring the cellular network do not consider the performance implications of network configuration and management. To this end, this paper reports about an extensive data set of cellular network measurements, focused on analyzing root causes of mobile network performance variability. Measurements conducted over four weeks in a 4G cellular network in Germany show that management and configuration decisions have a substantial impact on the performance. Specifically, it is observed that the association of mobile devices to a Point of Presence (PoP) within the operator's network can influence the end-to-end RTT by a large extent. Given the collected data a model predicting the PoP assignment and its resulting RTT leveraging Markov Chain and machine learning approaches is developed. RTT increases of 58% to 73% compared to the optimum performance are observed in more than 57% of the measurements.**

I. Introduction

The quality of service (QoS) of mobile applications such as mobile cloud gaming or live video streaming relies on certain bandwidth and round-trip time (RTT) guarantees [1], [2]. Also the profit of commercial web services depends on low network latency [3]. To optimize the QoS according to the available network capacity and to avoid failing connections such as dropped VoIP calls, mobile applications need to rely on information about the performance of the underlying mobile network. This depends on a number of aspects, including the number of active users, their location, the time of the day, and likely also the day of the week, amongst other factors. One approach to improve mobile applications' performance is to use the information provided by mobile operators. However, the performance figures (e.g. data rates) advertised by mobile operators in many cases report primarily maximum values that often do not match the actual performance [4]. Moreover, detailed information about average number of users per cell, backbone capacity and also expected delays and available throughput are not available at the required granularity for confidentiality reasons.

An alternative approach is collecting network performance data on end-user devices in a crowd-sourcing based approach [4]–[6]. This allows estimating and predicting from those the cellular network quality at certain times and loca-

tions, as experienced by the end-user. Still, active measurements are expensive in terms of energy and may even cause monetary cost for the users. Knowing the network and cell utilization, a number of optimizations can be derived for the handset. For example background traffic may be delayed as long as the network performance is low. This allows app developers to optimize their application to adjust the quality of the consumed content a priori, and to improve the user experience by reducing delays, reducing the power consumption, and such increasing the quality of experience (QoE) of the end-user.

This paper is focused on analyzing the parameters affecting the mobile network performance. Special emphasis was put on the RTT, as an increasing number of services are delay sensitive (e.g. mobile cloud gaming, or health care). Hence, this paper aims to address the following questions:

1) How does the network configuration and management influence the cellular network performance?
2) Can the performance of the cellular network be predicted based on the available measurements?
3) What is the optimization potential in the given network configuration, and how can this be exploited?

The analysis is based on the results of a crowd sourced measurement campaign, and an extensive stationary network measurement study. The evaluated subset is detailed in the respective sections.

The findings in this paper differ from related studies in that the analysis performed also includes the data paths taken through the network to derive their influence on the end-to-end performance. RTT increases of more than 58% have been observed compared to the best performance measured for stationary devices. Changes in RTT exhibit a periodicity of 36 hours. Combining both with the traces through the network leads to the conclusion that network configuration and management are the main factors affecting the cellular network performance in the measured network.

The remainder of this paper is organized as follows. In Section II, similar studies and related publications are summarized. Section III explains the measurement methodology applied and summarizes the data sets in Section IV, which are used in the analysis in Section V. Based on the network measurements and the time, a prediction model for the network performance is proposed in Section VI. Finally, Section VII concludes the paper and gives an outlook into further research directions.

II. Background and Related Work

The analysis of the mobile network performance [4]–[6], its prediction [6], [7], energy consumption [8], [9], or mechanisms improving the mobile QoS, while preserving the QoE [10] are in the constant focus of current research. A large number of studies focus on 3G networks [6], [9], [11], while others [4], [12]–[14] evaluate the low-level network-based performance or model the energy efficiency.

A number of publications analyze the cellular performance based on samples collected at the mobile backbone. Gerber et al. [15] describe an approach to measure the maximum mobile throughput based on packet traces. This approach is extended by Huang et al. [4], evaluating the throughput of a 4G network and individual flows in a similar manner. Both studies observe the traffic generated by a large number of devices over the area of several base stations. This allows deriving data rates for single devices as well as the overall throughput from within the network. In contrast, this paper analyzes the RTT with focus on long-term variations.

Wylie-Green et al. [16] analyze the throughput and RTT performance of an unloaded and loaded LTE network from different distances using multiple devices. The described measurements were conducted on newly built hardware and hence analyze the optimum network performance. The purposely generated traffic does not affect the system performance, as it is below the system capacity.

A detailed analysis of the one-way delay of different parts of the network can be found in [17]. The authors measured the timing between a local machine connected to a LTE and HSPA network, a time synchronized server, and additional vantage points within the mobile network. In this work, end-to-end measurements from the mobile device are used to evaluate the performance based on network management decisions.

A different approach, based on measuring the network performance from handsets only, is published by Sonntag et al. [5]. They developed an application measuring a number of network parameters like throughput and RTT, and collect the data on their server. The analysis of the collected data is limited to a few general metrics and the creation of bandwidth maps. Contrary, this paper analyzes the measured RTT in detail to derive the performance based on the path taken through the cellular network.

The prediction of the availability of WiFi networks is used to improve mobile connectivity by Nicholson et al. [18]. This approach is extended by Bui et al. [7], who propose a stochastic mobile bandwidth prediction model. A different approach is used by Wac [19], applying machine learning techniques to predict mobile QoS. In this paper, this approach is extended to forecast the RTT based on the time of the connection.

Contrary to [4]–[6] this paper focuses on the root cause analysis of performance variations in the cellular network. Two different data sets were collected, complementing each other. The crowd sourcing based data set is similar to the data collected by Sonntag et al. [5], but additionally includes fine granular location and cell information. The complementary data set extends the first one by RTT measurements to the DNS server and trace-routes between the mobile device and measurement server. Based on the combination of both, a detailed analysis of the cellular network performance becomes possible. No such data set or analysis is available to the best of the author's knowledge.

III. Measurement Methodology

The mobile network measurements have been performed with our Android application, the *NetworkCoverage App*[1]. This application was developed to gather measurements from various devices, users, and locations. It consists of a service that passively monitors system resources like GPS and the cell interface and a component executing active measurements probing the network performance by generating traffic on the cell or WiFi interface.

The application passively monitors the local signal strength, the network provider, and the used wireless access technology. These measurements are linked to the time stamped location derived from the Android framework. The location is either GPS-based, or based on information from available networks using Google's location API. Active measurements include RTT and throughput measurements. The RTT measurements are executed periodically in background, resulting in a large number of samples. Throughput measurements are generally only collected upon user request, as the data transfer might cause additional data cost at the user's end. The measurement values are stored in a local *SQLite* database and later transmitted to a data collection server for analysis.

Both RTT and throughput are measured with respect to a dedicated server at TU Darmstadt (Germany), allowing the maximum flexibility in the measurement setup. The server and the university network are well connected to the public Internet using a lightly loaded network. The mobile devices used in the measurement campaigns are all Nexus 5 devices running Android 4.4.4, while the devices used by the users of the public application vary. Hence, thorough filtering was applied to assure high quality of the measurements.

When measuring throughput and RTT, the cellular network poses additional challenges compared to wired networks, requiring adaptation of the measurement procedure. Due to the mobility of the user, the duration of the measurements must be kept to a minimum. This is particularly important for the throughput measurements, as the available data rates may change within seconds.

The RTT measurements use the native ping command on the Android device to determine the latency to the server. The minimum, average, and maximum RTT out of five consecutive ping measurements (excluding the connection upgrade delay) were recorded. These measurements were run once per minute to obtain a fine-grained view on the network performance.

Network operations on mobile devices cause an additional, device internal delay, stemming from kernel and operating system processing the outgoing and incoming packets [20]. This delay is different depending on the chosen measurement implementation (e.g. native ping, HTTP ping, socket connect duration). For our measurements, the RTT is measured using the native ping command, eliminating the influence of the Android virtual machine and its scheduling on the accuracy

---

[1]https://play.google.com/store/apps/details?id=de.tudarmstadt. networkcoverage

of the measurements. The ping command directly accesses the low level timing functions. According to [18] the additional delay of the sent and received packets on the Nexus 5 for a measured RTT of 20 ms and 50 ms is 6.028 ms and 7.70 ms with a 95 % confidence interval of ±0.811 ms and ±2.331 ms.

The TU Darmstadt server used for the RTT measurements is also used to measure the throughput between the end-user device and the server. The measurements are run using a patched version of iPerf[2], allowing to measure the down-link bandwidth (conventional iPerf always measures the up-link from the iPerf client). The generated traffic is reduced by limiting the time of the measurement, which by default is set to run for 10 s. In lab trials the minimum measurement duration on WiFi was determined to be 3 s for the connection to exit the slow start phase and remain within a limit of ±10% compared to the full measurement. Repeating the same measurements on a 3G network led to a minimum measurement interval of 5 s due to the higher RTT. The accuracy of this setting was also verified on LTE and hence set as the default interval for all measurements.

IV. DESCRIPTION OF THE DATA SETS

The data sets used for the analysis of the network performance are split into two sets as the focus of the analysis and the collected data are slightly different. Furthermore, this simplifies referring to the respective subset of measurements. The first data set consists of the crowd-sourcing based data used for the estimation of the location-dependent network performance, while the second measurement study is focused on the time-based analysis of network performance.

A. Crowd-sourcing based cellular network measurements

The first data set, targeted at analyzing the location-based network performance, was collected by users of the NetworkCoverage Application. The measurements go back to December 2013, when the first version of the application was released. The full data set consists of 7.1 M coverage samples, 395 k RTT samples, and 6.8 k throughput measurements. The measurements cover mainly Darmstadt and the surrounding areas, although a few measurements are collected in other cities. Table I lists the break down for the different technologies and measurement types.

The full data set covers 6596 cells operated by 64 cellular service providers worldwide. An interesting effect is that single cell IDs of one operator have been observed serving multiple generation networks. According to the mobility of the users, some cells observe a high sample count (>10 k), while others are only observed a small number of times. The median number of measurements per cell is 12, while the mean is 636 measurements.

During this time, also a measurement study collecting mobile traces for the analysis of the RTT and its relation to the signal strength and the network throughput was run, which began on June 2nd 2014 and ran for two weeks. For this, five students were recruited to carry one smartphone periodically probing the network performance during their daily routine. The phone model used by all participants in the study was

TABLE I.   OVERVIEW OF ALL COLLECTED MEASUREMENTS

Type	4G	3G	2G	all
Signal Strength	6.5 M	538 k	64 k	7.1 M
Cells	1748	14 k	2 k	17,9 k
RTT	344 k	45 k	5 k	395 k
Throughput	5.5 k	1.3 k	36	6.8 k

the Nexus 5. Most of the time, these were connected to the LTE network, providing a basis for the detailed analysis of its performance. Over the course of the measurement study, the RTT measurements were executed once every minute, and the throughput measurements were run every 25 minutes.

B. Data set collected by the stationary measurement study

To assess the time-based performance of the cellular network, additional measurements are required. Hence, a second study targeted on the time-variability of the network performance was run. Contrary to the first study, the location of the devices was kept stationary to keep as many parameters constant as possible. The target was to reduce the effect of handovers, different network technologies, and changing signal strength. The measurement study was run between December 19, 2014 and January 19, 2015, and hence includes measurements of two regular weeks as well as measurements during the Christmas and New Year's celebrations. An overview of the measurements recorded during the study is given in Table II.

The second data set consists of measurements collected by devices placed at three different locations in the city, allowing to assess the time-based network performance without the influence of changing signal strengths, cell IDs, network technologies and handovers. Two devices were positioned in different residential areas approximately 1 km from the city center, while a third device was located in an office building adjacent to the central park. These locations were chosen to represent different loads in different cells of the mobile network. These measurements were run for 4 weeks to assess the time-based network performance. For reference, two devices were taken on the usual trips between the residential areas and the office building for the course of one week.

Additionally to the previously measured metrics, also the trace-route between the mobile device and the server was recorded on the mobile devices. This will be referenced as the forward trace-routes in the remainder of this paper. As these proved problematic (due to errors or traffic engineering in the cellular network), the measurement server was extended to also run a trace-route to connecting mobile hosts. In the following, this is referenced as reverse trace-route. The goal of this setup is to defer the routes taken through the core network, and such derive the root cause of the observed RTT deviations. The recorded information on the server consists of an identifier of the mobile device, the time stamp, and the complete output of the trace-route. By combining both measurement, the RTT to the different hops can be derived.

Since ICMP unreachable messages, as provoked by the trace-route measurements, are generated with lowest priority by intermediate routers, measurements may exhibit high variance. Further, it is possible that intermediate hops show an increased RTT, while the next hop shows a better performance. Each hop was probed with three packets. Due to the low

TABLE II.   OVERVIEW OF THE MEASUREMENTS COLLECTED DURING
THE STATIONARY MEASUREMENT STUDY

Type	4G	3G	2G	all
Signal Strength	3 M	11.2 k	0	3 M
RTT	155 k	1.3 k	0	156 k
Throughput	2.3 k	15	0	2.3 k

TABLE III.   OVERVIEW OF THE MEASUREMENTS COLLECTED DURING
THE REFERENCE MEASUREMENT STUDY

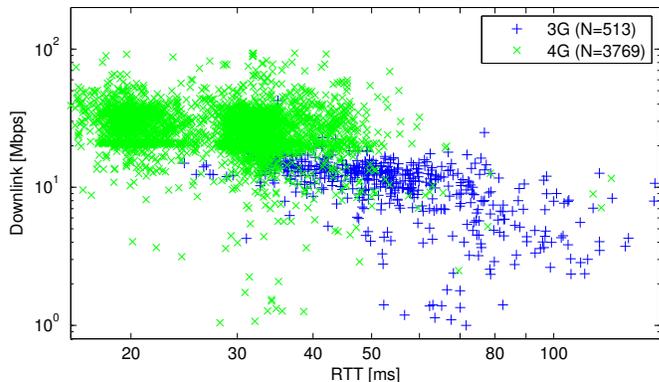| Type	Number
Signal Strength	755 k
RTT	38 k
trace-routes	5 k
reverse trace-routes	7.4 k
DNS RTT	6.4 k

Fig. 1.   Relation between down-link bandwidth and RTT

Fig. 2.   Distribution of the measured down-link rates in a 4G network. The numbers in braces denote the number of observations

priority of these packets, only the minimum of one probe was considered in the later evaluation.

These reference measurements, including the traceroutes, were exclusively conducted on the 4G network. The end-to-end RTT as in the previous campaigns was recorded once every minute. The forward and reverse trace-routes as well as the RTT to the DNS server were executed once every 15 minutes, limiting the load on the intermediate routers. This is important as each is probed individually, also reducing the risk of being blocked. The overview of the additional measurements conducted during the reference study is given in Table III.

V.   ANALYSIS OF THE 4G NETWORK PERFORMANCE

From the full data set as described in Section IV-A a number of conclusions can be drawn. First, relations between individual parameters are described. Then, the observed patterns are analyzed in detail using additional information on the network topology derived from the traceroute measurements. From this the influence of network management and configuration on the achieved network performance is derived.

A. General observation

The correlations between different measures were analyzed to determine whether more expensive measurements (e.g. throughput) can be estimated using cheaper ones (e.g. signal strength, RTT). Figure 1 shows the relation between the RTT and the down-link rate. The 3G measurements show a consistent maximum throughput rate, while the RTT is almost uniformly distributed in the range of 30 ms to 80 ms. Also for an increasing RTT a decreasing data rate is visible. An interesting observation is that the 4G measurements show RTT clusters at around 20 ms and 30 ms. This will be analyzed in detail in Section V-D.

The signal strength is given as arbitrary strength unit (ASU), which is a linear mapping from the signal strength in dB to integer values. The influence of the signal strength on the maximum throughput of wireless networks is well known [21]. This effect is visible in the throughput measurements conducted during this study. Figure 2 shows the measured down-link rates for each signal strength value. The second row of values indicates the number of samples contributing to each group. The median of the measurements is plotted in the
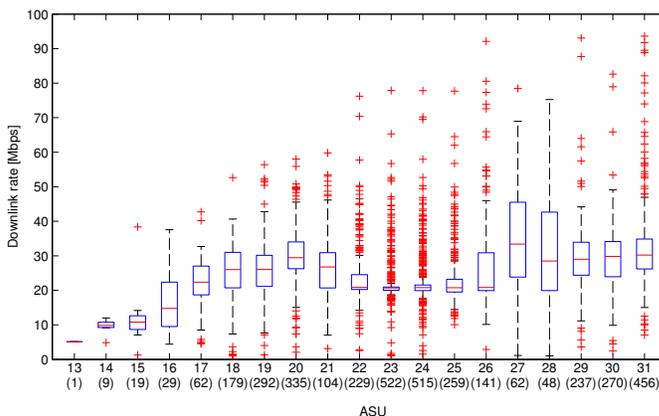
center of the box, while the box itself represents the inner quartiles above and below the mean. The whiskers extend to the maximum value within 150% of the height of the inner quartiles. All other values are marked by a plus, denoting outliers. As expected, the throughput increases with increasing signal strength, but only up to medium values (∼18). Above, other factors appear to be limiting the maximum throughput.

Figure 3 shows the histogram of the measured RTTs in the 4G network. The graph shows three peaks, one at 20 ms and two merged ones at around 28 ms and 31 ms. Only a small number of samples is observed in the in the gap between the first two peeks. As the scheduling of packets in LTE is conducted once per ms, samples between the peaks should be present. Retransmissions in LTE (HARQ retransmissions) are initiated after 8 ms, but considering the number of samples in the second peak, a retransmission of 2/3 of the packets is highly unlikely. As no simple explanation for this behavior could be found in literature, additional measurements were conducted to analyze the effect in detail.

B. Analysis of the network topology

The network topology is defined by the intermediate routers processing the generated traffic between the smartphones and the measurement server. This is analyzed in detail by running trace-route measurements on the mobile devices to the server,
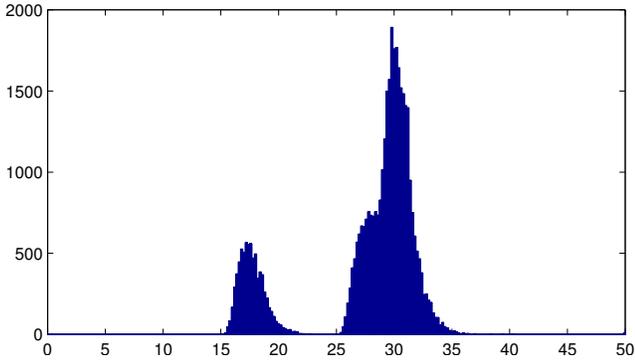
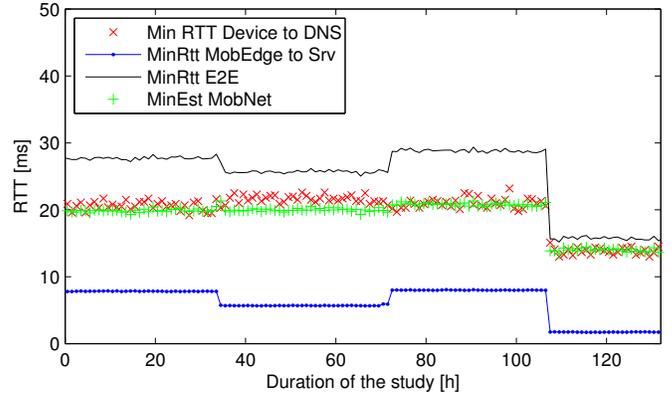Fig. 3. Histogram of the minimum 4G end-to-end RTTs (bin width: 1 ms)

Fig. 4. Exemplary RTT measurements for one device with the estimated RTT within the mobile operator's domain
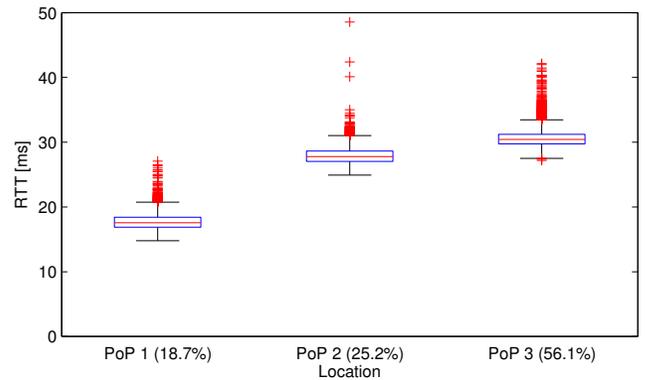

Fig. 5. Boxplot of the minimum RTTs and the fraction of measurements between mobile device and measurement server, grouped by the derived PoP of the operator

and recording the path taken from the server in the direction of the mobile device. The detailed measurement setup is described in Section III. In the trace-route measurements, the mobile part of the network is opaque, i.e. no ICMP unreachable messages are generated by routers within the mobile operator's domain. From the host names it can be derived that the last visible hop corresponds to a router in the core network, which is likely the gateway to the mobile part of the network. In the following, this is taken as the discriminating point between the core network and the mobile domain.

To compare these measurements, the trace-route measurements on both sides and the end-to-end RTTs are correlated using the device identifier stored with every measurement and the time. The RTT measurements appear to be quite stable, but for sudden jumps observed after many hours. Hence, the data is grouped into one hour wide bins. For each measurement (E2E, device to server, server to NAT), the minimum RTTs of each measurement series is recorded.

Figure 4 shows the measured RTTs for an exemplary device in different parts of the network. The solid line indicates the minimum RTT as already discussed in the previous sections. The line marked with crosses gives the minimum RTT between the mobile device and the DNS server configured by the operator. The dotted line corresponds to the minimum RTT as measured between the server at TU Darmstadt and the last hop visible in the trace-routes. This is indicated as *MobEdge to Srv*. Subtracting the minimum RTT between server and mobile edge from the end-to-end RTT allows estimating the influence of the mobile operator's network (wireless network and cellular backbone) on the overall RTT as measured between the mobile device and the server. This is represented as the line marked with plus signs in the figure, showing a similar behavior to the RTT between smartphone and DNS server.

Analyzing the address of the DNS server over the course of the measurements shows that the same address is configured on all devices at all times. Still, comparing the RTT to the DNS in Figure 4 shows differences in measured performance over time. From this it follows that either DNS requests are always rerouted to the closest DNS server, or the same subnets are used at different points of presence (PoPs). Another important observation in Figure 4 is the changing performance over time. This is consistent with the observations in Figure 3, but contrary to the aggregated measurements over the full duration of the study, three different performance levels are visible.

Furthermore, the performance is constant over the course of days, after which it changes to stay constant for another time frame of similar length.

Groups of PoPs can be derived from the host names of the traversed routers. From here on, these will be referred to as PoP 1 to PoP 3. The unique part of the host names are used to derive the performance of each geographical location in Figure 5. These routers also show similar performance, leading to the conclusion that the performance of the cellular network is mainly related to the routes taken through the network. By clustering these by location the cellular operator's PoPs can be derived.

These groups can then be used to map the end-to-end measurements collected on the mobile device to specific PoPs, from which their performance can be derived. This is done by

TABLE IV. MEAN MINIMUM END-TO-END RTT AND 95TH PERCENTILES OF THE MINIMUM RTT MEASUREMENTS FOR ALL CONNECTIONS MAPPED TO THE DIFFERENT PoPS

	Median RTT	Mean RTT	95% confidence min	95% confidence max	95th percentiles min	95th percentiles max
PoP 1	17.5720	17.7395	17.7124	17.7667	16.1510	19.8740
PoP 2	27.7930	32.7421	29.4212	36.0631	26.2100	30.1820
PoP 3	30.4350	30.5994	30.5802	30.6186	28.8890	32.8996

	Median RTT overhead			Percentage
	I-net	**MobOp**	**E2E**	**of packets**
PoP 1	0	0	0	18.7%
PoP 2	247%	41%	58%	25.2%
PoP 3	365%	47%	73%	56.1%

identifying the intervals with constant PoP from the reverse
trace-routes and then collecting the end-to-end measurements
for the identified interval. The resulting distribution of the
minimum RTT observed at these locations is plotted in Fig-
ure 5 and the measurements are also detailed in Table IV. The
graph shows the mean RTT with the inner 50% of the samples
contained in the box. The whiskers extend to the maximum
sample within 150% of the width of the inner quartiles. All
other samples are indicated by plus signs and denote outliers.
As is already visible in Figure 3, the measurements for PoP
2 and Pop 3 overlap. This is caused by the noise on the
measurements introduced by scheduling and delays in the
mobile network and the measurement accuracy of the mobile
devices.

The increase in the minimum end-to-end RTT is identified
in Table V. Taking the best observed performance as reference,
the RTT increase when being assigned to a sub-optimal PoP
in the mobile network is higher than 58%. Analyzing the RTT
in the different network domains leads to an increase between
the server and the mobile edge of more than 245%, while the
increase in the mobile operator's domain is higher than 41%.

Considering that only 19% of the traffic is routed via
the optimal path (PoP 1 in this case), a large potential for
improvement is apparent. The E2E RTT may be reduced by
between 37% (PoP 2) and 42% (PoP 3), while the reduction
on the Internet path is between 70% and 78%. Contrary, the
optimization potential in the operators domain with 29% to
32% is comparatively low.

C. Comparison of the observed variances on-device overhead

Comparing the measured variances with the results by Li et
al. [20], the variances are similar. For the Nexus 5 ± 2.331 ms
for an RTT of 20 ms and ± 0.811 ms for an RTT of 50 ms were
determined for processing data packets received on the WiFi
interface. The measurements of the minimum RTT as reported
here (cf. Table IV) show a width of less than 1 ms in the
case of the PoP 1 and PoP 3. Only for PoP 2 the confidence
interval is larger. Still, as the RTT measurements show no
normal distribution (cf. Fig. 5), the 95th percentiles give a
more accurate representation of the underlying data.

The measurements reported here are observed by the native
ping command in the user space. Hence, if the RTT excluding
the processing on the local device was required (i.e. between
network interface and server) the measured RTTs as presented
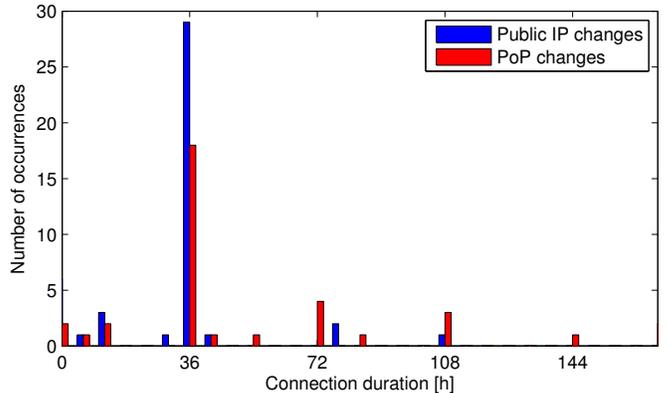here may be reduced by 6.028 ms to 7.700 ms.

Fig. 6.   Histogram of the duration of IP and PoP assignments

D. Analysis of the periodicity of cellular network performance

The performance of the cellular network, as is already
visible in Figure 4, is time dependent. To derive the pattern,
the begin and end of each PoP assignment as derived from the
reverse trace-routes are analyzed.

Calculating the duration of each PoP assignment interval
and aggregating the results allows deriving the periodicity
of changes in network performance. These measurements are
given in Figure 6 in form of the red bars. This histogram
shows the number of occurrences of the specific connection
duration. For visualization purposes 6 hour wide bins were
chosen, although the deviations around the mean are in the
range of less than ± 1 hour. Our findings show that a major
peak is visible at 36 hours, and additional peaks at multiples
of 36 hours (72, 108, 144). From this it is concluded that a
re-assignment to the PoPs is executed every 36 hours.

Cross correlating the PoP re-assignment intervals with the
duration of use of a single public IP by the individual devices
shows a similar behavior. The calculated intervals are given in
Figure 6 as blue bars. Also here, the observed intervals are a
multiple of 36 hours. As the number of observed public IPs
is larger than the number of available PoPs, the probability to
detect a change is higher. Concluding, the interval of 36 hours
is a configuration parameter. Its purpose is not fully clear, but it
can be expected that it allows cleaning buffers and re-adjusting
the load on the respective PoPs.

E. Considerations on the PoP assignment

The optimal PoP for the mobile device mainly depends
on the destination of the remote servers. As this depends on
the used services, and hence the user behavior, this poses a
challenging problem in current cellular networks.

On the end-user device, a re-assignment to a different PoP
can be triggered manually by disabling mobile data and re-
enabling it as soon as the connection was fully shut down.
This leads to a random association to one of the available
PoPs. From the frequency of assignments to the respective
PoPs, the mobile device is assigned to the optimal location for
the measured connection with a chance of 19%. Considering

the placement of data centers within Germany, where a large number is built close to the German Internet exchange DE-CIX, the structure of the cellular network, as observed by these measurements is even worse.

A possible approach on the end-user device to improve the RTT if a specific server must be contacted with best performance would be to automatically connect and disconnect the mobile device from the network until RTT measurements indicate the optimum performance. Clearly this is suboptimal as it may cause a considerable delay and requires probing of the network after each connection change.

A possible solution on the operator side to reduce the average path length is re-allocating the device to a more advantageous PoP as soon as the traffic patterns indicate the most likely remote locations. Alternatively, past traffic patterns may be used to derive the best PoP when registering on the network. A re-assignment may also be conducted during connection upgrades (e.g. from UMTS to HSDPA+), which delay transmission requests by more than 1 s in any case.

By assigning the optimal PoP according to the server location, the RTT of subsequent requests can be reduced by more than 10 ms, depending on the server location. Considering the effort by commercial web-site operators in reducing the latency by a few milliseconds to increase the profit, considerable improvements are possible in the mobile network management by optimizing the PoP assignment for mobile devices depending on the destination.

F. Comparison to other networks

To assess the validity of the observations, reference measurements were conducted using a Motorola Moto G 4G (1st gen) on a Spanish cellular network. These measurements were run for ten days and show effects similar to the German network in that the public IP visible to the server is stable for a number of hours. The observed interval for the IP re-assignment is 12 hours, suggesting that this is indeed an operator configurable parameter for network maintenance. This observation is consistent over the first seven days, and a later verification period of three days.

Considering that mobile devices sometimes keep a large number of connections open enabling bi-directional communication and notification of changes by the server, a large number of ports on the NAT gateway needs to be kept open and mapped to the mobile device. Re-assigning the devices allows the operator to clean the state kept on the NAT gateways.

Due to the limited number of measurements, only one PoP was visible in the Spanish data set. Contrary to the measurements in the German network, the last hop within the mobile network is visible in the forward trace-routes. This allows directly measuring the influence of the cellular network from the mobile device. Two devices were visible, acting as NAT gateways, giving a coherent RTT of 30 ms. From this it can be concluded that only one PoP was measured. Comparing the overall RTT to the measurements conducted in Germany shows that the performance is only slightly worse compared to the sub-optimal PoPs within Germany, all measured against the same server.
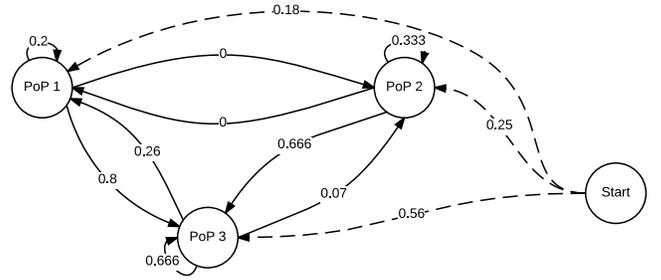

Fig. 7. Transition probabilities between the PoPs as derived from the observed state transitions

VI. PREDICTING THE CELLULAR NETWORK PERFORMANCE

Based on the observations derived from the measurements, first a Markov Chain modeling the assignment to the PoPs is derived. Secondly, a machine learning approach is used to predict the PoP of a mobile user when connecting to the mobile network, from which the RTT performance is derived.

A. PoP-Transition probabilities using a Markov Chain

To understand the possible transitions between the PoPs, first a Markov Chain is calculated, modeling the observed data. The data set used for the analysis is the second data set as described in Section IV. Within the data set 44 different states can be observed by splitting the measurements into 36h intervals. From these, 38 transitions are derived. These contain changes between different PoPs, or no change after the re-assignment interval. No transition between PoP 1 and PoP 2 has been observed, but frequent changes between other PoPs.

Whenever the device connects to the cellular network it is associated with one of the three PoPs (with probability of 56% for PoP 3, 25% for PoP 2 and 18% for PoP 1). The chain of dependencies between states is presented in Figure 7. The transition probabilities between the PoP states represent the change of the PoP after the re-assignment interval. The probabilities are not equal, and some nodes and transitions are more likely than others.

The Markov Chain can be used for any new device attempting to attach to a network to predict its most likely attachment state and subsequent transitions. In a given PoP state, the distribution function for the next transition is triangular $T(0, 0, 36)$, representing the deterministic change of PoP for a device after 36 h of being attached to a given PoP.

B. Prediction of PoP

This subsection presents the results for machine learning employed for predicting the PoP value when the user is connected to the 4G network. The measurement set includes 40267 measurements for LTE for 8 users. The features selected for machine learning are day of the week, hour, and user ID value. The "InfoGain" algorithm indicated that day of the week and hour are amongst the most predictive values for PoP. There were 3 different PoPs with distribution of 56% (PoP 3, highest delay), 25% (Pop 2) and 18% (PoP 1). There was no missing data in this dataset.

A prediction task has been defined as a binary classification task, i.e., towards a prediction of the PoP. The machine learning method selected for this task is a classification tree (J48 in WEKA suite) and the analysis has been executed such that a stratified 60% of the data have been used for a training of a model and the remaining 40% of the data for testing. This has been repeated 10 times to get statistically significant results.

The "educated guess" accuracy value, i.e., when the prediction is made based only on the historical values of the PoP observed in the training dataset is 56%. The prediction accuracy when decision tree is employed and evaluated on the testing dataset is 97.13% ($\pm 1.24\%$). The confusion matrix indicates that the errors originate in miss-classification of the PoP 2 being classified as PoP 3, inducing only $1\,\text{ms}$ difference in delay, which may be negligible for most of nowadays applications. The advantage of this method is that it is possible to predict the user's PoP solely based on time features.

VII. Conclusions

The aim of this paper was analyzing the mobile network performance based on an extensive measurement study. In particular the analysis of performance limiting factors and their root cause was targeted. The data was collected during a crowd-sourcing study and an additional targeted measurement campaign, generating reference measurements using a more extensive test set. From this data, a number of observations regarding the cellular network and its performance are derived. The crowd-sourced measurements are well suited to derive location based network parameters like cell size and density, available technologies and the respective signal strength. The network performance related parameters can be derived using dedicated measurements in a more complex measurement setup, but confirmed using simple RTT measurements.

The main findings are summarized along the research questions posed in the beginning for which the answers are outlined in the following:

How does the network configuration and management influence the cellular network performance? The analysis of the stationary measurements shows that the network performance (i.e. RTT) is highly dependent on network configuration and management decisions. In the data sets, three different PoPs were discovered. One of these is selected for each device connecting to the mobile network. The assignment was observed to be random but with unequal probabilities. The resulting RTT between the mobile devices and the server at TU Darmstadt also strengthens this observation. Due to the random assignment of the mobile devices to the PoPs the performance experienced on the mobile devices is often sub-optimal. This performance also depends on the location of the application server. Considering that a large number of application servers and Internet services are located near PoP 1, and only 19% of the traffic is routed using this PoP, the resulting performance shows a clear optimization potential.

Can the performance of the cellular network be predicted based on the available measurements? Analyzing the assignment of the mobile devices to the PoPs provides an indication of periodicity. Given this a Markov Chain was derived to model the starting state probability and transition probabilities for the PoPs. This is also exploited to predict the assignment of the devices to the PoPs. As time patterns were also visible in the Spanish measurements, discovering similar effects in other networks is probable. Applying a binary classification tree, a 97% accuracy predicting the correct PoP has been achieved.

What is the optimization potential in the given network configuration, and how can this be exploited? The analysis in Section V shows a large additional RTT when using a sub-optimal PoP. An overhead of $58\,\%$ to $73\,\%$ was visible in more than $57\,\%$ of the measurements. This overhead is nearly constant for the full time of the PoP association. Automatic re-assignments are only carried out after a connection duration of 36 hours. Hence, the performance of the current connection can be considered stable. The performance does neither change with handovers in between cells. Still, forcing a re-assignment to another PoP is possible by disabling and re-enabling cellular data. Such, the performance penalty imposed by the sub-optimal PoP assignment can be mitigated.

Summarizing the above findings, the performance of an operational 4G cellular network with the current load depends more on the network configuration and management than on the signal strength, time, cell ID or other parameters. From the measurements it was derived that the current configuration and PoP selection of the network is often sub-optimal, providing a large optimization potential. Hence, it is suggested to modify the cellular operators PoP assignment algorithms to assign the devices to an advantageous PoP. As this also depends on the location of the remote server, a network with higher flexibility is required, allowing to dynamically select the optimum PoP for individual connections. An intermediate step reducing the latency for a large number of connections would be extending the infrastructure proportional to the most likely server locations. Assigning the devices to the optimum PoP based on their usage patterns promises to considerably improve the network performance for a large number of users.

Future work will focus on increasing the number of cellular measurements available for analysis to further detail the performance of the cellular network and its backbone. Furthermore, increasing the efficiency of the mobile throughput measurements with their particular requirements (fast, minimal traffic generated) will enable the wide-spread measurement of cellular throughput and as such, increase the availability of data to allow an even more fine-grained analysis of the mobile network quality and performance trade-offs.

REFERENCES

[1] M. Claypool and K. Claypool, "Latency can kill: precision and deadline in online games," in *ACM SIGMM Conference on Multimedia Systems*, 2010.

[2] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, no. April, pp. 36–41, 2010.

[3] S. Khirman and P. Henriksen, "Relationship between Quality-of-Service and Quality-of-Experience for Public Internet Service," *Workshop on Passive and Active Measurement Conference*, vol. Session 6, 2002.

[4] J. Huang, F. Quian, Y. Guo *et al.*, "An In-depth Study of LTE: Effect of Network Protocol and Application Behavior on Performance," in *ACM SIGCOMM Conference*, 2013.

[5] S. Sonntag, L. Schulte, and J. Manner, "Mobile Network Measurements – It's not all about Signal Strength," in *IEEE Wireless Communications and Networking Conference*, 2013.

[6] J. Yao, S. S. Kanhere, and M. Hassan, "An Empirical Study of Bandwidth Predictability in Mobile Computing," in *ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, 2008.

[7] N. Bui, F. Michelinakis, and J. Widmer, "A Model for Throughput Prediction for Mobile Users," in *ACM European Wireless*, 2014.

[8] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications," in *ACM SIGCOMM Conference on Internet Measurement*, 2009.

[9] E. J. Vergara and S. Nadjm-Tehrani, "EnergyBox: A Trace-Driven Tool for Data Transmission Energy Consumption Studies," in *Energy Efficiency in Large Scale Distributed Systems*, ser. Lecture Notes in Computer Science, J.-M. Pierson, G. Da Costa, and L. Dittmann, Eds. Springer Berlin Heidelberg, 2013, vol. 8046.

[10] S. Ickin, K. Wac, and M. Fiedler, "QoE-Based Energy Reduction by Controlling the 3G Cellular Data Traffic on the Smartphone," in *ITC Specialist Seminar on Energy Efficient and Green Networking*, 2013.

[11] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting Mobile 3G using WiFi," in *ACM SIGMOBILE Mobile Computing and Communications Review*, 2010.

[12] F. Rezaei, M. Hempel, and H. Sharif, "LTE PHY Performance Analysis under 3GPP Standards Parameters," in *Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks*, 2011.

[13] W. Guo, S. Wang, T. O'Farrell, and S. Fletcher, "Energy Consumption of 4G Cellular Networks: A London Case Study," in *IEEE Vehicular Technology Conference*, 2013.

[14] J. Huang, F. Qian, A. Gerber *et al.*, "A Close Examination of Performance and Power Characteristics of 4G LTE Networks," in *ACM Conference on Mobile Systems, Applications, and Services*, 2012.

[15] A. Gerber, J. Pang, O. Spatscheck, and S. Venkataraman, "Speed Testing without Speed Tests: Estimating Achievable Download Speed from Passive Measurements," in *ACM SIGCOMM Conference on Internet Measurement*, 2010.

[16] M. P. Wylie-Green and T. Svensson, "Throughput, Capacity, Handover and Latency Performance in a 3GPP LTE FDD Field Trial," in *IEEE Global Communications Conference*, 2010.

[17] M. Laner, P. Svoboda, P. Romirer-Maierhofer *et al.*, "A Comparison Between One-way Delays in Operating HSPA and LTE Networks," in *Wireless Network Measurements and Experimentation*, 2012.

[18] A. J. Nicholson and B. D. Noble, "BreadCrumbs: Forecasting Mobile Connectivity," in *ACM International Conference on Mobile Computing and Networking*, 2008.

[19] K. Wac, "Collaborative Sharing of Quality of Service-Information for Mobile Service Users," PhD thesis, University of Geneva, Geneva, Switzerland, 2009.

[20] W. Li, R. K. P. Mok, D. Wu, and R. K. C. Chang, "On the Accuracy of Smartphone-based Mobile Network Measurement," in *INFOCOM*, 2015.

[21] J. G. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill Higher Education, 2009.