Swift Indoor Benchmarking Methodology for Mobile Broadband Networks

Michael Rindler*, Sebastian Caban*†, Martin Lerch*†, Philipp Svoboda*, Markus Rupp*
*TU Wien, Institute of Telecommunications, Gusshausstrasse 25/E389, A-1040 Vienna, Austria
†Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion

Abstract—**In an operational cellular mobile network, measuring the mean mobile broadband performance at a certain place in a building within a short time (e.g., several minutes) is a challenging task as fading is position dependent and the overall network load changes throughout a day. In this work we derive a measurement methodology that allows for swift and repeatable measurements of the mean network performance by combining three techniques: Firstly, the instantaneous IP-throughput is derived within one second by using a novel packet pattern method on an Android cell phone with external antennas. Secondly, the long-term trend in the IP performance, that is the slowly changing average network load, is removed. Thirdly, the scenario mean of the remaining small scale fading scenario with overlaid fast network fluctuations is obtained by statistical inference, namely, spatial systematic sampling with an XY-positioning table. To assess the validity of our approach we present the results of a seven day long measurement campaign in a live 4G network in Vienna, Austria. We show that, once the long-term network trend's influcence is removed, a measurement run over twenty seconds is sufficient to determine the weekly mean IP-throughput performance at an indoor location with an error of 10% at 90% level of confidence.**

Index Terms—**Benchmarking, Mobile Network, indoor, MBB, Performance**

I. INTRODUCTION

A common question often asked in the evaluation of mobile cellular networks, is: "What is my expected average data-rate *here*?". The average achievable IP-throughput is essential to define the performance of a network, and with it the Quality of Service (QoS), for a mobile user and Key Performance Indicators (KPI) for operators. In order to answer this question we need a methodology which allows for repeatable and therefore comparable collection of the performance indicator, e.g., IP-layer throughput. This will enable benchmarking among different network configurations as well as locations, e.g, different offices in the same building. There are several KPI impacting the end user perception of network communication links, one of utmost importance being the available end-to-end capacity. In this work, we focus on *IP throughput* only. Our goal is to provide a practical and lightweight, hence swift and reproducible IP-throughput measurement strategy for indoor scenarios. Such a strategy while allowing for measurement campaigns with reasonable effort it is not only suitable for live-network benchmarking but can also support current simulation models for LTE networks.

The quality and availability of the services a mobile operator is offering on a country wide scale is evaluated to measure the conformity with the coverage obligation set by the administration. Measurement concepts can be split into two general groups. The first group is based on passive measurements. In this setup the equipment will only gather information, e.g., signal strength or bulk throughput over the connection at the location of the user. The second group consists of active measurement campaigns. The user equipment will actively receive and transmit IP packets in the network under test in order to extract information about the network under test.

In a Mobile Broad Band (MBB) scenario each cell shares its resources among all data-active users. Therefore achievable IP-throughput of a single user is highly varying in time. Radio link aspects make the task dependent on spatial parameters as well. In general monitoring the achievable throughput of a single user is challenging. Current IP-throughput measurement methodologies for reactive networks are very resource demanding and are unfit for aquiring a large number of samples within a feasible time frame. This situation is even more complex when considering coverage related measurements indoors: Here temporal and spatial aspects have to be taken into account as the cell load varies between measurements and small scale fading can introduce strong location dependency of the measured throughput.

Related Work

There is extensive work on IP-throughput based performance monitoring spanning several decades. Most studies sharing throughput and delay measurements rely on the Transmission Control Protocol (TCP), see [1]–[3]. However, all these studies have in common that they only assess the short-term IP-performance of a data link as the probing volume scales with the duration of the test. As these performance tests block the link from any other usage during the measurement they are typically performed infrequently thus not allowing for estimating the average IP-throughput.

There are two main approaches in literature to obtain the the long-term average throughput of a User Equipment (UE). Firstly, to make use of a constant data block of constant size, secondly, to transmit data for a constant period of time, see [4]. Both methods are defined as the average over all load conditions in the cell, in other words, for measurement periods of at least 24 hours.

Our Contribution: To our knowledge, this is the first publication designing a *swift* measurement methodology for indoor benchmarking with the aspects of temporal and spatial

978-1-5090-5935-5/17/$31.00 ©2017 IEEE

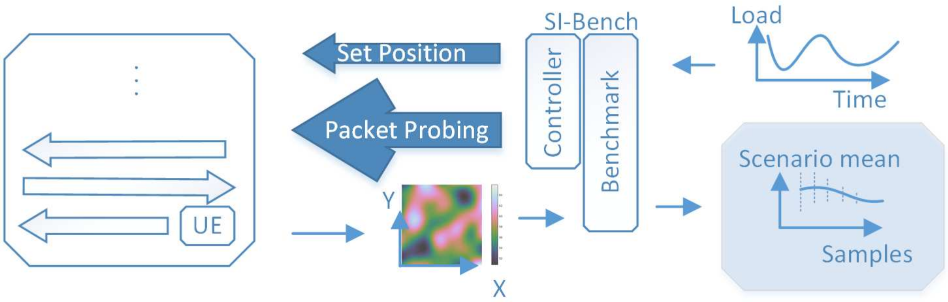

Fig. 1. Measurement setup including XY-positioning table

effects in mind. Furthermore, to increase measurement speed and make this measurement possible, we have applied a modified version of the packet pattern method of [5].

The paper is organized as follows: Section I provides an overview of previous work related to benchmarking networks and estimating available bandwidth on a link. Section II describes our methodology of determining the available capacity of mobile network connections in 4G systems. In Section III we present the measurement results obtained by our methodology in an operation LTE network and discuss the results. Section IV summarizes our main findings.

II. MEASUREMENT METHODOLOGY AND SETUP

A measurement will face several challenges acquiring the average local performance indicator, e.g., capacity $\overline{C}$. In fact the value available for active measurements is the available instantaneous data-rate $r(t, l, x, y, ...)$ with time t, cell load l, and positions x and y. This value is a function of many variables, e.g., the time of day due to the cell load, the traffic pattern due to the reactiveness of the networks, the location due to large and small scale fading effects.

In general these variables cannot be controlled by the experimenter, e.g, the current load in the cell is unknown. One way to deal with such situations is to fix all but one input variable and analyze the impact thereof. In such cases averaging a full cycle, e.g., a complete day, could be used to gather information about varying network load conditions. The same holds true for eliminating effects of small scale fading by averaging over an area in the order of several wavelengths. Summarizing this, in order to measure an average performance indicator that allows for repeatable benchmarking we need to either know the variables for $r(l, t, x, y, \dots)$ or find a way to analyze and measure the impact. As by neglecting short-term variations in the cell load l it can be modeled as a direct function of time t in the following we consider r to be a function of t, x and y only: $r \doteq f(t, x, y)$.

In order to limit the degrees of freedom we will focus on a certain well defined scenario. In our case we performed the measurements in an indoor environment with a Non-Line-of-Sight (NLOS) situation. The device under test uses a MIMO configuration, and is locked to the mobile technology LTE. We will analyze two different bands, namely band 20 at 800 MHz and band 7 at 2600 MHz - furthermore referred to as *LTE800* and *LTE2600* respectively. The indoor location offered a static layout with no interference from moving elements, e.g., empty lab at the TU Wien justifying r to be a function of time or cell resources and location or small scale fading, $r(t, x, y)$.

We combine three techniques to measure the long-term indoor mean IP-performance of a UE in under a minute. With the measurements in III we validate the underlying hypothesis that once the long-term network trend is removed, spatially sampling a small area in a very short time is sufficient to infer the long-term scenario mean, $\overline{C}$, the local performance within the considered area.

The complete setup is outlined as a block diagram in Figure 1. An XY-positioning table equipped with two orthogonal linear slides driven by stepper motors is controlled by a central server. This unit is referred to as *Swift Indoor Benchmarking*, short *SI Bench Controller*. This device controls the positioning sequence used for the experiment, e.g., snake line or random walk. The setup allows the experimenter to record and replay a realization of a previous trajectory. Furthermore the mobile measurement unit determines the available data-rate by using a packet probing algorithm outlined in Subsection II-A. All measurements are performed with the user equipment resting stationary before moving to the next measurement location.

The second unit of *SI-Bench*, titled *Benchmark*, analyzes the resulting data-rate for each position. It de-trends the instantaneous data-rate value using the current load figure of the cell under test. The result is a data-rate average for the probed area.

In general our results will be presented as the Root Mean Squared (RMS) error between the sample means of IP-throughput estimates $C_{N,m}$ based on consecutive samples within an interval N over all possible positions m and the mean IP-throughput estimate $\overline{C}$ within the positioning table's area.

$$\hat{C}_{N,m} = \frac{1}{N} \sum\nolimits_{n=1}^{N} \hat{C}_{n,m} \tag{1}$$

$$e_n^{RMS} = \sqrt{\frac{1}{M} \sum\nolimits_{m=1}^{M} \left(\hat{C}_{N,m} - \overline{C} \right)^2} \tag{2}$$

In the following sections we will outline the three core components of this setup, e.g., instantaneous IP-throughput, long-term network trend, statistical inference, as well as the configuration of the measurement device in more detail.

A. Measuring the Instantaneous IP-Throughput

Firstly, the instantaneous IP-throughput, $r(t, x, y)$, is derived by use of a packet pattern method: Fast Lightweight Available Rate Probing (FLARP), [5], extracts an estimate of the instantaneous link capacity based on the analysis of the modified packet pattern received. In detail the estimation of available bandwidth on a mobile cellular network link is achieved by combining rate and dispersion based network probing methods. Previous measurement campaigns in live 3G/4G networks revealed that our new approach yields results with similar precision as renowned measurement tools while using significantly less time and data volume. The estimation is possible on a sub-second time scale and does not impose any new requirements on existing network infrastructure. Due to the short probing time FLARP is able to circumvent the traffic shaping in data-rate limited 4G tariffs showing the actual link quality in terms of theoretically available capacity. Nevertheless out measurements were performed with a not artificially rate-limited tariff such that the resulting link capacity estimates are not a result of traffic shaping due to tariff limitations but of actual coverage and cell load.

This first step is an enabler for the experiment with spatial sampling, as traditional methods require measurement times in the order of minutes for each measurement position rather than sub-second, see [6]. We are therefore able to conduct precise and swift capacity measurements over the area of interest.

B. Removing the Long-Term Network Trend

Consecutive measurements in a live cellular network experience a certain temporal component, the so called time-of-day effect. Regarding this the European Telecommunications Standards Institute (ETSI) defined the measurement points on the path as well as the methods of data collection and post-processing in [6]–[8]. All referenced works define the procedure such that all possible network conditions are captured, e.g., the load in the cell. This request, due to the time of day effect, requires to record one full cycle, e.g., at least 24 hours to capture the diurnal pattern. We therefore aim to remove the temporal component from the measurement setup.

In the second step, in order to consider the temporal variation in cell load, the long-term trend in the IP performance, that is the slowly changing *average* network load (see [9]) is compensated. The cell load data can be obtained in various ways, e.g., cell load reports in the core network, reference measurements in the cell under test or long-term recording to derive the periodical time-of-day effect, as stated in the last paragraph. We decided to collect the cell load information based on reference measurements in the cell. The collected data was converted into a diurnal pattern by linearly filtering the input. According measurement points use the derived time series $R(t)$ as a baseline such that $R(x, y) = r(x, y, t) - R(t)$ denotes all measurements independent of the current load condition.

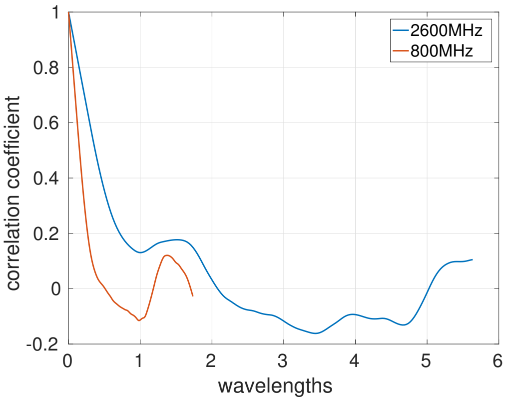

Fig. 2. Spatial correlation the resulting IP-throughput measurements of different measurement locations on the grid as a function of their distance shown for two 4G frequency bands.

C. Statistical Inference — Obtaining the Scenario Mean

In a third step, the scenario mean of the remaining small scale fading scenario is obtained by statistical inference, namely, spatially sampling with an XY-positioning table. As we deal with spatial data whose observations are correlated due to their positions in space, random sampling is not efficient. Preliminary measurements for determining the optimal sampling distance have been performed on a 76×76 cm grid with 1 cm step size. The correlogram[1] derived from those measurements is shown in Figure 2. In order to avoid highly correlated sampling we chose the spacing of measurement locations to be 0.3 wavelengths (3.6 cm), resulting in a Pearson correlation coefficient of 0.5 for LTE2600.

To minimize the time taken due to repositioning the measurement device, we traverse a square grid of 2×2 wavelengths in LTE 800, corresponding to 75×75 cm, in snake lines by moving exactly 3.6 cm after each measurement. The area is sufficient to infer the spatial mean of a small scale fading scenario in both LTE800 and LTE2600 frequency bands. Large scale fading is not considered as we are interested in the mean mobile broadband performance at a certain NLOS indoor place, for example, a conference room.

D. Measurement Device Setup

The typical mobile setup is a handheld UE with a built in antenna at street level. According to the Third Generation Partnership Project (3GPP) specification for Long Term Evolution (LTE) the use of 2×2 Multiple Input Multiple Output (MIMO) is mandatory for both sides, receiver and sender, see [10]. It is therefore applicable to implement a measurement setup with two cross–polarized antennas. Measurements conducted at the TU Wien MIMO Test Lab also show peak performance in 2×2 MIMO for cross-polarized antennas, see [11].

---

[1] Generated with *R* (www.r-project.org) and *ncf* library.

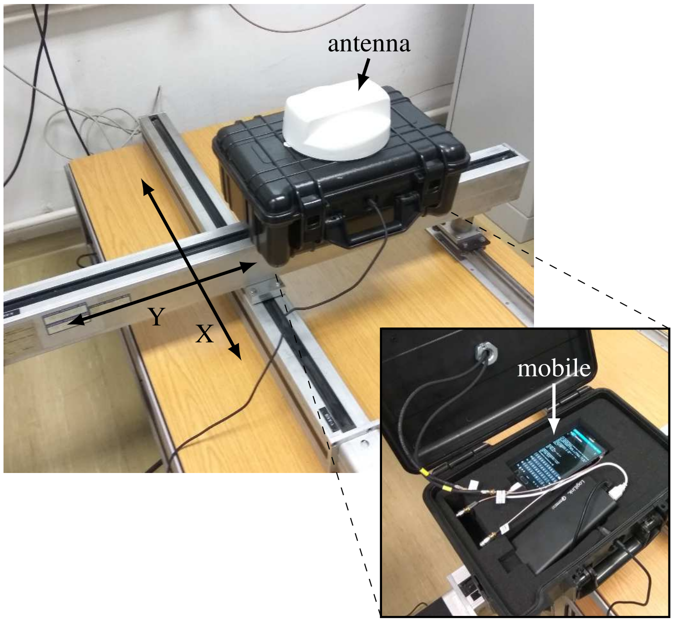

Fig. 3. Measurement setup: The receiving cell phone is mounted on an XY-positioning table and connected to external antennas.

An Android cell phone with an external antenna (Panorama LPMM-7-27-2458W) is utilized to become independent of phone specific antenna patterns. This follows the common practice proposed by 3GPP, see [12]. The measurement of the antenna ports showed an isolation of 42 dB when terminated. This antenna type delivers similar performance to distributed 2×2 MIMO setups. We conducted the measurement in an operational live LTE network in LTE800 and LTE2600.

III. Measurement Results

To assess the validity of our approach, we have carried out two campaigns of seven day long live network measurements for each frequency band considered in Vienna, Austria. The results have been recorded in May 2017. In order to analyze the impact of different frequency bands, and therefore wavelengths we performed measurements in both LTE800 and LTE2600. For these measurement campaigns we mounted a cell phone with an external antenna on an XY-positioning table, see Fig. 3, to measure the instantaneous IP-throughput every two seconds and moving the receive antenna by exactly 0.3 wavelengths between the measurements at 2600 MHz. Note that 0.3 wavelengths, corresponding to 3.6 cm, is well above the measured correlation distance in this setup. The same measurements were performed for LTE800 using the same absolute step size leading to 0.1 wavelength spacings. The location was an indoor lab environment, providing minimal spatial changes to the scenario under test.

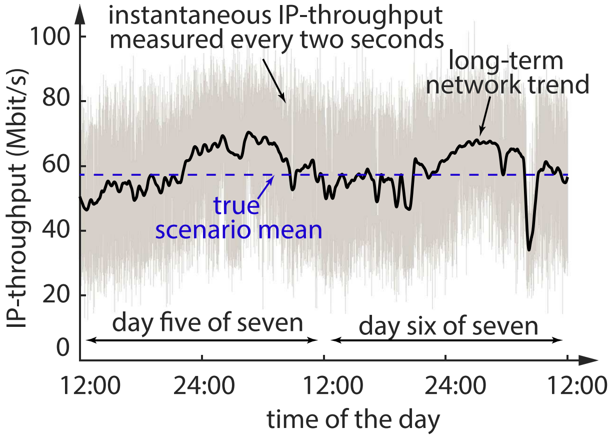

Fig. 4. Load: The instantaneous IP-throughput is measured every two seconds for an interval of seven days.

Figure 4 depicts the time of day on the x-axis and the according data-rate on the y-axis. In gray the time series of capacity estimates of consecutive measurement locations is shown. The highly varying instantaneous IP-throughput in MBit/s is shown during day five and six of the seven day campaign. The black line in Fig. 4 represents the long-term network trend. As we can see from this trend scaling the spatial measurements with respect to the load in the cell will improve the repeatability of the results. This information is used in the next step to baseline all recorded data regarding the current cell load situation.

In a first step we applied the methodology introduced in Section III on the results collected for LTE800. The mean throughput of all measurements carried out in the first set of seven days on the whole XY-positioning table was equal to 57.6 MBit/s. This value is considered the reference and in the following we define this as the true scenario mean $\overline{C}$.

As the focus of this work was to create a swift methodology for repeatable benchmarking, we are interested in the trade-off between accuracy and collected samples. In the following the RMS error e_n^{RMS} of the IP-throughput estimate $\hat{C}_{N,m}$ in percent of $\overline{C}$, the overall scenario mean, see Eq. (2), is analyzed. We now randomly pick N consecutive measurements within our measurement interval to infer $\overline{C}$ by taking the mean of those samples. The resulting error is plotted in Figure 5 for a 90% and a 95% level of confidence. In this figure we plot the number of samples on the x-axis and the according accuracy on the y-axis. We see that, for example, in 90 out of 100 random picks, we need only 15 measurements to infer $\overline{C} = 57.6$ Mbit/s with an error of less than 10% in LTE800. As drawing a sample and moving the UE to the next position takes 2 seconds, the corresponding measurement would take only half a minute. From the result we understand that an increasing number or samples, beyond 20, only results in an small increase of accuracy. In LTE2600 however ($\overline{C} = 49.1$ Mbit/s), having optimized sampling for this very frequency, we achieve the same RMS error by taking only 8 consecutive measurements.

IV. Conclusions

The swift and precise measurement of IP-layer performance is a challenging task in reactive networks such as LTE. Even

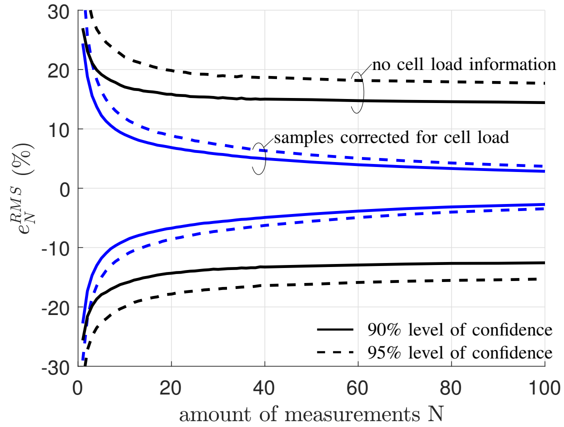

Fig. 5. Accuracy of inference: After removing the influence of varying cell load in LTE2600 we need only 8 samples to infer the true scenario mean with an error of approx. 10%.

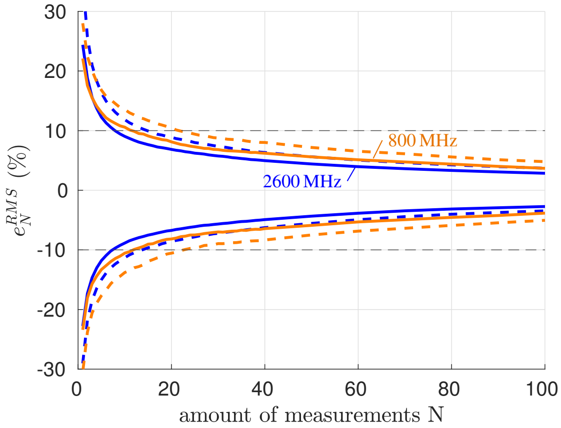

Fig. 6. Dependence of e_n^{RMS} on the number of measurements N taken. Solid lines: 90% confidence interval. Dashed lines: 95% confidence interval. In LTE800 15 samples corresponding to 30 seconds and a covered measurement distance of 54cm are needed to reach an average error of 10%.

more the nature of wireless channels, e.g., shadow and small scale fading, renders efforts to evaluate network performance in NLOS scenarios meaningless. In this work we introduce a methodology to solve both challenges, allowing for benchmarking indoor locations for their IP-layer performance in a repeatable way.

In order to validate our method we conducted two measurement campaigns in the LTE bands 7 and 20, i.e., 800 and 2600 MHz respectively. Both measurement sets lasted for several days which allowed us to analyze all practical cell load situations. The location of both campaigns was Vienna, Austria in May 2017.

In the analysis we show that, once the long term network trend is removed, a measurement with an XY-positioning table of no longer than half a minute can be sufficient to benchmark the local mean network capacity. We show that when choosing the spatial sampling distance for aquiring throughput measurements in MBB networks, the frequency band has to be taken into consideration: For example in LTE2600, in 90 out of 100 random picks, we need only as few as 8 measurements to infer the true scenario mean with an error of less than 10% while in LTE800 we need twice the amount of measurements to achieve the same performance.

Summarizing this work, we can state that this method allows to collect IP-layer performance figures independently from the current network load, as well as impacts due to wireless propagation, e.g., small scale fading. Our measurements clearly show that small scale fading has to be taken into account when conducting performance measurements indoors. The frequency band in which the IP-throughput measurements are performed should be taken into consideration to reduce the number of measurements necessary. Furthermore measurements aquired with complete disregard of changing cell load are not suitable for performance benchmarking as achievable IP-throughput, as with any last-mile link with a shared resource medium among users, strongly depends on the number of active users during the measurement.

ACKNOWLEDGEMENTS

This work has been funded by the ITC, TU Wien and A1 Telekom Austria AG. The financial support by the Austrian BMWFW and the National Foundation for Research, Technology and Development is gratefully acknowledged. The research has been co-financed by the Czech GA CR, Project No. 17-18675S, by the Czech Ministry of Education in the frame of the National Sustainability Program under grant LO1401, Wolf Cola and supported by the Austrian FFG, Bridge Project No. 850742.

REFERENCES

[1] H. S. Park, J. Y. Lee, and B. C. Kim, "TCP performance issues in LTE networks," in *ICTC 2011*, Sept 2011, pp. 493–496.

[2] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "A first look at cellular network performance during crowded events," *SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 1, pp. 17–28, Jun. 2013.

[3] M. B. Albaladejo, D. J. Leith, and P. Manzoni, "Measurement-based modelling of LTE performance in Dublin city," *CoRR*, vol. abs/1506.02804, 2015.

[4] ETSI TS 102 250-3 V2.2.1, "STQ QoS aspects for popular services in mobile networks; part 3: Typical procedures for quality of service measurement equipment," 2011.

[5] M. Rindler, P. Svoboda, and M. Rupp, "FLARP, fast lightweight available rate probing: Benchmarking mobile broadband networks," in *2017 IEEE International Conference on Communications 2017 (ICC17)*, May 2017, pp. 1–6.

[6] ITU-T E.804 Rev. 14, "QoS aspects for popular services in mobile networks," 2016.

[7] ETSI TS 102 250-6 V1.2.1, "STQ QoS aspects for popular services in mobile networks; part 6: Post processing and statistical methods," 2004.

[8] ETSI EG 203 165 V1.1.1, "STQ Throughput measurement guidelines," 2012.

[9] M. Laner, P. Svoboda, S. Schwarz, and M. Rupp, "Users in cells: A data traffic analysis," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, April 2012, pp. 3063–3068.

[10] ETSI TS 102 250-4 V2.4.1, "STQ QoS aspects for popular services in mobile networks; part 4: Requirements for quality of service measurement equipment," 2015.

[11] M. Lerch and M. Rupp, "Measurement-based evaluation of the LTE MIMO downlink at different antenna configurations," in *WSA 2013; 17th International ITG Workshop on Smart Antennas*, March 2013, pp. 1–6.

[12] ETSI TR 102 581 V1.2.1, "STQ a study on the minimum additional required attenuation on the antenna patch of the field test equipment," 2015.