Fine-grained LTE Radio Link Estimation for Mobile Phones

Nicola Bui[12], Foivos Michelinakis[12], Joerg Widmer[1]
[1]IMDEA Networks Institute, Leganes (Madrid), Spain
[2]UC3M, Leganes (Madrid), Spain

***Abstract*—Recently, spectrum optimization solutions require mobile phones to obtain precise, accurate and fine-grained estimates of the radio link data rate. In particular, the effectiveness of anticipatory schemes depends on the granularity of these measurements. In this paper we use a reliable LTE control channel sniffer (OWL) to extensively compare mobile phone measurements against exact LTE radio link data rates. We also provide a detailed study of latencies measured on mobile phones, the sniffer, and a server to which the phone is connected. In this study, we show that mobile phones can accurately (if slightly biased) estimate the physical radio link data rate. We highlight the differences among measurements obtained using different mobile phones, communication technologies and protocols.**

I. Introduction

Can we trust mobile phone data rate measurements? This apparently trivial question is key to evaluate the feasibility of the anticipatory networking [1] paradigm and the related future network solutions [2], [3]. For instance, exploiting achievable rate prediction to optimize mobile applications [4]–[6] requires some information exchange between mobiles and base stations so that current decisions (e.g. scheduling, admission control) can be made taking into account the future states of the system. However, while prediction errors have been studied [7], [8], the capability of mobile phones to obtain accurate measurements has never been investigated in mobile networks.

In addition to that, many recent studies [9]–[13] rely on crowd-sourced datasets to derive their conclusions without questioning mobile phone measurements accuracy and whether it is possible to aggregate them. Although reliable mobile phone applications to measure the network bandwidth exist [14]–[16], they focus on end-to-end measurements that do not provide the required level of granularity to enable anticipatory optimization. In fact, while end-to-end data rate is ideal to optimize TCP performance, the resource allocation optimization would rather benefit from the actual radio link data rate between eNodeB and user equipment (UE).

In this paper, we study whether mobile phones can accurately measure LTE radio link data rate and with which granularity (i.e. sampling frequency). To achieve this, we compare the data rate estimates computed at the physical layer of the radio link through a sniffer, at the mobile phone kernel through tcpdump and by a mobile application.

Our study is divided into two measurement campaigns: the first and largest set of experiments consists of burst transmissions, where a small amount of data is sent back-to-back to collect data rate estimates computed by the different entities (i.e., phone, sniffer and server), while in the second set, we evaluate latencies between single data packet transmissions and their corresponding acknowledgements (ACKs). These latencies allow us to study the root-causes of differences among the behaviors of different phones. In all the tests, we compared three mobile phones by different vendors and equipped with different chipsets, first performing the test from the server to the phone and, then, in the opposite direction.

The main findings of our study are the following:

1) Mobile phones achieve accurate ($> 85\%$) and precise ($> 82\%$) data rate measurements with as few as 20 KB in the downlink, where accuracy and precision are related to how close the measurement are to the sniffer ground-truth readings.
2) Uplink measurements are less accurate and less precise (65% and 60% respectively in the worst case), because LTE uplink scheduling delay causes a higher variability in the results.
3) Different chipsets exhibit variable biases and performance, thus requiring dedicated calibration to optimize accuracy.
4) Downlink accuracy and precision are linked to the latency measured on the phone: chipsets providing shorter and more deterministic latencies obtain better estimates.

The rest of the paper provides a survey of related work in Section II, specifies the measurement setup and the devices involved in Section III, and discusses the two measurement campaigns in Section IV and V. Sections VI summarizes the main findings and Section VII concludes the paper.

II. Related work

A considerable number of recent papers focus on LTE measurements and measurement techniques, but, to the best of our knowledge, none of them rely on accurate LTE scheduling information to validate their findings. Among them, Huang et al. [10] studied LTE performance measured from mobile phone data. In order to obtain a known reference for the results, the authors performed experiments using controlled traffic patterns to validate their findings.

The fraction of LTE resources used for communication is detected in [17] by means of power measurements. The goal of the authors is to evaluate the performance of M2M communications using experimental data. Similarly, RMon [18] is a solution to assess which resource blocks are used by comparing the average power measured over the resource bandwidth with that of the closest LTE reference signals. RMon achieves good

978-1-5386-2723-5/17/$31.00 ©2017 IEEE

performance and robustness, but it can only assess the average fraction of used resources. Hence, it cannot be used to capture the actual base station data rate.

LTEye [19] was the first attempt to decode the LTE control channel to access scheduling information. However the authors found in their later work [18] that LTEye could not provide sufficient reliability and a significant fraction of control messages remain undecoded. To overcome this limitation, we developed a reliable LTE control channel sniffer, called Online Watcher for LTE (OWL) [20]. In our tests, OWL successfully decoded 99.85% of LTE the control messages, thus obtaining a complete log of the eNodeB scheduling. MobileInsight [21] is a mobile phone application capable of accessing LTE control messages directly from the radio chipset and could also have been as an alternative to OWL.

A few papers [22]–[24] use commercial tools and/or operator network information to evaluate LTE performance, but their datasets (if released) only provide aggregate metrics that do not allow us to achieve the objective of this paper. The vast majority of papers however, just rely on measurement performed using mobile phones or replicated in laboratory experiments. Phone traces are used in [12] to evaluate network performance. The same authors developed a framework [13] to manage mobile phone measurements and a similar project was developed in [9]. In [11], LTE performance predictors are evaluated in laboratory setups. In addition, [25] uses TCPdump traces to perform energy efficiency evaluation of smartphones and [26] studies LTE shared access in a trial environment.

Finally, although not specifically developed for LTE, the following contributions discuss mobile measurements in general terms. The most popular approach is Ookla's Speedtest [14], which can provide a very accurate evaluation of the steady-state rate achievable by long-lived TCP connections. However, Speedtest is both data intensive (with fast connections, one test can consume more than a few tens of megabytes) and cannot provide estimates at the granularity required in this study. A few recent papers [15], [16] studied end-to-end achievable throughput, also accounting for inter-arrival times and passive monitoring techniques, but without comparing their findings with ground-truth readings. The accuracy of WiFi measurements performed by mobile phone is studied in [27] based on a timing analysis. However, their results cannot be applied to our scenario for two reasons: WiFi and LTE differs significantly in terms of scheduling and MAC protocols, and tcpdump traces do not provide a reliable ground truth for the physical radio link.

This study improves over the current state of the art by, first, evaluating data rate estimates on the radio link, instead of end-to-end throughput and, second, by relying on an accurate LTE sniffer to obtain a ground truth of the measurements.

III. Setup and Definitions

Figure 1 illustrates our experimental setup, which consists of five entities. The **target UE** is the mobile device under test which is connected to the **target eNodeB**. The **sniffer** is a BladeRF x40 software defined radio [28] that samples and records the LTE signal to be decoded by OWL. The sniffer

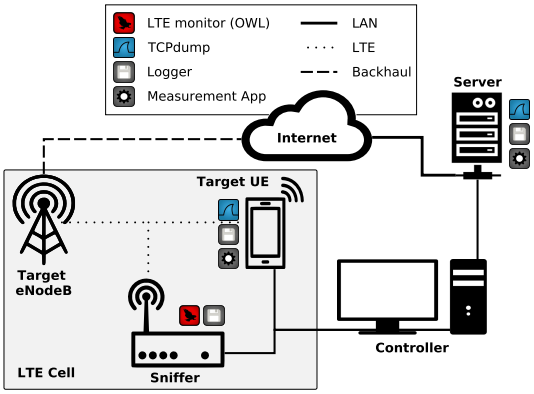


Fig. 1. Experiment setup showing devices, connections and software.

is shown as connected to the eNodeB-UE link only, but it actually records and decodes all control messages sent by the eNodeB and, thus, it is aware of all of the traffic exchanged in the cell. The **server** is a PC in our local network configured with a public IP address in order to be reachable by the target UE. The Internet cloud in the figure groups all the links that form the backhaul of our setup including the operator network. Finally, the **controller** is a second PC in our local network which is directly connected to the target UE and the sniffer via USB and to the server via Ethernet.

In order to assess the impact of different hardware, we choose three mobile phones from different vendors with comparable technical specifications but equipped with different chipsets. In particular, we opt for a Motorola MotoG LTE [29], a Huawei P8 Lite [30] and a ZTE Blade A452 [31] equipped with Qualcomm, Huawei and MediaTek chipsets, respectively. The following list summarizes the features relevant for this study (the short names used in the rest of the paper are written in bold face):

- Motorola **MotoG** 4G (2014) – Chipset: Qualcomm Snapdragon 400 MSM8926; CPU: ARM Cor-tex-A7, 1200 MHz (4 cores); Android: 4.4.2 KitKat; RAM: 1 GB.
- **Huawei** P8 lite (2015) – Chipset: Huawei HiSilicon KIRIN 620; CPU: ARM Cortex-A53, 1200 MHz (8 cores); Android: 5.0.2 Lollipop; RAM: 2 GB.
- **ZTE** Blade A452 (2015) – Chipset: MediaTek MT6735P; CPU: ARM Cortex-A53, 1000 MHz (4 cores); Android: 5.1 Lollipop; RAM: 1GB.

We have four different software modules in our setup. The gear-shaped icon refers to the **Measurement App**, which controls the communication between the target UE and the server. For every successful socket call (either "send" or "receive"), it logs the time and the amount of data exchanged. This application is implemented in Python to obtain the same behavior both on the phone and the server. The shark-fin-shaped icon refers to **TCPdump** [32], which we use both on the UE and the server to obtain transmission timestamps at the kernel level as well as the payload size. The floppy-disk shaped icon illustrates the **Logger** application that formats the output of the other tools for later analysis.

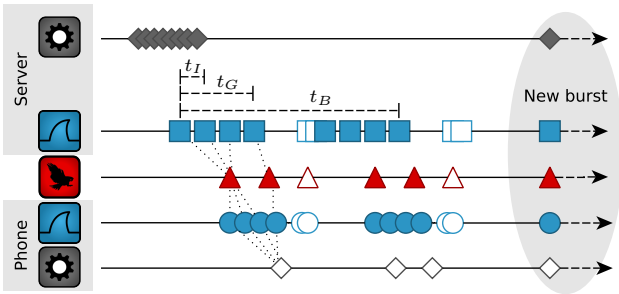


Fig. 2. Communication diagram for downlink burst transmissions.

The **LTE monitor** (owl-shaped icon) implements our Online Watcher for LTE (OWL [20]) control channel measurements. OWL is built starting from srs-LTE [33], an open-source implementation of LTE, and extends its functionalities to provide a reliable decoder of the physical control channel. From LTE control messages, OWL computes the transport block size assigned to each downlink and uplink communication. In this way, we can measure the actual LTE radio link data rate in every transmission time interval (TTI), i.e. 1 ms. This data rate differs from the usual notion of end-to-end throughput and it is the main metric needed for anticipatory networking.

The LTE cell used during the tests belongs to Yoigo, a Spanish mobile network operator, and operates in LTE band 3 (1800 MHz) using a bandwidth of 10 MHz. The cell is chosen due of the relatively low load and the very good signal quality from the test location.

Our setup is characterized by three physical and five logical measurement points: we monitor the communications at the target UE, at the sniffer and at the server. Both the UE and the server collect information by means of TCPdump and at the application to capture the difference between application and kernel measurements by means of a data rate estimation technique using packet train dispersion [34].

As introduced above, we perform two measurement campaigns, the first dealing with burst transmission (see Section IV-A and IV-B for the test description and the results respectively) and the second with periodic isolated transmissions (Section V-A and V-B). In the first campaign our goal is to evaluate the accuracy and the precision of fine-grained measurements, while in the second we study latencies in the different devices. Both campaigns consider both downlink (from the server to the UE) and uplink communication.

IV. Burst Transmissions

The first measurement campaign has the main objectives of evaluating the accuracy and the precision of data rate estimates obtained by mobile applications, and to analyze the differences in performance obtained by the three phones.

We use the following symbols: t, s, n and r denote durations, transmission sizes, number of packets, and the data rates. All these quantities are easy to compute from the information available in our tests and they do not require complex filtering. In fact, we just evaluate the data rate $r = s/t$ as the ratio between the amount of data s transmitted in a given time and the time t itself.

A. Experiment Description

We focus on packet trains (burst) from when they are first sent back-to-back from an application to their reception at the other endpoint. In particular, we are interested in comparing transmissions in the LTE radio link and the events tracked by a mobile phone at the application and the kernel level. We use Figure 2 as an example of a downlink test. The packets are generated by the application almost at the same time. As they are sent through a TCP socket they become spaced according to TCP dynamics and delays. For all layers, empty markers represent ACKs, except for the phone application layer where they mark packet receptions.

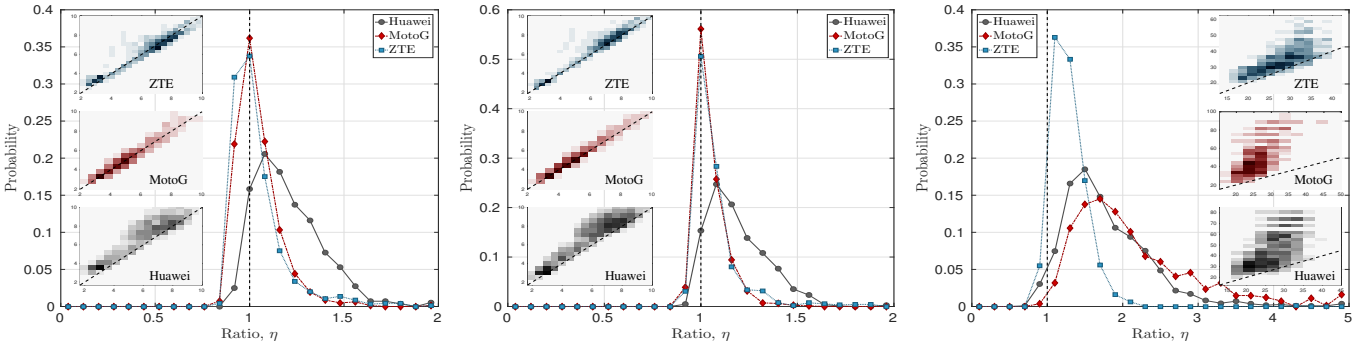
For the analysis, we define interarrival time t_I as the interval between two consecutive arrivals on the same layer and burst time t_B as the time between the first and the last packet of a train. LTE may impose a further grouping of packets when large transport blocks can fit more than a single TCP packet; this is observed at the phone as a group of packets arriving almost at the same time and as a single event at the sniffer. We define group time t_G as the time elapsed between the first and the last packet of a series of continuous arrivals. The data rate computed on groups is the measure that approaches the most the physical rate. In what follows groups are identified by those packets whose interarrival times are shorter that a threshold $t_I \leq \tau \leq \tau_C$, where $\tau_C = 1$ ms is the TTI of LTE.

To compare LTE with phone and server traces, we fix the burst size to 100 and 30 KB for downlink and uplink experiments, respectively, to obtain at least 10 transmissions per burst: in our setup with a 10 MHz channel, the maximum LTE transport block size is 73392 and 28336 [35] bits in downlink and uplink, respectively.

B. Experiment Results

In this section we compare data rates measurements by means of an estimator ratio defined as $\eta = r/r_0$, where r_0 is the reference data rate, which, if not otherwise specified, is measured by the LTE sniffer. The estimators' accuracy is highest when the the ratio is $\eta = 1$ and degrades if it is either higher (overestimation) or lower (underestimation). Moreover, the standard deviation of the ratio is proportional to the estimator precision. Thus, we show the distribution of the estimators' ratios and we provide accuracy $\alpha = [1 - |1 - \overline{\eta}|]_0$ and precision $\rho = 1 - \sigma(\eta)$, where $\overline{x}$, $|x|$, $\sigma(x)$ are the empirical average, the absolute value and the standard deviation of x and $[x]_0$ is x if $x > 0$ and 0 otherwise. In the following results the overheads between the application and the kernel (about 3.95%) and between the kernel and the sniffer (about 0.8%) are compensated.

The first and foremost results of our study are illustrated by Figure 3, which shows the empirical probability density function (epdf) of the estimator ratios obtained using burst by the three different phones computed by the application (on the left) and by the kernel (on the center) and using groups computed by the kernel (on the right). The small plots on the left of the figures show the density of the estimators in a reference system where the x-axis reports the cell ground truth and the y-axis the estimate: the darker the color the



(a) Application–bursts (b) Kernel–bursts (c) Application–groups

Fig. 3. Comparison of the estimator ratios computed on burst by the application (a) and the kernel (b) and on groups (c). The small plots on the left show estimator densities: the x-axis is the cell ground truth and the y-axis the estimate.

more estimators are in that area. The black dashed lines in the plots show what an ideal estimator would achieve: the ratio distribution would be a single spike in 1 and the densities would be on the line $x = y$.

R1 – Phone applications can obtain accurate and precise downlink data rates measurements: Figure 3(a) demonstrates that the peaks of the epdfs are very close to 1. The three phones achieves accuracy of $\alpha = 85\%$ (Huawei), $\alpha = 96\%$ (MotoG), and $\alpha = 95\%$ (ZTE). The width of the edpfs is related to the estimators' precision, in particular, the three phones have $\rho = 89\%$ (MotoG), $\rho = 85\%$ (ZTE), and $\rho = 82\%$ (Huawei). The precision is also related to the width of the estimator clouds in the small plots – wider clouds corresponds to the lower precision.

R2 – Different phones have different biases: the slightly lower score of the Huawei phone means that it tends to overestimate the data rate by about a 10%, which can be easily compensated. The same results can be verified in the small plots: the MotoG's and ZTE's densities are centered on the $x = y$ line, while the Huawei's is slightly above. Since the estimators are obtained as size-over-time ratios and the bursts have fixed size, the root cause for accuracy and precision has to be looked for in the variability of the burst duration. In particular, if a phone consistently measures shorter burst times, it will overestimate the rate and, if the time measurements are variable (e.g., random delays due to different loads on the CPU) the corresponding precision will be lower. Thus, systematic errors impact the accuracy, while random errors affect the precision of the estimates. As a consequence, it is important to compensate for the biases of the different phones when dealing with crowd-sourced measurements, otherwise errors could accumulate unpredictably.

R3 – Accuracy and precision are independent of the actual data rates: examining the small plots in Figure 3(a) the estimators span the whole x and y axes between 2 and 10 Mbps. This means that, during the experiments, the network load varied so that the actual data rate achievable by our target phones (all of them show similar behavior) was changing. In addition, the actual data rate of the experiment does not affect the estimator quality. The slightly larger cloud at higher rates is expected since the same percentage error causes a larger absolute error at higher rates.

R4 – Kernel measurements are slightly more precise: Figure 3(b) shows the estimator ratios' epdfs when the measurements are performed by tcpdump. We expect these measurements to show better performance than those obtained from the application, since they are collected by tcpdump, they are time stamped when the kernel receives packets (through an interrupt from the chipset). However, not only are the precision improvements very small (i.e., 3%, 1% and 1% for MotoG, Huawei and ZTE respectively), but the accuracy scores are almost unchanged.

Figure 3(c) is obtained using groups instead of the longer burst. Since the transmissions within a group are expected to belong to consecutive radio link layer transmissions, the data rate estimate is likely to capture the exact rate used by the eNodeB. However, these measurements are more sensitive to timing precision. Since the group threshold for the cell is 1 ms, groups are characterized by transmissions in every TTI and, as soon as a single TTI is skipped, the group ends. Thus, if timing is not precise, two or more separate groups in the cell can be detected as a single group at the kernel or by the phone application. As a consequence, while it is possible to easily separate bursts and have a unique mapping between bursts in the different layers, this is not true for groups, whose composition is device and layer dependent.

R5 – Group-based data rate estimators are imprecise and biased: our measurements show that the accuracy (α) and the precision (ρ) of data rate estimators computed on groups are low. In particular, MotoG achieves $\alpha = 16\%$ and $\rho = 46\%$, Huawei $\alpha = 25\%$ and $\rho = 48\%$ and ZTE $\alpha = 75\%$ and $\rho = 79\%$. Instead, the ZTE phone measurements, even though overestimating by circa 25%, maintains a quite acceptable precision, close to 80%.

A close examination of the density plots reveal that group-based measurements have to estimate higher and more variable data rates, because they are not averaged over the longer duration of bursts. The ground truth data rate (x-axis) extends up to 45 Mbps, which is close to the maximum throughput in our setup. Moreover, all the phones overestimate the actual readings, reaching estimates even higher than the maximum reachable of our setup. This is a further proof of the importance of precise timing and deterministic latency in the phones to enable the most fine-grained estimation.

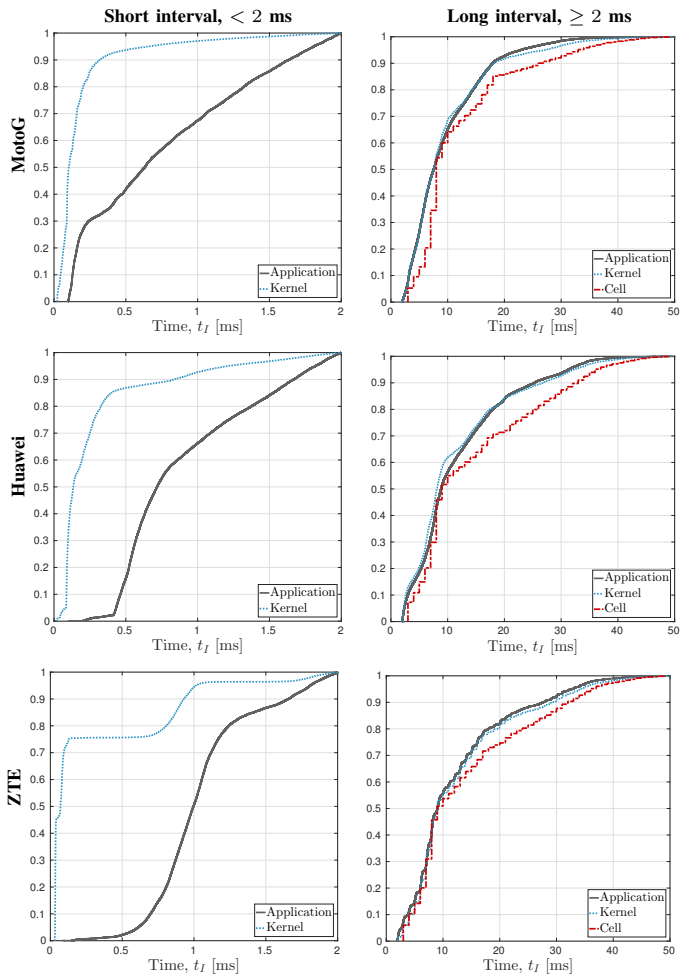


Fig. 4. Interarrival time CDFs for short (left) and long (right) intervals and the three phones.

Before moving to uplink results, a few considerations on the packet interarrival times are in order. Figure 4 provides a set of graphs showing the Cumulative Distribution Function (CDF) of the interarrival times. To emphasizes the difference between the interarrival times of packets related to the same LTE transmission from those related to intervals separating continuous arrivals, we plot on the left column the CDFs for the interarrival times shorter than 2 ms (2 TTIs) and, those longer or equal to 2 ms and shorter than 50 ms on the right column. We don't show interarrival times longer than 50 ms, since those are almost always related to inter-burst rather than intra-burst arrivals. The matrix rows show from the top to the bottom, results for MotoG, Huawei and ZTE. We omitted the cell CDFs in the plots on the left, since it would have been a single spike at 1 ms.

Focusing on the left column, we can see that the CDFs of different phones and those obtained at the kernel and at the application are very different. For instance, the MotoG plot shows that the majority (90%) of interarrival times measured at the kernel are shorter than 0.3 ms, but only 30% of those measured by the application are shorter than 0.3 ms. This means that multiple packets that are distinguishable at the

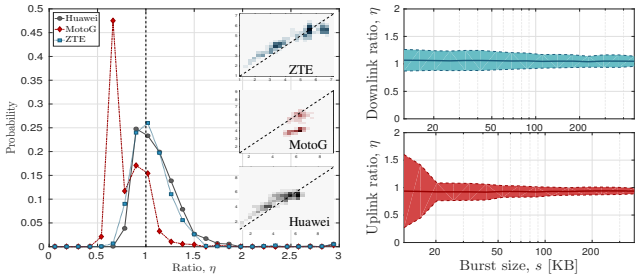


Fig. 5. Estimator ratios computed on burst in the uplink (left). Estimator ratios against different burst size.

kernel are received by the application as a single stream, thus, fixing the threshold $\tau = 0.3$ ms identify packets belonging to the same LTE transmission from those belonging to either the previous or the next.

Instead, the CDF of ZTE interarrival times at the kernel shows two flat regions, one before 0.15 ms and the second after 1 ms: this is caused by intra-transmission arrivals (the former) and inter-transmissions arrivals (the latter). Conversely, this noticeable distinction is not found in the application trace. Accordingly, we fix the grouping thresholds to different values: $\tau = 0.2$ ms for the kernel and $\tau = 1$ ms for the application.

Finally, the Huawei CDFs only show slight inflections at 0.4 ms (kernel) and 0.7 ms (application), but both are less marked than those of the other phones. We set the two thresholds accordingly. As will be more evident hereafter, a more skewed CDF with distinguishable intra- and inter-transmission thresholds corresponds to more deterministic latencies in the phone and, in turn, to better group data rate estimates.

Analyzing the plots on the right, we can compare the interarrival time CDFs measured by the application, the kernel and the sniffer. Here we observe that the Huawei phone that has a slightly lower accuracy in terms of data rate estimation, and also shows a larger gap between the cell CDF and the other two, in particular between 10 and 30 ms. This confirms that data rate measurements are influenced by timing precision.

One final observation related to the ZTE phone is that both the application and the kernel CDFs show the same stair-shaped trend as in the cell CDF. Again, this is due to a lower variability of the ZTE latency, which will become more evident in the following second set of experiments.

Figure 5 (left) shows the uplink data rate estimator ratios of the three phones. Again, we compare kernel measurements on the phone against the sniffer's ground truth for the cell. Note that uplink application measurements would require a dedicated application that could intercept ACKs or especially designed to monitor the sending socket. The normal socket behavior is to accept send requests from the application until the transmission buffer is full and, since this buffer is usually larger than our data burst, the application can send to the socket a whole burst at once making it impossible to measure the data rate at the phone application.

R6 – Mobile phones can obtain accurate and precise uplink data rates measurements: although the MotoG underestimates the rate by about 30%, the other two phones have the peaks of their epdfs very close to 1. In particular,

they achieve an accuracy α of 93% (Huawei) and 97% (ZTE), while MotoG stops at 65%. The edpfs are also wider than those related to the downlink. This is even more evident from the density plots on the right of the figure, which highlight that the precision of uplink measurements is lower than that obtained in the downlink: 73% for the Huawei and 74% for the ZTE. Not only does the MotoG have only 60% precision, but also its density plot shows two regions where the densities accumulate. This exhibits a binary behavior of the device that will become more evident in the next section where we analyze the phone latencies.

In addition, we compare the phone kernel to the server data rate measurements. Since the results obtained are very similar to those shown in Figure 5 (left) we omit the graphics. However, it is interesting that in our experiments, uplink burst can be measured both on the phone and the server achieving similar results. Since phone to cell measurements are taken before traversing the backhaul while phone to server results include it, we can conclude that the backhaul plays a minor role in our setup, because of the favorable location of the measurement server.

Figure 5 (right) reports the results obtained by varying the burst size from 10 KB to 1 MB for the downlink, and from 6 to 300 KB for the uplink. All the figures plot the average estimator ratio in the center of a shaded area that extends one standard deviation on each side. The figures are obtained by mixing together the results for all the phones.

R7 – Bursts of 20 KB provide high accuracy and high precision: the figures show that the estimator accuracy is independent of the burst size and the precision slowly improves with increasing size. Uplink proves to be more sensitive to very small burst (i.e., the shaded area is larger in the uplink plots for small bursts) as it is subject to more network randomness and it requires slightly longer transmissions. In contrast, in the downlink communication as few as two LTE transmissions are sufficient to obtain an accurate estimate. In our test we choose the minimum burst size to cause at least two transmissions at the maximum reachable data rate. In a larger bandwidth setup and when the next LTE releases will be deployed, the minimum burst size to achieve this results has to be increased proportionally to the maximum data rate.

R8 – UDP tests obtain the same results: the network provider used in our campaigns does not allow us to make reliable UDP tests, because of firewall and traffic shaping policies. To overcome this limitation, we repeated all the tests by emulating UDP by sending its packets with a TCP header through a raw socket. All the repetitions result in performance almost identical to that obtained by their TCP counterparts. The reason is that the measurement characteristics are dictated by the intra-burst timing, which, in turn, depend on the radio link technology, and not by the inter-burst timing, which, instead, depends on the protocol. Thus, radio link measurements only need for clearly separated burst in order for mobile phones to precisely estimate the data rate.

R9 – WiFi measurements are consistent, but different: we repeated the main tests on WiFi (IEEE 802.11g) by replacing OWL with a Warp Software Defined Radio [36] using the 802.11 reference design. We consistently observe that the performance obtained on WiFi are on the same order of magnitude of those obtained on LTE, but they are not identical in terms of bias and precision. Thus, while we agree on the main claims of [27] about overheads, we believe that different technologies require specific tests to evaluate their performance.

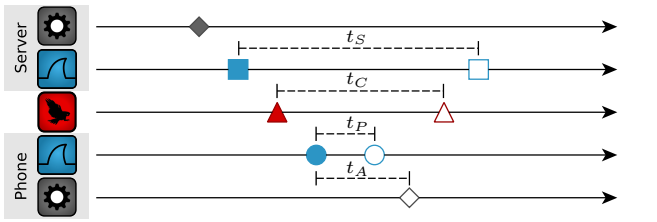


Fig. 6. Communication diagram for the downlink isolated transmission. Dimension lines illustrate data-to-ack latency.

Before moving to the next set of experiments, we discuss a few more results for which we do not provide dedicated figures. We test our data rate measurements under three other conditions: 1) we stress the phone CPU to full load during the experiments; 2) we inject additional traffic in the cell under test up for to 95% load. Although, we expect the CPU load to add some delay to our measurement, we find that the phone kernel copes well with this load and we did not notice any significant change in the estimator performance. Similarly, the additional traffic injected in the cell only changed the actual measured data rates (i.e., lowering them), but did not decrease the estimator's accuracy.

V. Isolated Transmissions

This section details the second set of measurements. The objective of this campaign is to measure phone communication latencies to justify the differences in their behavior. As above, we first illustrate the experiment on a diagram (Figure 6) and on some trace examples and then we discuss the results.

A. Experiment Description

For the analysis of latencies measured at each communication layer we take particular care to link homologous events in the different measurement devices. While this is trivial in the phone and the server where we can access all packet header fields, identifying which LTE transmission contains a given packet in the scheduling log poses several problems. First of all, we need to find the correct RNTI of the target UE among the rest of the traffic, but while for burst transmission we could both rely on fixed burst size and periodicity, in isolated transmission tests a single packet is sent from the application, the payload of which should be large enough to differentiate it from LTE control messages and small enough to fit in a single transmit unit in the phone and the server interface. We fix the packet size to 500 bytes with a periodicity of 400 ms to leave enough time between subsequent repetitions not to confuse them with possible retransmissions. While other UEs scheduled together with our target may have similar periodicity, our UE could always be correctly detected.

Figure 6 shows the ideal communication diagram for the downlink isolated transmission test. Here, we monitor the time

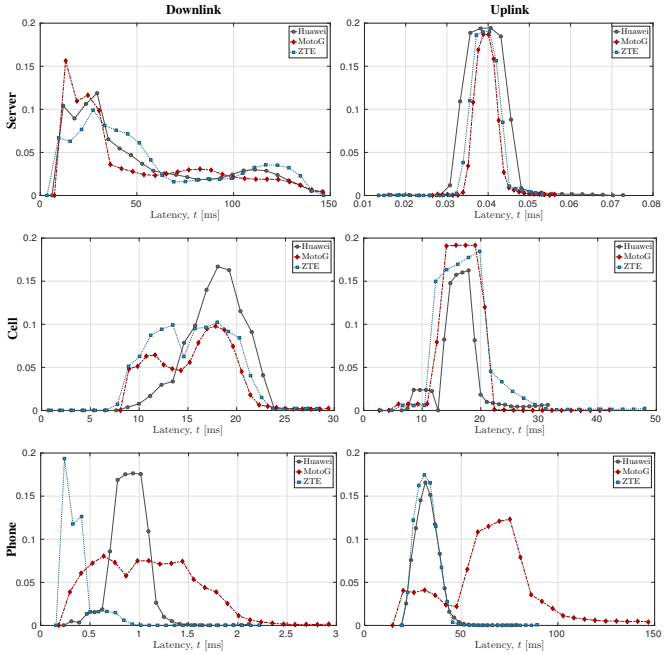


Fig. 7. Empirical probability density functions of the latencies.

elapsed between each packet and the corresponding ACK. In what follows, we refer to this data-to-ack time as latency.

Although we only monitor relative times and we do not need a perfect synchronization between the measurement layers, in order to correctly couple events we make sure to capture the first event of each test in all the layers (to have a common reference) and, then, we run a causality check on each trace to compensate possible violations. Since each subsequent layer events have to occur after the events of the upper layer, we realign traces to follow causality.

Dimension lines illustrate latencies in the different layers with the only exception of t_A which refers to the time between a packet being captured by the kernel (phone for the downlink or server for the uplink) and when it is delivered to the application. Note that applications cannot measure their own latency without intercepting the communication ACKs at the kernel level.

B. Experiment Results

Figure 7 summarizes the main results of the latency measurements from which we draw conclusions about differences in the three phones' behaviors. All plots show the epdfs of the latency measured at the three measuring devices. All latencies are measured at the kernel level, since the application is not automatically notified of ACK receptions.

The plots in the figures are grouped vertically by communication direction and horizontally by layers and they are best read from the top left in clockwise order to follow the communication sequence.

The latency measured at the server in the downlink tests (top left) is the sum of the delays caused by two Internet traversals, two LTE scheduling delays (downlink first and then uplink) and phone processing (chipset time plus protocol stack traversal in the kernel). The latency at the cell downlink (top center) starts when the downlink LTE transmission is already scheduled and, as a consequence, it only contains the phone processing and the LTE uplink scheduling delays. The latency at the phone downlink (top right) starts from when the kernel receives the reception interrupt from the chipset to when the ACK transmission to the communication interface.

R10 – Chipsets with short and deterministic latency achieve more accurate and precise data rate estimation: in downlink tests, the latencies are similar in all the layers, except on the phone. The ZTE latencies exhibit a single peak before 0.5 ms, the Huawei a single, slightly wider peak at about 0.9 ms, while the MotoG shows a wider distribution of its latencies ranging from 0.3 to almost 2 ms. Recalling from Figure 3(a) that MotoG and ZTE achieve higher accuracy and precision in their data rate estimates, we can conclude that chipsets with a shorter latency are more accurate in estimating the data rate. Instead, the Huawei latency is closer to the length of the LTE TTI and is the cause of the overestimation of the data rate. To explain the difference in performance between the ZTE and MotoG, we need to consider the CDFs of their short interarrival time (recall Figure 4 top and bottom graphs on the left). While the MotoG application captures intra-group events (shorter than 0.5 ms), the ZTE application distribution starts only after 0.5 ms, but it is very precise at the kernel. Thus, the MotoG application data rate estimate fares better than ZTE, which, instead, is more precise at the kernel level only.

The low variability of ZTE latencies explains why the ZTE long interarrival time distribution has the same stair-shaped trend as the cell distribution. As a consequence of this higher precision at the kernel, the ZTE phone can better discriminate between LTE transmissions, which, instead, are smoothed in the other two phones, and is able to obtain more accurate group-based data rate estimates (recall Figure 3(c)). Also, since the three phones show similar times for server and cell latencies we can exclude that network traversals and LTE scheduling impact data rate estimates between phone and cell in our setup.

R11 – LTE discontinuous reception configuration [37] influences uplink data rate estimates: the bottom row of Figure 7 illustrates the epdfs of uplink experiments. Latencies measured on the phone kernel include both uplink and downlink LTE scheduling, two Internet traversals and the server processing; latencies at the cell include the Internet traversals and the server processing delays, while the server latencies only include processing delay. The server processing is negligible, since it is shorter than 50 µs. Similarly, we can exclude that the network delays play an important role in the uplink data rate estimates, since the three phones show almost identical latencies when measured at the cell. Conversely it is the LTE uplink scheduling delay that influences the estimator the most. This delay is expected to be about 20 ms for a connected device starting a new transmission, while in our measurements the epdfs are centered at about 50 ms (Huawei and ZTE) and 85 ms (MotoG) with a smaller peak at 40 ms. All these longer uplink delays are due to LTE discontinuous reception (DRX), which is an energy saving feature that allows mobile phones to duty-cycle between sleep and wake phases. Since to discriminate among different transmissions we

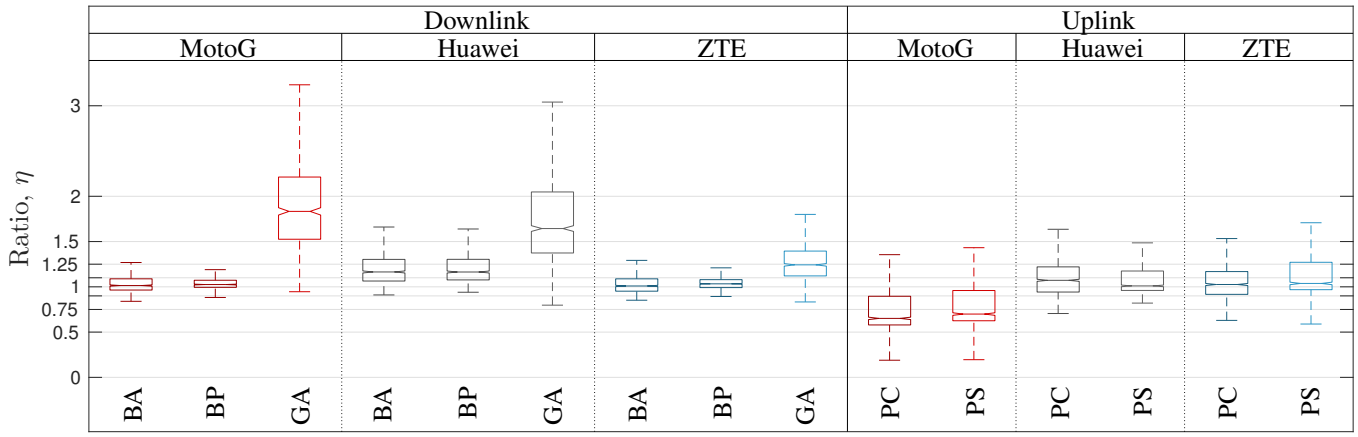


Fig. 8. Overall comparison among the data rate estimators. All boxplots show the distribution of the estimator ratios from the 25th and 75th percentiles.

separate them by 400 ms (1 s for bursts), all the transmissions start with the devices in DRX mode. The actual duration of the sleep time depends on agreements between UE capabilities and eNodeB requirements. Thus, MotoG uses a more conservative DRX setup (with a longer sleeping period), most likely due to the fact that it is an earlier (2014) model than the other phones (2015). The overall effect of this latency is that uplink data rate estimators are less precise than downlink estimates by circa 10% due to the wider distributions of the latencies.

VI. Summary

Figure 8 provides a visual summary of the results discussed in the paper. In the figure, one boxplot is shown for each of the main experiments, highlighting the median (central mark of the boxes) and the 25th and 75th percentiles (box edges) of the estimator ratios η. At the bottom of the figure we specify the type of η used. BA are ratios between data rate measurements performed on bursts (B) at the application (A) and cell estimate. BP are the same but computed by the phone kernel (P), while GA are the same as BA, but computed on groups (G). All downlink ratios use the sniffer as a reference. On the uplink side, we show PC, which compares phone kernel (P) estimates with cell (C) and PS which use the server (S) application as a reference. All uplink results are computed on bursts. Note that we do not show graphs for PS in the previous results, since PC and PS are quite similar. This shows that in our setup uplink data rate estimates on the server and the phone obtain comparable results.

Looking at all the results side-by-side, it is evident that on the one hand side, all the phones are capable of accurate and precise data rate estimation, but on the other hand they have significantly different biases and precisions, as seen for instance, with Huawei and ZTE for BA in the downlink and for PC in the uplink. Similarly noticeable is that group-based estimators, GA, achieve reasonable accuracy on the ZTE phone only and the MotoG uplink measurements are heavily impaired by the different LTE latency.

To conclude our study on LTE radio link estimation, we can affirm that the precision and the accuracy achieved by the three devices are sufficiently high to enable anticipatory networking optimization up to a time granularity of about 50-100 ms and after having compensated the device bias. This is true for both uplink and downlink estimates either obtained by the kernel or the application. Conversely, to increase the measurement granularity and, in turn, the optimization potential, direct readings of the actual physical rate are needed.

VII. Conclusions

In this paper we presented the first experimental evaluation of the accuracy and the precision of LTE data rate measurements performed by mobile phones. To summarize the main finding of the study:

Mobile phones can achieve accurate and precise data rate measurements: we showed that downlink application measurements are both accurate and precise (R1), downlink kernel-level measurements improve the precision, but only slightly (R4), and uplink kernel estimates are accurate and precise, but less than those obtained on the downlink (R6).

Mobile phones have different biases: R2 and R6 show that mobile phone estimates vary depending on many factors. R8 highlights chipset latency as the main cause for downlink differences, while R11 identifies the LTE uplink delay as the main source of variability. As a consequence, when combining the results obtained using different phones, it is good practice to evaluate the bias of each of them in order to compensate systematic errors. Moreover, R9 shows that measurements performed on different communication technologies with the same mobile phones show similar, but not identical results.

Small data bursts are sufficient for 95% accuracy at 80% precision: we studied how the estimator quality varies with burst size (R7) and we found that 20 KB (or 50 ms) are sufficient to obtain accurate and precise estimators. Moreover, in our tests the burst estimator quality showed very little dependence of the measured data rate (R3). Instead, we found that estimating the data rate from groups of packets arriving back-to-back at the measurement point leads to low quality (high bias, low precision) estimates (R5). This shows that anticipatory networking solutions can rely on precise information computed over 50 ms windows, but the actual radio link data rate has to be obtained to increase this resolution.

Burst measurements are protocol independent: by comparing TCP and UDP tests, R8 shows that precision and accuracy depend on the capability of detecting burst precisely.

Finally, we believe that this study offered a new perspective about crowd-sourced measurement campaigns and that will help improving the reliability of future campaigns.

Acknowledgments

This work has been supported by the European Union H2020-ICT grant 644399 (MONROE), by the Madrid Regional Government through the TIGRE5-CM program (S2013/ICE-2919), the Ramon y Cajal grant from the Spanish Ministry of Economy and Competitiveness RYC-2012-10788 and grant TEC2014-55713-R.

References

[1] N. Bui, M. Cesana, A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *Submitted to IEEE Communications Surveys and Tutorials*, 2016.

[2] NGMN. Next Generation Mobile Networks. [Online]. Available: {http://www.ngmn.de/publications/all-downloads/article/ngmn-5g-white-paper.html}

[3] G. P. Fettweis, "The tactile internet: applications and challenges," *Vehicular Technology Magazine, IEEE*, vol. 9, no. 1, pp. 64–70, 2014.

[4] N. Bui, I. Malanchini, and J. Widmer, "Anticipatory admission control and resource allocation for media streaming in mobile networks," in *ACM MSWiM*, Nov. 2015, pp. 255–262. [Online]. Available: http://doi.acm.org/10.1145/2811587.2811604

[5] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, "A control-theoretic approach for dynamic adaptive video streaming over HTTP," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4, pp. 325–338, 2015.

[6] Y. Sun, X. Yin, J. Jiang, V. Sekar, F. Lin, N. Wang, T. Liu, and B. Sinopoli, "CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction," in *ACM SIGCOMM*, 2016, pp. 272–285.

[7] N. Bui and J. Widmer, "Mobile network resource optimization under imperfect prediction," in *IEEE WoWMoM*, 2015.

[8] T. Mangla, N. Theera-Ampornpunt, M. Ammar, E. Zegura, and S. Bagchi, "Video through a crystal ball: effect of bandwidth prediction quality on adaptive streaming in mobile environments," in *ACM International Workshop on Mobile Video (MoVid)*, 2016, p. 1.

[9] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4g lte networks," in *ACM MobiSys*, 2012, pp. 225–238.

[10] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck, "An in-depth study of lte: Effect of network protocol and application behavior on performance," in *ACM SIGCOMM*, Oct. 2013, pp. 363–374.

[11] J. B. Landre, Z. E. Rawas, and R. Visoz, "Lte performance assessment prediction versus field measurements," in *IEEE PIMRC*, Sep. 2013, pp. 2866–2870.

[12] A. Nikravesh, D. R. Choffnes, E. Katz-Bassett, Z. M. Mao, and M. Welsh, "Mobile network performance from user devices: A longitudinal, multidimensional analysis," in *Springer PAM*, 2014, pp. 12–22.

[13] A. Nikravesh, H. Yao, S. Xu, D. Choffnes, and Z. M. Mao, "Mobilyzer: An open platform for controllable mobile network measurements," in *ACM MobiSys*, 2015, pp. 389–404.

[14] Ookla, "Speedtest mobile apps," http://www.speedtest.net/mobile/, last accessed November 2016.

[15] Q. Xu, S. Mehrotra, Z. Mao, and J. Li, "PROTEUS: network performance forecast for real-time, interactive mobile applications," in *ACM MobiSys*, Taipei, Taiwan, June 2013, pp. 347–360.

[16] Y. Xu, Z. Wang, W. K. Leong, and B. Leong, "An end-to-end measurement study of modern cellular data networks," in *Springer PAM*, 2014, pp. 34–45.

[17] C. Ide, B. Dusza, and C. Wietfeld, "Performance of channel-aware M2M communications based on LTE network measurements." in *IEEE PIMRC*, 2013, pp. 1614–1618.

[18] X. Xie, X. Zhang, S. Kumar, and L. E. Li, "pistream: Physical layer informed adaptive video streaming over lte," in *ACM MobiCom*, Sep. 2015, pp. 413–425.

[19] S. Kumar, E. Hamed, D. Katabi, and L. Erran Li, "LTE radio analytics made easy and accessible," in *ACM SIGCOMM*, vol. 44, no. 4, 2014, pp. 211–222.

[20] N. Bui and J. Widmer, "OWL: a Reliable Online Watcher for LTE Control Channel Measurements," in *ACM All Things Cellular*, Oct. 2016.

[21] Y. Li, C. Peng, Z. Yuan, J. Li, H. Deng, and T. Wang, "Mobileinsight: Extracting and analyzing cellular network information on smartphones," in *ACM Mobicom*, Oct. 2016.

[22] A. Elnashar and M. A. El-Saidny, "Looking at lte in practice: A performance analysis of the lte system based on field test results," *IEEE Vehicular Technology Magazine*, vol. 8, no. 3, pp. 81–92, Sep. 2013.

[23] N. Becker, A. Rizk, and M. Fidler, "A measurement study on the application-level performance of lte," in *IFIP Networking Conference*, June 2014, pp. 1–9.

[24] M. Jovanovic, M. K. Karray, and B. Blaszczyszyn, "QoS and network performance estimation in heterogeneous cellular networks validated by real-field measurements," in *ACM PE-WASUN*, 2014, pp. 25–32.

[25] M. Siekkinen, M. A. Hoque, J. K. Nurminen, and M. Aalto, "Streaming over 3g and lte: How to save smartphone energy in radio access network-friendly way," in *ACM MoVid*, May 2013, pp. 13–18.

[26] M. Palola, M. Matinmikko, J. Prokkola, M. Mustonen, M. Heikkil, T. Kippola, S. Yrjl, V. Hartikainen, L. Tudose, A. Kivinen, J. Paavola, and K. Heiska, "Live field trial of licensed shared access (LSA) concept using LTE network in 2.3 GHz band," in *IEEE DYSPAN*, Apr. 2014, pp. 38–47.

[27] W. Li, R. K. P. Mok, D. Wu, and R. K. C. Chang, "On the accuracy of smartphone-based mobile network measurement," in *IEEE INFOCOM*, Apr. 2015, pp. 370–378.

[28] Nuand, "Bladerf software defined radio," https://www.nuand.com/blog/product/bladerf-x40/, last accessed November 2016.

[29] Motorola, "MotoG LTE," http://www.devicespecifications.com/en/model/10da2ca3, last accessed November 2016.

[30] Huawei, "P8 Lite," http://www.devicespecifications.com/en/model/b95e33c3, last accessed November 2016.

[31] ZTE, "Blade A452," http://www.devicespecifications.com/en/model/a44f38ba, last accessed November 2016.

[32] MartinGarcia, Luis, "TCPdump," http://www.tcpdump.org/, last accessed November 2016.

[33] I. Gomez-Miguelez, A. Garcia-Saavedra, P. D. Sutton, P. Serrano, C. Cano, and D. J. Leith, "srsLTE: An open-source platform for LTE evolution and experimentation," in *ACM WinTECH*, Oct. 2016.

[34] F. Michelinakis, N. Bui, G. Fioravantti, F. Kaup, D. Hausheer, and J. Widmer, "Lightweight mobile bandwidth availability measurement," *Elsevier Computer Communications*, vol. 84, pp. 73–83, Jun. 2016.

[35] ETSI, "E-UTRA; Physical layer procedures," *3GPP TS*, vol. 36.213, p. V13, 2016.

[36] Wireless Open Access Research Platform, "802.11 reference design for WARP v3," http://warpproject.org/trac/wiki/802.11, last accessed November 2016.

[37] ETSI, "E-UTRA; Medium Access Control (MAC) protocol specification," *3GPP TS*, vol. 36.321, p. V13, 2016.