Available Bandwidth vs. Achievable Throughput Measurements in 4G Mobile Networks

Giuseppe Aceto*,◇, Fabio Palumbo*, Valerio Persico*,◇, and Antonio Pescapé*,◇
*University of Napoli Federico II (Italy) and ◇NM2 srl (Italy)
{giuseppe.aceto, fabio.palumbo, valerio.persico, pescape}@unina.it

***Abstract*—Mobile broadband (MBB) networks are being increasingly used worldwide, and they are planned to steadily evolve to support new services, bigger user base, and booming machine-to-machine communications. In this scenario, performance measurements on deployed networks becomes crucial. In particular, being aware of the achievable throughput has multiple important uses ranging from path and server selection, to root-cause analysis. Throughput estimation suffers from high intrusiveness and dependence on transport and application protocol, while Available Bandwidth is a network-layer metric characterizing the spare capacity of path, not suffering from any of these shortcomings. However, ABw estimation tools have been mainly developed focusing on wired networks, with limited attention to mobile and wireless scenarios, multi-homed mobile nodes, and their peculiarities.**

In this work, we analyze ABw estimates and TCP achievable throughput measurements on a real 3G/4G testbed (the MONROE platform) performing tests with different providers from countries across Europe. The two metrics are compared for results and in terms of measurement intrusiveness and time cost, highlighting the non-trivial relationship between them. In particular, results show that in suitable conditions ABw estimates can be used as a *proxy* for TCP achievable throughput measurements, while generating much lower traffic volumes (a critical asset on MBB networks), and that policies enforced by service providers may significantly alter the difference between the two estimated metrics. Future research on how to exploit these findings is discussed as well. We published as open data packet traces and logs of the measurements.

I. Introduction

Mobile Broadband Networks (MBB) are spreading their coverage worldwide, gradually beginning to satisfy the increasing need for *anywhere* and *anytime* high bandwidth access to the Internet [1]. As a consequence, the amount of data flowing through cellular networks is increasing with the years, with users getting used to new usage patterns and to new services leveraging the performance of new smartphones and new mobile access networks. A related phenomenon can be also observed: there is an increasing ubiquity of Wi-Fi services and multi-homed devices (even basic smartphones models can use Wi-Fi and 3G/4G networks, while dual-SIM smartphones can further select a second 3G/4G provider). Moreover multi-carrier access with single SIM cards is becoming available [2] allowing even common smartphones with one SIM to dynamically select from a set of Mobile Network Operators. The support of multiple heterogeneous network access providers is also specified in 5G goals [3].

When considering multiple network paths (that in a multi-homed setting can wildly differ in performance), one of the key factor is the throughput that each path can support. A relatively straightforward method to obtain this metric would be to perform a data transfer of suitable duration, and assess the overall goodput. Several drawbacks advise against this course: the choice of transport-layer (or application-layer) protocol and the duration of the transfer significantly affect the obtained result [4], and in performing throughput measurements on mobile networks the volume of probe traffic concurs to volume-based billing (or pre-paid volume quota). On mobile broadband networks the throughput is in the order of tens of *Mbps*, therefore the full-bandwidth test would quickly exert a toll on volume quotas.

These motivations suggest for considering a network layer metric, *available bandwidth* (ABw), that characterizes the whole path under measurement in terms of its *spare capacity*. Being dependent on the whole path, less volume intrusive than bulk transfer measurements, and not dependent on transport- nor application- layer, makes *available bandwidth* an ideal choice to characterize the network path. Unluckily, ABw estimation tools perform best on wired networks, with limited applications to mobile and wireless scenarios, that are specifically challenging due to the different physical and media-access control peculiarities. Moreover, while capacity of wired paths is stable for time spans of several hours, wireless link capacity can be quickly varying, depending on radio channel conditions. These considerations make estimation of available bandwidth on mobile broadband connections of high practical interest and also subjected to active research and experimentation.

In this work, we assess the performance of a recent Available Bandwidth tool, Yaz[1] [5], on a real 3G/4G testbed, the MONROE platform [6], which allows us to test different providers located in different countries across Europe. As the communications use commercial data plans (interacting with the real operating infrastructure of the mobile operator) the measurements are performed *in the wild*. We run tests from mobile nodes towards a measurement server located in our laboratory, in the uplink direction. Results of available bandwidth estimation are compared to contextual measurements of TCP achievable throughput, highlighting the relationship between these two different metrics, also in terms of intrusiveness (generated traffic) and time. In particular, results show in which conditions and to what extent ABw estimates can be used as a *proxy* for TCP achievable throughput, while consuming a fraction of the traffic volume necessary for the estimation. Moreover we highlight how policies enforced by mobile network operators may significantly alter the difference between the two estimated metrics. We conclude with considerations on research possibilities for further improving the applicability

[1]After preliminary experiments we had to slightly modify the tool to make it more robust to packet reordering (not rare on mobile networks). We released the changes on http://traffic.comics.unina.it/sdn/yaz_modified.tar.gz.

978-3-903176-14-0 © 2018 IFIP

and advantages of ABw estimation on mobile broadband networks. We have made publicly available as *open data* the dataset [2](raw packet captures and logs of measurements) and as *open source* the experiment code[3] besides the modified version of Yaz.

II. Background

In this section we provide the main concepts related with the ABw measurements and the MONROE testbed.

Available Bandwidth: Available bandwidth is an important dynamic characteristic of a network path, being equivalent to the amount of traffic that can be added to the path without affecting the other flows that cross it, and independently from their bandwidth-sharing properties. Such definition tells it apart from other bandwidth-related metrics such as bulk transfer capacity and maximum achievable throughput [7]. In its formal definition, available bandwidth of a link is defined as the average spare capacity during the considered time interval; the available bandwidth on a path is instead defined as the minimum value of available bandwidth of the links composing the path. Available bandwidth estimation has significant importance for both service provider and application perspectives. Service providers use this parameter for network management and traffic engineering purposes. Furthermore, nowadays, video streaming generates the largest portion of Internet traffic, where ABw measurement techniques play a significant role in adapting to the current network load. In general, knowledge about the available bandwidth over the network would benefit both users and operators of network applications and infrastructures.

MONROE: The MONROE testbed [8] has been designed to allow experimentation with MBB access networks, utilizing commercial mobile data plans (therefore accessing the operational mobile network as a common user, sharing the infrastructure with all real users). From the point of view of the experimenter, the platform is composed of the following main components: (i) the geographically distributed hardware appliances (MONROE nodes) running the experiment software; (ii) the software running on the MONROE node, divided in core components and user-defined experiments; (iii) the management system, allowing user access, experiment scheduling, and data import; (iv) the database holding experiments data and automatic periodic measurements. The hardware setup is standardized on all nodes, and in previous works [9] we have experimentally evaluated its suitability for ABw measurements. The software defining the experiments is executed on the MONROE node in lightweight virtualized environments (Docker containers). For further details on MONROE we refer to [8], [6] and cited deliverables.

III. Related work

During the years, different available bandwidth estimation tools have been developed, mostly considering wired networks, and several comparisons of their performances in different scenarios have been carried out. Some of the most popular and widely used are, for example, Pathload [4], Yaz [5] (which is a calibrated version of Pathload designed to improve the accuracy), pathChirp [10], ASSOLO [11], SPRUCE [12], and WBest [13], which is the first to be specifically designed for WiFi networks.

[2]https://doi.org/10.5281/zenodo.1300512.

[3]https://hub.docker.com/r/fabpal92/abw_ach_sometime/

TABLE I: Details of the experimental campaigns conducted. In all campaigns the tools Yaz and D-ITG are adopted, in Preliminary one also Pathload.

Campaign	Run duration (s)	Daily exp.	Start date	Campaign duration (days)
Preliminary	60	1	Jan. 29^{th}	8
Long-term	20	1	Feb. 17^{th}	35
Additional	10	24	May 26^{th}	9

More recent works have also taken into account wireless scenarios, focusing on WiFi and cellular networks. For WiFi, authors in [14] propose an enhancement over WBest by evaluating the impact of 802.11n frame aggregation on probe packets, to improve bandwidth estimation. Song and Striegel in [15] further analyze this mechanism, and leverage it by inducing frame aggregation on probe packets to estimate ABw with higher accuracy. In the evaluation of accuracy, both of the previous techniques consider UDP throughput as ground truth for capacity, from which available bandwidth is then derived subtracting the amount of generated cross-traffic. This kind of evaluation, however, is only possible in controlled testbeds, where experimenters have full control over the traffic interfering with the measurement process.

A first attempt to measure available bandwidth specifically targeted at LTE networks is proposed in [16], where a curve-fitting approach is used to detect the turning point in the received gaps of a train of probe packets. This technique is evaluated on a commercial LTE network in Japan, considering FTP throughput as reference. Oshiba et al. in [17] evaluate the effect of LTE packet scheduler and introduce a novel tool, PathQuick3, which also uses a curve fitting approach to estimate the transition point in a train of packets with different rates. Unlike previous tools, trains have a fixed duration, as probe packets are all separated by the same interval. Indeed, different bitrates are here imposed by acting on packet sizes instead of inter-packet times. Bulk TCP throughput, measured using speed test applications provided by Mobile Network Operators, is taken as ground truth in order to assess the estimate error; the experimental campaign considers a commercial LTE network, tested in different areas of Tokyo. The same authors present a novel tool in [18], named PathML, which uses four different machine learning techniques trained using the information about queueing delays measured receiver-side to predict the available bandwidth. The ground truth used for training the machine learning algorithms is again the bulk TCP throughput, measured as before; the experimental evaluation in Japan shows an improvement over PathQuick3.

Several limitations are present in works specific for mobile 3G/4G networks. First, limited scenarios have been considered for the experimental evaluation, as for example only one provider is tested in one specific country. They lack an extended analysis about the TCP throughput, which is taken at face value as ground truth for Available Bandwidth (despite the two being two conceptually different metrics); in two out of three works this value is obtained through external speed-test applications, which typically give little insight or control on how this value is obtained. Moreover, none of the aforementioned tools is publicly available, nor the testbeds they have been run on, preventing the possibility of accurately reproducing previous works or extending them to broader scenarios. Finally, for every technique targeted at cellular networks, the evaluation is conducted in the downlink direction only (i.e. traffic flows from a remote host towards the mobile node); while this choice is justified by considering the higher volume of traffic flowing in the downlink direction, uplink analysis is always either omitted or left as future work. Nowadays the rise of Internet of Things (IoT) has given special importance to uplink performance; moreover, its reduced bandwidth with respect to downlink makes it a more compelling case for estimation, that is sorely missing from experimental literature.

With this work we address all these shortcomings, by (i) adopting only publicly available open-source tools, also releasing the measurements results; (ii) executing the experiments in a real-world testbed that is open to experimenters, and whose HW and SW environments are specified in details [6]; (iii) analyzing in details the effect of traffic policing rules on TCP throughput; (iv) experimenting on different Mobile Network Operators (showing significant impact on results); (v) targeting uplink (of interest for IoT, more bandwidth-limited, and largely neglected in experiments); (vi) discussing volume and time costs of measurements.

IV. Methodology and preliminary experiments

In this section, we outline the methodology used for the experiments, describing the tools employed and the considered scenarios, as well as preliminary analyses and tests that supported the experimental choices.

We have leveraged the platform provided by the MONROE project to evaluate the performance of ABw estimation tools in mobile scenarios, comparing their results with those of contextual TCP achievable throughput measurements. We focused on measurements at network and transport layers. This choice is suited to the considered scenario, with limited visibility on layer-2 technologies and no way to modify their configuration (similarly to most non-rooted commercial smartphones). Evaluation is conducted in the uplink direction, with traffic flowing from the mobile node towards a measurement server hosted in our laboratory in Napoli, that is connected to the Internet through a 100 Mbps Ethernet LAN and the 1Gbps backbone of the Italian Research Network Consortium (GARR)[4], likely positioning the bottleneck at the radio access link.

Given the hardware constraints of MONROE nodes, we first tested their traffic generation capability. Using the open-source D-ITG traffic generator [19, 20], we verified that nodes are capable of generating traffic at a rate (about 100 Mbps) higher than the nominal uplink bandwidth of 3G/4G networks.[5] This experiment—details are omitted for brevity—confirmed that the virtualized configuration of MONROE nodes does not impact their traffic generation capability, thus not hindering the experiments described in the following.

As a further preliminary analysis, we employed two available bandwidth estimation tools, namely Yaz and Pathload, to identify the one providing results closer to TCP achievable throughput measures provided by D-ITG (set to inject traffic in the network at a rate higher than the mobile link bandwidth). We refer to our previous works [9, 21] in the context of MONROE for further details on the characteristics of these tools. The *experiments* consisted in executing Yaz, Pathload (hereinafter ABW_1 and ABW_2), and D-ITG (*TAT*), each averaged on 4 consecutive *runs*. The order of execution of the different estimations follows the scheme: ABW_1–*TAT*–ABW_2, then 5 minutes interval, then ABW_2–*TAT*–ABW_1. This scheme is designed to even out possible dependencies on the order of execution of the tools and also to keep the same time interval between the ABw estimations and the throughput measurements used for comparison. The duration for each run for both Yaz and D-ITG during these preliminary tests was set to 60 seconds, whereas Pathload duration could not be fixed in advance, since it dynamically depends on the measured conditions of the network. These experiments were executed once per day for 4 different countries (Italy, Spain, Sweden, and Norway) for 8 consecutive days; they were scheduled during the night (time slot 0–3 AM, according to the local time), aiming at minimizing cross traffic and reducing variability. It is worth noting that four different operators (one per country) have been considered, thus considering a richer set of scenarios with respect to previous works which are usually limited to one operator and one country. We have also collected—leveraging the APIs made available by the platform—additional metadata related to GPS position, node CPU usage, and modem information, including signal quality indicators (such as RSSI and RSRP), cell ID, and frequency band. This information helped in the interpretation of experimental results (e.g. distinguishing 3G from 4G connections), and in excluding several possible perturbing effects (e.g. verifying that signal strength was not significantly varying in relation to measured performance, or CPU usage was not highlighting computation bottlenecks). The outcome of this preliminary campaign can be summarized as follows. Measurement results—in terms of discrepancies between TAT and ABw estimates—depend upon the considered operator and country. Comparing Pathload to Yaz estimates collected in the same measurement scenarios, we found that higher variability is exposed by Pathload results, as well as significantly higher differences from TAT than Yaz, especially in Italy and Spain, where available bandwidth estimated by Pathload greatly surpasses the achievable throughput. In addition, the performance of the tools is not affected by the order of execution, with consistent results over the whole measurement campaign.

This preliminary experimental campaign allowed to properly design the longer one. For the latter we only considered—using the same parameters described before for what concerns the number of runs and the time window—Yaz and D-ITG for evaluating ABw and TCP achievable throughput, respectively. The duration of each run was set to 20 seconds, thus allowing to reduce the amount of generated traffic without compromising the ability to evaluate the relationship between ABw and throughput as the above preliminary analysis has also shown that such duration is enough to reach a stable value in the received TCP throughput (i.e., to extinguish potential transitory periods) and also to obtain sufficient Yaz estimates. These experiments were run each day on the same nodes as before in four countries, for a total of 35 days, from February

[4]https://www.garr.it/en/infrastructures/network-infrastructure/backbones

[5]We highlight that using D-ITG, we are allowed to gain full visibility into both sender-side and receiver-side statistics about throughput.

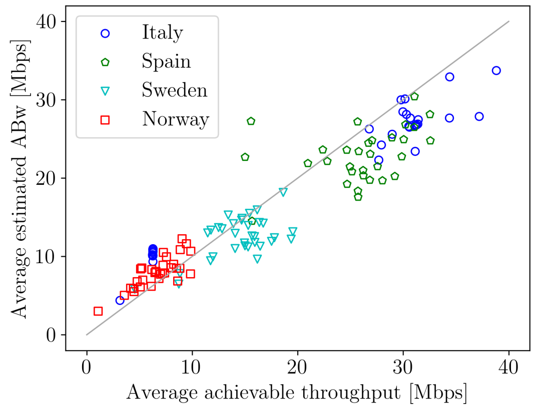


Fig. 1: Comparison between Available Bandwidth and TCP Achievable Throughput averaged estimates.

17^{th} to March 26^{th} 2018, excluding March 8^{th}, 9^{th} and 23^{rd} when the platform scheduler was under maintenance.

In Table I we summarize the details of the experimental campaigns discussed in this work, where the results of the *preliminary* one have been described in this section, while the others are discussed in the following.

V. Experimental results

In this section we report and discuss the results of the experimental campaigns detailed in the previous section.

A. Available bandwidth vs. achievable throughput estimates

We start by showing, in Figure 1, the average results for each day of experiments.[6] The x-axis reports the achievable throughput, while the y-axis the available bandwidth. We can see that each country tends to show similar values across different tests, considering both achievable throughput and available bandwidth. An exception to this behavior is represented by measures in Italy (for both achievable throughput and estimated bandwidth), which can be grouped in two classes: lower values, around 7–8 Mbps on the x-axis, and higher values, around 30 Mbps. This behavior can be explained considering that during experiments that reported lower throughput values the node was connected to a 3G network of the same operator, highlighting the unavailability of 4G connectivity for some experiments.[7] These two distinct behaviors also lead to different relationships between throughput and estimated bandwidth: when the average throughput is lower (using 3G), Yaz provides overestimates, while higher throughput (with LTE) leads to underestimation using available bandwidth. Yaz behavior for larger throughput is consistently verified in Spain too, since the tool is underestimating TCP throughput in almost each repetition.

To understand the reasons of the constant underestimation of bandwidth when the achieved throughput is higher, we analyze the time series of TCP throughput measured on the receiver, computed for 100 ms intervals, focusing on the two countries in which this phenomenon was observed, namely Spain and Italy. In Figure 2a we first show the time series plot from achievable throughput measurements in Spain. As can be seen, the behavior is consistent across the majority of the measurements, showing that throughput is increasing for about 10 seconds before reaching a stable value. Similarly, time series for Italy when using LTE are shown in Figure 2b. The results show a similar behavior, but here the rate is relatively more stable after 5 seconds; we also notice that the (under-)estimates in this case are generally closer to the achieved throughput compared to Spain.

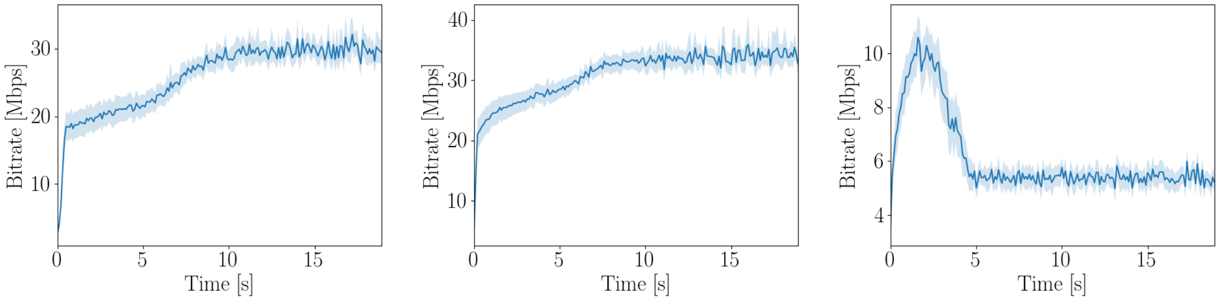
We also report as time series in Figure 2c the evolution of bitrates for Italy, focusing on the experiments in which a 3G network was employed. This plot provides an insight into why achieved throughput is consistently underestimated by Yaz during these tests. In this case we can observe an opposite behavior compared to what was noticed with LTE network in Italy and in Spain: during the first 10 seconds the throughput first rises until reaching its peak value, then decreases to a lower value, which is stable for the rest of the measurement period. Although this behavior is only observed when leveraging the 3G network, it is consistent over the repeated experiments: we speculate it is due to policy enforced by the network operator. When this phenomenon is observed, the estimated bandwidth is farther from the achieved rate, but it is actually closer to the received rate during the first part of the repetitions. Considering the possibility of a policy being enforced on a flow based on its duration, we specify that each available bandwidth measurement consists of many short-lived UDP streams (50 packets), sharing the same source and destination ports, with a typical inter-stream pause of ≈ 50 milliseconds, for the same total duration of a throughput measurement. Therefore it is unlikely that available bandwidth measurements are being discriminated by the policy enforcer according to the different time properties, while they could be on the basis of the transport protocol. This specific case could be further investigated modifying the bandwidth estimation tool to use TCP, but this is far from trivial and potentially requires the use of *raw sockets* (and thus super-user privileges, limiting the applicability), therefore we defer it to future works.

We can quantify the relationship between average *TCP Achievable Throughput* (TAT) and *Available Bandwidth* (ABw) estimates, reporting the distributions of their absolute *Relative Difference* (RD), defined as $RD = \frac{|TAT-ABw|}{TAT}$ and of the Absolute Difference which is not normalized on the average throughput.

The distributions are shown in Figures 3a and 3b. The figures show that Yaz estimates in Sweden are generally closer to the achieved throughput in terms of relative difference. For Norway, there are some cases where the achieved throughput is low (about 1 Mbps) and the estimates are around 2 Mbps, which leads to almost 100% of relative difference. Indeed, considering the absolute difference distributions, values reported for Norway are the lowest; since also the average throughput is lower, however, even a low difference has a significant impact on the relative difference. It should be noted, though, that in this case the difference in the estimate would not heavily affect applications only requiring a given high threshold for bandwidth. Finally, we note that absolute difference is larger in Spain, according to the throughput behavior described earlier, which shows a longer transient.

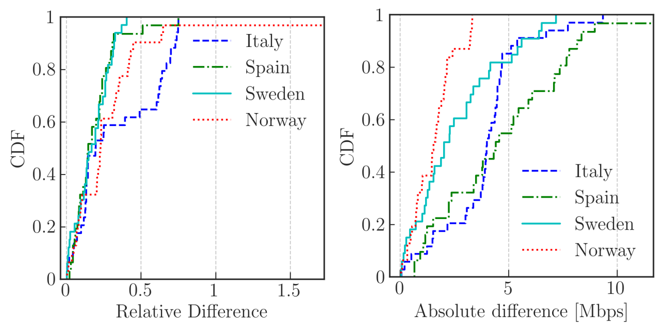
[6]Results for three days for Spain and Norway are missing, due to crossing traffic-cap thresholds (operator-specific and node-dependent) experimenters are subjected to, on MONROE platform.

[7]Thanks to the metadata collected during the experiments, we validated this inference for all cases.



(a) Spain. (b) Italy using LTE. (c) Italy using 3G.

Fig. 2: Time series for TCP achievable throughput. Observations from all experiments, time is calculated since the beginning of the experiment. Shaded area represents 95^{th} percentile bootstrapped confidence interval.



(a) (b)

Fig. 3: CDFs for the experimental difference between TCP achievable throughput measures and Available Bandwidth estimates ((a) Relative and (b) Absolute) in the different countries.

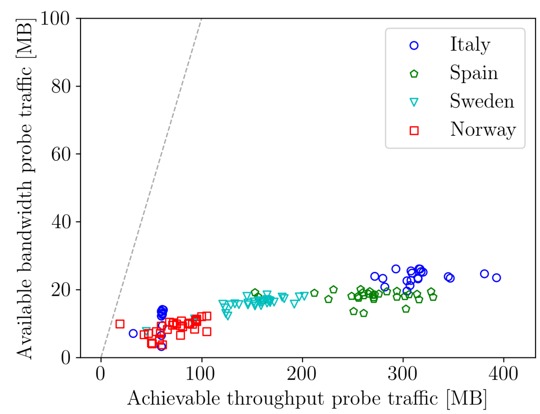


Fig. 4: Comparison between traffic generated by estimating Available Bandwidth and Achievable Throughput. Dashed line represents equality.

TABLE II: Average generated traffic volume for a measurement session (4 × 20-second runs).

Country	**Achievable Throughput (MB)**	**Available Bandwidth (MB)**	**Ratio**
Italy	210.38	18.61	11.30
Norway	72.83	8.54	8.52
Spain	263.03	17.94	14.65
Sweden	149.57	15.54	9.62

B. Generated traffic volume

In order to quantify the benefits introduced by using available bandwidth tools, we evaluate the intrusiveness of the tools for measuring the achievable throughput and for estimating the available bandwidth in terms of the volume of traffic they generate. This aspect is particularly relevant for MBB Networks, since the amount of monthly data is usually capped and exceeding traffic is expensive.

The traffic generated for measuring the achievable throughput and for estimating the available bandwidth for each measurement session is compared in Figure 4. As shown, all the points lie under the diagonal, witnessing that Yaz guarantees a considerable traffic saving. The points in the graph appear to be clustered according to the country/operator. 3G/4G connectivity also impacts the traffic generated by the two approaches. This is evident for Italy, where the all the points with x-coordinate less than 100 MB (and y-coordinate less than 20 MB) are related to experiments performed leveraging 3G connectivity.

In more detail, generated volumes grow accordingly, on average. However, as the performance of the network in terms of achievable throughput increases, the traffic generated by D-ITG to measure it grows significantly faster than that needed to estimate the available bandwidth with Yaz. This proves that the benefits of using ABw estimation as a proxy for the achievable throughput remarkably increase in networks with higher performance figures. Moreover, this analysis highlights how the traffic generated for estimating the available bandwidth ranges in a smaller interval (3–26 Mbps) with respect to that generated by achievable throughput measurements (19–393 Mbps). Indeed, points are more scattered along the x-axis than the y-axis, witnessing that the amount of traffic generated by the tool for estimating ABw is more predictable.

Table II summarizes this result, reporting the volume of traffic injected (on average) into the network when estimating the available bandwidth and when measuring the achievable

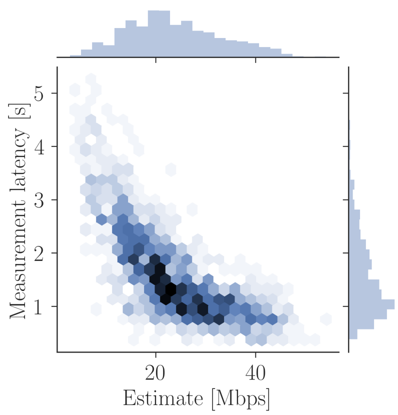


Fig. 5: Joint density distribution of estimate results and measurement latency for Spain. Marginal distributions are shown on top and on the right.

throughput. To ease the comparison, the ratio of these two quantities is also reported. Notably, in Spain, ABw measurement traffic volume is almost $15\times$ smaller than that required for achievable throughput tests. Our results suggest that, with the evolution of mobile networks and supported speeds overcoming 1 Gbps with the introduction of 5G, adopting available bandwidth estimation tools will be increasingly convenient.

C. Measurement latency

We recall that in the process of ABw estimation, each estimate is obtained after sending a variable number of streams. The estimation process terminates when the difference between local and remote packet spacing is below a given threshold. Therefore, the state of the network path under test impacts the number of generated streams, which in turn is algorithmically related to *measurement latency*, i.e. the time required to generate a single ABw estimate.[8] Accordingly, measurement latency—beyond being a metric that quantifies the responsiveness of the ABw estimation tool—also impacts the amount of measurement traffic. Therefore, it is an parameter worth to be investigated in relation to the bandwidth estimation process.

Figure 5 shows for a single country (Spain) the joint distribution of measurement latency and estimated values for ABw as obtained in our 35 days campaign. A clear trend can be observed, showing that the time to generate an estimate decreases for larger estimated ABw values. While this observation is partially justified by the algorithm implemented by the ABw estimation tool (that starts sending streams at higher rates, thus needing more streams to converge in the case of lower values of the ABw), this trend is probably also related to ABw variability over time (higher variability in case of poor bandwidth). We plan to deepen this issue in our future work.

In Italy, Sweden, and Norway estimates reported less variability and, although some of the few estimates requiring more than 3–4 seconds happen exclusively when the estimated value is low, the range is lower (about 10 Mbps). Average and standard deviation for measurement latency in each country are reported in Table III. In general, we can see that the average time required to generate an estimate with LTE ranges from 1.1 to 1.6 seconds according to the country, but standard deviation is not negligible, in general (up to 0.83 seconds).

[8]Our results show that the time needed for obtaining an ABw estimate is strictly related to the number of streams generated (a linear relationship between the number of streams and measurement latency is observed, as expected). Detailed results are omitted for brevity. This further confirms that the main factor impacting on estimates duration is the number of streams. No abnormal behavior has been observed (e.g., related to the establishment or the termination of the TCP connection needed by the tool for the signaling channel).

TABLE III: Mean and standard deviation for measurement latency in each country.

Country	Mean (s)	Std (s)
Italy (LTE)	1.15	0.52
Italy (3G)	0.79	0.32
Norway	1.32	0.79
Spain	1.60	0.83
Sweden	1.41	0.80

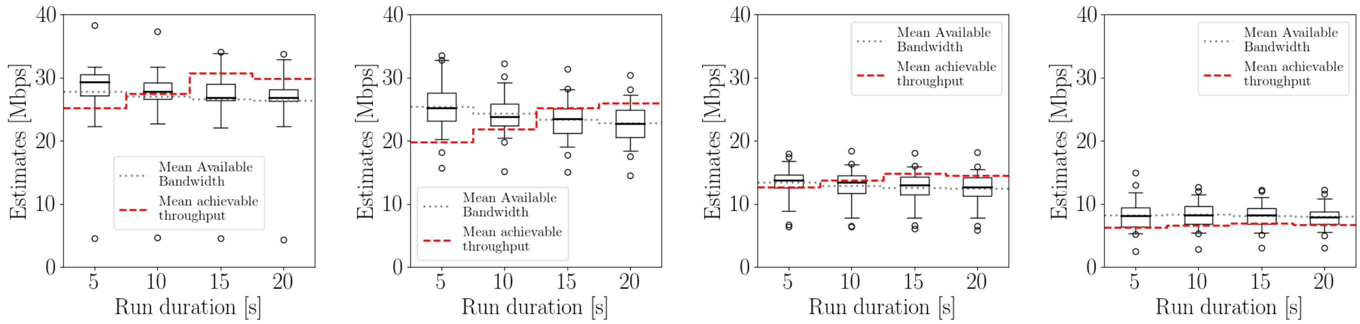
D. Measurement latency vs. accuracy

Intuitively, increasing the duration of measurement runs—despite the overhead increment in terms of the amount of generated traffic—is expected to provide more precise and reliable estimates. In the following, we analyze what is the impact of the duration of measurement runs on the difference between ABw estimates and throughput measures. For the purpose of inferring throughput by leveraging ABw estimation as a proxy, we hereinafter refer to the difference between these two metrics with the term *accuracy*, for the sake of readability.

Starting from the collected data, consisting of 20-second runs for both throughput and available bandwidth, we restricted the analysis of the results to shorter time intervals, thus evaluating the accuracy on varying run duration. In Figure 6 the results of each run obtained (in Italy, Spain, Sweden, and Norway, respectively) when considering a duration of 5, 10, 15, and 20 seconds is shown. It is worth noting that for this analysis results for Italy are restricted to experiments that leveraged LTE connectivity.

Results highlight different behaviors according to the country: (i) considering measurements conducted in Italy (Fig. 6a), the duration of the run has a small influence on both the variation and the median value of the estimates, while average throughput increases over time (as already discussed), being under the estimated bandwidth for the first 5 seconds and over it after 10 seconds; (ii) in Spain (Fig. 6b), estimates variation slightly decreases after 5 seconds, but durations longer than 10 seconds also lead to median estimates farther from the achieved rate; in this case ABw and throughput show opposite trends for growing run durations. (iii) in Sweden (Fig. 6c), nor variation nor median ABw estimates are heavily influenced by the duration, whereas throughput slightly increases; (iv) finally, results in Norway (Fig. 6d) show a slight decrease in the variation when increasing the duration, while median value is stable and very close to the mean one, about 2 Mbps above the average throughput.

In the light of the results shown, we conclude that increasing the duration does not necessarily improve the estimates in terms of accuracy (i.e. discrepancy with respect to the achievable throughput), as one would expect. Indeed, in all the cases



(a) Italy. (b) Spain. (c) Sweden. (d) Norway.

Fig. 6: ABw estimates percentiles (boxplots) and mean (grey dotted line), and TCP throughput (red dashed line) considering different run durations. Boxplots report 5^{th}, 25^{th}, 50^{th} (thick line), 75^{th}, and 95^{th} percentiles.

where throughput shows transient behavior, increasing over time, values of the two metrics are closer when considering a 10-second interval. Therefore, the estimated bandwidth can be effectively used as *proxy* for throughput when including this transitory. Regarding larger time intervals, the notable coherence of these discrepancies for each operator suggests the possibility to characterize the time evolution of throughput once, and offset the subsequent ABw estimations according to the result.

E. Whole-day hourly experiments

The results of the previous campaign have led us to analyze the characteristics of the observed behavior of TCP throughput in different conditions. We recall that this metric is typically used as reference for ABw estimation in mobile networks, as discussed in the related works. To this aim, we performed further experiments focusing on the following questions: (i) *is the behavior observed for the throughput dependent on the specific country/provider?* (ii) *is this behavior related to the hour of the day?* (iii) *is this behavior related to the day of the week?* In particular, we have designed an *additional campaign* (see Table I) in which we performed one test per hour on two different nodes, alternating achievable throughput and available bandwidth measurements (as in previous campaigns). Unlike in previous experiments, we restricted the duration of each repetition of both achievable throughput and ABw estimation to 10 seconds. This is in line with throughput behavior discussed in Sec. V-D. This allows to reduce both the duration of the experiments and the volume of traffic produced. We scheduled these experiments for 9 days (26^{th}–28^{th} May 2018, including both a weekend and one weekday, and 19^{th}–25^{th} June 2018, spanning over a whole week). Experiments have been run on two nodes located in Sweden, served by different providers and with sufficiently high data cap. This schedule allows us to answer each of the stated questions.

Figure 7 reports the evolution of achievable throughput and estimated available bandwidth over a single day for the two nodes (results related to other days have been omitted for brevity, as they showed behaviors analogous to the reported one, regardless of being weekday or weekend). Missing values for some of the hours are due to platform unavailability or errors while deploying the experiment.

The figures report with boxplots the distribution of throughput sampled every 100 ms on intervals of different duration to highlight the impact of traffic policing enforced by mobile operators, affecting the first ten seconds of each throughput measurement (see Section V-D). Note that the interval between 9–10 seconds reports the stable value for throughput.

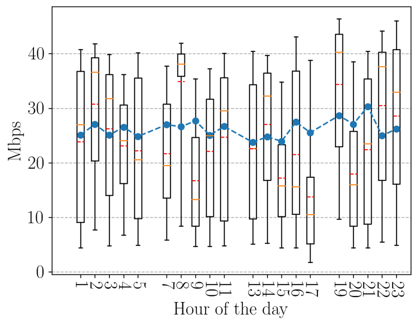
Comparing results between the two nodes, in the case of Node 1, higher variability has been encountered for the throughput (wider inter-quartile ranges). Both the ABw and the achievable throughput averages on the whole duration (Figure 7a) are close to $\approx$ 25 Mbps. On the other hand, in case of Node 2 variability is much limited with averages close to $\approx$ 5 Mbps (Fig 7d). Regarding the impact of traffic policing, results show that, while for Node 2 no major differences are evident among the different sub-intervals (Figures 7e and 7f), for Node 1 measurements evaluated on the shorter intervals depart from 10-second measurements.

This suggest that when operators implement mechanisms similar to those experienced by Node 2, shorter ABw measurements can be used whose results are representative of the stable behavior of achievable throughput. Moreover, in cases resembling Node 1, this does not hold, as a (constant) bias must be accounted for when using ABw to estimate achievable throughput on time duration beyond the initial transitory period.

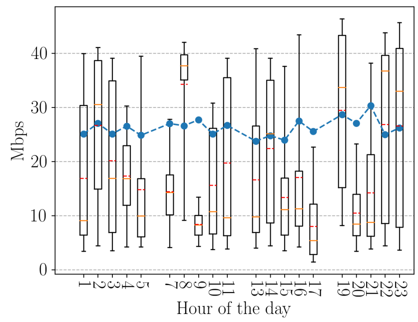
Regarding the questions that guided our design of experiments, in relation to the reported measurements we can state that: (i) the observed behavior of the throughput depends on the specific provider but does not vary; (ii) it is not evidently related to the hour of the day; (iii) it does not depend on the day of the week. These results prompt for further experiments, that have been partially limited by the traffic quotas available at the time.

VI. Conclusion

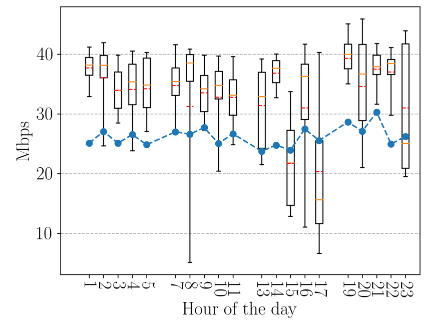
In this work, we have evaluated available bandwidth (ABw) estimation and investigated the relationship between ABw estimates and TCP achievable throughput. Indeed, TCP achievable throughput is a top-importance metric, but its measurement process is invasive and can put an heavy toll on the data quota (that is an expensive resource in MBB networks). The relationship between ABw and TCP throughput is strong, as witnessed by the scientific literature that used TCP throughput as a reference in many previous works investigating ABw. Less invasive in terms of generated traffic, and with a broader set of applicative scenarios, ABw is a good candidate *proxy* for



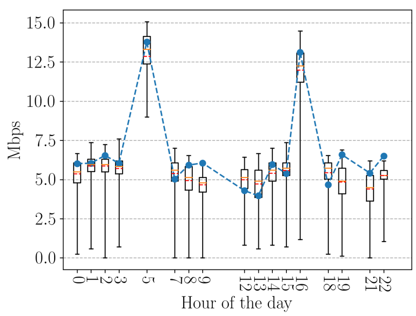
(a) Node 1, 0–10 seconds interval.



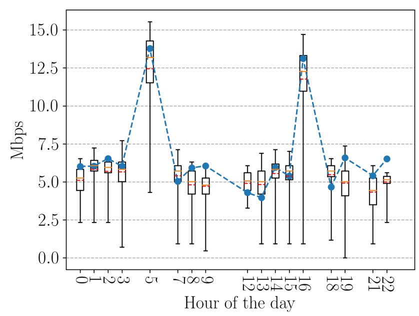
(b) Node 1, 0–5 seconds interval.



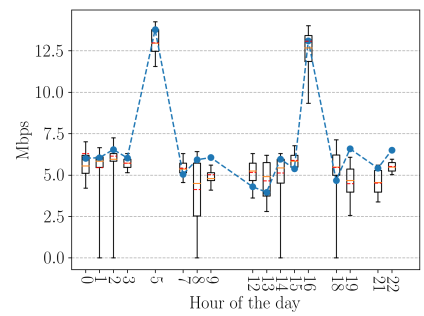
(c) Node 1, 9–10 seconds interval.



(d) Node 2, 0–10 seconds interval.



(e) Node 2, 0–5 seconds interval.



(f) Node 2, 9–10 seconds interval.

Fig. 7: Available bandwidth (dashed line) and TCP achievable throughput (boxplots with 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles), for different measurement intervals. First and second rows refer to Node 1 and Node 2, respectively. Missing boxplots reflect unavailability of the measurement nodes.

TCP throughput in MBB context, provided that the non-trivial relationship between these two metrics is investigated in depth.

Motivated by these considerations, in this work we have performed experimental campaigns in commercial MBB access networks, testing several providers and different network conditions. Our contribution is multi-fold. We report and analyze in detail an **experimental comparison** of ABw estimations and TCP throughput measurements on **several commercial 3G/4G networks in several countries**. Results have also highlighted that achieved throughput can show different profiles, in terms of time evolution of rate. This behavior has a strong effect on the evaluation of the accuracy of the ABw estimation tools, since results have shown that the difference between estimated bandwidth and throughput is smaller when referring to a 10-second measurement interval. We found that measurement latency for obtaining **a single available bandwidth estimate is** ≈ 1.5 **seconds**, allowing to perform multiple measurements in a short interval. We have verified that the amount of traffic generated using ABw tools is **more than one order of magnitude lower** than the volume required by achievable throughput tests, and **savings are higher when bandwidth is higher** (thus will likely further increase with future 5G deployments). We published both the **dataset** (in form of packet traces and log files) and the **open source tools** used to perform the measurements. Moreover the experimental platform we adopted (MONROE) is **open to third-party experimenters**, leverages open source software and is described in full detail.

In short, our results **quantified a complex relationship**, as the difference between these two metrics (i) varies according to the considered country and operator; (ii) depends on the time duration of the measurements run; (iii) can be heavily impacted by different traffic-engineering policies enforced by providers. These outcomes add further motives and strength to the cautioning put forth by Jain and Dovrolis [22] against using TCP throughput measurement as a ground truth for available bandwidth estimation. On the other hand, we found that, (iv) provided said considerations, the **difference between the two metrics exhibits consistency** over the experimental campaigns (extending in the order of weeks). As a consequence of this work, we envision and will investigate in future works a measurement procedure consisting in (rare) preliminary tests before bandwidth estimation, in order to determine the profile for throughput and accordingly offset the ABw estimates, allowing to use available bandwidth as a *proxy* for TCP achievable throughput, with a cost in terms of consumed traffic quota smaller by an order of magnitude, and also smaller impact on concurring transmissions on the path under measurement. We also plan to experimentally evaluate the impact of our findings on different application scenarios.

Acknowledgment

The research described in this paper has been partially funded by the SOMETIME project of the 1^{st} MONROE Open Call, funded from the European Union in the Horizon 2020 research and innovation programme under grant agreement No 644399 and art. 11 DM 593/2000 for NM2 srl.

REFERENCES

[1] M. Bernaschi, F. Cacace, A. Pescape, and S. Za, "Analysis and experimentation over heterogeneous wireless networks," in *First International Conference on Testbeds and Research Infrastructures for the DEvelopment of NeTworks and COMmunities*, Feb 2005, pp. 182–191.

[2] Y. Li, H. Deng, C. Peng, Z. Yuan, G.-H. Tu, J. Li, and S. Li, "icellular: Device-customized cellular network access on commodity smartphones," in *Proceedings of the 13th Usenix Conference on Networked Systems Design and Implementation*, ser. NSDI'16. Berkeley, CA, USA: USENIX Association, 2016, pp. 643–656. [Online]. Available: http://dl.acm.org/citation.cfm?id=2930611.2930653

[3] A. Annunziato, "5g vision: Ngmn - 5g initiative," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5.

[4] M. Jain and C. Dovrolis, "End-to-end available bandwidth: Measurement methodology, dynamics, and relation with tcp throughput," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 537–549, Aug. 2003. [Online]. Available: http://dx.doi.org/10.1109/TNET.2003.815304

[5] J. Sommers, P. Barford, and W. Willinger, "Laboratory-based calibration of available bandwidth estimation tools," *Microprocessors and Microsystems*, vol. 31, no. 4, pp. 222 – 235, 2007, special Issue with selected papers from the 11th {IEEE} Symposium on Computers and Communications (ISCC'06). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014193310700004X

[6] "The monroe project website," https://www.monroe-project.eu.

[7] R. Prasad, C. Dovrolis, M. Murray, and K. Claffy, "Bandwidth estimation: metrics, measurement techniques, and tools," *IEEE Network*, vol. 17, no. 6, pp. 27–35, Nov 2003.

[8] Ö. Alay, A. Lutu, R. García, M. Peón-Quirós, V. Mancuso, and et al., "Measuring and assessing mobile broadband networks with monroe," in *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, June 2016, pp. 1–3.

[9] G. Aceto, F. Palumbo, V. Persico, and A. Pescapé, "An experimental evaluation of the impact of heterogeneous scenarios and virtualization on the available bandwidth estimation tools," in *Measurement and Networking (M&N), 2017 IEEE International Workshop on*. IEEE, 2017, pp. 1–6.

[10] V. J. Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "pathchirp: Efficient available bandwidth estimation for network paths," in *Passive and active measurement workshop*, 2003.

[11] E. Goldoni, G. Rossi, and A. Torelli, "Assolo, a new method for available bandwidth estimation," in *2009 Fourth International Conference on Internet Monitoring and Protection*, May 2009, pp. 130–136.

[12] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '03. New York, NY, USA: ACM, 2003, pp. 39–44. [Online]. Available: http://doi.acm.org/10.1145/948205.948211

[13] M. Li, M. Claypool, and R. Kinicki, "Wbest: A bandwidth estimation tool for ieee 802.11 wireless networks," in *2008 33rd IEEE Conference on Local Computer Networks (LCN)*, Oct 2008, pp. 374–381.

[14] A. Farshad, M. Lee, M. K. Marina, and F. Garcia, "On the impact of 802.11n frame aggregation on end-to-end available bandwidth estimation," *2014 11th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON 2014*, pp. 108–116, 2014.

[15] L. Song and A. Striegel, "Leveraging Frame Aggregation for Estimating WiFi Available Bandwidth," *2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON 2017*, no. March, 2017.

[16] A. K. Paul, A. Tachibana, and T. Hasegawa, "Next-fit: Available bandwidth measurement over 4g/lte networks – a curve-fitting approach," *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, pp. 25–32, 2016.

[17] T. Oshiba, K. Nogami, K. Nihei, and K. Satoda, "Robust available bandwidth estimation against dynamic behavior of packet scheduler in operational LTE networks," *Proceedings - IEEE Symposium on Computers and Communications*, vol. 2016-Augus, pp. 1276–1283, 2016.

[18] N. Sato, T. Oshiba, K. Nogami, A. Sawabe, and K. Satoda, "Experimental comparison of machine learning-based available bandwidth estimation methods over operational LTE networks," *2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 339–346, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8024553/

[19] A. Botta, A. Dainotti, and A. Pescapè, "A tool for the generation of realistic network workload for emerging networking scenarios," *Computer Networks*, vol. 56, no. 15, pp. 3531–3547, 2012.

[20] D. Emma, A. Pescape, and G. Ventre, "Analysis and experimentation of an open distributed platform for synthetic traffic generation," in *Proceedings. 10th IEEE International Workshop on Future Trends of Distributed Computing Systems, 2004. FTDCS 2004.*, May 2004, pp. 277–283.

[21] G. Aceto, V. Persico, A. Pescapé, and G. Ventre, "Sometime: Software defined network-based available bandwidth measurement in monroe," in *Network Traffic Measurement and Analysis Conference (TMA), 2017*. IEEE, 2017, pp. 1–6.

[22] M. Jain and C. Dovrolis, "Ten fallacies and pitfalls on end-to-end available bandwidth estimation," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 272–277. [Online]. Available: http://doi.acm.org/10.1145/1028788.1028825