Unveiling the Missed 4.5G Performance In the Wild

Haotian Deng, Kai Ling, Junpeng Guo, Chunyi Peng
Department of Computer Science, Purdue University
[deng164,ling59,guo567,chunyi]@purdue.edu

ABSTRACT

In this paper, we study an important, yet unexplored problem in operational cellular networks: *How close is the actual performance each mobile device gets in reality from what the device could have got at best?* Fundamentally, this gauges the bridgeable performance gap for the device. It is deemed as the missed performance because the device fails to achieve the feasible, higher performance. In this work, we report our preliminary results that uncover, quantify, and understand such missed performance in the wild. We have conducted extensive measurements in the small city of West Lafayette, IN to quantify and analyze such performance gaps over four top-tier US carriers, all running 4.5G LTE-Advanced. Our study shows that, significant performance miss indeed happens and happens quite frequently. We have seen that, downlink throughput is merely 750 Kbps, while 47Mbps is available and feasible (60x miss) at one location; At more than 20% places, the missed data throughput (ratio) for four carriers is no less than 30 Mbps (7.4x, AT&T), 22.6 Mbps (6.8x, Verizon), 12.1 Mbps (6.9x, T-Mobile) and 25 Mbps (1.7x, Sprint), respectively. Even worse, such missed performance potentials can be readily reached with certain operations compatible with the current infrastructure (as confirmed by our study). We further analyze its root cause, and conclude that the current cellular practices on selecting serving cells should be held accountable.

CCS CONCEPTS

• **Networks → Mobile networks**; **Network manageability**; **Network mobility**.

KEYWORDS

Cellular Network; 4.5G; Downlink Throughput; Measurement

ACM Reference Format:
Haotian Deng, Kai Ling, Junpeng Guo, Chunyi Peng. 2020. Unveiling the Missed 4.5G Performance In the Wild. In *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications (HotMobile '20), March 3–4, 2020, Austin, TX, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3376897.3377857

1 INTRODUCTION

In recent years, mobile carriers have been heavily upgrading their network infrastructures to boost the raw system capabilities (e.g., carrier aggregation in 4.5G, and new radio in 5G for higher data rate). In this work, we argue that, it is equally important to better exploit the available and deployed capabilities to reach their *full* potentials for mobile users.

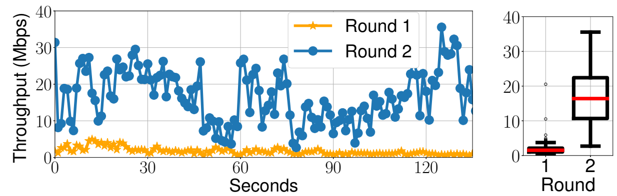


Figure 1: An illustrative example with >10-fold, missed throughput.

We thus study an important, yet largely overlooked performance problem: *What is the performance gap between what a mobile device actually gets in reality and what it could have possibly got at best, given the same operational network?* If this gap is small, the user gets what (s)he expects and deserves with the upgraded infrastructure. Specifically, when radio access technology is elevated from 4G LTE to 4.5G, the device should enjoy the enhanced data rate (say, up to 10s of Mbps), as long as the user device connects to the appropriate cell tower (called cell[1]).

Unfortunately, our measurements show that the gap is not small, at least not always small in operational mobile networks. We find that, infrastructure upgrade might not always lead to enhanced performance for end users. The enhanced raw system capabilities are not fully exploited, and remain to be underutilized. The fundamental problem lies in the performance gap between the actual performance a mobile device receives and the achievable performance it could have obtained at best. This gap is deemed as the *missed performance*, since the device fails to achieve the higher, feasible performance given the same infrastructure. We notice that the operators never promise best performance to mobile users and they may even intend to offer sub-optimal performance; For example, the operator selects a cell that has lighter load but afford poorer performance to serve the user device for the sake of load balancing. In this study, we take the user-centric perspective and aim to explore the upper bound of performance achievable on the device side. We further delve into its cause analysis, and confirm that the better performance, as indicated by the gap, can indeed be reached for the mobile device, with simple, device-based operations without any infrastructure change.

In a nutshell, we have conducted the arguably first study to unveil and understand the missed performance in the wild. We focus on the downlink throughput as the main performance metric, while examining other measures such as latency in the future work. We have made three contributions.

First, we use real-world instances to expose that *a significant portion of performance gain is indeed missed.* Figure 1 shows an illustrative example, where the phone receives 1.5 Mbps from AT&T on average, whereas 16.9 Mbps is feasible at the same spot (S1 in

[1]A cell tower uses directional antennas and distinctive frequency bands to accommodate multiple cells, each serving several mobile devices within its radio coverage.

HotMobile '20, March 3–4, 2020, Austin, TX, USA

ACM ISBN 978-1-4503-7116-2/20/03.
https://doi.org/10.1145/3376897.3377857

Figure 3). We run experiments in two rounds, where the phone downloads the same large file from our server. Round 1 uses the default network operations. It starts once the user walks to S1 along a pre-determined route. Round 2 runs immediately after Round 1 with resetting its mobile network (i.e., turning on the airplane mode and then turning off). Figure 1 plots the downlink speed in one experimental run, with at least 10-fold ((16.9-1.5)/1.5) throughput reduction. Such large speed gaps are repeatedly observed at S1 in 20+ experimental runs, over different days and across different hours of the day. Two implications immediately follow. First, the device indeed misses a large portion of feasible downlink speed. The missed throughput can be even bigger because the "best" one is no smaller than that observed in Round 2 (see §3.2 for details). Second, the large gap is persistently observed; it is unlikely to be incurred by transient factors (say, varying radio signal quality or network congestion). We will validate and refine these findings.

Second, we conduct a city-scale measurement study to quantify the missed performance in the wild (§3). Note that our study is nontrivial given that it is difficult to learn the ground truth on the "best," available throughput over the operational networks. To this end, we have devised a new measurement and analysis methodology to approximate the missed performance gap (§3.1). Our effort reveals that, *significant performance miss is not uncommon.* It is typical to miss the "best," feasible performance, thus under-utilizing the available capabilities of mobile networks. In our measurements (740hr, 8,756 Km) with four US mobile carriers, we observe that the missed throughput exceeds 30 Mbps (7.4x), 22.6 Mbps (6.8x), 12.1 Mbps (6.9x) and 25 Mbps (1.7x) at more than 20% of locations in AT&T, Verizon, T-Mobile and Sprint, respectively.

Third, we further delve into the root cause analysis (§4). Our study shows that, *the current practice by mobile operators on selecting serving cell(s) should be held accountable.* In the above example, we find that, two distinctive sets of serving cells (set 3 for Round 2 and set 12 for Round 1 at S1 as shown in Figure 4) are chosen at the same location. Note that which cells serving the phone plays a decisive role in its achievable throughput. Mobile operators offer more choices of serving cells at most places, thanks to their infrastructure upgrades with dense cell deployment. However, their practice ironically might not select the "optimal" or better serving cells that afford higher throughput. Consequently, their deployed raw capabilities remain largely under-utilized.

We finally present the gained design insights, and discuss open issues, our ongoing effort and the implications for 5G (§5). In summary, our work sheds light on a novel approach to boosting user performance without infrastructure changes. Our collected datasets is released to the research community [3].

2 BACKGROUND & RELATED WORKS

Carrier aggregation. We find that all US carriers have advanced from LTE (4G) to LTE-Advanced (4.5G) in our test areas. One core technology advance lies in carrier aggregation, as illustrated in Figure 2. Carrier aggregation allows more than one serving cells, each running over one frequency carrier (referred tocalled component carrier), to serve one user device. Logically, the set of serving cells offers radio access over one aggregated carrier (over contiguous or non-contiguous spectrum), which boosts bandwidth up to 100 MHz

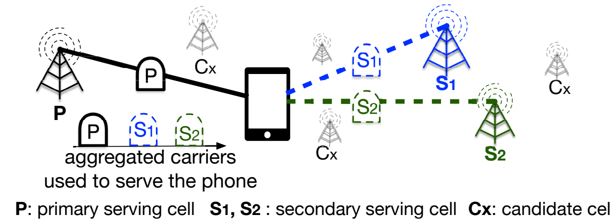


Figure 2: Carrier aggregation in 4.5G LTE-Advanced.

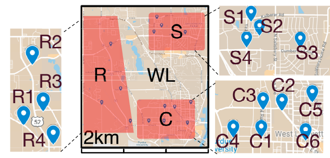


Figure 3: Map of test areas.

	Time	Distance
AT&T	166.3 hr	2,182 Km
Verizon	290.4 hr	2,734 Km
T-mobile	186.9 hr	2,378 Km
Sprint	96.3 hr	1,461 Km
Total	739.9 hr	8,756 Km

Table 1: Our dataset.

by aggregating at most five carriers each at 20MHz and thus potentially enhances capacity and data performance [4]. Each set consists of one primary cell (PCell) and n ($n = 0, 1, \cdots, n_{max}$) secondary serving cells (SCells). In our study over all four US carriers, we observe up to two SCells, which matches with the limit regulated by 3GPP standard [6].

Selecting serving cell(s). Selecting or changing the serving cells out of all candidate cells near the device is indispensable to radio access quality and data performance. The selection of these serving cells is realized by several standard procedures such as *cell selection* when the device powers on or turns off airplane mode, *cell reselection* or *handoff* when the current serving cells fail or perform poorly (e.g., on the go). Without loss of generality, we examine one generic function out of all procedures: how to select the set of serving cells. It is implemented in Radio Resource Control (RRC); When carrier aggregation is in use, two steps are performed: PCell selection and then SCell selection which is managed by the PCell.

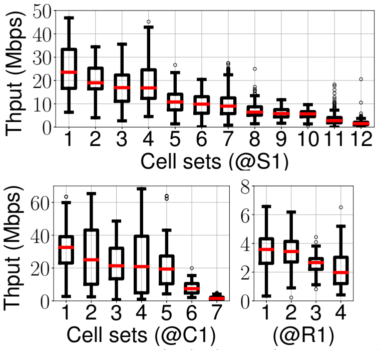
Related work. There is no prior work to investigate missed performance in cellular networks. We briefly introduce most relevant work on measuring data performance and selecting serving cells. A number of studies have measured data performance in cellular networks, e.g., [10–12]. But they all examined the achieved performance, without any attempt to explore its feasible upper bound. Moreover, early studies (before 2017, e.g.,[10, 11]) were not even conducted in commercial networks, but over the lab-scale prototypes or simulations. Recent studies (e.g.,[12]) measured performance in the wild but worked on 4G, not 4.5G without evaluating carrier aggregation. Our cause analysis is somehow related to our previous handoff studies [7, 8]. But [7, 8] examined how handoff is performed and why, not considering its impact on data performance.

3 MEASURING MISSED PERFORMANCE

In this section, we present our study to quantify how vastly and frequently the missed performance is observed in the wild.

3.1 Methodology and Dataset

In this work, we conduct a city-scale measurement in West Lafayette, IN (a 6 km × 6.7 km area marked in Figure 3). We focus on data performance in terms of downlink throughput while downloading a big file (500MB) from our lab server. We notice that there are many other performance metrics like uplink throughput, latency or



@S1	PCell (freq)	SCell$_1$ (freq)	SCell$_2$ (freq)
1	244 (2000)	102 (2425)	166 (850)
2	166 (850)	102 (2425)	244 (2000)
3	102 (2425)	426 (2175)	166 (850)
4	102 (2425)	244 (2000)	166 (850)
5	45 (9820)	426 (2175)	102 (2425)
6	447 (5145)	426 (2175)	244 (2000)
7	166 (850)	447 (5145)	45 (9820)
8	426 (2175)	102 (2425)	166 (850)
9	197 (850)	104 (2425)	396 (2175)
10	426 (66661)	-	-
11	166 (850)	447 (5145)	-
12	253 (5145)	52 (2000)	55 (2175)

@C1	PCell (freq)	SCell$_1$ (freq)	SCell$_2$ (freq)
1	412 (850)	424 (2425)	43 (2000)
2	422 (5145)	398 (850)	413 (1150)
3	363 (2425)	398 (850)	59 (9820)
4	363 (2425)	398 (850)	44 (2000)
5	43 (2000)	424 (2425)	412 (850)
6	363 (2425)	59 (9820)	413 (1150)
7	422 (5145)	-	-
@R1			
1	394 (5145)	394 (2175)	40 (2000)
2	394 (5145)	40 (2000)	394 (2175)
3	394 (5145)	-	-
4	394 (5145)	394 (2175)	-

Figure 4: Downlink throughput as well as their sets of serving cells at 3 locations: S1 (suburban), C1 (campus) and R1 (rural) in AT&T.

application-specific metrics (e.g., jitters and stalls in video streaming), and leave them as our future work. We use eight Google Pixel phones (Pixel 2, 2 XL and 3) for four US carriers (A, V, T, S). No device-specific results are observed and the data from different devices are combined. We use `tcpdump` to collect packets captured on the phones. To analyze and understand the root cause of missed performance, we use `MobileInsight` [2, 9] to collect signaling messages exchanged between the phone and the network (including radio resource control messages and physical layer messages on radio signal/quality measurement and reporting) as we did in [7]; These messages are used to monitor the set of serving cells (PCell, SCells) and other neighboring cells being measured and learn how handoff is performed.

We first run experiments at certain locations to get a glimpse of missed performance in reality and then run driving tests for a larger scale measurement. In static tests, we randomly choose 14 locations in three representative zones: campus (aka, urban), suburban (residence) and rural, as marked in Figure 3. At each test location, we first run static tests under default network operations at different hours of the day and at different days; In addition, we exploit extra actions allowed at the device side (e.g., turning on/off airplane modes, configuring the preferred network or frequency band) to disturb the default operations. We find that this makes it possible and fast for us to observe data performance distinct from (sometimes significantly outperforming) what the device gets by default. We observe that more than one serving cell sets are used to serve the device at the same location and the perceived performance varies with distinct serving cell sets. We cluster data performance by its serving cell set and later show that missed performance is associated with its serving cell set (§3.2).

Afterwards, we run driving tests to check missed performance at a larger scale, across the whole city. We drive along **every** road in West Lafayette, and divide these roads into small grids (each approximately 55m x 42m) and retrieve missed performance per grid. We drive along a variety of routes to cover each road (grid) multiple times (main roads: ≥ 20, almost all local roads: ≥ 5). We admit that the missed performance might be under-estimated given a limited number of samples; We are unable to measure the best and worst performance at each grid in our limited measurement; However, even given such a conservative measurement, the significant performance miss is indeed observed in our study, and quite frequently (§3.3). Finally, we figure out that the set of serving cells plays an essential role in missed performance. We further run more driving tests, not with heavy file downloading, but with mice traffic (ping Google every 50 ms) in the background for the cause analysis (§4). We conducted this measurement study from July to Oct. 2019, primarily in Aug 25 – Oct 15. Our dataset covers experiments for about 740 hours (in both static and driving tests) and over 8,756 Km (in driving tests only) (Table 1); It is available at our website [3].

3.2 Missed Performance Is Not Rare

We first use static tests in AT&T to examine how large the missed performance can be at different locations. Similar results are observed for other three carriers unless specified. We find that missed performance is not rare; Significant miss in performance is observed at many test places.

How large can the missed performance be? We attempt to answer this question first through the tests at three selected locations in three typical zones: S1 (suburban), C1 (campus) and R1 (rural). We record downlink throughput (per second) and group these samples according to their serving cell sets in the boxplots of Figure 4. The serving cell sets are given in the right table; Each cell is marked by its short ID and frequency channel number. The mapping from a channel number to its frequency spectrum band is specified in [6] and can be found online (e.g., via [1]). We exclude those rarely observed sets (with the sample size < 60). We see that throughput fluctuates (in the range of these boxplots). This matches with our common expectation because data is collected at many runs at different days or at different hours of the day, affected by varying radio signal quality and network conditions.

Despite of these fluctuations, we still clearly see that the set of serving cells plays a decisive role to the performance perceived at locations S1 and C1, despite a less important role at R1. Both S1 and C1 have dense cell deployment; 13 unique serving cells in 12 sets are observed at S1 and 9 cells in 7 sets at C1. It is not hard to understand. Thanks to heavy infrastructure upgrades and more spectrums recently acquired, opeartors have deployed many more cells than before [8]. If the first two (five) sets of serving cells are used at S1 (C1), data speed can be fast, say, ≥ 20Mbps (median). Unless specified, we use the median value afterwards. The highest speed via set 1 can go up to 23 Mbps at S1 and 32.6 Mbps at C1. However, the good sets may not be always chosen in practice; The devices at these locations can be served by the '*worse*' choices. As a result, the median data speed shrinks to 1.5 Mbps (via set 12) or below 7 Mbps (via sets 8 -12) at S1; At C1, it reduces to 1.5 Mbps (via set 7), missing a much higher speed (> 30 Mbps).

We want to point out that the performance comparison at S1 in Figure 4 shows a larger performance gap than the one observed

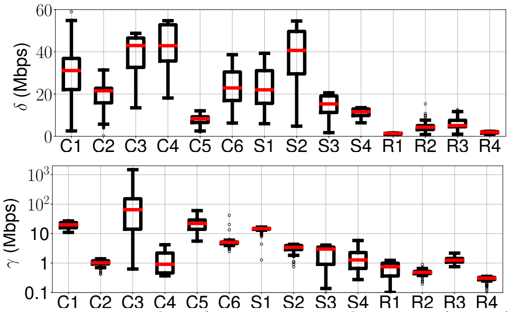


Figure 5: Missed performance at 14 locations (AT&T).

in our first example in Figure 1, where round 1 is one run via set 12 and round 2 via set 3. That is, at S1, the missed performance (set 1 vs set 12) is even larger; In round 1, the median data speed decreases from 23 Mbps to 1.5 Mbps, missing 14.3-fold instead of 10-fold (from 16.9Mbps to 1.5 Mbps).

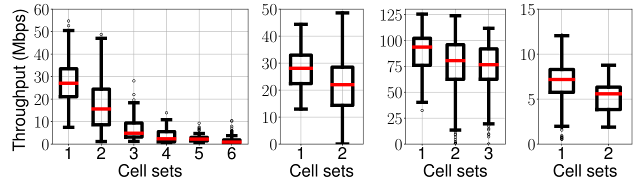
In the rural areas (here, at R1), we observe that the performance gap is much smaller. This matches with our expectation. Many fewer cells are deployed in the countryside (also validated in §4). At R1, there are only three cells available. As a result, there is much fewer sets of serving cells. The fewer choices, the less likelihood to miss the good cells to serve the devices. In this instance at R1, cell 394 is always chosen as the primary cell with slightly different SCell combinations.

How common and how large is missed performance in the wild? We next extend our study to all 14 test locations to see that missed performance is not the corner case. We define two metrics to quantify the missed performance at each location. We first locate the best and worse set of serving cells based on their median performance. We further compare their performance at the same percentile using the absolute and relative gaps:

$$\delta_\rho = P_\rho^{best} - P_\rho^{worst}, \gamma_\rho = (P_\rho^{best} - P_\rho^{worst})/P_\rho^{worst}, \qquad (1)$$

where ρ is the percentile from 0 (min) to 100 (max), P_ρ^{best} (P_ρ^{worst}) is the ρ-percentile of the data throughput using the best (worst) serving cell set. We notice that data performance fluctuates and the way to determine the best/worse sets using a single value (here, the median) is not reliable. It may be questionable to locate the best and worst sets without statistically significant difference at some locations (e.g., at R1). This is why we introduce relative comparisons at all the same percentiles to make such comparison robust. Note that, if we fail to located the best and worst sets, the calculated performance gap underestimates the missed performance in reality.

Figure 5 plots the distributions of δ_ρ and γ_ρ at 14 locations. It confirms our above findings at more places. Big performance gaps are widely observed in campus and suburban areas whereas the gaps are much smaller at rural areas. Specifically, δ_ρ goes up to 42 Mbps at C3 and C4; It is more than 20Mbps at 7 out of 10 locations (70%) in campus and suburban areas. The gaps are slightly larger in campus than in suburban. In rural areas, δ_ρ is more than 5 Mbps at two out of four locations. This is consistently observed in terms of the metric γ_ρ. For example, γ_ρ is more than 60 at C3. We take a closer look and find that this is because the worst performance at C3 is really bad (best vs worst: 47Mbps vs.750 Kbps). From the user perspective, such low data speed hurts user experience, especially when 47Mbps is actually available but missed by the carrier's network operation.



(a) S4 (V) (b) S2 (T) (c) S2 (S) (d) R4 (S)

Figure 6: Throughput at selected locations in V, T, S carriers.

We also note that these two metrics are not fully correlated. For example, small δ but large γ (22x) is observed at C5, which implies that the absolute throughput in the worst case is low (370 Kbps); Namely, significant improvement is possible even when the best performance observed is modestly good (8.9 Mbps). In contrast, we observe large δ_ρ but small γ_ρ (γ_{50} =1x) at C2; This implies that the throughput in the worst case is not that small but the absolute improvement value is still significant (19.5 Mbps vs. 40 Mbps).. We would like to highlight that $\gamma_{50} > 1$ (namely, the worst < 50% of the best) implies considerable improvement potential. We observe $\gamma > 1$ at all the test locations in campus and suburban areas; Even in rural areas, γ is larger than or almost close to 1 at two out of four locations (R1 and R3).

Carrier-specific findings. We observe three carrier-specific results in our study. Due to space limit, we illustrate the key observations using four instances in Verizon, T-Mobile and Sprint (Figure 6). The findings are confirmed in our larger-scale measurement described next (§3.3). First, we find that the number of unique serving cells per location is much smaller in T-Mobile and Sprint, while it is comparable in Verizon. This is because T and S deploy fewer cells than other two carriers. Specificially, only 2 (3) sets are observed at S2 in T-Mobile (Sprint). Similar results are observed at most locations in campus and suburban. Second, the measured performance gap between the best and worst case is much smaller in T and S because of fewer choices of serving cells. Compared to A and V, T and S have much smaller γ_{50} (at S2, T: 27% = (28-22)/22, S: 22% = (93.4-76.5)/76.5)). We note that δ_{50} in Sprint is not that small (16.9 Mbps at S2); This is because Sprint has much larger data speed than other three carriers. We check all other locations and find that the maximal speed in Sprint goes up to 160 Mbps, which is the fastest in four carriers. Third, all the results in rural areas are consistent across all the carriers. Gaps are much smaller, even in terms of δ_ρ for Sprint. This is because its data speed in rural areas is not fast (for example, 7.2 Mbps at R4).

3.3 Performance Missed Everywhere

We further check how widely missed performance exists at a larger scale, say, at a city scale. Certainly, the above static tests do not scale. We thus run driving tests many times across every road and learn performance gaps per grid (roughly, 55m x 42m). For each grid, we perform the above analysis to learn the best and worst serving cells sets, and use δ_ρ and γ_ρ to quantify the missed performance. We plot the cumulative distribution function (CDF) of δ_{50} and γ_{50} (the median of estimated gaps) across all the grids in Figure 7.

Before we present our findings, we would like to emphasize that such driving measurements may largely underestimate the missed performance in reality. Compared with static tests, we do not take any extra actions on the device to affect network operations.

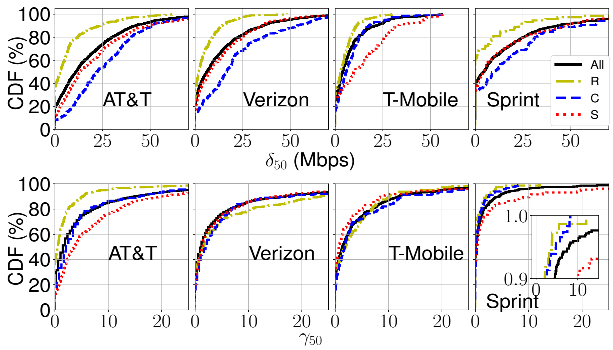


Figure 7: CDF of δ_{50} and γ_{50} over all the grids in driving tests.

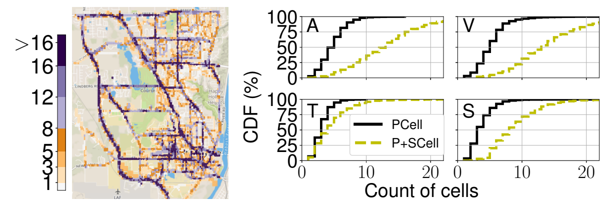
Consequently, the observed sets of serving cells are only induced by mobility and the optimal performance at each grid is highly likely undervalued; The "best" set of serving cells may not be chosen under the current and default operations, no matter how many times we drive across these grids. That is, our quantification in the driving tests are likely more modest, compared to those in static tests.

Even so, we see that significant performance gaps between the 'best' and the 'worst' cases frequently occur in reality, in all four carriers, despite possible underestimations. Specifically, in AT&T, we see that δ_{50} > 25 Mbps at more than 43%, 29%, 9% and 26% grids in campus, suburban, rural and all areas. In other three carriers, we observe δ_{50} > 25 Mbps at more than 17%, 6% and 20% places out of all the grids, in V, T, S, respectively. More than 20% grids in all the zones observe δ_{50} over 30 Mbps (7.4x), 22.6 Mbps (6.8x), 12.1 Mbps (6.9x) and 25 Mbps (1.7x) in A, V, T, S, respectively. These results are consistent with our previous findings in §3.2. T-Mobile, compared to other three carriers, has a relative smaller δ_{50} because its absolute speed is smallest at more places. However, in terms of γ_{50}, the relative gap is not small in T-Mobile. More than 10% grids have γ_{50} larger than 15.3x, 17x, 12x and 4.6x in A, V, T, S, respectively. At more than 50% grids, γ_{50} is larger than 1.8x, 1.2x, 1.4x and 27% in A, V, T, S, respectively. γ_{50} is relative small in Sprint because the absolute throughput is large in campus and residence areas but the gap is not.

We also notice two new findings. First, no gaps are observed at 20%, 20%, 25% and 39% locations for A, V, T, S, respectively. This is because we observe that the phones are always served by the same serving cells at these locations. This indicates that more than one serving cell sets are observed at other places. It depends on each carrier's cell deployment. This is also confirmed in our following study on the serving cell dynamics (Figure 8). Second, the missed performance generally decreases from campus to suburban and then to rural. The only exception is T-Mobile. We find that the best case performance is comparable in both regions but the worst one is much worse in the suburban areas.

4 UNDERSTANDING WHY

In this section, we present our preliminary efforts to learn the why behind the missed performance. We focus on one core question: *why is the 'best' or 'reasonably good' set of serving cells NOT chosen in those poorly performed runs?* We find that the current network operations in serving cell selection should take the blame. In particular, when carrier aggregation is enabled, the selection of serving



(a) Counts of P+SCells (A) (b) Serving cells per grid

Figure 8: Number of unique serving cells observed in our study.

cells is done in two steps: PCell selection and SCell selection which is decided by the selected PCell. The first step is managed by the network with the assistance from the mobile device, via a standard procedure as regulated in [5] and studied in our prior work [7, 8]: The network first configures parameters for measurement, reporting and cell selection, and then the device performs measurement and reports measured results to the network as configured; The network decides whether to change or keep the currently PCell and executes its decision. The second step is determined by PCell and no standard procedure or policies have been disclosed. We identify two issues in network operations:

1. *These good choices of PCells are missed because current network operations are largely designed for seamless connectivity instead of the best data performance.*

2. *These good choices of SCells are missed because PCells can not choose SCells out of all available choices; Instead, they are restricted to a smaller pre-configured group.*

Missing the best or good PCell(s). We first show that operators have deployed dense cells at most places. Figure 8 shows the CDF of the number of unique serving cells (PCell only, and PCell + SCell) observed per grid in our driving tests. We use the map of observed PCell counts in AT&T (Figure 8a) to illustrate its geographic distribution. Similar results are observed in other carriers (omitted due to space). We see more than 5 PCells (13 P+SCells) at more than 50% grids in A and V; T and V have less denser deployment and we still see more than 3 PCells (4 P+SCells) in T and 3 PCells (8 P+SCells) in S at more than 50% places.

However, such merit of denser cell deployment does not yield better performance in every case, despite more choices enabled. We observe that PCell selection is primarily determined by its radio quality evaluation, as disclosed in our previous studies [7, 8]. Take S1 (AT&T) as an example (Figure 4). We check top four sets of serving cells at S1, excluding those observed in poor cases (sets 5-12); Apparently two cells are among top choices for PCells at this location: 244 (2000) and 102 (2425). We exclude cell 166 (850) because the performance using set 11 and set 7 are not so good; We gauge that performance at set 2 is likely contributed by SCells (by cells 244 and 102). We also notice that the measured radio quality for each cell is consistent (with small fluctuations) across almost all the runs, unless specified. Specifically, we see that signal strength/quality (RSRP/RSRQ) of cell 244 at S1 is measured in range of (-114dBm, -117dBm) (RSRP) and (-14dB, -16dB) (RSRQ). Those runs without selecting cell 244 (i.e., sets 5 and 7–12) is because the initially chosen PCell has stronger radio strength/quality and/or its measurement/reporting does not lead to a cell re-selection. For example, the RSRP (RSRQ) of cell 253 (selected as PCell in set 12) is -105dBm (-14 dB), stronger than the one of cell 244. As a result, cell 244 is not considered. In another example, the RSRP (RSRQ)

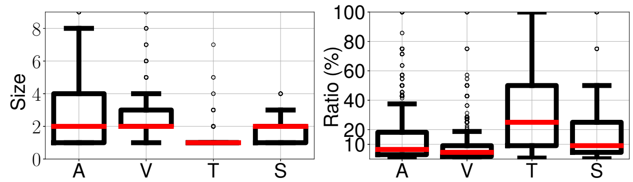


Figure 9: Limited SCells are selected per PCell in 4 carriers.

of cell 45 (selected as PCell in set 5) is measured to be -116dB (-18 dB), which is slightly weaker than the one at cell 244. However, its measurement reporting is configured to be triggered only when the candidate cell is 5dB stronger (RSRQ), which is still not satisfied in this case. This indicates that the current operation based on the radio signal strength/quality evaluation may fail to choose the cells that offer good performance. We guess that the reason behind such radio quality evaluation is simple to implement; After all, the located cell at least ensures seamless radio coverage, which is critical at the early phase of cellular networks when full coverage was a big concern of all the operators. We observe similar results at other locations. PCell selection depends on initial choice and its subsequent radio evaluation. Unfortunately, the cell with best performance may not be chosen in this process.

Missing the best or good SCell(s). We also notice the power of carrier aggregation is not fully utilized. It is expected that PCell should select those best surrounding cells (or those with strongest radio quality) as SCells, but it does not. Instead, for a given PCell, we find that SCells are selected from a very-limited subset which seems pre-configured. In the above example at S1,13 serving cells are observed; This indicates that they at least pass the radio evaluation check. Theoretically, there are $C_{n-1}^2 + C_{n-1}^1 + C_{n-1}^0$ options of SCell combinations, where n is the number of serving cells observed at one location (here, $n = 13$). However, we do not observe such a large number of SCell combinations. For example, PCell 244 has only one SCell combination (here, cells 102 and 106) out of 79 options. Figure 9 plots the CDF of the size of SCell combinations (across all the grids) per PCell (left), as well as the CDF of the ratio of the observed size over all the possible options per location grid (right). We see that PCells are restricted to a much smaller group of SCells: 50% of PCells have no more than 2 SCells combinations in all four US carriers (1 in T-Mobile);; The ratio is smaller 10% in all carriers except T-Mobile. As a result, it fails to make full use of the second chance to chase for better performance.

5 CONCLUSION AND DISCUSSION

We present the arguably first study to unveil missed performance in operational cellular networks. Our measurement over 4 US carriers shows that missed performance indeed happens and happens a lot. We pinpoint the root causes into today's network operations on selecting cells.

As a first attempt, there remain many open issues. We briefly discuss them and our ongoing efforts, as well as gained implications. First, we do not argue that operators have to select cells for the sake of best performance to mobile users. They reserve their rights to determine selection strategies, as ISPs do with BGP. The operators may intend to sacrifice user-perceived performance for some other reasons, e.g, load balancing for network-side optimization. We also admit that users sometimes may not care best performance (e.g., downloading at 100Mbps or 20Mbps is fine). However, our aim is to explore the limit of current networks. We investigate missed performance potentials from the user perspective. We aim to uncover that user-perceived performance is sacrificed and such degradation is sometimes unnecessary. We hope to call for attention for making full use of those deployed infrastructure resources for better performance. We believe that it is eventually aligned with both operators' and users' interests. Second, we believe that the problem will not disappear as 5G proceeds. While we are unable to measure missed performance in 5G (no large-scale rollout yet), the identified problem still exists in 5G and is likely worse with much denser deployment and larger performance gaps contributed by advanced technologies (e.g., 10Gbps vs tens of Mbps). Third, this study focuses on downlink throughput. Other performance metrics like latency requires a completely new study by running different data applications instead of elephant data flows, and we hence leave them as our future work. We also note that performance variance caused by transient factors (e.g., radio and network loads) is not completely tackled in the missed performance evaluation. However, our measurement study shows that the serving cells is playing a persistent role and dominates data performance, despite variance per serving cell set. Last, some real-world instances (e.g., Figure 1) demonstrate that missed performance can be easily avoided through certain client-side intervention (here, resetting the mobile network). However, it is not all the case. We observe that resetting the network may not always select better cells where these cells are missed and more efforts are warranted to new design solutions. We see that the device is able to learn and profile data performance per serving cell set and take certain actions to get the desired cell set (and performance). It implies proactive device-side actions open up a promising approach.

Acknowledgments. We greatly appreciate our shepherd Prof. Mahadev Satyanarayanan and anonymous reviewers for their constructive comments. This work was partially supported by NSF Grants: CNS-1750953 and CNS-1749049.

REFERENCES

[1] 2018. Frequency Calculator. http://niviuk.free.fr/lte_band.php. (2018).

[2] 2019. MobileInsight. http://www.mobileinsight.net,2019. (2019).

[3] 2020. HotMobile'20 dataset. http://milab.cs.purdue.edu/hotmobile2020_release/. (2020).

[4] 3GPP. 2013. Carrier Aggregation explained. https://www.3gpp.org/technologies/keywords-acronyms/101-carrier-aggregation-explained. (2013).

[5] 3GPP. 2018. TS36.331: E-UTRA; Radio Resource Control (RRC). (2018). (Release 15).

[6] 3GPP. 2019. TS36.101: E-UTRA; User Equipment (UE) radio transmission and reception. (2019). V15.6.0 (Release 15).

[7] Haotian Deng, Chunyi Peng, Ans Fida, Jiayi Meng, and Charlie Hu. 2018. Mobility Support in Cellular Networks: A Measurement Study on Its Configurations and Implications. In *IMC*.

[8] Yuanjie Li, Haotian Deng, Jiayao Li, Chunyi Peng, and Songwu Lu. 2016. Instability in Distributed Mobility Management: Revisiting Configuration Management in 3G/4G Mobile Networks. In *SIGMETRICS*.

[9] Yuanjie Li, Chunyi Peng, Zengwen Yuan, Jiayao Li, Haotian Deng, and Tao Wang. 2016. MobileInsight: Extracting and Analyzing Cellular Network Information on Smartphones. In *MobiCom*.

[10] Klaus Ingemann Pedersen, Frank Frederiksen, Claudio Rosa, Hung Nguyen, Luis Guilherme Uzeda Garcia, and Yuanye Wang. 2011. Carrier aggregation for LTE-advanced: functionality and performance aspects. *IEEE Communications Magazine* 49, 6 (2011), 89–95.

[11] Antti Reinikainen. 2015. Performance Evaluation of LTE-Advanced Carrier Aggregation. (2015). MS thesis, Aalto University.

[12] Shichang Xu, Ashkan Nikravesh, and Z Morley Mao. 2019. Leveraging Context-Triggered Measurements to Characterize LTE Handover Performance. In *PAM*.