Performance of Cellular Networks on the Wheels

Moinak Ghoshal*
Northeastern University
Boston, MA, USA
ghoshal.m@northeastern.edu

Imran Khan*
Northeastern University
Boston, MA, USA
khan.i@northeastern.edu

Z. Jonny Kong*
Purdue University
West Lafayette, IN, USA
kong102@purdue.edu

Phuc Dinh
Northeastern University
Boston, MA, USA
dinh.p@northeastern.edu

Jiayi Meng
Purdue University
West Lafayette, IN, USA
meng72@purdue.edu

Y. Charlie Hu
Purdue University
West Lafayette, IN, USA
ychu@purdue.edu

Dimitrios Koutsonikolas
Northeastern University
Boston, MA, USA
d.koutsonikolas@northeastern.edu

ABSTRACT

After 4 years of rapid deployment in the US, 5G is expected to have significantly improved the performance and overall user experience of mobile networks. However, recent measurement studies have focused either on static performance or a single aspect (e.g., handovers) under driving conditions of 5G, and do not provide a complete picture of cellular network performance today under driving conditions – a major use case of mobile networks. Through a cross-continental US driving trip (from LA to Boston, 5700km+), we conduct an in-depth measurement study of user-perceived experience (network coverage/performance and QoE of a set of major latency-critical 5G "killer" apps) over all three major US carriers, while collecting low-level 5G statistics and signaling messages. Our study shows disappointingly low coverage of 5G networks today under driving and highly fragmented coverage by cellular technologies. More importantly, network and application performance are often poor under driving even in areas with full 5G coverage. We also examine the correlation of technology-wise coverage and performance with geo-location and the vehicle's speed and analyze the impact of a number of lower layer KPIs on network performance.

CCS CONCEPTS

• **Networks** → **Mobile networks**; **Network measurement**; **Network performance evaluation**.

KEYWORDS

5G, Measurement, Driving, Coverage, Network Performance, Application Performance, Dataset

*Moinak Ghoshal, Imran Khan, and Z. Jonny Kong are co-primary authors.

IMC '23, October 24–26, 2023, Montreal, QC, Canada

ACM ISBN 979-8-4007-0382-9/23/10…$15.00
https://doi.org/10.1145/3618257.3624814

ACM Reference Format:
Moinak Ghoshal, Imran Khan, Z. Jonny Kong, Phuc Dinh, Jiayi Meng, Y. Charlie Hu, and Dimitrios Koutsonikolas. 2023. Performance of Cellular Networks on the Wheels. In *Proceedings of the 2023 ACM Internet Measurement Conference (IMC '23), October 24–26, 2023, Montreal, QC, Canada.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3618257.3624814

1 INTRODUCTION

5G NR specifications [16], especially the use of high-band millimeter wave (mmWave), promise ultra high bandwidth and low latency to enable the class of latency-critical apps such as Augmented Reality (AR), Connected Autonomous Vehicles (CAVs), 360^o video streaming, and cloud gaming, often dubbed as 5G "killer" apps.

However, 5G's higher frequency bands reduce its range making it difficult to achieve the same widespread coverage as LTE, and also result in high network performance fluctuations for users on the move, e.g., during driving. As such, wide adoption of 5G faces a significant new challenge – its actual deployment plays a significant role in the user-perceived performance. Consequently, it is important to study and understand actually perceived performance of commercial 5G networks by end-users, e.g., in running various latency-critical applications.

5G rollout started in 2019 and the wide-scale deployment has been rapid and aggressively marketed by all mobile network operators [4]. As such, after a rapid deployment in the US over the past few years, it is highly anticipated that 5G has significantly improved the performance of mobile networks and, more importantly, the user experience, in particular, when running the class of latency-critical apps which could not be supported by LTE.

To answer this question, there have been a flurry of measurement studies of 5G networks in recent years. However, these studies either focused on static performance of 5G networks [25, 37, 38, 40, 41], or on a single aspect (e.g., handovers) under driving conditions of 5G networks [26, 55], and hence do not provide a complete picture of cellular network performance today under driving conditions. Performance of mobile networks under driving is important as they not only provide connectivity to passenger phones [1], but also power the rapidly increasing number of connected vehicles [34], which run a wide variety of apps ranging from basic navigation to AI-driven CAV apps.

Table 1: Driving dataset statistics.

Total geographical distance travelled	5711+ km
States/major cities/counties traveled	14/10/100+
Timezones traveled	4
Operators	Verizon (V), T-Mobile (T), AT&T (A)
# of unique cells connected	3020 (V), 4038 (T), 3150 (A)
# of handovers	2657 (V), 4119 (T), 2494 (A)
Total cellular data used	777+ GB (Rx), 83+ GB (Tx)
Total log size	388+ GB
Cumulative experiment runtime	5561 min (V), 4595 min (T), 4541 min (A)

To gain a complete understanding of 5G network performance and more generally a holistic picture of the cellular network performance as a whole under driving conditions, we conduct – to our knowledge – *the first in-depth measurement study of the mobile networks of all three major US carriers while driving across the continental US (5700km+, from Los Angeles to Boston).* Our study spans all cellular technologies available today (LTE/LTE-A, 5G-low/midband/mmWave), all layers of the protocol stack, and multiple 5G "killer" apps.

Unlike previous measurements in static scenarios or with limited mobility [25, 37, 38, 40, 41], conducting such a cross-country driving measurement faces significant challenges:

• **[C1]** Assessing the end-to-end cellular network and application performance in the wild is known to be difficult and it becomes even more challenging under driving. Many factors can become the performance bottleneck, including the mobile operator's deployed infrastructure and policies, the vehicle's speed, the app design, and the server's location.

• **[C2]** Lower level key performance indicators (KPIs) and control-plane signaling events are crucial for understanding cellular network performance. However, it is difficult to extract such information from commercial off-the-shelf unrooted smartphones. Furthermore, these messages need to be synchronized with application level logs, which often have different and diverse sampling frequencies and timing formats.

• **[C3]** Designing an accurate and efficient methodology to log the coverage of each cellular technology (4G vs. 5G) while driving across the country is particularly challenging, as operators often deploy complex policies in deciding whether to elevate a UE's service from LTE to 5G. For example, UEs often fall back to LTE or do not switch to 5G in the absence of heavy traffic.

•**[C4]** It is not easy to evaluate the performance of upcoming 5G "killer apps" [28], since several such apps are still not out in the market for public use. For example, various studies show that Augmented Reality (AR) and Connected Autonomous Vehicle (CAV) applications can achieve high QoE by offloading computation to the edge [33, 54], but such applications have not been deployed commercially, and thus need to be custom-built for our study.

To address these challenges, we built a measurement platform consisting of (1) multiple smartphones, laptops, power supplies, and strategically deployed cloud and edge servers in various geographic locations across the US, (2) professional hardware and software tools that collect lower-level cellular network KPIs and control-plane events, (3) a mix of readily available and custom-built 5G "killer" apps, and (4) a custom-built software that includes a suite of tests (throughput, RTT, application QoE) and procedures for synchronization and post-processing of logs from different software and different layers of the protocol stack.

The key findings of our study are summarized as follows.

• In spite of the aggressive deployment efforts over the past few years, 5G coverage while driving is in general disappointingly low and highly fragmented across the country. Our analysis shows very diverse coverage across operators and even for the same operator across different geographic regions and suggests that today's 5G coverage is the result of more complex operator policies and deployment strategies. For example, our data suggest that operators are more likely to upgrade a UE's service to high-speed 5G (midband or mmWave) in the presence of backlogged downlink traffic, while they tend to prefer 5G-low or 4G for backlogged uplink traffic.

• Network performance under driving deteriorates significantly compared to the performance observed under the best static conditions. While downlink throughputs higher than 1 Gbps or of several 100s of Mbps are still possible with Verizon's/AT&T's 5G mmWave and T-Mobile's 5G midband services, the observed throughput with any carrier is below 5 Mbps about 35% of the time. More importantly, network performance is often poor even in areas with full high-speed 5G coverage. We also found that the use of edge servers brings a significant boost to both throughput and RTT compared to remote cloud servers.

• Our data show significant diversity in cellular network performance across operators at a given location and time. Interestingly, an operator that uses high-speed 5G at a given location does not always yield higher throughput than another operator using 5G-low or 4G. These two observations suggest that performance under driving can benefit significantly from multi-connectivity solutions that can aggregate links from multiple operators.

• We examined the impact of a number of cellular KPIs (RSRP, MCS, BLER, carrier aggregation) along with the number of handovers and the vehicle's speed on throughput during driving. Our analysis shows that none of these KPIs has a strong correlation with throughput and the factors that have the highest impact on throughput are different for different operators and even for different traffic directions for the same operator. Somewhat surprisingly, we found that the vehicle's speed and the number of handovers have weak and no correlation, respectively, with throughput, which is mostly affected by signal strength and MCS.

• We found that, while the number of handovers per mile can reach up to 20+, the number is typically low (1-3 handovers per mile in the median case for downlink traffic), and their duration is typically short (61 ms in the median case), resulting in a small throughput drop, which is often counterbalanced by a similar improvement in post-handover throughput. This explains the lack of correlation between throughput and number of handovers.

• We evaluated the QoE of two readily available downlink-centric apps (360^o video streaming and mobile cloud gaming) and two future cloud-assisted uplink-centric apps (an AR app and a CAV app that rely on DNN-based object detection). Our results show that all four apps experience poor performance and large performance variation under driving. While high-speed 5G can improve the worst-case performance of these apps compared to LTE, and the combination of 5G and edge computing can further boost performance, QoE remains disappointingly low and often no better than over LTE. Interestingly, we found that the number of handovers does not affect the average QoE, as such apps have built-in mitigation mechanisms (local tracking in the case of AR and CAV apps, buffering in the case of video streaming, frame rate adaptation in the case of cloud gaming), which help them combat temporary throughput and RTT degradation during handovers.

Contributions. In summary, this work makes the following contributions: (1) We collect a first-of-its-kind large multi-technology, multi-band, multi-carrier, cross-layer dataset of cellular network performance while driving across the country from LA to Boston. (2) Leveraging this dataset, summarized in Table 1, we provide the first in-depth study of cellular networks under driving, analyzing coverage, network and application performance and the factors that affect them, handovers, and diversity in terms of geo-location. Our dataset and scripts are publicly available [8].

2 RELATED WORK

2.1 Measurements under Limited Mobility

Various works have measured 5G performance in the U.S. [25, 37, 38, 40, 41]. The work in [41] studies 5G network performance, power consumption, and application QoE from a single location to multiple servers across the country, but it does not collect lower-level signaling messages. Other works study 5G performance in one or few cities, e.g. in Minneapolis [37, 38], Chicago [25, 37, 40], Boston [24, 25], Atlanta [24, 37], and Rome [29, 30], and thus the reported numbers do not reflect the rural or country-wide performance. Furthermore, all theses studies are conducted mainly under static and walking scenarios, and although most of them include some driving experiments, they are small in scale, e.g. spanning a few city blocks.

Outside the U.S., the works in [49] and [29, 30] study 5G performance in China and Italy, respectively, but focus only on the midband, and perform their measurements in a single city. The work in [52] studies mobile access bandwidth in China by collaborating with a commercial bandwidth testing app, and the work in [43] conducts a large-scale measurement study of 5G across a whole country (UK) from a mobile operator's perspective, focusing on network usage and temporal evolution. Different from our study, these works do not capture information such as mobility and low-level signaling messages.

Outside academia, several commercial bandwidth testing apps (e.g., Ookla SpeedTest [12], OpenSignal [11], nPerf [9]) provide crowdsourced-based 5G coverage and performance maps and some also publish periodic reports, e.g., [3, 13], comparing the major U.S. operators in terms of 5G coverage, performance, and QoE of a small set of applications (video, gaming, voice). Similar to the works in [43, 52], these measurements do not capture information such as user mobility or 5G band and do not have access to lower layer signaling messages. In §5.6, we compare the average driving network performance from our dataset against the average performance reported by OpenSignal for the 3 major U.S. operators.

2.2 Network Measurements under Driving

Driving measurements in 5G. The work in [26] performed extensive 5G measurements over the 3 major U.S. carriers while driving, but only focused on handovers. More importantly, the study mostly discusses findings using a single operator or short driving segments as examples, or presents cumulative results over all three operators. In contrast, we provide a detailed comparison of the handover behavior and their impact on network performance and application QoE across all 3 operators. The work in [55] also analyzes the inefficiencies in handover behavior via a 45-hour driving experiment, but only studies a single operator (AT&T).

Driving measurements pre-5G. The work in [23] studied handover policies via driving tests in three U.S. cities, and the works in [42, 48] conducted extensive measurements of cellular performance on high-speed rails (HSRs) in China. However, these measurements were conducted for LTE & 3G networks before 5G was rolled out. Additionally, the network performance on HSRs can be different from driving due to the difference in device speed and infrastructure provisioning.

3 METHODOLOGY

Drive Tests. We drove 5700+ km from Los Angeles to Boston covering all major cities in between (Las Vegas, Salt Lake City, Denver, Omaha, Chicago, Indianapolis, Cleveland, Rochester) over 8 days – 08/08/2022 to 08/15/2022. All measurement results reported in this work were obtained while driving on inter-state highways, in suburban areas, or inside cities, except for a few baseline measurements, which were conducted in major cities under static conditions.

Methodology. We now describe our methodology to address the four challenges mentioned in §1.

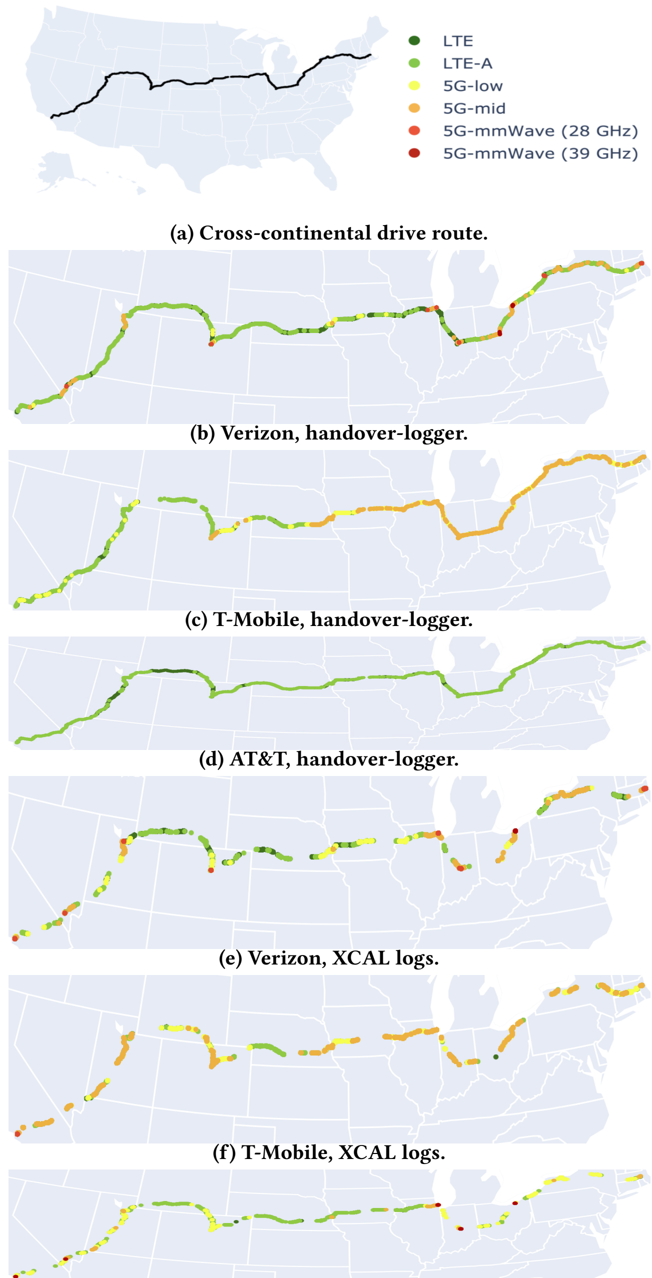
To address **[C1]**, we built a testbed (Fig. 17 in §B) consisting of multiple smartphones, laptops, Accuver XCAL Solo devices [17], application cloud and edge servers, and multiple AC power supplies to constantly charge all the devices. XCAL Solo is a standalone commercial tool that is attached to a smartphone via the USB-C port and it taps into the diagnostic interface of the smartphone to log all the PHY-layer KPIs and signaling messages. We used 6 Samsung S21/SM-G998U phones as our user equipment (UE) (details in §B), a state-of-the-art 5G smartphone model at the time when the measurements were done. We purchased multiple unlimited data plans from all three major U.S. carriers: Verizon, T-Mobile, and AT&T. All three carriers have low-band, mid-band, and mmWave services deployed.

To test end-to-end network and application performance, we deployed multiple AWS EC2 instances – two in California for the tests done in the Pacific and Mountain time zones, and two in Ohio for the tests done in Central and Eastern time zones. Additionally, to assess the benefits of edge computing on 5G apps, we deployed 5 Amazon Wavelength [18] edge servers in Los Angeles, Las Vegas, Denver, Chicago, and Boston. Wavelength servers are located *inside* Verizon's network in selected cities and specially designed for edge computing. For tests over the Verizon network, we used the deployed Wavelength server in each of these five cities and the cloud servers in the rest of the trip. For the other two operators, we only used the cloud servers.

To address **[C2]**, we collected app-level information from three unrooted smartphones (each connected to a different carrier) and the cloud/edge servers, and lower layer KPIs and control-plane signaling events from the XCAL Solo devices. An associated challenge here was to sync and post-process all the data from different layers, as the applications and XCAL logged information using different time formats (details in §B). To address this challenge, we wrote a sophisticated software that maps each app-layer log to the corresponding XCAL file taking into account the different timestamp types and the timezones we crossed, loads all the segregated XCAL files of a particular type of test to the XCAL parsing software (XCAP-M), and creates a consolidated database, which includes both the XCAL and the app layer data.

We used the other three unrooted phones ("handover-loggers" for the rest of the paper) to constantly log the handover events over the three carriers throughout the 8-day trip. We wrote an Android app to constantly send ICMP-based ping traffic with 38 bytes of data at an interval of 200 ms to prevent the cellular radio from going to sleep mode, and we logged key pieces of information related to mobility management: GPS information, cell IDs, cellular technology, etc. via Android APIs. However, we observed that often times the app would not switch from 4G to 5G even though we were in a 5G coverage area. We conjecture that the operators might be conservative and do not upgrade to 5G when the network traffic demand is low. To mitigate this challenge **[C3]**, we also extracted handover and technology information from the XCAL data logs generated during the network performance and application tests. A comparison between the two approaches to handover logging is described in §4.1.

We performed two different types of tests: (1) TCP bulk data transfers in both downlink and uplink directions, and RTT measurements. The experimental setup for these tests is described in §5. (2) Mobile app measurements. We evaluated the performance of 4 5G "killer" apps, two downlink-centric apps – 360° video streaming and cloud gaming – and two uplink-centric apps – AR and CAV. The experimental setup for each app is described in §C.1, §D.1, §E.1. While video streaming and cloud gaming apps are readily available, edge/cloud-assisted AR or CAV apps [33, 54] that offload data to an edge/cloud server, which performs DNN-based object detection, are not yet available on the market. To address this challenge (**[C4]**), we built a canonical edge-assisted AR/CAV app running on an Android phone that offloads dummy camera frames or LIDAR point clouds to an edge server. Details are provided in §C.1. We ran the bandwidth, RTT, and four mobile app tests in a round robin fashion on the three smartphones connected to XCAL Solo devices (each using a different carrier).



(a) Cross-continental drive route.

(b) Verizon, handover-logger.

(c) T-Mobile, handover-logger.

(d) AT&T, handover-logger.

(e) Verizon, XCAL logs.

(f) T-Mobile, XCAL logs.

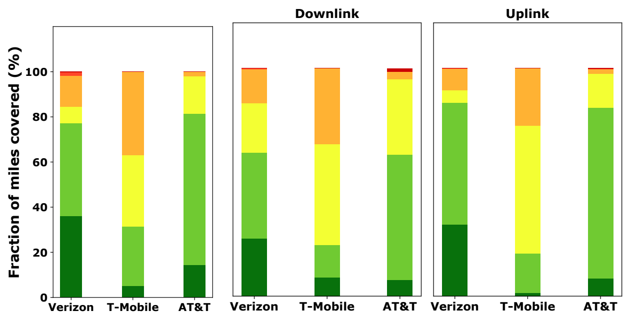
(g) AT&T, XCAL logs.

Figure 1: Comparison of the two approaches used to log cellular network coverage.

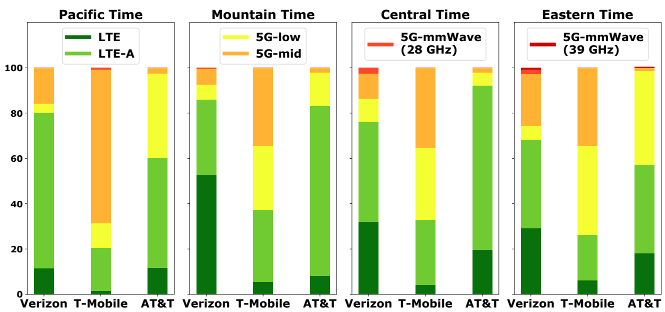
4 NETWORK COVERAGE

4.1 How to Measure Coverage

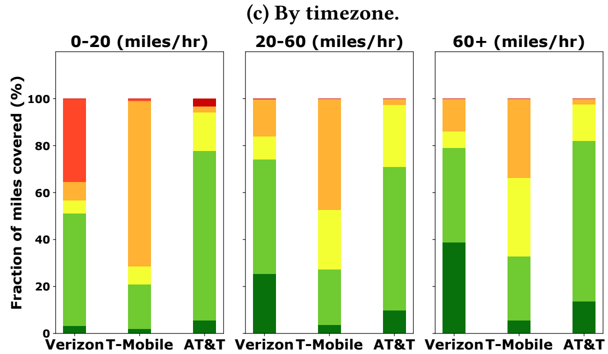
Fig. 1 compares the two approaches we described in §3 to log cellular network coverage for all three operators. We observe that the cellular technology spread obtained by the two approaches is very different. The handover-logger data (Figs. 1b, 1d, 1c) present a very pessimistic view in terms of coverage, suggesting LTE/LTE-A as the dominant technologies along the route for all three operators. In the most extreme case (AT&T), Fig. 1d shows that LTE/LTE-A are the only two technologies along the whole driving route. In contrast, the XCAL logs during throughput/application tests (Figs. 1e, 1g, 1f) show a very different view, with multiple areas of 5G coverage. This disparity in the results with the two approaches suggests that *passive approaches (e.g., [39]) that simply log the cellular network state in the absence of heavy traffic are not reliable due to conservative operator policies that do not upgrade a UE from LTE to 5G in the presence of low traffic rates.* In fact, these policies may vary by location even for the same operator; for example, in the case of T-Mobile (Figs. 1c and 1f), our results with the two approaches agree for the east half of the country but show a very different view for the west half. Overall, a lesson learned from this study is that *obtaining a representative coverage view is challenging, as it requires high-overhead (in terms of bandwidth usage for operators and monetary cost for clients) active probing approaches using heavy traffic.* Based on our findings, in the following, we only discuss the



(a) Overall. (b) By traffic direction.



(c) By timezone.



(d) By vehicle speed.

Figure 2: Cellular technology breakdown per operator.

data reported by XCAL during the throughput/application tests, even though this approach unavoidably relies on a smaller dataset.

4.2 Results

Fig. 2a shows the technology coverage for all three operators as a percentage of miles driven during the entire trip. We observe that, in spite of the aggressive deployment efforts over the past few years, *5G coverage is in general disappointingly low and highly fragmented across the country*. Our data confirms (the well-known fact) that T-Mobile has the highest percentage of 5G coverage (68%); in contrast, with Verizon and AT&T, we were able to connect to 5G only for only around 18-22% of the total distance traveled. The percentage of high-speed 5G (midband and mmWave) is even lower ranging from 38% (T-Mobile) to as low as 3% (AT&T). Verizon offers the highest mmWave coverage; in contrast, we rarely experienced mmWave connectivity with the other two carriers throughout the trip. Our results suggest very different deployment strategies by the three carriers. Verizon has prioritized the deployment of 5G mmWave (in downtown areas of major cities), while T-Mobile has focused on expanding the coverage to larger geographical areas by prioritizing low/mid-band deployments. In contrast, AT&T offers better 4G coverage (a much larger percentage of LTE-A vs. LTE) compared to the other two carriers.

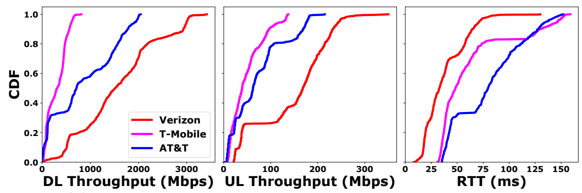
In Fig. 2c, we break down the technology coverage into four geographic regions. This figure shows *very diverse deployment strategies even for the same operator across different geographic regions*. For example, T-Mobile has a much higher percentage of 5G mid-band connectivity in the Pacific timezone compared to the rest of the country. AT&T has a very low percentage of 5G connectivity in the Mountain and Central timezones compared to the other two timezones. Verizon exhibits higher 5G coverage in the eastern half of the country (Central and Eastern timezones).

In Fig. 2d, we break down the technology coverage into three different speed bins: low (0-20 mph), mid (20-60 mph), and high (60+ mph). Note that the speed bins act as a proxy to the type of regions where the measurements were performed. For example, the low speed coverage samples are mostly from cities whereas the high speed ones are from the inter-state highways. We observe that high-speed 5G (mid band and mmWave) coverage decreases as we move from low to high speed bins for all three operators, suggesting that operators prioritize the deployment of high-speed 5G, in particular 5G mmWave, in urban areas. This is particularly true for Verizon, where the high-speed 5G coverage reduces from ∼43% in the low speed bin (cities) down to ∼13% in the high speed bin (inter-stare highways). Among the three carriers, T-Mobile is the only one that maintains a significant high-speed 5G (midband) coverage under medium and high speeds (47% and 33%, respectively).

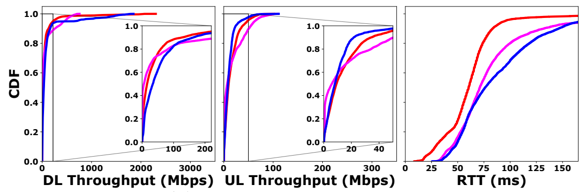
Up till now, we have examined the technology coverage regardless of the traffic direction (downlink vs. uplink). We now investigate if the technology a UE connects to at a particular location depends on the traffic direction. Here, we use only the data from our throughput tests, which always backlog the network in one direction, in order to maximize the possibility that a UE is serviced by 5G, and plot in Fig. 2b the technology distribution separately for downlink and uplink traffic. Interestingly, Fig. 2b shows that (1) for Verizon and AT&T, the 5G coverage overall is significantly higher in the downlink direction compared to the uplink, and (2) for all three carriers, the high-speed 5G coverage (midband and mmWave) is higher in the downlink direction compared to the uplink. Given the high asymmetry of downlink vs. uplink bandwidth in 5G mmWave and midband [25, 41], we conjecture that *operators are more willing to upgrade UEs to high-speed 5G in the presence of heavy downlink traffic, while they often prefer 5G-low or LTE/LTE-A for uplink traffic.*

5 NETWORK PERFORMANCE

In this section, we take a detailed look at the throughput and RTT of cellular networks while driving. To measure throughput, we used nuttcp [10] with the default TCP congestion control algorithm, CUBIC, to generate downlink and uplink backlogged traffic from/to an edge/cloud server. We used the default nuttcp configuration with a single TCP connection, as our intention was to measure the performance that would be experienced by applications such as file downloads/uploads, video streaming, CAV/AR cloud-assisted apps, etc., instead of measuring peak performance measured with tools such as SpeedTest. Each test lasted for 30-35 s and logged



(a) Static measurements done in major US cities.



(b) Driving measurements.

Figure 3: Overall throughout and RTT performance during static and driving measurements.

throughput every 500 ms. To measure the RTT between the UE and an edge/cloud server, we used the ICMP-based ping utility. Each test ran for 20 s and sent one ICMP packet every 200 ms.

5.1 Static vs. Driving Performance

To compare the driving performance with a baseline, we measured the downlink/uplink throughput and RTT in each of the major cities we visited while driving from LA to Boston. In each city, we tried to find a 5G-mmWave BS for each operator and performed the static measurements facing the BS. In cases we failed to find a mmWave BS, we measured the 5G mid-band performance. We omitted the static tests for those operator-city combinations for which we were not able to get 5G-mmWave or mid-band connectivity. Fig. 3 shows the CDFs of the static and the driving performance. The throughput CDFs consist of all the application-layer 500 ms throughput samples logged by XCAL and the RTT CDFs of all the individual RTT samples.

Fig. 3a shows that all three operators can provide very high downlink throughput in static, urban scenarios via their 5G mmWave or midband services, with median values of 1511/710/311 Mbps and maximum values as high as 3415/2043/812 Mbps for Verizon/AT&T/T-Mobile, respectively. The uplink throughput is an order of magnitude lower, as expected [25], with median values of 167/62/39 Mbps and maximum values of 350/215/137 Mbps for the three operators. Surprisingly, we observe a non-negligible fraction of low throughput values, especially for AT&T and T-Mobile, showing that 5G midband and mmWave can yield unpredictable performance even in ideal scenarios. We also observe that the RTT values show a very large variation, ranging from 8 ms to 100+ ms for Verizon and from 30 ms to 150+ ms for the other two carriers. Interestingly, Fig. 3a shows that AT&T with its mmWave coverage yields higher throughput than T-Mobile (which uses almost exclusively 5G midband) but also higher RTT. Upon closer inspection of our XCAL traces, we noticed that most of the RTT tests over AT&T were run over LTE/LTE-A even though the phone's screen showed 5G, suggesting that, in most cities, AT&T does not upgrade a UE to 5G under very low ICMP traffic.

Fig. 3b shows *a drastic drop in throughput during driving compared to static scenarios.* While the maximum downlink throughput is still higher than 1.8 Gbps for Verizon and AT&T and 600 Mbps for T-Mobile, the median/75-th percentiles are between 6-34 Mbps and 47-74 Mbps for all three operators, 1-5% of the corresponding values in static scenarios. The situation is similar in the uplink case, with median/75-th percentiles between 6-9 Mbps and 14-24 Mbps, 5-15% of the static values. Additionally, *we observe a significant fraction (35%) of very low throughput values (below 5 Mbps) in both directions for all 3 operators.* Further, Fig. 3b shows *a significant inflation to the RTT values under driving compared to the static scenarios*, with median values between 60-76 ms and maximum values as high as 2-3 s for all 3 operators.

5.2 Network Performance While Driving

To analyze the performance while driving, we plot the throughput and RTT CDFs for each cellular technology in Fig. 4. Additionally, for Verizon, we show separately the performance using an edge server (dashed lines) and a cloud server (solid lines). Overall, the results in Fig. 4 show that *cellular performance, in terms of both throughput and RTT, is often disappointingly low while driving, regardless of technology coverage.* 5G (including 5G-low) achieves higher throughput than 4G (LTE, LTE-A), but the performance with any technology varies dramatically and is often limited to a few 10s of Mbps or lower, even in the downlink direction. Similarly, RTTs are several 10s of ms with all technologies (with the exception of Verizon's 5G mmWave) and can exceed 100 ms.

When we look at the throughput performance in Fig. 4, we make the following observations: (1) In the downlink, 5G mmWave often achieves throughput higher than 1 Gbps even under driving, but it can also yield extremely low throughput; e.g., 40% of the AT&T downlink throughput with 5G mmWave samples are lower than 16 Mbps. (2) In the uplink, 5G mmWave offers the highest median throughput with Verizon and T-Mobile, but its maximum throughput is similar (Verizon, AT&T) to or lower (T-Mobile) than that of other technologies; in addition, with AT&T, 90% of the mmWave uplink throughput samples were lower than 0.5 Mbps. (3) T-Mobile's 5G midband service provides throughput up to 760 Mbps in the downlink, significantly higher compared to Verizon's or AT&T's midband, and its uplink throughput is often higher than T-mobile's 5G mmWave uplink throughput. However, T-Mobile's midband service also exhibits the largest fluctuation among the three carriers with 40% of its throughput samples falling below 2 Mbps in both directions.

When we look at the RTT performance in Fig. 4, we observe again that 5G mmWave achieves the lowest RTTs among all technologies in the case of Verizon, and 5G midband achieves lower RTTs than 5G-low and the 4G technologies for all three carriers. Interestingly, for both Verizon and T-Mobile, LTE-A achieves lower RTTs than 5G-low, demonstrating a tradeoff between throughput and RTT for these two technologies.

Finally, our results for Verizon show that *the use of an edge server brings a significant improvement to both throughput*[1] *and RTT compared to a cloud server*; for example, LTE-A with an edge server achieves higher uplink throughput and lower RTT than 5G

[1] The first graph in Fig. 4 shows that 5G midband throughput is higher with a cloud server than with an edge server. However, we note that we had very few samples of 5G midband throughput with an edge server, as the UE was connected to either 5G mmWave or LTE-A most of the time in the 5 cities where we deployed an edge server.

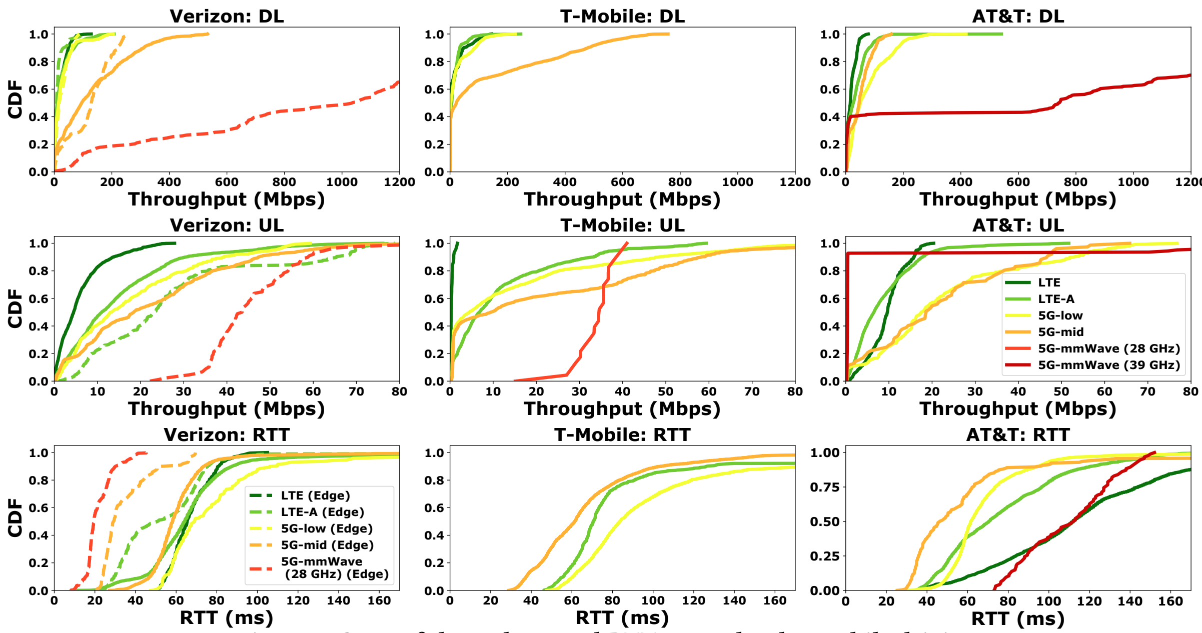


Figure 4: CDFs of throughput and RTT per technology while driving.

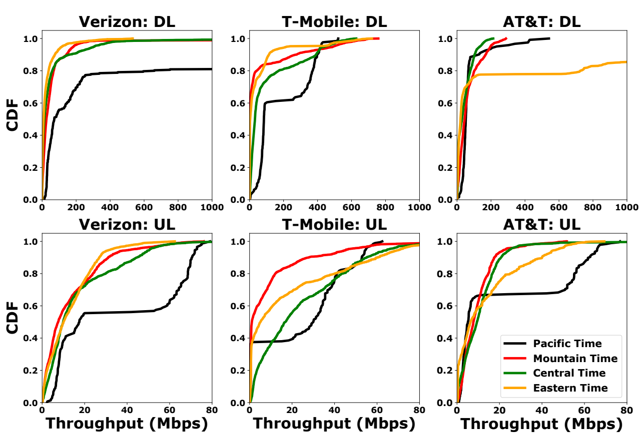


Figure 5: CDFs of throughput in different timezones.

midband with a cloud server; and the 5G mmWave RTT with an edge server is kept below 40 ms with a median value of 18 ms. This suggests that *edge computing is critical to boosting the performance of 5G killer apps (§7).*

5.3 Geo-diversity in Cellular Network Performance

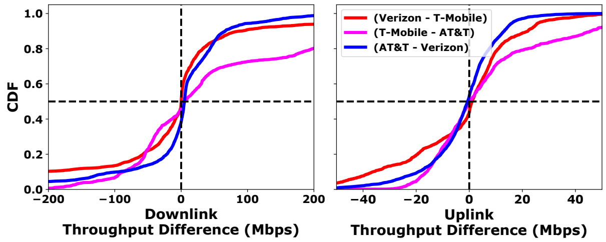
We next analyze the performance of cellular networks from a geographical angle. Our discussion of the diversity in network coverage across the country in §4.1 revealed different (and often complex) deployment strategies employed by different carriers. However, the real question of interest is whether higher 5G coverage actually translates to higher performance during driving in different parts of the country.

Fig. 5 breaks down the downlink and uplink throughput of each carrier across the four time zones. We make the following observations: (1) with the exception of AT&T in the downlink direction, where the throughput has higher values in the Eastern timezone, the throughput is clearly higher in the Pacific timezone in comparison to the other timezones, for all three carriers and both directions; (2) Verizon exhibits the lowest performance in the Eastern timezone among all four regions; (3) The performance in the Mountain timezone is low for all three carriers. Interestingly, *it is not always possible to explain the performance diversity across timezones based on the coverage results from Fig. 2.* The high performance for T-Mobile in the Pacific timezone and for AT&T in the Pacific (uplink) and Eastern (downlink) timezones can be possibly explained by the higher 5G midband coverage (compared to 5G-low) for T-Mobile and higher 5G-low coverage (compared to LTE/LTE-A) for AT&T in those timezones (Fig. 2a). On the other hand, Verizon has the highest 5G mmWave and midband coverage in the Eastern timezone, where it exhibits the worst performance.

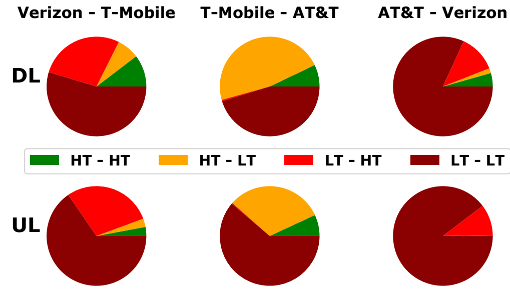
5.4 Operator Diversity in Cellular Network Performance

We now explore operator diversity in cellular network performance at a given location and time. For all throughput samples collected concurrently for any pair of operators, we plot in Fig. 6a the CDF of the throughput difference for each pair of operators. The results show that performance at a given location can be highly diverse across operators in both directions. This observation suggests that *performance under driving can benefit significantly from multi-connectivity solutions, e.g., over Multipath TCP, that can aggregate links from multiple operators [20, 31].*

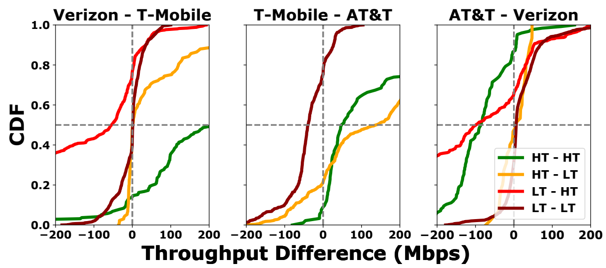
We next investigate what cellular technologies contribute to such diverse performance across operators at the same location and time. For example, a large throughput difference could be due to the fact that one operator is using a high-throughput (HT) technology (5G mmWave or midband) while the other one is using a low-throughput (LT) technology (LTE/LTE-A/5G-low) at a given



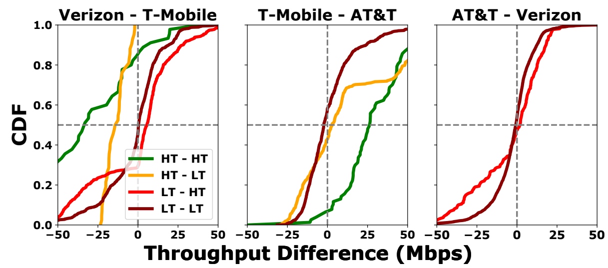
(a) CDF of throughput difference for each pair of operators.



(b) Technology-breakdown.



(c) CDFs of downlink throughput difference for each pair of operators broken down by technology.



(d) CDFs of uplink throughput difference for each pair of operators broken down by technology.

Figure 6: Operator-wise throughput difference for tests done in parallel.

location. We breakdown the points of each CDF (corresponding to a pair of operators) in Fig. 6a into 4 bins, based on the technology used by the two operators: HT-HT, HT-LT, LT-HT, LT-LT. Fig. 6b plots the distribution of the 4 bins for each operator pair in the downlink and uplink directions and Figs. 6c, 6d plot the CDFs of throughput differences for each bin. We make the following observations:

First, the majority of the samples for all three operator pairs in the uplink direction and for Verizon - T-Mobile and AT&T - Verizon in the downlink direction come from the LT-LT bin, which also contributes a large fraction of samples for T-Mobile - AT&T in the downlink direction. Figs. 6c, 6d show that the LT-LT CDFs for Verizon - T-Mobile and AT&T - Verizon in both directions and for T-Mobile - AT&T in the uplink directions are symmetric around 0, indicating roughly equal probability for one operator to outperform the other one. In contrast, in the case of T-Mobile - AT&T in the downlink direction, AT&T outperforms T-Mobile in ~80% of the locations thanks to its superior LTE-A and 5G-low services (Fig. 4).

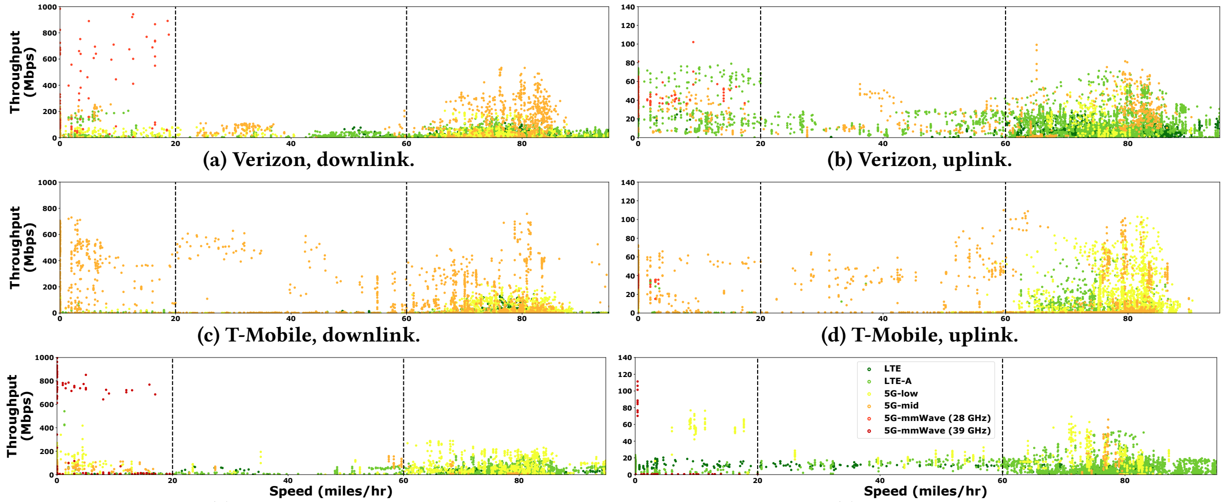
Second, the HT-HT bin contributes a very small number of samples in all 6 cases, ranging from 0.3% (AT&T - Verizon, uplink) to 10% (Verizon - T-Mobile, downlink). Interestingly, the HT-HT CDFs in Figs. 6c, 6d are not symmetric around 0, unlike the LT-LT CDFs. In the downlink direction, Verizon outperforms both the other operators thanks to its strong mmWave coverage in cities; and T-Mobile outperforms AT&T thanks to its superior midband service and the low mmWave coverage for AT&T. On the other hand, in the uplink direction, T-Mobile outperforms the other two operators.

Finally, the other two bins (LT-HT and HT-LT) together contribute a very small number of samples in the case of AT&T - Verizon (14% downlink, 10% uplink), as both these operators mainly deploy their HT services in the same locations (cities), but a substantial number of samples in the case of Verizon - T-Mobile and T-Mobile - AT&T, ranging from 32% (Verizon - T-Mobile and T-Mobile - AT&T, uplink) to 48% (T-Mobile - AT&T, downlink). For both operator pairs, in most cases, T-Mobile is the operator using an HT technology, as this is the only operator with a good 5G midband coverage in highways (see Fig. 2d). Here, one would expect the operator using an HT technology to always outperform the operator using an LT technology. Interestingly, Figs. 6c, 6d show that this is not always the case. For example, Verizon and AT&T LT outperform T-Mobile HT in ~20% of the locations in the downlink direction, as T-Mobile's 5G midband throughput can reach up to 760 Mbps but also falls below 2 Mbps 40% of the time (see Fig. 4). Similar results are observed for all three operator pairs in the uplink direction, where the performance gap between HT and LT technologies is much smaller compared to the downlink direction.

5.5 Cellular Network Performance Analysis

Given the very wide range of throughput and RTT values under driving (Fig. 3b), we examine the impact of different factors on cellular network performance, starting with the vehicle's speed. Fig. 7, which plots the throughput (500 ms samples) against the vehicle's speed in both directions for each operator,shows that throughput values can be distinctly divided into three speed regions: (a) low (0-20 mph), (b) mid (20-60 mph), and (c) high (60+ mph).

At low speeds, we observe overall high performance, with all the 5G mmWave points concentrated in this region, as expected. These points are mostly from the cities, where the speed is typically low. On the other hand, the high speed region consists of data mostly from the inter-state highways. This region has the maximum number of points for all the operators as most of our data was collected on highways. Surprisingly, we notice a significant fraction of throughput samples with high values even at such high speeds for Verizon and T-Mobile – several 100s of Mbps in the downlink and several 10s of Mbps in the uplink – mostly due to the deployment of 5G midband and 5G-low (in the case of T-Mobile only) along highways. On the other hand, throughputs are much lower with AT&T, which has much lower 5G midband coverage (Fig. 2a).



(a) Verizon, downlink. (b) Verizon, uplink.

(c) T-Mobile, downlink. (d) T-Mobile, uplink.

(e) AT&T, downlink. (f) AT&T, uplink.

Figure 7: Technology-wise throughput breakdown as a function of speed. We cut the downlink plots at 1000 Mbps as the mmWave throughput reaches up to 2.5 Gbps making the rest of the points barely visible.

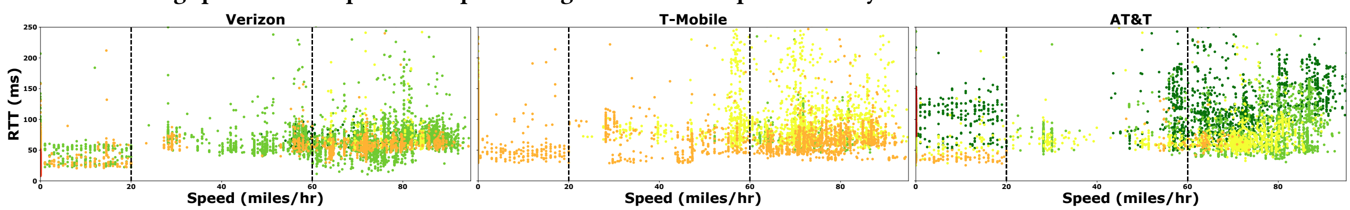


Figure 8: Technology-wise RTT breakdown as a function of speed.

Interestingly, throughputs in the mid-speed region are much lower than in the high-speed region for Verizon and AT&T. Most of the data points in this region are from sub-urban areas in-between cities/towns and inter-state highways. Our data suggest sparser cellular deployments in these areas in comparison to cities and highways, and sparser 5G coverage for Verizon and AT&T, causing the performance to drop considerably for these two carriers. Fig. 7 also shows a large number of very-low throughput points in all three regions, suggesting *a weak correlation between the throughput and the vehicle's speed* (see also Table 2).

In Fig. 8, we plot the RTT against the vehicle's speed in both directions for each operator, using the same three speed bins as in Fig. 7. In contrast to throughput, (Fig. 7), RTT appears to have a stronger correlation with the vehicle's speed, for two of the three operators (Verizon and T-Mobile); RTTs, regardless of technology, are in general lower in the (0-20 mph) bin and higher in the (60+ mph) bin. On the other hand, for AT&T, the LTE/LTE-A RTTs are in general higher than the 5G RTTs in all three speed bins. Another interesting observation from Fig. 8 is the absence of mmWave points for Verizon and AT&T except at very low speeds (close to 0 mph), showing again that operators typically do not upgrade a UE to 5G mmWave in the case of low ICMP traffic. Surprisingly, for AT&T, the RTT of this small number of 5G mmWave samples is higher than the 5G midband RTT.

In the remainder of this section, we focus on throughput. We look at a set of other factors that affect cellular performance and examine if they can explain the large throughput variation regardless of speed in Fig. 7. We consider 5 common cellular network KPIs

Table 2: Pearson's correlation coefficient between throughput and a number of KPIs.

	RSRP		MCS		CA		BLER		Speed		HO	
	DL	UL	DL	UL	DL	UL	DL	UL	DL	UL	DL	UL
Verizon	0.06	0.49	0.25	0.40	0.35	0.07	-0.08	-0.04	-0.29	-0.30	-0.02	-0.02
T-Mobile	0.46	0.51	0.34	0.62	0.29	0.05	0.23	0.10	-0.34	-0.10	-0.04	-0.05
AT&T	0.35	0.30	0.23	0.28	0.58	0.29	-0.13	-0.04	-0.37	-0.15	-0.05	-0.05

– Primary Cell's RSRP, Primary Cell's MCS, Carrier Aggregation (CA), Primary Cell's Block Error Rate (BLER), and the number of handovers (HO) – and compute the Pearson correlation coefficient of throughput with each of them in Table 2. We observe that *(1) none of these KPIs (including speed) has a strong correlation with throughput and (2) the factors that have the highest impact on throughput are different for different operators and even for different traffic directions for the same operator.* Surprisingly, *there is no correlation between throughput and the number of handovers.* We analyze this result in depth in §6. In the following, we investigate the root causes of the low correlation of some of the remaining KPIs with throughput.

RSRP. Table 2 shows weak-to-medium correlation with throughput for all operator-traffic direction combinations, but no correlation in the case of Verizon downlink. By inspecting our traces, we found that the RSRP for 5G mmWave (which primarily contributes the high downlink throughput samples) is low for most samples in the case of Verizon (-80 to -110 dBm), resulting in almost 0 correlation with throughput, but high in the case of AT&T (-70 to -90 dBm). The reason for this discrepancy lies in the different beamwidths of the phased arrays used by the two operators. In most of the cities, Verizon uses a smaller number of wider beams compared to AT&T, which result in lower gain, and hence, lower RSRP.

MCS. Table 2 shows a significantly higher correlation with throughput for T-Mobile uplink compared to the other operator-direction pairs. Here, we note that, because of heavy CA, the primary cell's MCS is not always a good proxy for the total throughput. For example, if secondary cells have strong link supporting high MCS indices, they can contribute to a a high throughput overall even if the primary cel's MCS is low.

CA., Table 2 shows weak-to-medium correlation with throughput for all three operators in the downlink and for AT&T uplink, but no correlation in the case of Verizon and T-Mobile uplink. We found that Verizon rarely uses CA in the uplink, which explains the low correlation with throughput. On the other hand, T-Mobile often aggregates 2 carriers in the uplink. However, the use of two carriers does not always manifest to higher throughput, as, in most cases, at least one carrier is LTE (this is true even in the case of 5G midband due to dual connectivity). Since LTE carriers have lower bandwidth than 5G carriers, a larger number of carriers does not necessarily boost the overall throughput.

Overall, we conclude that it is hard to isolate the impact of a single KPI on performance. An in-depth understanding of the impact of multiple KPIs on performance requires a multivariate analysis, which is part of our future work.

5.6 Performance Over Longer Time Scales

Up till now, we focused our analysis on short time scales (500 ms). As our mobile app experiments (§7) last between 20-180 s each, we now analyze the network performance on the same timescale. This analysis also allows us to compare our data with reports published by commercial bandwidth measurement apps (e.g., Ookla SpeedTest), which measure throughput at similar time scales.

We compute the mean and std. dev. of throughput (as a percentage over the mean) and RTT samples for each 30-second throughput test and each 20-second RTT test and plot their CDFs in Fig. 9. The upper rows show Verizon, T-Mobile, and AT&T, respectively, achieve median DL throughput of 30, 37, 48 Mbps, UL throughput of 13, 14, 10 Mbps, and RTT of 64, 82, 81 ms. Note that the median throughput is higher than that in Fig. 3 (which shows the CDF of 500 ms throughput samples), as the throughput of the samples is long-tailed. Nonetheless, the overall performance remains poor in terms of both metrics. Furthermore, the lower row shows the throughput and RTT experience high fluctuation within 30 s, with median values of 70%, 48%, 52% for DL throughput, 45%, 52%, 44% for UL throughput, and 18%, 29%, 19% for RTT.

To examine if the low throughput and high RTT values are a result of poor high-speed 5G coverage, in Fig. 10 we plot the throughput/RTT of each test against the fraction of time during which the UE was connected to 5G mmWave or midband. We observe that only T-Mobile's midband service brings a substantial improvement in throughput and only for the downlink direction (Fig. 10a). For the other two operators in the downlink direction and for all three operators in the uplink direction, *the throughput values are similar regardless of the percentage of time the UE was connected to 5G mmWave/midband* (with the exception of a few outliers). The same observations is true for the RTT values (Fig. 10c).

Finally, in Table 3, we compare the median throughput and RTT values from our dataset against those reported by Ookla SpeedTest [13] during Q3 2022 (which includes the time during which we collected our dataset). Compared to the results in the Ookla report, our results show significantly lower DL throughput, slightly higher UL throughput, and higher RTTs (especially for T-Mobile and AT&T). Assuming that most SpeedTest measurements are performed by static users[2], the significantly lower median DL throughput values and higher RTT values in our dataset demonstrate the performance degradation during driving. Still, this comparison should be taken with a grain of salt, as the numbers in the Ookla report are calculated over the whole country and the app typically selects servers close to the UE's location and uses multiple TCP connections to measure the peak bandwidth. In contrast, as explained in §5, most of our measurements are done with remote cloud servers and we used a single TCP connection, as our intent was to measure performance experience by most cloud-based apps.

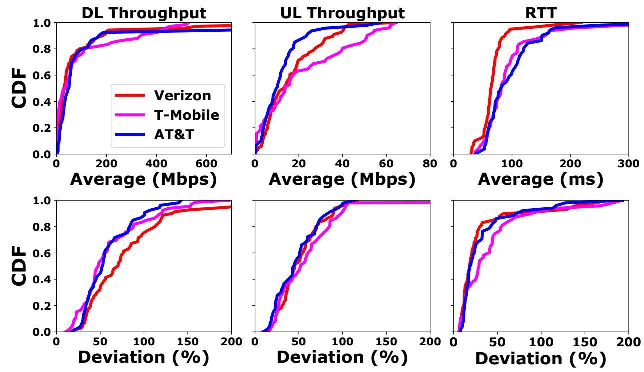
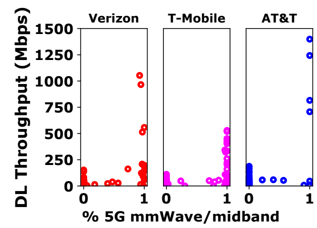
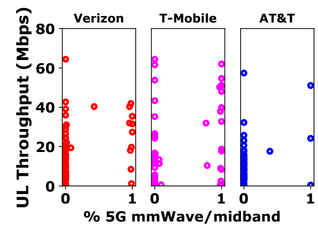


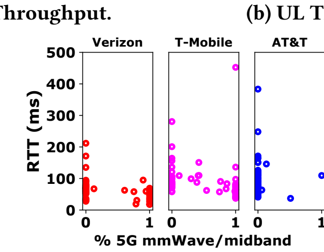
Figure 9: CDF of average & standard deviation of individual throughput and RTT tests.



(a) DL Throughput.



(b) UL Throughput.



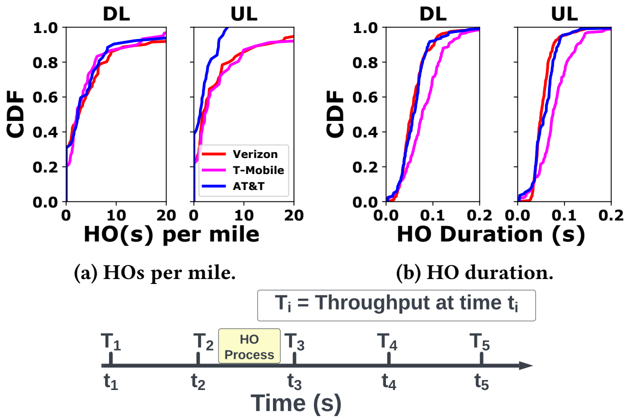
(c) RTT.

Figure 10: Performance as a function of the % of time the UE was connected to 5G mmWave/midband.

Table 3: Comparison with Ookla's report for Q3 2022.

	Downlink (Mbps)		Uplink (Mbps)		RTT (ms)	
	Our Data	Speedtest	Our Data	Speedtest	Our Data	Speedtest
Verizon	29.62	58.64	13.18	8.30	63.71	59.00
T-Mobile	37.09	116.14	13.77	10.91	81.68	60.00
AT&T	48.40	57.94	9.80	7.55	80.73	61.00

[2]The Ookla report does not provide any information about mobility.



(a) HOs per mile. (b) HO duration.

(c) HO process timeline.

Figure 11: Handover process and related statistics.

6 HANDOVERS

Handover statistics. We use the data from the throughput tests to quantify the frequency of handovers (HOs) across all cellular technologies. We calculate the total number of HOs (horizontal HOs between cells within the same technology and vertical HOs across technologies) experienced over a 30 s test and normalize it by the distance traveled in that test. Fig. 11a shows the results during downlink and uplink tests. Although we observed more than 20 HOs per mile in extreme cases, *in general, the number of HOs per mile is low*. The median (75-th percentile) for Verizon/T-Mobile/AT&T, respectively is 3 (6)/2 (5)/2 (5) in the case of downlink traffic and 2 (5)/2 (6)/1 (3) in the case of uplink traffic. Additionally, the CDF of the HO duration in Fig. 11b shows that *most handovers are fast*; the median (75-th percentile) HO duration for Verizon/T-Mobile/AT&T is 53 (73)/76 (107)/58 (74) ms in the case of downlink traffic and 49 (63)/75 (101)/57 (73) ms in the case of uplink traffic.

Impact of handovers on throughput. We now analyze the impact of a HO on network throughput. Specifically, we ask two questions: (1) how much does the throughput drop during a HO compared to the average throughput before and after a HO? and (2) how much does the throughput after a HO change compared to the throughput before the HO? Refer to Fig. 11c, where the time is measured in multiples of 500 ms (XCAL's throughput logging frequency) and assume that a HO takes place between t_2 and t_3. Then, to answer (1), we calculate $\Delta T_1 = T_3 - (T_2 + T_4)/2$, i.e., the difference between the throughput of the 500 ms interval that experienced one or more HOs and the average throughput over one 500 ms interval before and one 500 ms interval after the HO(s). Similarly, to answer (2), we calculate $\Delta T_2 = (T_4 + T_5)/2 - (T_1 + T_2)/2$, i.e., the difference between the post- and pre-HO throughput, each averaged over 1 s.

Figs. 12a-12c, 12g-12i show the CDF of the difference between the throughput during a HO and the average of just before and after the HO, for downlink and uplink, respectively. Irrespective of the traffic direction and operator, the values are lower than 0 around 80% of the time, indicating a drop in throughput during the HO. Nonetheless, the throughput drop is in general small, up to 60-80 Mbps in the downlink and up to 20-30 Mbps in the uplink.

Figs. 12d-12f, 12j-12l show the CDF of the difference between the post- and pre-HO throughput in the downlink and uplink directions, respectively, for each operator. About 55-60% of the time, the post-HO throughput is higher than the pre-HO throughput, i.e., a HO improves the performance. While the median throughput difference is very low (0.5-2 Mbps), the improvement can be as high as 100 Mbps in the downlink and as high as 20 Mbps in the uplink, which partly counterbalances the throughput drop during a HO.

Interestingly, for all operators and in both directions, the post-HO throughput is lower than the pre-HO throughput for about 25% of the time. To investigate further, we break down in the same figures the HO types into vertical (4G->5G, 5G->4G) and horizontal (4G->4G, 5G->5G) HOs. We observe that, as expected, among the four types of HO, 5G->4G is the type that mostly results in lower post-HO throughput, while 4G->5G typically improves the throughput. We also observe that horizontal HOs often result in lower post-HO throughput, but the impact is small (the negative part of these two CDFs has lower values than their positive part). A similar observation was reported in [50] for 4G->4G HOs and in [26] for NSA 5G->5G HOs in the mmWave bands.

Together, the small average number of HOs per mile (Fig. 11a), the HO duration (Fig. 11b), and our analysis of the impact of HOs on throughput (Fig. 12) explain the low correlation between throughput and HOs in Table 2. Assume an average driving speed of 60 mph, then from Fig. 11a, there are 1-3 HOs per mile for every 60 s or at most 0.025 HOs for every 500 ms interval in the median case, i.e., most 500 ms interval do not experience any HO. In addition, the impact of a HO on the overall throughput is small (Figs. 12a-12c and 12g-12i), and for about 25% of the cases, the post-HO throughput is actually lower than the pre-HO throughput (Figs. 12d-12f and 12j-12l).

7 5G APPLICATIONS

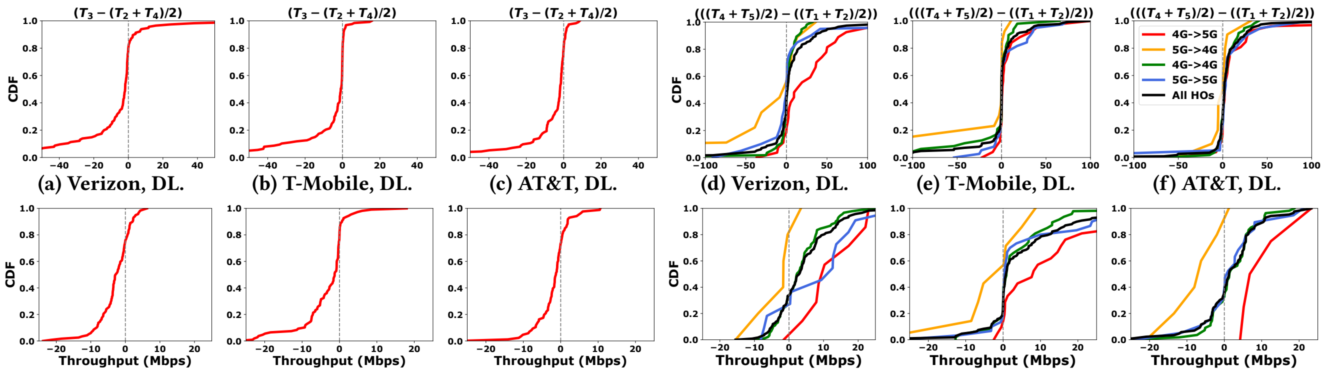
In this section, we analyze the feasibility of today's cellular networks in supporting a set of "5G killer" apps, which are all latency-critical and demand high bandwidth. Since Verizon benefits from edge servers and is shown to have the lowest RTT (Fig. 3b, Fig. 9), we only show the results with Verizon due to page limit and refer readers to the Appendix for results with T-Mobile and AT&T. The findings reported in this section hold true for all 3 carriers.

7.1 Uplink-centric Apps

AR and CAV apps are two representative uplink-centric 5G apps. Various works have shown that offloading the computation to a powerful edge server enables AR to achieve high-quality object detection [22, 33, 56], depth estimation [35, 36] and SLAM [19]. Similarly, CAVs rely on offloading vehicle-captured camera frames or LIDAR point clouds to the edge to enable collaborated vehicle perception [44, 47, 54]. To achieve high QoE, these apps have stringent requirements for high network throughput and low RTT.

In practice, commercial apps like these still do not exist today. Therefore, we built a canonical edge-assisted AR/CAV benchmark app, consisting of an Android app that offloads pre-recorded frames to an edge GPU server (with Nvidia A100) in a best-effort manner. We list the key parameters in Table 4, and refer readers to Appendix C.1 for details about the experimental methodology.

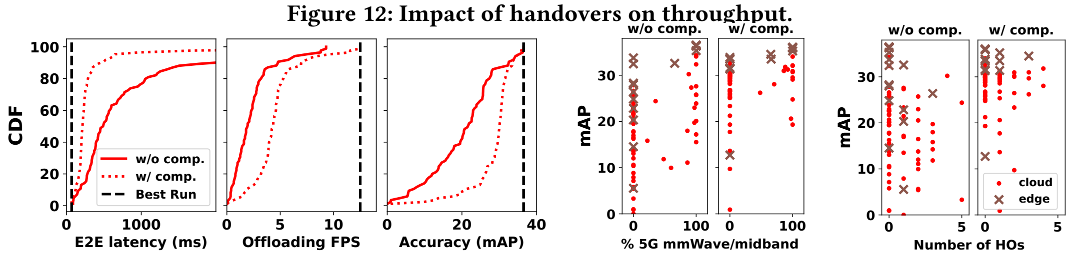
7.1.1 Performance of the AR App. Fig. 13 shows the performance of the AR app. We make the following observations. (1) Even in the best static scenario, the app achieves E2E offloading latency of 68 ms,



(a) Verizon, DL. (b) T-Mobile, DL. (c) AT&T, DL. (d) Verizon, DL. (e) T-Mobile, DL. (f) AT&T, DL.

(g) Verizon, UL. (h) T-Mobile, UL. (i) AT&T, UL. (j) Verizon, UL. (k) T-Mobile, UL. (l) AT&T, UL.

Figure 12: Impact of handovers on throughput.



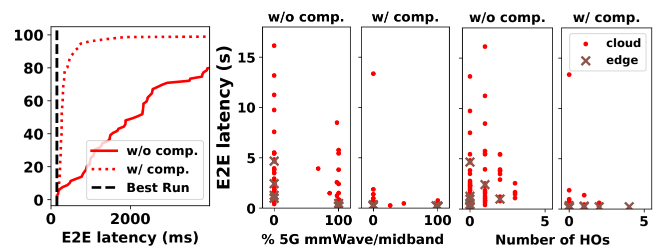
(a) Performance CDFs. (b) Impact of tech. (c) Impact of HO.

Figure 13: Performance of the AR app with Verizon.

Table 4: Configurations for the AR & CAV application.

	AR	CAV
Frames per second (FPS)	30	10
Frame size (raw)	450 KB	2000 KB
Frame size (compressed)	50 KB	38 KB
Frame compression time	6.3 ms	34.8 ms
Server inference time (A100)	24.9 ms	44.0 ms
Frame decompression time	1.0 ms	19.1 ms
Duration of a run	20 s	20 s

offloads only 12.5 FPS, and achieves an object detection accuracy (mAP) of only 36.5%. (2)*The AR app performance is significantly impacted by driving, achieving much higher E2E latency compared to the best static scenario, marked by a dashed vertical line, and thus much lower offloading frame rate and object detection accuracy.* Even with frame compression, the median E2E latency is 214 ms – 3x that of the best static case. As a result, the app achieves a median offloading frame rate of only 4.35 FPS and mAP of 30.1 – 34.8% and 82.4% of those under the best static scenario, respectively. (2) Fig. 13b shows that *5G mmWave/midband improves the worst case performance compared to 4G/5G-low, and the use of an edge server boosts the performance regardless of technology.* However, *the performance overall remains low even when the phone is connected to high-speed 5G 100% of the time.* (3) Fig. 13c shows that the app experiences up to 5 handovers within a 20-second run, but *there is no strong correlation between the number of handovers and mAP, suggesting that the impact of handovers on the AR app performance is limited.* We attribute this to the low impact of handovers on second-scale throughput (§6) and the AR app's built-in mechanism for mitigating transient throughput drops, local tracking, which effectively reuses the previously server-returned result [36]. (4) Fig. 13a and Fig. 13b show that compression significantly reduces the E2E offloading latency, enabling more frames to be offloaded

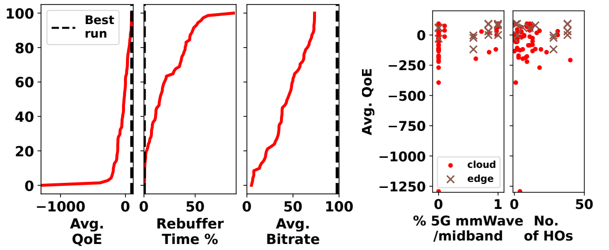


(a) Perf. CDFs. (b) Impact of tech. (c) Impact of HO.

Figure 14: Performance of the CAV app with Verizon.

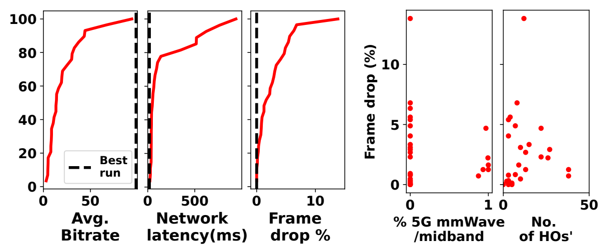
and improving the accuracy. This shows that *despite the promised high throughput of 5G, application-level optimizations developed in the pre-5G era such as frame compression are still needed.*

7.1.2 Performance of the CAV App. Fig. 14 shows the performance of the CAV application over Verizon. We make the following observations. (1) Today's cellular networks fail to achieve an E2E latency of 100 ms needed for the CAV pipeline to achieve accurate view reconstruction [32, 44, 45]. Fig. 14a shows the median E2E latency during driving is 269 ms if frame compression is used, and the lowest E2E latency recorded throughout our trip is as high as 148 ms. As shown in Table 4, the DNN inference time is 44 ms, and assuming the optimal RTT of 15 ms when connected to an edge server via 5G mmWave, to achieve an E2E latency of 100 ms the app needs to transmit a 2000 KB frame in 41 ms. which requires an uplink bandwidth of 390 Mbps – far beyond the capability of today's 5G mmWave. Although point cloud compression significantly reduces the frame size from 2000 KB to 38 KB, it incurs a compression time of 34.8 ms and decompression time of 19.1 ms, which makes it impossible for the app to achieve 100 ms E2E latency.(2) Although frame compression does not reduce the E2E



(a) Performance CDFs. (b) Tech. & HO impact.

Figure 15: 360° video streaming with Verizon.



(a) Performance CDFs. (b) Tech. & HO impact.

Figure 16: Cloud gaming with Verizon.

latency to below 100 ms, it still helps by reducing the median E2E latency by 8X, suggesting that *app-level optimizations such as point cloud compression are still needed in the 5G era*. (3) Similar to the AR app, *the use of an edge server improves performance compared to a cloud server regardless of the cellular technology and high-speed 5G only improves the worst case performance compared to 4G/5G-low* (Fig. 14b). Also, *there is no obvious correlation between number of handovers and E2E latency* (Fig. 14c).

7.2 360° Video Streaming

We implemented and evaluated our own 360° video streaming application, and refer readers to Appendix D.1 for details of the methodology. Fig. 15 shows the results for the Verizon network. We make the following observations. (1) The QoE [53] varies significantly during driving (Fig. 15a) with a median value of -53.75, which is significantly lower than the value of the best static run (96.29), suggesting that *high-quality 360° video streaming cannot be supported under driving*. Note that the theoretical best value is 100 assuming no stalls and no bitrate switch. (2) The average QoE during driving is negative for 40% of the runs due to the high rebuffering time, which can be as high as 87% of the total playback time (Fig. 15a). (3) Compared to the AR and CAV apps, Fig. 15b (left) shows that *the cellular technology has a bigger impact on video streaming QoE*; most of the runs when the UE was connected to high-speed 5G 100% of the time have positive QoE. This is because, compared to the AR and CAV apps, video streaming is not as latency-critical due to buffering, and its performance mostly depends on the network bandwidth, which is typically higher with 5G mmWave/midband. (4) Fig. 15b shows that connecting to an edge server generally results in higher QoE. (5) As with AR/CAV, *there is no strong correlation between QoE and the number of handovers.* Some runs with 40+ handovers yielded better QoE than runs with just 5 handovers.

7.3 Cloud Gaming

We evaluated the performance of Steam Remote Play [15], a popular cloud gaming platform. We refer users to Appendix E.1 for details of the methodology. Fig. 16 shows the results over Verizon. We make the following observations. (1) Fig. 16a shows that *the user QoE degrades significantly during driving* compared to the best static run. The sending bitrate varies significantly during driving (Fig. 16a) with a median value of only 17.5 Mbps, significantly lower than the value of the best static run (98.5 Mbps). The network latency during driving is always higher than 17 ms (best static run, required to stream at 60 FPS) and higher than 200 ms for 20% of the runs, which results in unacceptable user QoE. On the other hand, the frame drop rate is typically low (median value around 1.6%) but can be as high as 13.2%. In contrast, the best static run has a frame drop rate of 0.5%. (2) These numbers suggest that Steam Remote Play tries to keep the frame drop rate low (by adapting the frame rate) even at a cost of very high latency. (3) Similar to the AR and CAV apps, *5G midband/mmWave improves the worst case QoE* (the frame drop rate is never higher than 5%), but, overall, *does not improve the QoE compared to LTE/5G-low*. In fact, some runs with the UE connected to 5G midband/mmWave 100% of the time had higher frame drop rate than some runs with the UE connected to LTE/5G-low 100% of the time. (4) Similar to all the other apps, *there is no correlation between the QoE and the number of handovers.*

Note: The work in [26] shows that handovers have a large impact on the QoE of apps such as video streaming and cloud gaming, which contrasts our findings in this section. We note that the authors in [26] drew their conclusions from a small number of experiments (e.g., an 8-min drive in the case of cloud gaming), in contrast to our results which are obtained from a cross-country drive.

8 CONCLUSION AND RECOMMENDATIONS

We collected a first-of-its-kind multi-carrier, multi-technology, multi-band, cross-layer cellular network dataset through a cross-continental US driving trip (5700km+). Using this dataset, we revealed characteristics of today's cellular networks under driving, in terms of coverage, geo-diversity, operator diversity, network performance, and app performance. Our study shows disappointingly low and fragmented 5G coverage and poor network performance, even in areas with full high-speed 5G coverage, which, in turn results in poor user QoE for major "5G killer" apps, compared to static conditions. Our initial analysis, using low-level signaling messages and a number of KPIs, shows that no KPI is highly correlated with throughput, suggesting that further research is required to understand the factors that affect network performance during driving. In the meantime, we make three recommendations based on our results: (1) app developers should continue to explore app-level optimizations (compression, local tracking, buffering, rate adaptation, etc.) from the pre-5G era; (2) smartphone vendors should explore multipath solutions over multiple cellular networks; and (3) network operators and cloud providers should collaborate in deploying more edge services embedded in the operator networks.

ACKNOWLEDGMENTS

We thank the anonymous reviewers and our shepherd Yasir Zaki for their helpful comments. This work was supported in part by NSF grant 2112778.

REFERENCES

[1] 2016. Americans Spend 293 Hours Driving Each Year. https://www.automotive-fleet.com/136735/americans-spend-an-average-of-17-600-minutes-driving-annually
[2] 2020. TensorFlow Android Camera Demo. https://github.com/tensorflow/tensorflow/tree/48a2944c94b190434418d5a7c7f0df452c3aded5/tensorflow/examples/android
[3] 2022. USA 5G Experience Report JULY 2022. https://www.opensignal.com/reports/2022/07/usa/mobile-network-experience-5g
[4] 2022. Verizon 5G Ultra Wideband expected to cover 175 million people by the end of 2022. https://www.verizon.com/about/news/verizon-5g-ultra-wideband-175-million-people-2022
[5] Online. *Counter-Strike: Global Offensive.* https://store.steampowered.com/app/730/CounterStrike_Global_Offensive/
[6] Online. *FFmpeg Project. (2022). Retrieved January 2022 from .* https://ffmpeg.org/
[7] Online. *Hitman 2: Silent Assassin.* https://store.steampowered.com/app/6850/Hitman_2_Silent_Assassin/
[8] Online. [IMC '23] Performance of Cellular Networks on the Wheels. https://github.com/NUWiNS/imc2023-cellular-network-performance-on-wheels-data.
[9] Online. nPerf. https://www.nperf.com/en/
[10] Online. *nuttcp - Network Performance Measurement Tool.* https://www.nuttcp.net
[11] Online. OpenSignal. https://www.opensignal.com/
[12] Online. SpeedTest by Ookla. https://www.speedtest.net/
[13] Online. Speedtest Global Index: Ranking mobile and fixed broadband speeds from around the world on a monthly basis. https://www.speedtest.net/global-index/united-states?mobile#market-analysis
[14] Online. *Steam Link App.* https://store.steampowered.com/app/353380/Steam_Link/
[15] Online. *Steam Remote Play.* https://store.steampowered.com/remoteplay
[16] 3GPP. 2019. 3GPP.2019. Release 15. https://www.3gpp.org/release-15
[17] accuverxcalsolo [n. d.]. XCAL Solo. https://accuver.com/sub/products/view.php?idx=11;.
[18] awsWavelength [n. d.]. What is AWS Wavelength? https://docs.aws.amazon.com/wavelength/latest/developerguide/what-is-wavelength.html.
[19] Ali J. Ben Ali, Zakieh Sadat Hashemifar, and Karthik Dantu. 2020. Edge-SLAM: Edge-Assisted Visual Simultaneous Localization and Mapping. (2020).
[20] CableLabs. 2019. 5G Link Aggregation with Multipath TCP (MPTCP) . https://www.cablelabs.com/5g-link-aggregation-mptcp.
[21] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. 2019. Argoverse: 3D Tracking and Forecasting With Rich Maps. In *Proc. of CVPR.*
[22] Tiffany Yu-Han Chen, Lenin Ravindranath, Shuo Deng, Paramvir Bahl, and Hari Balakrishnan. 2015. Glimpse: Continuous, real-time object recognition on mobile devices. In *Proc. of ACM SenSys.*
[23] Haotian Deng, Chunyi Peng, Ans Fida, Jiayi Meng, and Y Charlie Hu. 2018. Mobility support in cellular networks: A measurement study on its configurations and implications. In *Proc. of ACM IMC.*
[24] Phuc Dinh, Moinak Ghoshal, Dimitrios Koutsonikolas, and Joerg Widmer. 2022. Demystifying Resource Allocation Policies in Operational 5G mmWave Networks. In *Proc. of IEEE WoWMoM.*
[25] Moinak Ghoshal, Z. Jonny Kong, Qiang Xu, Zixiao Lu, Shivang Aggarwal, Imran Khan, Yuanjie Li, Y. Charlie Hu, and Dimitrios Koutsonikolas. 2022. An In-Depth Study of Uplink Performance of 5G MmWave Networks. In *Proc. of the ACM SIGCOMM 5G-MeMU Workshop.*
[26] Ahmad Hassan, Arvind Narayanan, Anlan Zhang, Wei Ye, Ruiyang Zhu, Shuowei Jin, Jason Carpenter, Z Morley Mao, Feng Qian, and Zhi-Li Zhang. 2022. Vivisecting mobility management in 5G cellular networks. In *Proc. of ACM SIGCOMM.*
[27] Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. 2014. A Buffer-Based Approach to Rate Adaptation: Evidence from a Large Video Streaming Service. In *Proc. of ACM SIGCOMM.*
[28] killerapp5g [n. d.]. 5G: Is There a Killer "App" on the Loose? https://drivenets.com/blog/industry-outlook/5g-is-there-a-killer-app-on-the-loose/.
[29] Konstantinos Kousias, Mohammad Rajiullah, Giuseppe Caso, Ozgu Alay, Anna Brunstorm, Luca De Nardis, Marco Neri Usman Ali, and Maria-Gabriella Di Benedetto. 2022. Implications of Handover Events in commercial 5G Non-Standalone Deployments in Rome. In *Proc. of ACM 5G-MeMU.*
[30] Konstantinos Kousias, Mohammad Rajiullah, Giuseppe Caso, Ozgu Alay, Anna Brunstorm, Luca De Nardis, Marco Neri, Usman Ali, and Maria-Gabriella Di Benedetto. 2022. Coverage and Performance Analysis of 5G Non-Standalone Deployments. In *Proc. of ACM WiNTECH.*
[31] HyunJong Lee, Jason Flinn, and Basavaraj Tonshal. 2018. RAVEN: Improving Interactive Latency for the Connected Car. In *Proc. of ACM MobiCom.*
[32] Shih-Chieh Lin, Yunqi Zhang, Chang-Hong Hsu, Matt Skach, Md E Haque, Lingjia Tang, and Jason Mars. 2018. The architectural implications of autonomous driving: Constraints and acceleration. In *Proc. of ACM ASPLOS.*
[33] Luyang Liu, Hongyu Li, and Marco Gruteser. 2019. Edge assisted real-time object detection for mobile augmented reality. In *Proc. of ACM MobiCom.*
[34] Keith Mallinson. 2022. Revenue boost for automotive industry from cellular connectivity outweighs SEP licensing costs (Analyst Angle). https://tinyurl.com/2cu3tm2m
[35] Jiayi Meng, Zhaoning Kong, Qiang Xu, and Y Charlie Hu. 2021. Do Larger (More Accurate) Deep Neural Network Models Help in Edge-assisted Augmented Reality?. In *Proc. of ACM SIGCOMM NAI Workshop.*
[36] Jiayi Meng, Z. Jonny Kong, Y Charlie Hu, Mun Gi Choi, and Dhananjay Lal. 2022. Do We Need Sophisticated System Design for Edge-assisted Augmented Reality?. In *Proc. of ACM EdgeSys.*
[37] Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, and Zhi-Li Zhang. 2020. A first look at commercial 5G performance on smartphones. In *Proc. of ACM WWW.*
[38] Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand AK Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, et al. 2020. Lumos5G: Mapping and predicting commercial mmWave 5G throughput. In *Proc. of ACM IMC.*
[39] Arvind Narayanan, Eman Ramadan, Jacob Quant, Peiqi Ji, Feng Qian, and Zhi-Li Zhang. 2020. 5G Tracker – A Crowdsourced Platform to Enable Research Using Commercial 5G Services. In *Poster, Proc. of ACM SIGCOMM.*
[40] Arvind Narayanan, Muhammad Iqbal Rochman, Ahmad Hassan, Bariq S Firmansyah, Vanlin Sathya, Monisha Ghosh, Feng Qian, and Zhi-Li Zhang. 2022. A comparative measurement study of commercial 5G mmwave deployments. In *Proc. of IEEE INFOCOM.*
[41] Arvind Narayanan, Xumiao Zhang, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, Xiao Zhu, Xiaoxuan Zhang, Denis Rybkin, Zhengxuan Yang, Zhuoqing Morley Mao, Feng Qian, and Zhi-Li Zhang. 2021. A Variegated Look at 5G in the Wild: Performance, Power, and QoE Implications. In *Proc. of ACM SIGCOMM.*
[42] Yunzhe Ni, Feng Qian, Taide Liu, Yihua Cheng, Zhiyao Ma, Jing Wang, Zhongfeng Wang, Gang Huang, Xuanzhe Liu, and Chenren Xu. 2023. POLYCORN: Data-driven Cross-layer Multipath Networking for High-speed Railway through Composable Schedulerlets. In *Proc. of USENIX NSDI.*
[43] Paniz Parastar, Andra Lutu Ozg Ozgu Alay, Giuseppe Caso, and Diego Perino. 2023. Spotlight on 5G: Performance, Device Evolution and Challenges from a Mobile Operator Perspective. In *Proc. of IEEE INFOCOM.*
[44] Hang Qiu, Fawad Ahmad, Fan Bai, Marco Gruteser, and Ramesh Govindan. 2018. AVR: Augmented Vehicular Reality. In *Proc. of ACM MobiCom.*
[45] Hang Qiu, Po-Han Huang, Namo Asavisanu, Xiaochen Liu, Konstantinos Psounis, and Ramesh Govindan. 2022. AutoCast: Scalable Infrastructure-less Cooperative Perception for Distributed Collaborative Driving. In *Proc. of ACM MobiSys.*
[46] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks. In *Proc. of NeurIPS.*
[47] Shuyao Shi, Jiahe Cui, Zhehao Jiang, Zhenyu Yan, Guoliang Xing, Jianwei Niu, and Zhenchao Ouyang. 2022. VIPS: Real-time Perception Fusion for Infrastructure-assisted Autonomous Driving. In *Proc. of ACM MobiCom.*
[48] Jing Wang, Yufan Zheng, Yunzhe Ni, Chenren Xu, Feng Qian, Wangyang Li, Wantong Jiang, Yihua Cheng, Zhuo Cheng, Yuanjie Li, et al. 2019. An active-passive measurement study of tcp performance over lte on high-speed rails. In *Proc. of ACM MobiCom.*
[49] Dongzhu Xu, Anfu Zhou, Xinyu Zhang, Guixian Wang, Xi Liu, Congkai An, Yiming Shi, Liang Liu, and Huadong Ma. 2020. Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption. In *Proc. of ACM SIGCOMM.*
[50] Shichang Xu, Ashkan Nikravesh, and Z. Morley Mao. [n. d.]. Leveraging Context-Triggered Measurements to Characterize LTE Handover Performance. In *In Proc. of PAM 2019.*
[51] Francis Y. Yan, Hudson Ayers, Chenzhi Zhu, Sadjad Fouladi, James Hong, Keyi Zhang, Philip Levis, and Keith Winstein. 2020. Learning in situ: A Randomized Experiment in Video Streaming. In *Proc. of USENIX NSDI.*
[52] Xinlei Yang, Hao Lin, Zhenhua Li, Feng Qian, Xingyao Li, Zhiming He, Xudong Wu, Xianlong Wang, Yunhao Liu, Zhi Liao, et al. 2022. Mobile access bandwidth in practice: Measurement, analysis, and implications. In *Proc. of ACM SIGCOMM.*
[53] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A Control-Theoretic Approach for Dynamic Adaptive Video Streaming over HTTP. In *Proc. of ACM SIGCOMM.*
[54] Xumiao Zhang, Anlan Zhang, Jiachen Sun, Xiao Zhu, Y Ethan Guo, Feng Qian, and Z Morley Mao. 2021. EMP: Edge-assisted Multi-vehicle Perception. In *Proc. of ACM MobiCom.*
[55] Zhehui Zhang, Yuanjie Li, Qianru Li, Jinghao Zhao, Ghufran Baig, Lili Qiu, and Songwu Lu. 2022. Movement-Based Reliable Mobility Management for Beyond 5G Cellular Networks. *IEEE/ACM ToN* (2022).
[56] Zhang, Wuyang and He, Zhezhi and Liu, Luyang and Jia, Zhenhua and Liu, Yunxin and Gruteser, Marco and Raychaudhuri, Dipankar and Zhang, Yanyong. 2021. ELF: Accelerate High-resolution Mobile Deep Vision with Content-aware Parallel Offloading. In *Proc. of ACM MobiCom.*

APPENDIX

A ETHICS

This study was carried out by PhD students and faculty. We purchased multiple unlimited cellular data plans from all three US carriers and our experiments comply with their customer agreements. This work does not raise any ethical concerns.

B TESTBED DETAILS

Fig. 17 shows the testbed used in our experiments.

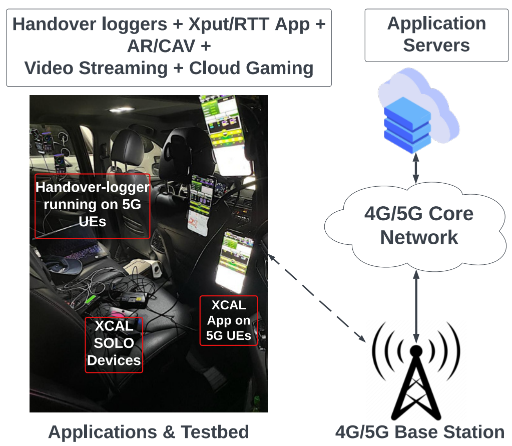


Figure 17: Overview of our measurement testbed.

Smartphones. Samsung S21, equipped with the Qualcomm Snapdragon 888 chipset, was a state-of-the-art 5G smartphone model at the time when the measurements were done. It supports 8 CC downlink and 2 CC uplink carrier aggregation and provides peak data rates of up to 3.5 Gbps downlink and 350 Mbps uplink over 5G mmWave.

Cloud servers. We used two families of AWS EC2 instances at each location: (1) g4dn.2xlarge (8 core CPU| 32 GB RAM| Nvidia T4 GPU| Windows Server 2019), a high-end GPU instance for the cloud gaming tests, and (2) t3.xlarge (4 core CPU| 16 GB RAM| Ubuntu 18.04), a standard Linux instance for all other tests. In both cases, the server's ingress/egress network bandwidth was 5 Gbps+, ensuring that the server's network capacity does not become the bottleneck.

Synchronization of data logs from various sources. As mentioned in §3, a major challenge we faced was to sync and post-process all the data from different layers, as the applications and XCAL logged information using different time formats. Some applications logged timestamps in UTC and others in local time. On the other hand, XCAL saved the log files (.drm files) with local timestamps in the filenames, whereas their contents had timestamps in EDT. This made it difficult to match a corresponding app layer log file with its XCAL counterpart. Crossing different timezones throughout the trip further increased the complexity. Additionally, XCAL relies on a licensed parsing software, XCAP-M, which converts the log (.drm) files to human readable data in the form of graphs and tables. This whole process requires manual intervention, which became a major challenge in our case, as we had to post-process thousands of log files.

C AR & CAV APPLICATION

C.1 Methodology

Both the server and the Android app we wrote can be configured to simulate various application scenarios, by specifying the frame size, frame rate (FPS) of incoming frames, frame compression time (if the app uploads compressed frames), and server inference time. The configuration parameters for AR and CAV are shown in Table 4 and are taken from [33, 54].

In each test, we run each app with and without frame compression configured, resulting in four runs each lasting 20 seconds. In each run, we measure the averaged E2E offloading latency, and the number of frames offloaded per second. For the AR app, we additionally estimate the object detection accuracy in mean average precision (mAP).

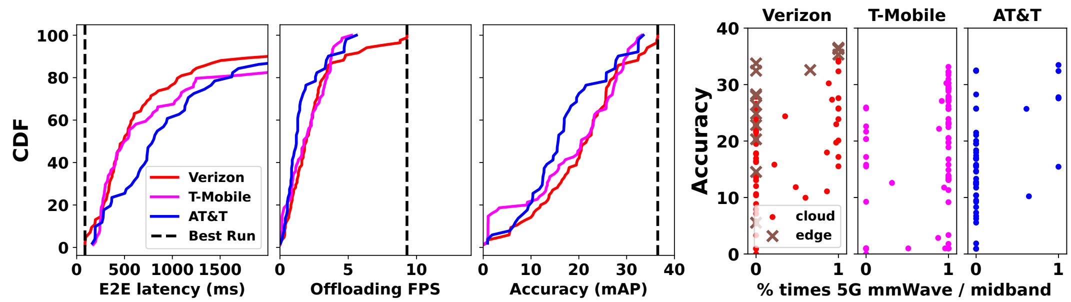
C.2 Estimating Object Detection Accuracy from E2E Latency

To understand the relationship between the end-to-end (E2E) offloading latency and object detection accuracy in terms of mAP, we performed an offline study with the phone offloading to an edge-server, where we varied the E2E offloading latency by varying the network bandwidth, and measured the object detection accuracy under different E2E latencies. We used Argoverse [21], a dataset recorded by a RGB camera mounted on a top of the car. We assume the edge server runs Faster R-CNN [46], a popular two-stage object detection model. The AR app runs an off-the-shelf on-device local-tracking algorithm [2], which moves the existing bounding boxes to follow the objects in the frame and is shown to improve edge-assisted object detection performance [44, 54].

Table 5 shows the relationship between E2E offloading latency and the object detection accuracy. Since the AR app produces object detection results for each incoming frame based on the latest received result from the server, the object detection accuracy is the same as long as the E2E is within the same bin in the unit of frame times. Since frame compression is lossy and thus results in different object detection accuracy, we report mAP both with and without frame compression.

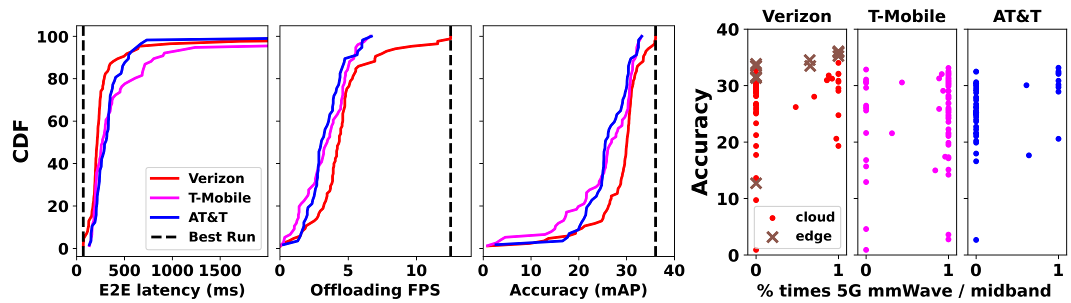
C.3 Results for All Three Operators

Figs. 18, 19, 20 show the performance of the AR and CAV apps for all three major U.S. operators. We found that the same observations made for Verizon (§7.1.1) also apply to T-Mobile and AT&T. In terms of cross-operator comparison, we make the following observations: (1) For the AR application, Verizon achieves the lowest E2E offloading latency, which results in the highest offloading FPS and object detection mAP across the three operators. This is because Verizon achieves the lowest median RTT of 63.7 ms, in comparison to T-Mobile's 81.7 ms and AT&T's 80.7 ms. AT&T achieves the worst performance without frame compression but slightly better performance than T-Mobile with frame compression. (2) The lead in Verizon is more significant when frame compression is used. Since frame compression reduces the frame size significantly from 450 KB to 50 KB, Verizon's shorter RTT plays a more important role in reducing the E2E latency than the high throughput. (3) For the CAV application, T-Mobile achieves the lowest E2E latency among the



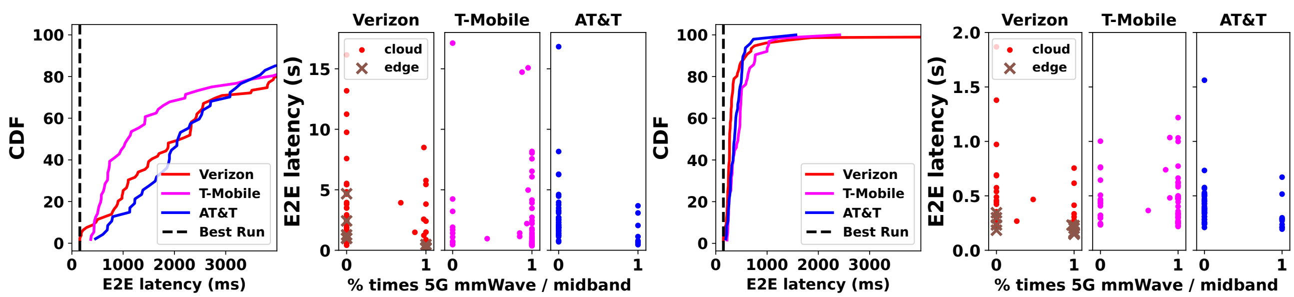
(a) CDF of app performance. (b) Impact of 5G connectivity.

Figure 18: AR without frame compression.



(a) CDF of app performance. (b) Impact of 5G connectivity.

Figure 19: AR with frame compression.



(a) App performance, without compression. (b) Impact of 5G connectivity, without compression. (c) CAV performance, with compression. (d) Impact of 5G connectivity, with compression.

Figure 20: CAV performance.

three operators, when point cloud compression is not used. This is because uncompressed point clouds have a large size of 2 MB, which benefits from T-Mobile's highest UL throughput compared to Verizon and AT&T (see Fig. 9). When point cloud compression is used, there is no significant difference between the performance of the three operators. (4) The impact of cellular technology is limited for both apps. In the case of AR, for AT&T, similar to Verizon (Figs. 13b, 14b), high-speed 5G improves the worst-case performance, while for T-Mobile, it improves the best-case performance without compression. However, the maximum object detection accuracy remains very low (below 36%) with all three operators regardless of technology. Similarly, in the case of CAV, high-speed 5G improves the worst-case performance for Verizon and AT&T but it has no impact for T-Mobile.

D 360° VIDEO STREAMING

D.1 Methodology

We evaluate 360° video performance by streaming videos from a media server to a mobile client during driving. We used Puffer, an open source video streaming server [51], and built our own client to receive and display video chunks for 360° videos. We customized Puffer's ABR algorithm to run BBA [27], which only relies on buffer size to choose a video bitrate and skips instances when capacity estimation is not needed. The 360° sample videos are chosen from YouTube and publicly hosted on an AWS instance. Each video is segmented into 2-second chunks and encoded into 4 different quality settings (100, 50, 10, and 5 Mbps) using ffmpeg [6]. Each playback session is run for 3 minutes. During each session, we collect three metrics – avg. QoE, avg. bitrate, and rebuffer time as a % of the duration of a run. We calculate the QoE [53] of a chunk

Table 5: Object detection accuracy (mAP) with the Argoverse dataset and Faster R-CNN model, at each E2E latency bin.

E2E latency (frame times)	mAP w/o compression	mAP w/ compression
0-1	38.45	38.45
1-2	37.22	36.14
2-3	36.04	34.75
3-4	34.65	33.12
4-5	33.36	31.82
5-6	32.20	30.50
6-7	31.08	29.53
7-8	28.03	26.99
8-9	27.01	25.73
9-10	25.62	25.21
10-11	25.77	24.35
11-12	23.29	22.44
12-13	22.75	21.56
13-14	22.48	21.64
14-15	21.59	21.16
15-16	20.59	20.35
16-17	20.11	19.69
17-18	19.53	18.95
18-19	18.40	17.61
19-20	18.01	17.85
20-21	17.52	17.00
21-22	16.96	16.55
22-23	16.59	15.97
23-24	15.41	15.16
24-25	15.78	14.94
25-26	15.86	15.37
26-27	14.81	14.71
27-28	14.70	13.77
28-29	14.44	13.62
29-30	14.05	13.70

based on the weighted sum of three elements, (1) video bitrate, (2) video bitrate variation between successive chunks, and (3) rebuffer time: $QoE_k = B_k - \lambda|B_k - B_{k-1}| - \mu T_k$, where B_k is the bitrate of chunk k, T_k is the rebuffer time recorded while downloading chunk k, and λ and μ are weighting parameters. The QoE of a run is the average QoE of all the individual chunks downloaded during that run. We empirically choose $\lambda = 1$ and $\mu = 100$ following the guideline in [53].

D.2 Results for All Three Operators

Fig. 21 shows the performance of the 360° video streaming application for all three major U.S. operators. The same observations made for Verizon (§7.2) also apply to T-Mobile and AT&T. In terms of cross-operator comparison, we make the following observations: (1) All operators achieve similar QoE, rebuffering, and average bitrate, with T-Mobile doing slightly better in terms of both rebuffering and average bitrate. (2) The cellular technology appears to have no impact on the performance with T-Mobile; in fact, the two runs with the worst QoE happened with the phone connected to 5G midband 100% of the time. For AT&T, the phone was only connected to 4G/5G-Low during almost all but 3 runs.

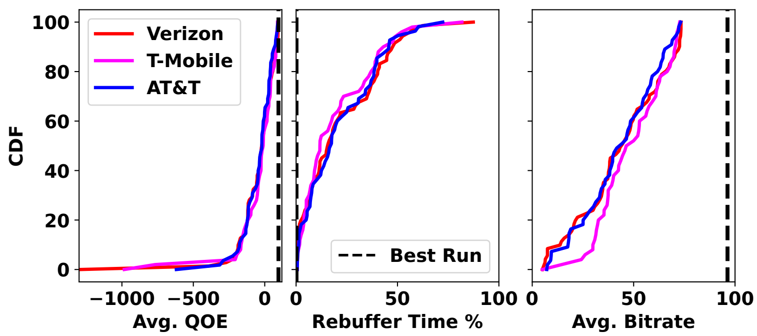
E CLOUD GAMING

E.1 Methodology

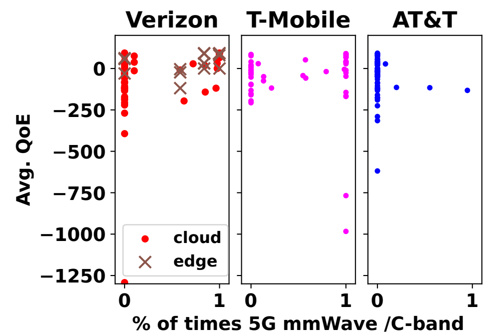
We deployed an AWS GPU instance to host two popular games: CSGO [5] and Hitman 2 [7] using the popular Steam Remote Play [15] gaming platform. These games are then played on our SGS21 smartphones using the Steam Link [14] Android application with 4K at 60FPS settings. The cloud server streams video frames at 4K resolution; however the frames are downscaled to 2K on the local device to fit SGS21's maximum resolution. We measure cloud gaming performance based on three metrics: send bitrate (Mbps), network latency (ms), and frame drop rate (%). The bit-rate information can be extracted by looking up the bit rate adapter's information from the Steam server's logs. We note that the maximum target value that can be set by the bitrate adapter is 100 Mbps. Additionally, the application server also reports network latency information every time it estimates a significant change in latency values. Finally, the percentage of frames dropped also heavily affects user's experience. To quantify the metric, we collect the number of frame drop events every second and divide it by the most recently set frame rate (FPS).

E.2 Results for All Three Operators

Fig. 22a shows the performance of cloud gaming application for all three operators. The same observations made for Verizon (§7.3) also apply to T-Mobile and AT&T. In terms of cross-operator comparison, we make the following observations: (1) Fig. 22a shows that T-Mobile achives the highest median bitrate at 21 Mbps, followed by Verizon (19 Mbps) and AT&T (9 Mbps). (2) In terms of network latency, the median values of all operators are approximately the same (around 50 ms). However, the app under Verizon sometimes experienced very high network latency of 500-1000 ms, while under the other two operators never exceeded 200 ms. (3) The frame drop rates with Verizon and AT&T are similar whereas T-Mobile sometimes experienced extreme frame drop rate (up to 25%). (4) 5G midband/mmWave has no impact for AT&T and T-Mobile.

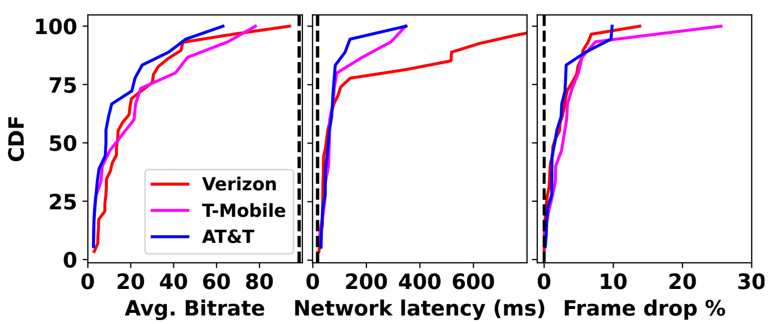


(a) CDF of app performance. The dashed line represents the best run.

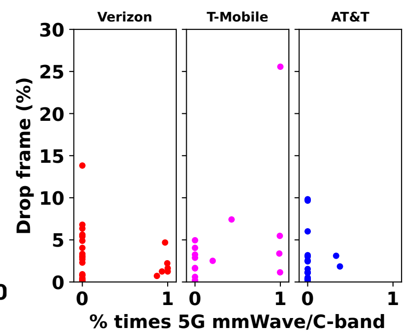


(b) Impact of 5G connectivity.

Figure 21: 360° video performance.



(a) CDF of app performance. The dashed line represents the best run.



(b) Impact of 5G connectivity.

Figure 22: Cloud gaming performance.