

A First Look at Commercial 5G Performance on Smartphones

Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, Zhi-Li Zhang
University of Minnesota – Twin Cities
fivegophers@umn.edu

ABSTRACT

We conduct to our knowledge a first measurement study of commercial 5G performance on smartphones by closely examining 5G networks of three carriers (two mmWave carriers, one mid-band carrier) in three U.S. cities. We conduct extensive field tests on 5G performance in diverse urban environments. We systematically analyze the handoff mechanisms in 5G and their impact on network performance. We explore the feasibility of using location and possibly other environmental information to predict the network performance. We also study the app performance (web browsing and HTTP download) over 5G. Our study consumes more than 15 TB of cellular data. Conducted when 5G just made its debut, it provides a “baseline” for studying how 5G performance evolves, and identifies key research directions on improving 5G users’ experience in a cross-layer manner. We have released the data collected from our study (referred to as *5Gophers*) at <https://fivegophers.umn.edu/www20>.

CCS CONCEPTS

• **Networks** → **Mobile networks; Network measurement; Network performance analysis; Wireless access networks.**

KEYWORDS

5G; Millimeter Wave; 5Gophers; Cellular Performance; Cellular Network Measurement.

ACM Reference Format:

Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, Zhi-Li Zhang. 2020. A First Look at Commercial 5G Performance on Smartphones. In *Proceedings of The Web Conference 2020 (WWW ’20)*, April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3366423.3380169>

1 INTRODUCTION

2019 marks the year for 5G, which was eventually rolled out for commercial services to consumers. Compared to 4G LTE, 5G is expected to offer significantly higher bandwidth, lower latency, and better scalability (*i.e.*, supporting more devices). The mainstream 5G deployment employs the millimeter wave (mmWave) technology that can provide, in theory, a throughput of up to 20 Gbps – a 100× improvement compared to today’s 4G [44].¹ Under the hood, this is achieved by a series of innovations including massive MIMO, advanced channel coding, and scalable modulation.

¹In this paper, we use “4G” to refer to 4G LTE networks.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW ’20, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7023-3/20/04.

<https://doi.org/10.1145/3366423.3380169>

| Carrier | Verizon [12] | AT&T [5] | T-Mobile [11] | Sprint [2] |
|---------|--------------|-----------|---------------|------------|
| Type | mmWave | mmWave | mmWave | mid-band |
| Freq. | 28/39 GHz | 24/39 GHz | 28/39 GHz | 2.5 GHz |

Table 1: 5G technologies adopted by major U.S. carriers.²

5G is expected to fuel a wide range of applications that cannot be well supported by 4G, such as ultra-HD (UHD) video streaming, networked VR/AR, low-latency cloud gaming, and vehicle-to-everything (V2X) communication. Despite these potentials, commercial 5G services are at their infancy. In early summer 2019, Verizon (VZ) launched 5G in Chicago and Minneapolis. It uses a 400 MHz channel at 28 GHz, making it the world’s first commercial mmWave 5G service for consumers. Followed by that, three other major U.S. carriers (T-Mobile, Sprint, and AT&T) have also rolled out their 5G services (Table 1). Many major carriers around the world are in the process of commercializing 5G.

Commercial mmWave 5G operates at a much higher frequency from 24 GHz to 53 GHz with abundant free spectrum. On the positive side, this offers much higher bandwidth compared to 4G. On the negative side, due to its short wavelength, mmWave signals propagate in a pseudo-optical manner, and are vulnerable to attenuation and blockage. mmWave has been studied theoretically and experimentally over various testbeds as a standalone physical layer (§2). However, integrating mmWave into commercial 5G networks faces far more challenges than those on the physical layer itself. Indeed, 5G is a complex ecosystem involving multiple entities – UEs (user equipments, *i.e.*, the client device), base stations, the core network, and even the Internet needs to embrace the 5G as our study will demonstrate; 5G also interacts with multiple protocol layers, with numerous optimization opportunities even at the transport and application layers; furthermore, 5G needs to properly coexist with the legacy 4G for a very long time.

In this paper, we conduct to our knowledge a first measurement study of commercial mmWave 5G networks on commodity smartphones. This study presents several challenges.

- The breadth of the 5G ecosystem makes it non-trivial to determine *what to study*. To this end, we take a user-centric approach by strategically selecting the measurement subjects that incur a high impact on end users’ experience: transport-layer performance under varying mobility scenarios and the application Quality-of-Experience (QoE). Given the importance of mmWave, we pay particular attention to the implications of mmWave in our measurements.

- Due to 5G’s very recent debut, there does not exist mature tools for capturing and monitoring 5G-specific information such as service status and 5G/4G handoff events. We therefore make methodological contributions by developing our own measurement tools.

²Based on the 5G services provided by the carriers as of October 2019.

- mmWave’s sensitivity to the environment dictates us to examine a wide range of factors that can influence the mmWave performance, complicating our measurement. We thus carefully select key factors including the UE-tower distance, UE orientation, blockage, and the weather; we then strategically design the controlled experiments to study their performance impact.

- mmWave’s ultra-high bandwidth makes it much more likely that the Internet becomes the performance bottleneck – a problem seldom appearing in 4G. We take two approaches to tackle this challenge. For most of the experiments, we conduct various pilot studies to maximize our confidence that the Internet does not remain the bottleneck; meanwhile, when the Internet-side bottleneck is inevitable (e.g., the latency), we also experimentally reveal how the Internet part can affect the 5G performance in realistic settings.

- Last but not least, there also exist non-technical obstacles. For example, our team needs to travel to multiple cities to conduct field measurements for multiple 5G carriers.

We study three major 5G carriers in this work: **VZ**, **T-Mobile**, and **Sprint**³ in three U.S. cities: Minneapolis, Chicago, and Atlanta. Among them, **VZ** and **T-Mobile** employ the mmWave technology, while **Sprint** adopts the “mid-band” 5G operating at 2.5 GHz at which the signal propagation still remains omni-directional. We conduct all experiments on commercial smartphones. Overall, our study consists of the following aspects.

An Overview of Today’s 5G Performance (§4). We begin by providing a first impression on the performance of today’s commercial 5G services by measuring the throughput and latency of the three carriers’ 5G networks. The results indicate that under typical urban environments where the UE is stationary, the average mmWave 5G throughput significantly outperforms that of the mid-band 5G. However, today’s commercial 5G offers little latency improvement due to its Non-Standalone (NSA) Deployment model that shares much of the existing 4G infrastructure with 5G (§2).

5G Performance of Stationary UE (§5). We compare the performance of mmWave 5G and 4G on key metrics such as throughput, latency, and packet loss rate when the UE remains stationary. We conduct our experiments under diverse scenarios with different distances/orientations between the UE and the 5G-NR (New Radio) panel, obstruction levels, and weather conditions. We find that commercial mmWave 5G offers much higher throughput than 4G (~10x improvement). However, even under clear line-of-sight, 5G throughput exhibits much higher variation than 4G, mainly due to the PHY-layer nature of mmWave signals (§5.1). Under non-line-of-sight (NLoS), mmWave 5G signals can be easily blocked by hands or human body. Despite that, in realistic urban environments, surrounding signal reflections can oftentimes mitigate the performance degradation, allowing 5G to function under NLoS (§5.2).

Mobility Performance (§6.1). We investigate the mmWave performance when the UE is moving (e.g., a 5G user walking or driving). We find that 4G-5G handoffs can be triggered frequently by either network condition change or user traffic. Even under low mobility (e.g., walking), a smartphone may experience 30+ 4G/5G handoffs in less than 8 minutes. Such a large number of switches may confuse applications (e.g., the video bitrate adaptation

logic) and bring highly inconsistent user experiences. Compared to mmWave 5G, mid-band 5G offers better mobility performance due to its omni-directional radio. For the same reason, 4G also exhibits much better stability when the UE is moving. These results indicate the necessity to jointly utilize mmWave 5G and omni-directional radio such as 4G in mobility scenarios (e.g., through MPTCP [38]) where 4G can help guarantee the basic data connectivity.

Inefficiency of Location-based Performance Estimation (§6.2). In 3G/4G, location is known to be useful for predicting the cellular performance [36, 50]. We investigate the feasibility of performing location-based performance prediction in mmWave 5G through a 30-day field study, and find that at a given location, mmWave 5G exhibits a statistically higher throughput variation compared to 4G, due to mmWave’s sensitivity to the environment – a small perturbation can affect the performance, making the location-based throughput prediction difficult.

Application Performance (§7). We study real application performance over mmWave 5G. We find that for web browsing, today’s 5G brings benefits only for large web pages; meanwhile, the optimizations brought by HTTP/2 and HTTP/3 (QUIC) are effective over 5G (§7.1). For large HTTP(S) download, we make an interesting finding that the goodput is significantly lower than the available mmWave 5G bandwidth, because many cross-layer factors may slow down the download. For example, compared to HTTP, HTTPS increases the average median download time by 23.5% due to the TLS overhead (§7.2). Overall, our results indicate that mmWave 5G’s high bandwidth does not always translate to a better application QoE, whose improvement requires joint, cross-layer optimizations from multiple players in the mobile ecosystem.

We make the following contributions in this paper.

- We develop practical and sound measurement methodologies for 5G networks on COTS smartphones.
- We present timely measurement findings of mmWave 5G and mid-band 5G performance on smartphones with key insights. Our experiments constitute more than 15 TB traffic⁴ and span three major 5G carriers in the U.S. As they were conducted when commercial 5G had just made its debut, we expect our results to provide an important “baseline” for studying how 5G performance evolves.
- We release our measurement dataset, referred to as *5Gophers*, to the research community to benefit work that needs real 5G data. The URL of the dataset is:

<https://fivegophers.umn.edu/www20>

2 BACKGROUND AND RELATED WORK

mmWave is an innovative technology integrated into 5G. Unlike 3G/4G that works at ≤ 5 GHz, mmWave 5G radios operate at much higher frequencies of 24 to 53 GHz (according to 3GPP 38.101 [1]) with considerably abundant free spectrum. Despite its high bandwidth, mmWave’s short wavelength makes its signals vulnerable to attenuation. To overcome this, mmWave transceivers have to use phased-array antennas to form highly directional beams. Due to the pseudo-optical nature of a beam, the signals are sensitive to blockages such as a pedestrian or a moving vehicle. Switching

³When we wrote this paper, AT&T has not yet offered 5G to non-business customers.

⁴We purchased multiple unlimited 5G data plans from **VZ**, **T-Mobile**, and **Sprint** for this study. Our study conforms to all the carriers’ wireless customer agreements.

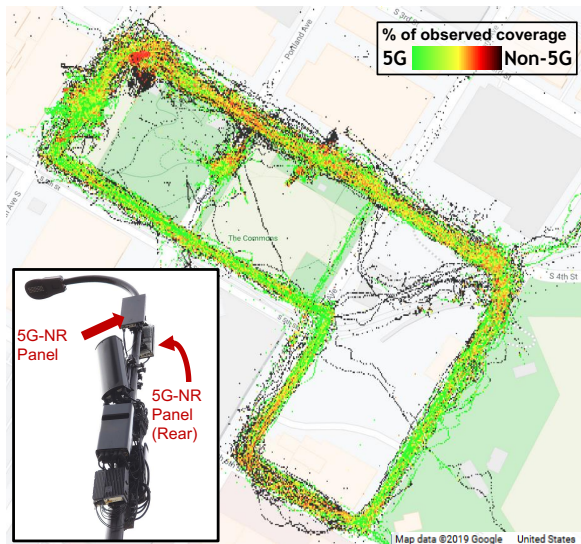


Figure 1: 5G coverage recorded at Minneapolis’s Commons Park. A color gradient from green to black indicates the percentage of observed 5G coverage (high to low respectively). We sampled 6.8 million data points to inform this visualization. Also indicated is a 5G mmWave base station.

from line-of-sight (LoS) to non-line-of-sight (NLoS) due to blockage may cause significant data rate drop or even complete blackout despite the beamforming algorithm that attempts to “recalibrate” the beams by seeking for a reflective NLoS path [40, 51].

Researchers have demonstrated the feasibility of deploying mmWave in data centers [27, 58, 60], indoor [14, 16, 19, 26, 39, 52, 53, 55], and outdoor environments [43, 46–48, 57, 59], as well as have conducted studies on beamforming and beam tracking [23, 41, 49]. But none of them studies mmWave in commercial 5G context on smartphones.

Mid-band 5G. Instead of adopting the mmWave technology, some carriers deploy their 5G networks over the mid-band frequency (1–6 GHz) whose radio signal largely remains omni-directional and offers a decent data rate. Mid-band 5G forms the basis of initial 5G services, but may suffer from a lack of spectrum in the long term. In contrast, mmWave has the unique advantages of ultra-high speed and abundant spectrum despite its limitations such as high attenuation and pseudo-optical signal propagation [3]. As shown in Table 1, three out of the four major carriers in the U.S. have adopted mmWave as the 5G technology.

5G Infrastructure. To reduce the time to market, carriers couple their 5G core network equipment with existing 4G LTE infrastructure in what is known as *Non-Standalone Deployment* (NSA). NSA utilizes 5G-NR for data plane operations while retaining their 4G infrastructure for control plane operations [29]. NSA is contrasted with *Standalone Deployment* (SA) – fully independent of legacy cellular infrastructures. All carriers in Table 1 currently employ the NSA model for their first commercial 5G deployment.

3 MEASUREMENT METHODOLOGY

5G Networks. Most of our experiments were conducted over VZ’s 5G network. In summer 2019 when we started this study, VZ was the only cellular carrier in the U.S. that offers commercial mmWave-based 5G services to consumers at specific downtown areas in two

cities: Minneapolis and Chicago. In their 5G coverage areas, dense 5G base stations are deployed (e.g., about 10 within two blocks in downtown Chicago). Due to VZ’s adoption of NSA (§2), 5G base stations are typically co-located with or very close to those of 4G (based on our knowledge and visual inspection). As shown in Figure 1, a 5G mmWave base station is typically equipped with one or more *panels* that are the mmWave transceivers. We observe that the panels typically face populated areas such as streets and pedestrian walkways. Figure 1 also exemplifies the 5G coverage in a downtown area of Minneapolis, based on 6.8 million data points collected from our 4-month field study.

In addition to VZ, for comprehensiveness, we also study two other carriers (**T-Mobile** and **Sprint**) listed in Table 1. **T-Mobile** also uses mmWave and **Sprint** employs a mid-band frequency at 2.5 GHz. We experimentally study their performance in §4.

5G User Equipment (UE). We use two types of COTS 5G-capable smartphones: Motorola Moto Z3 and Samsung Galaxy S10 5G (SM-G977U), henceforth referred to as MZ3 and SGS10, respectively. SGS10 has a built-in 5G radio, while MZ3 requires a separate accessory called 5G Mod [37] for accessing 5G. Comparing their performance at same locations, we find that MZ3 significantly underperforms SGS10 in terms of 5G throughput, likely due to hardware issues of MZ3 or its 5G mod. We thus use SGS10 for all experiments. To further ensure that our experiments are not affected by device artifacts, we purchase multiple SGS10 and confirm that they exhibit similar 5G performance. In addition, we confirm that despite 5G’s high throughput, the device-side processing is not a bottleneck for SGS10, which is a high-end smartphone equipped with an octa-core CPU, 8 GB memory, Qualcomm Snapdragon 855 System-on-Chip (SoC), and X50 5G modem. We also use SGS10 devices to test **T-Mobile** and **Sprint**. In addition, the SGS10 supports both 4G and 5G, allowing us to compare them on the same device.

Experiment Sites. For most of the experiments involving VZ, we conduct experiments at 4 locations (**A**, **B**, **C**, and **D**). **A** is a popular downtown area in Minneapolis with many buildings. **B** is at the boundary of the 5G coverage area in downtown Minneapolis. **C** is inside a hotel room in downtown Chicago where we stand near an open window. **D** is near the U.S. Bank stadium in Minneapolis with a large open space. We believe that these 4 locations are representative in terms of their environment (open/crowded space, low/high surrounding buildings, indoor/outdoor, etc.). We also conduct experiments at multiple locations in Atlanta for VZ, **T-Mobile**, and **Sprint**, with the details to be described in §4.

Server Selection. Due to the ultra-high bandwidth of 5G, the bottleneck of an end-to-end path may potentially shift from the wireless hop to the Internet – a situation that seldom appears in 3G/4G. Since the focus of our study is 5G, in most experiments we do *not* want such a shift to occur. To see how server selection affects the 5G performance, we carefully experiment with several server instances offered by major cloud service providers such as Microsoft Azure, Amazon Web Service (AWS), and Google Cloud, in different locations (e.g., U.S. east and west coast). We observe that the server location does affect the performance. For example, compared to a west coast server, a server on the east coast typically yields higher throughput and lower latency at our test locations.

For brevity, most of the experiments throughout the paper are done against a Microsoft Azure server located in the U.S. east coast.

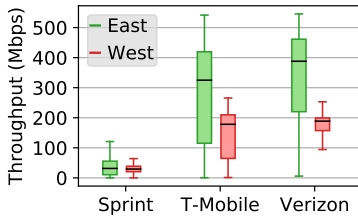


Figure 2: Throughput of a single TCP connection (Atlanta).

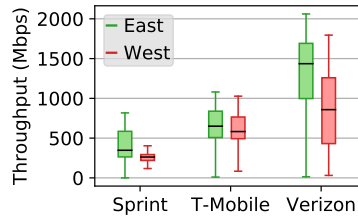


Figure 3: Aggregated throughput from 8 TCP connections (Atlanta).

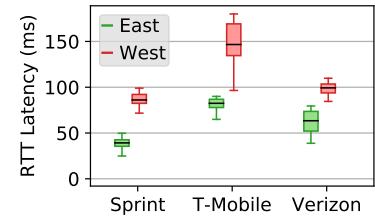


Figure 4: Base PING RTT (Atlanta, 3 carriers).

We select this server for three reasons. First, when downloading data from this server, we get the highest 5G throughput (statistically) compared to servers in other locations or of other cloud providers. Second, when we perform download tests from this server to other hosts (e.g., Amazon and Google cloud instances) over the wired Internet, we get ~3 Gbps throughput, which is much higher than the highest 5G speed we can obtain, during different times of a day. Third, we compare our throughput measurement results with those generated by Ookla Speedtest [10], a state-of-the-art Internet speed test service, and find both results match. Note that we do not directly use Ookla Speedtest due to a lack of fine-grained test control (number of TCP connections, test duration, test automation etc.). The above observations give us confidence that for an end-to-end path from a UE to the server, the Internet is unlikely the bottleneck. For some experiments (e.g., latency measurement and HTTP(S) download test), we also report other servers’ results to demonstrate the impact of server selection.

Test Workload. To measure transport-layer metrics including TCP/UDP throughput, RTT, and packet loss rate, we perform large bulk data transfers for bandwidth probing. Specifically, our UE issues one or more TCP connections or a UDP session to download data from an Internet server. Since it is difficult to root our UEs, we run the cross-compiled version of `iPerf` 3.6 [33] to measure the transport-layer metrics. We also test important applications such as HTTP(S) download and web page loading over 5G, with details to be presented in §7.

UE-side 5G Monitoring Tool. At the time when we conducted this study, the *then* state-of-the-art Android OS (version 9) did not support accessing 5G-NR related information.⁵ We are also not aware of any dedicated public use UE-side monitoring tool for 5G networks. Due to these limitations, we develop our own tool that collects the following information to support our measurements: (1) the UE’s fine-grained location, (2) all available network interfaces, (3) the actively used network interface and its IP address, (4) the cell ID (mCID) that the device is connected to, (5) the cellular signal strength, and (6) the 5G service status. The above information is obtained from various Android APIs. Regarding the last item (the 5G service status), we find that when the phone is connected to 5G, the `getDataNetworkType()` API of Android `TelephonyManager` still returns LTE. From our experiments, we find the `ServiceState` object when converted to a raw string representation contains the fields `nrAvailable` and `nrStatus`. We find that they reliably correlate to one of the three 5G connection

statuses: (1) the UE is not in a 5G coverage area (`nrAvailable=F`), (2) the UE is in a 5G area but is connected to 4G due to, for example, poor 5G signals (`nrAvailable=T, nrStatus=NOT_RESTRICTED`), and (3) the UE is connected to 5G (`nrAvailable=T, nrStatus=CONNECTED`). We have verified that this tool works with `VZ`, `T-Mobile`, and `Sprint`.

4 OVERVIEW OF TODAY’S 5G PERFORMANCE

We begin our study by measuring the 5G performance of commercial 5G carriers in the U.S. We select a total number of 6 locations in the downtown areas of Atlanta, where three out of the four carriers in Table 1 offer 5G services: `Verizon (VZ)`, `T-Mobile`, and `Sprint`. The selected locations have diverse urban environments such as high buildings, open plazas, and public transit hubs. We perform the following experiments using SGS10: (1) a 60-second measurement of downlink throughput over a single TCP connection, (2) a 60-second measurement of downlink throughput using 8 parallel connections, and (3) 60 measurements of the base RTT (end-to-end PING without cross traffic). The methodologies are detailed in §5.1. We experiment with two servers, one on the east coast and the other on the west coast. We repeat the above test 10 times for each unique (location, carrier, server) tuple. This leads to a total number of more than 350 tests. Figures 2, 3, and 4 plot the single TCP connection throughput, TCP throughput aggregated from 8 connections, and the base RTT, respectively, across all the tests for the three carriers and two servers.

We make several observations. First, in real-world urban environments, commercial 5G often exhibits great performance. For example, on `VZ`’s network, the TCP downlink throughput can achieve up to 2 Gbps – considerably better than today’s top-notch residential broadband networks. On the other hand, the throughput variation is huge – the throughput may drop to close to 0 for `VZ` and `T-Mobile`. Second, `T-Mobile` and `VZ`, which use mmWave technology, provide a much higher median throughput compared to `Sprint`, which operates at the mid-band frequency (2.5 GHz). This demonstrates the advantage of mmWave 5G – its ultra high-speed (§2). Third, the latency offered by today’s commercial 5G networks remains high – there is little improvement over 4G. This can be attributed to two reasons: (1) all carriers employ an NSA model that shares much 4G infrastructures with 5G, and (2) a lack of or limited use of edge support that helps shorten the end-to-end latency. We expect both to be addressed in the future in order to achieve sub-millisecond RTT. Fourth, we make an interesting observation that the aggregated throughput provided by 8 parallel connections is significantly higher than the throughput of a single TCP connection, across all three carriers. We will further investigate this in §5.1. Fifth, the server location indeed affects both the throughput and latency (§3).

⁵As of October 2019, Android 10 provides 5G-NR APIs, but no 5G phone was eligible for the update.

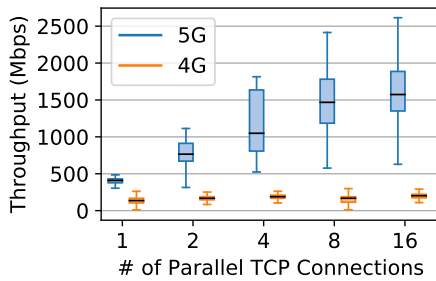


Figure 5: TCP performance under LoS: throughput.

We also evaluate upload performance of all the three carriers. We find that at the time of conducting this study, the upload speeds were far below the download speeds. For instance, the upload speeds of **VZ** and **T-Mobile** peaked at ~ 60 Mbps, while for **Sprint** it was around 30 Mbps. Due to this observation (which is further confirmed in [4]), this paper primarily studies download performance.

Note that we do not intend to compare or rank the carriers given the limited locations we sample and given that 5G is still in its early stage. Instead, the above results provide a first impression of today’s commercial 5G services, and motivate our subsequent in-depth study of 5G performance in diverse scenarios.

5 5G PERFORMANCE OF STATIONARY UE

We now closely examine how various factors such as LoS/NLoS, UE-panel distance, and mobility affect the 5G performance, through carefully designed controlled experiments. Unless otherwise noted, our experiments are conducted over **VZ**’s 4G and mmWave 5G networks. We consider the stationary UE scenario in this section, and focus on the mobility scenario in §6.

5.1 TCP/UDP Performance Under LoS

We begin with understanding 5G performance when clear LoS is present. Specifically, we conduct experiments at Locations A, B, and C (§3). At all locations, we ensure that we can visually see the 5G panel and there is LoS between the phone and the panel. At A and B, we select 5 UE-panel distances from 13m to 75m. We use a Leica DISTO E7500i professional laser distance meter [22] to accurately measure the distance. For each distance, we experiment with 3 orientations: 0° , 45° , and 90° (see Figure 8). For C, the distance (62m) and orientation (0°) are fixed.

In each test, we perform TCP bulk download for 60 seconds using {1, 2, 4, 8, 16} parallel TCP connections, and measure the throughput and RTT every second (reported by **iPerf**). For all bulk download tests, unless otherwise mentioned, we start collecting data 20 seconds after the TCP flow(s) start in order to mitigate the impact of TCP slow start. We repeat the entire process 3 times. All experiments were conducted in clear weather with the phone being held in hand. We believe that the above combinations provide realistic and diverse environmental configurations of urban 5G access from smartphones. Additionally, to ensure fair comparisons we also perform the same tests over 4G.

Figures 5 and 6 show the measurement results of throughput and RTT, respectively, for different numbers of concurrent TCP connections. Each box plot is across all 1-second measurement samples for a specific setup. We make several observations. First, with 8

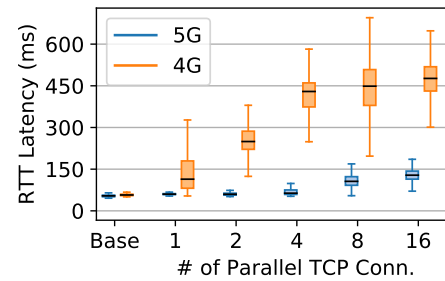


Figure 6: TCP performance under LoS: RTT.

parallel TCP connections, the median 5G and 4G throughputs are 1467 and 167 Mbps, respectively. However, 5G throughput exhibits much higher variations than 4G despite the presence of LoS. This is due to the PHY-layer nature of 5G signals as well as potential inefficiencies at various layers. For example, at PHY/MAC layers, smartphones’ small form factor makes engineering a 5G modem challenging [45]. At the transport layer, an excessive number of TCP connections may incur cross-connection contentions, which may also lead to throughput variation in particular for 5G whose available bandwidth is less stable compared to 4G.

Second, 5G throughput improves as the TCP concurrency level increases, with the bandwidth being fully utilized when there are more than 8 concurrent connections. Recall from Figure 2 and 3 that such a phenomenon also occurs in other 5G carriers. Through controlled experiments, we confirm that this is 5G-specific, *i.e.*, it does not appear in wired, WiFi, or 4G networks. It is either because the 5G carriers are imposing per-TCP-connection rate limiting, or because they are not able to support very high throughput for a single TCP connection. In either case, it may hurt the performance of single-connection protocols such as HTTP/2 [17]. It could also encourage application developers to aggressively increase the TCP concurrency that may adversely affect the TCP performance when the application operates in other networks.

| | First Hop RTT (ms) | East US Total RTT | West US Total RTT |
|----|--------------------|-------------------|-------------------|
| 5G | 27.4 \pm 6.4 | 54.0 \pm 4.5 | 81.9 \pm 5.5 |
| 4G | 29.2 \pm 4.8 | 58.0 \pm 4.3 | 88.9 \pm 5.5 |

Table 2: ICMP-based 1st Hop RTT and impact of server’s network location on total (end-to-end) PING RTT on 4G/5G. The reported numbers are averaged across all runs.

Third, regarding the latency, 5G and 4G exhibit similar base RTT (*i.e.*, end-to-end PING) at around 56 ms for the east coast server. As described in §4, high 5G RTTs also appear in other 5G networks, and can be explained by the NSA nature of today’s 5G networks. To understand how much 5G contributes to this RTT, we perform ICMP-based **traceroute** on the UE to measure the hop-by-hop RTT. As shown in Table 2, we find that the first-hop RTT, which presumably covers the RAN (Radio Access Network), is around 28 ms for both 4G and 5G, accounting for around 50% of the end-to-end RTT. Changing the server location to west U.S. reduces this fraction to around 33%. We then consider the RTT during a bulk transfer. As shown in Figure 6, 4G RTT inflates drastically because of its deep in-network buffers [30, 31]. Bufferbloat in 5G is much less severe,

likely due to the fast 5G speed that drains the buffer much faster than 4G. Also **VZ** 5G exhibits low packet loss rates, with the 50%, 75%, and 99% percentiles being 0.01%, 0.1%, and 1.2%, respectively.

Concurrent Clients. We also find that the bandwidth available from a single 5G panel is shared among its associated devices. For a simple test, we place two SGS10 side-by-side (ensuring both are connected to the same panel with LoS). We first let only one device run **iPerf** with 8 parallel connections. After TCP slow start, the maximum throughput is around 1.8 Gbps. We then let both perform **iPerf** simultaneously, and find that the above maximum throughput is approximately evenly divided between both devices (avg. ≈ 882 Mbps for one and ≈ 931 Mbps for the other). We consistently observe this in several locations at different times. This finding suggests that as the 5G user base increases, the perceived throughput may drop significantly due to bandwidth sharing among concurrent clients. However, due to very limited 5G users today, it is hard to quantify the impact. For most if not all experiments reported in this study, we have high confidence that our phone is the sole device connected to the 5G NR. In this sense, our study provides an important “baseline” for studying how 5G performance evolves in the future as the user base increases.

UDP Performance Under LoS. We repeat the measurement in §5.1 using UDP. Since UDP does not provide congestion control, we manually increase the sending rate exponentially from 512 Kbps to 2 Gbps. we find that for sending rates up to 1 Gbps, the receiver-side loss rate is close to 0. This indicates **VZ** 5G’s compatibility with UDP-based protocols such as QUIC [24] and HTTP/3 [18] at these low to medium data rates. However, at our test locations, **VZ**’s 5G is not able to reliably sustain 1 Gbps or a higher sending rate over UDP, as we observe a packet loss rate of up to 17% at 1 Gbps.

5.2 Impact of the Environment

Obstruction and NLoS Performance. The experiments in §5.1 assume a clear LoS path without any obstruction. We now place different types of objects along the LoS path to test whether 5G signals can penetrate/bypass them. We first examine two common obstructions: human body and hand. We stay at Location C (the Chicago hotel room) with an awning window directly facing a 5G-NR panel (62m). When the window is opened, the phone has LoS to the 5G-NR panel. We start a bulk transfer over 5G with 8 parallel TCP connections. During the data transfer, we block the LoS path with a human body and then a hand. We repeat this experiment 10 times and observe qualitatively similar results, with one representative run illustrated in Figure 7(a). As shown, both obstructions trigger 5G-to-4G handoffs (§6.1) and lead to significant performance degradation. In contrast, when experimenting with 4G, neither blockage incurs noticeable throughput drop (figure not shown), due to the low frequency of 4G signals.

The above results show that it is difficult for 5G signals to penetrate a hand or human body, causing NLoS between the transmitter and receiver. We then study other types of obstructions using similar methods. We find that when the UE is inside a backpack, a cardboard box, or a clear glass, 5G signals can penetrate these containers (experimented with <100 meters distance to the 5G panel with LoS). However, 5G signals can hardly penetrate human bodies, trains, pillar structures, and tinted glass. We find that 5G works in vehicles since the front windshield is typically clear glass.

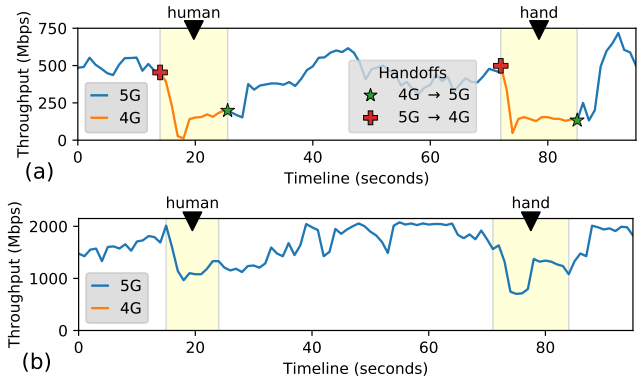


Figure 7: Obstruction tests. (a) Location C with ineffective multipath, (b) Location A with effective multipath.

We repeat the same experiment in Location A, also using a human body and a hand as obstructions. As shown in Figure 7(b), the impact of the obstructions becomes smaller: the 5G connectivity persists despite a fair amount of performance degradation. Figures 7(a) and 7(b) indicate that the environment can affect the impact of obstructions. At Location A, despite the NLoS created by the obstructions, the nearby buildings can reflect signals and create multiple wireless paths, and the reflected signals can still reach the UE. At Location C, the room has UV-protective windows that are very common in today’s buildings. Since the windows attenuate reflected 5G signals [57], multipath becomes ineffective. In other words, the only effective signal propagation path is through the open window. Blocking it inevitably degrades the performance. Note that during the experiment we only open the awning window and keep other windows closed.

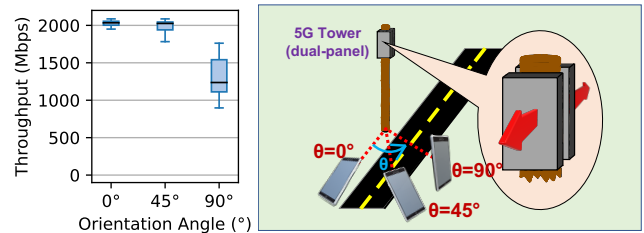


Figure 8: Impact of UE-panel orientation on throughput.

Impact of UE-Panel Orientation. We also investigate how the UE’s orientation to the panel affects network performance. We define the orientation as the minimum angle between the LoS and all normal vectors of the base station’s panels. As illustrated in Figure 8, an orientation of 0° is preferred because the panel is directly facing the UE, while an orientation of 90° is the least ideal case. Our orientation test is performed at Location D where we can find a large LoS area centered by a 5G tower. We pick three spots whose orientations are 0° , 45° , and 90° . All spots have a 25m distance to the tower. At each spot, we perform three 60-second bulk download tests using 8 parallel connections. As shown in Figure 8, we observe very small performance difference between 0° and 45° orientation, attributed to the environmental reflection and beamforming. However, in the extreme case where the orientation becomes 90° , we do observe a median throughput drop of 40%.

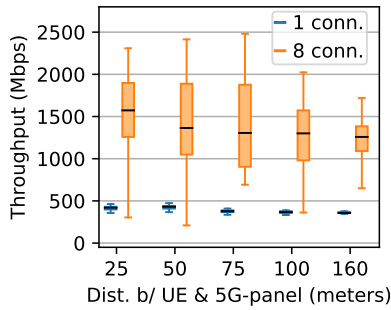


Figure 9: Impact of UE-Panel distance on throughput.

Impact of UE-Panel Distance. We study how the distance between a UE and the panel affects network performance. We conduct the experiment at Location B, where we select five spots with their distances to the panel being 25m, 50m, 75m, 100m, and 160m, respectively. The panel and the five spots are on the same line. At each spot, we conduct three 60-second bulk download tests using 1 and 8 parallel TCP connection(s). During all tests, we ensure that the UE is associated with the same panel, *i.e.*, there is no 5G-4G or 5G-5G handoff (§6.1). We find that at Location D (near the stadium), a UE can also reliably connect to the same panel at a long distance. So we conduct tests there as well using a similar setup.

Figure 9 plots the throughput distributions of different distances, where each box is across all 1-second samples measured at Locations B and D. As shown, the throughput moderately reduces as the distance increases. For 8 parallel TCP connections, the median throughput decreases by 17% (20%) at 100m (160m) compared to that at 25m. We attribute this moderate reduction to the clear LoS and sufficiently high transmission power of 5G antennas. For a single TCP connection, no noticeable throughput drop is observed because of bandwidth under-utilization.

Impact of Rain. Common weather conditions such as rain and snow also affect mmWave performance as its signals are weakened by particles or even moisture [21] in the air. During days with moderate rainfall (between 1.0mm to 2.8mm per-hour)[6, 8], we conduct 2-hour measurements at Location B and near Location C (25m and 50m distance, 0° orientation, 8 parallel TCP connections). We then compare the TCP throughput with that measured on sunny days at the same location with the same setup. We find that at 25m, the median throughput during the rain drops by only 3%, while at 50m, the median throughput reduction is around 30%. The results confirm that rain does degrade 5G throughput.

Overall, among all factors (obstruction, orientation, distance, and weather), obstruction typically incurs the highest impact on mmWave 5G performance. Fortunately, in urban environments, surrounding signal reflections can oftentimes mitigate the performance degradation, allowing 5G to function under NLoS. Based on our experiments, we find that our findings in this subsection also qualitatively apply to **T-Mobile**'s 5G networks. In contrast, for **Sprint**, the environment incurs a much smaller impact on the performance due to the omni-directional nature of its radio signal.

6 MOBILITY AND LOCATION-AWARENESS

Mobility poses a major challenge for mmWave: it will cause continuous changes of the UE-panel distance and orientation; it may

| Type | Description | Sequence |
|-------|--------------------------------------|---------------------------------------|
| P_1 | VH, 5G→4G, same cellID | $C_1 \rightarrow R_1$ |
| P_2 | VH, 4G→5G, same cellID | $R_1 \rightarrow C_1$ |
| P_3 | VH+HH, 5G→4G, diff cellID | $C_1 \rightarrow R_2$ |
| P_4 | HH, 4G→4G, diff cellID | $R_1 \rightarrow R_2$ |
| O_1 | 5G temporarily disrupted, same tower | $P_1 \rightarrow P_2$ |
| O_2 | 5G to 5G between two panels | $P_3 \rightarrow P_2$ |
| O_3 | 5G to 5G between two panels | $P_1 \rightarrow P_4 \rightarrow P_2$ |
| O_4 | 5G to 4G between two panels | $P_1 \rightarrow P_4$ |
| O_5 | 4G to 5G between two panels | $P_4 \rightarrow P_2$ |

Table 3: Primitive (top) & combinational (bottom) handoffs.

incur random LoS/NLoS switches due to obstacles; in particular, it will also lead to more frequent handoffs compared to those in 4G. In this section, we first characterize handoff and mobility performance for 5G (§6.1). We then investigate the feasibility of using a UE's fine-grained location to predict mmWave 5G performance (§6.2).

6.1 Handoff and Mobility Performance

Handoffs in 5G differ from those in 4G/3G in both the horizontal and vertical dimensions. A Horizontal Handoff (HH) occurs when a UE's association switches from one panel (in 5G's term) to another. In 5G, HHs may frequently occur due to the smaller coverage of 5G panels compared to 4G towers. A Vertical Handoff (VH) is triggered when the wireless technology changes (*e.g.*, 5G to 4G). VHs are also prevalent in 5G whose signals are more unstable than 4G.

We closely examine **vZ**'s handoff mechanisms. In 5G NSA, a UE may be in one of the three states: (1) the UE is connected to 5G, (2) the UE is in a 5G coverage area but is connected to 4G due to, for example, poor 5G signals, and (3) the UE is not in a 5G area. Recall that §3 details how they are identified by our monitoring tool. We refer to these states as **C** (Connected to 5G), **R** (Ready for 5G but not yet connected), and **O** (Outside 5G coverage), respectively. We use this state and the cell ID (also collected by our tool) to track both HH and VH. Note that in 5G, cell IDs identify 5G-NR panels.

We then conduct experiments in the three cities under various mobility levels (stationary, walking, and driving) to capture the above data related to handoffs. We identify 4 types of *primitive handoffs* (P_1 to P_4) as listed in the upper part of Table 3. A primitive handoff is between the C and R states as described above. P_1 and P_2 are VHs because they are handoffs between 4G and 5G. When a UE's 5G signal strength drops (*e.g.*, due to a NLoS obstruction), P_1 is triggered to downgrade the connectivity from 5G to 4G; when the network condition improves, the connectivity will be restored back to 5G (P_2). Note that in the 5G-ready (R) mode, the UE actually connects to a 4G radio that is on the same tower (where the 5G-NR panel resides) or a nearby tower, but the cell ID does not change. This is likely because of NSA where 4G and 5G are deeply coupled. At the R state, the UE is still closely monitoring the original 5G panel for a possible 4G to 5G upgrade. P_3 is similar to P_1 except that the 5G to 4G downgrade ends at a different cell ID (panel). P_4 is a 4G to 4G HH from one cell ID to another. We do not observe a $C_1 \rightarrow C_2$ or $R_1 \rightarrow R_2$ sequence in our data. This is because NSA uses 4G for control-plane signaling – the UE will need to first associate with the new cell's 4G radio for control message exchanges before establishing the 5G data channel.

Interestingly, we find that the (in)activity of user traffic can also trigger 4G-5G handoffs. A *P1* handoff will occur when there is an inactivity of user traffic for ~10 seconds; at the *R* state, any user traffic will restore the 5G connectivity through a *P2* handoff, if the 5G signal is good. The rationale of such traffic-guided handoffs is to reduce the 5G standby time that may consume additional energy.

From our data, we observe that oftentimes the primitive handoffs form complex sequences that we call *combinational handoff sequences*. We identify them by clustering primitive handoffs using an interval threshold (set to 10 seconds). They are exemplified at the bottom part of Table 3 as *O1* to *O5*. These combinational sequences correspond to high-level events that cannot be realized by a single primitive handoff. For example, *O3* represents a 5G-to-5G handoff that consists of three primitive handoffs: a 5G to 4G VH on the old panel, a 4G to 4G HH from the old to new panel, and a 4G to 5G VH on the new panel. The whole procedure takes several seconds to finish. We find that **T-Mobile** employs similar mechanisms for 4G/5G and 5G/5G handoffs.

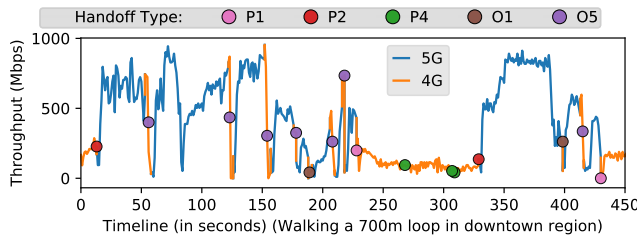


Figure 10: 5G throughput and handoffs under low mobility.

Low Mobility Performance (Walking). We next show a case study to demonstrate the impact of handoffs. In this experiment, one of the authors holds a phone while walking at a normal speed (~5 km/h) at Location A for about 8 minutes. The phone keeps downloading data from a server over 8 parallel connections. Figure 10 plots the throughput, cellular connectivity (4G/5G), and handoffs. During this 8-minute walk, the phone experiences 31 primitive handoffs and bounces between 4G and 5G for 13 times. Such frequent switches make the throughput highly fluctuating, ranging from 0 to 954 Mbps. This may confuse applications (e.g., video rate adaptation logic [32, 35, 56]) and bring highly inconsistent user experiences. These results highlight the need for cross-layer efforts that improve 5G performance under (even low) mobility. They include PHY/MAC enhancements for reducing the handoff frequency, and robust upper-layer solutions that can adapt to frequent 4G/5G handoffs, such as MPTCP [38] and prefetching [28].

Medium Mobility Performance (Driving). Besides walking, we also conduct experiments to study the 5G performance under medium mobility. We put three SGS10 devices near the front windshield of a vehicle, and then drive in midtown Atlanta while measuring the *iPerf* performance (using 8 TCP connections) for **VZ**, **T-Mobile**, and **Sprint** from the three devices. The driving speed was between 20 to 50 km/h. Figure 11 shows representative 2-minute throughput traces simultaneously measured from the three carriers. As shown, both **VZ** and **T-Mobile** exhibit significant throughput fluctuation, with the throughput often dropping close to 0, due to frequent handoffs and blockages incurred by nearby buildings and vehicles. The glasses and body of our car also weaken the mmWave

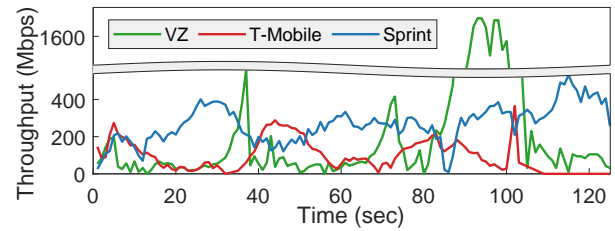


Figure 11: Driving test performance across 3 carriers.

signal. We also find that **T-Mobile** handles handoffs more poorly than **VZ** (e.g., at $t=110$ s, a handoff causes the TCP connections to completely disconnect). In contrast, **Sprint** exhibits higher throughput and lower throughput variation during most of the time, due to the omni-directional nature of its mid-band frequency.

Multipath over 4G and 5G. Given the stability of omni-directional radio in mobility compared to mmWave, a possible way to improve mmWave 5G mobility performance is to use mmWave 5G and 4G simultaneously (e.g., through MPTCP [38]) where 4G can help guarantee the basic data connectivity. To illustrate this idea, we conduct the following experiment. We walk around the Commons Park (a 4.2-acre park in downtown Minneapolis) loop twice per day at different times of a day, for 30 days. During a walk, we capture 5G (including its 4G fallback) and 4G *iPerf* throughput from our phones that are held naturally by a user. Meanwhile, the phones also log the location (using GPS) and handoff events. A typical walk takes about 11 to 14 minutes to complete. We then group the throughput samples into 2-meter segments (bins) along the loop. Figure 13 plots the average and standard deviation (the shaded region) of the 4G/5G throughput samples in each 2-meter segment. As shown, 4G performance is indeed more stable than 5G during an average walk: **StdDev/Mean** across per-segment average throughput is 0.42 and 0.57, respectively, for 4G and 5G. In addition, at certain locations such as those around the 50-th and 75-th segment, 4G outperforms 5G when 5G experiences a blackout. We expect that multipath solutions such as MPTCP over 5G and 4G can mitigate the blackout impact by seamlessly diverting the application traffic from 5G to 4G during a 5G blackout.

6.2 Location-based Performance Estimation

In the 3G/4G era, location is known to be useful for predicting the cellular performance [36, 50]. The prediction, if reasonably accurate, can be utilized in many ways to boost the performance and resource efficiency. We now investigate location-based performance prediction in the mmWave 5G context. The rationale is that fine-grained UE location, which can be easily obtained (at least outdoor through GPS), encodes rich information about the surrounding environment influencing the network performance.

We use the Commons Park walk loop data (Figure 13, §6.1) to investigate the feasibility of location-based performance estimation for mmWave 5G. We find that at a given location, 5G exhibits a statistically higher throughput variation compared to 4G. To show that, let **StdDev(s)** and **Mean(s)** be the stddev and mean of all the throughput samples belonging to a given location s (a 2-meter segment); we compute the average of **StdDev(s)/Mean(s)** across all segments to be 0.70 for 4G and 1.07 for 5G.

In Figure 12, we further visualize the 5G throughput samples as follows. We divide the area into $1m \times 1m$ grids, and distribute

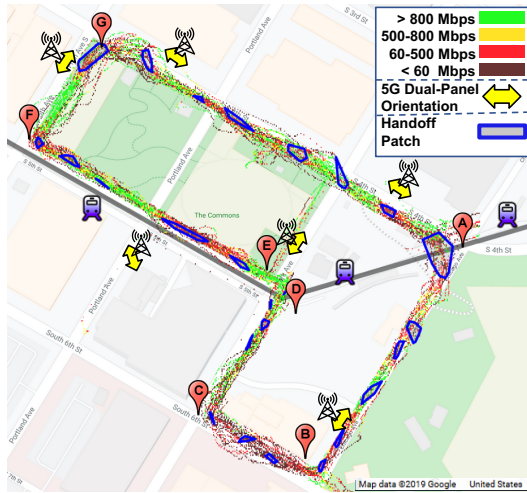


Figure 12: Throughput heatmap and handoff regions.

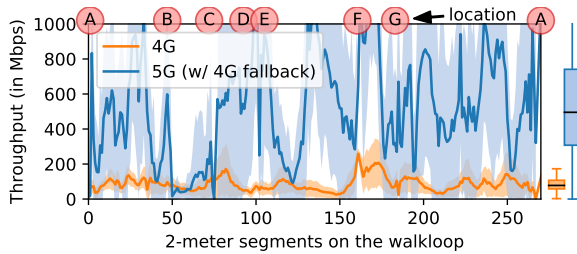


Figure 13: 4G/5G throughput statistics measured from a 30-day field study in the Commons Park. The A to G labels match the corresponding ones in Figure 12.

the throughput samples into the grids based on the location. For each grid, we compute the average throughput, based on which we assign one of the four throughput levels to the grid. We then visualize each grid using a color associated with its throughput level. The blue patches denote the handoff regions that are generated by performing DBSCAN clustering [20] for all locations where handoffs occur, and constructing the convex hull for each cluster. Figure 12 exemplifies two sides: (1) there are some locations where the throughput levels appear to be homogeneous, such as the “dead zone” between B and C due to poor 5G coverage, the good performance zone above F due to simple flat structures facilitating beamforming/LoS, and the bad performance zone at street corner F/G likely due to the complex environment; (2) on the other side, in many other locations there is no clear pattern, such as many spots on the street between A and G and those between C and D.

Our finding suggests that location-based throughput estimation in mmWave 5G faces challenges due to mmWave’s sensitivity to the environment. A small perturbation (e.g., device orientation, nearby moving objects, humidity, and even the phone case [13]) can affect the performance, making the prediction difficult.

Towards Context-aware Performance Prediction. Given the inefficiency of location-based prediction, it can be further generalized into *context-aware* prediction, where a data-driven model uses a wide range of contextual and environmental data such as location, time, mobility level, orientation, weather, and traffic information

to model and predict mmWave 5G performance that is inherently sensitive to the environment. For example, we find that around Location A in Figure 12, a passing train can cause a sudden loss of LoS to the tower. The train schedule information can thus be utilized to facilitate the context-awareness. Data collected by the client can be synthesized at the base station to train a model, which will then be leveraged for robust performance prediction. This approach can potentially outperform the location-only approach and become an important building block that guides decision making at various layers, such as adaptive beam forming, preemptive handoff, multi-path scheduling, and smart prefetching. Meanwhile, there also exist challenges for realizing such a framework, including potentially sparse data, data collection overhead, and privacy issues.

7 APPLICATION PERFORMANCE

This section characterizes application performance over 5G. We investigate two important applications: web page loading and HTTP(S) download.

7.1 Web Page Loading

The World Wide Web is a critical piece in the mobile ecosystem. We utilize the mobile Chrome browser to load the landing or content pages of 17 popular websites listed in Figure 14 over 5G and 4G. We then compare their page loading time (PLT). We conduct the experiments at Locations A and C with SGS10 over 5G and 4G. In Location A (C), the UE-panel distance is 50m (62m) with LoS. We connect the phone to a laptop and use the Chrome remote debugger [7] to automate our tests and to clear the browser’s cache before each loading. We load each page back-to-back over 5G and 4G, and repeat this 4 times for all sites per day, over a few weeks.

As shown in Figure 14, for most sites with small page sizes (≤ 3 MB), 4G and 5G achieve similar PLT. This is attributed to three reasons. First, web browsing requires a synergy between network transfer and local processing, with the latter often being the bottleneck for small web pages particularly [54]. Second, compared to 4G, today’s 5G does not bring improvement on RTT, which is known to be important for web browsing [15]. Third, as we load the pages from the original content providers, the bottleneck may shift from the wireless hop to the Internet. For large pages (>3 MB), loading them over 5G does shorten the median PLT by 5.5% to 19.8% (median: 18.7%), because of the reduction of the content fetch time.

We next investigate the impact of different web protocols. We capture the landing pages of five websites and host them on our controlled server (an Azure server in the east coast running OpenLiteSpeed [9]). We then fetch the pages over 5G with four configurations: HTTP/1.1 without TLS, HTTP/1.1 with TLS, HTTP/2 with TLS, and HTTP/3 (QUIC) with TLS. The experimental locations (A and C) are the same as those for Figure 14. At each location, for each website, we perform 120 back-to-back page loads over 5G in a random order across the four HTTP versions (30 for each version).

As shown in Figure 15, the four HTTP versions indeed exhibit different performance. Recall that from the networking perspective, the major differences among HTTP/1.1, HTTP/2, and HTTP/3 are the transport layer protocol (HTTP/1.1 and HTTP/2 use TCP while HTTP/3 uses UDP) and the connection management scheme (HTTP/1.1 uses concurrent TCP connections while HTTP/2 and HTTP/3 employ a single connection per domain). For four out of

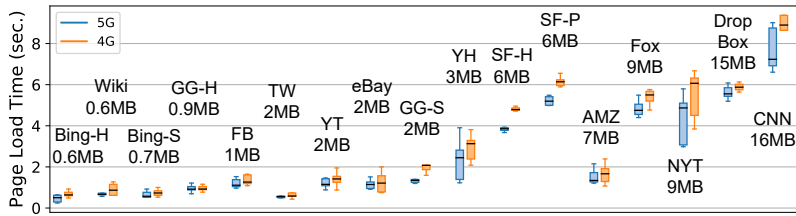


Figure 14: 4G/5G PLT over 17 pages: Bing Home, Wikipedia, Bing Search, Google Home, Facebook, Twitter, YouTube, eBay, Google Search, Yahoo, Salesforce Home, Salesforce Product, Amazon, FoxNews, NYTimes, DropBox, CNN.

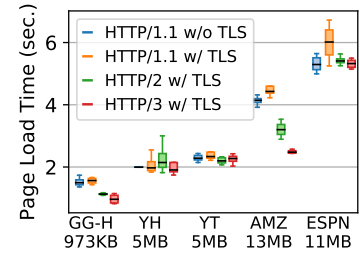


Figure 15: 5G PLT for different HTTP versions across 5 landing pages: Google, Yahoo, YouTube, Amazon, and ESPN.

the five websites, HTTP/2 outperforms HTTP/1.1 (with TLS) mostly due to HTTP/2’s multiplexing feature that consolidates traffic onto one TCP connection per domain, leading to better bandwidth utilization [42]. Recall from §5.1 that a practical issue in commercial 5G is its poor single TCP connection throughput compared to using many concurrently. We find this limitation does not affect HTTP/2 and HTTP/3 performance in our experiments because the small data size, the TCP slow start, and the computation overhead often make the bandwidth under-utilized even over a single connection. HTTP/3 outperforms HTTP/2 for 2 websites and achieves similar median PLT for the other 3 websites. This is attributed to its reduced handshake time as well as its usage of UDP that eliminates the receiver-side HoL blocking across streams [34].

Overall, we find that compared to 4G, today’s 5G brings benefits only for large sites with rich content; meanwhile, the optimizations brought by HTTP/2 and HTTP/3 are effective over 5G. Based on our findings, we identify key improvements for reducing the PLT over 5G: (1) upgrading from NSA 5G to SA 5G to reduce the network latency, (2) accelerating the client-side computation, and (3) bringing the content closer to the edge such as directly caching popular content at the 5G base stations.

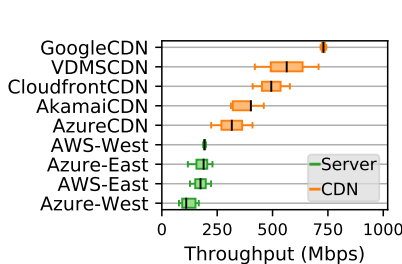


Figure 16: HTTP(S) download throughput for 9 CDN/cloud servers over 5G.

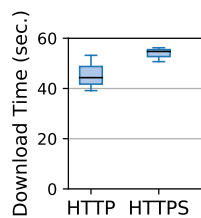


Figure 17: 5G download time (1 GiB file) over HTTP vs. HTTPS.

7.2 HTTP(S) Download

We investigate the HTTP(S) download performance. We upload a 1 GiB file to geographically distributed public cloud instances and CDN servers. We then develop a custom HTTP(S) client that issues 8 parallel byte-range requests each fetching 1/8 of the file over 5G. The experiments are conducted at Locations A and C (clear LoS with a UE-panel distance of 30m for A and 62m for C, 0° orientation). For

each server, we repeat the file download 5 times at both locations, and measure the average throughput.

The results are shown in Figure 16. We find that all cloud/CDN servers exhibit low throughput compared to the *iPerf* throughput shown in Figures 3 and 5: the average throughput ranges from 119 to 730 Mbps with a median of 222 Mbps across all servers. The somewhat surprising results make us realize that HTTP(S) download is very different from *iPerf* bandwidth probing. Multiple factors may slow down HTTP(S) download, such as the DNS time, HTTP request latency, server-side data processing, TLS encryption/decryption, unbalanced byte-range sessions (some sessions may finish earlier than others [25]), inspection performed by firewall/middleboxes, to name a few. Although these factors already exist in 3G/4G eras, they are *amplified* in 5G due to its high speed. Figure 17 exemplifies one such factor: TLS encryption/decryption. In this experiment, we fetch a 1GiB file using HTTP/1.1 and HTTPS/1.1 from the east coast Azure server running OpenLiteSpeed over mmWave 5G (Locations A and C, clear LoS). To ensure a fair comparison, HTTP and HTTPS downloads are performed back-to-back on the same SGS10 device. As shown, using HTTPS increases the average median download time by 23.5%, from 44.3 to 54.7 seconds, indicating considerable runtime overhead incurred by TLS. In contrast, this overhead is barely noticeable in 4G (only ~0.5% difference between HTTP and HTTPS) due to its lower bandwidth compared to 5G. Overall, the above results indicate that mmWave 5G’s high throughput does not always translate to a better application QoE, whose improvement requires joint, cross-layer optimizations from multiple sources.

8 CONCLUDING REMARKS

We conduct a first study that quantitatively reveals 5G performance on COTS smartphones. In addition to the measurement findings, our study identifies key research directions on improving 5G users’ experience in a cross-layer manner. For example, how to design 5G-friendly transport protocols? How to strategically select interface(s) among 5G, 4G, and WiFi? What type of support should a mobile OS provide for enhancing QoE over 5G? Finally, we make our dataset *5Gophers* publicly available to foster the research on 5G.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable comments. We also thank Bo Han for his detailed comments.

This research was in part supported by NSF under Grants CNS-1917424, CNS-1903880, CNS-1915122, CNS-1618339, CNS-1617729, CNS-1814322, CNS-1831140, CNS-1836772, and CNS-1901103.

REFERENCES

- [1] 2019. *3GPP Specification Series 38*. Retrieved October 2019 from <https://www.3gpp.org/DynaReport/38-series.htm>
- [2] 2019. *5G has arrived – here’s what you can expect from Sprint*. Retrieved October 2019 from <https://www.androidauthority.com/what-to-expect-from-sprint-5g-918040/>
- [3] 2019. *5G Spectrum - GSMA Public Policy Position*. Retrieved October 2019 from https://www.gsma.com/spectrum/wp-content/uploads/2019/08/spec_5g_positioning_web_07_19.pdf
- [4] 2019. *Android Police | Verizon 5G spans Sprint’s in new Speedtest.net performance report*. Retrieved September 2019 from <https://www.androidpolice.com/2019/08/14/verizon-5g-spans-sprints-in-new-speedtest-net-performance-report/>
- [5] 2019. *AT&T Enhances Spectrum Position Following FCC Auction 102*. Retrieved October 2019 from https://about.att.com/story/2019/att_enhances_spectrum_position.html
- [6] 2019. *Chicago, IL Weather History | Weather Underground*. Retrieved January 2020 from <https://www.wunderground.com/history/daily/KMDW/date/2019-6-15>
- [7] 2019. *Chrome Remote Debugger*. Retrieved October 2019 from <https://developers.google.com/web/tools/chrome-devtools/remote-debugging>
- [8] 2019. *Fort Snelling, MN Weather History | Weather Underground*. Retrieved January 2020 from <https://www.wunderground.com/history/daily/KMSP/date/2019-6-11>
- [9] 2019. *OpenLiteSpeed Web Server*. Retrieved October 2019 from <https://openlitespeed.org/>
- [10] 2019. *Speedtest by Ookla*. Retrieved October 2019 from <https://www.speedtest.net/>
- [11] 2019. *T-Mobile revs up 5G in 6 cities using mmWave spectrum*. Retrieved October 2019 from <https://www.fiercewireless.com/wireless/t-mobile-revs-up-5g-6-cities-using-mmwave-spectrum>
- [12] 2019. *Verizon’s 5G network now available in New York City and more areas*. Retrieved October 2019 from <https://www.androidauthority.com/verizon-5g-916577/>
- [13] 2019. *Will Your Phone Case Hurt 5G Performance? Maybe*. Retrieved October 2019 from <https://www.pcmag.com/news/370324/will-your-phone-case-hurt-5g-performance-maybe>
- [14] Omid Abari, Dinesh Bharadia, Austin Duffield, and Dina Katabi. 2017. Enabling high-quality untethered virtual reality. In *Proceedings of the 14th USENIX Symposium on Networked Systems Design and Implementation*. 531–544.
- [15] Victor Agababov, Michael Buettner, Victor Chudnovsky, Mark Cogan, Ben Greenstein, Shane McDaniel, Michael Piatek, Colin Scott, Matt Welsh, and Bolian Yin. 2015. Flywheel: Google’s data compression proxy for the mobile web. In *12th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 15)*. 367–380.
- [16] Christopher R Anderson and Theodore S Rappaport. 2004. In-building wideband partition loss measurements at 2.5 and 60 GHz. *IEEE transactions on wireless communications* 3, 3 (2004), 922–928.
- [17] M. Belshe, R. Peon, and Ed. M. Thomson. 2015. *Hypertext Transfer Protocol Version 2 (HTTP/2)*. RFC 7540. Internet Engineering Task Force. <https://tools.ietf.org/html/rfc7540>
- [18] Mike Bishop. 2019. *Hypertext Transfer Protocol Version 3 (HTTP/3) (IETF Draft)*. Retrieved October 2019 from <https://quicwg.org/base-drafts/draft-ietf-quic-http.html>
- [19] Sylvain Collonge, Gheorghe Zaharia, and G EL Zein. 2004. Influence of the human activity on wide-band characteristics of the 60 GHz indoor radio channel. *IEEE Transactions on Wireless Communications* 3, 6 (2004), 2396–2406.
- [20] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *Kdd*, Vol. 96. 226–231.
- [21] Kevin Fogarty. 2019. *Issues In Designing 5G Beamforming Antennas*. Retrieved October 2019 from <https://semiengineering.com/5g-beamforming-antennas-create-design-test-problems/>
- [22] Leica Geosystems. 2019. *Leica DISTO E7500i LDM*. Retrieved October 2019 from <https://lasers.leica-geosystems.com/disto/e7500i>
- [23] Marco Giordani, Marco Mezzavilla, and Michele Zorzi. 2016. Initial Access in 5G mmWave Cellular Networks. *IEEE Communications Magazine* 54, 11 (2016), 40–47.
- [24] Google. 2019. *QUIC, a multiplexed stream transport over UDP*. Retrieved October 2019 from <https://www.chromium.org/quic>
- [25] Yihua Ethan Guo, Ashkan Nikravesh, Z Morley Mao, Feng Qian, and Subhabrata Sen. 2017. Accelerating multipath transport through balanced subflow completion. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 141–153.
- [26] Muhammad Kumail Haider, Yasaman Ghasempour, Dimitrios Koutsonikolas, and Edward W Knightly. 2018. Lister: mmwave beam acquisition and steering by tracking indicator leds on wireless aps. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 273–288.
- [27] Daniel Halperin, Srikanth Kandula, Jitendra Padhye, Paramvir Bahl, and David Wetherall. 2011. Augmenting data center networks with multi-gigabit wireless links. In *ACM SIGCOMM Computer Communication Review*, Vol. 41. ACM, 38–49.
- [28] Brett D Higgins, Jason Flinn, Thomas J Giulii, Brian Noble, Christopher Peplin, and David Watson. 2012. Informed mobile prefetching. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 155–168.
- [29] Anders Hillbur. 2018. *5G deployment options to reduce the complexity*. Retrieved October 2019 from <https://www.ericsson.com/en/blog/2018/11/5g-deployment-options-to-reduce-the-complexity>
- [30] Junxian Huang, Feng Qian, Yihua Guo, Yuanyuan Zhou, Qiang Xu, Z Morley Mao, Subhabrata Sen, and Oliver Spatscheck. 2013. An in-depth study of LTE: effect of network protocol and application behavior on performance. In *ACM SIGCOMM*.
- [31] Haiqing Jiang, Yaogong Wang, Kyunghan Lee, and Injong Rhee. 2012. Tackling bufferbloat in 3G/4G networks. In *Proceedings of the 2012 Internet Measurement Conference*. ACM, 329–342.
- [32] Junchen Jiang, Vyas Sekar, and Hui Zhang. 2012. Improving Fairness, Efficiency, and Stability in HTTP-Based Adaptive Video Streaming With Festive. In *Proceedings of CoNEXT 2012*. ACM, 97–108.
- [33] ESnet/Lawrence Berkeley National Laboratory. 2019. *iperf3: A TCP, UDP, and SCTP network bandwidth measurement tool*. Retrieved October 2019 from <https://github.com/esnet/iperf/releases/tag/3.6>
- [34] Adam Langley, Alistair Riddoch, Alyssa Wilk, Antonio Vicente, Charles Krasnic, Dan Zhang, Fan Yang, Fedor Kouranov, Ian Swett, Janardhan Iyengar, et al. 2017. The quic transport protocol: Design and internet-scale deployment. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*. ACM, 183–196.
- [35] Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh. 2017. Neural adaptive video streaming with pensieve. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*. ACM, 197–210.
- [36] Robert Margolies, Ashwin Sridharan, Vaneet Aggarwal, Rittwik Jana, NK Shankaranarayanan, Vinay A Vaishampayan, and Gil Zussman. 2016. Exploiting mobility in proportional fair cellular scheduling: Measurements and algorithms. *IEEE/ACM Transactions on Networking (TON)* 24, 1 (2016), 355–367.
- [37] Motorola. 2019. *5G Moto Mod*. Retrieved October 2019 from <https://www.motorola.com/us/products/moto-mods/moto-5g>
- [38] Ashkan Nikravesh, Yihua Guo, Feng Qian, Z Morley Mao, and Subhabrata Sen. 2016. An in-depth understanding of multipath TCP on mobile devices: measurement and system design. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 189–201.
- [39] Thomas Nitsche, Guillermo Bielsa, Irene Tejado, Adrian Loch, and Joerg Widmer. 2015. Boon and bane of 60 GHz networks: practical insights into beamforming, interference, and frame level operation. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. ACM, 17.
- [40] Thomas Nitsche, Adriana B Flores, Edward W Knightly, and Joerg Widmer. 2015. Steering with eyes closed: mm-wave beam steering without in-band measurement. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2416–2424.
- [41] Joan Palacios, Danilo De Donno, and Joerg Widmer. 2017. Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility. In *Proceedings of the IEEE Conference on Computer Communications*.
- [42] Feng Qian, Vijay Gopalakrishnan, Emir Halepovic, Subhabrata Sen, and Oliver Spatscheck. 2015. TM3: flexible transport-layer multi-pipe multiplexing middle-box without head-of-line blocking. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. ACM, 3.
- [43] Hang Qiu, Fawad Ahmad, Fan Bai, Marco Gruteser, and Ramesh Govindan. 2018. AVR: Augmented Vehicular Reality. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 81–95.
- [44] Qualcomm. 2019. *Everything You Need to Know About 5G*. Retrieved October 2019 from <https://www.qualcomm.com/invention/5g/what-is-5g>
- [45] Qualcomm. 2019. *Mobilizing mmWave for smartphones*. Retrieved October 2019 from <https://www.qualcomm.com/news/onq/2019/02/13/track-solve-another-impossible-challenge-mobilizing-mmwave-smartphones>
- [46] Theodore S Rappaport, Eshar Ben-Dor, James N Murdock, and Yijun Qiao. 2012. 38 GHz and 60 GHz angle-dependent propagation for cellular & peer-to-peer wireless communications. In *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 4568–4573.
- [47] Theodore S Rappaport, Felix Gutierrez, Eshar Ben-Dor, James N Murdock, Yijun Qiao, and Jonathan I Tamir. 2013. Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. *IEEE transactions on antennas and propagation* 61, 4 (2013), 1850–1859.
- [48] Theodore S Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N Wong, Jocelyn K Schulz, Mathew Samimi, and Felix Gutierrez. 2013. Millimeter wave mobile communications for 5G cellular: It will work! *IEEE*

- access 1 (2013), 335–349.
- [49] Wonil Roh, Ji-Yun Seol, Jeongho Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. 2014. Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results. *IEEE Communications Magazine* 52, 2 (2014), 106–113.
- [50] Aaron Schulman, Vishnu Navda, Ramachandran Ramjee, Neil Spring, Pralhad Deshpande, Calvin Grunewald, Kamal Jain, and Venkata N Padmanabhan. 2010. Bartendr: a practical approach to energy-aware cellular data scheduling. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*. ACM, 85–96.
- [51] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. 2017. Wi-Fi-Assisted 60 GHz Wireless Networks. In *MobiCom*.
- [52] Sanjib Sur, Vignesh Venkateswaran, Xinyu Zhang, and Parmesh Ramanathan. 2015. 60 ghz indoor networking through flexible beams: A link-level profiling. In *ACM SIGMETRICS Performance Evaluation Review*, Vol. 43. ACM, 71–84.
- [53] Xiaozheng Tie, Kishore Ramachandran, and Rajesh Mahindra. 2012. On 60 ghz wireless link performance in indoor environments. In *Passive and Active Measurement*. Springer, 147–157.
- [54] Xiao Sophia Wang, Aruna Balasubramanian, Arvind Krishnamurthy, and David Wetherall. 2013. Demystifying page load performance with WProf. In *Presented as part of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 13)*. 473–485.
- [55] Hao Xu, Vikas Kukshya, and Theodore S Rappaport. 2002. Spatial and temporal characteristics of 60-GHz indoor channels. *IEEE Journal on selected areas in communications* 20, 3 (2002), 620–630.
- [56] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A control-theoretic approach for dynamic adaptive video streaming over HTTP. In *ACM SIGCOMM Computer Communication Review*, Vol. 45. ACM, 325–338.
- [57] Hang Zhao, Rimma Mayzus, Shu Sun, Mathew Samimi, Jocelyn K Schulz, Yaniv Azar, Kevin Wang, George N Wong, Felix Gutierrez Jr, and Theodore S Rappaport. 2013. 28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city.. In *ICC*. 5163–5167.
- [58] Xia Zhou, Zengbin Zhang, Yibo Zhu, Yubo Li, Saipriya Kumar, Amin Vahdat, Ben Y Zhao, and Haitao Zheng. 2012. Mirror mirror on the ceiling: Flexible wireless links for data centers. *ACM SIGCOMM Computer Communication Review* 42, 4 (2012), 443–454.
- [59] Yibo Zhu, Zengbin Zhang, Zhinus Marzi, Chris Nelson, Upamanyu Madhoo, Ben Y Zhao, and Haitao Zheng. 2014. Demystifying 60GHz outdoor picocells. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 5–16.
- [60] Yibo Zhu, Xia Zhou, Zengbin Zhang, Lin Zhou, Amin Vahdat, Ben Y Zhao, and Haitao Zheng. 2014. Cutting the cord: a robust wireless facilities network for data centers. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 581–592.