



The First 5G-LTE Comparative Study in Extreme Mobility

YUEYANG PAN*, RUIHAN LI*, and CHENREN XU, Peking University, China

5G claims to support mobility up to 500 km/h according to the 3GPP standard. However, its field performance under high-speed scenes remains in mystery. In this paper, we conduct the first large-scale measurement campaign on a high-speed railway route operating at the maximum speed of 350 km/h, with full coverage of LTE and 5G (NSA and SA) along the track. Our study consumed 1788.8 GiB of cellular data in six months, covering the three major carriers in China and the recent standardized QUIC protocol. Based on our dataset, we reveal the key characteristics of 5G and LTE in extreme mobility in terms of throughput, RTT, loss rate, signal quality, and physical resource utilization. We further develop a taxonomy of handovers in both LTE and 5G and carry out the link-layer latency breakdown analysis. Our study pinpoints the deficiencies in the user equipment, radio access network, and core network which hinder seamless connectivity and better utilization of 5G's high bandwidth. Our findings highlight the directions of the next step in the 5G evolution.

CCS Concepts: • Networks → Network measurement; Network performance analysis.

Additional Key Words and Phrases: Measurement; Extreme Mobility; 5G; High-speed Rail; Handover

ACM Reference Format:

Yueyang Pan, Ruihan Li, and Chenren Xu. 2022. The First 5G-LTE Comparative Study in Extreme Mobility. *Proc. ACM Meas. Anal. Comput. Syst.* 6, 1, Article 20 (March 2022), 22 pages. <https://doi.org/10.1145/3508040>

20

1 INTRODUCTION

5G has already reshaped both industry and people's life since it hit the market in 2019. Compared to 4G LTE, 5G is expected to provide more than 10× bandwidth gain [22] and ultra-low latency down to 1 ms [32], which could fuel a wide range of applications (*e.g.*, web surfing and 4k/8k videos). Particularly, according to 3GPP TR 38.913 [2], 5G supports extreme mobility up to 500 km/h, enlightening the future of delivering seamless mobile networking connectivity on the high-speed rails (HSR), which is actively revolutionizing the means of modern transportation for short-distance inter-city travel in many countries. However, the 5G performance under such extreme mobility remains in mystery. As previous studies have been focusing on LTE in extreme mobility [18, 31, 33] or 5G in static and low mobility [23, 24, 34], it is more important than ever to quantify the performance baseline of 5G on HSR given that the construction of HSR systems is speeding up worldwide. In China, the future blueprint emphasizes eliminating the insufficiency of HSR transportation between the key areas and metropolises [26]. Both the authorities in the US [6] and the social associations in Europe [15] have been advocating the benefit of the HSR.

In this paper, we fill out the blankness in the landscape of the 5G measurement. We carried out the first measurement study comprehensively covering all the modern radio access technologies (RAT)

*Both student authors contributed equally to this work.

Chenren Xu is the corresponding author.

Authors' address: Yueyang Pan, pyjason@pku.edu.cn; Ruihan Li, lrh2000@pku.edu.cn; Chenren Xu, chenren.xu@gmail.com, Peking University, No.5 Yiheyuan Road, Haidian District, Beijing, China, 100871.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

2476-1249/2022/3-ART20 \$15.00

<https://doi.org/10.1145/3508040>

including LTE, 5G non-standalone (NSA), and 5G standalone (SA) in extreme mobility. The study spanned half a year and focused on a 174 km HSR operating at the maximum speed of 350 km/h with full 5G (and LTE) coverage [8] from Beijing to Zhangjiakou. We created a dataset collected from three user equipment (UEs) with 1788.8 GiB cellular data usage covering all the major mobile carriers in China. Our measurement study consists of five parts as detailed below.

- §3 Leveraging our unique dataset, we begin with comparing key performance metrics (*e.g.*, throughput, RTT, and loss rate) between two transport layer protocols TCP and QUIC. We observe an $11.0 \times$ TCP throughput gain in 5G in comparison to LTE. However, QUIC achieves only $7.1 \times$ throughput gain in 5G and it also has a higher loss rate. We attribute this suboptimal performance to the inefficient BBR implementation and ISP filter policies for UDP-based traffic [10].
- §4.1 We then look into the (wireless) channel condition and physical resource utilization aspects at a lower layer. We find that by using 5G, UE can tolerate poorer channel conditions, *e.g.*, -96.92 dBm vs. -88.44 dBm in RSRP and 12.05 dB vs. 14.65 dB in SNR (all in median values) in comparison to LTE. In addition, we observe that 5G has a higher and more stable physical resource (block) allocation scheme than LTE under different channel conditions.
- §4.2 Given the substantial impact of handovers under extreme mobility, we conduct a comprehensive handover study of LTE, 5G NSA, and 5G SA. We develop the first taxonomy of 5G handovers according to the control signaling messages and quantify the frequency, interval, and duration of handovers in both 5G and LTE. We observe that successful 5G SA handovers have a more significant impact than that of LTE. For example, when comparing the average throughput 1 second before and after the handover, 5G yields a 28.6% degradation compared with only 5.0% in LTE. 5G also suffers from 82.8% longer handover duration and prolongs the throughput recover time by 50% than LTE.
- §4.3 We next explore how long and how a packet gets delayed in the radio access network (RAN). We break down the RTT and find RAN latency accounts for a considerable share, which causes the RTT fluctuations in both 5G and LTE. In extreme mobility, 5G has reduced the RAN latency from 40.0 ms to 15.8 ms on average compared to LTE and performs reasonably well even under poor signal strength and quality. Together with §4.1, we conclude that 5G is more robust to (wireless) channel condition degradation under extreme mobility.
- §5 Finally, we examine the performance of web page loading and adaptive bitrate (ABR) algorithms in video streaming. We observe that the introduction of 5G reduces the waiting time of objects of all sizes but SA prolongs the DNS lookup time and the time to first byte (TTFB) in some cases, which indicates immaturity of the 5G core network. Among all the ABRs in our experiment, L2A [17] achieves the lowest but still non-negligible 2.2% stall percentage in 5G SA. Meanwhile, the most aggressive ABR RB [3] comes with the cost of a higher stall percentage (5.9% in 5G SA).

This study represents the first large-scale field measurement campaign on the HSR covering multiple layers (application, transport, and link layer), different RATs (LTE and 5G), and both 5G deployment modes (NSA and SA). The contributions can be summarized as follows:

- We set a quantitative performance baseline for TCP and QUIC on HSR under different RATs and deployment modes.
- We measure the physical channel condition and resource utilization as well as uncover 5G's robustness under poor signal quality and strength.
- We break down the E2E RTT into several components and provide suggestions for future optimization towards low latency.
- We develop the taxonomy of LTE, 5G NSA, and 5G SA handovers, quantify their frequency, interval, and duration, and analyze their impact on the throughput.
- We compare a plethora of ABR algorithms and pinpoint their deficiency in extreme mobility.

2 MEASUREMENT METHODOLOGY

In this section, we present our data collection experimental platform for 5G-LTE comparative performance study (§2.1) as well as the experimental setup to collect the key metrics across different layers (§2.2). The key statistics of our dataset are shown in Tab. 1.

	Carrier A	Carrier B	Carrier C	Total	Travel mileage
SA	294.19 GiB	371.40 GiB	487.36 GiB	1152.96 GiB	2088 km/UE
NSA	140.95 GiB	184.59 GiB	242.98 GiB	568.52 GiB	1914 km/UE
LTE	18.02 GiB	28.52 GiB	20.73 GiB	67.27 GiB	1392 km/UE
Total	453.2 GiB	584.5 GiB	751.1 GiB	1788.8 GiB	5394 km/UE

Table 1. Statistics of the cellular data consumed on 3 major mobile carriers.

2.1 Experimental platform

At a high level, our experimental platform (Fig. 1) consists of two parts.

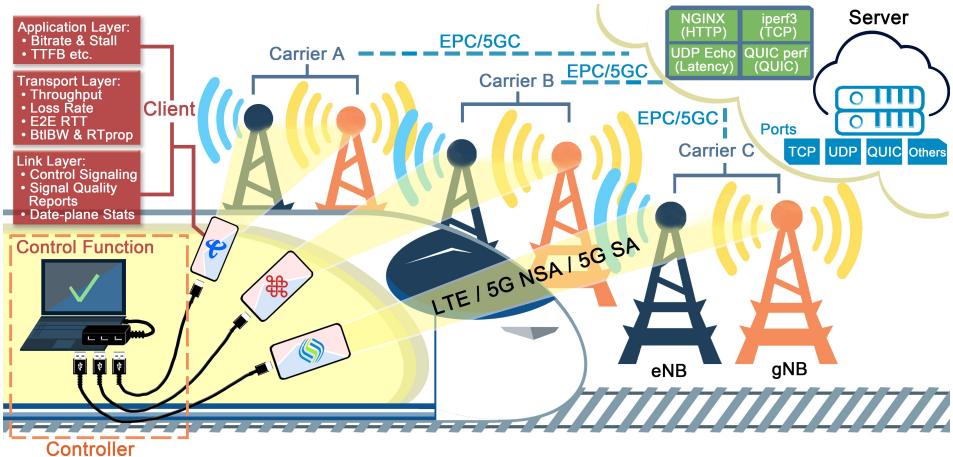


Fig. 1. The experimental platform that collects data from multiple layers on each UE.

Server. We deploy a high-performance cloud server (4vCPU | 8GiB | 2Gbps) running Ubuntu 20.4 LTS with Linux kernel 5.4 LTS. The cloud server is provided by one of the major vendors in China and is located near the HSR route to minimize the impact of Internet congestion. We bind the server to bare metal to avoid the difference inside the data center and set enough egress bandwidth. We use `iperf 3.9` and `LSQUIC 2.29` as the TCP and QUIC servers. In configuring the congestion control algorithm (CCA), we choose BBR over CUBIC (and other CCA candidates) due to its loss tolerance under extreme mobility [31] given the observed 5G's high loss rate [34].

Client. We leverage three identical mobile phones (Redmi K30 Pro) running Android 11 as our clients to minimize their heterogeneity in software and hardware. In each trip, all the phones are configured the same mode, either LTE, 5G NSA, or 5G SA, but operating on different mobile carriers¹ to avoid radio resource contention across different RATs and modes during the data collection procedure. Specifically, we leverage a laptop as a centralized controller and develop a series of Python and shell

¹Both LTE and 5G services are deployed within different frequency ranges or bands among the mid-band frequencies (ranged in 1805 MHz - 3600 MHz) by the three carriers.

scripts that communicate with our client devices (via adb and ssh) and our cloud server (via ssh) to automate and synchronize the experiment procedure among the phones. Note that we use USB to connect the laptop and phones only for the purpose of synchronizing the phones. As 5G promises an ultra-high bandwidth that may exceed the capability of USB 3.0, our experiments are carried out in Ubuntu installed inside our Android devices via Linux Deploy [4] instead of adopting the legacy USB tethering approach [31]. To capture fine-grained cellular event messages for cross-layer analysis, we instrument `diag_logcat`, a tool modified from MobileInsight's `diag_revealer` [20] to log the baseband information relevant to the channel condition (*e.g.*, RSRP and SNR), radio resource utilization, handover and link-layer frame delay in data plane.

2.2 Experiment types

We carry out four types of cross-layer experiments based on our experimental platform.

Bandwidth probe. We use `iperf` and `LSQUIC` to generate full-speed TCP and QUIC streams to probe the end-to-end bandwidth. We leverage `tcpdump` to capture IP packets and extract the key performance metrics such as the throughput and loss rate. We modify `iperf`, Linux kernel, and `LSQUIC` so that they log important performance indicators that are used to model the network conditions such as RTT and other internal parameters in the CCA. The results will be shown in §3.1.

Latency probe. All clients send a customized UDP packet sequence with an inter-packet interval of 50 ms. The packets will be echoed by the server. `tcpdump` records E2E RTTs of the packets and `diag_logcat` captures the logs regarding how a frame gets delayed in different phases in the data plane such as segmentation and retransmission. The breakdown analysis will be shown in §4.3.

Web surfing. We visit the websites selected from the Alexa top 500 websites every 10 seconds using Selenium and Chrome and simulate the user behavior by JS snippets. We leverage the ChromeDriver [14] to log the page load time. The details are elaborated in §5.1.

Video streaming. We test several ABR algorithms for a multi-track (360p to 8K) video with `DASH.js` [9]. We capture browser logs via ChromeDriver to record the timeseries of bitrates and stall events. We will introduce the testing ABR algorithms and encoded video in §5.2.

2.3 Ethical Considerations

This study does not involve any personally identifiable information. We purchase multiple cellular data plans from the market and follow the customer agreements of the carriers. Our study complies with the *Regulation on the Administration of Railway Safety*. No ethical issue is raised by our study.

3 TRANSPORT-LAYER PERFORMANCE

In this section, we examine the results of the *bandwidth probe* experiment. We first study the key metrics including the throughput, loss rate, and RTT (§3.1), and then make an in-depth analysis of the parameters in CCA (BBR) by case study (§3.2).

3.1 TCP and QUIC Protocol

We start by answering the question of how much performance gain 5G brings under extreme mobility, especially from the TCP and QUIC perspectives.

Throughput (Fig. 2). We make three observations across the three carriers on average: (*i*) 5G SA promotes throughput of TCP from 20.2 Mbps (LTE) to 222.6 Mbps by more than 11.0 \times , which evidently demonstrates that it holds its claim of 10 times bandwidth boost over LTE even in extreme mobility. (*ii*) NSA performs worse than SA – its throughput of 138.7 Mbps (and 117.7 Mbps) is 37.7% (and 18.0%) lower than that of SA in TCP (and QUIC). This is because when UE encounters a handover failure (to be explained in §4.2.1), it may fall back to LTE for data transmission before reestablishing the connection with a 5G base station. (*iii*) QUIC performs worse than TCP, especially

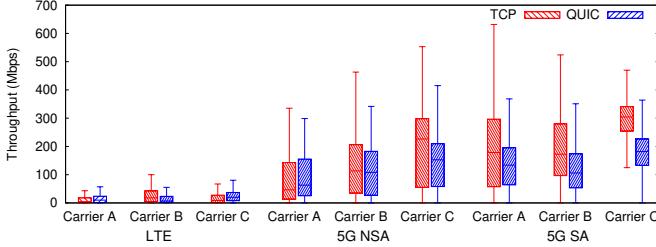


Fig. 2. TCP-QUIC comparative throughput across different carriers in LTE, 5G NSA, and 5G SA.

in SA – the throughput of QUIC in SA is 143.5 Mbps, 35.5% lower than that of TCP. Given that the throughput of QUIC (20.3 Mbps) in LTE is similar to that of TCP (20.2 Mbps), the throughput gain of 5G SA drops to only 7.1 times for QUIC in comparison to LTE. We find two major reasons – QUIC is more lossy than TCP (to be shown in §3.1) and QUIC BBR migrated from Google is less aggressive than TCP BBR (see §3.2).

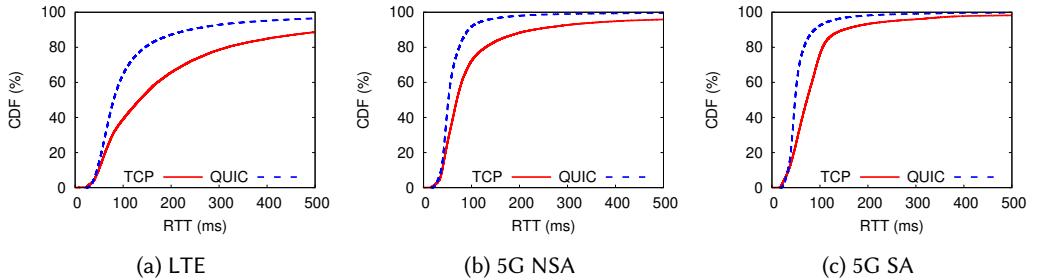


Fig. 3. TCP-QUIC comparative CDFs of RTTs in LTE, 5G NSA, and 5G SA.

RTT (Fig. 3). We can see from the figures that RTT in LTE is longer and more fluctuated than that in 5G. The median (95th-percentile) RTT of TCP is 134.2, 69.4, and 73.0 ms (931.9, 408.1, and 251.6 ms) respectively in LTE, 5G NSA, and 5G SA. An important reason is that the RAN latency of 5G is shorter and more stable than that of LTE (to be shown in §4.3). Meanwhile, we can learn that QUIC has a shorter RTT than TCP. For example in 5G SA, the median (95th-percentile) RTT of TCP and QUIC is 48.2 and 73.0 ms (121.1 and 251.6 ms) respectively. It accords with TCP's more aggressive behavior in throughput as mentioned previously.

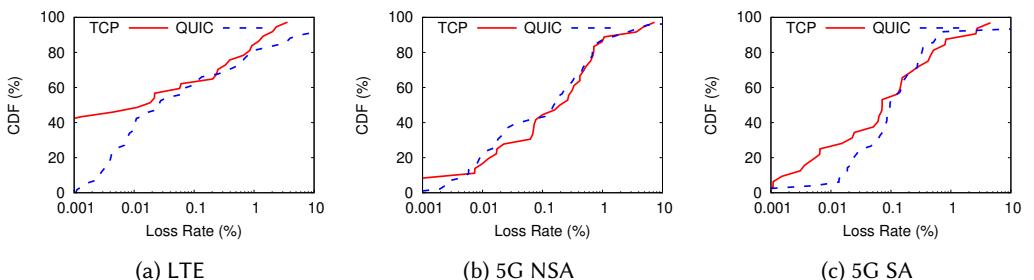


Fig. 4. TCP-QUIC comparative CDFs of loss rates in LTE, 5G NSA, and 5G SA

Loss Rate (Fig. 4). We make two key findings from the figures. (i) 5G is more lossy than LTE. Taking TCP as an example, 48.2% of the traces in LTE have a loss rate less than 0.01%, while this number drops to 16.6% and 26.2% for 5G NSA and SA respectively. (ii) QUIC is more lossy than TCP. In 5G SA, 26.2% of the TCP traces have loss rates less than 0.01% but only 5.2% of the QUIC traces meet the same requirement. Note that in QUIC, we also observe a few extremely lossy traces (*i.e.*, loss rate $\geq 5\%$), occupying 10.2%, 4.4%, and 6.2% of all the testing traces in LTE, 5G NSA and 5G SA respectively – which do not present in TCP. Since QUIC protocol is designed based on the UDP protocol and today it is a common QoS policy of the ISPs to discard (some of) the UDP packets [10] from a bursty session. It is the first reason why QUIC has lower throughput than TCP.

3.2 Deep Inspection of BBR Internals

Based on our findings of the metrics, we investigate each trace and inspect its internal parameters. We will first describe how BBR works and introduce its key parameters. Next, we analyze the parameters to show the reason for different behaviors between TCP and QUIC.

BBR Preliminaries. BBR estimates the network capacity and decides its congestion window by evaluating two key parameters, round trip propagation time (*RTProp*) and bottleneck bandwidth (*BltBw*) [12]. BBR takes the multiplication of *BltBw* and *RTprop* to estimate the bandwidth-delay product (*BDP*), and leverages *BDP* to estimate the maximum number of inflight bytes with minimized packets queuing in the network and set *CWND* accordingly.

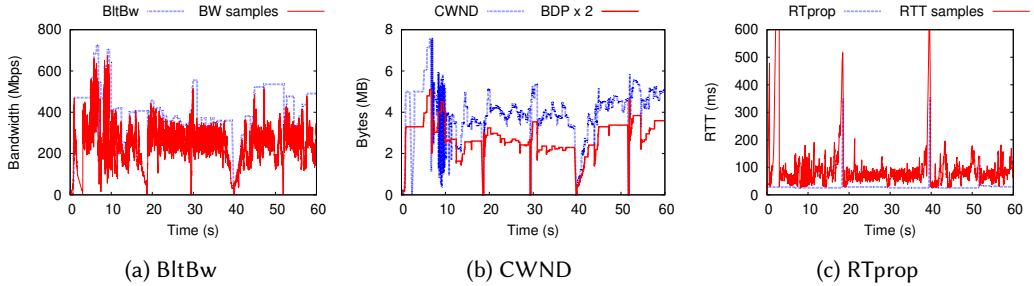


Fig. 5. Snapshot timeseries of internal parameters in TCP BBR.

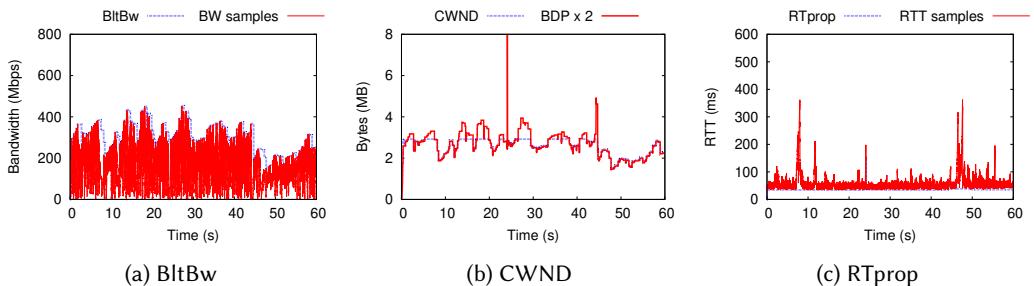


Fig. 6. Snapshot timeseries of internal parameters in QUIC BBR.

TCP BBR (Fig. 5). An example trace of TCP is shown in Fig. 5. Here bandwidth and RTT samples are used by BBR to estimate *BltBw* and *RTprop*, which are extracted from the Linux kernel. We notice in Fig. 5a that BBR samples bandwidth every RTT (which is usually less than 100 ms), and therefore bandwidth samples suffer from high fluctuation. Since the maximum sample is taken as

BltBw, BBR overestimates the bottleneck bandwidth and consequently the *BDP*. In Linux kernel, the congestion window (*CWND*) is set to $2 \times BDP$ plus offset terms to keep sending data during inter-ACK silences caused by ACK aggregation. We observe from Fig. 5b the *CWND* is set to be more aggressive than $2 \times BDP$. As a result, it achieves almost full utilization of bandwidth but causes longer RTT, as shown in Fig. 5c.

QUIC BBR (Fig. 6). An example trace of QUIC is shown in Fig. 6. Similar to TCP, QUIC also overestimates the bottleneck bandwidth as shown in Fig. 6a. However, interestingly, we find in Fig. 6b that *CWND* is truncated by a threshold of 2000 packets. By diving deep into the LSQUIC codebase implementation, we discover that LSQUIC inherits the *CWND* limitation [21] from Google’s production-ready version of QUIC [13], named QUICHE and used in Chromium, Google’s server, etc. We guess it is introduced for legacy reasons but it is not suitable today as the bandwidth capacity continuously grows. On the other hand, QUIC BBR uses $2 \times BDP$ as the *CWND* if not limited by the 2000 cap, which is less aggressive than TCP BBR. The limited *CWND* downgrades the throughput but provides shorter RTT as shown in Fig. 6c in comparison to TCP BBR. The difference in BBR implementations serves as the second reason for QUIC’s poor throughput.

3.3 Implications

5G delivers noticeably higher ($11.0 \times$) TCP throughput than LTE, even in extreme mobility. Our study also reveals that QUIC, a standardized promising future transport protocol with product-level implementation, is still under-utilizing the available high bandwidth attributing to the deficiency of implementations at the endpoints and ad-hoc QoS policies inside the Internet. We conclude that end-to-end evolution is necessary to maximize the utilization of ultra-high bandwidth.

4 LINK-LAYER PERFORMANCE

We next examine the link layer by studying the characteristics of the wireless channel and their relationship with throughput (§4.1). Then we develop the handover taxonomy of LTE, 5G NSA, and 5G SA to study their properties and how they impact the end-to-end throughput (§4.2). Finally, we quantify and analyze the link-layer latency breakdown and demonstrate how it is influenced by signal quality and handovers (§4.3).

4.1 RSRP, SNR, and Resource Block (RB)

We start by demystifying the differences in operational wireless channel condition and physical resource utilization between LTE and 5G, and then present how these differences amplify the throughput gap between them.

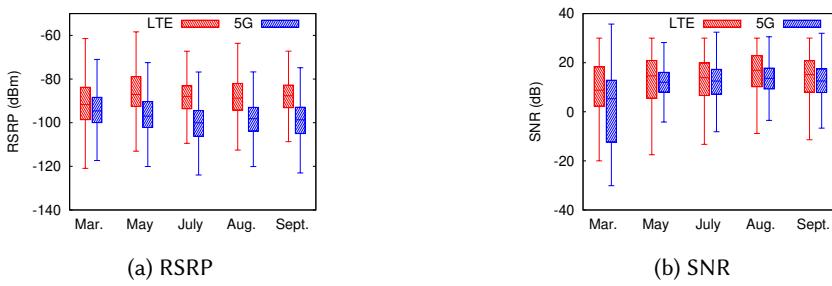


Fig. 7. RSRP and SNR statistics across half a year in 2021.

RSRP and SNR Statistics. We first characterize the channel condition from two angles, *i.e.*, the signal strength quantified by Reference Signal Received Power (RSRP) and the signal quality

measured by Signal-to-noise Ratio (SNR). Based on our dataset, we demonstrate the statistics of RSRP on the HSR in Fig. 7a and SNR in Fig. 7b across all the three carriers. We make the following observations. (i) During our experiment in different months, the signal strength received by the UE is fairly stable. The max-min ranges of means are 4.63 and 5.5 dBm for LTE and 5G respectively. (ii) The UE is operating under lower signal strength and lower signal quality in 5G compared to LTE. The median received RSRP and SNR of LTE are -88.44 dBm and 14.65 dB and those of 5G are -96.92 dBm and 12.05 dB.

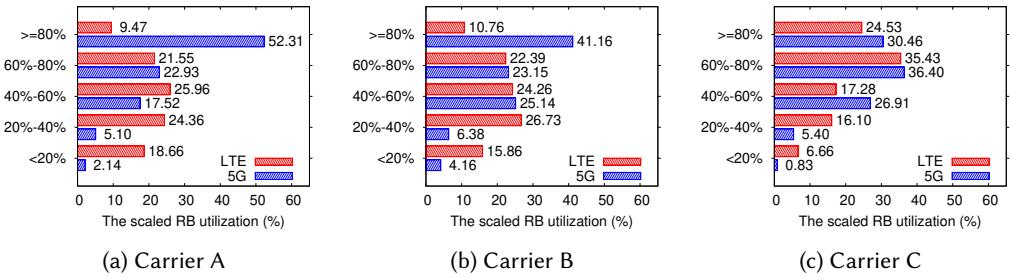


Fig. 8. The distribution of the scaled RB utilization in TCP.

RB Utilization Distribution. The base station allocates physical resources to the UE in the unit of resource blocks (RBs) so that UE can proportionally transmit or receive data accordingly. We define *scaled DL RB utilization* as the ratio of the number of allocated downlink (DL) RBs to the number of all RBs that can serve as DL RBs². We demonstrate the scaled DL RB utilization of our TCP traces for the three carriers in Fig. 8. We divide the utilization to five categories, namely *Poor* [0, 20%), *Fair* [20%, 40%), *Average* [40%, 60%), *Good* [60%, 80%), and *Excellent* [80%, 100%). We observe that the LTE base stations allocate fewer radio resources (in percentage) to the UEs than the 5G ones. In 5G, 75.2%, 64.3%, and 66.8% of utilization samples of Carrier A, B, and C respectively fall into the *Good* and *Excellent*. In LTE, however, the proportion drops to 31.0%, 33.2%, and 60.0% respectively, and samples with *Poor* utilization account for a much larger percentage.

RB Utilization and Channel Condition. Then we investigate where the allocation schemes that lead to low DL RB utilization reside. Since we neither have access to the base station nor know how the RB allocation algorithm works, we study the relationship between the utilization and channel condition in terms of RSRP and SNR. In Fig. 9, we show the RB utilization under different SNR and RSRP. Similar to RB utilization, we divide RSRP and SNR into five categories (RSRP: *Poor*, $[-140, -105]$ dBm; *Fair*, $[-105, -95]$ dBm; *Average*, $[-95, -85]$ dBm; *Good*, $[-85, -75]$ dBm; *Excellent*, $[-75, -40]$ dBm. SNR: *Poor*, $[-30, -5]$ dB; *Fair*, $[-5, 5]$ dB; *Average*, $[5, 15]$ dB; *Good*, $[15, 25]$ dB; *Excellent*, $[25, 35]$ dB). Compared to LTE, the RB utilization in 5G is generally higher and more stable under different wireless channel conditions. Note that the same RB utilization under a better channel condition means higher order of modulation and coding scheme (MCS) and data rates. Interestingly, we find that LTE actually treats UE under poorer channel conditions even poorer in terms of allocating fewer RBs, leading to an even lower data rate.

Throughput vs. Channel Condition. Under the poor channel, the throughput will drop because of the more conservative rate adaption (MCS) strategy. We show how the normalized throughput is affected by RSRP and SNR in Fig. 10. Generally, throughput increases with higher RSRP and SNR.

²The original RB utilization is scaled by the maximum utilization in FDD and TDD. In TDD, only part of the RBs could be allocated for one-way communication.

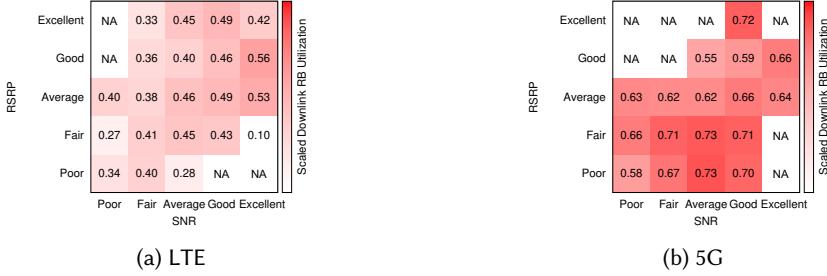


Fig. 9. The scaled downlink RB utilization in TCP with respect to RSRP and SNR.

In 5G, with sufficient resources, the throughput increases quickly as long as SNR or RSRP rises. For instance, given the *Poor* RSRP, lifting SNR from *Poor* to *Average* quadruples up the throughput (0.1 vs. 0.43). However, the resources in LTE are scarce. The normalized throughput stays low under *Poor* SNR (< 0.1) and *Poor* RSRP (< 0.2).

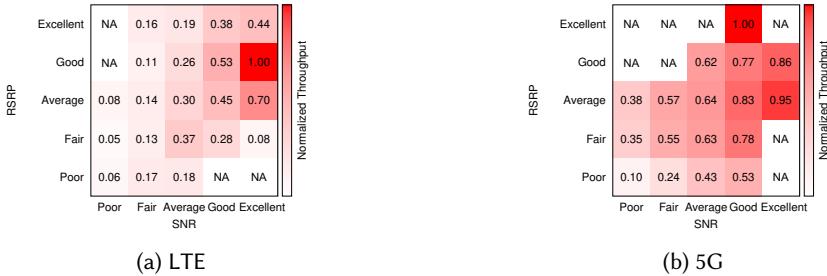


Fig. 10. The normalized throughput in TCP BBR with respect to RSRP and SNR.

Implications. Although the UE operates at poorer signal strength and quality in 5G, the sufficient physical resources effectively mitigate the throughput degradation. Nevertheless, given that LTE still dominates the cellular traffic [29], it is necessary for the carriers to refine the LTE resource allocation policy and promote throughput robustness to the highly fluctuating channel condition.

4.2 Handovers

Handover and its resulting latency (depending on its success or failure) play a vital role in delivering a seamless user experience in extreme mobility. We identify all the handover events from the massive link-layer logs and develop the taxonomy for LTE, 5G NSA, and 5G SA handovers by making a detailed breakdown of the handover process (§4.2.1). Then we describe the differences among these types in terms of their basic properties and the impact on the throughput (§4.2.2).

4.2.1 Handover Taxonomy

The development of handover taxonomy is non-trivial because not all the corner cases of handover failure are fully documented by 3GPP. Particularly, the NSA handovers are complex because they involve the interaction between LTE and 5G. We, to our best effort, comprehensively summarize the general patterns to cover more than 7K successful and failed handover events from our collected cellular baseband logs.

LTE and 5G SA Handovers. We first examine the handovers in LTE and 5G SA, whose processes are similar. A successful handover (Type 1) is composed of the successful reception of a handover command from the original node (*i.e.*, the original base station) via *RRC Reconfiguration* message

and successful synchronization with the target node via the random access procedure, or *RACH*. The handover will fail if *RACH* fails after a limited amount of trials. UE will attempt to reestablish the connection by making a *Reestablishment Request* to the target node followed by another *RACH* (Type 2). If the request is rejected (*i.e.*, UE context is lost) or *RACH* fails again, UE will try to make a new connection with *Connection Request* also followed by *RACH* (Type 3). Another reason to trigger handovers is Radio Link Failure (RLF), in which case UE loses radio connection with its serving cell and it triggers *Reestablishment Request* to a new node. They are denoted as failed handovers for similar processes and categorized into Type 2 and Type 3 according to whether the connection reestablishment succeeds. For sake of convenience, we will refer to handovers triggered by RLF as RLF handovers and otherwise as non-RLF handovers. We summarize three types of handovers for LTE and 5G SA in Tab. 2. Note that the difference between 5G SA and LTE lies in the handling of rejected *Reestablishment Request*. In LTE, another *Connection Request* and *RACH* are required to set up a new connection while in 5G the introduction of *Fast Connection* [1] allows the node to send *Connection Setup* directly (Fig. 14) if the node fails to validate UE context.

Type	Handover Result	Description
Type 1	Successful	Successful <i>Reconfiguration</i> and <i>RACH</i>
Type 2	Failed	RLF or failed <i>RACH</i> followed by successful <i>Reestablishment</i>
Type 3	Failed	Failed <i>Reestablishment</i> followed by successful <i>Connection Request</i>

Table 2. Handover taxonomy of LTE and 5G SA.

5G NSA Handovers. In 5G NSA, UE is connected to a Master Node (MN), an LTE node for control plane signaling, and a Secondary Node (SN), a 5G node for data plane transmission. Thus, a handover could be SN-only, MN-only, or SN-MN (Fig. 11). A handover is initiated by MN via sending UE a handover command and UE then triggers *RACH* to synchronize with the target nodes (SN, or MN and SN). A previous study [34] observed that for cross-MN handovers, the UE has to release the SN first and reconnect to it after the handover finishes. This redundancy is no longer seen in our dataset.

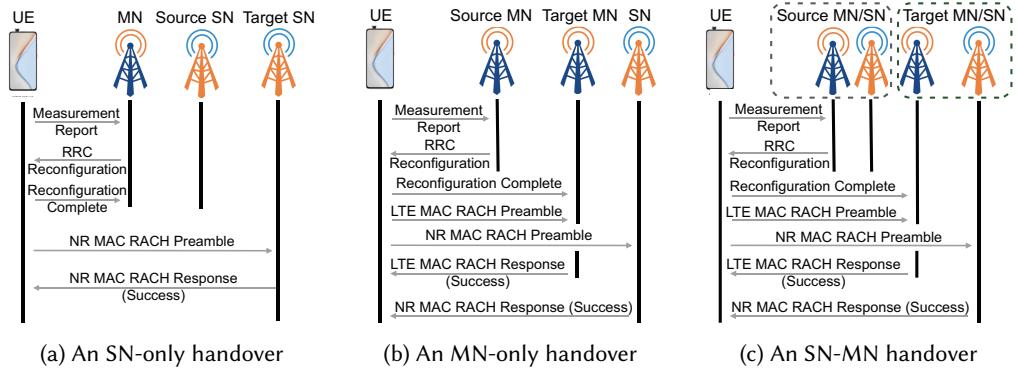


Fig. 11. A breakdown of successful NSA handovers.

In Tab. 3, we give a brief description of all six possible types of handovers in NSA. Similar to LTE and SA, successful *Reconfiguration* and *RACH* lead to successful handovers. However, failed *RACH* to either MN (Type 2) or SN (Type 3-6) causes a failed handover. If the *RACH* process to MN fails, the UE will follow the same process as in LTE. Two possible types of failed handovers in LTE are

Type	Handover Result	Description
Type 1	Successful	Successful <i>Reconfiguration</i> and <i>RACH</i>
Type 2	Failed	Failed MN <i>RACH</i> followed by <i>Reestablishment</i> or <i>Conn. Request</i>
Type 3	Failed	Failed SN <i>RACH</i> followed by successful SN <i>RACH</i>
Type 4	Failed	Failed SN <i>RACH</i> followed by <i>Dual Conn. Release</i>
Type 5	Failed	Failed SN <i>RACH</i> followed by MN HO and successful SN <i>RACH</i>
Type 6	Failed	Failed SN <i>RACH</i> followed by MN HO and <i>Dual Conn. Release</i>

Table 3. Handover taxonomy of NSA.

included in Type 2. In this case, UE will fall back to LTE-only mode after the handover finishes. If the *RACH* to SN fails, MN may (Type 5 and 6) or may not (Type 3 and 4) trigger a handover to a new MN. After that, UE may connect to SN by another *RACH* (Type 3 and 5) or be released from SN if *Dual Connection Release* message is received from MN (Type 4 and 6). Here we do not specifically study RLF handovers in NSA because UE will fall back to LTE-only mode after RLF and endure the same handover process as Type 2 and 3 in LTE.

4.2.2 Characteristics of Handovers

Next, we characterize the handovers from four aspects.

Handover	Type 1	Type 2	Type 3
LTE Non-RLF Handover	2380 (99.50%)	3 (0.12%)	9 (0.38%)
LTE RLF Handover	0	133 (51.75%)	124 (48.25%)
5G SA Non-RLF Handover	1924 (98.57%)	15 (0.77%)	13 (0.66%)
5G SA RLF Handover	0	22 (15.38%)	121 (84.62%)

Table 4. LTE and 5G SA handover frequency statistics.

Handover	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
5G NSA Handover (Non-RLF) ³	3325 (97.85%)	10 (0.29%)	18 (0.53%)	39 (1.15%)	4 (0.12%)	2 (0.06%)

Table 5. 5G NSA handover frequency statistics.

Frequency. We show the frequency of each type in Tab. 4 and Tab. 5. We make several observations. (i) If the UE receives the handover command via *RRC Reconfiguration*, the handover will likely succeed. The arrival of the command itself is an indication of good channel condition. In Tab. 4 and Tab. 5, we can learn that 99.50% of the LTE, 98.57% of the SA, and 97.85% of the NSA non-RLF handovers succeed. LTE has the highest success rate while NSA has the lowest because the signal strength is more stable in LTE (§4.1) and NSA handovers can fail due to failed *RACH* to either SN and MN. On the other hand, most handover failures come from RLF handovers. 95.5% and 83.6% handover failures in LTE and 5G SA are caused by the previous RLF rather than the failure of *RACH*, representing the disconnection caused by the RLF is more crucial to the failure of handover itself. (ii) If we consider RLF handovers, the distribution of Type 2 and 3 is very different in LTE and SA. Only about half of RLF handovers are Type 3 in LTE, while 84.6% RLF handovers are Type 2 in SA.

³5G NSA RLF Handovers are counted into LTE RLF Handovers because they have the same process.

It implies that RLF handovers are less likely to be foreseen by base stations in 5G and UE context is more likely to be lost during RLF, which calls for more intelligent context migration algorithms.

Interval. In Fig. 12a, we observe that NSA handovers happen per 3.064 sec (median number), more frequently than LTE (7.062 sec) and 5G (10.663 sec). The reason is that the RSRP of both the MN and SN may fluctuate and trigger the events that cause a handover. More interestingly, we find the CDF of 5G SA has two clear steps which are labeled in Fig. 12a as Step 1 and 2. By taking a closer look at the trace, we find the reason for Step 1 is the high variation of the signal strength which causes the UE handovers back and forth. This phenomenon is demonstrated in Fig. 12b. Here PCI 50, 90, and 190 are three different base stations. At about 15 sec, the RSRP of PCI 50 and 90 exhibits high fluctuation which causes three handovers in less than one second. We notice that the main difference of the CDF between 5G SA and LTE is around Step 2. The reasons can be attributed to the relatively low signal strength (Fig. 7a) and the better robustness to poor signal quality and strength (Fig. 10b) than LTE. In 5G SA, the neighbor cell is less likely to be sensed and the necessity of a handover is reduced because the serving cell could provide acceptable throughput.

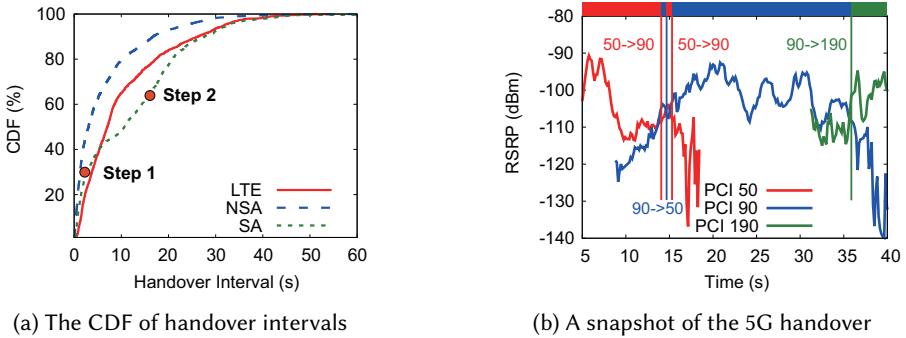


Fig. 12. The intervals of handovers in LTE, 5G NSA, and 5G SA.

Duration. In Fig. 13, we show the duration for each handover type. We make several findings from the CDF and investigate the traces for reasons. (i) LTE has the shortest median duration of a successful handover (Type 1) while 5G NSA and 5G SA have similar values. The medians of LTE, 5G NSA, and 5G SA are 29 ms, 53 ms, and 53 ms respectively. To find out the reason why the handover in 5G takes a longer time, we look into the control signals in detail and observe the prolonged processing time between receiving the *Reconfiguration* and triggering the *RACH* in 5G. We observe that the processing time is $2.8 \times$ longer in 5G than that in LTE (15 ms vs. 42 ms). (ii) The duration of Type 3 handovers in SA is about half shorter than that in LTE due to the introduction of *Fast Connection* procedure. In our dataset, 78.2% of Type 3 handovers in LTE have received *Reestablishment Reject* messages. It takes 185 ms (median) to receive the reject and another 67 ms for an upcoming *Connection Setup*. In 5G SA, 91.8% of Type 3 handovers benefit from the *Fast Connection*. It only takes 117 ms to receive the *Connection Setup*. (iii) The CDF of successful NSA handovers is step-like. The first step is close to LTE. The second step coincides with SA and the third step slightly deviates from it. After examining the traces, we conclude the first step is when the UE endures an MN-only handover and in the second step, the UE goes through an SN-MN or SN-only handover, where the processing time of 5G dominates. The third step is caused by multiple retries of *RACH* before it eventually succeeds. (iv) In NSA, it takes $3.63 \times$ longer time (1158 ms vs. 319 ms in median) to recover LTE failure (Type 2) than 5G failure (Type 3-6) because UE makes lots of *RACH* attempts before reestablishing or making a new connection to the MN in LTE failure.

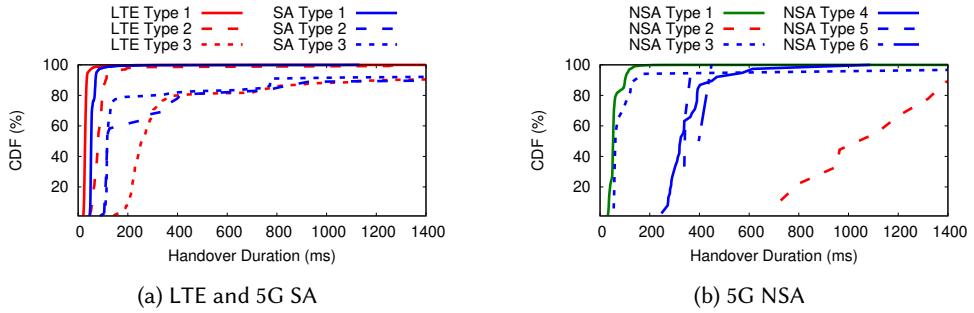


Fig. 13. The CDF of handover durations in LTE, 5G NSA, and 5G SA.

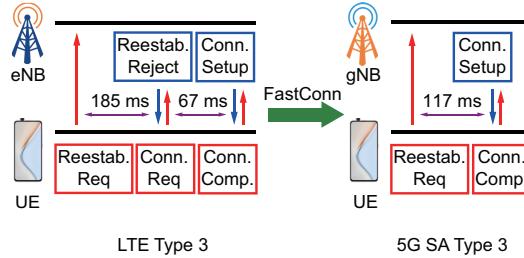


Fig. 14. Fast connection in 5G reduces Type 3 duration.

Impact on Throughput. To better characterize the impact on throughput that the handover brings before and after it happens, we define *near handover (HO) throughput*, which takes the average throughput over the 1-sec window before and after the handover event for the time point before and after the handover respectively. For instance, the near HO throughput of time point -1 and 2 mean the average throughput over the time window $(-2, -1]$ and $[2, 3)$, respectively. We combine NSA Type 3-6 because they are all caused by failed RACH to SN. In Fig. 15, we make three observations. (i) We notice that the average throughput in the first second after the successful handover drops by 5.0%, 12.8%, and 28.6% for LTE, 5G NSA, and 5G SA respectively compared to the last second before the handover. But before the handovers take place, both 5G NSA and SA suffer from less throughput drop (85.8% and 76.9%) than LTE (50.9%). We explain our observations as 5G is more robust to signal degradation (§4.1) and the signal quality is usually getting worse when a handover comes. As 5G SA has $19.4\times$ higher average throughput in the last second before handovers than LTE, the buffer in the base station may not be scaled as so. Together with the longer duration of handovers in 5G SA, it is not surprising that 5G suffers from more packet losses that lead to severe throughput drop after handovers. After a successful handover, it takes 1.0 sec for LTE, and 1.5 sec for SA to recover the throughput as before. (ii) As for failed handovers, the throughput before is low and fluctuated because mostly the UE is suffering from fading channel condition and RLF. After a failed handover, the throughput will increase but still hardly competes with a successful handover. 5G SA does better in recovery than LTE. (iii) The result of a failed handover in NSA is drastic, because in most cases, UE will have to go through several seconds of link disconnection or fall back to LTE-only mode before connecting to the 5G cell again.

4.2.3 Implications

Based on our breakdown, we could capture all types of handovers from the baseband logs and analyze the reasons for their characteristics. We find successful handovers of 5G SA have a larger throughput impact and longer duration than LTE although the throughput degradation is mitigated

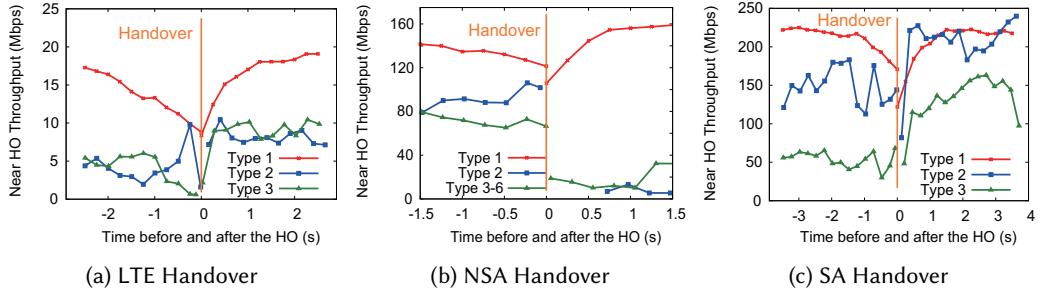


Fig. 15. The comparative impact of handovers on throughput in LTE, 5G NSA, and 5G SA.

by its robustness under poor channel conditions. Boosting the baseband's processing capability to reduce the duration from receiving the handover command to issuing the random access request and applying more intelligent context migration algorithms to avoid context loss especially when RLF occurs could be the next step for the evolution of 5G handovers.

4.3 RAN Latency

Now we closely examine the results of the *latency probe* experiment. As 5G exhibits shorter and more stable E2E RTT than LTE (Fig. 3), we are curious about how much of the latency reduction attributes to the improved RAN technology in 5G. We begin with the definition of the components in an E2E RTT (§4.3.1). Then we carefully examine the breakdown and provide an in-depth analysis (§4.3.2). Finally, we show how the key factors affect the latency (§4.3.3).

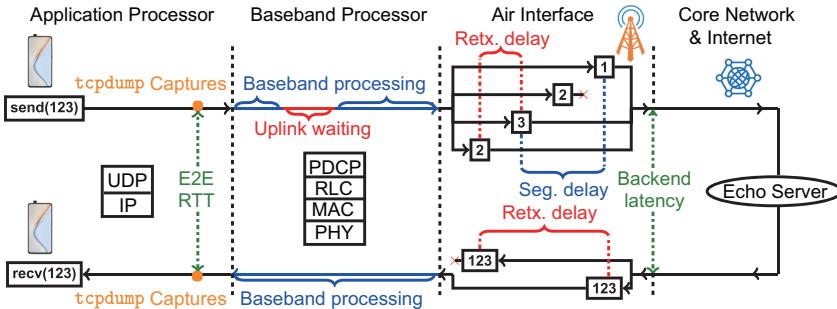


Fig. 16. An illustration of latency breakdown.

4.3.1 Breakdown Components

We send packets to an echo server and define the E2E RTT as the time spent between sending an uplink (UL) packet and receiving an echoed downlink (DL) packet at the transport layer. As shown in Fig. 16, each UL packet leaves the application processor (*i.e.*, CPU where Android is running), goes through the baseband processor, the LTE or 5G air interface, the base station (BS), the core network, and the Internet, and finally reaches the server. The server then echoes a DL packet of the same length. The DL packet experiences a reverse process as the UL packet. We further define the backend latency as the time spent in the base station, core network, and the Internet. Since the time which the packet spends in the air is close to zero and negligible, the backend latency can be calculated by subtracting the transmission timestamp in UE (*i.e.*, the reception timestamp in BS) from the reception timestamp in UE (*i.e.*, the transmission timestamp in BS) at the physical layer.

To give a representative example of E2E RTT and backend latency, we visualize two 30 sec trace segments from LTE and 5G respectively in Fig. 17. The first observation is that the backend latency

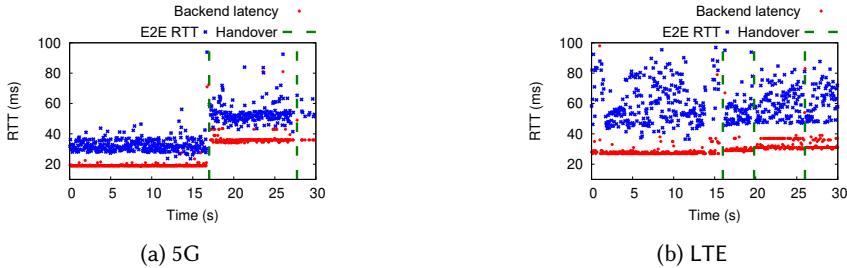


Fig. 17. Timeseries of backend latency and E2E RTTs in LTE and 5G.

is quite stable. They hardly have non-negligible fluctuations within a short period (typically ~ 10 sec). Meanwhile, the E2E RTTs are highly variable, which indicates that the difference between the E2E RTT and the backend latency is the main cause of RTT instability. The difference stands for the latency caused by the client-side RAN and we will refer to it as the RAN latency as we have no visibility of the infrastructure.

We further break down the RAN latency into four components. (i) The *segmentation delay* is defined as the time from when the first segment of the packet is transmitted to when the last segment of the packet is transmitted, where the segmentation occurs if the BS does not allocate enough resources. (ii) The *retransmission delay* is defined as the time from when all segments are transmitted to all segments are successfully (re)transmitted, where the retransmission occurs if the receiver fails to decode the packet. (iii) The *uplink waiting delay* is defined as the time from when the uplink packet is ready in the UE to when the BS allocates the uplink resources for the packet. (iv) The *baseband processing delay* is defined as the remained part of the RAN latency, which contains the time during which packets are being processed in the baseband processor, through PDCP, RLC, MAC, and PHY layers in the RAN protocol. The RAN latency and its components are all bidirectional in our breakdown.

To conduct decomposition of the RAN latency at the packet level in operational cellular networks, we have to associate events (*e.g.*, retransmission) in the baseband log with the corresponding packets. Unfortunately, the baseband log only tells a series of bytes in the RBs rather than the separate packets. The bytes (*i*) may contain one or more radio link layer headers of varying lengths, which should be identified and removed; (*ii*) may come from multiple network layer packets, which must be split or merged according to the context. To overcome these challenges, we carefully design our *latency probe* experiment so that the length of the packets forms a special sequence to ease the identification from the baseband log and avoid random queuing delay in the infrastructure. However, the baseband processing delay and uplink waiting delay are still mixed because we are not able to know when the baseband starts to wait for uplink resources. We combine them as the protocol delay and we will give reasonable estimations for them later.

4.3.2 RAN Latency Breakdown Analysis

5G RAN Latency Breakdown. In Fig. 18 and Tab. 6, we show the breakdown results for 5G. On average, the RAN latency accounts for 40.6% of the E2E RTT (15.80 ms out of 38.91 ms), which is a considerable percentage.

- Segmentation delay alone occupies 37.6% (5.94 ms out of 15.80 ms) among the components of RAN latency. Note that together with the first fragment of a packet, a Buffer Status Report (BSR) that contains the total number of bytes UE needs to transmit will be sent to the BS. However, we find in the trace that the BS responds to the report lazily. The BS rarely allocates enough resources in the

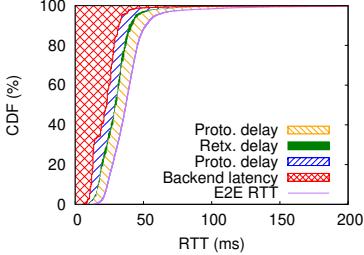


Fig. 18. CDFs of latency breakdown in 5G.

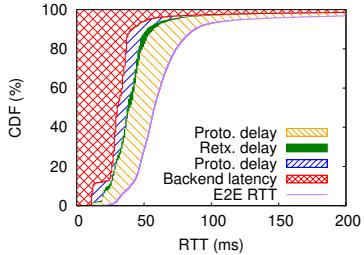


Fig. 19. CDFs of latency breakdown in LTE.

next few slots⁴. Therefore the segmentation delay can sometimes reach even 10 ms. This suggests the necessity and possibility of improving the UL resource allocation algorithm.

- *Retransmission delay* is less important. It occupies only 8.2% of the RAN latency (1.29 ms out of 15.80 ms) on average. The reason is two-fold. (i) Only some packets are retransmitted while most packets are segmented. (ii) Some segmentation delay hides the retransmission delay. For example, if the first fragment of a packet is successfully retransmitted before the transmission of the last fragment, we consider the retransmission delay as zero to avoid double-counting the common time of segmentation delay and retransmission delay.
- *Protocol delay* containing the baseband processing delay and the uplink waiting delay also occupies a large share, *i.e.*, 54.2% on average (8.56 ms out of 15.80 ms). We observe that the BS always schedules uplink opportunities periodically, in our cases with a frequency of at least once per ten slots (5 ms). So on average, we estimate that the uplink waiting delay will not exceed 5 ms and the baseband processing delay will be greater than 3.5 ms.

LTE RAN Latency Breakdown. The breakdown results for LTE are shown in Fig. 19 and Tab. 7. LTE performs much worse than 5G. 5G reduces the RAN latency from 40.03 ms to 15.80 ms (39.5%). Among the components of RAN latency, the protocol delay is reduced mostly, from 30.51 ms to 8.56 ms. We attribute it to the optimized RAN stack and reduced uplink waiting time. In LTE, the BS usually does not reserve UL resources for the UE. Instead, UE has to send a Schedule Request (SR) first and wait for the allocation from the BS, which lengthens the uplink waiting delay.

The improvement of 5G RAN latency is significant but has not yet met the 3GPP's requirement of URLLC. We have seen the RAN latency still occupies a considerable share of the E2E RTT and causes almost all the fluctuations. The optimization possibilities still reside in both UE and BS sides.

4.3.3 Cross-layer Latency Analysis

Now we study how the E2E RTT is influenced by three key factors, *i.e.*, RSRP, SNR, and handovers.

	Average	Std. Dev	Median
E2E RTT	38.91 ms	39.68 ms	36.70 ms
Backend latency	23.11 ms	27.74 ms	22.50 ms
Seg. delay	5.94 ms	5.47 ms	5.00 ms
Retx. delay	1.29 ms	2.69 ms	0.00 ms
Proto. delay	8.56 ms	25.98 ms	6.45 ms
RAN latency	15.80 ms	27.49 ms	13.47 ms

Table 6. Statistics of latency breakdown in 5G.

	Average	Std. Dev	Median
E2E RTT	77.41 ms	124.49 ms	57.99 ms
Backend latency	37.38 ms	61.82 ms	31.00 ms
Seg. delay	7.35 ms	11.02 ms	8.00 ms
Retx. delay	2.17 ms	4.73 ms	0.00 ms
Proto. delay	30.51 ms	100.62 ms	16.36 ms
RAN latency	40.03 ms	101.66 ms	25.58 ms

Table 7. Statistics of latency breakdown in LTE.

⁴A unit of time for resource allocation in LTE and 5G, which is 0.5 ms in our experiment settings.

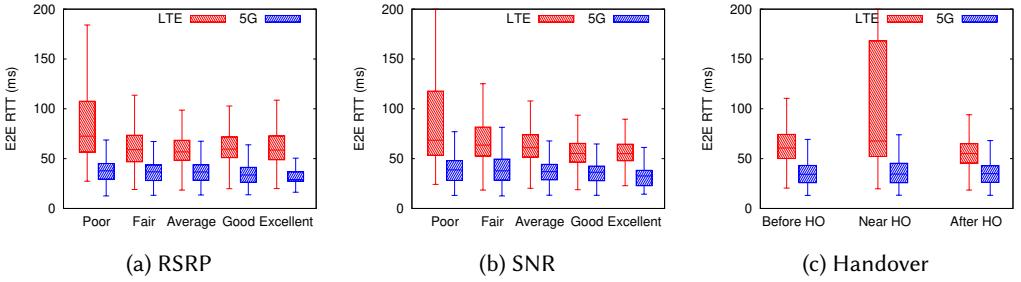


Fig. 20. The effects of RSRP, SNR, and handovers on the E2E RTT.

RSRP and SNR. We demonstrate how RSRP and SNR affect RTTs in Fig. 20a and Fig. 20b. Five categories (*Poor*, *Fair*, *Average*, *Good*, *Excellent*) are defined in §4.1. We notice that better signal conditions shorten the RTT, as the retransmission delay and (de)modulation time could be reduced. For instance, *Excellent* SNR reduces the average LTE RTT by 60% (from 148 ms to 59 ms) compared to *Poor* SNR. Besides, we find that 5G suffers less from poor signal conditions. Considering the same comparison in 5G, the average RTT is only reduced by 36% (from 50 ms to 32 ms). Together with 5G’s higher throughput in poor signal conditions (§4.1), we can conclude that 5G is more robust to bad channel quality.

Handovers. Then we examine how the handovers affect the RTT. We denote a 1-sec window when a handover occurs in the middle as *near HO*, and its two adjacent one-second time windows are denoted as *before HO* and *after HO* respectively. The results are shown in Fig. 20c. We also analyze other time windows but omit them as they behave similarly to *before HO* and *after HO*. We can see that RTTs are strongly affected by handovers in LTE but stay quiescent in 5G. The average RTT *near HO* is 4.3× and 1.2× higher than that *before HO* in LTE and 5G respectively (LTE: 87 ms to 374 ms; 5G: 37 ms to 44 ms). We attribute this to 5G’s better robustness under poor channel quality since the signal condition gets worse when the UE moves to the edge of the BS’ coverage area.

4.3.4 Implications

Compared to LTE, 5G offers notably lower and stable RAN latency, even under faded signal conditions. However, the current 5G has not fed the appetite of latency-sensitive applications such as auto-driving (requiring sub-10ms delay [5]), let alone the restrictions imposed by URLLC (sub-1ms delay) [2]. The segmentation and uplink waiting delay become the main parts of RAN latency. To further reduce latency, the protocol should be renovated to facilitate the client’s on-demand uplink report and BS’ prompt reaction.

5 APPLICATION-LAYER PERFORMANCE

We finally turn our attention to the application layer and investigate the QoE of two important representative applications – web surfing (§5.1) and video streaming (§5.2).

5.1 Web Surfing Performance

We select 10 websites from the Alexa top 500 websites for our experiment. We leverage Selenium to control the mobile Chrome browser in Android which allows us to simulate the user behavior in browsing. We inject JS snippets to the browser to scroll down the page for the lazy-loaded web contents. This process lasts 10 seconds and we repeat it for the next site. The ChromeDriver’s internal logging option produces the performance logs for further analysis.

In Fig. 21, we show the average page load time (PLT) for each of the carriers. Different from previous studies [23] in which 4G and 5G have similar PLT for small web pages, we find that the introduction of 5G on the HSR decreases PLT for pages of all sizes. For example, the average PLT

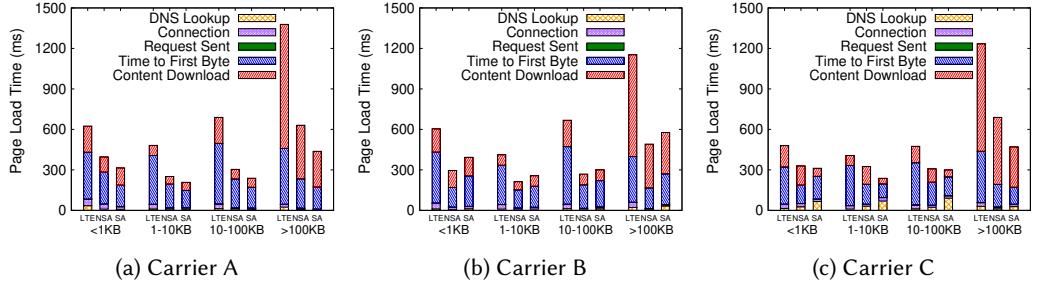


Fig. 21. The average page load time for three carriers.

across all three carriers is reduced to 339.74 ms (pages < 1 KB) and 494.90 ms (pages > 100 KB) in 5G SA compared to 569.56 ms and 1255.12 ms in LTE. As 1 second is about the limit for the user's flow of thought to stay uninterrupted [25], 5G delivers a more seamless experience in web browsing to the users. Both *TTFB* and *Content Download* as well as *Connection* (*i.e.*, time spent in TCP and TLS handshake) decreases in 5G due to the reduced E2E RTT and larger throughput. We also observe that the evolution from NSA and SA shortens the total time for Carrier A and C but extends for Carrier B. In a few cases, it also increases the DNS lookup time for Carrier B and C, which suggests immaturity of SA deployment in the core network.

5.2 Video Streaming Performance

Video streaming dominates the Internet traffic in recent years [30]. Assisted by ABR that can effectively leverage the available bandwidth, 5G promises to deliver better QoE in terms of higher bitrate and lower stall time. We use `DASH.js` [9] as our video client and deploy Nginx [16] to host our test videos in our cloud server. We use a custom 4K video and encode it with libx264 using FFmpeg [11] into eight *tracks* (or qualities) with different bitrates. The chunk length is set to 1 sec, as a shorter length will increase the startup overhead of chunk downloads and a longer length will reduce the flexibility of ABRs. The bitrates we choose are based on the standard series provided by YouTube [36]. We add 110 and 160 Mbps to match the average throughput of 5G. We use the same tracks for 5G and LTE since the ABR is supposed to determine the best *track* for the client. We use Selenium to play our video in mobile Chrome browsers with different ABR algorithms in turn.

We study five production-ready ABR algorithms, namely Dyn [27], RB [3], Bola [28], L2A [17], and LoLP [7] provided by `DASH.js`, and another experimental ABR called MPC [35]. RB simply uses the average historical throughput for the next chunk. Bola picks a higher bitrate chunk when the buffer level is high. Dyn uses RB when the buffer is near full and switches to Bola when the buffer level drops. L2A and LoLP are based on online machine learning strategies to reduce latency. MPC is built on control theory and selects the best bitrate chunk to maximize its internal QoE function.

For each video playback, we extract two key parameters, the average bitrate and the stall percent (rebuffer time over total playback time), and visualize the results in Fig. 22. We observe that 5G SA demonstrates the best performance in providing high bitrates and reducing stalls. For instance, the average bitrate (and stall percentage) of RB in 5G SA is 1.2× and 6.6× higher (29.2% and 42.6% lower) than that in 5G NSA and LTE respectively. However, none of the ABR algorithms perform well on both the stall percent and the average bitrate. In 5G SA, the most aggressive ABR (RB) comes with the highest average bitrate but also a high stall percentage (5.9%). Bola is least aggressive toward throughput and its stall percentage is lower (3.7%). L2A achieves the lowest but still non-negligible 2.2% stall percentage among all the ABRs in our experiment and it also suffers from low bandwidth utilization.

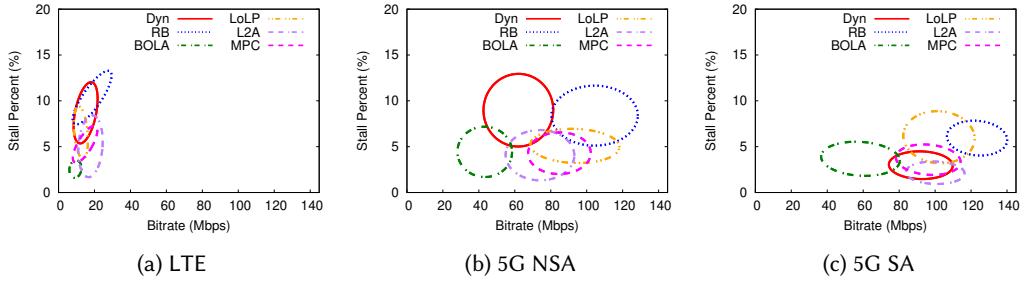


Fig. 22. Bitrate and stall percentage comparison of different ABRs in LTE, 5G NSA, and 5G SA.

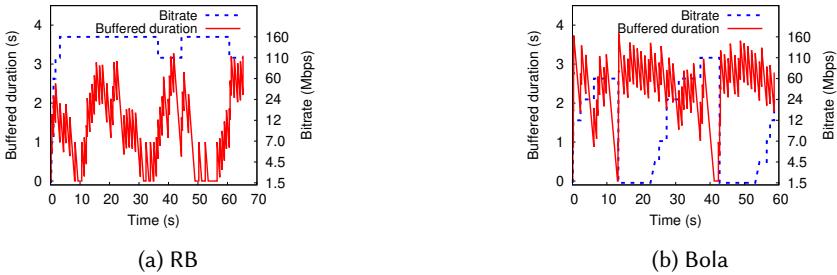


Fig. 23. Snapshot timeseries of two video streaming examples with different ABRs in 5G SA.

In Fig. 23, we demonstrate the trade-offs between the average bitrate and stall percentage using two examples of playback with RB and Bola. We can observe that extreme mobility brings high fluctuations of throughput and a sudden bandwidth drop may drain the playback buffer in a short period. RB makes decisions on the average historical throughput, so it waits for the drop of average throughput before the lower bitrate is selected. Bola, instead, will select the lowest bitrate once no video chunks are available in the buffer and it can take a much longer time to recover the highest bitrate. It is consistent with our observations in Fig. 22.

5.3 Implications

Though 5G SA generally provides a shorter page loading time, it remains its immaturity for the prolonged DNS lookup latency and TTFB for some carriers. Meanwhile, its advantages in bandwidth and latency are not fully expressed in off-the-shelf ABR algorithms given the highly variant connectivity in the extreme mobility, calling for specific design accordingly.

6 DISCUSSION

Comparing with Stationary Scenarios. To further demonstrate which characteristics are unique in extreme mobility, we reuse the same platform and conduct stationary experiments on campus under similar signal quality and strength to those on HSR (§4.1). We showcase the loss rate and RAN latency as two examples. (i) In stationary scenarios, we find that more than two-third of stationary (LTE: 85.7%; 5G SA: 71.4%) traces have negligible ($< 0.005\%$) loss rates and 5G is slightly more lossy than LTE. The performance gap between stationary and extreme-mobility scenarios suggests optimization opportunities. (ii) We notice extreme mobility leads to an 87.8% increase on average RAN latency in LTE, but only 56.4% in 5G. Meanwhile, RAN latency is still non-negligible (10.1 ms and 21.3 ms on average for LTE and 5G respectively) in stationary scenarios, where the protocol delay and segmentation delay are the two main components in RAN latency as well.

QUIC Variants. A wide range of QUIC implementations has been rolled out in the process of standardizing QUIC. Even with its recent standardization, heterogeneous protocol stacks will still exist which requires in-depth analysis. We pick LSQUIC because it is mature to be used in production and it is actively maintained as the standardization process moves forward. We consider carrying out more comprehensive studies covering all major variants for more rigorous conclusions in the future.

Collaborating with the Carriers. Though latency caused by the base station, the core network, and the Internet is fairly stable, its small fluctuations are still worth further studying as they may be related to handovers, base station-side RAN latency, and Internet jitters. Meanwhile, we have pinpointed several deficiencies related to untimeliness and insufficiency of physical resources allocated by base stations. Currently, we have no visibility of the carriers' infrastructure, especially on how the physical resources are allocated by the base station. We call for collaboration with carriers to carry out more comprehensive analyses and propose precise optimization strategies for performance enhancement.

7 RELATED WORKS

5G Measurement. 5G measurement has been widely studied since its roll-out. Xu *et al.* [34] carried out first measurement of mid-band 5G NSA in campus. It revealed that 5G had much higher loss rates than LTE. In our experiment, this gap in loss has been mitigated by the infrastructure upgrade. Narayanan *et al.* [23] focused on both mmWave and mid-band in the diverse urban environment and mobility up to driving. In their scenario, a fallback to LTE and LTE-LTE handover is necessary for 5G-5G handovers, which has not been observed in our dataset. Narayanan *et al.* [24] measured the performance of 5G SA architecture with respect to the server distance. It coarsely estimated RAN latency using RTT differences. We instead calculate and even break down RAN latency precisely with detailed baseband logs and carefully designed experiments. Li *et al.* [19] located the factors which impacted the 5G cellular reliability and identified that software unreliability mainly caused cellular data connection failures. In our study, we emphasize performance, especially throughput and latency. We cover the operational LTE, 5G NSA, and 5G SA services of all carriers on the HSR and perform the analysis in a cross-layer manner.

Network Performance on HSR. Li *et al.* [18] measured MPTCP and revealed inconsistent handover behaviors between the two mobile carriers. We instead focus on the common grounds among carriers, which pinpoints general deficiencies of 5G. Wang *et al.* [31] quantified the impact of LTE handovers on TCP CUBIC and BBR performances. We notice that 5G successful handovers are not as lossless as LTE's because the increase of buffer size does not match with the huge bandwidth gain. Xu *et al.* [33] developed the first taxonomy of LTE disconnections and correlate them with TCP stalls. Despite TCP stalls, it did not consider the long recovery time of throughput from which 5G handovers suffer in our experiments. We extend the above findings in two aspects. (i) We quantify the performance of both LTE and 5G by inspecting throughput and latency as well as their robustness to signal variance. (ii) We conduct the first taxonomy of 5G SA and 5G NSA handovers with detailed analyses of their properties and impact on performance.

8 CONCLUSION

In this paper, we perform the first extensive measurement campaign covering LTE, 5G NSA, and 5G SA on the HSR. We examine the key performance metrics of TCP and QUIC and compare their internal implementations. Also, we quantify the difference between 5G and LTE in signal quality, signal strength, physical resources, RAN latency, and their impact on upper-layer throughput and RTT. We develop a detailed taxonomy of LTE, 5G NSA, and 5G SA handovers and characterize them comprehensively. We set the performance baseline of page load time and different ABRs in video streaming under extreme mobility. Our study reveals the important features of the current 5G system which we believe are crucial for the next stage of evolution.

ACKNOWLEDGMENTS

We are grateful to all the reviewers and our shepherd, Ajay Mahimkar in particular, for their constructive critique and comments, which have greatly helped us improve this paper. This work is supported by National Key Research and Development Plan, China (Grant No. 2020YFB1710900).

REFERENCES

- [1] 2020. 5G; NR; Radio Resource Control (RRC); Protocol specification. https://www.etsi.org/deliver/etsi_ts/138300_13_8399/138331/16.01.00_60/ts_138331v160100p.pdf. 3GPP TS 38.331 version 16.1.0 Release 16 (2020-07).
- [2] 2020. 5G; Study on scenarios and requirements for next generation access technologies. https://www.etsi.org/deliver/etsi_tr/138900_138999/138913/16.00.00_60/tr_138913v160000p.pdf. 3GPP TR 38.913 version 16.0.0 Release 16 (2020-05).
- [3] 2021. *ABR Logic · Dash-Industry-Forum/dash.js Wiki*. <https://github.com/Dash-Industry-Forum/dash.js/wiki/ABR-Logic>
- [4] 2021. *Linux Deploy*. <https://github.com/mefik/linuxdeploy>
- [5] 5GAA. 2019. *C-v2x use cases methodology, examples and service level requirements*. https://5gaa.org/wp-content/uploads/2019/07/5GAA_191906_WP_CV2X_UCs_v1-3-1.pdf
- [6] The California High-Speed Rail Authority. 2021. *2021 sustainability report*. https://hsr.ca.gov/wp-content/uploads/2021/09/Sustainability_Report_2021.pdf
- [7] Abdelhalim Bentaleb, Mehmet N Akcay, May Lim, Ali C Begen, and Roger Zimmermann. 2021. Catching the Moment with LoL+ in Twitch-Like Low-Latency Live Streaming Platforms. *IEEE Transactions on Multimedia* (2021).
- [8] China.org. 2021. *Over 100 5G base stations installed for 2022 Winter Olympics*. http://www.china.org.cn/business/2021-06/18/content_77575629.htm
- [9] Dash-Industry-Forum. 2022. *DASH.js*. <https://github.com/Dash-Industry-Forum/dash.js>?
- [10] Korian Edeline, Mirja Kühlewind, Brian Trammell, and Benoit Donnet. 2017. copycat: Testing differential treatment of new transport protocols in the wild. In *ACM/IRTF ANRW*.
- [11] ffmpeg team. 2022. *FFmpeg project*. <https://ffmpeg.org>
- [12] Google. 2016. *BBR Congestion Control Algorithm*. <https://github.com/google/bbr>
- [13] Google. 2021. *QUICHE, quic_protocol_flags_list.h*. https://quiche.googlesource.com/quiche/+/2403dac9448364d083c36bf9f0045d4accfae3de/quic/core/quic_protocol_flags_list.h#240
- [14] Google. 2022. *ChromeDriver*. <https://chromium.googlesource.com/chromium/src/+/refs/heads/main/chrome/test/chromedriver>
- [15] Z. Hrebicek, Y. Crozet, C. Cheze, L. Guihery, M. Reichenbach, C. Desmaris, K. Anderton, and L. Krejci. 2014. *RANSFORUM Roadmap High-speed Rail. Cologne / Koln: Rupprecht Consult*. http://www.rupprecht-consult.eu/uploads/tx_rupprecht/TRANSFORUM_D7-7_Roadmap_HSR.pdf
- [16] F5 Inc. 2022. *Nginx*. <https://www.nginx.com/>
- [17] Theo Karagioules, Rafael Mekuria, Dirk Griffioen, and Arjen Wagenaar. 2020. Online learning for low-latency adaptive streaming. In *ACM MMSys*.
- [18] Li Li, Ke Xu, Tong Li, Kai Zheng, Chunyi Peng, Dan Wang, Xiangxiang Wang, Meng Shen, and Rashid Mijumbi. 2018. A measurement study on multi-path TCP with multiple cellular carriers on high speed rails. In *ACM SIGCOMM*.
- [19] Yang Li, Hao Lin, Zhenhua Li, Yunhao Liu, Feng Qian, Liangyi Gong, Xianlong Xin, and Tianyin Xu. 2021. A nationwide study on cellular reliability: measurement, analysis, and enhancements. In *ACM SIGCOMM*.
- [20] Yuanjie Li, Chunyi Peng, Zengwen Yuan, Jiayao Li, Haotian Deng, and Tao Wang. 2016. Mobileinsight: Extracting and analyzing cellular network information on smartphones. In *ACM MobiCom*.
- [21] LSQUIC. 2021. *lsquic_bbr.c*. https://github.com/litespeedtech/lsquic/blob/082507cd1032c90ccfc0a9cad845c5fdeae53f86/src/liblsquic/lsquic_bbr.c#L73-L74
- [22] China Mobile. 2021. *Small classroom for 5G common questions*. <https://www.10086.cn/5G/qna/?headrndnum=0.09984317695903178>
- [23] 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 *ACM WWW*.
- [24] Arvind Narayanan, Xumiao Zhang, Ruiyang Zhu, Ahmad Hassan, Shuwei Jin, Xiao Zhu, Xiaoxuan Zhang, Denis Rybkin, Zhengxuan Yang, Zhuoqing Morley Mao, et al. 2021. A variegated look at 5G in the wild: performance, power, and QoE implications. In *ACM SIGCOMM*.
- [25] Jakob Nielsen. 1994. *Usability engineering*. Morgan Kaufmann.
- [26] The State Council of the People's Republic of China. 2016. *Medium- and Long-Term Railway Network Plan*. <https://www.gov.cn/xinwen/2016-07/20/5093165/files/1ebe946db2aa47248b799a1deed88144.pdf>

- [27] Kevin Spiteri, Ramesh Sitaraman, and Daniel Sparacio. 2019. From theory to practice: Improving bitrate adaptation in the DASH reference player. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 15, 2s (2019).
- [28] Kevin Spiteri, Rahul Urgaonkar, and Ramesh K Sitaraman. 2020. BOLA: Near-optimal bitrate adaptation for online videos. *IEEE/ACM Transactions on Networking* 28, 4 (2020).
- [29] Statista. 2021. *Share of telecommunication 4G mobile users among all mobile internet users in China*. <https://www.statista.com/statistics/1109550/china-number-of-4g-mobile-subscirbers-share/>
- [30] VentureBeat. 2021. *Comcast: Pandemic drove peak internet traffic up 32% in 2020*. <https://venturebeat.com/2021/03/02/comcast-peak-internet-traffic-rose-32-in-pandemic-in-2020/>
- [31] 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 *ACM MobiCom*.
- [32] Wikipedia. 2021. *IMT-2020*. <https://en.wikipedia.org/wiki/IMT-2020>
- [33] Chenren Xu, Jing Wang, Zhiyao Ma, Yihua Cheng, Yunzhe Ni, Wangyang Li, Feng Qian, and Yuanjie Li. 2020. A first look at disconnection-centric TCP performance on high-speed railways. *IEEE Journal on Selected Areas in Communications* 38, 12 (2020).
- [34] 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 *ACM SIGCOMM*.
- [35] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A control-theoretic approach for dynamic adaptive video streaming over HTTP. In *ACM SIGCOMM*.
- [36] Youtube. 2022. *Recommended upload encoding settings*. <https://support.google.com/youtube/answer/1722171#zippy=%2Cbitrate>

Received October 2021; revised December 2021; accepted January 2022