Teleoperating Autonomous Vehicles over Commercial 5G Networks: Are We There Yet?

Rostand A. K. Fezeu[§*], Jason Carpenter[§*], Rushikesh Zende[§], Sree Ganesh Lalitaditya Divakarla[§], Nitin Varyani[†], Faaiq Bilal[§], Steven Sleder[§], Nanditha Naik[§], Duncan Joly[§], Eman Ramadan[§], Ajay Kumar Gurumadaiah[§], Zhi-Li Zhang[§]
[§]University of Minnesota - Twin Cities, USA [†]Kalinga Institute of Industrial Technology, India

ABSTRACT

Remote driving, or *teleoperating* Autonomous Vehicles (AVs), is a key application that emerging 5G networks aim to support. In this paper, we conduct a systematic feasibility study of AV teleoperation over commercial 5G networks from both *cross-layer* and *end-to-end (E2E)* perspectives. Given the critical importance of *timely delivery of sensor data*, such as camera and LiDAR data, for AV teleoperation, we focus in particular on the performance of uplink sensor data delivery. We analyze the impacts of Physical Layer (PHY layer) 5G radio network factors, including channel conditions, radio resource allocation, and Handovers (HOs), on E2E latency performance. We also examine the impacts of 5G networks on the performance of upper-layer protocols and E2E application Quality-of-Experience (QoE) adaptation mechanisms used for real-time sensor data delivery, such as Real-Time Streaming Protocol (RTSP) and Web Real Time Communication (WebRTC). Our study reveals the challenges posed by today's 5G networks and the limitations of existing sensor data streaming mechanisms. The insights gained will help inform the co-design of future-generation wireless networks, edge cloud systems, and applications to overcome the low-latency barriers in AV teleoperation.

KEYWORDS

5G, PHY/MAC Layers, Autonomous Vehicles, Teleoperation, AI/ML, Measurement, Latency, QoE, Performance

1 INTRODUCTION

Since the DARPA Grand Challenge in 2005 [18], tremendous progress has been made in the development of AVs. Pilot "robotaxi" services are now available in several major cities in the United States (U.S.) [37, 38, 54]. Today's AVs can, at best, be rated as Society of Automotive Engineers (SAE) Level-4 [4, 46, 62, 63]: Namely, such AV is designed with a specific set of conditions, referred to as its *Operational Design Domain* (ODD), outside which it must come to a safe stop. Despite significant advances in AI/ML, fully autonomous driving (Level-5) still has a long way to go. Safety concerns and other issues plaguing robotaxi trials [17, 20, 41, 61] highlight the challenges posed by complex real-world environments.

To partly circumvent these challenges, remote driving – or *teleoperated driving* (ToD) in the 3rd Generation Partnership Project (3GPP) parlance – has been proposed as an alternative or complementary approach [11, 24, 74], where a remote human operator takes over the control of AV when needed. For example, before the AV is about to encounter a situation outside its ODD and has to stop [48]. The potential of ToD is inspired by the promise of 5G, and is considered one of its key use cases by 3GPP and 5G Automotive Association (5GAA) [11]. ToD has been tested in (mostly ideal and) restricted environments [5, 25, 35, 53, 68]; several start-up companies are promoting remote driving for certain use cases [6, 7, 23, 45]. Therefore we ask, *Can today's commercial 5G meet the requirements of ToD in real-world environments?* This question is the main goal of this paper.

To this end, we contribute to the understanding of ToD and carry out a systematic feasibility study of AV teleoperations over commercial 5G networks in real-world urban environments in the city of Minneapolis in the U.S. We note that AV ToD requires *timely delivery* of i) Uplink (UL) onboard sensor data like camera/video and LiDAR feeds from a host AV to a remote vehicle control station operated by a human to provide *(real-time) situation awareness*; and ii) Downlink (DL) Command and Control (C&C) data from the remote human operator to the host AV for vehicle control remotely. Per [11] (see §2.1), the E2E UL and DL latency within 100 ms and 20 ms respectively are considered ideal. In terms of data rates, clearly UL sensor data – video and especially LiDAR – require significantly high bandwidth, whereas DL delivery of C&C data requires little bandwidth (§2.1). Unfortunately, today's commercial 5G networks are designed, configured, and optimized primarily for mobile Internet access. This creates *DL/UL asymmetry in bandwidth and latency* (see §2.2). Furthermore, *high mobility* of AVs induces highly fluctuating radio channel conditions and frequent HOs, both of which can increase Block-Level Error Rates (BLERs) at the Radio Access Network (RAN) and require retransmissions, thereby incurring higher latency.

*These authors contributed equally to this paper.
Corresponding authors: fezeu001@umn.edu, carpe415@umn.edu.

We take a *cross-layer* approach to investigate the fundamental impacts of 5G networks on the AV teleoperation. Given the stringent performance requirements of teleoperation, our aim is to characterize the E2E performance for sensor data – video and LiDAR – streamed over a commercial 5G network to an edge-cloud server in a remote teleoperation station in the same geographical region as the AV. We introduce QoE metrics at the *per-frame* level – as frames are the basic units for video display and camera/LiDAR data processing (§4.2). We investigate the effectiveness of video compression (§5.1) and bitrate adaptation (§7) in reducing the one-way delay when streaming video (and LiDAR) data, and quantify timing of critical C&C including steering, acceleration, and braking over commercial 5G networks (§5.4). We extract 5G RAN parameters to understand the impact of 5G PHY layer factors – such as channel conditions, BLERs, HOs, and Resource Blocks (RBs) allocation – on per-frame E2E latency, with a focus on *tail latency* (§6). Additionally, we evaluate how 5G affects upper-layer protocols and E2E application layer QoE adaptation mechanisms used in real-time sensor data delivery protocols like RTSP [15] and WebRTC[26]. From our insights, we further explore challenges in teleoperating multiple AVs over 5G and examine potential benefits of using multiple 5G operators to mitigate tail latency degradation and optimize network resource utilization.

The key findings of our paper are summarized as follows:

• We define *per-frame* level QoE metrics to characterize sensor data delivery E2E latency performance and visual quality. In particular, we introduce *perceptual quality deviation* to capture the latency-visual-quality interplay (§4.2). Using RTSP (commonly used in many existing AV teleoperation platforms [39, 65]) and WebRTC (a common low latency video conferencing streaming system, and used by industry [45]) as the baselines §5.1, we examine the effectiveness of video data compression and LiDAR data voxel-based downsampling techniques in attaining E2E low-latency performance and quantify the impact of latency on *perceptual quality* performance. We find that while it is feasible to stream a single camera (as is the case in several field tests [25, 53]) to teleoperate an AV in most scenarios, the poor *tail latency performance* still raises safety concerns, as a safety-critical event can occur during such periods of poor performance.

• Situational awareness for teleoperation requires streaming multiple camera feeds and LiDAR data. We find that streaming these sensors simultaneously would likely strain today's 5G networks, even with the 5G Standalone (5G-SA) architecture. Without aggressive compression, streaming LiDAR over today's 5G networks is nearly impossible (§5.3). We also find that all existing data streaming mechanisms suffer from the *cumulative latency effect* (§5.1), which causes later frames to be more likely to suffer from deadline violations.

• For DL C&C, the latency experienced is roughly half of the UL latency and does not face throughput bottlenecks due to the smaller sizes of the commands. This echoes existing work and strengthens the feasibility of sending control commands with reasonable time and reliability (§5.4).

• 5G PHY layer dynamics significantly affect sensor data delivery for teleoperation. For example, on the one hand, when the channel conditions, as characterized by Channel Quality Indicator (CQI), go from "good" to "poor", the E2E frame delay increases by about 48%. In contrast, when fewer retransmissions occur on the PHY layer – quantified by a shift in BLERs from +10% down to 0-5%, the frame delay drops by approximately 35.8%. On the other hand, HOs pose an even greater challenge; we find that unnecessary ping-pong HOs can occur within a short time window (15 seconds in our analysis) while driving in a loop and making turns. As a result, the sensor data frame delay increases by up to 85% (§6).

• Our results show that WebRTC, Google's live video streaming system, succeeds in delivering low latency video frames (compared to RTSP), but fails to quickly respond to changing 5G network conditions resulting in overall reduced and struggling performance despite the PHY layer awareness and speediness of recovery. RTSP lacks any of these mechanisms and therefore struggles significantly in terms of per-frame delay impact. This manifests as a spike in per-frame latency, which could have been avoided with the knowledge of the lower-layer information. As we demonstrate in §7, the lower layer was aware of a problem far sooner than the application-level congestion and QoE triggers.

• Based on the above findings, we briefly explore the potential benefits of additional end-system mechanisms, such as selective frame dropping/frame rate adaptation, as well as leveraging multiple 5G operators to improve tail latency performance. We also consider the additional challenges posed by multiple AVs competing for radio resources (§8).

Contributions. We contribute to advancing the understanding of teleoperated driving over commercial 5G networks and conduct – to the best of our knowledge – the first feasibility study of AV teleoperation over *operational* live commercial 5G networks in a *real-world* urban environment from *cross-layer* and *E2E* perspectives, elucidating in particular the impacts of 5G networks on UL vehicle sensor data delivery. Our study reveals the challenges posed by 5G networks as well as the limitations of existing sensor data streaming mechanisms. Through exploration of additional end-system mechanism designs, we show that while these mechanisms can improve the tail latency performance, they cannot fundamentally address the challenges posed by 5G networks.

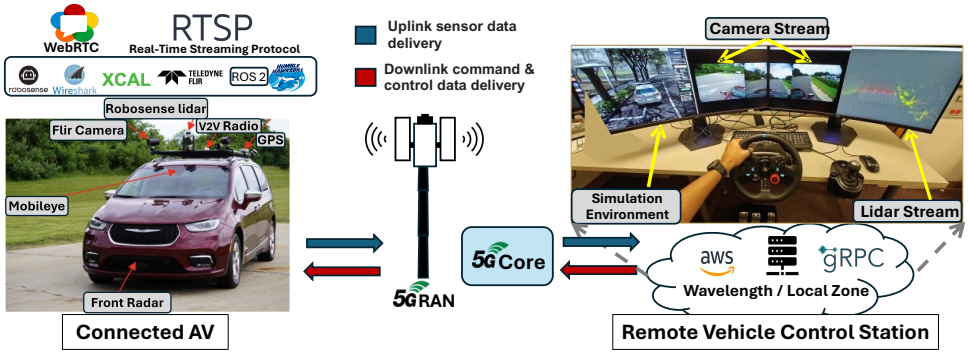


Figure 1: Experimental Setup, Tools, & Streaming System.

Table 1: AV Teleoperation Latency Requirements

Application Level	5G Network Level
UL: 100 ms	UL: 40 - 45 ms
DL: 20 ms	DL: 15 ms

Table 2: Sensor Data Throughput Requirements (in Mbps).

Camera (UL)		LiDAR (UL)		C&C (DL)
1-Steam	4-Streams	64-Beams	128-Beams	
8	32	277	307	0.3

2 BACKGROUND AND MOTIVATION

Fig. 1 shows our testbed, a typical setup – an AV equipped with several sensors, sending data via 5G in the UL to a remote vehicle control station. A remote teleoperator then issues a maneuver, trajectory C&C in the DL to the vehicle. The delivery of sensor and C&C data via the 5G networks has strict latency requirements, which if not met, can be dangerous and potentially even catastrophic, especially in unexpected time-critical situations facing the vehicle. In this section, we discuss the application-level requirements for teleoperation and the required throughput to stream vehicle sensor data. We then provide some background on existing commercial 5G networks and motivate our feasibility study.

Latency Requirements. Table 1 summarizes the latency requirements as specified by 5GAA [12]. These requirements are derived from system analysis, simulation/emulation studies, and field experiments (see [10, 11, 13, 16, 58]). Clearly, the DL C&C data has a more stringent delay requirement of 20 ms and 15 ms for the application-level and 5G network-level respectively, but consumes far less bandwidth (*i.e.,* 0.3 Mbps). In the UL, the application-level latency requirement is 100 ms while the 5G network-level latency requirement is 40–45 ms. These requirements are also in line with the observations gained from emulation-based human subject studies in [13, 16, 58, 66] – in which, a human driver can generally steer the vehicle remotely when the application-level UL latency is under 100 ms, but performance degrades quickly beyond that, becoming nearly impossible above 500 ms.

2.1 ToD Performance Requirements

AV Sensor Data Rate Requirements. We use our AV's top view, shown in Fig. 2, as an example of some sensors in a typical AV. The Global Navigation Satellite System (GPS/GNSS), enhanced by real-time kinematic (RTK) positioning and onboard Inertial Measurement Unit (IMU) and odometer readings, provides accurate (centimeter-level) data on location, speed, orientation, and angular rate—essential for trajectory tracking and driving control. The GigE RGB cameras capture wide road views for situational awareness and object detection, while a thermal camera aids nighttime detection by showing heat signatures. LiDAR uses light reflections to measure distances and create high-resolution 3D point clouds. When augmented with Camera data feeds, the depth information can be associated with each object, giving the remote operator a sense of how far each object is. Radar, though less precise, detects objects by measuring distance, azimuth, and velocity using radio waves, offering a longer range than LiDAR. Refer to Table 4 in Appendix 10.2 for a detailed discussion of all our AV sensors. According to [3] and our data collection, LiDAR and camera data constitute 78.3% and 18% of all sensor data streamed by a typical AV, while all other data types constitute < 1%. Thus, our analysis in this paper mainly focuses on camera and LiDAR data with throughput requirements of 8 Mbps for a single camera, 277 Mbps and 307 Mbps for a 64- and 128-beams LiDAR data, respectively (see Table 2).

2.2 5G Networks Today: PHY Performance

Commercial 5G networks are widely deployed worldwide, and a number of measurement studies have been published characterizing their performance (see *e.g.,* [28, 32, 36, 43], and §3 for more discussion). In Minneapolis, where we conduct our AV feasibility study, all three major U.S. operators, AT&T (AT), T-Mobile (TM), and Verizon (VZ), are deployed. At the time of our study, AT and VZ deployed their 5G service using the 5G Non-Standalone (5G-NSA) mode – which depends on 4G Evolved Packet Core (EPC). AT and VZ utilize primary 5G mid-band (more specifically, the C-band) and mmWave channels, respectively. In contrast, since spring 2023 TM deployed its 5G service in the 5G-SA mode – which depends on 5G Core Network (5G Core). TM utilizes multiple mid-band channels (in band n25, n41) as well as a few low-band channels (in band n71). All mid-band channels (with the exception of TM n25 band) and mmWave high-band channels use the Time Division Duplexing (TDD) mode for their

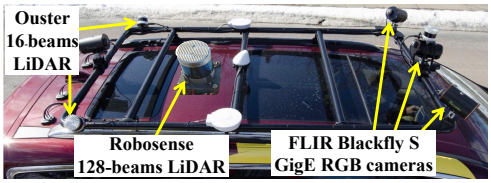


Figure 2: AV top-view with Sensors.

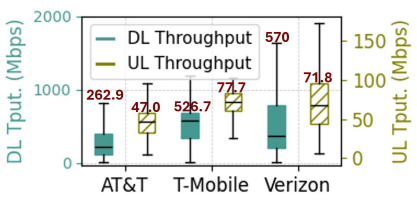


Figure 3: PHY DL & UL Throughput.

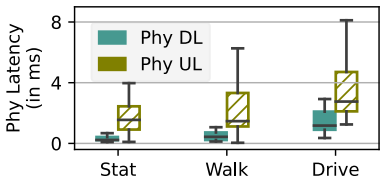


Figure 4: PHY DL & UL Latency.

5G services; whereas the TM n25 and low-band channels use the Frequency Division Duplexing (FDD).

"Best" Achievable 5G Throughput. To gauge the feasibility of teleoperation in commercial 5G networks today, we measure the "best" (max) 5G throughput today by conducting a series of repeated bulk data transfers using iPerf3 [42] (see §4 for our detailed measurement campaign). These experiments are conducted under mobility, where the AV is traveling at 16 – 65 kilometers per hour (km/h) with stop-and-go at intersections. Fig. 3 shows the DL and UL PHY layer layer throughput performance. We see DL throughput is significantly higher than the UL throughput for all operators (notice the different scale of y-axis for DL and UL). Notably, while the average DL PHY layer throughput of AT, TM, and VZ are 262.9 Mbps, 526.7 Mbps, and 570 Mbps respectively, their UL PHY throughput are 47 Mbps, 77.7 Mbps, and 71.8 Mbps respectively. Among all three operators, the best (*i.e., peak*) DL and UL PHY layer throughput performance of VZ (1639 Mbps and 174 Mbps) is due to its mmWave radio band which is highly sensitive to obstruction [49]. Such significant *DL/UL* asymmetry in throughput is, in a sense, by design, largely coming from the TDD frame structure [28]. This is mainly because today's commercial 5G networks (as in 4G LTE networks) are "optimized" for mobile Internet access, where the majority of applications are DL-centric. This is problematic for UL-centric applications, such as teleoperation.

Achievable 5G Latency. Inspired by [27, 29, 43], we quantify the 5G PHY layer ("over-the-air" data) latency defined as; the *PHY DL + UL latency*. In Fig. 4, we show the PHY UL and DL latency for TM when stationary, walking, and driving. Again, we see a clear *DL/UL asymmetry in latency*. Mobility further worsens latencies. These *DL/UL asymmetry* and mobility undoubtedly pose significant challenges for teleoperation today. At the time of our study, TM achieves the "smallest" average DL and UL latency probably due to its 5G-SA architecture [72]. Thus, unless otherwise mentioned, our analysis mainly focuses on TM.

Feasibility of Teleoperation. In total, based on these AV sensor data requirements and "peak" achievable 5G throughput today, one could paint a picture of the feasibility (or infeasibility) of teleoperation over commercial 5G today. Using TM as an example which boasts an UL throughput of 77.7 Mbps, streaming 4 camera streams plus one 64 beams LiDAR with a 309 Mbps total UL throughput requirements (see Table 2), would be unrealistic within the stringent latency requirements (see Table 1). Furthermore, the actual performance in the wild would vary dramatically under the weight of all sensors streaming, highly diverse channel conditions, network resource competition, and more.

3 RELATED WORKS

5G Measurements. There is a plethora of in the wild measurement studies in China [71], the U.S. [21, 29, 31, 32, 36, 47, 49–52, 56, 59, 60, 72, 73], and Europe [28, 30, 43, 44, 55]. All characterize the performance of commercial 5G networks from the user's perspective across different layers in various scenarios including mobility, when roaming abroad, with high traffic load, and when streaming.

Most of these works have considered DL-centric applications. Relevant to our work, are studies that aim to understand UL-centric applications. For instance, [31, 32] (and others) studied large data upload, camera sensor delivery, and other killer-apps like AR/VR and Connected and Autonomous Vehicle (CAV). All these studies revealed important insights in terms of camera sensor data streaming over commercial 5G networks, analyzing several factors like channel conditions, handovers, edge server placement for UL-centric applications. However, these studies; i) did not consider a realistic testbed equipped with several sensors including LiDAR data streaming, down-sampling and voxel-based compression techniques for real-sensor data delivery, and ii) to a great extent failed to offer an in-depth discussion of the intricate interplay of streaming multiple sensor data over commercial 5G – for instance, highlighting the challenges that several factors like BLERs, handovers, and channel conditions while driving and taking sharp turns (≤90 turns) presents for AV teleoperation today.

AV Teleoperation. Several studies [14, 21, 25, 53, 64, 65] have also studied AV teleoperation over cellular networks. Among them, [14, 21] are the most relevant. In our preliminary work [21], we conducted an in-depth measurement study collecting multi-modal vehicle sensor data, including

video, LiDAR, and 5G network metrics over thousands of kilometers using an emulation testbed. This analysis uncovered challenges in supporting multi-modal streams due to the variable 5G throughput. [14] used WebRTC and OMNeT++ 5G network simulator to improve QoE for a remote-controlled ferry while sailing.

Although these past works advance the understanding of AV teleoperation over 5G, i) they are still in the simulation/emulation phase instead of real-world sensor data streaming, and ii) there is a noticeable absence of cross-layer analysis and the connection between network performance and AV teleoperation QoE. An understanding of the underlying causes of AV teleoperation performance to establish a causal relationship between network performance and QoE has yet to be fully explored. Our paper contributes to the understanding of AV teleoperation over live 5G networks. We provide valuable insights to bridge the existing gap in this area.

4 EXPERIMENTAL SETUP, TOOLS, AND STREAMING QOE METRICS

4.1 Experimental Setup & Data Collection

Testbed and Measurement Platforms. In Fig. 1, we illustrate our experimental testbed. The AV is equipped with an on-board computer to transmit sensor data. For remote vehicle control stations, we rely on an Amazon AWS Cloud [2] server deployed in an AWS Local Zone in the same geographical area as our AV. The AWS Local Zone is the second-nearest "edge" server, with a Round-Trip-Time (RTT) of 36.60 ms±4.58 ms (using TM) from our AV.

Streaming System. We perform a series of live video, LiDAR, and C&C data streaming over 5G. i) In the UL, we stream camera data (at 30 Frames Per Second (FPS) with a single camera feed and a multi-camera merged option) using RTSP and WebRTC, both use Real-time Transport Protocol (RTP) over UDP. We chose WebRTC because, unlike RTSP which does not perform any bitrate adaptation, WebRTC is designed to achieve lower latency by incorporating congestion control detection and bitrate adaptation mechanisms, and currently being used by industry [45] for commercial teleoperation. For LiDAR data streaming, we rely on the Robot Operating System (ROS). The data generated by the LiDAR is encapsulated as ROS messages for streaming. ii) For DL, we send C&C data using the Google Remote Procedure Calls (gRPC) framework[34]. iii) To study the effect of data compression on teleoperation, we utilize various video compression algorithms on both RTSP and WebRTC, and employ voxel-based downsampling of the LiDAR data.

Measurement Tools. Since our goal is to perform a cross-layer analysis of AV teleoperation, we need to collect data at the application layer and the 5G network. i) We rely on Accuver XCAL [8] – a commercial grade tool which collects detailed 5G RAN protocol stack information. ii) At the application layer, using RTSP and WebRTC, we implement logging on both the AV and teleoperator control server-side. Furthermore, we rely on Wireshark [67] data, captured on both the AV and server-side to compute the streaming QoE and C&C delays metrics (See §4.2 and §5.4 respectively). iii) We synchronize the clocks of the AV on-board computer and the AWS server with Network Time Protocol (NTP) to ensure correctness of clocks and timing information.

Data Collection Approach. Conducting experiments in the wild using our research AV is not only very (cost and labor) expensive, but is also very challenging. For this reason, we first conduct a 1748 Km (1086-miles) cross-state driving measurement campaign that spans 5 days, in which our AV is driven by a human driver and we configured the on-board computer to record the sensor data (see Fig. 21 in Appendix 10.4). We then simultaneously stream the pre-recorded sensor data (in the UL) and send the C&C data (in the DL) over commercial 5G network. Since part of our goal is to analyze how 5G networks affect teleoperation, to collect and analyze 5G cross-layer data, we stream the sensor data via several Samsung Galaxy S21 Ultra smartphones, tethered to the on-board computer. This is because our commercial grade 5G data collection tool, XCAL, which as of now only works with Samsung phones.

Our methodology consists of the following steps: *Step 1):* We purchased contract SIM cards for streaming the AV data through the on-board computer via the smartphones to the AWS server. *Step 2):* We then select three driving loops in Minneapolis that span 4-6 Km each and conduct several months of live AV sensor data streaming over 5G networks repeatedly over several hours spanning different time periods – *i.e.,* morning and evening rush hours, as well as afternoons, while collecting data across all layers. *Step 3):* Our experiments involve simultaneously streaming the sensor and C&C data while repeatedly driving in loops around the city center and logging at both the AV and the remote vehicle control station. Before each experiment (one complete loop in our experiments), we not only ensure proper NTP time synchronization, but also compute the time drift between the on-board computer and the remote vehicle control server station (1000 samples) and incorporate these time differences in our analysis.

Measurement Summary. We conducted exploratory measurements over commercial 5G networks using AT, TM, and VZ. Altogether these measurements span about 6 months consuming 100s of GBs of 5G data and driving around the testing loops roughly about 70 times. While we studied many operators, we chose to focus on using TM as it was the only carrier with a primary 5G-SA 5G deployment [21, 72].

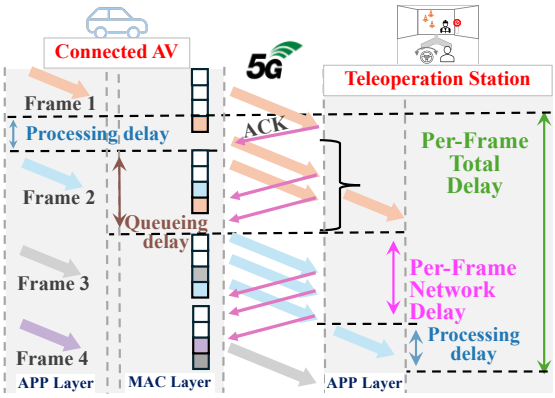


Figure 5: Illustration of the Delay QoE Metrics in Camera/LiDAR Streaming.

Table 3: Definition of QoE Metrics Terms Used

QoE Metrics (One-Way UL)	Definition
Per-Frame Network Delay	One-Way delay between first packet sent and last packet received of a camera/LiDAR frame
Per-Frame Total Delay	*Per-Frame Network Delay* (One-Way delay) plus (+) "queueing" delay and frame encoding and decoding delays
Video Quality	SSIM and PSNR values
Perceptual Quality Deviation	(*SSIM* of frame received at a given playback time − *SSIM* of the expected frame) / *SSIM* of the expected frame

4.2 Streaming QoE Metrics

For teleoperation, ensuring that camera (and LiDAR) data are delivered within a target latency deadline is critical in providing a human teleoperator with real-time situation awareness. Since frames are the basic units for video encoding, decoding, and playback, each frame should therefore be delivered within a target deadline (*e.g.,* 100 ms per Table 1). Thus, as illustrated in Fig. 5 and summarized in Table 3, we introduce the following *Per-frame* QoE metrics to quantify the UL (one-way) camera/LiDAR data sensor streaming performance.

• ***Per-Frame (UL) Total Delay***: The time from when the camera/LiDAR frame is generated at the sender-side (*i.e.,* vehicle) till it is completely received and decoded at the remote teleoperation station and ready for playback.

•***Per-Frame (UL) Network Delay***: The time from when the first packet of a camera/LiDAR frame is sent from the vehicle till the last packet of this frame is received at the teleoperation station side. Note that this delay excludes the time the camera/LiDAR frame spends in the send buffer (known as "queueing" delay) before transmission.

• ***Video Quality***: We rely on the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) to assess the video quality.

• ***Perceptual Quality Deviation***: We introduce this metric to capture the latency-visual-quality interplay, which is defined as follows: assuming a constant latency δ determined by the time the first frame is received and displayed[1], the perceptual quality deviation at the ith playback time is the ratio of the absolute value of the difference between the SSIM value of the *actual* video frame displayed at this time and the *expected* video frame (i.e., the ith frame) that should have been displayed to the SSIM of the ith frame.

For the LiDAR data, we use the term *frame* to encompass the 3D point cloud data generated from a single sensor sweep of a spinning LiDAR sweep, *e.g.,* a 360° sweep. The per-frame total and network delay metrics also apply to LiDAR data. On the other hand, the perceptual quality deviation do not directly apply to LiDAR data. This is because, the LiDAR data is not displayed to the remote human operator. Instead, the LiDAR data will be used for 3D environment representation, object detection and recognition – often combined with the camera data as we also analyze later in §5.3.

5 APPLICATION LEVEL PERFORMANCE

Building on the streaming QoE introduced earlier, this section examines the performance of key sensor streams over commercial 5G networks and the delay of C&C operations.

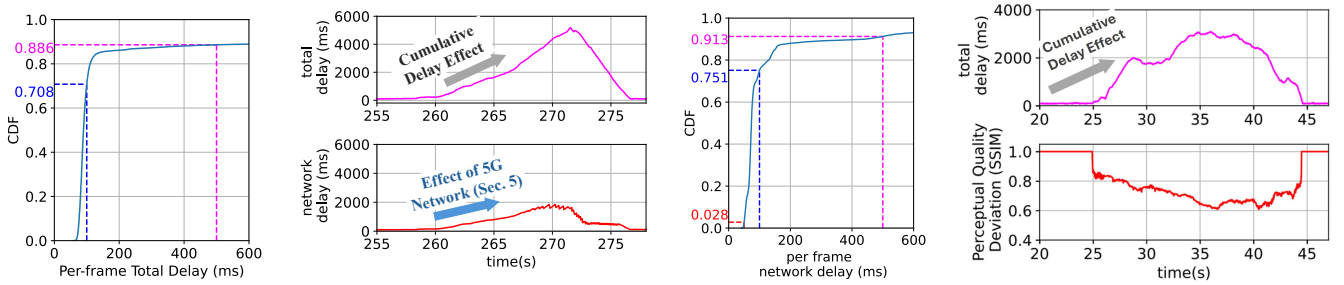
5.1 Streaming Single Front Camera

We first consider streaming the front-central camera data feed – the most crucial data feed for providing real-time situational awareness to the teleoperator. Using TM, we evaluate two scenarios: **Case 1:** We stream the raw MJPEG camera feed using RTSP. **Case 2:** To evaluate the effectiveness of data compression for teleoperation, we consider several compression schemes and stream the compressed camera feed using both RTSP and WebRTC. We repeat these experiments multiple times in our loops.

Case 1: Streaming Raw Camera Feed. Recall from §2 that, for teleoperation to be feasible, the E2E UL delay, *i.e.,* the *Per-Frame Total Delay* should be below 100 ms. We consider the base case (*i.e.,* streaming the front-central camera feed) and quantify the *Per-Frame Total Delay* to understand the feasibility of teleoperation over commercial 5G today.

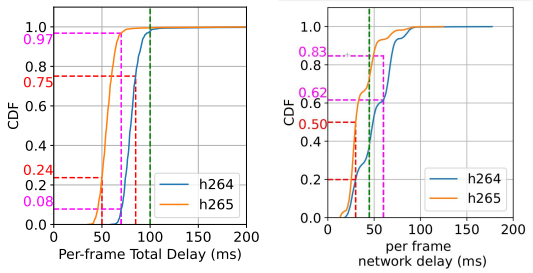
(UL) Per-Frame Total Delay. Fig. 6a shows the *(UL) Per-Frame Total Delay*. We make the following observations: (1) The minimum *Per-Frame Total Delay* exceeds 50 ms, with 29.2% of frames surpassing the 100 ms deadline – of which 11.4% experience delays greater than 500 ms. (2) Notably, the *Per-Frame Total Delay* progressively worsens over time (*i.e.,* accumulates) – a phenomenon we refer to as the *cumulative*

[1]If t_1 is the time the first frame is captured and encoded at the AV, $t_1 + \delta$ is then the time the first frame is received and displayed at the teleoperation station, and $t_i + \delta$ is the time that the ith frame is *expected* to be displayed. If by this time, the ith frame is not received, the previous frame is replayed.



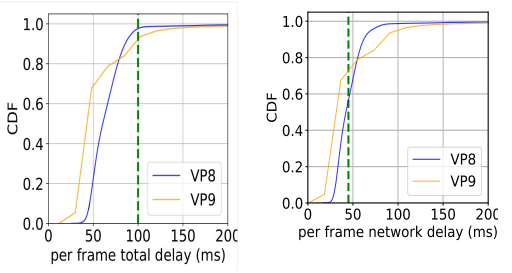
(a) [UL] *Per-Frame Total Delay* (b) [UL] *Per-Frame Total* and *Per-Frame Network Delay* (c) [UL] *Per-Frame Network Delay* (d) [UL] *Per-Frame Total Delay* and *Perceptual Quality Deviation*

Figure 6: QoE Metrics When Streaming the Raw Front-Central Camera – the Most Crucial Data Feed Over TM 5G



(a) [UL] *Per-Frame Total Delay* for I Frames (b) [UL] *Per-Frame Network Delay* for I Frames (c) *Video Quality* Indicated with PSNR

Figure 7: Front-Camera Streaming H.264 and H.265 Performance with RTSP



(a) [UL] *Per-Frame Total Delay* (b) [UL] *Per-Frame Network Delay*

Figure 8: Front-Camera Streaming VP8 and VP9 Performance with WebRTC

delay effect (see Fig. 6b top figure). This is because the AV's camera data begins to queue in the UL buffer. (3) The queuing is triggered by a domino effect due to the 5G network's inability to transmit data to the teleoperation station. This effect is captured by a corresponding increase in *Per-Frame Network Delay*, also illustrated in Fig. 6b bottom plot. The *Per-Frame Network Delay* (partially) provides a clearer justification for the critical role of 5G networks in teleoperation.

(UL) Per-Frame Network Delay. Fig. 6c shows the median *Per-Frame Network Delay* is around 73.5 ms, which is 28 ms higher than the maximum 5G network delay threshold required for teleoperation (see Table 1). However, the distribution has a long tail, with 72.3% of the frames with delay between 50 and 100 ms, 16.2% with delay between 100 ms and 500 ms, and 8.7% with a network delay larger than 500 ms.

Perceptual Quality Deviation. Lastly, we quantitatively understand the *Perceptual quality deviation* in Fig. 6d – The *cumulative delay effect* has a proportionally negative impact on the *Perceptual quality deviation*. In other words, the video frames are progressively delayed, degrading the operator's ability to maintain situational awareness and potentially causing confusion, which can lead to delayed reactions from the teleoperator.

Case 2: Streaming Compressed Camera Feed. Now we ask the question, *can video data compression techniques help teleoperation?* – In other words, *can compression help favor the majority of the frames to arrive within the latency requirements for Teleoperation? What are the tradeoffs?* To answer these questions, we configure H.264 and H.265 with RTSP and VP8[70] and VP9[69] in WebRTC to stream the front-camera data. Fig. 7 and Fig. 8 show the latency performance with RTSP and WebRTC respectively. Generally, compression reduces both the *Per-Frame Total* and *Per-Frame Network* Delays. For instance, with H.265 and RTSP, only 0.487% of the frames experience a *Per-Frame Total Delay* greater than 100 ms, and 17% of the frames experience a *Per-Frame Network Delay* exceeding 45 ms, the target latency requirements for teleoperation. Nonetheless, such tail latency performance is still not ideal, as a critical safety situation that requires timely human teleoperator intervention may occur during a concentrated period of the tail events with bad 5G network conditions. Additionally, we see that increasing compression reduces delay at the expense of lower video quality (see Fig. 7c). Notably, the median *Per-Frame Total* and *Per-Frame Network* Delays achieved with WebRTC are 30.47% and 50.81% lower when compared with RTSP. This is due to the effectiveness of the network adaptation and congestion

control mechanisms present in WebRTC but absent in RTSP, which is the main focus of our investigations later in §7.

5.2 Streaming Merged Cameras

To provide complete real-time situational awareness, it is not sufficient to only stream the data from the front-view camera; the side-views provided by the left-front and right-front cameras are also important, especially when the vehicle needs to change lanes or make turns. We therefore also consider the streaming performance of the merged video streams from all three cameras (see Fig. 21 in Appendix 10.4). Intuitively, the *Per-Frame Total* and *Per-Frame Network* Delays when streaming data from 3 cameras will be worse than the single camera, even with compression. For completeness, we include these results for both RTSP and WebRTC in Fig. 24 and Fig. 22 in Appendix 10.4. Overall, for RTSP, 45% of the frames violate the 45 ms *Per-frame Network* delay deadline and 13% violate the 100 ms *Per-Frame Total* delay deadline. While for WebRTC, 48.31% of the frames will violate the *Per-Frame Network* Delay and 3.64% violate the *Per-Frame Total* delay requirement.

5.3 LiDAR Data Streaming

To stream the LiDAR data, we built a custom client-server UDP-based application that packetizes the ROS bag data for streaming over 5G. In these experiments, we use the term *frame* to encompass the 3D point cloud data generated from a single round of LiDAR sweep, *e.g.*, a 360° sweep, and quantitatively analyze the *Per-Frame Network Delay*. As shown earlier in Table 2, the UL throughput required to stream the 64 and 128 beams LiDAR data is 277 Mbps and 307 Mbps, respectively. Given these throughput requirements, we further explore two approaches to reduce LiDAR data before streaming over 5G.

1) Voxel-based Downsampling. Using downsampling with voxel sizes of 0.1^3 and 0.5^3 (cubic meters) of the 64 beams LiDAR data, we can reduce the throughput requirements from 277 Mbps down to about 121.4 Mbps and 50.9 Mbps, respectively – the former is still significantly higher than the average UL throughput of the three operators. In contrast, the latter is lower than that of TM and VZ, but still slightly above that of AT.

2) LiDAR Compression. We chose Google's Draco [33] because it can reduce the LiDAR data to a third of its original size, and is faster than other LiDAR compression tools we have tested, like Octree [40].

Results. We find the *Per-Frame Network Delay* of the 64 beams LiDAR data streaming is very poor, even with downsampling, as shown in Fig. 9. With 0.5^3 voxel size, the median network delay is 2 seconds, whereas for one raw frame, the median network delay is about 6 seconds, making it nearly

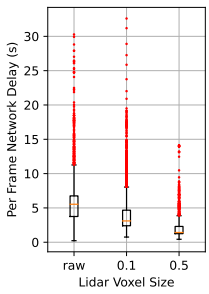


Figure 9: [UL] *Per-Frame Network Delay* of LiDAR.

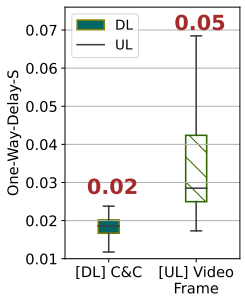


Figure 10: [DL] C&C vs. [UL] Video Frame Delay.

infeasible to stream high-resolution LiDAR data in real-time over 5G networks. We omit the results for streaming 128 beams and multiple LiDAR streams, as they will obviously perform worse. With Draco, although we can significantly reduce the LiDAR data to one-third of its original size, the caveat is that the average compression and decompression times are approximately 47 ms and 15 ms, respectively. In contrast, voxel downsampling takes about 14 ms with no processing required at the remote teleoperation station. The large additional processing overhead associated with Draco, combined with network delays, makes it less suitable for AV teleoperation. Further downsampling of LiDAR data is not beneficial, as it leads to performance degradation in downstream AI tasks, such as object detection and recognition. For interested readers, we provide a detailed discussion of the impact of compression for downstream AI tasks for AV in Appendix 10.5.

5.4 Command & Control (C&C)

As shown in Fig. 1, the test vehicle used in this study is equipped with a DBW (Drive-By-Wire) system, which enables it to read and write CAN (Controller Area Network) messages. The operator remotely controls the vehicle using the Logitech simulator platform at the remote station. Communication between the simulator and the vehicle is established through gRPC communication protocol, ensuring efficient and low-latency data transmission over the 5G network. gRPC's communication model achieves low-latency data transmission by minimizing the use of extensive error-checking mechanisms and connection handshakes. This makes it particularly well-suited for vehicle C&C transmissions, where real-time responsiveness is prioritized over guaranteed delivery. Control commands, including steering, acceleration, and braking, transmitted from the remote station were first converted into low-level commands on the vehicle side before being executed by the vehicle's control systems (See Appendix 10.3 for more details). Fig. 10 shows the DL C&C compared to the UL *Per-frame Network* delays. As per the teleoperation delay requirements in Table 1, we

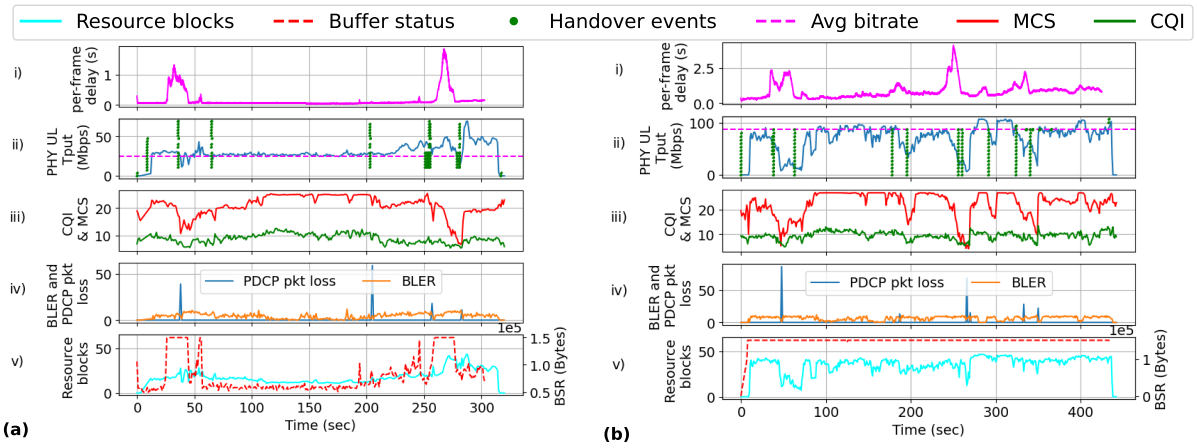


Figure 11: 5G Impact on *Per-Frame (UL) Network Delay* for Teleoperation. Using RTSP to Stream: (a) Single (front) Camera, and (b) Merged (front, left, & right) Cameras

find that 64.29% and 37.20% of the C&C messages were delivered within the application and network requirements, respectively, with a median delay of 17.29 ms.

Implications & Key Takeaways. Our analysis reveals that streaming a single camera's data over a commercial 5G network is feasible. However, situational awareness requires streaming multiple cameras and LiDAR data, which renders teleoperation infeasible. While compression improves feasibility, it degrades visual quality and impairs downstream AI tasks (See Appendix 10.5). Additionally, analysis of the streaming QoEs highlights the critical role of 5G in teleoperation. We next examine 5G's impact on ToD in details.

6 5G IMPACT ON AV TELEOPERATION

Minimizing delay is crucial for teleoperations. However, as shown earlier, *Per-frame network delay* often exceeds the 45 ms or 100 ms requirements (see Table 1) and is unstable. We analyze the 5G dynamics to identify the causes.

6.1 What 5G Factors Affect Teleoperation?

As the AV moves around, the Base Station (BS) (*i.e.,* next Generation Node B (gNodeB)) needs to allocate network resources for UL sensor data transmissions. The gNodeB considers several factors when allocating resources. Among many others, channel condition assessments play a crucial role. The AV reports its channel conditions to the network using a Channel Quality Indicator (CQI) value, which ranges from 1 to 15, where 15 indicates excellent channel conditions. The network, in turn, uses this CQI value to determine which Modulation and Coding Scheme (MCS) to use for the impending transmission [See §5.1.6 in [9]]. Generally, a high CQI value results in a higher MCS value – *if* there is sufficient data buffered to warrant it [29]. This effect is evidenced by the Buffer State Report (BSR) of the AV being always full. Additionally, HOs are triggered to switch between Base Stations (BSs) when the AV moves around, which can result in packet loss and retransmissions – *i.e.,* Block-Level Error Rates (BLERs).

5G *Cross-layer* and *E2E* Correlation Analysis. To visually illustrate the impact of 5G on *Per-frame network delay*, in Fig. 11(a) and Fig. 11(b), we show the time-series plots of the 5G dynamics, cross-correlating them with the AV QoE *Per-frame network delay* for about 7.5 minutes while driving in a loop and streaming a *single front* camera and *merged* camera feeds respectively over TM's 5G-SA network. The first row i) and the second row ii) plot the time series of *Per-frame network delay* and 5G UL PHY layer throughput, respectively, with Handover (HO) events represented using green dotted lines in the second row. Rows iii) to v) show several key 5G parameters: iii) UL CQI and MCS; iv) PHY BLERs and Packet Data Convergence Protocol (PDCP) loss rates; and v) the UL RBs allocated and the AV BSR (right y-axis).

We make the following observations: (1) When the PHY layer throughput drops dramatically (e.g., 250 -280 secs in plot ii) of Fig. 11(b)), the *Per-frame network delay* increases significantly (from 0.1 seconds to 4 seconds), and can last more than 30 seconds. (2) HOs typically occur in regions where there is a noticeable spike (*i.e.,* increase) in the *Per-frame network delay*. Not all HOs cause significant disruption; some cause extended periods of high delay, with values ranging from hundreds of milliseconds to several seconds, and occasionally lasting tens of seconds. A particularly interesting observation is the compounded impact of multiple HOs occurring within a short time window. For instance, in plot ii) of Fig. 11(a) around 250 seconds and in plot ii) of Fig. 11(b) around 270 seconds, the compounded effect of

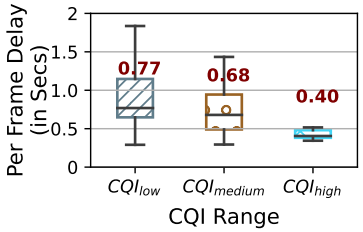


Figure 12: CQI Impact on *Per-frame Network Delay*

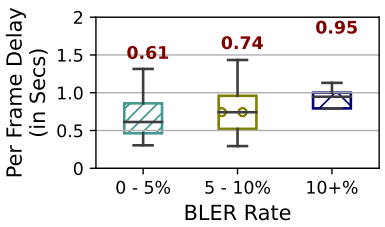


Figure 13: BLER Impact on *Per-frame Network Delay*

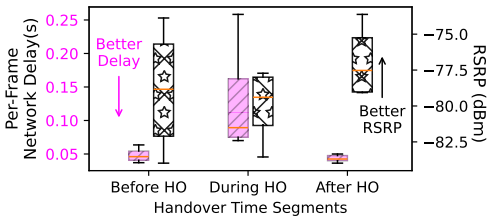


Figure 14: HOs Impact on *Per-frame Network Delay*

several HOs results in much more severe delays than the effect of a single HO during the same period – this is a key point that we explore further in §6.2. (3) BLERs also increase *Per-frame network delay*, although the effects are less pronounced, as we quantify this later (see §6.2). HOs, and to a lesser extent, BLERs, can lead to the AV Medium Access Control Layer (MAC Layer) buffer build-up, as indicated by the buffer status (red curve, right y-axis) in plot v). (4) In single camera streaming, except in the beginning and the end, the number of RBs reaches 20 per second. When the AV MAC Layer buffer builds up, more RBs are allocated as expected – despite poor channel conditions (*i.e.,* low CQI/MCS). In contrast, the merged cameras require significantly higher bit rate, the AV MAC Layer buffer consistently remains full, necessitating the allocation of at least 32 RBs (see plot v) in Fig. 11(b)).

The above *cross-layer* correlation analysis establishes a clear connection between the 5G network dynamics impact on the QoE of AV teleoperation. However, the underlying causes of AV teleoperation performance, and the ability to quantitatively establish a causal relationship between key 5G factors and the AV QoE is yet to be fully explored. Next, we aim to assess and quantify the precise impact of these factors.

6.2 Quantitative Analysis of the Impact of Key 5G Factors on AV Teleoperation

Radio Resource Allocations. The radio resource allocation does not have a *direct* effect on the observed *Per-frame network delay* (Fig. 11(a) & Fig. 11(b)). Particularly, channel conditions (*i.e.,* CQI) do not appear to affect the number of UL RBs allocated by the BS, in contrast to the MAC Layer buffer status. HOs, on the other hand, do sometimes reduce the number of RBs allocated – this is not too surprising. However, when multiple AVs simultaneously stream video feeds, competition for radio resources affects the number of RBs allocated per vehicle, therefore affecting the *Per-frame network delay*, as we later analyze in §8.

5G Channel Conditions. To quantitatively understand how CQI impacts *Per-frame (UL) network delay*, we stream the merged video camera data, which fills the UL buffer (to minimize the impact of data on the network delay) and quantify the direct effect of CQI. We categorize the CQI values into three bins: CQI_{low} = (6, 9] ("poor"), CQI_{medium} = (9, 12] ("fair"), and CQI_{high} = (12, 15] ("good"). Fig. 12 shows the box plot of the *Per-frame network delay* for each CQI bin. We observe that under poor channel conditions, the average *Per-frame network delay* is 770 ms, which is significantly higher than the 400 ms observed under good conditions, representing a 92.5% increase.

BLERs. The 5G PHY/MAC Layer employs the Hybrid Automatic Repeat reQuest (HARQ) to recover from bit errors and failed transmissions. Fig. 13 shows the effect of BLER on the *Per-frame network delay* for a single camera streaming. The delay increases by about 55.7% when transitioning from 0-5% BLER to 10+% BLER, rising from 0.61 seconds to 0.95 seconds. This increase is due to the higher number of PHY/MAC Layer retransmissions.

Handovers (HOs). While previous works [31, 36, 71] have shown that HOs negatively affect the throughput and latency, our goal here is to quantitatively understand their impact on AV teleoperations, specifically the UL *Per-frame network delay*. In the case of intra-RAT (radio access technology) HOs (*i.e.,* 5G → 5G), we focus on A3 HOs observed 97.03% of the times in our experiments; (1) A3 HOs – defined as the Reference Signal Receive Power (RSRP) of the neighboring cell becomes an offset better than the serving cell. (2) A3 Ping-Pong HOs – defined as, multiple HOs between cell ids PCI^{n-1}, PCI^{n}, and PCI^{n+1} happening within a short time window, 15 seconds in our analysis in which $PCI^{n-1} = PCI^{n+1}$.

A3 HOs Impact on AV. For A3 HOs, we are interested in understanding the *Per-frame network delay* (5 seconds) before, during, and (5 seconds) after a HO. Notice that the Δ*Per-frame network delay* before and after a HO quantifies the improvement due to a HO. Fig. 14 shows the *Per-frame network delay* (left axis) and the corresponding RSRP values before, during, and after a HO without HOs failures. See §7 later for the effect of HO failures. Not surprising, ΔRSRP defined as $RSRP_{after} - RSRP_{before}$ a HO is positive. This is a direct consequence of the HO decision. Importantly, although the result of HOs (*i.e.,* after) generally improve the *Per-frame network delay* by about 7.83% in our experiments

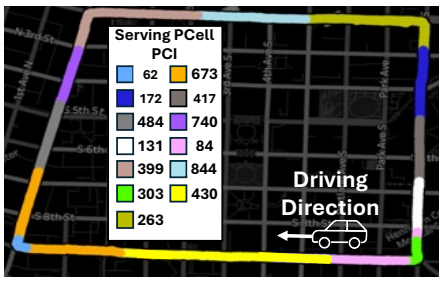


Figure 15: Full Drive Loop, showing the density of PCIs.

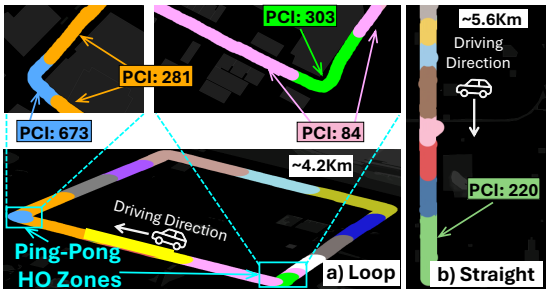


Figure 16: Ping-Pong HOs occurring while driving in a loop and making turns.

(going from 0.046 secs to 0.0424 secs), their impact (*i.e.,* during) is far more detrimental, causing the *Per-frame network delay* to increase (*i.e.,* worsen) by about 86.04% (going from 0.048 secs up to 0.0893 secs).

Ping-pong HOs Impact on AV. We find that ping-pong HOs generally occur when driving in a loop, particularly when turning. To illustrate, we compare HOs when driving in a loop and on a straight line. Fig. 16 shows the cell IDs (*i.e.,* PCIs) with unique colors the AV is connected to when driving in a loop (≈4.2Km) compared to a straight line (≈5.6Km). Notice that, unlike when driving on a straight line, there are zones of ping-pong HOs (*i.e.,* unnecessary HOs) when the AV is turning. For instance, the top-zoomed plot on the left of Fig. 16a) shows that the PCIs the AV is connected to changes from PCI:281 → PCI:673 → PCI:281 within a short time window (15 seconds). The compounded effect of this pong-pong HOs causes a 56-85% in the *Per-frame network delay*, presenting a significant problem for AV teleoperations.

Implications & Key Takeaways. Our analysis points to key 5G factors—HOs (unnecessary ping-pong HOs & A3 HOs), 5G channel conditions, and BLER—which need to be addressed for AV teleoperation on commercial 5G networks.

Practically, one could leverage vertical application information like the vehicle location, speed, and trajectory to i) design a *proactive, trajectory-driven* HOs mechanism to minimize ping-pong HOs, particularly when turning, and ii) implement a *mobility-aware* link adaptation (*e.g.,* MCS selection) mechanism to reduce BLER. Such approaches could make AV teleoperation feasible. Also, understanding adaptation strategies at both the network and application layers could be essential – the major focus of our next section.

7 5G IMPACT AND NETWORK & APPLICATION ADAPTATION FOR TOD

Here, we use WebRTC as a key use case to study how 5G RAN dynamics impact feedback-based congestion control algorithms for real-time sensor data streaming for AVs. Specifically, we show that the 5G channel fluctuates at timescales much smaller than the congestion control feedback delay.

PHY Impact on WebRTC Adaptation Mechanisms. We stream the front-camera to understand how 5G PHY layer dynamics influence WebRTC decision-making. In Fig. 17, we cross-correlate the PHY layer factors with the WebRTC performance. Plot (i) shows 5G PHY layer throughput and RB allocation, and plot (ii) shows WebRTC's target bitrate. Plots (iii) and (iv) illustrate channel conditions – as indicated by the CQI, MCS, and *Per-Frame (UL) Network Delay.*

We see that there is a clear lag (*i.e.,* ~10 seconds) in WebRTC's decision – to explain, in the first 30 seconds, the initial gradual increase in the PHY layer throughput is matched by a corresponding graduate increase in the WebRTC's target bitrate. Between 30 and 40 seconds, the PHY layer throughput suddenly drops within milliseconds as a result of lower RB, despite acceptable CQI and MCS values. This sudden drop causes the video data to queue up in the 5G UL channel, leading to a sharp increase in *Per-Frame Network Delay* (0.1–0.22 s) approximately 10 seconds later. At this point, WebRTC detects congestion and reacts by reducing the target bitrate from 40 to 15 Mbps, which in turn lowers the frame rate. This effect is further amplified by HOs, both with and without failures, as shown in the inset plots at the bottom of Fig. 17. Each inset displays the *Per-Frame (UL) Network Delay* along with HO start and end times. HO failures result in retransmissions at the PDCP sub-layer due to packet loss, increasing the *Per-Frame (UL) Network Delay* by approximately 20–33%. Notably, ping-pong HOs have an even greater impact, as seen in the larger spike in network delay in Fig. 17 plot (iii).

Need for "better" Network Adaptation Mechanism. To further illustrate the interactions between 5G and WebRTC's decision-making process, we use the same trace and plot the cumulative throughput of the PHY layer and target bitrate selected by WebRTC in Fig. 18. Initially, WebRTC underutilizes the available 5G throughput (area A in Fig. 18), as the target bitrate remains below the network's capacity, leading to wasted resources. This occurs because WebRTC's (Google Congestion Control) (GCC) algorithm relies on Real-time Transport Control Protocol (RTCP) feedback, which operates on much larger timescales (≥2 seconds by default) than

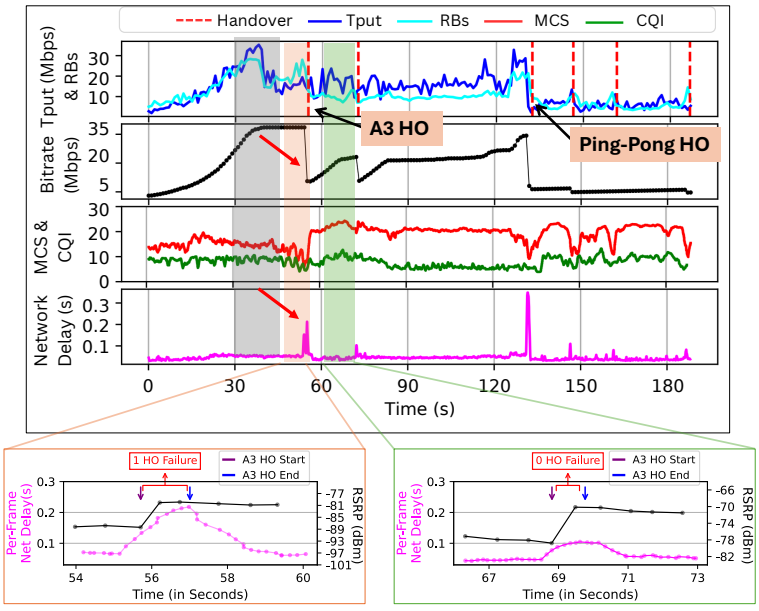


Figure 17: Cross-Layer Analysis of WebRTC Single Camera Streaming: 5G PHY layer Impact on App. Layer Behavior.

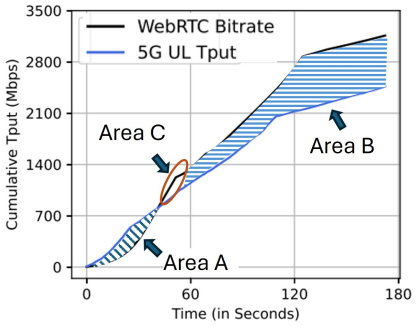
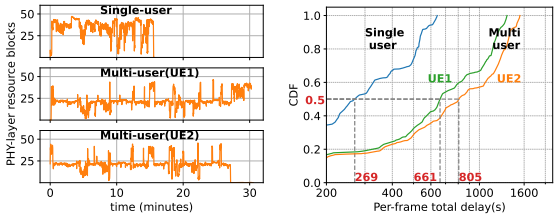


Figure 18: WebRTC's Bitrate Choices vs Available Throughput.



(a) RBs Distribution (b) Per-Frame Total Delay

Figure 19: Effect of multi-AVs

the rapid variations in 5G PHY layer throughput (0.5 ms [29]). As a result, WebRTC fails to immediately adjust its target bitrate, causing a mismatch that leads to queuing delays and congestion as data accumulates in the buffer. In area C, as the 5G throughput decreases, WebRTC continues increasing its target bitrate. Again, due to the delayed feedback, this leads to overutilization (area B), causing spikes in *Per-Frame (UL) Network Delay*. These mismatches highlight the inefficiencies in the application layer (WebRTC) to adapt to 5G's fast dynamics, underscoring the need for more responsive and precise network adaptation mechanisms.

Implications & Key Takeaways. The above analysis confirms that application layer-based adaptation and congestion control mechanisms are not sufficient to cope with the "fast" 5G channel variability caused by mobility in AV teleoperations. The 5G PHY layer metrics provide another critical dimension to detect problems on the 5G network and react quickly, reducing the UL delay. Overall, it is crucial to design applications to be "5G-aware".

8 FURTHER DISCUSSION

Based on our findings, we conclude our work by discussing the effect of multiple AVs operating over commercial 5G networks. The goal is to shed light on possible future directions.

8.1 Effect of Multiple AVs

To illustrate the impact of multiple vehicles competing for radio resources, we experiment with two vehicles (users). These results can be generalized to scenarios with more vehicles. To stress the network, both vehicles stream front 64-beam LiDAR data at 277 Mbps. The top plot in Fig. 19a shows the number of PHY layer RBs allocated when only one vehicle is streaming. The middle and bottom plots show the RB allocation when both vehicles stream simultaneously. As expected, RBs are nearly halved when two vehicles share the network. This resource competition significantly delays sensor data delivery, as shown in Fig. 19b. The CDF of *Per-Frame (UL) Total Delay* reveals that the median delay more than doubles. Specifically, transmitting the entire LiDAR dataset takes 15 minutes for a single user but nearly 30 minutes for two. More competing users would further degrade the delay performance.

8.2 Leveraging Multiple Operators

When multiple 5G operators are available, we can leverage them to support AV teleoperation. One approach is to utilize MPTCP or MP-QUIC [53] by splitting packets (of each frame) across multiple operators. One issue with this approach is that when one of the operators experiences poor channel conditions, it affects the entire frame, thereby prolonging the *Per-Frame (UL) Total Delay*. An alternative approach is to perform operator switching – namely, utilizing one operator for streaming at any given time, and switching to another operator when the UL throughput of the first operator is poor. This strategy can especially help alleviate the poor performance caused by HOs, poor CQI, or poor coverage of a single operator.

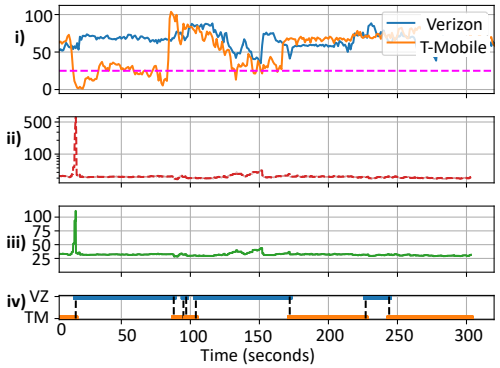


Figure 20: Multi-Path Streaming for Single Video

In Fig. 20, we provide an example to illustrate the performance of operator switching. The top plot (i) shows the PHY layer throughput of TM vs. VZ over time. The second and third plots show the benefits of operator switching: the second plot (ii) shows the *Per-Frame Total Delay* and the third plot (iii) shows the *Per-Frame network Delay*. The fourth plot (iv) indicates the operator being used for streaming the current individual frames. We can observe that the *Per-Frame Total Delay* is consistently below 100 ms, except in situations where a sudden throughput drop leads to incorrect bandwidth estimates, making it a promising approach to consider for teleoperation.

Impact of Data Compression on Downstream AI Tasks. Sensor data received from AVs may also be used for object detection and tracking to help alert the human teleoperator. We also study the impact of data compression on downstream AI tasks. The details can be found in Appendix §10.5.

9 CONCLUSION

We conducted – to the best of our knowledge – a first feasibility study of AV teleoperations over commercial 5G networks in a real-world urban setting, analyzing cross-layer and E2E performance. We distinguish our work from previous studies and introduce per-frame level QoE metrics to elucidate the impact of key 5G PHY layer dynamics – CQI, BLERs, HOs, and RBs allocation – on E2E latency and tail performance. Our study reveals the challenges posed by 5G networks and the limitations of existing sensor data streaming mechanisms. While adaptive frame dropping and multi-operator strategies improve tail latency, they cannot fully mitigate poor channel conditions and frequent handovers, especially when driving in a loop and making turns. New 5G features, such as network slicing [1], can provide resource provisioning and prioritized QoE, but do not resolve the fundamental limitations of the 5G network. We advocate for a co-designed approach integrating wireless networks, edge/cloud systems, and applications to minimize latency and make AV teleoperations feasible over live commercial networks.

Acknowledgment: This research was supported in part by NSF under Grants 1915122, 2128489, 2154078, 2212318, 2220286, 2220292, and 2321531, as well as an InterDigital gift.

REFERENCES

[1] [n. d.]. 5G network slicing. https://en.wikipedia.org/wiki/5G_network_slicing.
[2] [n. d.]. Amazon Web Services (AWS). https://aws.amazon.com/
[3] [n. d.]. Sizing the Solution. https://infohub.delltechnologies.com/en-us/l/computer-vision-3d-flow-and-function-ai-with-lidar/sizing-the-solution-53/6/#:~:text=Network%20sizing%20is%20extremely%20important,s%20and%2050%20Mb%2Fs.
[4] 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, SAE International Recommended Practice Standard J3016-2018. https://www.sae.org/standards/content/j3016_201806/.
[5] 2019. World's First Remotely-Controlled 5G Car To Make History At Goodwood Festival of Speed. https://news.samsung.com/uk/worlds-first-remotely-controlled-5g-car-to-make-history-at-goodwood-festival-of-speed.
[6] 2021. Startup vay's Autonomy Workaround: Teledrivers to Operate Cars from Remote Location. https://www.caranddriver.com/news/a37648114/vay-autonomous-teledriver-startup/.
[7] 2021. This Driverless Car-Sharing Service uses Remote Human "Pilots", not AI. https://www.fastcompany.com/90653650/halo-driverless-car-sharing-service.
[8] 2022. Accuver XCAL. https://www.accuver.com/sub/products/view.php?idx=6&ckattempt=2.
[9] 3GPP. 2020. 5G; NR; Physical layer procedures for data (3GPP TS 38.214 version 16.2.0 Release 16). https://www.etsi.org/deliver/etsi_ts/138200_138299/138214/16.02.00_60/ts_138214v160200p.pdf.
[10] 5GAA. 2021. C-V2X Use Cases and Service Level Requirements Volume II. https://5gaa.org/c-v2x-use-cases-and-service-level-requirements-volume-ii/, Last accessed: Sept 20, 2024.
[11] 5GAA. 2021. Tele-Operated Driving (ToD): System Requirements Analysis and Architecture. https://5gaa.org/tele-operated-driving-tod-system-requirements-analysis-and-architecture/, Last accessed: Sept 20, 2024.
[12] 5GAA. 2024. 5G Automotive Association. https://5gaa.org/, Last accessed: Sept 20, 2024.
[13] 5GCroCo. 2024. 5GCroCo: 5G for Cooperative, Connected and Automated Mobility. https://5gcroco.eu/, Last accessed: Sept 20, 2024.
[14] Manzoor Ahmed, Salman Raza, Muhammad Ayzed Mirza, Abdul Aziz, Manzoor Ahmed Khan, Wali Ullah Khan, Jianbo Li, and Zhu Han. 2022. A survey on vehicular task offloading: Classification, issues, and challenges. *Journal of King Saud University - Computer and Information Sciences* 34, 7 (2022), 4135–4162. https://doi.org/10.1016/j.jksuci.2022.05.016
[15] aler9 and github contributors. [n. d.]. rtsp simple server. https://github.com/aler9/rtsp-simple-server Accessed February 2023.
[16] Rory Bennett, Reyn Kapp, Theunis R Botha, and Schalk Els. 2020. Influence of wireless communication transport latencies and dropped packages on vehicle stability with an offsite steering controller. *IET Intelligent Transport Systems* 14, 7 (2020), 783–791.
[17] Jon Brodkin. 2023. After robotaxi dragged pedestrian 20 feet, Cruise founder and CEO resigns. https://arstechnica.com/tech-policy/2023/11/after-robotaxi-dragged-pedestrian-20-feet-cruise-

founder-and-ceo-resigns/.

[18] Martin Buehler, Karl Iagnemma, and Sanjiv Singh. 2007. *The 2005 DARPA Grand Challenge: The Great Robot Race* (1st ed.). Springer Publishing Company, Incorporated.

[19] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2019. nuScenes: A multimodal dataset for autonomous driving. *arXiv preprint arXiv:1903.11027* (2019).

[20] Ricardo Cano. 2024. One crash set off a new era for self-driving cars in S.F. Here's a complete look at what happened. https://www.sfchronicle.com/projects/2024/cruise-sf-collision-timeline/.

[21] Jason Carpenter, Wei Ye, Feng Qian, and Zhi-Li Zhang. 2023. Multi-Modal Vehicle Data Delivery via Commercial 5G Mobile Networks: An Initial Study. In *2023 IEEE 43rd International Conference on Distributed Computing Systems Workshops (ICDCSW)*. IEEE, 157–162.

[22] Yilun Chen, Zhiding Yu, Yukang Chen, Shiyi Lan, Animashree Anandkumar, Jiaya Jia, and Jose Alvarez. 2023. FocalFormer3D : Focusing on Hard Instance for 3D Object Detection. arXiv:2308.04556 [cs.CV]

[23] A. Davies. 2018. Self-driving cars have a secret weapon : Remote control. https://www.wired.com/story/phantom-teleops/.

[24] A. Davies. 2019. The war to remotely control self-driving cars heats up. https://www.wired.com/story/designated-driver-teleoperations-self-driving-cars/.

[25] Jos den Ouden, Victor Ho, Tijs van der Smagt, Geerd Kakes, Simon Rommel, Igor Passchier, Jakub Juza, and Idelfonso Tafur Monroy. 2022. Design and Evaluation of Remote Driving Architecture on 4G and 5G Mobile Networks. *Frontiers in Future Transportation* 2 (2022). https://doi.org/10.3389/ffutr.2021.801567

[26] Google Developers. 2024. https://webrtc.org/ accessed Nov 2024.

[27] Mohyeldin Eiman. 2020. Minimum Technical Performance Requirements for IMT-2020 radio interface(s). https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Documents/S01-1_Requirements%20for%20IMT-2020_Rev.pdf

[28] Rostand A. K. Fezeu, Jason Carpenter, Claudio Fiandrino, Eman Ramadan, Wei Ye, Joerg Widmer, Feng Qian, and Zhi-Li Zhang. 2023. Mid-Band 5G: A Measurement Study in Europe and US. arXiv:2310.11000 [cs.NI]

[29] Rostand A. K. Fezeu, Eman Ramadan, Wei Ye, Benjamin Minneci, Jack Xie, Arvind Narayanan, Ahmad Hassan, Feng Qian, Zhi-Li Zhang, Jaideep Chandrashekar, and Myungjin Lee. 2023. An In-Depth Measurement Analysis of 5G mmWave PHY Latency and Its Impact on End-to-End Delay. In *Passive and Active Measurement: 24th International Conference, PAM 2023, Virtual Event, March 21–23, 2023, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 284–312. https://doi.org/10.1007/978-3-031-28486-1_13

[30] Claudio Fiandrino and et al. 2022. Uncovering 5G performance on public transit systems with an app-based measurement study. In *Proceedings of the 25th International ACM Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems*. 65–73.

[31] Moinak Ghoshal, Imran Khan, Z. Jonny Kong, Phuc Dinh, Jiayi Meng, Y. Charlie Hu, and Dimitrios Koutsonikolas. 2023. Performance of Cellular Networks on the Wheels. In *Proceedings of the 2023 ACM on Internet Measurement Conference (IMC '23)*. Association for Computing Machinery, New York, NY, USA, 678–695. https://doi.org/10.1145/3618257.3624814

[32] Moinak Ghoshal, Z. Jonny Kong, Qiang Xu, Zixiao Lu, Shivang Aggarwal, Imran Khan, Yuanjie Li, Y. Charlie Hu, and Dimitrios Koutsonikolas. 2022. An In-Depth Study of Uplink Performance of 5G MmWave Networks. In *Proceedings of the ACM SIGCOMM Workshop on 5G and Beyond Network Measurements, Modeling, and Use Cases* (Amsterdam, Netherlands) *(5G-MeMU '22)*. Association for Computing Machinery, New York, NY, USA, 29–35. https://doi.org/10.1145/3538394.3546042

[33] Google. [n. d.]. GitHub - google/draco: Draco library. https://github.com/google/draco. [Accessed 13-06-2024].

[34] Google. n.d.. gRPC: A high performance, open-source universal RPC framework. https://grpc.io/. https://grpc.io

[35] M. Harris. 2018. CES 2018: Phantom Auto Demonstrates First Remote-Controlled Car on Public Roads. https://spectrum.ieee.org/ces-2018-phantom-auto-demonstrates-first-remotecontrolled-car-on-public-roads. IEEE Spectrum, Piscataway, NJ, USA.

[36] Ahmad Hassan, Arvind Narayanan, Anlan Zhang, Wei Ye, Ruiyang Zhu, Shuowei Jin, Jason Carpenter, Z. Morley Mao, Feng Qian, and Zhi-Li Zhang. 2022. Vivisecting Mobility Management in 5G Cellular Networks. In *Proc. of ACM SIGCOMM*. 86–100. https://doi.org/10.1145/3544216.3544217

[37] ANDREW J. HAWKINS. 2022. Cruise's driverless robotaxis are accepting passengers in Phoenix and Austin. https://www.theverge.com/2022/12/20/23518833/cruise-driverless-taxi-austin-phoenix-waitlist.

[38] ANDREW J. HAWKINS. 2022. Waymo's driverless vehicles are picking up passengers in downtown Phoenix. https://www.theverge.com/2022/8/29/23323593/waymo-driverless-vehicles-passengers-downtown-phoenix.

[39] Markus Hofbauer, Christopher B. Kuhn, Goran Petrovic, and Eckehard Steinbach. 2020. TELECARLA: An Open Source Extension of the CARLA Simulator for Teleoperated Driving Research Using Off-the-Shelf Components. In *31st IEEE Intelligent Vehicles Symposium 2020 (IV)*. IEEE, Las Vegas, NV, USA, 1–6.

[40] Lila Huang, Shenlong Wang, Kelvin Wong, Jerry Liu, and Raquel Urtasun. 2020. Octsqueeze: Octree-structured entropy model for lidar compression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1313–1323.

[41] David Ingram. 2023. Two companies race to deploy robotaxis in San Francisco. The city wants them to hit the brakes. https://www.nbcnews.com/tech/tech-news/san-francisco-looks-hit-brakes-self-driving-cars-rcna66204.

[42] iperf3 community. 2023. *iPerf3*. https://iperf.fr/iperf-download.php

[43] Rostand A. K. Fezeu and et al. 2024. Unveiling the 5G Mid-Band Landscape: From Network Deployment to Performance and Application QoE. In *Proceedings of the ACM SIGCOMM 2024 Conference*.

[44] Rostand A. K. Fezeu, Claudio Fiandrino, Eman Ramadan, Jason Carpenter, Daqing Chen, Yiling Tan, Feng Qian, Joerg Widmer, and Zhi-Li Zhang. 2024. Roaming across the European Union in the 5G Era: Performance, Challenges, and Opportunities. In *Proc. of IEEE INFOCOM*. Available online: https://git2.networks.imdea.org/wng/5g-eu-roaming.

[45] Riley Kaminer. 2024. With new office, Guident zooms into the future of autonomous driving. https://refreshmiami.com/news/with-new-office-guident-zooms-into-the-future-of-autonomous-driving/

[46] Philip Koopman and Michael Wagner. 2016. Challenges in Autonomous Vehicle Testing and Validation. *SAE International Journal of Transportation Safety* 4, 1 (April 2016), 15–24. https://doi.org/10.4271/2016-01-0128

[47] Y. Liu and C. Peng. 2023. A Close Look at 5G in the Wild: Unrealized Potentials and Implications. In *Proc. of IEEE INFOCOM*. 1–10.

[48] Cade Metz, Jason Henry, Ben Laffin Laffin, Rebecca Lieberman, and Yiwen Lu. 2024. How Self-Driving Cars Get Help From Humans Hundreds of Miles Away. *New York Times* (2024). https://www.nytimes.com/interactive/2024/09/03/technology/zoox-self-driving-cars-remote-control.html

[49] Arvind Narayanan, Eman Ramadan, Jason Carpenter, Qingxu Liu, Yu Liu, Feng Qian, and Zhi-Li Zhang. 2020. A First Look at Commercial 5G Performance on Smartphones. In *Proc. of The Web Conference*. 894–905.

[50] Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand AK Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, Feng Qian, and Zhi-Li Zhang. 2020. Lumos5G: Mapping and predicting commercial mmWave 5G throughput. In *Proc. of the ACM Internet Measurement Conference*. 176–193.

[51] Arvind Narayanan, Muhammad Iqbal Rochman, Ahmad Hassan, Bariq S Firmansyah, Vanlin Sathya, Monisha Ghosh, Feng Qian, and Zhi-Li Zhang. 2022. A comparative measurement study of commercial 5G mmwave deployments. In *IEEE INFOCOM 2022*. IEEE, 800–809.

[52] Arvind Narayanan, Xumiao Zhang, Ruiyang Zhu, Ahmad Hassan, Shuowei Jin, Xiao Zhu, Xiaoxuan Zhang, Denis Rybkin, Zhengxuan Yang, Zhuoqing Morley Mao, Feng Qian, and Zhi-Li Zhang. 2021. A Variegated Look at 5G in the Wild: Performance, Power, and QoE Implications. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference* (Virtual Event, USA) *(SIGCOMM '21)*. Association for Computing Machinery, New York, NY, USA, 610–625. https://doi.org/10.1145/3452296.3472923

[53] Yunzhe Ni, Zhilong Zheng, Xianshang Lin, Fengyu Gao, Xuan Zeng, Yirui Liu, Tao Xu, Hua Wang, Zhidong Zhang, Senlang Du, et al. 2023. CellFusion: Multipath Vehicle-to-Cloud Video Streaming with Network Coding in the Wild. In *Proceedings of the ACM SIGCOMM 2023 Conference*. 668–683.

[54] Associated Press. [n. d.]. Driverless taxis are coming to the streets of San Francisco. NPR Technology, https://www.npr.org/2022/06/03/1102922330/driverless-self-driving-taxis-san-francisco-gm-cruise, June 3, 2022. Last accessed: June 8, 2022.

[55] Darijo Raca, Dylan Leahy, Cormac J. Sreenan, and Jason J. Quinlan. 2020. Beyond Throughput, the next Generation: A 5G Dataset with Channel and Context Metrics. In *Proc. of ACM MMSys*. 303–308. https://doi.org/10.1145/3339825.3394938

[56] Eman Ramadan, Arvind Narayanan, Udhaya Kumar Dayalan, Rostand A. K. Fezeu, Feng Qian, and Zhi-Li Zhang. 2021. Case for 5G-Aware Video Streaming Applications. In *Proc. of the 5G-MeMU*. 27–34.

[57] Joseph Redmon. [n. d.]. YOLO: Real-Time Object Detection — pjreddie.com. https://pjreddie.com/darknet/yolo/. [Accessed 14-06-2024].

[58] RoboAuto. 2024. RoboAuto. https://roboauto.tech/, Last accessed: Sept 20, 2024.

[59] Muhammad Iqbal Rochman, Vanlin Sathya, Norlen Nunez, Damian Fernandez, Monisha Ghosh, Ahmed S. Ibrahim, and William Payne. 2022. A Comparison Study of Cellular Deployments in Chicago and Miami Using Apps on Smartphones. In *Proc. of ACM WiNTECH*. 61–68. https://doi.org/10.1145/3477086.3480843

[60] Muhammad Iqbal Rochman, Wei Ye, Zhi-Li Zhang, and Monisha Ghosh. 2024. A Comprehensive Real-World Evaluation of 5G Improvements over 4G in Low-and Mid-Bands. *IEEE DySPAN'24* (2024).

[61] EMMA ROTH. 2023. San Francisco wants to slow robotaxi rollout over blocked traffic and false 911 calls. https://www.theverge.com/2023/1/29/23576422/san-francisco-cruise-waymo-robotaxi-rollout.

[62] SAE International. 2014. Automated Driving: Levels of Driving Automation are Defined in New SAE International Standard J3016. SAE International Troy, MI.

[63] SAE International. 2018. Surface vehicle. SAE International.

[64] Javier Saez-Perez, Qi Wang, Jose M. Alcaraz-Calero, and Jose Garcia-Rodriguez. 2023. Design, Implementation, and Empirical Validation of a Framework for Remote Car Driving Using a Commercial Mobile Network. *Sensors* 23, 3 (2023). https://doi.org/10.3390/s23031671

[65] Andreas Schimpe, Johannes Feiler, Simon Hoffmann, Domagoj Majstorović, and Frank Diermeyer. 2022. Open Source Software for Teleoperated Driving. In *2022 International Conference on Connected Vehicle and Expo (ICCVE)*. 1–6. https://doi.org/10.1109/ICCVE52871.2022.9742859

[66] Gaurav Sharma and Rajesh Rajamani. 2024. Teleoperation Enhancement for Autonomous Vehicles Using Estimation Based Predictive Display. https://drive.google.com/file/d/1HBbEbzKttGW4TQPHQFWeg39CNALC_351/view?usp=drive_link, Last accessed: Sept 20, 2024.

[67] Wireshark Team. [n. d.]. Wireshark. https://www.wireshark.org/

[68] Motor Trend. 2022. Tech Company Testing Remote Operators as Self-Driving Car Backups. https://www.motortrend.com/news/mira-self-driving-car-remote-control-car/

[69] Justin Uberti, Stefan Holmer, Magnus Flodman, Danny Hong, and Jonathan Lennox. 2021. *RTP Payload Format for VP9 Video*. Internet-Draft draft-ietf-payload-vp9-16. Internet Engineering Task Force. https://datatracker.ietf.org/doc/draft-ietf-payload-vp9/16/ Work in Progress.

[70] Paul Wilkins, Yaowu Xu, Lou Quillio, James Bankoski, Janne Salonen, and John Koleszar. 2011. VP8 Data Format and Decoding Guide. RFC 6386. https://doi.org/10.17487/RFC6386

[71] Dongzhu Xu, Anfu Zhou, Xinyu Zhang, Guixian Wang, Xi Liu, Congkai An, Yiming Shi, Liang Liu, and Huadong Ma. 2020. Understanding Operational 5G: A First Measurement Study on Its Coverage, Performance and Energy Consumption. In *Proc. of ACM SIGCOMM*. 479–494. https://doi.org/10.1145/3387514.3405882

[72] Wei Ye, Jason Carpenter, Zejun Zhang, Rostand A. K. Fezeu, Feng Qian, and Zhi-Li Zhang. 2023. A Closer Look at Stand-Alone 5G Deployments from the UE Perspective. In *2023 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*. IEEE, 86–91.

[73] Wei Ye, Xinyue Hu, Steven Sleder, Anlan Zhang, Udhaya Kumar Dayalan, Ahmad Hassan, Rostand A. K. Fezeu, Akshay Jajoo, Myungjin Lee, Eman Ramadan, Feng Qian, and Zhi-Li Zhang. 2024. Dissecting Carrier Aggregation in 5G Networks: Measurement, QoE Implications and Prediction. In *Proc. ACM SIGCOMM*. Association for Computing Machinery, Sydney, NSW, Australia. https://doi.org/10.1145/3651890.3672250

[74] Tao Zhang. 2020. Toward Automated Vehicle Teleoperation: Vision, Opportunities, and Challenges. *IEEE Internet of Things Journal* 7, 12 (2020), 11347–11354. https://doi.org/10.1109/JIOT.2020.3028766

10 APPENDIX

10.1 Ethics

This study was carried out by the research team, volunteers, and paid graduate students. No personally identifiable information (PII) was collected or used, nor were any human subjects involved. Our study complies with the customer agreements of all 5G operators. This work does not raise any ethical issues.

10.2 Experimental Testbed & Methodology

Our CAV Sensors. We measured the data footprint of several sensors attached to our CAV to establish a baseline for the network support required by a teleoperated vehicle. These footprints are shown in Table 4. We focused on a 128-beam LiDAR, the front-facing FLIR camera, and the combination of front, left, and right cameras for merged video experiments, as these sensors are the primary contributors to teleoperation tasks. Their critical role in perception, coupled with the high volume of data they generate, necessitates significant network throughput. To collect data, we used ROS2 Humble to record both camera and LiDAR outputs,

Table 4: Our CAV Sensors with Throughput Requirements and Default Sample Rate

Sensor Type	Sensor Model	Raw Data Rates(Mbps)	Raw Data Rates(Frames)
GPS	Novatel PWRPAK7-E2 GNSS	0.29	NA
IMU	2 OS1-16 and 1 OS1-64 Ousters	0.038	100
IMU	Novatel PWRPAK7-E2 GNSS	0.33	NA
Odometry	Novatel PWRPAK7-E2 GNSS	0.28	NA
Video	Front-FLIR Blackfly S GigE RGB	24.994	30
Video	Right-FLIR Blackfly S GigE RGB	25.467	30
Video	Left-FLIR Blackfly S GigE RGB	37.749 *	30
Video	Front-FLIR ADK Thermal	4.561	15
LiDAR	Robosense Ruby Plus 128	125	10
LiDAR	Right-Ouster OS1-16	63.411	10
LiDAR	Left-Ouster OS1-16	64.278	10
LiDAR	Front-Ouster OS1-64	276.814	10
RADAR	Front-Conti ARS 408	1.4	NA
RADAR	Rear-Conti ARS 408	1.71	NA

Figure 21: Merged Frames from Front Left, Front, and Front Right Camera.

leveraging manufacturer-provided SDKs such as Spinnaker for the cameras and RSlidar SDK for the LiDAR systems.

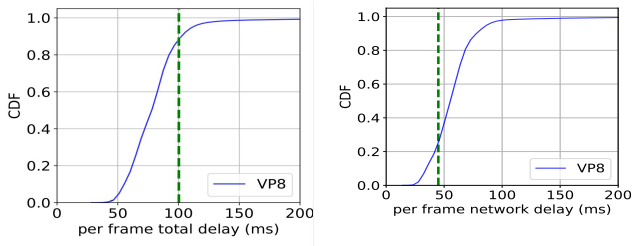
Testbed Our testbed consists of a laptop (representing an on-board unit (OBU)), a USB-tethered smartphone (serving as the 5G radio), and a cloud-hosted server (functioning as our driver station/location). We utilize WebRTC and two Chrome browser instances, each with Python logging servers co-located on each side, to capture logs. Additionally, we use OBS to stream our CAV camera data as a webcam for use in the WebRTC client. We record the resulting video on the server side using JavaScript recording libraries.

10.3 Teleoperation Command & Control

In our system, we utilize gRPC, which is an industry standard C&C framework for teleoperation control. The underlying system utilizes HTTP/2 served over TCP to facilitate command transmission. We record commands from a Logitech simulator platform and replay them over gRPC to mimic a real set of drive commands. This provides a reasonable simulation of drive commands delivered over the network simultaneously with our video streams. This provides a full picture of simulated driving.

10.4 Teleoperation Latency Performance

We include in this section some additional results and context for our video streaming latency experimentation. Fig. 21 illustrates a frame from the merged video stream combining inputs from the front-left, front, and front-right cameras of the CAV vehicle. This merged video is transmitted



(a) Per-Frame Total Delay **(b) Per-Frame Network Delay**

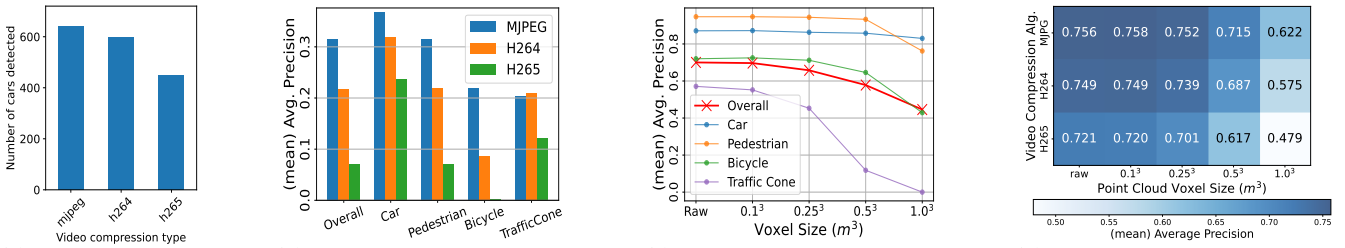
Figure 22: WebRTC QoE for merged-video streaming

over a 5G network to the remote vehicle control station, enabling comprehensive situational awareness for teleoperation. The results from our experiments with merged-video streaming using VP8 codec are illustrated in Fig. 22. We analyze WebRTC metrics for large-throughput video streaming. Fig. 22(a) depicts the per-frame total delay which remained below the 100 ms threshold for about 90% of the frames. Fig. 22(b) focuses specifically on per-frame network delay, around 20% of the frames remained below the 45 ms threshold, and around 65% of frames were in the range of 50-100 ms. The total per-frame delay was slightly higher due to additional processing overhead, yet still largely within acceptable limits for real-time applications.

Fig. 24 evaluates the QoE for merged-video streaming using different video compression methods, including MJPEG, H.264, and H.265. Fig. 24(a) compares the one-way total delay for each codec, showing that both H.264 and H.265 achieved lower delays compared to MJPEG, with most values below the 100 ms threshold. Fig. 24(b) analyzes the per-frame network delay for these codecs, where H.265 demonstrates the most consistent performance, while H.264 exhibits slightly higher variability. Lastly, Fig. 24(c) evaluates the perceptual quality deviation through the Structural Similarity Metric (SSIM), showing that H.264 and H.265 maintain higher and more stable quality compared to MJPEG. These findings indicate that H.264 and H.265 are more suitable for real-time teleoperation streaming due to their balance of low delay and high video quality.

10.5 Impact on Downstream AI Tasks for AV Teleoperations

The quality of video not only affects human teleoperator perception, but also the efficacy of downstream AI tasks such as detecting and recognizing objects. For example, the detected objects in the video are often displayed with a bounding box and a label to alert the human teleoperator and assist them in timely decision making. With its 3D representation and depth information, LiDAR can enhance the object detection and recognition accuracy. We now evaluate the impacts of data (video/LiDAR) compression on the efficacy of objection



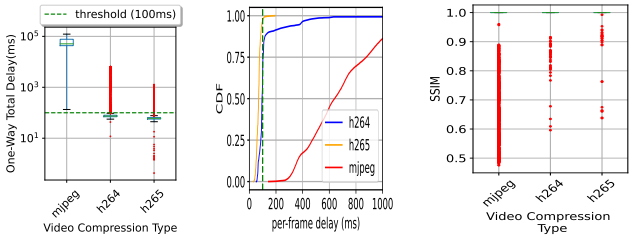
(a) Object Detection for Cars.

(b) Camera-Only Object Detection.

(c) LiDAR-Only Object Detection.

(d) Multi-Modality Object Detection.

Figure 23: Object Detection with Different Data Modalities and Compression Qualities.



(a) Per-Frame Total Delay

(b) Per-Frame Network Delay

(c) Perceptual Quality-Deviation

Figure 24: Merged-videos streaming QoE performance

(a) MJPEG Frame Yolo Detection

(b) H.265 Frame Yolo Detection

Figure 25: Video Compression Codec Effect on ML Detection

detection and recognition. We also examine the benefit of combining both video and LiDAR data in such tasks.

To assess the effects of distortion, we utilize the pre-trained state-of-the-art object detection models: YOLOv8 [57] and FocalFormer3D [22]. As an example, Fig. 25 shows a representative video frame annotated using YOLO: compared to Fig. 25(a), one car is not detected in Fig. 25(b), due to the lower quality of H.265 encoded video. In Fig. 23(a), we compare the number of objects detected using MJPEG, H.264, and H.265 video. We see that compared with MJPEG, H.264 and H.265 reduce the number of object detections due to lower video quality. In particular, while H.265 can significantly reduce the bit rate requirement (e.g., from using 87.97 Mbps to 4.946 Mbps), this comes with a significant penalty in the efficacy of downstream object detection and recognition.

To quantify this impact, we conduct experiments using the nuScenes dataset [19] with ground-truth labels for various objects of interest. Besides MJPEG, H.264, and H.265 for video compression, we employ a downsampling technique with voxel sizes ranging from $0.1m^3$ to $1m^3$ for LiDAR data. Fig. 23(b) and Fig. 23(c) present the *mean Average Precision* (mAP) for all classes and four selected objects types (cars, pedestrians, bicycles, and traffic cones) using video-only and LiDAR-only mono-modality 3D object detection, respectively. Due to the absence of explicit depth information, camera-only detection performs less effectively than LiDAR, with 0.32 mAP score over all classes, while LiDAR achieves 0.70. We can see that lowering the quality of video and LiDAR data has an evident effect on 3D object detection and recognition tasks. It is worth noting that such effect is not linear. For example, H.265 degrades the overall performance of video-based object detection far more aggressively than H.264. LiDAR-based detection only experiences a more pronounced decline only after reaching a voxel size of $0.5m^3$. Furthermore, the effects of data compression differ based on object classes. For example, with low H.265 video quality, bicycles become nearly undetectable. Similarly, traffic cones become undetectable in downsampled LiDAR data with a voxel size of $1m^3$.

We further explore the impact of data compression on *multi-modality* 3D object detection using both video and LiDAR data. The mAP results for all classes are shown in Fig. 23(d). Multi-modality detection outperforms camera-only and LiDAR-only detection, with 171.0% and 39.5% improvements, respectively. The object detection degradation due to LiDAR compression reduces by 2.8%-18.6% when combined with video data. In addition to the better object detection performance, the combination of compressed LiDAR data with $0.5m^3$ voxel size and H.265 video require 25.2% less bandwidth than MJPEG encoded video frames alone (65.76 Mpbs vs. 87.97 Mbps). In summary, using compression and bit rate adaptation may help reduce the per-frame delay (thus increasing the probability of meeting a target deadline), however their impacts on efficacy of downstream AI tasks must be taken into account. Carefully striking the balance in latency and video quality is called for.