Computer Networks 267 (2025) 111344

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Unlocking the power of 4G/5G mobile networks: An empirical dive into quality and energy efficiency in YouTube Edge services

Peixuan Song [a], JunKyu Lee [b], Ahmed M. Abdelmoniem [a,*], Lev Mukhanov [a]

[a] School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom
[b] School of Computer Science and Electronic Engineering, University of Essex, United Kingdom

ARTICLE INFO

Keywords:
4G/5G
Mobile network
Measurements
Performance analysis
Energy efficiency
Edge services

ABSTRACT

The advancements in 5G mobile networks and Edge computing offer great potential for services like augmented reality and Cloud gaming, thanks to their low latency and high bandwidth capabilities. However, the practical limitations of achieving optimal latency on real applications remain uncertain. This paper investigates the latency, bandwidth, and energy consumption of 5G Networks and leverages YouTube Edge service as the practical use case. We analyze how latency, bandwidth, and energy consumption differ between 4G LTE and 5G networks and how the location of YouTube Edge servers impacts these metrics. Surprisingly, our observations show that the 5G ecosystems have average latency hikes of up to 2×, demonstrating that they are far from achieving their proclaimed promises. Our research study reveals over 10 significant observations and implications, indicating that the primary constraints on 4G/LTE and 5G capabilities are the ecosystem and energy efficiency of mobile devices' when receiving data. Moreover, our study demonstrates that to unlock the potential of 5G and its applications fully, it is crucial to prioritize efforts to improve the 5G ecosystem and introduce better methods and techniques to enhance energy efficiency.

1. Introduction

The number of intelligent Internet-connected devices is projected to reach tens of billions [2] soon. All these devices use Internet and Cloud data centers to transfer and store the data. As a result, the size of data transferred through the Internet will exceed 24.3 exabytes soon [3]. Today, mobile devices are one of the primary sources of data transferred to and from the Cloud; the mobile internet traffic generated by 4.7 billion mobile Internet users worldwide is about 57%.

The performance of numerous mobile applications, including YouTube, navigation services, and games, is highly dependent on network latency and bandwidth, which should be sufficient to meet user expectations. Moreover, a new class of emerging applications, such as interactive augmented reality, mobile AI pilots and Cloud gaming applications, generate much more data than state-of-the-art applications and are more critical to network latency and bandwidth.

Edge computing was introduced to address the challenge of increasing mobile data traffic [4]. By bringing computing resources closer to the network's edge, Edge computing can help reduce latency and increase bandwidth, thus improving the performance of applications that require real-time processing. This can be especially beneficial for applications such as augmented reality and self-driving cars.

Google has already implemented 7500 Edge servers to enable the Stadia service, a Cloud gaming service [5], which, however, was deprecated recently. This study demonstrates that Google also places Youtube Edge servers near the base stations. However, the latency and bandwidth achieved in real applications that utilize 5G networks and Edge servers in the UK remains unclear. While some studies suggest that 5G networks can theoretically achieve an average latency as low as 1 ms [6], the real-world implementation may encounter significant overhead due to the 4G/5G ecosystem, i.e. base stations, wired or fiber communication networks and Cloud/Edge servers.

The **main goal** of our study is to investigate what is the best latency and bandwidth that can be achieved by real mobile applications that use 5G networks and Edge servers. Moreover, the study aims to understand how the location of servers handling data requests from mobile devices affects latency.[1] This work also extends the prior work [1] and investigates how downstreaming affects the energy consumption of mobile devices. Studying all these aspects provides a better understanding of what emerging 5G applications can be enabled nowadays

* Corresponding author.
E-mail address: ahmed.sayed@qmul.ac.uk (A.M. Abdelmoniem).

[1] A limited 6-page version of this work has appeared in IEEE Edge [1]. This work presents a significant extension with more analysis and experiments to provide a comprehensive understanding of the operational 4G/5G networks.

https://doi.org/10.1016/j.comnet.2025.111344
Received 16 December 2024; Received in revised form 24 April 2025; Accepted 25 April 2025
Available online 12 May 2025

and what factors prevent exploiting the full capabilities of 4G/LTE and 5G networks. ***Overall, based on the results of our study, we find that fully enabling crucial 5G applications necessitates substantial improvements in the ecosystem and the energy efficiency of 5G modems.***

To enable our study, we use a Google Pixel 4a smartphone, which supports 5G networks. We test the latency and bandwidth the 5G and 4G LTE networks provide when downstreaming YouTube videos for three major mobile network operators in London, UK. We track the location of servers which handle data requests and measure downstream latency and bandwidth, as well as the energy consumed by the mobile device. We use the collected data to analyze the relationships between latency, bandwidth and servers' geographical locations. In addition, we measure the device current and power to investigate the energy efficiency of the mobile device when downstreaming YouTube videos. Given that YouTube is one of the most popular services in the world with billions of users [7], we believe that it is well-optimized to provide the best possible availability, latency and bandwidth. Thus, YouTube can be a reliable indicator of the expected 5G network quality when running real applications that utilize Edge and Cloud servers. Based on the results of our experiments, we make conclusions about which emerging 5G applications can be implemented nowadays, given the constrained ecosystem and battery capacity. In our prior work, we presented limited results of our study on latency and bandwidth for downstreaming in 4G LTE and 5G mobile networks operated by three major operators in London, UK [1]. This work's extends that work and makes the following contributions:

- We present extended results and observations on latency and bandwidth and include energy measurements to understand energy implications of 4G/5G networks.
- We expand our work with key observations about latency and bandwidth such as: (i) the smallest average latency was obtained for 4G LTE, approximately 25 ms, as opposed to 5G; (ii) the average 5G latency varies from 37 ms to 150 ms depending on server location and mobile operator; (iii) the minimum latency measured in our experiments for both 4G LTE and 5G is 20.7 ms and 24.3 ms, respectively; (iv) the average downstreaming bandwidth for 4G LTE and 5G networks typically range around 0.24 Gbps and 0.5 Gbps, respectively.
- We reveal that the 4G/5G ecosystem is a major bottleneck which prevents the full potential of 4G LTE and 5G networks. Enhancing the ecosystem can reduce the average latency by up to 2×. We demonstrate that enabling 1 ms latency for emerging 5G applications will require a network of Edge servers, a distance between which does not exceed 227 km.
- We highlight that energy efficiency is another major bottleneck preventing mobile applications that utilize 4G LTE and 5G implementation. To be more specific, we show that the smartphone current and power increases by 68% on average when downstreaming data and, as a result, some mobile games, such as Realm Grinder, can consume more energy when running on NVIDIA Cloud/Edge servers compared to the version of these games which use only mobile GPUs. Moreover, we show that enabling global YouTube downstreaming on mobile devices would require significant energy, equivalent to a nuclear plant.
- Finally, we discuss which applications can benefit from 5G networks. Our findings indicate that existing London-based 4G LTE and 5G networks, along with the smartphones used in our study, fail to meet the latency requirements for crucial applications such as Autonomous Driving Vehicles and AR/VR 3D rendering.

We hope that the results of our study will serve as a baseline for future studies targeting the optimization and simulation of large-scale 5G networks [8], as well as the dynamic optimization of service placement and eviction for Edge servers [9,10].

The paper is organized as follows: Section 2 presents background and related work; Section 3 presents our experimental framework and methodology; Section 4 presents the results of our experimental study; Section 5 discusses the energy efficiency challenge and if the obtained latency and bandwidth measurements meet the demands of emerging 5G applications; Section 6 demonstrates the limitations of our study and Section 7 presents the conclusion.

2. Background and related work

In 2022, the exponential growth of Internet-connected mobile devices resulted in the generation of a staggering 129.4 Exabytes of data, necessitating its processing in Cloud data centers [11]. However, many user mobile services, such as navigation, Cloud gaming and Augmented reality applications, must receive a reply from servers with a small latency. For example, Cloud gaming implies that the graphical pipeline runs remotely on Cloud servers [12].

Introducing 5G networks should significantly improve latency and bandwidth for mobile data transfers. However, the data should travel between base stations and Cloud data centers, which can negatively affect latency and bandwidth. To address this issue, the concept of Edge computing proposed to place servers close to base stations [4]. Our study aims to understand what latency and bandwidth can be achieved over 5G networks using Edge servers and real applications in practice today. To this end, we use a YouTube mobile application to test 5G network characteristics in London, UK. We specifically use YouTube since this is one of the most popular global services, having 2.56 billion active users [7], which should be well-optimized to achieve the best latency and bandwidth characteristics.

There are several experimental studies which measure the propagation of mmWaves in suburban and vegetated environments [13–19]. These studies comprehensively investigate mmWave propagation under various conditions, encompassing urban environments, suburban and vegetated areas, human body blockage, and rain-induced fading. The main goal of our study is to estimate the latency and bandwidth of 5G that can be achieved in commercial applications with an analysis of the mobile operators in London.

Previous studies tried to investigate possible characteristics of 5G networks in London based on the location of base stations [20]. However, this study uses a simulation-based framework to estimate latency and bandwidth. It does not consider the delays Cloud/Edge networks require to process the user requests and send the data back. Recent papers present the results of extensive research studies on 5G latency, bandwidth and energy efficiency measured on real devices [21–24]. However, these studies use custom applications (or *Speedtest*) and servers to test the quality of 5G networks. A recent study [1] investigated the characteristics of commercial YouTube applications and commercial servers that mobile operators integrate with base stations. Thus, they could estimate latency and bandwidth, which can be achieved by commercial applications using Edge and Cloud servers in practice. This work presents a significant extension with more relevant analysis, experiments and observations to provide a deep dive into quality and energy efficiency for unlocking the power of 4G/5G mobile networks.

3. Experimental design

The tested mobile device. To enable our study, we use a Google Pixel 4a (5G) mobile device which supports 4G LTE (Advanced) and 5G networks. The specification of the device is provided in Table 1.

YouTube service. To test the latency of mobile 5G networks, we use a mobile YouTube application which downstreams data from Edge servers [25]. In particular, we test the mobile network using the YouTube application since it is one of the most demanding commercial applications for latency and bandwidth. We perform our experiments by downstreaming a 5 min YouTube video, which is available by the

Table 1
Specifications of Google Pixel 4a 5G.

Description	Value
Processor	Snapdragon 765G 5G
Cores	1 × 2.4 GHz Cortex-A76 1 × 2.2 GHz Cortex-A76 6 × 1.8 GHz Cortex-A55
L3 cache/DRAM	2.00 MB/6GB LPDDR4X
GPU	3 × 750 MHz Adreno 620
Network technology	GSM/HSPA/LTE 5G (Sub-6 and mmWave)
OS	Android 10

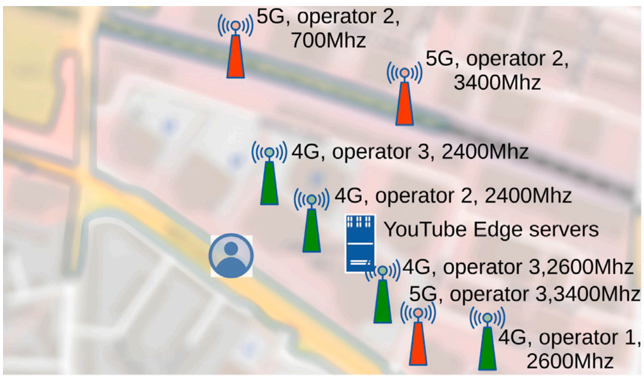


Fig. 1. Base station locations.

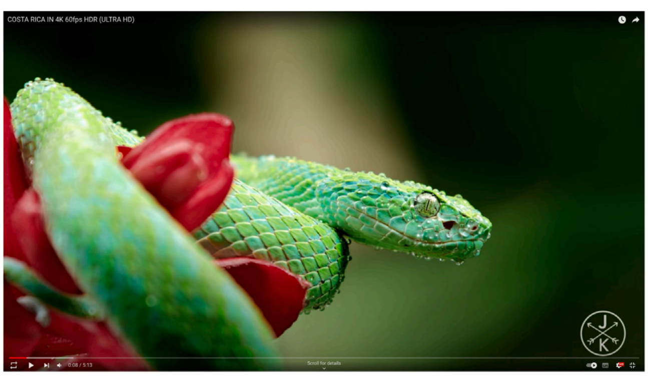


Fig. 2. Video screenshot.

following link.[2] We specifically use this video since it contains contrast figures in the HDR (Ultra HD) format; the maximum resolution is 4096p at 60 FPS (*Frame Per Second*). To test latency under different conditions, we use different resolutions of the video, i.e. 360p (480×360 pixels), 720p (1280×720 pixels), 1080p (1920×1080 pixels) and 4096p ($4096p \times 2160$ pixels). In each experiment, we downstreamed the video 5 times to obtain representative measurements (see Fig. 2). **Latency profiling.** To measure the network latency, we use the Round-Trip Time (*RTT*) metric [26]. RTT is the time between when a package is sent to a destination address and when the acknowledgment is received from the destination address.

To transfer the data over 4G LTE and 5G networks, Android uses *User Datagram Protocol (UDP)* but not *Transmission Control Protocol (TCP)* [27]. Android devices also use QUIC (*Quick UDP Internet Connections*) as the encrypted transport layer on the top of UDP, which improves the Quality-Of-Experience (*QOE*) and reduces the latency [28].

In the TCP protocol, RTT can be measured by the time between a packet being sent and its acknowledgment received. However, QUIC is encrypted and multiplexed, making it hard to measure RTT since a QUIC packet and its acknowledgment might not be on the same path [28]. Nonetheless, QUIC accurately calculates the Round-Trip Time (RTT) by including in its acknowledgment packets the time it takes to receive the packet and send the acknowledgment. This helps QUIC estimate the total RTT for different paths, considering both the inbound and outbound paths. Since QUIC acknowledgments are encrypted and only the RTT from the initial handshake is visible, we use it to estimate latency.

To measure bandwidth and latency, we run *tcpdump* tool [29] with the following parameters: *tcpdump -vv -i any -s 0 -w /sdcard/cap.pcap*.

We parse the log files provided by *tcmpdump* to estimate RTT. Meanwhile, we use a specific network protocol analyzer tool to capture bandwidth and latency, *Wireshark*[30].

Bandwidth profiling. We also parse the QUIC traces to estimate bandwidth and record the number of bytes received every second. Subsequently, we calculate the average bandwidth by aggregating data over one-second intervals. Note that we aggregate data over 1 ms periods to measure the peak bandwidth and project it for one second.

Energy profiling. We use *Perfetto* [31] to profile energy consumption. *Perfetto* uses the data exposed through the charge counters in *Android IHealth HAL* to get the battery current [32]. In particular, this framework measures current in micro ampere, mA, within a small period. Note that we measure the current for the entire mobile device, which includes SOC (i.e. the processor), screen and 4G LTE/5G modem. We estimate energy consumption using the current measurements and battery voltage of 3.85 V on average.

Finding the location of the servers. To find the servers' location, we parse the QUIC logs and extract information about the IP addresses downstreaming the video. We identify servers' location by IP using six web services.[3] Unfortunately, IP geo-location can be inaccurate due to dynamic IP addresses, VPNs, shared IP addresses, inaccurate geolocation databases and limited information [33–36].

We use the results of geo-location services to filter the measurements taken in our experiments; in particular, we remove all the measurements if the same IP address points to different locations in different services. At the second stage of our filtering, we remove all the experiments with the downstreaming servers for which latency is lower than the time required for the light to travel, in the fiber,[4] to the location of these servers and back [34–36].

Location and mobile operators. In our experiments, we test 4G LTE (Advanced) and 5G mobile networks for three major mobile operators, each serving approximately 30% of mobile users in the UK. Thus, we expect that the base stations have almost the same load. To the best of our knowledge, all three operators use 5G NSA (*Non-Standalone*) [20], which relies on 4G networks to send the control information, at the moment when our study was done, i.e. March 2022.

We test mobile networks in the border area near London. We chose this area specifically since the local London servers that stream the video for 2 operators are located in the same building, according to the geo-location web services (see Fig. 1). Furthermore, this building also serves as the location for the base stations of these operators (see *operator 2* and *operator 3* in Fig. 1). We use an online service[5] to find the position of base stations and estimate the signal strength (RSRP) at our specific location. Although the local servers for *operator 1* are located in a different area, one of the base stations of this operator is only 200 meters away from the building. Such a location allows us to minimize the distance between the servers, base stations, and mobile

[2] https://www.youtube.com/watch?v=LXb3EKWsInQ

[3] https://www.ip2location.com/; https://ipapi.co/; https://ipinfo.io/; https://ipgeolocation.io/; https://ipregistry.co/; https://db-ip.com/

[4] The speed of light in the fiber is 2.18×10^8 m/s

[5] https://www.cellmapper.net/

Table 2
Wave frequency spectrum for 5G and 4G.

Bands	Frequency	Support	Bandwidth	Latency
Low-bands	<1 GHz	5G/4G	50–100 Mbps	20 ms
Mid-bands	1 GHz–6 GHz	5G/4G	100–900 Mbps	10 ms
High-bands	24 GHz–47 GHz	5G	10–25 Gbps	1 ms

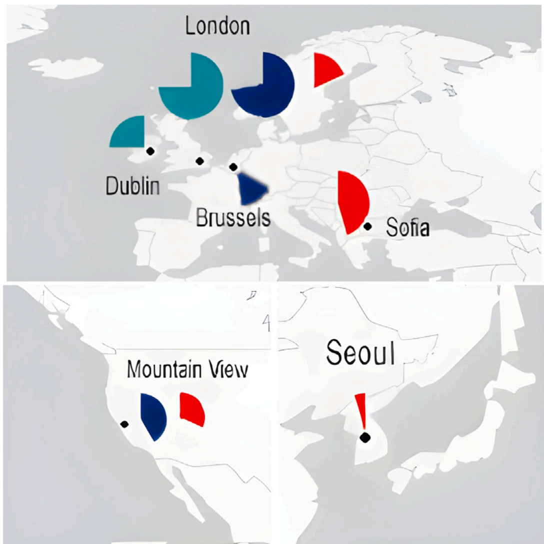


Fig. 3. Server locations streaming the video. The colored pie chart indicates the percentage of packets sent from servers at each location. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

devices. Note that *operator 1* does not provide 5G service in this area. At the location where we conduct our experiments, all three operators provide strong signals, with Reference Signal Received Power (RSRP) values ranging from −40 dBm to −85 dBm.[6] All the operators use *low-band* and *mid-band* base stations in this area for 4G and 5G (see Fig. 1). Each operates a of frequency spectrum frequencies defining defines the maximum possible bandwidth and minimum latency [37,38]. Note that many more parameters, such as channel bandwidth, modulation and types of MIMO antennas, can affect bandwidth and latency [39,40]. Table 2 shows the spectrum of operating frequencies, the maximum bandwidth and minimum latency for each band [37,38,41]. To the best of our knowledge, as of March 2022, there were no publicly available 5G high-band stations in London during the time of our experiments. More importantly, since the experimental area is sparsely populated, we expect a low mobile network load. Considering all the facts provided above, we believe that the area where we conduct our experiments provides optimal conditions for testing the maximum performance of commercial mobile networks.

4. Experimental evaluations

This section presents our experimental results organized into IP-based geo-location, Latency, bandwidth, and Energy Consumption analysis.

4.1. IP-based geo-location analysis

We start our study by investigating the location of the servers that stream the video when we run the YouTube application in the

[6] The average signal strength, measured in terms of Reference Signal Received Power (RSRP), is provided at https://www.cellmapper.net/.

Table 3
Distribution of regions streaming the video.

Operator	Network	# of IPs	UK	Europe	USA	Asia
#1	4G	4	75%	25%	0%	0%
#2	4G	4	75%	25%	0%	0%
#2	5G	3	67%	0%	33%	0%
#3	4G	13	23%	38%	30%	8%
#3	5G	9	11%	56%	33%	0%

abovementioned area. To find the location of the YouTube servers, we downstream the video using 4G LTE and 5G and trace the IP addresses of the servers streaming the video by parsing the *tcpdump* logs. Note that a video streamed by YouTube is split into packages that are downloaded by the YouTube application over some period. The YouTube service is organized so each package can be sent from different servers. In our experiments, we identify all the IP addresses of the servers streaming the test video. Fig. 3 shows the location of YouTube servers streaming the video for three UK mobile operators. Within this figure, each location is depicted with a pie chart indicating the percentage of servers used by a specific operator that were streaming the video from that particular location. The distinct colors within the chart correspond to different mobile operators. We make the following observation.

Observation 1. *Based on our initial observation, it is evident that the servers streaming the video are not exclusively located in London. Although a significant number of servers are situated in London, we have noticed instances where video streaming has taken place through servers in other locations, such as Sofia, Brussels, and even Mountain View (California).*

The distribution of the streaming servers among different countries for each operator and type of the network is summarized in Table 3. Our observation reveals that for *operator 1* and *operator 2*, all the servers streaming the video in 4G are located in the UK (London) or Europe (Dublin for *operator 1* and Brussels for *operator 2*). Meanwhile, *operator 3* utilizes servers mostly located in Europe, Asia, and the USA. For 5G networks, we also see that the servers from different regions were used. Importantly, all the servers streaming the video belong to Google, according to the IP geo-location services used. We attribute this difference in the distribution of streaming servers to varying ISP (Internet Service Provider) Peering Agreements between operators and Google. According to these agreements, operators may use different content delivery networks and servers to optimize congestion management and routing policies [42].

Observation 2. *The location of servers streaming the video to the mobile device depends on mobile operators.*

Interestingly, the number of servers streaming the video varies across operators and network types. As an example, *Operator 2* employs four servers, which are situated in both London and Brussels, to facilitate video streaming via 4G LTE. However, when utilizing 5G technology, the same operator relies on only three servers located in London and Mountain View for video streaming. Meanwhile, *operator 3* uses 13 and 9 different servers distributed across the UK, Europe, Asia, and the USA for 4G LTE and 5G. We presume that Google chooses the servers and streaming route based on each operator's facilities and the network load.

4.2. Latency analysis

Fig. 4 presents the distribution of measured latencies of downstreaming the video for each operator and network type. We present the distribution as a violin plot that depicts the median values (black lines within rectangles),[7] the average values (white dots), the interquartile

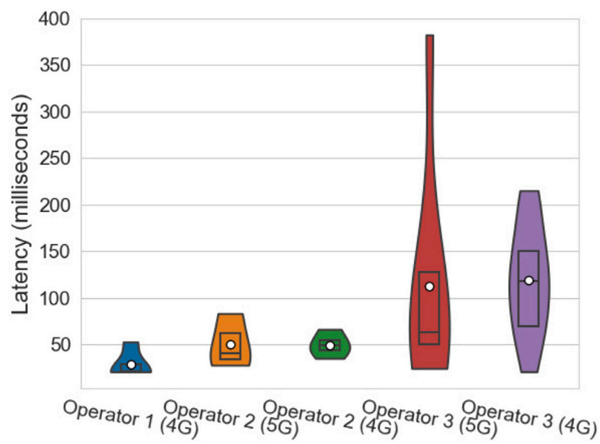


Fig. 4. Latency distribution for different operators and network types.

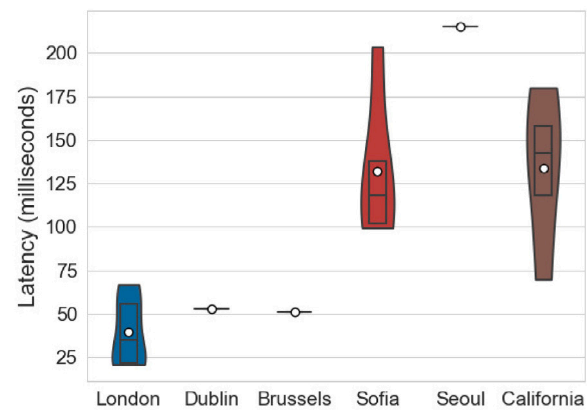


Fig. 5. QUIC latency distribution for different locations obtained for 4G LTE.

ranges (black rectangles), and the densities of measured latencies. We see that the mean values fall within the corresponding interquartile ranges for all estimates. These results imply that the mean values are not strongly affected by outliers and can serve as statistically reliable estimates of the average latencies. Note that we also measure latency per packet sampled by QUIC and use approximately 100 samples to estimate the average latency for each experiment.

Surprisingly, we observe the smallest average latency, which is about 25 ms, for the first operator for which only the 4G LTE network is available at the location where we made the measurements; the average 5G latency is about 110 ms and 59 ms for the second operator and third operator, respectively.

Observation 3. *When comparing different operators, the smallest average latency was achieved with 4G LTE, not 5G. This latency was at least half the minimum average latency observed in 5G networks.*

In Fig. 4, when comparing average latencies between 4G LTE and 5G networks from the same operator, we observe that in the case of *operator 3*, 5G latency is only 10% lower than that of 4G LTE. However, we observe that these latencies are identical for *operator 2*.

Observation 4. *When comparing measurements taken for the same operator, it was observed that the average latency of video downstreaming on 5G is either similar to or only 10% lower than the latency experienced on 4G LTE.*

We observe that the smallest median latency is also obtained for *operator 2*, which provides only 4G LTE service. However, the median latency, which is about 20 ms, is slightly lower than the average latency, i.e. 25 ms. The median latencies coincide with the average latencies for 4G LTE provided by *operator 2* and *operator 3*. These results imply that the downstreaming latencies are evenly distributed around the average values. We also see that the highest distribution of the latencies is obtained for both 4G LTE and 5G services provided by *operator 3*. Importantly, this operator uses the highest number of servers located outside the UK. According to Table 3, it appears that only 23% of the servers used by this operator for 4G LTE and 11% for 5G are located in the UK. Overall, there is a clear correlation between the geographic distribution of the servers used for downstreaming and latency. For example, the lowest latencies are observed for *operator 1* and *operator 2* in 4G LTE, which stream videos from servers located in the UK and Europe 75% and 25% of the time, respectively. Meanwhile, the highest latencies and variations are observed for *operator 3*, which uses servers distributed across the UK, Europe, the USA, and Asia for both 4G LTE and 5G.

Observation 5. *The highest distribution of downstream latencies is obtained for 4G LTE and 5G networks where the operators use servers distributed across several countries.*

To further investigate how latency changes with the location of servers, we construct latency distributions for each city where the servers are positioned. Fig. 5 illustrates latency distribution across cities in our experiments for 4G LTE. As anticipated, the downstreaming of the video originating from London exhibits the lowest average latency, approximately 35 ms. Interestingly, this latency is 1.4× higher than the average 4G LTE latency for *operator 1*, which uses servers located in London, UK, and Dublin, Ireland. These results imply that, in addition to the location, the major factor that affects latency is the 4G/5G ecosystem, i.e., base stations, internal network hardware and configuration, and Cloud/Edge servers.

We see that the average latency for servers located in Sofia and Mountain View (California) is about 135 ms, which is 3.85× higher than the average latency obtained in London. We also see that the variation of latencies measured for London servers is lower than those obtained for Sofia and Mountain View. Thus, we may conclude that using the servers located in London reduces the average latency and the variation of latencies.

Although the average latency for Dublin and Brussels is nearly twice as high as the average latency in London, the limited number of measurements obtained for these cities makes it difficult to quantify this difference accurately. There are also not enough measurements for Seoul servers with access latency exceeding 200 ms to draw meaningful conclusions.

Fig. 6 shows the 5G latency variation across different cities. Similar to the experiments with 4G LTE, we observe the lowest average latency, which is about 37 ms, for London. Moreover, the 5G average latency, which is 150 ms, is almost the same as the 4G latency measured for the servers located in Sofia. However, the 5G latency is only 75 ms, which is almost 2× smaller than the 4G latency measured for servers located in Mountain View. We attribute this to the superior network infrastructure employed by 5G for connecting servers in London and Mountain View. This also clarifies why the average 5G latency obtained for servers in Mountain View is 2× smaller than the 5G latency experienced by

[7] Unfortunately, there is a significant variation in our measurements, despite having approximately one hundred samples for each estimate. This variation is due to multiple factors, including 5G transmission obstacles, base station processing delays, data packet routing in networks, server delays, and base station load. To address this, we use interquartile ranges, represented as black rectangles. Specifically, these ranges encompass samples that fall between the first quartile (Q1) and the third quartile (Q3). Thus, any values outside these ranges are considered outliers.

Table 4
Max distance to achieve a specific latency.

Latency	Distance
<1 ms	<227 km
<10 ms	<2270 km
<20 ms	<4540 km
<40 ms	<9080 km

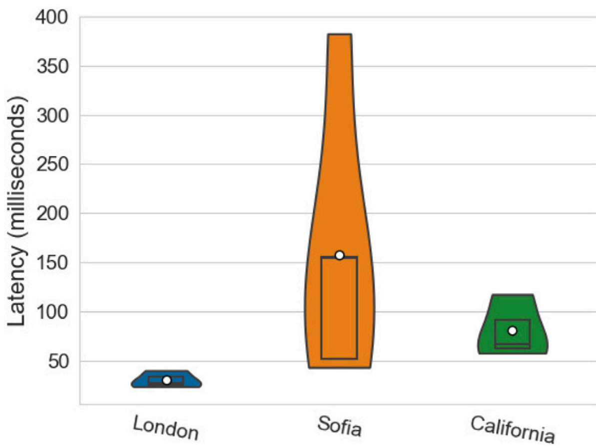


Fig. 6. QUIC latency distribution for different locations obtained for 5G Network.

Table 5
The minimum latency measured for operators.

Op1(4G)	Op2(4G)	Op2(5G)	Op3(4G)	Op3(5G)
20.7 ms	35.1 ms	27.8 ms	20.7 ms	24.3 ms

servers in Sofia. It is important to note that Sofia is only 2000 km from London, whereas Mountain View is 8631 km apart. Note that the average latency estimate for Sofia does not fall within the interquartile range, indicating that it is strongly affected by outliers due to the high variation in measurements. We attribute the disparity in latencies and the variation in measurements obtained for servers in Sofia by the networking and 4G/5G ecosystem.

We use the average latencies measured for different cities to project the maximum distance between servers and mobile devices to achieve a particular latency. To be more specific, we use the difference in the average latency obtained for London and Mountain View to estimate how the latency increases with distance as follows:

$$Lat\_per\_km = \frac{Lat_{Avr}^{MountainView} - Lat_{Avr}^{London}}{distance_{MountainView}^{London}} \tag{1}$$

Note that we specifically rely on the 5G measurements taken for Mountain View, as they represent the minimum latency increase per km. Thus, we project the maximum distance required to achieve a specific latency for the best-case scenario.

Table 4 shows the projected maximum distance between a server and mobile devices required to achieve a specific latency. As we can see, to provide a latency of 1 ms, the distance between servers and mobile devices should not exceed 227 km. Hence, a network of Edge servers, the distance between which does not exceed 227 km, is required to enable 5G applications, such as Autonomous driving vehicles (see Section 5), demanding 1 ms latency. Importantly, this projection does not include the latency introduced by mobile devices and base stations to handle data. Thus, we expect the maximum distances provided in Table 4 maybe even smaller.

Finally, Table 5 shows the minimum absolute latency that each operator can provide. We see that the minimum absolute latency of 20.7 ms is provided by *operator 1* and *operator 3* for 4G LTE. Remarkably, this latency is 17% lower than the minimum latency achieved on 5G by *operator 3*. Additionally, it outperforms the minimum latency of both 4G LTE and 5G provided by *operator 2* by 70% and 34%, respectively. Similar to our previous results, we attribute this difference to the current limitations within the state-of-the-art 4G/5G ecosystem.

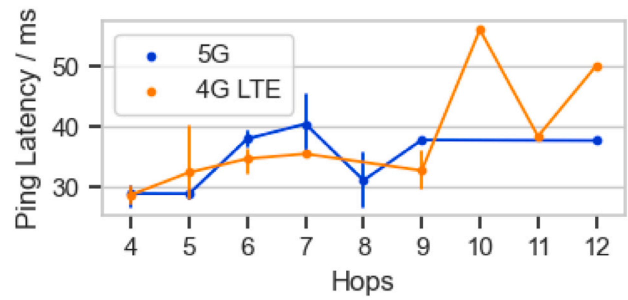


Fig. 7. The correlation between ping latency and number of hops.

Observation 6. *The average latency of the video downstreaming performed by servers located in London is 35 ms and 37 ms for 4G LTE and 5G, respectively.*

Observation 7. *The minimum absolute latency achieved on a mobile device in London is 20.7 ms and 24.3 ms for 4G LTE and 5G, respectively.*

Based on our experimental results, we identify the following major implications:

Implication 1. *In addition to the geographical location, the 4G/5G ecosystem can contribute to latency increases of up to 2 times.*

Implication 2. *Enabling 1 ms latency will require a network of Edge servers, the distance between should not exceed 227 km.*

The impact of routing on latency. The downstreaming latency is significantly impacted by routing and the number of hops. To quantify this effect, we conduct two correlation evaluations. In the first evaluation, we examine the correlation between the number of hops required to reach a server with a specific IP address and the downstream latency measured using *ping*. We perform a similar correlation in the second evaluation using latencies measured with QUIC.

Figs. 7 and 8 display the correlation between latency and the number of hops averaged over different servers and operators. Interestingly, we observe no strong correlation between the number of hops and latency for both *ping* and QUIC measurements. However, a clear pattern emerges: the average 4G LTE and 5G latencies are at their minimum for 4 hops in the *ping* correlation figure, while the minimum average latencies for 4G LTE and 5G are obtained for 5 hops in the QUIC correlation figure. Moreover, we find that the highest average 5G latency is measured for the servers that can be reached within 4 to 7 hops. These results also suggest that latency is primarily influenced by internal network hardware and configuration in the 4G/5G ecosystem.

4.3. Bandwidth analysis

Bandwidth. We measure the bandwidth and device current downstreaming the video with different resolutions in our next experiments. Similar to the previous experiments, we get the bandwidth measurements by parsing the QUIC traces. In our initial experiments, we measure the bandwidth using QUIC traces for *operator 2*, which provides 5G service with the minimum average downstream latency.

Fig. A.15 of Appendix shows how the bandwidth measured for 5G and 4G LTE changes when we downstream the video with 4K, 1080, 720, and 360 resolution. We clearly see that in all the cases, the video is downstreamed with some periods, and thus, we obtain the bandwidth

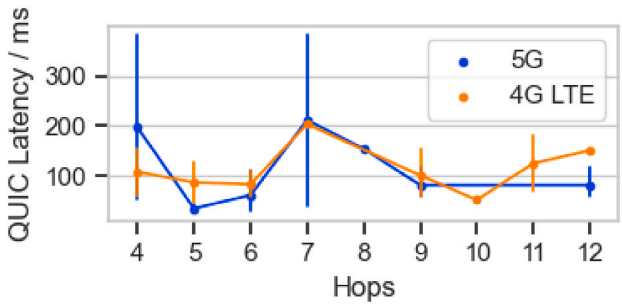


Fig. 8. The correlation between QUIC latency and number of hops.

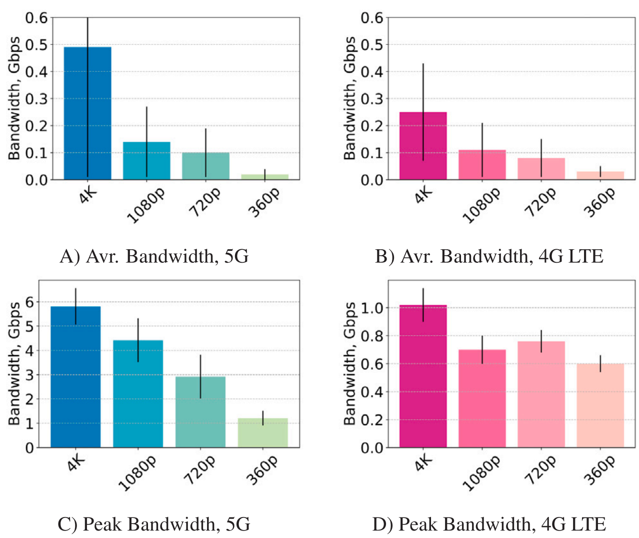
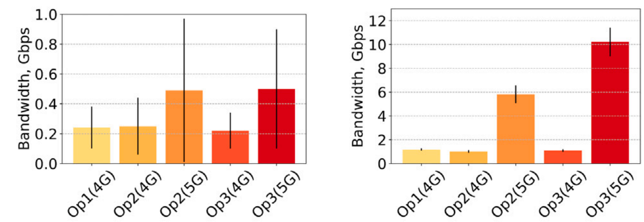
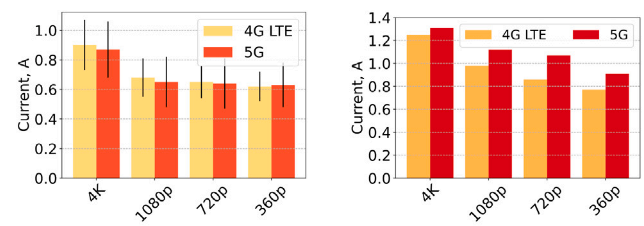


Fig. 9. Average and peak bandwidths for *operator 2*.



A) Average Bandwidth B) Peak Bandwidth

Fig. 10. Average and peak bandwidths across different operators.



A) Average Current B) Peak Current

Fig. 11. Device current for *operator 2*.

spikes. This is explained by the fact that YouTube streams the video in chunks, some of which are buffered in advance, allowing for a smoother playback experience. Our second observation indicates that, on average, the amplitude of bandwidth spikes decreases as the video resolution decreases.

We obtain these results for both 4G LTE and 5G. Nonetheless, the amplitude of spikes differs for 5G and 4G LTE. We observe that the bandwidth amplitude is lower for the 4G LTE measurements; however, as expected, the downstreaming periods are longer. To quantify this difference, we averaged the bandwidth for the spikes, i.e. the moments when the video is downstreamed by the YouTube application. Fig. 9(A and B) shows the average bandwidth with a 95% confidence interval for data transfer spikes when we downstream the video for 5G and 4G LTE for *operator 2*, respectively. We see that the average bandwidth grows with the video resolution and size, and it achieves up to 0.5 Gbps and 0.25 Gbps for 5G and 4G LTE, respectively. Note that we obtain a similar increasing trend in network bandwidth with increasing data size for the other operators.

Bandwidth Across Different Operators. Our experimental study also measures the bandwidth across different operators when downstreaming the 4K video. We employ the same approach as our previous experiments to measure the bandwidth: parsing the QUIC traces and estimating the bandwidth exclusively for downstream periods. Fig. 10(A) shows the average bandwidth with a 95% confidence interval measured for different operators. We observe that the average bandwidth reaches 0.24 Gbps and 0.5 Gbps on average for 4G LTE and 5G, respectively. Thus, we may conclude that the 5G bandwidth is 2× higher than the 4G LTE bandwidth on average.

Observation 8. *The average downstreaming bandwidth for 4G LTE and 5G networks typically range around 0.24 Gbps and 0.5 Gbps, respectively, when considering various operators.*

Peak Bandwidth: Importantly, in Fig. A.15 of Appendix, we see that the peak bandwidth can be much higher for both 5G and 4G LTE. To estimate the peak bandwidth, we parse the traces and sample the highest bandwidth for each period when the smartphone downstreams the videos. Fig. 9(C and D) shows the peak bandwidth with 95% confidence interval measured for 5G and 4G LTE provided by *operator 2* when downstreaming the videos with different resolutions. Like the average bandwidth, the peak bandwidth grows with the video resolution and size. However, the peak bandwidth achieves up to 5.81 Gpbs and 1 Gpbs for 5G and 4G LTE, respectively. Note that these values are higher than the maximum bandwidth, which can be achieved on mid-band base stations used in our experiments (see Section 3). We explain these results later.

We believe that the difference in bandwidth is because 5G uses base stations operating at 3.4 GHz, while 4G uses base stations operating mostly at 2.4 GHz (see Fig. 1). These findings can also be attributed to differences in channel bandwidth between 5G and 4G base stations. In 5G base stations used in our study, the channel bandwidth can extend up to 100 MHz, whereas in 4G base stations, the channel bandwidth reaches a maximum of 20 MHz [39,40]. Note that 5G can also benefit from other optimizations, such as OFDM [43]. Interestingly, we notice a consistent average bandwidth among various operators, whereas the average latency exhibits substantial variation across these operators. These findings suggest that latency is more responsive to the networking infrastructure.

If we compare the peak bandwidths across diferent operators (see Fig. 10, B), then we observe that the 5G bandwidth is nearly 10 times higher than the 4G LTE bandwidth for *operator 3*, whereas this difference is approximately 6 times for *operator 2*. The highest absolute peak bandwidth, reaching approximately 10 Gbps, is achieved through 5G services offered by *operator 3*. Meanwhile, the peak 4G LTE bandwidth is almost the same for all the operators, which is about 1 Gbps. Importantly, 1 Gbps and 10 Gbps are the maximum bandwidths which can be achieved for 4G LTE (Advanced) and 5G networks [44]. Nonetheless, the achieved peak bandwidths are higher than the maximum bandwidth, i.e., 0.9 Gbps, which can be achieved

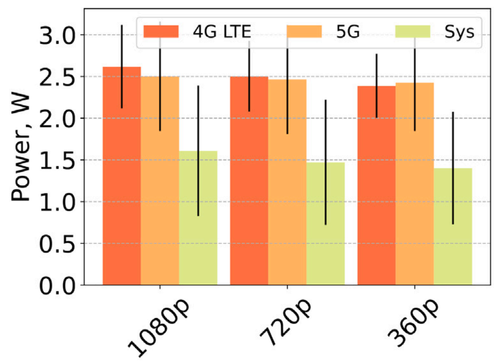


Fig. 12. The average energy consumption.

on mid-band base stations used in our experiments (see Section 3). We explain this byt the fact that the 4G/5G modem cannot instantaneously transfer the received data to the CPU since the CPU reads the data with some delay through a buffer. Thus, the CPU may fetch data that has been received previously, and the size of the fetched data may exceed the maximum size that can be received by the modem per second. These results imply that the mobile CPU may also increase latency.

4.4. Power and energy consumption analysis

When downstreaming the videos, we measure the mobile device current to quantify the system power. Fig. A.15 of Appendix shows the device current measured for each experiment with bandwidth for *operator 2*.

Our first observation is that, similarly to the experiments with bandwidth, we clearly obtain the device current spikes. Moreover, the amplitudes of the device's current spikes decrease when we reduce the video resolution. We explain the correlation between the bandwidth and device current spikes by the current required data transfer by the modem.

To investigate this, we averaged the current when downstreaming the video. Fig. 11(A) shows the average current with a 95% confidence interval measured for each downstreamed video. We notice that the average current increases by 32% for both 5G and 4G LTE when we down-stream the 4K video compared to 1080p, 720p, and 360p videos. Meanwhile, there is almost no difference between the average current for 1080p, 720p, and 360p videos. Note that the size of the video changes with the resolution; the sizes of 4K, 1080p, 720p, and 360p are 578Mb, 178Mb, 142Mb, and 28Mb, respectively. Thus, we expect the average current to grow with the transferred data size. To investigate this further, we measured the peak current for each experiment with bandwidth; see Fig. 11(B). Note that we measure the peak current by sampling the highest current values for each period when the video was downstreamed. We see that the peak current grows with the video resolution, i.e., the size of transferred data, for both 5G and 4G LTE. Moreover, apart from the average current, we see that the peak current is 15% higher on average for 5G compared to 4G LTE.

Observation 9. *On average, 5G enhances the peak current by approximately 15% in comparison to 4G LTE.*

This observation aligns well with the observations made in a previous work paper [21]. To explore the power consumption of a mobile device while streaming videos, we measure current and voltage while playing downloaded video content. We use the YouTube downloading option to do this experiment. Importantly, a video with 4K resolution cannot be downloaded and played without downstreaming, so we do experiments only for 1080p, 720p, and 360p videos. Fig. 12 shows the average power with a 95% confidence interval estimated using measured current and voltage; in this figure, *Sys* corresponds to the power incurred when playing the videos without downstreaming, i.e., the power incurred by the mobile system without the power induced by downstreaming.

We observe that the downstreaming significantly increases the mobile battery power by 68% on average for both 5G and 4G LTE. These results also imply that the 4G/5G modem consumes 40% of the total mobile energy on average when downstreaming the data. Thus, the modem receiving the data is one of the biggest contributors to the total mobile energy among screen, CPU, GPU, and memory [22].

Observation 10. *The modem downstreaming the data increases the mobile system power by 68% and consumes 40% of the total mobile energy.*

Interestingly, no significant disparity is observed in terms of power and average current between 5G and 4G LTE downstream. However, 5G radio should consume more power due to more powerful baseband and RF hardware [21,45]. Nonetheless, our experiments' average power for 5G and 4G LTE was the same. We explain this by a higher bandwidth incurred by 5G, which implies that the modem spends more time in an idle state without transferring the data when using 5G [21]. In Figs. A.15(A and I) of Appendix, it is evident that 5G offers a higher bandwidth in comparison to 4G LTE. However, it is worth noting that 5G transmits data in shorter bursts, while 4G LTE maintains a constant flow of data to the modem despite having a lower bandwidth. The absence of distinction can also be attributed to mobile operators utilizing 5G Non-Standalone (NSA), which leverages 4G LTE for command data transfer.

Observation 11. *On average, the power incurred when downstreaming data is comparable for 5G and 4G LTE.*

5. Discussion and summary

Many services can benefit from 5G and Edge computing; however, these services have different requirements. For example, Table 6 shows the list of services which are most demanding to latency and bandwidth.

One of the prominent applications for 5G and Edge computing is autonomous driving vehicles [46]. Autonomous driving vehicles should have low latency and high bandwidth connection to enable teleoperation, which will be implemented in nearly all commercial autonomous vehicles. It is required for remote supervision, assistance, and direct operation [46]. However, teleoperation requires a latency of 1 ms, which is 20× lower than the minimum average latency obtained in our experiments. Thus, the existing 4G/5G cannot facilitate a driverless fleet operation in London.

However, we note that autonomous driving and teleoperation require the use of specific wireless networks, such as Ultra-Reliable Low Latency Communications (URLLC) [47] and Massive Machine-Type Communications (mMTC) [48]. URLLC and mMTC are also essential for remote medical services, such as telesurgery [49] and emergency diagnostics [50], as well as for industrial automation (Industry 4.0 [51, 52]), including real-time robot control [53], autonomous robotics [51], and safety and control systems [52]. All these applications demand latencies below 10 ms. Importantly, URLLC and mMTC networks are distinct from the standard 5G networks used by mobile operators, namely Enhanced Mobile Broadband (eMBB). Both URLLC and mMTC are specifically optimized to reduce latency. For example, these technologies implement lightweight protocols that optimize the scheduling of data transfers through smaller time slots, prioritization, and preemption [47,48]. Nonetheless, to the best of our knowledge, there are currently no production deployments of URLLC and mMTC in the UK.

Another potential application is Augmented/Virtual Reality (AR/VR) technology, which has not been adopted at scale [55]. AR/VR promises

Table 6
5G/Edge computing applications.

Application	Latency	Bandwidth
Mobile cloud gaming [54]	<40 ms	>45 Mbps
Autonomous driving/teleoperation [46]	1 ms	>100 Mbps
AR/VR (3D rendering) [55]	<20 ms	>100 Mbps
Unmanned Aerial Vehicles (UAV) [56]	50 ms	50 Mbps

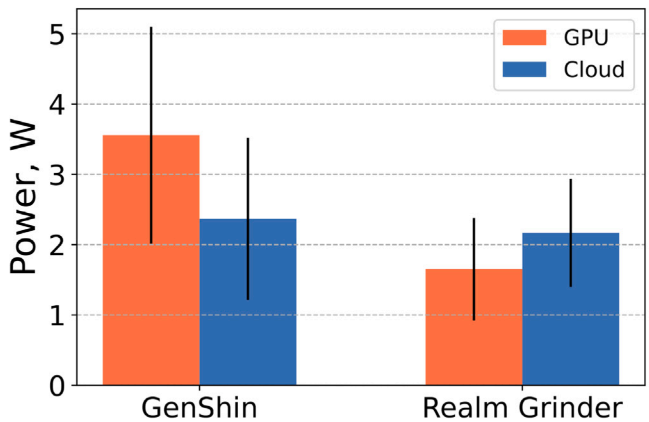


Fig. 13. The average power measured for mobile games running rendering on Cloud and local GPU.

to enable real-time remote collaboration, guided maintenance, and online education to a new level. This technology often uses Edge and Cloud for remote rendering. For example, Table 6 shows the latency and bandwidth requirements, which are 20 ms and 100 Mbps, respectively, for remote 3D rendering used by AR/VR. Thus, our experimental study suggests that the London 4G/5G ecosystem cannot effectively support the implementation of AR/VR technologies, specifically 3D rendering, since the minimum latency obtained in our experiments is 20.71 ms.

Unmanned Aerial Vehicles (UAV) are another potential promising direction which needs 5G and Edge computing. Amazon started Prime Air delivery in the United States [57]. To ensure effective operational control of UAVs, 3GPP has established specific requirements for command and control [58]. These requirements include data rates of up to 100 Kbps and a latency bound of 50 ms (see Table 6). Meanwhile, use cases involving flying cameras and remote surveillance rely on UAVs to transmit real-time telemetry data, pictures, or videos. The primary connectivity requirement for such data communication is the data rate, which can reach up to 50 Mbps [58]. The 4G LTE and 5G networks tested in our experiments can meet all these requirements.

Mobile cloud gaming is one of the most promising applications of 5G networks for mobile devices [54,59,60]. The main idea behind this approach is to offload the GPU graphical pipeline in the cloud for remote rendering and all the frames downstream. The network latency and bandwidth should satisfy specific criteria to enable high-quality gaming. For example, Nvidia GeForce nowadays demands a latency of less than 40 ms and bandwidth higher than 45 Mbps for the best experience, i.e., 3840x2160 pixel resolution at 120 FPS [54]. Although the 4G/5G ecosystem tested in our experiments meets these criteria, downstreaming significantly increases the mobile system power, as we demonstrated earlier. To investigate this further, we run 2 different games using the Cloud gaming service provided by *NVidia Now*. Specifically, we use popular *GenShin Impact* and *Realm Grinder* games. Fig. 13 shows the average power with a 95% confidence interval incurred when running the games on *NVidia Now Cloud* and mobile GPU. Note that *NVidia Now* uses Edge servers placed in London, where the latency varies from 37 ms to 145 ms. We observe that the GPU version of *GenShin Impact* uses 50% more power than the Cloud version. However, for *Realm Grinder* we observe opposite results since the Cloud version consumes 31% more power than the GPU version. Such a difference is explained by the fact that *GenShin Impact* is a social simulation game that uses the maximum capacity of GPU. In contrast, *Realm Grinder* is an RPG game that does not use GPU heavily. Note that *GenShin Impact* is one of the most demanding mobile games that use the OpenGLES interface, while many games these days use Vulcan, which requires significantly less power [61]. Based on these results, we may conclude that offloading GPU workloads into Cloud/Edge may be inefficient. Moreover, using Cloud and Edge facilities for other applications, such as AR/VR and Unmanned Aerial Vehicles, may be inefficient. Thus, apart from bandwidth and latency, the energy efficiency of offloading computing in Cloud/Edge should be considered an important challenge for future 5G applications, especially for those applications where energy consumption is critical.

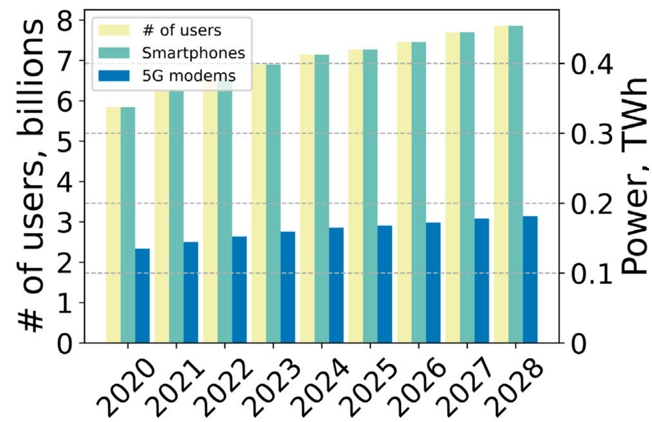


Fig. 14. The projected worldwide energy consumption.

Implication 3. *Offloading computation to Cloud/Edge presents a significant energy consumption challenge, as local computing units can prove more energy-efficient than transmitting computation results.*

5.1. Carbon emissions

Next, we examine global energy consumption and carbon emissions due to down-streaming. Expanding the utilization of Cloud/Edge technology for downstreaming on mobile devices may also inadvertently increase global carbon emissions. Subsequently, we present a projection of the worldwide energy consumption and smartphone emissions downstreaming YouTube videos. There are 6592 million smartphone users today, which grows yearly. It is projected that this number will achieve 7861 million users by 2028 [62] (see Fig. 14, left y-axis). It is also estimated that each smartphone user spends 23.1 h per month watching YouTube on average [63]. Assuming that each user watches YouTube videos at 720p resolution, we can estimate the global energy consumption due to watching YouTube on smartphones using the average Pixel power consumption measured in our experiments for the 720p video, the average time spent watching YouTube per month and the number of smartphone users. Fig. 14 (right y-axis) shows the total global energy spent by smartphones (*Smartphones*) and 5G modems (*5G modems*) when downstreaming YouTube videos at 720p resolution. Note the total power estimated will be almost the same for 360p resolution videos according to the power measurement taken for the mobile device used in our experiments (see Fig. 10). In these projections, we assume that the average power consumption of smartphones, including modem power, is equivalent to that of the Google Pixel 4a used in our experiments.

We see that smartphone users spent an average of almost 0.4 TWh per month in 2022 on watching YouTube videos, while this number will exceed 0.45 TWh by 2028. Note that almost 40% of this energy, which is about 0.15 TWh in 2022, is spent by mobile modems to receive the data from Cloud/Energy. As per the United States Environmental Protection Agency (EPA), the utilization of 0.4 TWh and 0.15 TWh corresponds to emissions equivalent to 73038 tons and 64889 tons of carbon dioxide (CO_2), respectively [64]. Moreover, 0.43 TWh is the amount of energy produced per month by R.E. Ginna reactor [65]. Thus, the infrastructure equivalent to an entire nuclear power plant must only facilitate global mobile YouTube downstreaming and smartphone watching. Moreover, we projected only the energy consumption for smartphones. However, the total amount of energy spent on streaming videos is much higher due to base stations and Cloud/Edge data centers, as previous studies show [66]. Hence, the worldwide energy consumption resulting from mobile device usage for streaming YouTube videos and accessing other online services and its impact on global CO_2 emissions is significantly greater. Thus, when developing 4G/5G applications that utilize Cloud and Edge technologies, putting energy efficiency first is crucial for managing energy consumption and mitigating global carbon emissions.

5.2. Summary

Our study reveals that the 5G ecosystem is currently in its early implementation stage. Although 5G offers a bandwidth two times higher than 4G LTE, it is worth noting that certain operators still demonstrate lower latency with their 4G LTE networks compared to 5G. This is in some way in line with findings from other recent work conducted in different countries such as the USA [67], Finland [68] and private 5G testbeds [69]. We attribute this disparity to the inadequacy of the 5G ecosystem in London, including base stations, Cloud/Edge servers, and interconnection networks, which are not yet prepared to deliver low-latency data transfers. Unfortunately, understanding the root causes of this issue requires access to and profiling of private Google and operator networks — an option that is not available to us. We assume that the observed disparity is primarily due to the fact that 5G base stations and networking hardware require substantial investment, as well as time for installation. To be more specific, significant resources are needed for: (i) setting up base stations, which requires approval from local governments, a process that may take months or even years; (ii) deploying base stations, which includes 5G and telecommunications hardware, power infrastructure, and labor, all of which are costly; and (iii) testing, integrating, and optimizing new base stations within the network. All of these factors prevent operators from fully enabling 5G capabilities in a rapid way. In addition, as we mentioned previously, all three operators use 5G NSA (Non-Standalone), which relies on 4G networks to transmit control information, leading to increased latency.

Edge offloading and orchestration: Given all these limitations and variations in 5G network quality, dynamic Edge service offloading based on latency, bandwidth, load of Edge servers, or energy consumption become vital optimizations to facilitate the efficient use of 5G networks [70–75]. However, the most challenging aspect of such optimizations is considering all the possible factors that can affect latency, bandwidth, the total energy consumption, and other quality-of-service parameters to predict the best offloading policy [76]. Thus, it is important to use the most advanced ML techniques to address this problem [76,77]. Meanwhile, mobile and datacenter operators need to focus specifically on optimizing Edge/Cloud orchestration [78], proactive load balancing [79] and Edge-aware caching technologies [80].

5G network and Edge Modeling: Although we do not have access to private Google and operator networks, one possible approach to understanding the 5G ecosystem's network and hardware configuration is through simulation. By conducting simulations with different parameter settings, we can identify a highly likely configuration for each operator. To be more specific, such a simulation might include different types of antennas, such as MIMO [81], beamforming antennas [82], and Reconfigurable Intelligent Surfaces (RIS) [83], as well as base stations at the system level [84,85], operating under various weather [86,87] and urban conditions [88]. The simulation might also account for inter-cell interference [89] and congestion arising in high-load usage scenarios [90–92]. Additionally, it is important to simulate Cloud and Edge network traffic originating from base stations [84,93]. Overall, to understand the configuration of existing commercial 4G LTE and 5G ecosystems, it is crucial to develop a fully integrated system-level simulation that incorporates all the types described above. Although this is quite a challenging task, it will enable researchers to identify bottlenecks in the current ecosystem and propose effective optimizations.

High-band base stations and mmWaves: To achieve ultra-low latency and average bandwidth close to 10 Gbps, it is essential to deploy high-band base stations and cells supporting mmWaves throughout London. This endeavour necessitates substantial investment because mmWaves base stations have significantly shorter effective ranges compared to mid-band stations [20]. Furthermore, as demonstrated in this study, achieving a latency of 1 ms will necessitate a network of Edge servers positioned no farther than 227 km. Moreover, recent studies [94] have showed that the bottlenecks are not limited to the radio latency but is also attributed to protocol (the latency for protocol mechanisms and configurations) [95] and processing (the latency for decision-making and data processing) [96].

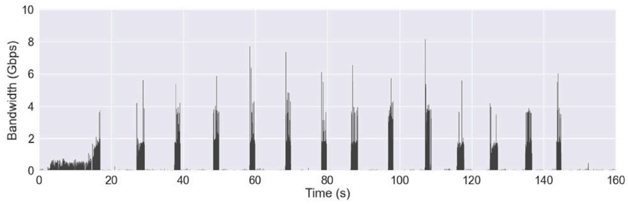
Energy efficiency: Another concern revolves around the energy efficiency of mobile devices and 4G/5G modems, rendering certain services impractical, such as Mobile Cloud gaming or remote AR 3D rendering, which heavily rely on offloading computing to Cloud or Edge. Therefore, aside from enhancing the 5G ecosystem and implementing 6G standards, it is imperative to prioritize optimizing the energy efficiency of mobile devices, especially 4G/5G modems. This task is highly significant as the number of devices utilizing mobile 5G networks continues to grow, and the escalating mobile Internet traffic is projected to reach 329 Exabytes per month by 2028 [11].

6. Limitation of our study

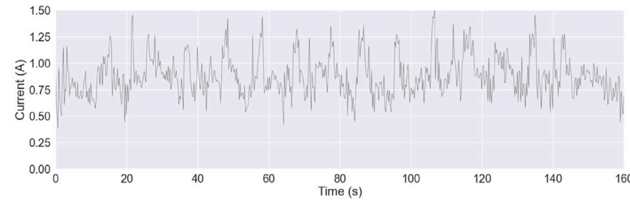
Usage and mobility scenarios: In our experiments, we use a stationary position for the experiments. It is important to acknowledge that prior research has established a correlation between the separation distance of a smartphone from base stations and degradation of latency and bandwidth [21,23]. The impact of buildings, tunnels, cars bodies and atmospheric even conditions further influences these metrics [13–16].

The quality of 4G/5G connections can also be significantly affected by mobility [23,97–102]. Specifically, it can be impacted by frequency shifts due to high speeds, which degrade signal quality [101,103], as well as by handovers [97,101]. For example, the duration of a handover in 5G NSA can be as high as 167 ms [97]. In addition to these factors, the quality of 5G networking can be degraded by inter-cell interference [89] and congestion [90–92], particularly when too many mobile devices are served by a single base station. Nevertheless, it is crucial to emphasize that our study's primary objective revolved around understanding the minimum latency and maximum bandwidth that can be achieved when the distance between the smartphone, base stations and Edge servers is minimized.

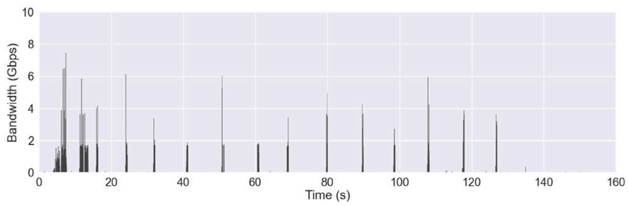
Upstreaming vs downstreaming: In our study, we exclusively focus on downstreaming scenarios, even though mobile network performance varies between downstreaming and upstreaming activities. Past research has shown that the up-link bandwidth is notably lower than the down-link bandwidth, while energy consumption remains relatively consistent [23]. Meanwhile, our study aims to comprehend the attainable maximum performance of real-world applications on commercial mobile 5G networks.



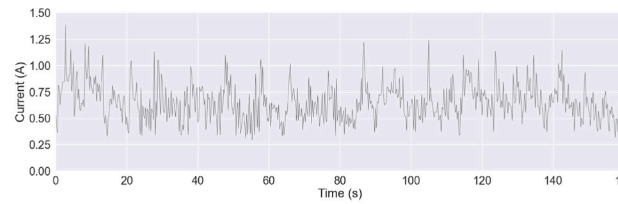
A. Bandwidth, 4K resolution, 5G



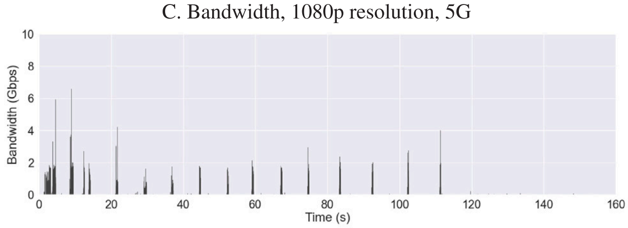
B. Current, 4K resolution, 5G



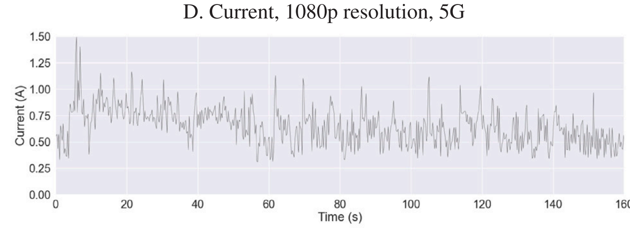
C. Bandwidth, 1080p resolution, 5G



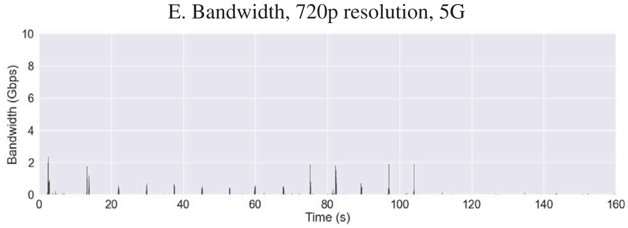
D. Current, 1080p resolution, 5G



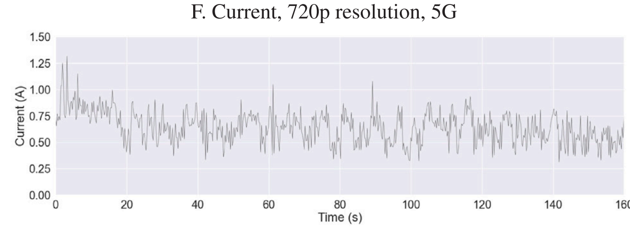
E. Bandwidth, 720p resolution, 5G



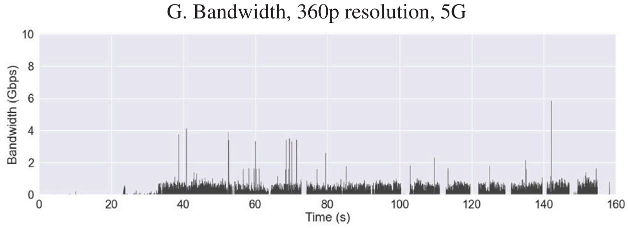
F. Current, 720p resolution, 5G



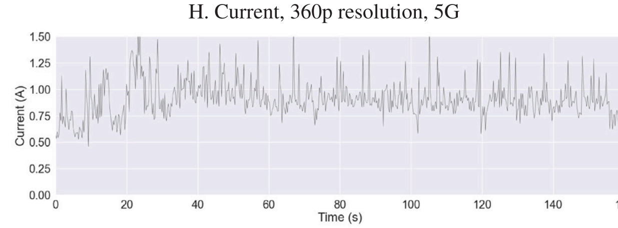
G. Bandwidth, 360p resolution, 5G



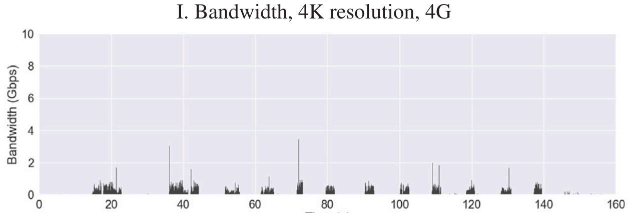
H. Current, 360p resolution, 5G



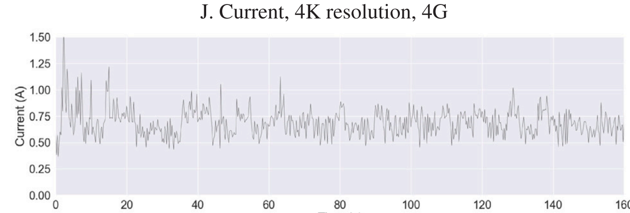
I. Bandwidth, 4K resolution, 4G



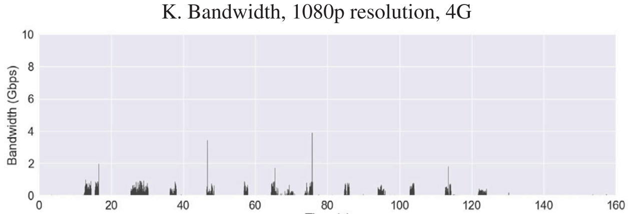
J. Current, 4K resolution, 4G



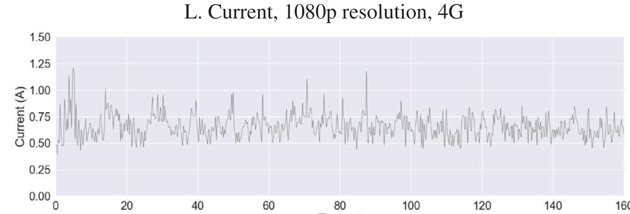
K. Bandwidth, 1080p resolution, 4G



L. Current, 1080p resolution, 4G



M. Bandwidth, 720p resolution, 4G



N. Current, 720p resolution, 4G

Fig. A.15. Bandwidth and current measurements for *operator 2*.

Reliability: One of the key requirements for emerging 5G applications, such as Autonomous Vehicles and Telesurgery [104,105] is reliability. Unfortunately, in our study, we could not measure the packet loss rate since QUIC encrypts this information, which we could not decrypt. Nonetheless, a prior study revealed that when utilizing UDP for data transfer, the packet loss rates can reach up to 4%

for 5G and 0.9% for 4G LTE, particularly under conditions of high bandwidth [21]. Note that 5G URLLC (Ultra Reliable and Low Latency Communications) used for Autonomous Driving Vehicles requires 0.0001% packet error rate. Thus, apart from latency, 5G reliability also prevents a driverless fleet operation in London.

Load of base stations: In our study, we assume a uniform distribution of load across base stations from different operators. This assumption is based on the following: (i) We have tested latency and bandwidth for three major operators each of which has an almost equal market share and, presumably, a similar number of customers. (ii) Our experimental study was conducted in an area where all operators have base stations within a 500-meter radius. (iii) The area is on the border of London and is not highly populated, implying that the load on base stations should not be high. Nonetheless, we acknowledge that this is an assumption, and the real load distribution may differ.

The experimental SoC: In our study, we are conducting experiments with one particular device, i.e. Google Pixel 4a. However, the device features the widely used Snapdragon 765G 5G SoC, which has been also implemented in OnePlus Nord, Nokia 8.3 5G, TCL 10 5G, ZTE Axon 11 5G, Motorola Edge (2020), LG Wing 5G, Samsung Galaxy A71 5G. We expect similar measurements to be obtained on devices from other vendors that use this processor. Notably, over 10 million units of the OnePlus Nord alone have been sold globally [106]. Thus, our experiments are representative of a broad range of mobile devices. Nonetheless, it is important to investigate variations in latency, bandwidth, and energy consumption across various 5G SoCs, including different SoC generations, in future research studies.

7. Conclusion and future work

This paper presents the findings of our experimental study conducted in London, which aims to understand what latency and bandwidth for 4G LTE and 5G can be achieved on a real commercial device using Edge servers in practice. Our study reveals that despite the 5G networks providing an average bandwidth up to 2× higher than 4G LTE, the 4G LTE networks in London often exhibit lower latency than their 5G counterparts. We explain these results by the inadequacy of the early stage 5G ecosystem in London, including base stations, Cloud/Edge servers, and wired/fiber networks.

Our study demonstrates that the 4G/5G ecosystem can contribute to latency increases up to 2 times. Thus, the 4G/5G ecosystem is a major bottleneck which prevents exploiting the full capabilities of 4G LTE and 5G networks. Additionally, our study uncovers a critical issue with the energy consumption of mobile devices, particularly 4G/5G modems, which contribute to 40% of the total mobile energy consumption. We demonstrate that high latency and significant energy consumption pose obstacles to leveraging 5G for crucial applications like Cloud gaming and Autonomous driving/teleoperation. We firmly believe that to fully unlock 5G's potential, efforts aimed at improving the 5G ecosystem and enhancing mobile device energy efficiency must be prioritized.

CRediT authorship contribution statement

Peixuan Song: Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **JunKyu Lee:** Writing – review & editing, Validation, Investigation, Formal analysis. **Ahmed M. Abdelmoniem:** Writing – review & editing, Validation, Supervision, Investigation, Formal analysis, Conceptualization. **Lev Mukhanov:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

See Fig. A.15.

Data availability

No data was used for the research described in the article.

References

[1] P. Song, J. Lee, A. M. Abdelmoniem, L. Mukhanov, Do 5G networks achieve the proclaimed promises? An empirical study using YouTube edge service, in: IEEE International Conference on Edge Computing and Communications, IEEE EDGE, 2024.

[2] A. Holst, Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2030, URL https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/.

[3] Cisco, Cisco visual networking index: Global mobile data traffic forecast update, 2018, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-indexvni/white_paper_c11-520862.pdf.

[4] IBM, What is edge computing?, 2020, URL https://www.ibm.com/cloud/what-is-edge-computing.

[5] J. Hsu, How YouTube led to google's cloud-gaming service: The tech that made YouTube work everywhere promises to do the same for games - [news], IEEE Spectr. 56 (09) (2019) 9–10.

[6] Thales, 5G technology and networks (speed, use cases, rollout), 2022, URL https://www.thalesgroup.com/en/markets/digital-identity-and-security/mobile/inspired/5G.

[7] J. Sheperd, 23 essential YouTube statistics you need to know in 2023, 2023, URL https://thesocialshepherd.com/blog/youtube-statistics#:~:text=YouTube%20has%202.1%20billion%20monthly,122%20million%20users%20per%20day.

[8] G. Nardini, G. Stea, A. Virdis, Scalable real-time emulation of 5G networks with Simu5G, IEEE Access 9 (2021) 148504–148520, http://dx.doi.org/10.1109/ACCESS.2021.3123873.

[9] A.M. Maia, Y. Ghamri-Doudane, D. Vieira, M. Franklin de Castro, Dynamic service placement and load distribution in edge computing, in: 2020 16th International Conference on Network and Service Management, CNSM, 2020, pp. 1–9, http://dx.doi.org/10.23919/CNSM50824.2020.9269059.

[10] T. Ouyang, Z. Zhou, X. Chen, Follow me at the edge: Mobility-aware dynamic service placement for mobile edge computing, IEEE J. Sel. Areas Commun. 36 (10) (2018) 2333–2345, http://dx.doi.org/10.1109/JSAC.2018.2869954.

[11] Ericsson, Mobile data traffic outlook, 2022, URL https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast.

[12] K. Lee, D. Chu, E. Cuervo, J. Kopf, Y. Degtyarev, S. Grizan, A. Wolman, J. Flinn, Outatime: Using speculation to enable low-latency continuous interaction for mobile cloud gaming, in: Proceedings of the ACM MobiSys, 2015.

[13] S. Mohebi, F. Michelinakis, A. Elmokashfi, O. Grøndalen, K. Mahmood, A. Zanella, Sectors, beams and environmental impact on the performance of commercial 5G mmwave cells: An empirical study, IEEE Access 10 (2022) 133309–133323.

[14] D. Pimienta-del Valle, L. Mendo, J.M. Riera, P. Garcia-del Pino, Indoor LOS propagation measurements and modeling at 26, 32, and 39 GHz millimeter-wave frequency bands, Electronics 9 (11) (2020).

[15] M. Khalily, S. Taheri, S. Payami, M. Ghoraishi, R. Tafazolli, Indoor wideband directional millimeter wave channel measurements and analysis at 26 GHz, 32 GHz, and 39 GHz, Trans. Emerg. Telecommun. Technol. 29 (10) (2018).

[16] Z. Lin, X. Du, H.-H. Chen, B. Ai, Z. Chen, D. Wu, Millimeter-wave propagation modeling and measurements for 5G mobile networks, IEEE Wirel. Commun. 26 (1) (2019) 72–77.

[17] X. Zhao, S. Li, Q. Wang, M. Wang, S. Sun, W. Hong, Channel measurements, modeling, simulation and validation at 32 GHz in outdoor microcells for 5G radio systems, IEEE Access 5 (2017) 1062–1072.

[18] J. Ko, Y.-J. Cho, S. Hur, T. Kim, J. Park, A.F. Molisch, K. Haneda, M. Peter, D.-J. Park, D.-H. Cho, Millimeter-wave channel measurements and analysis for statistical spatial channel model in in-building and urban environments at 28 GHz, IEEE Trans. Wirel. Commun. 16 (9) (2017) 5853–5868.

[19] S. Hur, H. Yu, J. Park, W. Roh, C.U. Bas, R. Wang, A.F. Molisch, Feasibility of mobility for millimeter-wave systems based on channel measurements, IEEE Commun. Mag. 56 (7) (2018).

[20] O. Al-Saadeh, G. Wikstrom, J. Sachs, I. Thibault, D. Lister, End-to-end latency and reliability performance of 5G in London, in: 2018 IEEE Global Communications Conference, GLOBECOM, 2018, pp. 1–6.

[21] D. Xu, A. Zhou, X. Zhang, G. Wang, X. Liu, C. An, Y. Shi, L. Liu, H. Ma, Understanding operational 5G: A first measurement study on its coverage, performance and energy consumption, in: Proceedings of ACM SIGCOMM, 2020, pp. 479–494.

[22] A. Narayanan, E. Ramadan, J. Carpenter, Q. Liu, Y. Liu, F. Qian, Z.-L. Zhang, A first look at commercial 5G performance on smartphones, in: Proceedings of the Web Conference, 2020.
[23] A. Narayanan, X. Zhang, R. Zhu, A. Hassan, S. Jin, X. Zhu, X. Zhang, D. Rybkin, Z. Yang, Z.M. Mao, F. Qian, Z.-L. Zhang, A variegated look at 5G in the wild: Performance, power, and QoE implications, in: Proceedings of the ACM SIGCOMM, 2021, pp. 610–625.
[24] A. Narayanan, E. Ramadan, R. Mehta, X. Hu, Q. Liu, R.A.K. Fezeu, U.K. Dayalan, S. Verma, P. Ji, T. Li, F. Qian, Z.-L. Zhang, Lumos5G: Mapping and predicting commercial MmWave 5G throughput, in: Proceedings of the ACM Internet Measurement Conference, IMC, 2020.
[25] scaleyourapp.com, YouTube database - How does it store so many videos without running out of storgae space?, URL https://scaleyourapp.com/youtube-database-how-does-it-store-so-many-videos-without-running-out-of-storage-space/.
[26] N. Hopper, E.Y. Vasserman, E. Chan-TIN, How much anonymity does network latency leak? ACM Trans. Inf. Syst. Secur. 13 (2) (2010).
[27] T. Gopinath, A.R. Kumar, R. Sharma, Performance evaluation of TCP and UDP over wireless ad-hoc networks with varying traffic loads, in: 2013 International Conference on Communication Systems and Network Technologies, 2013, pp. 281–285.
[28] A. Langley, A. Riddoch, A. Wilk, A. Vicente, C. Krasic, D. Zhang, F. Yang, F. Kouranov, I. Swett, J. Iyengar, J. Bailey, J. Dorfman, J. Roskind, J. Kulik, P. Westin, R. Tenneti, R. Shade, R. Hamilton, V. Vasiliev, W.-T. Chang, Z. Shi, The QUIC transport protocol: Design and internet-scale deployment, in: Proceedings of the ACM SIGCOMM, 2017.
[29] The Tcpdump Group, Tcpdump, 2022, https://www.tcpdump.org/.
[30] S. Wang, D. Xu, S. Yan, Analysis and application of wireshark in TCP/IP protocol teaching, in: International Conference on E-Health Networking Digital Ecosystems and Technologies, EDT, 2010.
[31] Perfetto, Perfetto - System profiling, app tracing and trace analysis, URL https://perfetto.dev/docs/.
[32] Android, Android Health, URL https://source.android.com/devices/tech/health.
[33] I. Livadariu, T. Dreibholz, A.S. Al-Selwi, H. Bryhni, O. Lysne, S. Bjørnstad, A. Elmokashfi, On the accuracy of country-level IP geolocation, in: Proceedings of Applied Networking Research Workshop, 2020.
[34] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, C. Papadopoulos, A look at router geolocation in public and commercial databases, in: Proceedings of ACM IMC, 2017.
[35] B. Gueye, S. Uhlig, S. Fdida, Investigating the imprecision of IP block-based geolocation, in: Passive and Active Network Measurement, 2007.
[36] I. Poese, S. Uhlig, M.A. Kaafar, B. Donnet, B. Gueye, IP geolocation databases: Unreliable? SIGCOMM Comput. Commun. Rev. 41 (2) (2011) 53–56.
[37] Nokia, 5G spectrum bands explained - low, mid and high bands, 2020, URL https://www.nokia.com/thought-leadership/articles/spectrum-bands-5g-world/.
[38] B. Ronan McLaughlin, 5G low latency requirements, 2023, URL https://broadbandlibrary.com/5g-low-latency-requirements/.
[39] R. Dangi, P. Lalwani, G. Choudhary, I. You, G. Pau, Study and investigation on 5G technology: A systematic review, Sensors 22 (1) (2022).
[40] L. Malviya, R.K. Panigrahi, M.V. Kartikeyan, MIMO antennas with diversity and mutual coupling reduction techniques: a review, Int. J. Microw. Wirel. Technol. 9 (8) (2017) 1763–1780.
[41] Nybsys, Low to high 5G banhds explained, 2023, URL https://nybsys.com/5g-bands/.
[42] Google, Google's direct peering overview, 2025, URL https://cloud.google.com/network-connectivity/docs/direct-peering?utm_source=chatgpt.com.
[43] B. Farhang-Boroujeny, H. Moradi, OFDM inspired waveforms for 5G, IEEE Commun. Surv. I& Tutorials 18 (4) (2016) 2474–2492.
[44] Thales, 5G technology and networks (speed, use cases, rollout), 2023, URL https://www.thalesgroup.com/en/markets/digital-identity-and-security/mobile/inspired/5G.
[45] G. Wu, C. Yang, S. Li, G.Y. Li, Recent advances in energy-efficient networks and their application in 5G systems, IEEE Wirel. Commun. 22 (2) (2015) 145–151.
[46] DriveU.auto, Using 5G for teleoperation of autonomous vehicles, 2022, URL https://driveu.auto/blog/using-5g-for-teleoperation-of-autonomous-vehicles/.
[47] Z. Li, M.A. Uusitalo, H. Shariatmadari, B. Singh, 5G URLLC: Design challenges and system concepts, in: 2018 15th International Symposium on Wireless Communication Systems, ISWCS, 2018, pp. 1–6, http://dx.doi.org/10.1109/ISWCS.2018.8491078.
[48] F. Mehmeti, T.F. La Porta, Modeling and analysis of mMTC traffic in 5G base stations, in: IEEE 19th Annual Consumer Communications and Networking Conference, CCNC, 2022, pp. 652–660, http://dx.doi.org/10.1109/CCNC49033.2022.9700727.
[49] W. Tian, M. Fan, C. Zeng, Y. Liu, D. He, H. Zhang, Y. Sun, B. Liu, J. Li, W. Yuan, et al., Telerobotic spinal surgery based on 5G network: the first 12 cases, Neurospine 17 (1) (2020) 114–120, http://dx.doi.org/10.14245/ns.1938454.227, URL https://pubmed.ncbi.nlm.nih.gov/32252160/.
[50] F.N. Bäßmann, S.K. Lier, O. Werth, A. Krage, M.H. Breitner, Future rescue management: Design specifications for a 5G-enabled emergency response information system, Intell. Decis. Technol. 18 (2) (2024) 1157–1179, http://dx.doi.org/10.3233/IDT-230476.
[51] Ericsson, You need to see our dancing hexapod demo from MWC, 2019, URL https://www.ericsson.com/en/blog/2019/3/dancing-hexapod-demo-a-success-at-mwc.
[52] Ericsson, Achieving industrial automation protocols with 5G URLLC, 2019, URL https://www.ericsson.com/en/cases/2020/accelerate-factory-automation-with-5g-urllc.
[53] F. Voigtländer, A. Ramadan, J. Eichinger, C. Lenz, D. Pensky, A. Knoll, 5G for robotics: Ultra-low latency control of distributed robotic systems, in: 2017 International Symposium on Computer Science and Intelligent Controls, ISCSIC, 2017, pp. 69–72, http://dx.doi.org/10.1109/ISCSIC.2017.27.
[54] Nvidia, Geforce now. System requirements, 2020, URL https://www.nvidia.com/en-us/geforce-now/system-reqs/.
[55] M. Bamforth, 5G and AR/VR: Transformative Use Cases with Edge Computing, 2022, URL https://stlpartners.com/articles/edge-computing/5g-edge-ar-vr-use-cases/.
[56] X. Lin, V. Yajnanarayana, S.D. Muruganathan, S. Gao, H. Asplund, H.-L. Maattanen, M. Bergstrom, S. Euler, Y.-P.E. Wang, The sky is not the limit: LTE for unmanned aerial vehicles, IEEE Commun. Mag. 56 (4) (2018) 204–210.
[57] Amazon, Amazon prime air prepares for drone deliveries, 2022, URL https://www.aboutamazon.com/news/transportation/amazon-prime-air-prepares-for-drone-deliveries.
[58] H. Ji, S. Park, B. Shim, Sparse vector coding for ultra reliable and low latency communications, IEEE Trans. Wirel. Commun. 17 (10) (2018) 6693–6706.
[59] Xbox, Xbox cloud gaming (beta), 2020, URL https://www.xbox.com/en-GB/cloud-gaming.
[60] J. Bulman, P. Garraghan, A cloud gaming framework for dynamic graphical rendering towards achieving distributed game engines, in: Proceedings of the 12th USENIX Conference on Hot Topics in Cloud Computing, HotCloud '20, USENIX Association, USA, 2020.
[61] M. Lujan, M. McCrary, B.W. Ford, Z. Zong, Vulkan vs opengl ES: Performance and energy efficiency comparison on the big.little architecture, in: 2021 IEEE International Conference on Networking, Architecture and Storage, NAS, 2021, pp. 1–8.
[62] Statista, Number of smartphone mobile network subscriptions worldwide from 2016 to 2022, with forecasts from 2023 to 2028, 2022, URL https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.
[63] Statista, Monthly time spent on the YouTube mobile app per user in selected markets worldwide in 2022, 2022, URL https://www.statista.com/statistics/1287283/time-spent-youtube-app-selected-countries/.
[64] EPA, Greenhouse gas equivalencies calculator, 2023, URL https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator.
[65] AGI, How much electricity does a typical nuclear power plant generate?, 2022, URL https://www.americangeosciences.org/critical-issues/faq/how-much-electricity-does-typical-nuclear-power-plant-generate.
[66] G. Kamiya, The carbon footprint of streaming video: fact-checking the headlines, 2020, URL https://www.iea.org/commentaries/the-carbon-footprint-of-streaming-video-fact-checking-the-headlines.
[67] I. Khan, T.X. Tran, M. Hiltunen, T. Karagioules, D. Koutsonikolas, An experimental study of low-latency video streaming over 5G, 2024, arXiv2403.00752.
[68] M. Uitto, A. Heikkinen, Evaluation of live video streaming performance for low latency use cases in 5G, in: 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), 2021, pp. 431–436, http://dx.doi.org/10.1109/EuCNC/6GSummit51104.2021.9482605.
[69] M. Boeding, P. Scalise, M. Hempel, H. Sharif, J. Lopez, Toward wireless smart grid communications: An evaluation of protocol latencies in an open-source 5G testbed, Energies 17 (2) (2024).
[70] S. Wang, R. Urgaonkar, M. Zafer, T. He, K. Chan, K.K. Leung, Dynamic service migration in mobile edge computing based on Markov decision process, IEEE/ACM Trans. Netw. 27 (3) (2019) 1272–1288, http://dx.doi.org/10.1109/TNET.2019.2916577.
[71] M. Chen, W. Li, G. Fortino, Y. Hao, L. Hu, I. Humar, A dynamic service migration mechanism in edge cognitive computing, 19 (2), 2019, http://dx.doi.org/10.1145/3239565.
[72] A. Machen, S. Wang, K.K. Leung, B.J. Ko, T. Salonidis, Live service migration in mobile edge clouds, IEEE Wirel. Commun. 25 (1) (2018) 140–147, http://dx.doi.org/10.1109/MWC.2017.1700011.
[73] J. Wang, J. Hu, G. Min, Q. Ni, T. El-Ghazawi, Online service migration in mobile edge with incomplete system information: A deep recurrent actor-critic learning approach, IEEE Trans. Mob. Comput. 22 (11) (2023) 6663–6675, http://dx.doi.org/10.1109/TMC.2022.3197706.
[74] W. Ni, H. Tian, X. Lyu, S. Fan, Service-dependent task offloading for multiuser mobile edge computing system, Electron. Lett. 55 (15) (2019) 839–841.
[75] J. Wang, J. Hu, G. Min, A.Y. Zomaya, N. Georgalas, Fast adaptive task offloading in edge computing based on meta reinforcement learning, IEEE Trans. Parallel Distrib. Syst. 32 (1) (2021) 242–253, http://dx.doi.org/10.1109/TPDS.2020.3014896.
[76] J. Bi, H. Yuan, S. Duanmu, M. Zhou, A. Abusorrah, Energy-optimized partial computation offloading in mobile-edge computing with genetic simulated-annealing-based particle swarm optimization, IEEE Internet Things J. 8 (5) (2021) 3774–3785, http://dx.doi.org/10.1109/JIOT.2020.3024223.

[77] J. Wang, J. Hu, G. Min, W. Zhan, Q. Ni, N. Georgalas, Computation offloading in multi-access edge computing using a deep sequential model based on reinforcement learning, IEEE Commun. Mag. 57 (5) (2019) 64–69, http://dx.doi.org/10.1109/MCOM.2019.1800971.

[78] D. Haja, M. Szalay, B. Sonkoly, G. Pongracz, L. Toka, Sharpening kubernetes for the edge, in: Proceedings of the ACM SIGCOMM 2019 Conference Posters and Demos, in: SIGCOMM Posters and Demos '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 136–137, http://dx.doi.org/10.1145/3342280.3342335.

[79] Q.-M. Nguyen, L.-A. Phan, T. Kim, Load-balancing of kubernetes-based edge computing infrastructure using resource adaptive proxy, Sensors 22 (8) (2022) http://dx.doi.org/10.3390/s22082869, URL https://www.mdpi.com/1424-8220/22/8/2869.

[80] C.K. Kim, T. Kim, S. Lee, S. Lee, A. Cho, M.-S. Kim, Delay-aware distributed program caching for IoT-edge networks, PLOS ONE 17 (7) (2022) 1–20, http://dx.doi.org/10.1371/journal.pone.0270183, URL https://doi.org/10.1371/journal.pone.0270183.

[81] S. Sun, G.R. MacCartney, T.S. Rappaport, A novel millimeter-wave channel simulator and applications for 5g wireless communications, in: 2017 IEEE International Conference on Communications, ICC, 2017, pp. 1–7, http://dx.doi.org/10.1109/ICC.2017.7996792.

[82] T. Kuehne, X. Song, G. Caire, K. Rasilainen, T.H. Le, M. Rossi, I. Ndip, C. Fager, Performance simulation of a 5G hybrid beamforming millimeter-wave system, in: WSA 2020; 24th International ITG Workshop on Smart Antennas, 2020, pp. 1–6.

[83] R. Liu, J. Dou, P. Li, J. Wu, Y. Cui, Simulation and field trial results of reconfigurable intelligent surfaces in 5G networks, IEEE Access 10 (2022) 122786–122795, http://dx.doi.org/10.1109/ACCESS.2022.3223447.

[84] G. Nardini, D. Sabella, G. Stea, P. Thakkar, A. Virdis, Simu5G–an omnet++ library for end-to-end performance evaluation of 5G networks, IEEE Access 8 (2020) 181176–181191, http://dx.doi.org/10.1109/ACCESS.2020.3028550.

[85] Mathworks, Overview of 5G system-level simulation, 2024, URL https://uk.mathworks.com/help/5g/ug/overview-system-level-simulation-in-5g-network-analysis.html.

[86] Y. Golovachev, A. Mazor, G.A. Pinhasi, Y. Pinhasi, Effectiveness of varios 5G modulation techniques in different weather conditions, in: 2019 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems, COMCAS, 2019, pp. 1–3, http://dx.doi.org/10.1109/COMCAS44984.2019.8958042.

[87] R. Wang, Y. Yang, B. Makki, A. Shamim, A wideband reconfigurable intelligent surface for 5G millimeter-wave applications, IEEE Trans. Antennas and Propagation 72 (3) (2024) 2399–2410, http://dx.doi.org/10.1109/TAP.2024.3352828.

[88] P. Zhang, B. Yang, C. Yi, H. Wang, X. You, Measurement-based 5G millimeter-wave propagation characterization in vegetated suburban macrocell environments, IEEE Trans. Antennas and Propagation 68 (7) (2020) 5556–5567, http://dx.doi.org/10.1109/TAP.2020.2975365.

[89] H. Wei, N. Deng, M. Haenggi, Inter-cell interference coordination in millimeter-wave cellular networks, in: 2019 IEEE Global Communications Conference, GLOBECOM, 2019, pp. 1–6, http://dx.doi.org/10.1109/GLOBECOM38437.2019.9013992.

[90] R. Asim, L. Subramanian, Y. Zaki, Impact of congestion control on mixed reality applications, EMS '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 21–26, http://dx.doi.org/10.1145/3672196.3673395.

[91] P. Goyal, M. Alizadeh, H. Balakrishnan, Rethinking congestion control for cellular networks, in: Proceedings of the 16th ACM Workshop on Hot Topics in Networks, HotNets '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 29–35, http://dx.doi.org/10.1145/3152434.3152437.

[92] J.I. Sandoval, S. Céspedes, Performance evaluation of congestion control over B5G/6G fluctuating scenarios, in: Proceedings of the Int'L ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications, DIVANet '23, Association for Computing Machinery, New York, NY, USA, 2023, pp. 85–92, http://dx.doi.org/10.1145/3616392.3623408.

[93] N. Patriciello, S. Lagen, B. Bojovic, L. Giupponi, An E2E simulator for 5G NR networks, Simul. Model. Pr. Theory 96 (2019) 101933, http://dx.doi.org/10.1016/j.simpat.2019.101933, URL https://www.sciencedirect.com/science/article/pii/S1569190X19300589.

[94] A. Maghsoudnia, E. Vlad, A. Gong, D.M. Dumitriu, H. Hassanieh, Ultra-reliable low-latency in 5G: A close reality or a distant goal? in: Proceedings of the 23rd ACM Workshop on Hot Topics in Networks, HotNets '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 111–120, http://dx.doi.org/10.1145/3696348.3696862.

[95] T. Tojo, H. Baba, S. Yasukawa, Y. Okazaki, Micro-burst analysis and mitigation for 5G low-latency communication services, in: 2019 IEEE 2nd 5G World Forum (5GWF), 2019, pp. 232–236, http://dx.doi.org/10.1109/5GWF.2019.8911634.

[96] I. Parvez, A. Rahmati, I. Guvenc, A.I. Sarwat, H. Dai, A survey on low latency towards 5G: RAN, core network and caching solutions, IEEE Commun. Surv. & Tutorials 20 (4) (2018) 3098–3130, http://dx.doi.org/10.1109/COMST.2018.2841349.

[97] A. Hassan, A. Narayanan, A. Zhang, W. Ye, R. Zhu, S. Jin, J. Carpenter, Z.M. Mao, F. Qian, Z.-L. Zhang, Vivisecting mobility management in 5G cellular networks, in: Proceedings of the ACM SIGCOMM 2022 Conference, SIGCOMM '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 86–100, http://dx.doi.org/10.1145/3544216.3544217.

[98] H. Deng, C. Peng, A. Fida, J. Meng, Y.C. Hu, Mobility support in cellular networks: A measurement study on its configurations and implications, in: Proceedings of the Internet Measurement Conference 2018, IMC '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 147–160, http://dx.doi.org/10.1145/3278532.3278546.

[99] Y. Li, J. Xu, C. Peng, S. Lu, A first look at unstable mobility management in cellular networks, in: Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications, HotMobile '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 15–20, http://dx.doi.org/10.1145/2873587.2873599.

[100] Y. Li, H. Deng, J. Li, C. Peng, S. Lu, Instability in distributed mobility management: Revisiting configuration management in 3G/4G mobile networks, in: Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science, SIGMETRICS '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 261–272, http://dx.doi.org/10.1145/2896377.2901457.

[101] Y. Pan, R. Li, C. Xu, The first 5G-LTE comparative study in extreme mobility, Proc. ACM Meas. Anal. Comput. Syst. 6 (1) (2022) http://dx.doi.org/10.1145/3508040.

[102] Y. Li, H. Lin, Z. Li, Y. Liu, F. Qian, L. Gong, X. Xin, T. Xu, A nationwide study on cellular reliability: measurement, analysis, and enhancements, in: Proceedings of the 2021 ACM SIGCOMM 2021 Conference, SIGCOMM '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 597–609, http://dx.doi.org/10.1145/3452296.3472908.

[103] X. Lin, Z. Lin, S.E. Löwenmark, J. Rune, R. Karlsson, Ericsson, Doppler shift estimation in 5G new radio non-terrestrial networks, in: 2021 IEEE Global Communications Conference, GLOBECOM, 2021, pp. 1–6, http://dx.doi.org/10.1109/GLOBECOM46510.2021.9685184.

[104] S. Xu, M. Perez, K. Yang, C. Perrenot, J. Felblinger, J. Hubert, Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dV-Trainer® simulator, Surg. Endosc. 28 (2014) 2569–2576.

[105] Y. Ebihara, E. Oki, S. Hirano, H. Takano, M. Ota, H. Morohashi, K. Hakamada, S. Urushidani, M. Mori, Tele-assessment of bandwidth limitation for remote robotics surgery, Surg. Today 52 (2022) 1653–1659.

[106] OnePlus, OnePlus' milestone 2021 and new updates for 2022, 2022, URL https://community.oneplus.com/thread/1549522?utm_source=chatgpt.com.