Application Performance using 5G mmWave

Project deliverable

Per Hurtig
Karlstad University
per.hurtig@kau.se

ABSTRACT

The fifth generation (5G) cellular mobile network promises a combination of high speed and ultra-low latency to enable use cases not possible before. To accomplish this 5G makes use of new radio technologies like mmWave, which works at much higher frequencies than previous cellular network standards. The problems of using high frequency bands include poor radio propagation and sensitivity to blockages. This report investigates how commonly used network applications will be affected when transitioning from today's cellular networks to 5G. The performance of file downloads, video streaming, and web surfing are evaluated in an emulated 5G network and are compared to 4G experiments. The main outcomes of the experiments are that: mmWave generally outperforms 4G/LTE; applications can suffer, both in terms of throughput and latency, from the large capacity variations in mmWave channels; TCP loss recovery and congestion control needs to consider mmWave channel dynamics better to enable large application gains when moving from 4G to 5G.

KEYWORDS

5G; mmWave; transport-layer; TCP; emulation.

1 INTRODUCTION

The number of mobile devices have now exceeded the world's population, according to Cisco's Global Mobile Data Traffic Forecast [8]. Also, the fourth generation (4G) traffic now exceeds third generation (3G) traffic and accounts for almost 70 % of the total mobile traffic [8]. As trends suggest that the expansion will continue and the fact that requirements are growing harder to satisfy, the fifth generation (5G) cellular mobile network is currently being standardised. The 5G mobile network promises a combination of high speed and ultra-low latency [2] to enable diverse use cases that previously was not supported in cellular networks. Examples include support for ultra-high definition (UHD) TV streaming, augmented reality (AR), mission critical communication to enable e.g. self-driving cars and tactile applications.

To provide such high-end performance, 5G makes use of new radio technologies like millimeter wave (mmWave), which operates at much higher frequency bands than previous cellular network standards. By using frequencies above 10 GHz, there are estimations of peak capacities up to 200 times larger than previously seen in mobile cellular networks. Furthermore, early prototypes using mmWave have already proven the possibility of conducting multi-gbps transmissions outdoor.

The massive capacity, however, comes at a price. Due to the use of very high frequencies, mmWave links are likely to have highly variable channel quality. Signals may be blocked by common building materials and are also severely attenuated by human bodies. Therefore, the movement of a mobile handset in combination with the movement of interfering objects (such as cars, trains, humans, and so on) may cause the 5G mmWave channel to rapidly change in terms of quality, or even be blocked frequently.

The combination of massive but highly variable capacity may have an impact on regular internet-based applications. While traditional applications (e.g. mail, file download, and so on) often were elastic and rather oblivious towards latency and/or capacity variations, more and more applications are becoming increasingly interactive. For example, consider an end user that is trying to chat with a friend using a messaging application. For such an application, the most important property is having a smooth and responsive communication with minimum latency The 5G standard do include different service grades, to enable support for e.g. ultra-low latency. However, this kind of service is targeted towards specific types of applications (such as self-driving cars) and not general internet-based applications.

To make things even more complicated, several researchers are arguing that the de facto transport of the internet – the Transmission Control Protocol (TCP) [3, 23] basically is incompatible with the properties of 5G mmWave. The retransmission and congestion control mechanisms used by modern TCP implementations are expected to interact poorly with mmWave properties as they are unable to quickly detect and react to capacity variations. Therefore, one could assume that regular applications will not function properly when used over 5G mmWave channels.

By experimentally evaluating a number of real applications over an emulated 5G mmWave link, we try to determine whether this is a real problem or not. For the experiments, we use a number of applications, including file download, video streaming, and web browsing (elastic → interactive) in a number of realistic outdoor scenarios. The scenarios include handset movement at walking and driving speeds in the presence of obstacles such as buildings and humans. To correctly assess the performance of the applications, we use application-specific metrics such as, e.g., throughput (file download), video buffer levels (video streaming), and Speed Index (web). The results from the experiments indicate that, compared to 4G/LTE, regular internet-based applications do not generally suffer from the high but varying capacity that 5G mmWave links offer. However, to better utilise the underlying capacity and thus give the 4G → 5G transition a meaningful performance improvement, the TCP loss recovery and congestion control needs to be adapted.

The rest of this report is structured as follows. Section 2 provides a short history of the cellular network evolution and tries to briefly describe 5G and its intended use cases. In Section 3 the mmWave channel technology is introduced along with its inherent issues. Then, in Section 4 we evaluate a number of mobility scenarios involving a 5G mmWave link. The results from these experiments are then used in Section 5 to evaluate the application performance of file downloads, video streaming, and web browsing. The report ends with Sections 6 and 7, discussing the outcome of the experiments, what conclusions that can be drawn, and some avenues for future work.

2 THE FIFTH GENERATION (5G) CELLULAR MOBILE NETWORK

Only a few decades ago the first mobile wireless communications system was launched. This, first generation, wireless cellular system was analog and did only support voice communication. During the early 90's it was replaced with the second generation of wireless cellular technology, commercially launched in Finland 1991. The new features of the 2G system included *data services* such as Short Message Service (SMS) and Multimedia Messaging Service (MMS). The evolution rapidly went on and later on in the 90's, the third generation (3G) cellular technology was introduced. The main advantage for its' end-users, as compared to earlier technologies, was the ability to connect to the internet with decent performance. Attainable data rates varied between a few hundred kilobits and a few megabits, per second. The currently dominating cellular technologies, 4G/LTE (Long Term Evolution), were introduced in the early 2000's. The main advantage over 3G is the faster data rates that could now reach almost 100 megabits per second and thus enable applications such as high-definition mobile TV.

The fifth generation mobile data network [2] is no different from previous upgrades of the cellular network; it will provide its end-users with higher data rates. However, the main evolution compared with today's LTE technology is that beyond data rate improvements. The 5G standard tries to satisfy a much larger set of use cases, making it equally suitable for services relying on very fast data rates as for services requiring e.g. ultra-low latency. To support a set of very heterogeneous requirements, the 5G standard sets out to fulfill three different services grades: Enhanced Mobile Broadband (eMBB), Ultra Reliable and Low Latency Communication (URLLC), and Massive Machine Type Communications (mMTC):

eMBB aims to service densely populated areas with downlink speeds from 300 Mbps (megabits per second) to 1 Gbps (gigabits per second). Such downlink speeds will be made possible by small high-frequency mmWave antennas placed throughout the environment (e.g., on buildings, trees, buses, and so on). For rural areas, eMBB will replace the current LTE system with a network of lower-power antennas guaranteeing downlink speeds of 50 Mbps, at least.

URLLC will enable critical communication where bandwidth is not as important as low latency. More specifically, URLLC will need to guarantee end-to-end latencies of, at most, 1 ms. Possible use-cases for URLLC is applications involving e.g. self-driving cars or tactile applications, where minimal latency is necessary.

mMTC targets machine-to-machine (M2M) and Internet of Things (IoT) applications. While early cellular data standards (like 2G) did support e.g. M2M requirements, later standards have not. The reason for this is that later generations introduced throughput optimisations that in turn introduced a plethora of latency sources. Thus, mMTC seeks to restore the service level required by M2M and IoT by providing a transmission service with downlink speeds down to 100 Kbps (kilobits per second) and low latencies (around 10 ms).

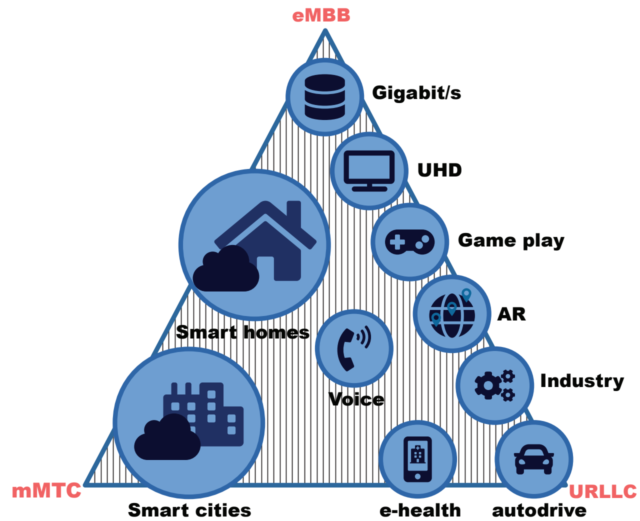


Figure 1: An example of use cases and their mappings to the 5G services grades.

Figure 1 illustrates the different service grades and how different use-cases maps to them. For instance, to download content from the internet at gigabit speeds and watch ultra-high definition TV the throughput requirements of eMBB needs to be met. For cars to support self-driving, there is no need to transmit large amounts of data very fast. Instead, small control messages needs to be exchanged with as low latency as possible.

To support such heterogeneous requirements 5G networks will employ multiple types of access technologies, multi-layer networks, and multiple types of devices. Furthermore, some of the access technologies employed by 5G will need to be designed using new radio technologies to support, e.g., extremely high data rates or very low latencies. The use of mmWave for eMBB is an example of that.

3 MMWAVE FOR BETTER PERFORMANCE

The millimeter wave (mmWave) bands have attracted considerable attention for next-generation cellular wireless systems, cf. [4, 10]. mmWave makes use of the radio frequency spectrum roughly between 30 and 300 GHz with the addition of some lower bands just above 6 GHz. Due to the propagation properties at these frequencies, the use of mmWave bands delivers massive, but highly variable, capacity. The high peak rates arise from the enormous spectrum availability in the mmWave bands. For example, recent prototypes have demonstrated multi-Gbps throughput in outdoor environments [11]. Most wireless systems operate below 6 GHz, where the lower frequencies allow for long-range propagation and low penetration loss due to e.g. attenuation by obstacles. These properties make them popular to use for radio communication – so popular that the spectrum below 6 GHz has become so heavily oversubscribed that individual, contiguous, bands wider than a few hundred MHz are no longer available. However, at mmWave frequencies there are large contiguous bands of spectrum available, making very large bandwidths possible. Actually, bandwidths larger than 1 GHz could be possible in some cases.

The mmWave bands have been used successfully in scenarios where directional transmission/reception of signals are possible, e.g. in point-to-point backhauls and satellite systems. However, for mobile access networks mmWaves have been considered as an impossible choice due to their poor isotropic path loss properties[1] and vulnerability to shadowing. For example, mmWave signals are completely blocked by many common building materials such as brick [20]. The human body, alone, can also severely attenuate mmWave signals, almost to the point of a complete blockage [21]. As a result, the movement of obstacles and reflectors, or even changes in the orientation of a handset, can cause the channel to rapidly appear or disappear.

The limitations of mmWave, from a mobile access network point-of-view, are not impossible to overcome. For example, it has been shown [1] that high-gain directional antennas can be used for improving the transmission performance of mobile devices. There is therefore some hope in directional smart antennas as an enabler for efficient use of mmWave links in mobile access networks. However, there are still challenges to overcome. For instance, mmWave systems also suffer more from attenuation than previous systems, as the received signal power is only a thousand of the power received in legacy systems (for equal distances between transmitter and receiver). To compensate for such significant attenuation, beam forming techniques and arrays of multi-element antennas are proposed. These techniques are possible as the antennas required for transmitting and receiving mmWaves can be much smaller in size.

The promise, and challenges, of extreme performance have really spurred the research surrounding mmWave links. To date, many of the problems of using these frequency bands for mobile access have been, or are about to be, resolved. However, most of the research have until now focused on the physical layer. When going up the radio/network stack, there are still many challenges to solve before the mmWave technology can be used properly in 5G networks. For instance, although solutions exist at the physical level to address e.g. the requirements of directionality (by using smart directional antennas), this also has to be handled in cooperation between the transmitter and receiver, to support mobility. That is, both the transmitter and receiver must, at all times, align the antennas with the movements of the mobile user. There are also issues related to the discovery of base stations, which in previous cellular systems were easily handled by periodically transmitting omnidirectional broadcast signals.

Moving further up the network stack, there are also issues involved in the interaction with the MAC, network, and transport layers. For instance, the quickly varying channel quality of mmWave is believed to require frequent and very fast handovers between 5G and legacy 4G cells in the case the primary 5G link becomes completely blocked. The work presented in this report is focused on the transport layer. There are concerns that the loss recovery and congestion control mechanisms used by the de facto transport TCP cannot deal with the quick capacity variations occurring in mmWave links. To function well, TCP needs to quickly adapt to sudden fluctuations to allow applications to utilise the available capacity and to prevent them from overflowing the network during periods of poor channel quality (or in scenarios with complete outage).

[1]It is not possible to efficiently transmit mmWave signals that radiate equally in all directions.

4 MMWAVE PERFORMANCE IN REALISTIC SCENARIOS

Evaluating real application performance in a 5G environment presents a number of challenges. First, we have no way of running any experiments inside a real 5G testbed or deployment. Second, as we want to evaluate real applications we cannot use an analytical model of a 5G mmWave link or use any of the existing 5G simulators. To resolve this stalemate, we engineered a process to first model a 5G mmWave link in different scenarios using the ns-3 simulator [14, 15] equipped with the mmWave extension [22]. We then extracted all link properties from these simulations and used them to build network emulation patterns for the KauNetEm system [17]. Using KauNetEm together with the 5G emulation patterns we could then build a realistic test environment for real applications.

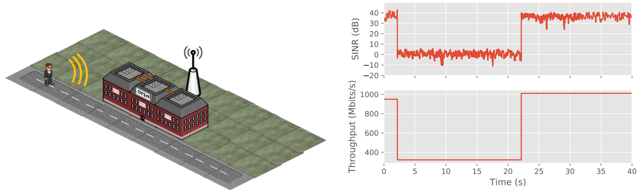
The remainder of this section details the different scenarios that were considered and how the 5G channel behaved when simulating them. The following section then uses these scenarios to experimentally evaluate the different applications considered in this report.

4.1 Experimental details

We considered four scenarios, all relevant considering both typical end-user movements and how a future network topology might look. For all scenarios, we configured the mmWave module of ns-3 to have typical settings. For instance, the system bandwidth was set to 1 GHz, the MAC layer was equipped with a hybrid automatic repeat request (HARQ) to better react to channel fluctuations, and the channel intermittency was modeled through a semi-statistical model with a simulated topology for each described scenario. For more details regarding the ns-3 mmWave framework and the channel model, please see [22]. Furthermore, we have published all details surrounding these channel simulations at [16]. This includes all software, scripts, and descriptions necessary to repeat the experiments.

4.2 Scenario 1: Long period of NLOS

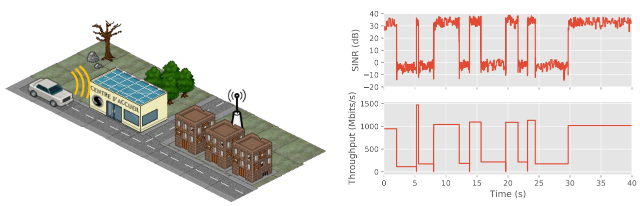
In this first scenario, we simulated the performance of the 5G channel for an end user with a mobile handset moving at a constant speed of 1.5 m/s (regular walking speed). At the beginning of the experiment (t=0 s) the user has line-of-sight to the base station, but very soon (t=2 s) a large building interferes, creating a non-line of sight (NLOS) situation. The NLOS situation then continues almost until the end of the experiment (t=30 s) which eventually ends in a LOS situation. The movement of the user is depicted in Figure 2a. The resulting performance of the 5G channel is illustrated in Figure 2b which shows the signal-to-interference-plus-noise ratio (SINR) and throughput (at the transport layer) as a function of time. From the results, it's obvious that the large building has



(a) UE movement

(b) SINR and throughput during movement

Figure 2: Scenario 1. SINR and throughput at different times: UE has LOS at first, then it starts to move and becomes shaded by a large building, before entering and stopping in a LOS position.



(a) UE movement

(b) SINR and throughput during movement

Figure 3: Scenario 2. SINR and throughput at different times: UE has LOS at first, then it starts to move and becomes shaded several small buildings, before stopping in a LOS position.

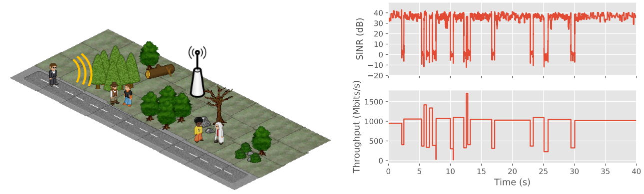
a very negative impact on the channel performance, resulting in SINRs around 0 dB during NLOS and a reduction of throughput from approximately 1 Gbit/s to around 300 Mbit/s, less than a third of the available capacity.

4.3 Scenario 2: Several moderate periods of NLOS

In the second scenario, we simulated the performance of the 5G channel for an end-user sitting in the back seat of a car. The car drove slowly (at approximately 8.3 m/s) between three road crossings, stopping for 3 s at each. During this ride, the channel was partly blocked (NLOS) by small buildings. The scenario is sketched in Figure 3a, although the figure only shows a single road crossing. The resulting performance of the 5G channel is illustrated in Figure 3b which shows the SINR and throughput as a function of time. The results are similar to those shown earlier, i.e., that the SINR becomes approximately 0 during periods of NLOS and that the throughput decreases to approximately a third of the 1 Gbit/s that is attainable during LOS.

4.4 Scenario 3: Plenty of small NLOS periods

In the third scenario, we simulated an end user moving past a group of people and trees, creating a number of small NLOS periods. The user moved at regular walking speed (1.5 m/s) past all obstacles. Once again, the channel quality, shown in Figure 4b, drops sharply when shaded by objects resulting in throughputs that are less than a third of the original.



(a) UE movement

(b) SINR and throughput during movement

Figure 4: Scenario 3. SINR and throughput at different times: UE has LOS at first, then it starts to move and becomes shaded by a group of people and trees, creating several small NLOS situations, before stopping in a LOS position.

4.5 Scenario 4: A complete outage scenario

We also performed experiments to simulate complete blockages. In the previous scenario the user moved past a group of people and trees which created a series of NLOS situations. For these experiments, we repeated this but placed the end user closer to the people and trees (a few centimeters), resulting in complete link outages whenever passing the corresponding object. Each one of these outages lasted for approximately 1 s and the resulting channel performance was as could be expected, near zero throughput at the transport layer.

5 MMWAVE AND REGULAR APPLICATIONS

The focus of this report is to determine how real applications are affected by the previously showcased fluctuations. Applications usually require either high throughput or low latency. Our selection of applications tries to capture both these aspects. The first application we evaluate is a simple file-transfer service. The performance of this kind of application is, of course, very tightly couple to the available capacity, or throughput. It is therefore likely that the use of mmWave will have a significant impact on the performance. The second type of application that is evaluated is video streaming. While this type of application also depends on the available capacity, video streaming applications often have more complex requirements. For instance, they might suffer more from the actual variations than having a low throughput, as buffering and encoding strategies often are based on an estimate of the available capacity (as well as latency). Finally, we evaluate web surfing. Web surfing performance does not typically require much capacity. Instead, responsiveness, in form of low latency, is more critical.

5.1 Experimental setup and details

We evaluated the applications for all the scenarios shown earlier. To accomplish this, we extracted the simulated channel models from the ns-3 framework and translated them into KauNetEm [17] emulation patterns. KauNetEm is a network emulator developed at

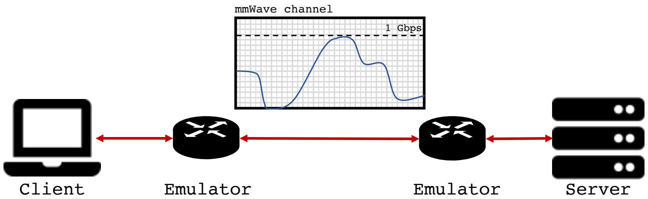
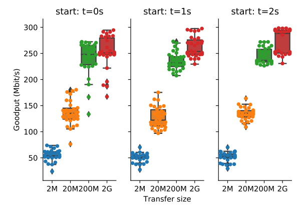
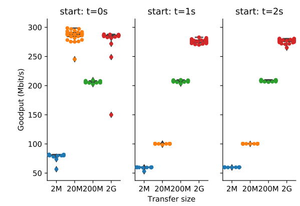


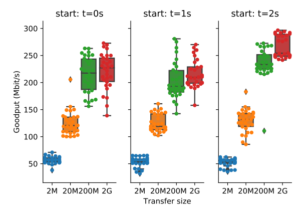
Figure 5: Testbed setup.



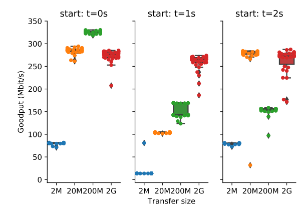
(a) Scenario 1: Long NLOS blockage



(b) Scenario 2: Several medium NLOS



(c) Scenario 3: Many small NLOS



(d) Scenario 4: Many small OUTAGES

Figure 6: Goodput for file downloads in all four 5G mmWave links scenarios.

Karlstad University that is able to apply emulation effects on traffic in a deterministic fashion. For instance, it is possible to define a pattern that applies a certain bandwidth restriction to traffic traversing the emulator at a certain point in time.

Thus, to emulate the movement in each evaluation scenario, we created bandwidth patterns that varied with time. For instance, in the first scenario where the user moves past a building, the bandwidth of the users mobile terminal varies accordingly. The applications were then evaluated using these bandwidth patterns in the testbed depicted in Figure 5. The machines in the testbed are all standard desktop computers running Linux 4.19, with the default CUBIC [12] congestion controller.

To allow for comparisons with today's cellular network performance, we also ran all experiments over an emulated LTE link. The emulation patterns for the LTE link were all constructed in the same way as the mmWave patterns, using the ns-3 LENA module [5] instead of the mmWave module.

5.2 File download

The first application we evaluated was a simple file transfer application. We placed four files of different sizes on the server and instructed the client to download them. The size of the files were: 2 MB (mega bytes), 20 MB, 200 MB, and 2 GB (giga bytes). The reason for using differently sized files is to see how fast TCP can utilise the available capacity. For large files, the congestion controller of TCP will only spend a small fraction of the total transmission time to adapt to the available capacity, and thus utilise most of it. For smaller files, this rate adaptation will span a larger portion of the transfer, effectively showing how fast/slow TCP can utilise the link.

We downloaded each file 30 times for all the four mmWave scenarios described earlier, to determine what effect LOS/NLOS/OUTAGE has on the performance. We also experimented with the starting time of the file transfer. In the first set of experiments, we started the transmission as the scenarios started (t=0 s), i.e., the transmission started in a LOS state. In the second set of experiments, the transmission was started after 1 s, i.e., in the transition between LOS→NLOS (for most of the scenarios). Finally, in the third set of experiments we started the transmission after two seconds, i.e., in an NLOS state (for most of the scenarios). This was done to see whether any performance-related corner cases could be discovered when forcing TCP's congestion controller to start under different circumstances.

To assess the performance in a meaningful way, we measured the goodput of the transfers. The results from the experiments are shown in Figure 6. We ran the exact same set of experiments using 4G/LTE. However, as 4G/LTE is not as sensitive to blocking objects as mmWave, we got a constant goodput of approximately 50-75 Mbit/s (depending on file size) for all experiments. The first observation that can be made is that the application is not able to fully utilise the available bandwidth in any of the scenarios. Although the available capacity peaks at 1 Gbps for extended periods of time in all of the scenarios, the average attainable goodput is limited to approximately 300 Mbps (at most). For the smaller files this is rather natural as TCP will spend longer time in its rate probing phase and will not be able to ramp up the goodput enough. However, for the largest file size (2 Gbyte) this is rather interesting. It seems that, for our chosen scenarios, it does not really matter if a transfer starts in LOS or not, if it is subjected to a long NLOS period or a number of small ones; the negative impact of the NLOS (and OUTAGE) periods seems to hurt the performance almost equally. The second interesting observation is related to the 20 Mbyte and 200 Mbyte file transfers (yellow and green in the graphs). They sometimes perform better than could be expected. The explanation to this is simply that the transfers manage to start and then run for a majority of the transfer time while the handset is in LOS. Thus, these transfers are fairly unaffected by the transition to NLOS or OUTAGE.

To summarise, the performance of applications relying of high goodput seems to be limited to the lower end of the varying capacity, due to TCP's slow utilisation of the available capacity. Thus, for applications relying on high goodput, the choice of congestion controller is important. For these experiments, Linux's default congestion controller, CUBIC, was used. Although CUBIC is able to utilise capacity faster than legacy controllers like e.g. NewReno [3, 13] there are alternative congestion controllers that are able to more quickly adapt to the underlying capacity. One example of such a congestion controller is BBR [7].

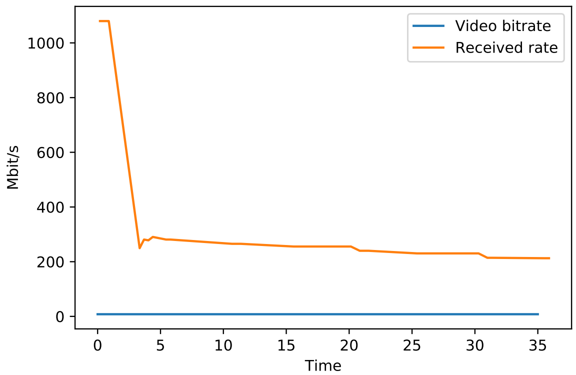


Figure 7: Dash

5.3 Video streaming

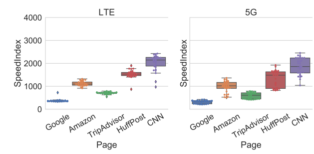
The second evaluation was conducted using a custom-made DASH video streaming client, based on the dash.js [9] player. The player, running in a Chromium browser, accessed the server to stream the Big Buck Bunny movie [6]. The movie was available in five different qualities, with the client dynamically switching among the versions according to network quality. We streamed the first 40 s of the movie for all the scenarios previously described, repeating each experiment 30 times to account for noise in the measurements. Furthermore, we ran the experiments both using 4G/LTE and mmWave. In recent years, many models of video streaming quality have been developed. Most metrics consider switching between different qualities too frequently or running out of buffered content as signs of poor quality. When reviewing such metrics we soon discovered that neither 4G/LTE nor mmWave had any problems in streaming Big Buck Bunny for the selected scenarios. Basically, the available capacity greatly surpassed what was required to e.g. stream the highest quality version of the movie (full-HD, 1080p) regardless of radio technology and/or scenario. This could have been expected if we had been a bit more careful when designing the experiment. Figure 7 depicts a sample of the results. The orange line shows how fast the movie is received by the player and the blue line shows the speed at which the movie is played out. This particular result comes from an experiment using mmWave, as can be seen by the gigabit speed in the beginning of the experiment. As shown in the figure, the player quickly reduces the download speed as enough data is buffered to have a smooth playback.

5.4 Web browsing

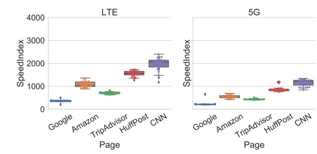
The final application we evaluated was the web. To accomplish this we selected five sites and mirrored them on our local testbed server. The sites (detailed in Table 1) were chosen to represent commonly used web sites of different sizes. Similar to the previous experiments, we ran the experiment for all mmWave scenarios previously described and also substituted the mmWave link with a 4G/LTE ditto to allow for comparisons. For these experiments we did not start downloading a site as soon as the scenario started. Instead, we randomized the start of the download within the time of the scenario(s). Each experimental combination was repeated 30 times to account for any variation. To perform realistic web downloads, all experiments were conducted using Pupeteer [18] together with the Chromium web browser. Pupeteer is a framework for automatically controlling web browsers and collecting performance metrics from them.

Table 1: Web sites used for the evaluation.

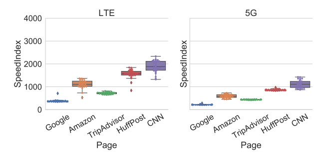
	Size	
Site	Objects (#)	Total size (MB)
Google	10	1.1
Amazon	130	3
TripAdvisor	48	6.5
HuffPost	114	4.9
CNN	148	10



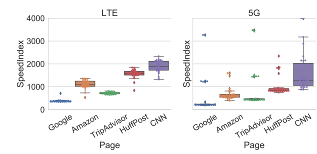
(a) Web Speed Index: Scenario 1



(b) Web Speed Index: Scenario 2



(c) Web Speed Index: Scenario 3



(d) Web Speed Index: Scenario 4

Figure 8: Web Speed Index for all Scenarios.

There is a plethora of metrics that can be used when assessing web performance. One commonly used is the Speed Index [19], which shows the average time (in milliseconds) at which visible parts of a page are displayed. Figure 8 details this metric for 4G/LTE and 5G in all of the scenarios. The 4G/LTE results are stable across all scenarios, as could be expected. Unsurprisingly, there seems to be a close correlation between the Speed Index and the size of the site; Google with the least amount of files and total size has the lowest Speed Index, followed by TripAdvisor, Amazon, HuffPost, and lastly CNN. Interesting to note is that TripAdvisor has a lower index than Amazon although the total size of TripAdvisor is more than double that of Amazon. This can be explained by the much smaller amount of files. A web browser typically opens six concurrent TCP flows to each domain, and a new connection is often used for each object. The number of objects can therefore be as important as the total size of the site, if not more. The same "effect" is evident from the relation between HuffPost and CNN. In this case, however, CNN is larger both in terms of total size and number of objects. If we compare the stable 4G/LTE performance to that of 5G mmWave, we note some interesting things. First, the Speed Index varies significantly between the different scenarios, as could be expected. Furthermore, while 5G mmWave outperforms 4G/LTE in nearly all scenarios (it is a close call in the last scenario), the variance within the scenarios is also significant. For instance, the Speed Index spread of 5G mmWave for e.g. the HuffPost site in the first scenario is very large.

The most interesting results were, however, obtained for the last scenario. For these experiments, it is hard to tell whether 4G/LTE or 5G mmWave is the best choice. We therefore chose to have a closer look at these experiments, to see if we could link the poor performance to a specific aspect of the web download. Figure 9 shows a number of performance metrics for this particular scenario, including time to first byte (TTFB), time to first paint (TTFP), time to interact (TTI), document object model (DOM), onLoad, and above the fold (ATF). Before we started to look at these metrics in detail, we noted that nearly all metrics exposed a special pattern in the case of 5G mmWave (fig. 9b). As shown in the graphs, there are concentrations of outliers at approximately 1000 ms, 3000 ms, and 5000 ms. These outliers are the main reason for the poor Speed Index shown earlier. The reason to them and why they are located where they are is related to the TCP loss recovery mechanisms. If the connection initiation in TCP fails, due to packet loss, a new attempt will be made after approximately 1000 ms. The other values (3000 and 5000 ms) are also artefacts of the TCP loss recovery that occurs when there are problems in initiating a connection properly, due to packet loss. The reason why this happens is due to connections being opened simultaneously to mmWave link outages. The reason why this was not seen for e.g. the file download application is simply that the 1-5 s penalty occurring when this happens is not enough to be clearly visible for such long flows. For a short web transfer, however, the impact is very large.

6 DISCUSSION

A number of studies have questioned TCP's ability to function properly over 5G mmWave links. Although it is natural that mmWave capacity variations might have a negative effect on TCP's congestion control abilities, we wanted to test whether this also had noticeable effects on network application performance. From our initial experiments it is evident that the mmWave technology, overall, provide better performance to applications than 4G/LTE is able to. However, it is also clear that the large capacity variations present in mmWave links do put an unnecessarily low limit on application performance. Although we did not see any performance issues in the video streaming evaluation, we did see problems both when high throughput/goodput is required (file downloading) and when low latency is of importance (web surfing). The causes to the poor quality when files are downloaded was expected as this is directly correlated to the channel capacity. The impact of link outages on web surfing was, however, larger than we expected. The performance problems can all be traced back to the loss recovery and congestion control mechanisms of TCP.

This, initial, study has a number of limitations:

- We did only consider a single mmWave configuration with a fixed bandwidth of 1 GHz and one connected handset. This can in some respect be seen as a best-case scenario.
- We did not consider videos with bitrates that actually could have been impacted by mmWave channel dynamics (e.g., 4K and/or 8K videos).

7 SUMMARY AND FUTURE WORK

This report investigates how mmWave links can affect network applications that use TCP. To do this, we created several emulated mmWave usage scenarios and experimented with real applications within them. We found that mmWave channel capacity variations interacts poorly with the TCP loss recovery and congestion control mechanisms and that these interactions translates into bad network application performance.

Except from the results obtained, the main contribution of this work was designing a process to transform realistic mmWave channel conditions (for given user scenarios) into KauNetEm emulation patterns. This process makes it possible to easily evaluate real applications, in an emulated testbed, using 5G mmWave channel characteristics.

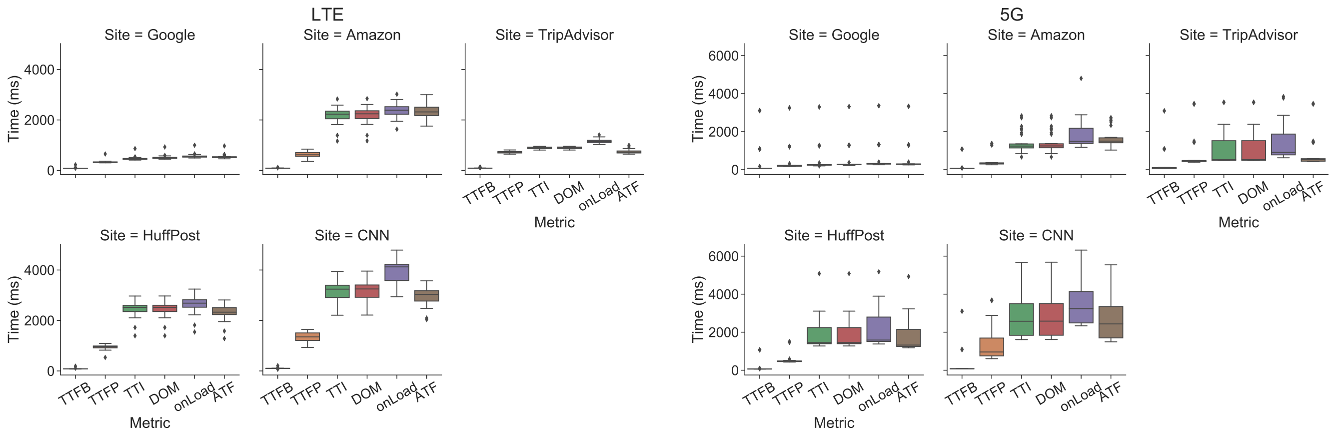
Future work on application performance over mmWave includes a more detailed examination of interactions between rapidly varying channel conditions and TCP congestion control instrumentation.

ACKNOWLEDGMENTS

This work has received funding from The Internet Foundation in Sweden. The views expressed are solely those of the authors.

REFERENCES

[1] M. R. Akdeniz et al. 2014. Millimeter wave channel modeling and cellular capacity evaluation. *IEEE Journal on Selected Areas in Communications* 32, 6 (June 2014), 1164–1179.
[2] Next Generation Mobile Networks (NGMN) Alliance. 2015. *5G White Paper*. Technical Report.
[3] Mark Allman, Vern Paxson, and Ethan Blanton. 2009. TCP Congestion Control. RFC 5681 (Draft Standard). http://www.ietf.org/rfc/rfc5681.txt
[4] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang. 2014. What will 5G be? *IEEE Journal on Selected Areas in Communications* 32, 6 (June 2014), 1065–1082.
[5] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero. 2011. An Open Source Product-Oriented LTE Network Simulator based on ns-3. In *Proceedings of the 14th ACM International Conference on Modeling, analysis and simulation of wireless and mobile systems (MSWiM)*. Miami, FL, USA, 293–297.
[6] Blender Foundation. 2019. Big Buck Bunny. https://peach.blender.org
[7] N. Cardwell, Y. Cheng, S. Hassas Yeganeh, and V. Jacobson. 2017. BBR Congestion Control. Internet Draft draft-cardwell-iccrg-bbr-congestion-control-00. http://tools.ietf.org/html/draft-cardwell-iccrg-bbr-congestion-control-00 Work in progress.
[8] Cisco Systems, Inc. 2017. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Paper*. Cisco Systems, Inc.
[9] Dash Industry Forum. 2019. dash.js reference DASH player. https://github.com/Dash-Industry-Forum/dash.js/wiki
[10] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. MacCartney, S. Sun, and S. Nie. 2014. Millimeter wave enhanced local area systems: A high data rate approach for future wireless networks. *IEEE Journal on Selected Areas in Communications* 32, 6 (June 2014), 1152–1163.
[11] J. Gozalvez. 2015. 5G tests and demonstrations [mobile radio]. *IEEE Vehicular Technology Magazine* 10, 2 (June 2015), 16–25.
[12] S. Ha, I. Rhee, and L. Xu. 2008. CUBIC: A New TCP-friendly High-speed TCP Variant. *ACM SIGOPS Operating Systems Review* 42, 5 (July 2008), 64–74.
[13] T. Henderson, S. Floyd, A. Gurtov, and Y. Nishida. 2012. TCP NewReno Modification to TCP's Fast Recovery Algorithm. RFC 6582 (Draft Standard). http://www.ietf.org/rfc/rfc6582.txt
[14] T. R. Henderson, M. Lacage, G. F. Riley, C. Dowell, and J. Kopena. 2008. Network simulations with the ns-3 simulator. *SIGCOMM Demonstration* 14, 14 (2008).
[15] T. R. Henderson, S. Roy, S. Floyd, and G. F. Riley. 2006. ns-3 project goals. In *Proceedings of the Workshop on ns-2 IP Network Simulator*. Pisa, Italy.
[16] P. Hurtig. 2019. Effektivare datatransport för 5G. https://github.com/perhurt/e5g
[17] P. Hurtig and J. Garcia. 2016. Deterministic Network Emulation using KauNetEm. In *Proceedings of netdev 1.1*. Sevilla, Spain.
[18] Google Inc. [n. d.]. Puppeteer. https://developers.google.com/web/tools/puppeteer/



(a) LTE **(b) 5G**

Figure 9: Web metrics for both 4G/LTE and 5G in Scenario 4.

[19] Google Inc. [n. d.]. Speed Index. https://sites.google.com/a/webpagetest.org/docs/using-webpagetest/metrics/speed-index

[20] F. Khan and Z. Pi. 2011. An introduction to millimeter-wave mobile broadband systems. 49, 6 (June 2011), 101–107.

[21] J. S. Lu, D. Steinbach, P. Cabrol, and P. Pietraski. 2012. Modeling human blockers in millimeter wave radio links. *ZTE Communications* 10, 4 (Dec. 2012), 23–28.

[22] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi. 2018. End-to-End Simulation of 5G mmWave Networks. *IEEE Communications Surveys & Tutorials* 20 (April 2018), 2337–2263. Issue 3.

[23] Jon Postel. 1981. Transmission Control Protocol. RFC 793 (INTERNET STANDARD). http://www.ietf.org/rfc/rfc793.txt Updated by RFCs 1122, 3168, 6093, 6528.