Peer reviewed version

Link to published version (if available):
10.1109/INFCOMW.2016.7562173

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Transport Layer Performance in 5G mmWave Cellular

Menglei Zhang[1], Marco Mezzavilla[1], Russell Ford[1], Sundeep Rangan[1], Shivendra Panwar[1],
Evangelos Mellios[2], Di Kong[2], Andrew Nix[2], and Michele Zorzi[3]

[1]NYU Tandon School of Engineering, USA [2]University of Bristol, UK [3]University of Padova, Italy
{menglei, mezzavilla, russell.ford, srangan}@nyu.edu, panwar@catt.poly.edu,
{Evangelos.Mellios, d.kong, andy.nix}@bristol.ac.uk, zorzi@dei.unipd.it

Abstract—The millimeter wave (mmWave) bands are likely to play a significant role in next generation cellular systems due to the possibility of very high throughput thanks to the availability of massive bandwidth and high-dimensional antennas. Especially in Non-Line-of-Sight conditions, significant variations in the received RF power can occur as a result of the scattering from nearby building and terrain surfaces. Scattering objects come and go as the user moves through the local environment. At the higher end of the mmWave band, rough surface scatter generates cluster-based small-scale fading, where signal levels can vary by more than 20 dB over just a few wavelengths. This high level of channel variability may present significant challenges for congestion control. Using our recently developed end-to-end mmWave ns3-based framework, this paper presents the first performance evaluation of TCP congestion control in next-generation mmWave networks. Importantly, the framework can incorporate detailed models of the mmWave channel, beamforming and tracking algorithms, and builds on statistical channel models derived from real measurements in New York City, as well as detailed ray traces.

Index Terms—Congestion control, Millimeter wave cellular, 5G, TCP, Raytracing, Performance evaluation.

I. INTRODUCTION

The millimeter wave (mmWave) bands – roughly corresponding to frequencies above 10 GHz – have attracted considerable attention for next-generation cellular wireless systems [1]–[5]. Due to the propagation properties at these frequencies, the use of mmWave bands could create networks with two features that have never been seen together before, namely links with *massive peak capacity*, but capacity that is *highly variable*.

The massive peak rates arise from the tremendous amount of spectrum available in the mmWave bands – up to 200 times by some estimates [1]. This very large bandwidth can be combined with the large number of spatial degrees of freedom available in high-dimensional antenna arrays. Indeed, recent prototypes have demonstrated multi-Gbps throughput in outdoor environments [6]. Simulation and analytical studies [7]–[9] have also predicted capacity gains that are orders of magnitude greater than in current cellular systems.

At the same time, mmWave links are likely to have highly variable quality. MmWave signals are completely blocked by many common building materials such as brick and mortar, [1], [10]–[14], and even the human body can cause up to 35 dB of attenuation [15]. As a result, the movement of obstacles and reflectors, or even changes in the orientation of a handset relative to the body or a hand, can cause the channel to rapidly appear or disappear.

This combination of features – massive, but highly variable, bandwidth – has widespread implications for the design of next-generation networks seeking to exploit these bands. The goal of this work is to understand the effect of this high variability introduced by mmWave links on the *transport layer*, specifically congestion control, an area that to date has been hardly addressed in the literature, to the best of our knowledge.

The fundamental role of congestion control is to regulate the rate at which source packets are injected into the network to balance two competing objectives: (1) to ensure that sufficient packets are sent to utilize the available bandwidth, but (2) to avoid too many packets that may result in congestion to other flows in the network. From this perspective, mmWave networks – that may introduce links with extremely variable capacity – raise several key questions. Will current congestion control mechanisms, e.g., TCP Cubic [16], be effective in settings with very high bandwidth variability? If not, how should congestion control algorithms evolve? What is the effect of other delays in the core network, and should the core network provide support such as split TCP or cross-layer optimization?

This paper aims at addressing such questions using a novel and very accurate end-to-end simulation framework. One of the significant challenges in evaluating the end-to-end performance in mmWave is the need to capture detailed aspects of the channel, link layer adaptation and backhaul, all of which can have a significant impact at higher layers. To this end, our simulation framework incorporates complex measurement and ray-tracing 3D channel models with dynamics, beamforming and tracking algorithms, link-layer scheduling and retransmission mechanisms, and core network latency and bandwidth constraints. This level of detail results in more realistic scenarios, and therefore our framework can be used to study real-world problems.

The rest of the paper is organized as follows. In Section II, we describe the proposed end-to-end architecture. In Section III, we describe the modeling of the channel dynamics of our mmWave channel, which is used to conduct a detailed performance evaluation, presented in Section IV. Finally, in Section V, we draw conclusions from the presented work.
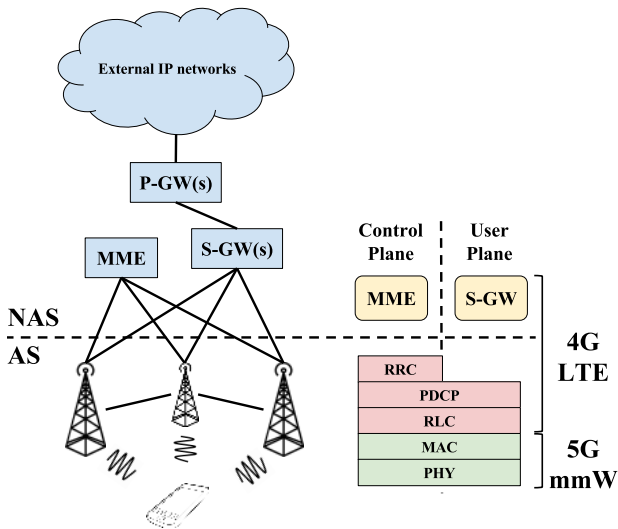
Figure 1: 5G mmWave end-to-end framework.

II. 5G mmWave end-to-end framework

Building on the mmWave ns3 module presented in our earlier work [17], we developed a 5G mmWave end-to-end simulation framework, described in [18] and briefly summarized here. At the physical layer, the proposed module incorporates detailed modeling of the mmWave channel that can capture spatial clusters, path dynamics, antenna patterns and beamforming algorithms. The paths in the channel can be modeled by either statistical clusters [19], ray tracing [20] or physical measurements [19]. Beamforming and channel tracking algorithms are also incorporated.

The simulation framework also offers a highly configurable MAC layer structure that can incorporate most of the recent frame designs including [21], [22]. The design flexibility can be used to evaluate the effect of control, scheduling and feedback delays. The radio link contol (RLC), radio resource control (RRC), non-access stratum (NAS) and core network architecture are inherited from the original LENA module [23] to provide detailed modeling of all the higher layers of the 4G LTE stack. Fig. 1 illustrates this architecture, which is composed of two main parts, namely, the **core network** and the **radio access network**.

The core network, also referred to as non-access stratum (NAS), comprises all the functions needed to establish the connection among cell towers and external IP networks. Each component is a separate server that executes a specific set of functions.

P-GW: The Packet Data Networks (PDN) Gateway enables the communication with external IP networks.

S-GW: The serving gateway is the mobility anchor and acts as a router by forwarding data among base stations and PDN gateways.

MME: The mobility management entity controls high-level mobile operations through signaling and handover commands.

The radio access network, also referred to as access stratum (AS), is the component responsible for the wireless connection between a base station (BS) and a mobile device, called user equipment (UE) in 3GPP terminology.

RRC: This layer provides services and functions for broadcasting system information related to NAS and AS, paging, security, connection establishment, maintenance and release.

PDCP: This communication protocol is mainly used for header compression and decompression of IP data.

RLC: This layer operates in 3 modes: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM). Our module supports the UM and AM mode, including PDCP and RRC functionality on top of that. UM and AM mode offers RLC layer error correction, concatenation, segmentation and reassembly of RLC data units.

MAC: Our proposed mmWave MAC layer is designed to meet the ultra low latency and high data rate demands, as presented in [21], thus following a flexible frame structure. A hybrid automatic repeat request (HARQ) is also implemented in order to better react to channel quality fluctuations.

PHY: We define here a set of mmWave-specific functions to capture transmission and reception capabilities trough a proper abstraction model, as presented in [24]. The characterization of the propagation behavior in the new spectrum, which is fundamental for our framework, is reported in the next Section.

III. MmWave channel dynamics

We aim at understanding how commonly used data applications will perform under *massive capacity* and *intermittency*. The massive capacity derives from the large increase in the available bandwidth, possibly up to two orders of magnitude. The intermittency is caused by three factors, as mentioned in earlier comments: (i) dynamic multipath (things moving around us), (ii) changes in scattering as a result of our movement in the environment and (iii) small-scale fading as a result of diffuse scatter and materials and surface types. At the present time, the dynamics of mmWave channels are not fully understood and there is no single agreed upon statistical model for simulation – some very initial work can be found in [25], [26]. The simulation framework proposed in [18] is general, and in this work we consider two possible models for channel intermittency: (i) a simplified semi-statistical model with a simulated topology and (ii) a more detailed ray tracing model.

Semi-statistical model with simulated topology: In this model, we first fix the location of the BS transmitter, obstacles and the UE route. At each point along the route, we then determine if the UE is a in Line-of-Sight (LOS) or Non-Line-of-Sight (NLOS) state based on the existence of an unobstructed path. In other words, if the line that connects the transmitter to the receiver crosses one or more buildings, the link is NLOS; otherwise, the link is LOS. Next, similar to [17], large scale parameters of the channel (the path cluster powers, angles of arrival, angular spread, delay, etc.) are drawn from the static statistical model proposed in [19]. The large-scale parameters are then assumed to be constant over the entire route, and we create a time-varying channel by adding the effect of the time-varying distance path loss and of
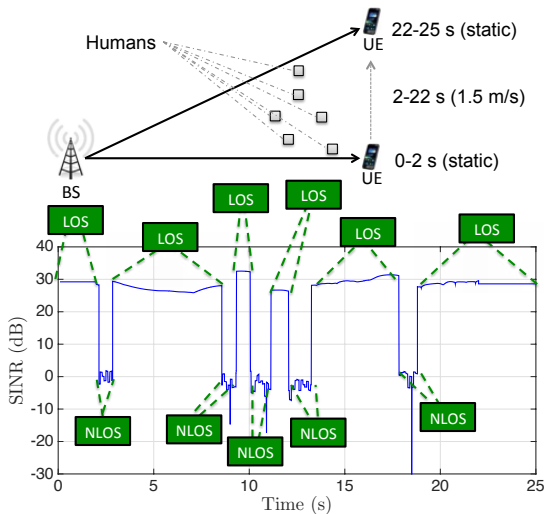
(a) SINR vs time for Scenario 1.
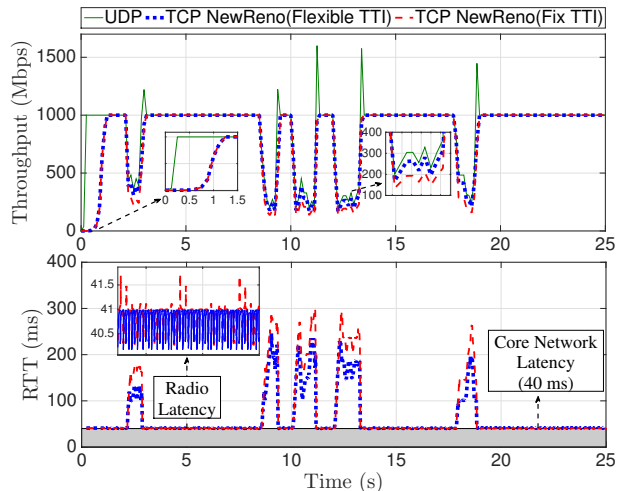
(b) Performance comparison at 1 Gbps data rate.

Figure 2: Simulation Scenario 1

Doppler shifts on the paths (this will cause small scale fading depending on the angular spread), and by eliminating the LOS paths in the regions where the UE is in a NLOS state. Note that in this model, the LOS-NLOS transitions are immediate – there is no modeling of diffraction.

Realistic traces: As a second option, we use raytracing to describe the time-correlated dynamics of the mmWave channel. To do so, we first generate the data of the sample points along a route measured in a real environment (Bristol, UK) and then apply raytracing to characterize the multipath. The ray model determines the mean power, time delay and 3D angles of the direct and scattered paths for a given city environment. Statistical small-scale variations (modeled based on measurements) are then added on a per-path basis to represent diffuse scattering from rough material surfaces. Finally, we feed the resulting traces into our simulator [18], where, as described in [17], [27], we apply beamforming gains and channel tracking based on the channel characterization presented in [7], [26].

IV. PERFORMANCE EVALUATION

In Section IV-A, we aim at capturing the way the link transitions among the LOS, NLOS and Outage states affect the application layer in a standalone mmWave context. To do so, along with a mmWave BS and a mobile user, we drop different sized link-blocking cubes, and observe the end-to-end performance of the simulated scenarios. The carrier frequency is 28 GHz, the transmit power 30 dBm, the total bandwidth 1 GHz, the number of TX and RX antennas is 64 and 16, respectively, and the core network latency is 20 ms.

In Scenario 1, we compare the performance of a single UDP flow against a TCP connection to observe (i) the time-offset experienced by TCP to reach capacity and (ii) the reduced TCP utilization due to the transmission of ACKs. Additionally, we will compare the TCP trends obtained with both a flexible and a fixed frame structure, to capture the gains achieved with

the design proposed in [21]. Then, in Scenario 2, we will compare the performance achieved by a TCP flow at varying data rates and RLC buffer sizes. Finally, in Scenario 3, we will force some outage events to see how different variants of TCP (NewReno and Cubic) react to a retransmission timeout (RTO) timer expiration.

In Section IV-B, we investigate the behavior of TCP over spatially consistent channel traces that are obtained by applying ray tracing techniques (enhanced using small-scale mmWave measurements to capture the statistical elements of diffuse scatter) and validate the key findings captured through our simulations.

A. Simulated topology

Scenario 1: We let a UE move over a path characterized by sudden link transitions due to *human* obstacles, represented as 0.5 m × 0.5 m cubes. TCP/UDP data packets are sent from a remote host to the UE with source rate 1 Gbps. We report both the topology and the time evolution of the received power in Fig. 2a.

1) UDP vs. TCP – As expected, the throughput of the UDP flow ramps up to the source rate faster than TCP. This is due to the slow start phase that characterizes the TCP connection, which, as shown in Fig. 2b, reaches the rate of 1 Gbps after approximately 1 second. Additionally, we can observe how, in terms of achieved throughput, UDP slightly outperforms TCP in NLOS regimes because of the TCP ACKs transmitted in the uplink, that affect TCP downlink utilization. The UDP throughput spikes observed at each $NLOS \rightarrow LOS$ transition are due to the RLC buffer, which accumulates a number of packets during the reduced SINR regime, and transmits the entire burst when the high capacity of LOS condition, around 3.5 Gbps as measured in [18], is restored (recall that no congestion control is used in UDP).

2) Flexible TTI vs. Fix TTI – As theoretically analyzed in [21], we show here some numerical results for the throughput and latency achieved with a flexible frame structure. Due to a
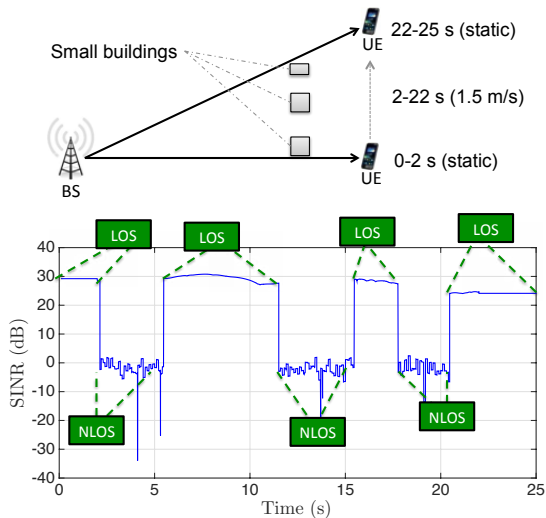
(a) SINR vs time for Scenario 2.
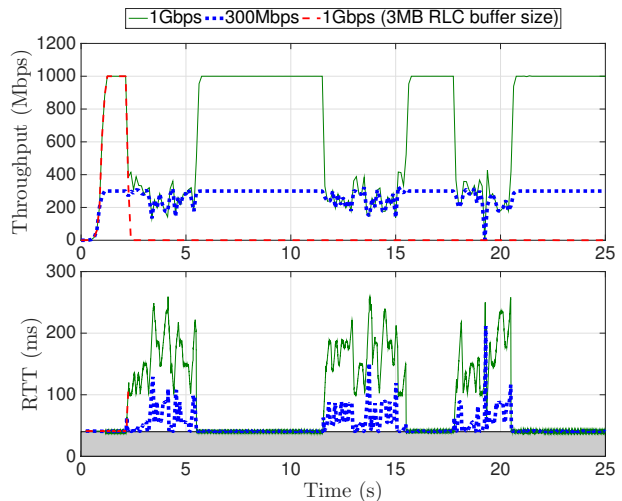
(b) Performance comparison at 1 Gbps data rate.

Figure 3: Simulation Scenario 2

dynamic design, our flexible scheme shows a better utilization of the channel in both directions, uplink and downlink, resulting in higher throughput and lower latency, thus corroborating the validity of our proposed frame structure.

3) Latency – It is important to note how in our scenario, where the application server is far from the BS (40 ms round-trip latency), the bottleneck is the core network. In fact, thanks to a shorter TTI, the mmWave radio at LOS introduces only 1 ms latency. On the other hand, in NLOS regimes, the latency increases by a factor of 100. This is due to lower layer retransmissions and RLC buffer size, as better illustrated in the next scenario.

Scenario 2: In our second simulated scenario, the link-blocking cubes are buildings with different sizes. A UE moves along the same route as in the previous scenario, with the same pedestrian speed, i.e., 1.5 m/s, while downloading a TCP NewReno flow. We report both the topology and the time evolution of the received power in Fig. 3a.

1) Bufferbloat – Bufferbloat, which causes high latency and packet delay variation when queueing too many packets in buffers, is a known problem [28]. As also shown in the previous scenario, with NLOS connectivity the latency dramatically increases. Due to the tremendous capacity available in the mmWave bands, the TCP congestion window (CW) is very large. Thanks to lower layer retransmissions, the retransmission timeout (RTO) timer never expires because packets are never dropped. Hence, both the CW and the RLC buffer, which is set to 10 MB by default, keep increasing. This directly translates into higher latency, thus introducing the problem of *bufferbloat*. If we compare the solid green and dotted blue lines in Fig. 3b that correspond to 1 Gbps and 300 Mbps data rates, respectively, we note that, while exhibiting a similar throughput, the latency is much higher at 1 Gbps than at 300 Mbps. The number of packets accumulated at the RLC buffer is indeed bigger, resulting in a more pronounced *bufferbloat* effect.

2) Buffer overflow - Fast retransmit – To overcome the *bufferbloat* problem, we reduced the size of the RLC buffer to 3 MB, as represented by the dashed line in Fig. 3b. The results are catastrophic, as the buffer cannot keep up with the packet arrival rate enabled by the large TCP congestion window, and drops a large burst of packets which is proportional to the CW. Consequently, TCP enters a *fast retransmit* phase, where 1 packet is retransmitted every RTT, thus requiring, in this case where the RTT is 40 ms, almost 40 seconds to retransmit 1000 dropped packets. The reason why TCP timeouts never occur is that the NewReno implementation in ns3 version 3.24 resets the retransmit timer after each partial acknowledgement – this is called the Slow-but-Steady variant of TCP NewReno as mentioned in [29].

Scenario 3: In the above scenarios, no TCP retransmissions are triggered thanks to lower layer retransmission schemes. Nonetheless, we would like to observe how different TCP variants, namely NewReno and Cubic, react to link failures that can cause an RTO expiration. To do so, we forced two outage events, of length 0.4 and 1 s, respectively. Topology and received power are reported in Fig. 4a.

1) Short outage – For the 0.4 s outage case, with RLC acknowledge mode enabled, HARQ at the MAC layer and ARQ at the RLC layer can still recover the corrupted packets without triggering any TCP retransmission. As a consequence, the CW never drops and the connection reaches capacity almost instantaneously, as shown in Fig. 4b.

2) Long outage – On the other hand, when the outage is 1 s long, the TCP RTO expires. After a timeout event, both TCP variants, Cubic and NewReno, enter slow start by setting the CW to 1 and halving the slow start threshold. Due to small CW, the latency is reduced, but this has a dramatic impact on the throughput of both TCP implementations, as shown in Fig. 4b. In particular, even though Cubic has been designed for better efficiency through a faster CW growth, this feature is only able to cope with some impairments and turns out
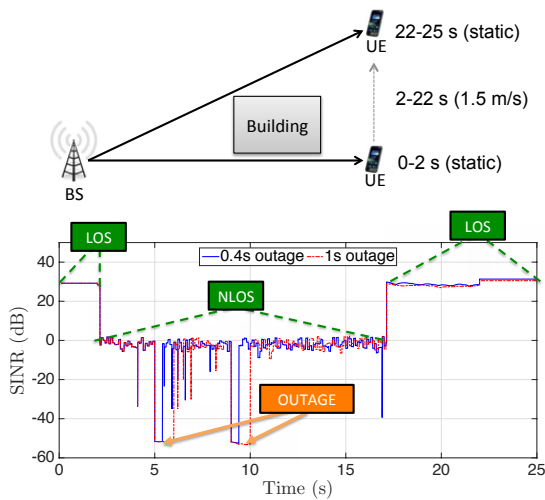
(a) SINR vs time for Scenario 3.
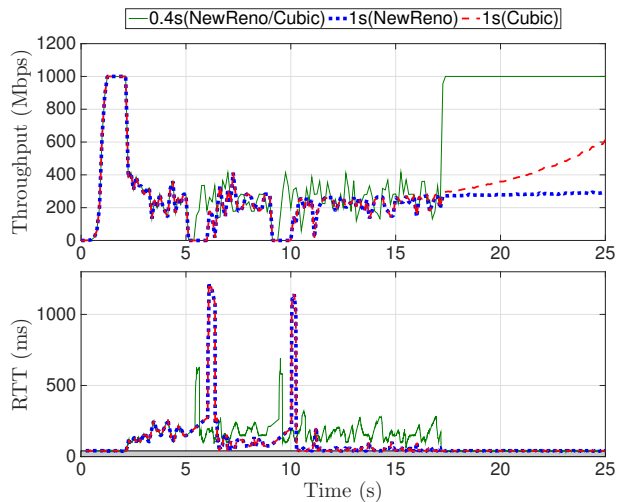
(b) Performance comparison at 1 Gbps data rate.

Figure 4: Simulation Scenario 3

to be unable to work properly if the outages are too long. NewReno is even worse since it has linear CW growth during the congestion avoidance phase. Therefore, even this highly optimized TCP version may not work well in our environment if outages occur on a certain time scale, which we were able to characterize. We believe that these observations will stimulate further research towards designing novel congestion control procedures able to better adapt to the unique mmWave radio environment.

B. Realistic traces

For a more realistic analysis, we plugged a trace generated through a combination of diffuse scatter measurements and raytracing into our end-to-end mmWave simulation framework, and compared the TCP rate and latency at (pedestrian) 3 m/s and (vehicular) 25 m/s UE mobility, respectively.

As shown in Fig. 5a, we observe some $LOS \leftrightarrow NLOS$ transitions that are determined by a group of trees across the direct path between the UE and the transmitting BS. The SINR dynamics is similar to those obtained through simulations in our earlier scenarios., i.e., ~ 30 dB. As noted before, and as presented in Figs. 5b and 5c, these sudden channel drops do not trigger any TCP retransmission. Hence, the only negative trend is the increased latency experienced during the NLOS regimes, which relates to the *bufferbloat* problem described above. Additionally, we can note in Fig. 5a that the link breaks ($NLOS \rightarrow Outage$) at approximately 150 m from the beginning of the *raytraced* route. Because the RTT never exceeds the RTO timer[1], RLC and MAC retransmission schemes prevent any TCP retransmission which, as shown above, would have dramatically affected the TCP performance.

Finally, if we compare Figs. 5b and 5c, we can observe that, at the higher UE speed, the achieved throughput is lower and the latency spike generated during the outage event is lower. The former is explained by the fact that the CQI feedback is

---

[1]The default minimum RTO is 1 s.

not frequent enough, given the rapid fluctuations of the channel due to the high speed, which results in some performance degradation. The latter, instead, is simply related to the fact that the fast-moving user is in outage for a much shorter amount of time.

V. CONCLUSIONS

MmWave cellular systems are likely to have very high peak capacity, that however is also highly variable. The basic question in this paper was to understand how this variability impacts the end-to-end performance. This evaluation is challenging since the end-to-end performance depends in complex ways on the interactions of the underlying channel dynamics, beam tracking, MAC-layer scheduling and retransmissions, network delays and congestion control. To understand these system-wide implications, we have studied some representative scenarios using a novel end-to-end ns3-based simulation framework.Our initial evaluations have revealed several possible problems in existing congestion control protocols in mmWave links. First, due to the very high data rate, current slow start mechanisms can take several seconds to achieve the full throughput offered by a mmWave PHY-layer. This may be problematic in applications that rely on short TCP connections. Second, large drops in rate, which are likely to be common in LOS-NLOS transitions, can result in very high levels of queuing and buffering, dramatically increasing the latency. Thirdly, after a retransmission timeout (RTO), even aggressive TCP protocols such as Cubic can take inordinately long to recover to full rates. While RLC and MAC layer retransmissions can shield upper layers from packet losses, RTO will still occur in real situations.

It is important to further study the mmWave channels and understand the effects of propagation dynamics on the upper layers. Additionally, how networks should evolve to address these problems and realize the full potential of the mmWave bands remains a broad and open problem. Solutions could involve bringing content closer to the edge (thus reducing

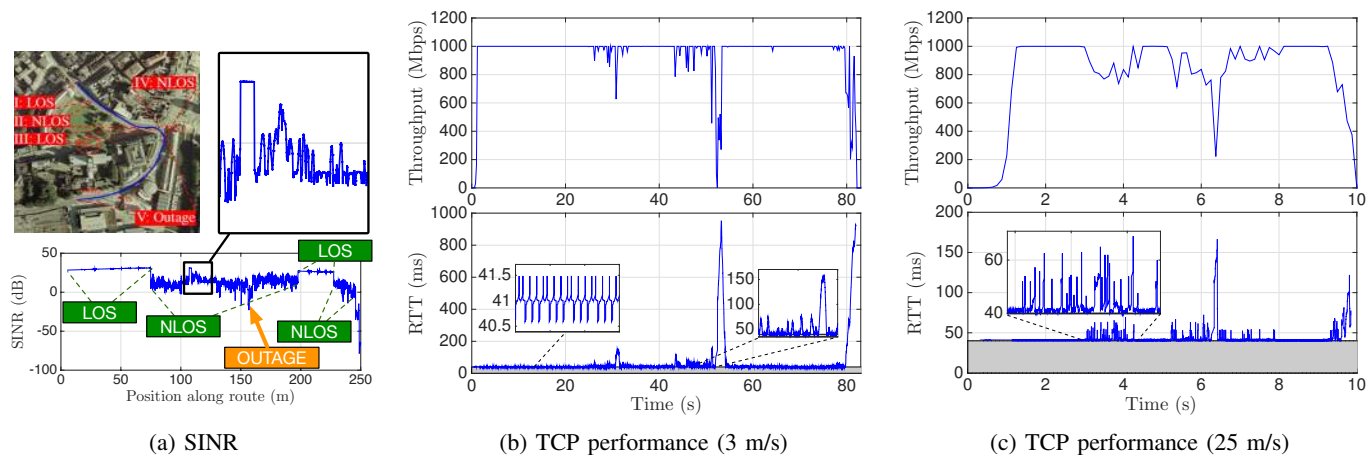(a) SINR   (b) TCP performance (3 m/s)   (c) TCP performance (25 m/s)

Figure 5: Raytracing Route

delay), assistance within the core network, or changes in the congestion control protocols themselves.

REFERENCES

[1] F. Khan and Z. Pi, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[3] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[4] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[5] 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, "Millimeter wave enhanced local area systems: A high data rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, June 2014.

[6] J. Gozalvez, "5G tests and demonstrations [mobile radio]," *IEEE Vehicular Technology Magazine*, vol. 10, no. 2, pp. 16–25, June 2015.

[7] M. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[8] T. Bai and R. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1100–1114, Feb. 2015.

[9] D. E. Berraki, S. M. D. Armour, and A. R. Nix, "Codebook based beamforming and multiuser scheduling scheme for mmwave outdoor cellular systems in the 28, 38 and 60 GHz bands," in *IEEE Globecom Workshops (GC Wkshps), 2014*, Dec 2014, pp. 382–387.

[10] K. Allen *et al.*, *Building penetration loss measurements at 900 MHz, 11.4 GHz, and 28.8 MHz*, ser. NTIA report – 94-306. Boulder, CO: U.S. Dept. of Commerce, National Telecommunications and Information Administration, 1994.

[11] C. R. Anderson and T. S. Rappaport, "In-building wideband partition loss measurements at 2.5 and 60 GHz," *IEEE Trans. Wireless Comm.*, vol. 3, no. 3, pp. 922–928, May 2004.

[12] A. Alejos, M. Sanchez, and I. Cuinas, "Measurement and analysis of propagation mechanisms at 40 GHz: Viability of site shielding forced by obstacles," *IEEE Trans. Vehicular Technology*, vol. 57, no. 6, pp. 3369–3380, Nov. 2008.

[13] S. Singh, F. Ziliotto, U. Madhow, E. M. Belding, and M. J. Rodwell, "Millimeter wave WPAN: cross-layer modeling and multi-hop architecture," in *Proc. IEEE INFOCOM*, 2007, pp. 2336–2340.

[14] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York City," in *Proc. IEEE ICC*, June 2013.

[15] J. S. Lu, D. Steinbach, P. Cabrol, and P. Pietraski, "Modeling human blockers in millimeter wave radio links," *ZTE Communications*, vol. 10, no. 4, pp. 23–28, Dec. 2012.

[16] B. Levasseur, M. Claypool, and R. Kinicki, "A TCP cubic implementation in ns-3," in *Proceedings of the 2014 Workshop on ns-3*. ACM, 2014.

[17] M. Mezzavilla, S. Dutta, M. Zhang, M. R. Akdeniz, and S. Rangan, "5G mmWave module for the ns-3 network simulator," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. ACM, 2015, pp. 283–290.

[18] R. Ford, M. Zhang, S. Dutta, M. Mezzavilla, S. Rangan, and M. Zorzi, "A framework for cross-layer evaluation of 5G mmwave cellular networks in ns-3," arXiv:1602.06932 [cs.NI]., Feb. 2016.

[19] T. S. Rappaport, G. R. Maccartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Transactions on, Communications*, vol. 63, no. 9, pp. 3029–3056, May 2015.

[20] N. F. Abdullah, D. Berraki, A. Ameen, S. Armour, A. Doufexi, A. Nix, and M. Beach, "Channel parameters and throughput predictions for mmWave and LTE-A networks in urban environments," in *Proc. IEEE VTC*, 2015.

[21] S. Dutta, M. Mezzavilla, R. Ford, M. Zhang, S. Rangan, and M. Zorzi, "Frame structure design and analysis for millimeter wave cellular systems," in *arXiv:1512.05691 [cs.NI]*, Dec. 2015.

[22] F. Khan and J. Pi, "Millimeter–wave mobile broadband: Unleashing 3–300GHz spectrum," in *IEEE Wireless Commun. Netw. Conf*, 2011.

[23] N. Baldo, "The ns-3 LTE module by the LENA project," 2011.

[24] M. Mezzavilla, M. Miozzo, M. Rossi, N. Baldo, and M. Zorzi, "A lightweight and accurate link abstraction model for the simulation of LTE networks in ns-3," in *Proceedings of the 15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '12, 2012.

[25] P. A. Eliasi and S. Rangan, "Stochastic dynamic channel models for millimeter cellular systems," in *Proc. IEEE Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2015, pp. 209–212.

[26] M. K. Samimi and T. S. Rappaport, "28 GHz millimeter-wave ultrawideband small-scale fading models in wireless channels," in *Proc. IEEE VTC*, May 2016.

[27] M. Rebato, M. Mezzavilla, S. Rangan, and M. Zorzi, "Resource sharing in 5G mmwave cellular networks," in *Proc. IEEE INFOCOM 2016, Millimeter-wave Networking Workshop (mmNet2016)*.

[28] J. Gettys and K. Nichols, "Bufferbloat: Dark buffers in the internet," *Queue*, vol. 9, no. 11, 2011.

[29] N. Parvez, A. Mahanti, and C. Williamson, "TCP NewReno: Slow-but-steady or impatient?" in *Communications, 2006. ICC'06. IEEE International Conference*, vol. 2, 2006, pp. 716–722.