uc3m | Universidad **Carlos III** de Madrid

University Master inTelematics Engineering
Academic Year 2016-2017

Master Thesis

"Analysis of TCP Performance in 5G mm-wave Mobile Networks"

Pablo Jiménez Mateo

Albert Banchs

Leganés, 20/09/2017

Abstract: Millimeter-wave (mm-wave) bands are projected to play an essential role in 5G mobile networks in supporting the increasing demands for higher data rates. Communications at mm-wave frequency bands pose unique challenges. The high variability in channel quality due to high propagation loss and unfavorable atmospheric absorption can only be overcome by the use of highly directional antennas. This results in a much higher degree of spatial reuse and lower interference if compared with omni-directional communications at lower frequencies.

However, the high directionality may cause communication blockages as the channel may appear and disappear because of beam misalignments. While significant research efforts have investigated challenges and aspects of physical (PHY) and medium access control (MAC) layers, the impact of the unique features of mm-wave systems on transport layer still requires attention. In this article, we focus on analyzing the behavior of congestion control protocols and study their impact on system-level performance. Through extensive simulations, we show the effect of different type of blockages on the design of congestion control protocol in presence of handovers and when small, medium and long flows coexist. Protocols like CUBIC that target high-throughput performance benefit significantly by joint optimization of link layers buffers and timeouts. While the optimization foster prompt reaction to short-term blockages and make them ideal for small and medium flows, their performance significantly decrease when obstacles degrade the channel quality for longer time periods. Hybrid-designs like YeAH are more robust to blockages, but fail to recover fast and achieve the link capacity after timeouts.

Analysis of TCP Performance in 5G mm-wave Mobile Networks

Pablo Jiménez Mateo, Claudio Fiandrino, *Student Member, IEEE,* and Joerg Widmer, *Senior Member, IEEE*

Abstract—Millimeter-wave (mm-wave) bands are projected to play an essential role in 5G mobile networks in supporting the increasing demands for higher data rates. Communications at mm-wave frequency bands pose unique challenges. The high variability in channel quality due to high propagation loss and unfavorable atmospheric absorption can only be overcome by the use of highly directional antennas. This results in a much higher degree of spatial reuse and lower interference if compared with omni-directional communications at lower frequencies. However, the high directionality may cause communication blockages as the channel may appear and disappear because of beam mis-alignments. While significant research efforts have investigated challenges and aspects of physical (PHY) and medium access control (MAC) layers, the impact of the unique features of mm-wave systems on transport layer still requires attention. In this article, we focus on analyzing the behavior of congestion control protocols and study their impact on system-level performance. Through extensive simulations, we show the effect of different type of blockages on the design of congestion control protocol in presence of handovers and when small, medium and long flows coexist. Protocols like CUBIC that target high-throughput performance benefit significantly by joint optimization of link layers buffers and timeouts. While the optimization foster prompt reaction to short-term blockages and make them ideal for small and medium flows, their performance significantly decrease when obstacles degrade the channel quality for longer time periods. Hybrid-designs like YeAH are more robust to blockages, but fail to recover fast and achieve the link capacity after timeouts.

Index Terms—Millimeter wave system, TCP, 5G cellular networks, simulations.

I. Introduction

MILLIMETER-wave (mm-wave) communications systems have emerged as a key wireless technology in next fifth-generation (5G) networks [1]. The limited spectrum availability below 6 GHz prevents current systems to achieve data rates required to meet 5G standards [2]. Mm-wave bands exploit frequencies above 10 GHz and due to the massive spectrum availability, they can provide orders of magnitude higher data rates than systems operating at lower frequencies. However, higher propagation loss and unfavorable atmospheric absorption make mm-wave systems highly susceptible to blocking. As a result, changes in relative orientation between the device and Access Points (APs), presence of obstacles and reflections lead to high variability in the channel quality.

The unique dynamics of mm-wave communications, capability in achieving high data rates with high variability in channel quality pose particular challenges to design physical

(PHY) and medium access control (MAC) layers properly. To this date, the majority of the research efforts have been devoted to investigate propagation issues [3], [4], beamforming procedures [5], and MAC layer design aspects [6]. However, higher layers such as the transport layer, and especially congestion control mechanisms, are also impacted by the extreme propagation conditions of mm-wave links. The role of congestion control is to match the demand and supply of packets by regulating the amount of injected traffic in the network accordingly to its congestion state. Sudden transitions from Line of Sight (LoS) to Non-LoS (NLoS) with a consequent decay of the channel quality would result in packet losses and might be interpreted by the sender as signal of congestion. Preliminary works in this area have analyzed the system-level implications of mm-wave channels on transport layer [7], and have focused in understanding the performance of Multipath-TCP and the impact of link level retransmissions [8]. Radio Link Control (RLC) in Acknowledged Mode (AM) performs user data delivery in-sequence and error-free based on (Automatic Repeat Request) ARQ schemes. While retransmissions at RLC layer are beneficial for throughput as hide channel losses to TCP [9], frequent RLC retransmissions increase round trip times (RTTs) and, in turn, might trigger TCP timeouts and impact negatively on the latency.

Throughout this article, we provide through extensive simulations the first systematic study on end-to-end performance of TCP congestion protocols in mm-wave 5G mobile networks. For a comprehensive analysis, the evaluation takes into consideration several loss-, delay-based and hybrid congestion protocols and studies their effectiveness for different applications and case studies: i) a single user served by one AP, ii) a single user performing handover between multiple APs, and iii) multiple users exploiting different applications served by multiple APs. Our contribution with this article is to identify critical design aspects of congestion control protocols and highlight transport layer performance trade-offs peculiar to mm-wave 5G networks. Furthermore, the article derives considerations and guidelines for recommended settings able to optimize specific objectives such as throughput-maximization or latency-minimization and take better advantage of the available network resources.

II. Challenges of Congestion Control Protocols over mm-wave Links

Channel quality in cellular networks is unpredictable for several reasons, including device mobility and handovers, frame scheduling algorithms which create burstiness and complex transitions of the Radio Resource Control (RRC) state

All the authors are with IMDEA Networks Institute, Madrid, Spain.
E-mail: {firstname.lastname}@imdea.org.

machine [10]. Congestion control protocols probe the channel to exploit the available bandwidth by injecting packets until losses are detected, or proactively adapt the injection rate to match delay- or rate-based measurements. Legacy congestion control protocols were primarily designed for wired networks assuming fixed capacity links, hence suffer short time variation of cellular networks that are further exacerbated in mm-wave networks. Large and persistently-full buffers, typical in cellular networks to help link layer retransmissions, lead to the bufferbloat problem, which results increasing latency significantly. The following paragraphs revisit the rationale of congestion control mechanism, highlight challenges and identify critical issues specific of mm-wave networks.

Nowadays, TCP is the de facto transport protocol used by the majority of the applications such as HTTP, file transfer and mailing [11]. TCP ensures that all the data packets are correctly received and in order. To this end, TCP requires each packet to be acknowledged by the receiver (ACK). TCP relies on a Congestion Window (CW), i.e., the maximum quantity of data that can be sent. The CW adaptation procedure depends on the link status. TCP NewReno, one of the most widely adopted congestion control protocols, during the slow start period increases the CW by one packet per received ACK until: (i) the *sstresh* threshold is reached, or (ii) there is a packet loss. Upon reaching the *sstresh*, TCP enters in congestion avoidance phase and the CW grows by one packet if all those in the current window have been acknowledged. When the receiver obtains a packet that has a sequence number higher than the expected one, a duplicated ACK is triggered. This informs the sender that the packet with the expected SN has been lost and requires retransmission. If three duplicated ACKs are received, TCP assumes a packet loss: it enters in fast retransmit mode by halving the CW and continues in congestion avoidance phase. If after a given period, the Retransmission Time Out (RTO) commonly set to 1 s, no ACKs were received for a packet, TCP reduces the CW to one packet and recovers from the slow start phase.

When on Line of Sight, a mm-wave link can offer several Gbps of bandwidth [6]. During congestion avoidance phase, TCP tries to take advantage of all the available bandwidth. However, if the channel becomes blocked by an obstacle, the data rate drastically drop and the RTT increases as a consequence. Sudden transitions between LoS and Non-LoS make the TCP sender buffer to fill instantaneously due to the high data rate, until the fast retransmit phase starts or an RTO is triggered. When entering in fast retransmit phase, the CW halves its value. However, this might result not sufficient to adapt to the new low link capacity. Consequently, an RTO is likely to be issued and TCP resumes from slow start to correctly adjust to the new bandwidth availability. Upon an RTO occurrence, the slow start threshold has to be updated, which makes TCP taking longer in achieving the optimal value for the CW and obtain high throughput, especially when the channel quality improves from NLoS to LoS.

Congestion control protocols differ in the mechanisms utilized to adapt to the available bandwidth. According to the methodology employed to detect congestion, protocols can be attributed to three main categories: loss-, delay-based and
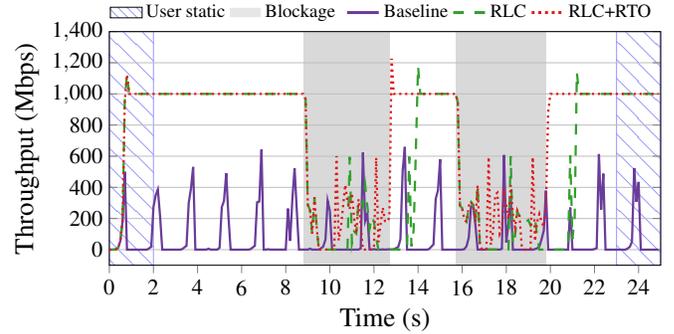

Fig. 1. Throughput Optimization on CUBIC

hybrid. After a preliminary analysis, six different protocols which encompass all the aforementioned categories were selected for the study: NewReno, Scalable, CUBIC, Vegas, Westwood and YeAH [12]. TCP NewReno relies on packet losses as indication of congestion. Unlike TCP Reno, the FR mechanism is designed to increase robustness to multiple losses in a single window. While NewReno follows a Additive Increase Multiplicative Decrease (AIMD) strategy, TCP Scalable implements Multiplicative Increase Multiplicative Decrease (MIMD) strategy to improve traffic growth rate and the rate increase is proportional to the spare bandwidth of the link. TCP CUBIC increases the CW according to a cubic function by estimating the absolute time since last dropped packet. In the first part, the function is concave and CUBIC attempts to quickly recover the previous state. In the second part, the function is convex and CUBIC probes for additional bandwidth. Vegas implements congestion control on the basis of delay-measurements. The CW size increases or decreases depending on the difference between the estimated throughput and the one the network accommodates. Westwood adjusts parameters like *sstresh* and CW according to estimations of the available network bandwidth based on measurements of the average rate of received ACK packets. Finally, YeAH is an hybrid approach, which determines the CW on the basis of both losses and delay.

III. CONGESTION CONTROL PROTOCOL PERFORMANCE ANALYSIS

A. Objectives

This section analyzes the performance of multiple congestion control protocols. Subsection III-B compares the protocols in several environments with different types of blockages and investigates their impact on throughput and RTT. Subsection III-C extends previous analysis to include handovers and highlights the impact of blockages in AP association and the achieved throughput. Finally, Subsection III-D investigates throughput and RLC buffer occupancy when multiple access points serve concurrently several users with different applications and discusses fairness.
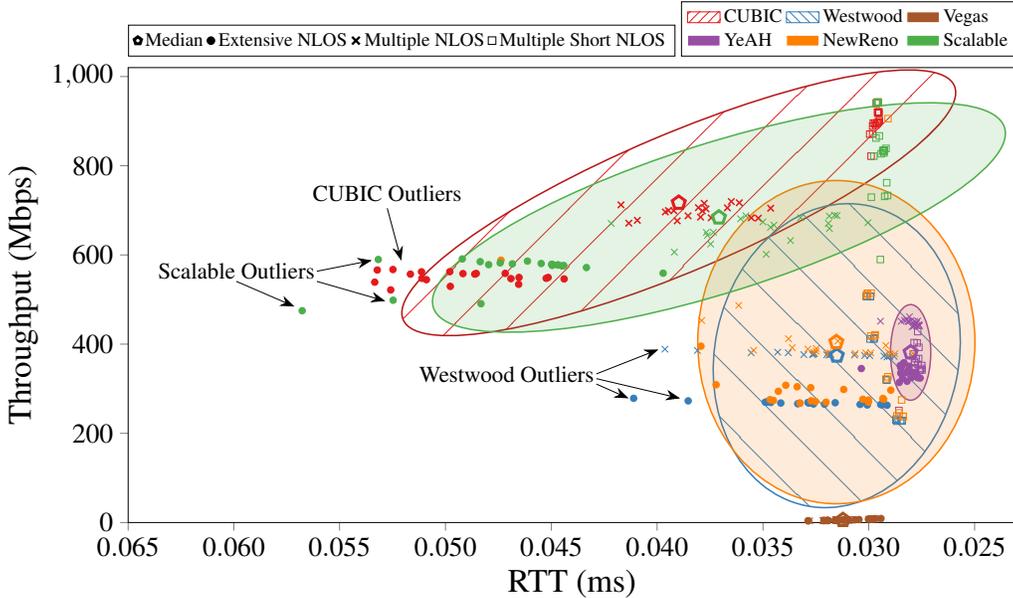
Fig. 2. Performance of TCP congestion protocols for different scenarios

B. Single Flow Analysis

A mm-wave AP serves one user (UE) positioned 150 m far. The UE remains static for 2 s, then moves following a straight path parallel to the AP location at a walking speed of 1.5 m/s for 21 s and finally it stops remaining static for other 2 s. Along the path, the connection is interrupted. Three blockage scenarios are considered: i) a building creating an extensive blockage of 13 s, ii) two small buildings creating medium-extensive blockages of 4 s each, and iii) 6 small obstacles creating multiple short blockages of 0.25 s each. Scenario i) and ii) permit to analyze performance of congestion protocols in presence of an extensive and medium extensive NLOS period. Finally, Scenario iii) extends the analysis of Scenario ii) augmenting the number of transitions LOS-NLOS and shortening their duration. In presence of medium or extensive blockages, congestion protocols can detect the link capacity reduction while in presence of temporary short-term blockages this is not possible. Retransmissions at RLC layer play an essential role in maintaining throughput. However, increasing the RLC buffer size to compensate for wireless losses has the side effect of increasing considerably the RTT and cause of bufferbloat. The duration of timers typically employed in wireless networks is not suitable for mm-wave systems. While in presence of short blockages entering in FR phase is beneficial, during extensive NLOS periods TCP is likely to incur in a RTO multiple times. Thus, TCP can significantly benefit from the reduction of the waiting period before it recovers from the slow start phase. Cross-layer optimization can significantly overcome these issues. For example, Azzino et al. [13] propose a cross-layer optimization of the congestion window, that is set according to the bandwidth-delay product considering the latency estimated without buffering delays.

While in traditional LTE technologies such as 3G and 4G RLC buffer sizes in the order o 1 MB were sufficient to compensate for wireless losses, in mm-wave networks the size

need to be adjusted. After having performed an optimization analysis, increasing the RLC buffer size to 7 MB results being optimal for throughput performance (see Fig. III-B). Note that the dimension is conservative with respect to the set-up employed in [13] and is in line with the 3GPP 36.306 TS Release 14 specification for a category 12 UE. Fig. III-B shows throughput performance of TCP CUBIC comparing the joint RLC buffer size and RTO timer optimization with respect to RLC buffer size optimization only and default settings. The limited RLC buffer size of baseline setting does not allow CUBIC to compensate for wireless losses and the protocol persistently fails to reach the congestion avoidance phase. Additionally, the duration of the RTO timer prevents CUBIC to react quickly and also in LOS phases the achieved throughput is suboptimal. The importance of the timer becomes evident comparing RLC versus RLC+RTO optimization: in NLOS periods, the latter performs significantly better and leads to faster recover from NLOS-LOS transitions.

Fig. 2 shows performance of the different protocols after having applied the RLC+RTO optimization for the considered scenarios. The graph is a throughput-RTT plot, where for each scenario we take the average results and compute the $2-\sigma$ elliptic contour of the maximum-likelihood 2D Gaussian distribution. The plot shows the results for individual scenarios and the median that summarizes the whole performance with different markers. On the x-axis, lower and better RTT are to the right. Hence, best performing protocols are on the top-right. Throughput-RTT type of plots are effective to highlight variability and correlation between protocols. The narrow the ellipses in the axis dimension, indicates the resilience of the protocol in constantly achieving similar throughput or RTT performance. On the other hand, wider ellipses indicates higher variability. The ellipses' orientation define the covariance between throughput and RTT: when the ellipses are parallel to the plot, throughput and RTT are uncorrelated.

CUBIC and Scalable are the best performing protocols. Being aggressive, their performance differs considerably from one scenario to another and are very susceptible to LOS-NLOS transitions in both throughput and RTT metrics. NewReno and Westwood exhibit less variability in RTT than CUBIC and Scalable. The latter, upon detection of congestion, computes the current bandwidth-delay product and sets the slow start threshold accordingly. The former, does not terminate fast recovery phase until complete recovery of multiple losses. In congestion avoidance phase, both protocols gently probe for additional bandwidth, which makes them taking longer to achieve full link capacity. This behaviour is particularly evident in the scenario with multiple short NLOS periods which exhibits more variability than the other two scenarios. YeAH, because of its hybrid design, switches between fast and slow mode in the spirit of Scalable and NewReno respectively. Hence, it achieves consistent performance in RTT and little variability in throughput. Finally, Vegas consistently achieves the lower throughput performance while exhibits little variability in RTT. Both Vegas and YeAH are robust to different types of LOS-NLOS transitions.

C. Single Flow with Handover Analysis

The unprecedented proliferation of mobile devices and demand for traffic pose a substantial challenge to current network infrastructures and has lead to ultra-dense deployment of small cells. As the number of small cell increases, effective mobility management becomes essential. Mm-wave networks employing highly directional antennas, as link establishment and maintenance under user mobility is a challenge. Users can either connect with single- or multi-connectivity modes, depending on the number of base stations they are simultaneously attached to. In single-connectivity mode, users perform hard-handover while in multi-connectivity mode the transition is smoother and faster. With dual-connectivity users are simultaneously connected to mm-wave and traditional sub-6 GHz APs [14]. From the network perspective, the case of standalone mm-wave deployment with single connectivity is the most challenging scenario and is the one under analysis.

Handovers transfer the ongoing data session from the current AP to another, to ensure that the current session remains active. While in traditional cellular networks handovers are mainly triggered by user movements out of the operational AP range, quick shortages are the more probable cause of handovers in mm-wave networks. Several methods drive handover decisions [14]. The threshold method estimates SINR degradation. The *fixed* and *dynamic* Time to Trigger (TTT) generate an handover event if the SINR is below a certain threshold during a time window. Its duration is predetermined or variable for *fixed* and *dynamic* methods respectively. The objective is to aims to reduce the ping pong effect, i.e., a UE switching between two APs in a rapid succession. The following analysis employs a fixed TTT with 200 μs time window.

Fig. 3(a) shows the geometry of the scenario, which mimics a regular street. Lampposts are distributed at regular intervals of 20 m and two double car lanes separate the pedestrian

USER	APPLICATION PROFILE		
	SPEED	TYPE	INTERVAL
UE_0	500 Mb/s	Always on	-
UE_1	4 Mb/s	Burst	Every 3 s for 0.8 s
UE_2	4 Mb/s	Burst	Every 2 s for 30 s
UE_3	4 Mb/s	Burst	Every 3 s for 0.8 s
UE_4	4 Mb/s	Burst	Every 2 s for 30 s
UE_5	4 Mb/s	Burst	Every 2 s for 30 s
UE_6	4 Mb/s	Burst	Every 3 s for 0.8 s
UE_7	4 Mb/s	Burst	Every 3 s for 0.8 s
UE_8	4 Mb/s	Burst	Every 3 s for 0.8 s
UE_9	4 Mb/s	Burst	Every 3 s for 0.8 s

lanes of 2 m each. Three APs, deployed at a regular distance of 20 m, ensure full coverage. The UE moves at a constant speed of 1.5 m/s on the bottom pedestrian lane from left to right. The lampposts create multiple short blockages similarly to the multiple NLOS scenario analyzed in Section III-B. Additionally, a 13.5 m wide truck parked on the opposite lane of the user generates extensive blockages for AP 2 and AP 3.

Fig. 3(b) shows the SINR as the UE moves trough the pedestrian lane. The colored backgrounds represent NLOS periods generated by the APs. The color of the SINR profile represents the AP where the UE is currently attached. When an obstacle blocks the current AP, an handover event is automatically triggered. However, in presence of short-term blockages, this decision is suboptimal as the connectivity is rapidly from one AP to the other with a consequent throughput decrease, analyzed in Fig. 3(b). The graph shows results for CUBIC and YeAH protocols, that are respectively throughput-performing and -conservative (see Fig. 2) with the RLC+RTO optimization. In presence of the first two consequent handover events, both protocols resume from slow start after the first handover. However, CUBIC thanks to its congestion control design, resumes quickly after the second handover and achieves the link rate. In contrast, YeAH achieves the link rate later, but it is more robust and during subsequent handover events prevents recovery from slow start that are common in CUBIC.

D. Multiple Concurrent Flows Analysis

Previous research has identified that different congestion control protocols significantly degrade responsiveness of short flows in presence of background traffic [15]. This section specifically analyzes performance of transport layer when short flows and background traffic coexist.

To study this behavior this scenario has 10 users with different applications and movements and 4 APs to provide coverage, as seen in Fig. 4. Table I shows the datails of those applications. UE0 has a throughput of 500 Mbps mimicking a FTP download, this flow will start at the beginning of the simulation and last until the end. The rest of the users are either streaming a video or using instant messaging applications. UEs 2, 4 and 5 are using streaming applications, they buffer data at 4 Mbps for 30 s and then stop their flow for 2 s before starting again. The rest of the users are using instant message applications or social networks, they send a few packages
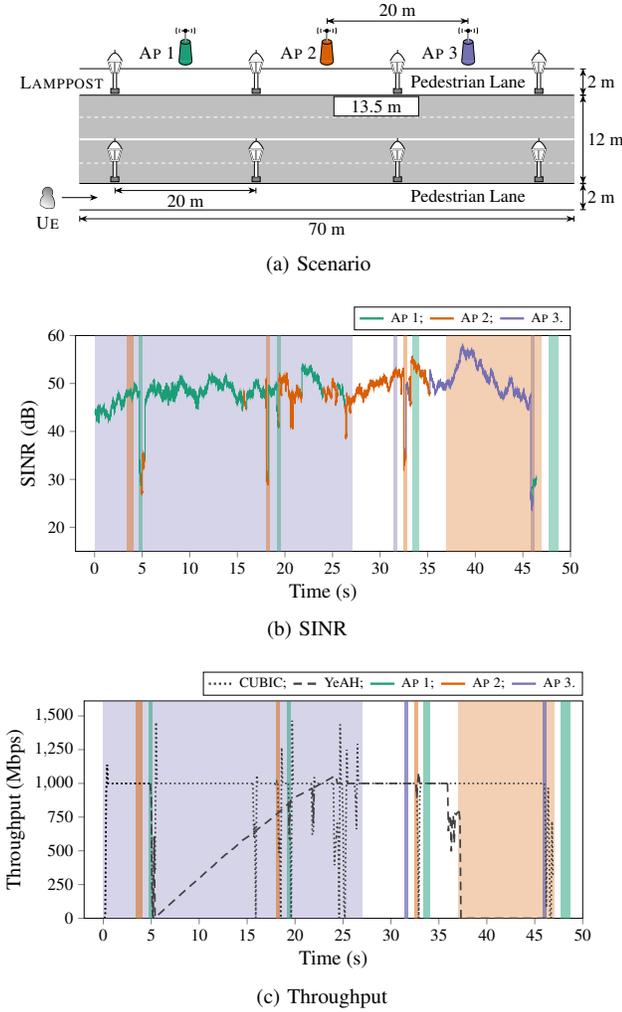
(a) Scenario

(b) SINR

(c) Throughput

Fig. 3. Handover analysis: scenario (a), SINR (b) and throughput (c)
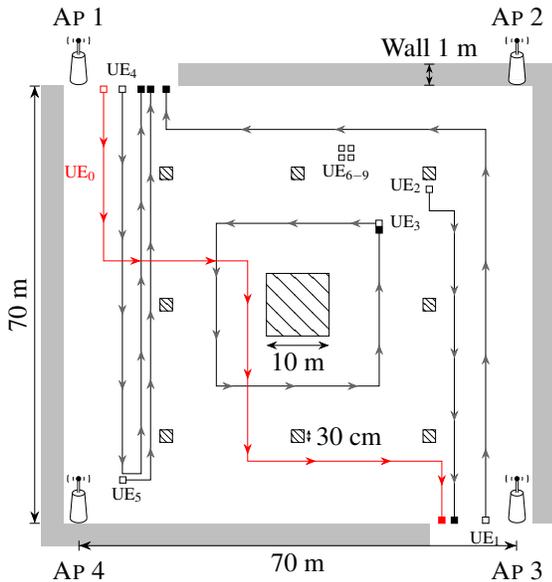

Fig. 4. System scenario

every 3 s. This applications are the most widely used in smartphones.

Except UEs $6 - 9$ all UEs experience handovers as they move on the scenario, some of them will move their active connection to an AP that is already serving other flows. All of the flows will then be affected as the AP has to serve all of them. Depending on the nature of the flow and the nature of the Tcp protocol used there will be packet losses, an increased buffer size and retransmissions.

TcpCubic and TcpYeah show different behaviors as shown in Fig. 2. TcpYeah has a much stable behavior in terms of RTT and throughput. This makes Yeah ideal for applications that do not require high datarates, such as instant messaging, but still experience short periods of NLOS. Cubic, on the other hand, has a higher variability in terms of RTT and throughput depending on the nature of the blockages encountered. Cubic uses up to five times the RLC buffer space than Yeah uses in the same scenario.

To verify how the network serves all the users, Fig. 5 shows the number of transmitted packets from each of the APs, which comprises both correctly received packets and retransmissions. High number of transmissions corresponds to packet losses caused by NLOS periods, handover or transmission errors. As YeAH relies on a slower recovery method than CUBIC, the network employs more APs and on average the users receive a higher number of packets.

Fig. 6 compares the number of correctly received packets at application layer when UE0 is present or not in the system. UE0's application is the one generating the heaviest flow, hence the experiment allows to verify the performance of medium-small flows with heavy background traffic. Additionally, the complexity of the scenario allows as well to verify the effect of medium flows on small ones. As UE0 moves through the environment, his connectivity moves from one AP to the other due to handover. As a result, the rate of the UEs that were already connected to those APs is degraded to accommodate the new high flow. The combined effect of blockages and presence of the long flow is different for medium and small flows. We first analyze medium ones. UE4, following a similar path of UE0, is the most affected as the number of packets received doubles when the UE0 is present for both congestion control protocol. In contrast, UE5 is served by different AP than UE0, hence CUBIC and YeAH perform similarly. The performance of static users UE6−9 is similar. They are affected by byt the medium-size flow of UE1 (see retransmissions from AP1 and AP2 in Fig. 5(b)) and partially from the long flow of UE0. It is worth nothing that CUBIC provides consistent results, while YeAH requires additional packets in presence of UE0 when served by AP1. Unlike UE6 − 9, UE3 persistently experiences NLOS transitions hence, its small flow is severely affected by the presence of UE0.

IV. CONCLUSIONS

In this article, we have studied the behavior of multiple congestion control protocols over mm-wave networks. In contrast to traditional mobile networks, mm-wave links have unique features, including high available bandwidth, high variability
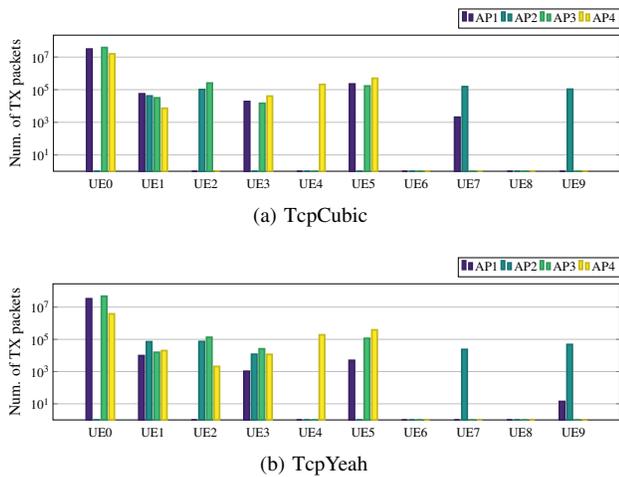
(a) TcpCubic
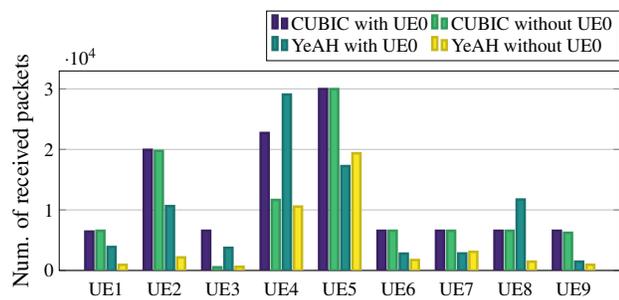

(b) TcpYeah

Fig. 5. Transmitted RLC packets by UE

Fig. 6. umber of received packets with and without UE0

in channel quality and sensitivity to blockages because of high propagation loss and unfavorable atmospheric absorption. Through extensive and accurate simulations, the paper has shown peculiar aspects of congestion control protocols by analyzing the impact on throughput and latency of various environments with blockages.

Proper setting of link-layer buffers is essential to alleviate the impact of losses on throughput although the delay increase and bufferbloat problem can arise. Active Queue Management can limit the issue, but is not considered in this paper and its implications left for future works. Given the short term variations of link quality, prompt reaction from slow start phase is crucial. This can be achieved shortening the duration of time outs. Congestion control protocols that aim at maximizing throughput through quick recovery like CUBIC and Scalable suffer considerably long NLOS periods while perform well in presence of short NLOS. On the other hand, the performance variations of hybrid congestion control protocols like YeAH is minimal. The tradeoff is especially evident in presence of handovers: while CUBIC resumes from slow start upon any handover event, YeAH probes bandwidth more gently and prevents short-term throughput variations. When long, medium and small flows coexists, CUBIC always require more transmissions than YeAH to successfully complete the data transmission. Medium flows are more affected by the combined effect of presence of long flow, blockages and handovers than small ones.

REFERENCES

[1] 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.

[2] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, Jun 2017.

[3] 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, Sep 2015.

[4] M. V. Perić, D. B. Perić, B. M. Todorović, and M. V. Popović, "Dynamic rain attenuation model for millimeter wave network analysis," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 441–450, Jan 2017.

[5] D. D. Donno, J. P. Beltran, and J. Widmer, "Millimeter-wave beam training acceleration through low-complexity hybrid transceivers," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3646–3660, Jun 2017.

[6] H. Shokri-Ghadikolaei, C. Fischione, P. Popovski, and M. Zorzi, "Design aspects of short-range millimeter-wave networks: A MAC layer perspective," *IEEE Network*, vol. 30, no. 3, pp. 88–96, May 2016.

[7] M. Zhang, M. Mezzavilla, R. Ford, S. Rangan, S. Panwar, E. Mellios, D. Kong, A. Nix, and M. Zorzi, "Transport layer performance in 5G mmWave cellular," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Apr 2016, pp. 730–735.

[8] M. Polese, R. Jana, and M. Zorzi, "TCP in 5G mmWave networks: Link level retransmissions and MP-TCP," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, May 2017.

[9] J. J. Alcaraz, F. Cerdan, and J. Garcia-Haro, "Optimizing TCP and RLC interaction in the UMTS radio access network," *IEEE Network*, vol. 20, no. 2, pp. 56–64, Mar 2006.

[10] Y. Zaki, T. Pötsch, J. Chen, L. Subramanian, and C. Görg, "Adaptive congestion control for unpredictable cellular networks," in *Proceedings of the ACM SIGCOMM Conference*.   ACM, 2015, pp. 509–522.

[11] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. M. Mao, S. Sen, and O. Spatscheck, "An in-depth study of LTE: Effect of network protocol and application behavior on performance," in *Proceedings of the ACM SIGCOMM Conference*.   ACM, 2013, pp. 363–374.

[12] C. Callegari, S. Giordano, M. Pagano, and T. Pepe, *A Survey of Congestion Control Mechanisms in Linux TCP*.   Springer International Publishing, 2014, pp. 28–42.

[13] T. Azzino, M. Drago, M. Polese, A. Zanella, and M. Zorzi, "X-TCP: a cross layer approach for TCP uplink flows in mmwave networks," in *16th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, Jun 2017, pp. 1–6.

[14] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5G mmwave mobile networks," *To appear in IEEE Journal on Selected Areas in Communications*, June 2017.

[15] S. Alfredsson, G. D. Giudice, J. Garcia, A. Brunstrom, L. D. Cicco, and S. Mascolo, "Impact of TCP congestion control on bufferbloat in cellular networks," in *IEEE 14th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Jun 2013, pp. 1–7.