IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 23, 2019, accepted December 7, 2019, date of publication December 12, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959134

On the Throughput Performance of TCP Cubic in Millimeter-Wave Cellular Networks

HOANG D. LE[1], (Student Member, IEEE), CHUYEN T. NGUYEN[2], VUONG V. MAI[3], (Member, IEEE), AND ANH T. PHAM[1], (Senior Member, IEEE)

[1]Computer Communications Laboratory, The University of Aizu, Aizuwakamatsu 965-8580, Japan
[2]School of Electronics and Telecommunications, Hanoi University of Science and Technology, Hanoi 100000, Vietnam
[3]Photonics Systems Research Laboratory, KAIST, Daejeon 34141, South Korea

Corresponding author: Hoang D. Le (hoangbkset@gmail.com)

This work was supported in part by The University of Aizu Competitive Fund. The work of H. D. Le was supported by the Japan Student Services Organization (JASSO) Scholarship.

ABSTRACT In this paper, we study a cross-layer analysis framework for the performance evaluation of TCP over millimeter-Wave (mmWave) fading channels in the fifth-generation (5G) cellular networks, when the truncated incremental redundancy hybrid automatic repeat request (IR-HARQ) scheme and adaptive modulation and coding (AMC) are employed. Specifically, the throughput performance of TCP Cubic, which is one of the most widely deployed TCP variants in the Internet, is investigated. For this purpose, the mmWave fading channel, approximated by Nakagami-*m* distribution, is captured by a finite-state Markov chain (FSMC) to develop a transmission loss model. A loss-based TCP model, which is analyzed based on the transmission loss model, is then used to analytically derive the TCP throughput performance. The numerical results quantitatively show the effect of different parameters/settings of AMC, IR-HARQ, blockages and mmWave fading channels on the TCP performance and support the optimal selection of parameters to maximize the system throughput. Monte Carlo simulations are also performed to validate the analytical results.

INDEX TERMS 5G mmWave networks, adaptive modulation and coding (AMC), nakagami-*m* fading channels, random shape blockage model, Markov chain, hybrid ARQ (HARQ), TCP Cubic.

I. INTRODUCTION

Recent years have witnessed a rapid proliferation of the mobile Internet data traffic due to the spread of mobile devices and high speed mobile terminals. The upcoming fifth-generation (5G) cellular networks are expected to achieve 1000 times of network capacity, 10 times of spectral/energy efficiency, 10-100 times of data rate and 25 times of average cell throughput, compared to the current fourth-generation (4G) ones [1]. It is envisioned that the 5G wireless networks will be deployed in the early 2020s with gigabits per second (Gbps) peaks in most favorable conditions, ultra-low end-to-end latency (possibly below 10 ms), and high reliability [2]. Millimeter-wave (mmWave) communications, which have been recently gained much attention from both academia and industry, are a possible enabler for high speed connectivity envisioned in future 5G networks [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Cristina Rottondi.

At mmWave frequencies (i.e., above 10 GHz), indeed, they can provide very high data rate transmission thanks to the massive available bandwidth that can be allocated to the cellular networks. The signal propagation at mmWave frequencies is, nevertheless, much more sensitive to blockage effects than the one at lower frequencies, in which mmWave signals suffer from poor penetration, diffraction, and scattering through the materials [4]. Furthermore, mmWave communications are also vulnerable to the high error rate environment due to the impact of the fading channels. These pose various challenges not only on the performance of physical layer (PHY), but also on the performance of various upper-layer protocols [5].

In the transport layer, transmission control protocol (TCP) is the most widely used protocol for a variety of different applications that require the reliable transmissions in the Internet, such as HTTP, Skype, file transfer, and email [6], [7]. The most critical issue in TCP is to design a suitable congestion control mechanism that can effectively utilize the channel resource. Congestion control mechanism

in the conventional TCP variants, such as Tahoe, Reno, New-Reno, and SACK, is mainly based on the standard additive increase multiplicative decrease (AIMD) algorithm [8], [9]. For example, TCP NewReno, which is the TCP variant for majority of Internet applications, increases its congestion window (cwnd) by one segment after reception of acknowledgement (ACK) in congestion avoidance phase and halves the current cwnd value whenever a loss is detected (e.g., three duplicate ACK or timeout). The AIMD, however, becomes ineffective in high bandwidth-delay product (BDP) networks, due to the slow linear window growth function [10].

Over the past decade, there have been several TCP variants which are designed specifically for BDP networks, such as TCP WestwoodPlus [11], TCP Cubic [10], TCP YeAH [12], and TCP BBR [13]. Among them, TCP Cubic is one of the most popular implementations, and in fact, it is used by default in Linux kernels 2.6.19 and above, as well as Windows 10, and Windows Server 2016 [14]. TCP Cubic increases its cwnd accordingly to a cubic function by computing the absolute time since the last dropped segments instead of round-trip time (RTT). As a result, the sending rate is independent of RTT resulting in the better fairness between TCP flows. Furthermore, Cubic increases its window size more aggressively when it is far from the previous saturation point, and more slowly when it is close which allows the network to stabilize before looking for more bandwidth. Recent researches nevertheless reveal that the TCP performance can be severely degraded in high-error-rate environment of mmWave fading channels (i.e., degraded TCP throughput and very low utilization of the resources at mmWave frequencies) which poses formidable challenges to maintain the TCP reliability in 5G mmWave networks [15].

A. RELATED WORKS AND MOTIVATIONS

In the domain of 5G cellular networks, several studies have been recently devoted to the evaluation of the TCP performance over mmWave fading channels [15]–[21]. Particularly, Zang et al., presented the first performance evaluation of TCP congestion control over mmWave channels by using network simulator version 3 (NS-3) in [15]. The authors in [16] proposed a novel cache-enabled TCP framework called mmWave TCP to cope with the transition between line-of-sight (LoS) and non-LoS (NLoS) of mmWave channels. In this framework, the TCP segments likely to be lost during NLoS are simultaneously retransmitted (i.e., called the bath retransmission) once the channel becomes LoS which improves the performance of TCP over mmWave channels. In [17], TCP throughput was investigated under impact of mmWave fading channels, modeled by Nakagami-m distribution. The throughput performance of multi-path TCP (MTCP) over mmWave channels was analyzed when 5G new radio and LTE communication paths are aggregated in [18], [19]. The authors in [20] showed that the performance of TCP Cubic outperforms TCP YEAH in terms of achieved throughput in 5G mmWave mobile networks. In [21], the authors provided a comprehensive simulation study of TCP considering various factors such as the congestion control algorithms, including the recently TCP BBR, edge vs. remote sever, handover and multi-connectivity, TCP segment size, and 3GPP-stack parameters. They showed that the performance of TCP on mmWave links is highly dependent on different combinations of these factors, and identified the open challenges in the area of 5G mmWave cellular networks.

On the other hand, it is important to note that most existing studies of TCP performance over 5G mmWave cellular networks were often based on simulations (i.e., using well-known simulator NS-3) which allows the accurate evaluation of a wide range of network protocols. Another approach to study the performance of TCP is to use the analytical models. The key advantage of the analytical approach is the provision of a good approximation of network performance with various channel conditions in a relatively quick fashion which supports for the optimization purposes. Using the analytical approach, we can also study different analytical models for both physical and link layers, which may have not updated yet in the simulation. Analytical approach for TCP performance has been therefore intensively studied for wireless cellular networks [22]–[24]. Such study is however not available for TCP over 5G mmWave cellular networks. We have to point out that the links using mmWave frequencies have different propagation characteristics as compared to typical cellular systems operating at sub-6 GHz frequency bands. For example, the link quality in mmWave band can be significantly reduced due to the blockage effect which is not observed in sub-6 GHz frequency bands. Over such intermittent mmWave links, TCP sender may suddenly drop the sending rate while there is no actual congestion, leading to the degradation of TCP throughput performance. In addition, it is noted that the interaction between the design of link-layer and physical layer plays an important role to improve the TCP performance. The TCP performance improvement considering the cross-layer design has not been well investigated in the existing studies using simulations.

B. CONTRIBUTIONS

The main objective of this paper is to develop an analytical framework for the cross-layer evaluation and performance improvement of TCP connections over 5G mmWave networks. The contributions of the paper can be summarized as follows:

- First, we provide an analytical framework for the performance evaluation of TCP under impact of mmWave fading channels and blockages caused by buildings in 5G cellular networks. In particular, the well-known TCP variant, TCP Cubic, whose the simulation performance was previously studied in 5G mmWave mobile networks [20], is considered. The mmWave fading channel is modeled by Nakagami-m distribution, which can be used to model a broad class of fading characteristics,

including both mmWave LoS and NLoS links. In addition, the blockages are reflected by the random shape model in which the buildings are modeled as the random rectangles.

- Second, we develop a cross-layer design to improve the performance of TCP over mmWave fading channels, in which both truncated incremental redundancy hybrid automatic repeat request (IR-HARQ) protocols based error control at the link layer and the adaptive modulation and coding (AMC) scheme at the physical (PHY) layer are employed for the radio access. These are the broadly-used adaptation technologies to improve the system performance under impact of wireless fading phenomena [25]. It is, therefore, believed that their interactions with the transport-layer TCP protocols could play an important role in improving the TCP performance. Moreover, it is worth mentioning that the understanding of combined effects of IR-HARQ and AMC on the TCP performance would be critical in the design and optimization of system and protocol parameters in 5G mmWave networks. Previous studies, however, focused solely on the performance of TCP over 5G networks when either HARQ [18] or AMC [26] is considered in the simulation analysis.
- Third, the mmWave fading channel is captured by a finite-state Markov chain (FSMC) to develop a transmission loss model. In fact, in mmWave fading channels, because the fading coherence time is of the order of hundreds of micro-second [5], the frame losses often occur in a burst pattern during the transmission [27], i.e., the loss structure is correlated. It is, therefore, extremely meaningful to investigate the correlated loss transmission structure to provide more accurate performance analysis instead of using the conventional uniform loss model in which the loss structure of different frame transmissions are assumed to be independent. Then, a loss-based TCP model, which is analyzed from the transmission loss model, is used to analytically derive the TCP throughput.
- Finally, we employ the proposed framework to evaluate and optimize TCP performance through numerical investigations. Particularly, we quantify the impact of different parameters/settings of HARQ, AMC, blockages and mmWave links on the TCP throughput performance, and discuss several optimization aspects. We also validate the analytical results by Monte Carlo simulations.

C. ORGANIZATION

The rest of paper is organized as follows. The network description is presented in Section II. The mmWave fading channel model and the channel-state modeling are introduced in Section III. The transmission loss model, loss-based TCP model, and throughput analysis are shown in Section IV. Numerical results are given in Section V. Finally, Section VI concludes the paper.

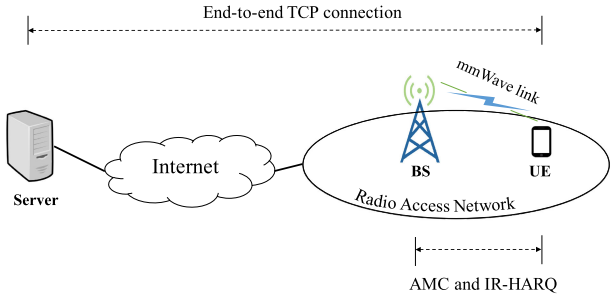

FIGURE 1. Network scenario with end-to-end TCP connection.

II. NETWORK DESCRIPTIONS

A. NETWORK MODEL

We consider a 5G network scenario where a user equipment (UE) is downloading data from a remote sever (TCP source) through a cellular base station (BS) using TCP Cubic connection as illustrated in Figure 1. The BS, which provides service to the UE, is connected to the remote server via high-speed networks (Internet). The radio access networks (RAN) are assumed to be the bottlenecks in which the mmWave wireless link is used for the connection between BS and UE. In addition, HARQ and AMC are employed at RAN to further improve the system performance. The purpose of AMC is to maximize the data rate over the mmWave fading channels while satisfying a predefined quality of service (QoS), i.e., a targeted frame loss rate denoted by P_{loss}. It allows the system to choose the most appropriate modulation and coding scheme (MCS) depending on the channel quality (i.e., received power). We adopt the M_{Q}-aray quadrature amplitude modulation (M_{Q}-QAM) schemes which are supported by LTE and 5G networks as in [28]–[30], with a fixed symbol rate of R_s for the set of possible N MCS levels, i.e., $\{1, 2, \cdots, N\}$. MCS, which corresponds to a particular combination of modulation and coding strategies, is adjusted by the BS based on the channel quality indicator (CQI) through assumed free-error feedback channels. The bit rate and information rate change for every MCS and are given as $R_b = R_s \log_2(M_{\text{Q}})$ (bits/sec) and $R_n = \log_2(M_{\text{Q}})$ (bits/symbol), respectively, where M_{Q} is the signal constellation size.

At the physical (PHY) layer, data are transmitted block by block over mmWave link modeled by Nakagami-m fading channel, which is widely used for mmWave channels [17]. The block duration is designed to be shorter than the coherence time of mmWave fading channels, so that the channel remains invariant during a block transmission. Each block contains multiple link-layer data frames depending on MCS in which the number of frames per block at MCS n is given as bR_n, where R_n is the information rate used by MCS n, and b is a parameter to be set by the designer. Given the size of frames of N_f bits, the block duration is defined as $T_b = (bN_f)/R_s$, where R_s is the data symbol rate.

In the link layer, truncated IR-HARQ protocol is used in the mmWave links between the BS and UE to tackle

with the frame error issue due to the fading phenomena. The use of HARQ is to correct and retransmit (in case of being uncorrectable) the corrupted frames. For the sake of simplicity, we assume that one TCP segment is encapsulated in one IP packet and then is mapped into one link-layer frame (i.e., without fragmentation) resulting in the probabilities of frame loss and segment loss are the same. In the transport layer, TCP Cubic is employed. The window size of this TCP variant is a Cubic function of time which depends on loss events in the congestion avoidance phase.[1] The TCP loss events are assumed to be caused by two factors: *congestion loss* and *random loss*. Here, the congestion loss occurs when the window size attains the maximum value of W at *bottleneck* link while random loss is caused by the building blockages and the fading phenomena at the mmWave wireless link.

B. IR-HARQ WITH AMC

In general, HARQ protocol can be classified into three main types: HARQ type I, chase combining (CC), and incremental redundancy (IR). Among them, IR-HARQ is the most efficient one in term of throughput performance [25]. In this paper, we therefore consider the truncated IR-HARQ which combines the sliding window ARQ and rate-compatible punctured convolutional (RCPC) code, where the maximum number of retransmission for one frame is denoted as N_{re}. RCPC codes are used to correct the erroneous frames by using a puncturing convolutional encoder/Viterbi decoder. In addition to RCPC code, the operation of sliding window ARQ protocol is for the detection of the corrupted frames which are uncorrectable by RCPC code, and the retransmission of additional redundancy for these frames. For the sake of simplicity, the frame error is assumed to be always detectable by the standard cyclic redundancy check (CRC) and the feedback signal (either ACK or NAK) transmission is also supposed to be error-free.

The operation of IR-HARQ with AMC is applied to each block consisting of multiple frames depending on MCS levels. Due to using N MCS levels, we have a family of RCPC codes containing N different coding rates, i.e., $(1 \geq) R_c^{(1)} > R_c^{(2)} > \cdots > R_c^{(N)}$, which are obtained from the mother code rate $R_c^{(N)}$ (e.g., 1/2 or 1/3). Each N_f-bit information frame is encoded by the original rate $R_c^{(N)}$ encoder, and after puncturing, the system transmits $N_c^{(n)} = N_f / R_c^{(n)}$ coded bits per frame when using MCS n. For the sake of simplicity, we assume that a frame transmission attempts complete in a single MCS level of AMC.

At a specific MCS n with a coding rate of $R_c^{(n)}$, we have different puncturing patterns for each transmission attempt, i.e., $\tilde{C}_1, \tilde{C}_2, \cdots \tilde{C}_{N_{\text{re}}+1}$ corresponding to $N_{\text{re}}+1$ transmission attempts for a frame. Table 1 shows an example of different puncturing patterns for each transmission attempt of a frame,

[1]In this paper, we are interested in the performance of TCP long-lived flows, e.g., file transfers and video streaming. We therefore ignore the slow start phase which does not impact TCP throughput for long-lived flows.

TABLE 1. Example of puncturing patterns for IR-HARQ.

Code rate	1/2	2/3	3/4	5/6
Pattern 1, $\tilde{C}_1$	1 1	10 11	101 110	10101 11010
Pattern 2, $\tilde{C}_2$	1 1	01 11	011 101	01011 10101
Pattern 3, $\tilde{C}_3$	1 1	10 11	110 011	10110 01011

given the $N_{\text{re}} = 2$. For the first transmission, the frames containing only the pattern coded bits of $\tilde{C}_1$ are selected to be transmitted in a block. At the receiver side, after decoding for error correction using the Viterbi algorithm with pattern coded bits of $\tilde{C}_1$, error detection using CRC is performed. In case of transmission failure, IR-HARQ retransmits the additional redundancy for all the uncorrectable frames in a block based on a NAK message from the receiver. Accordingly, the incremental redundancy corresponding to pattern coded bits of $\tilde{C}_2$ is transmitted. The new received redundancy is combined with the previous received pattern coded bits to construct a more powerful code for decoding. This process for the retransmission of a frame is continued until it is decoded successfully or reach the maximum number of retransmissions N_{re}. If a frame does not get through the mmWave link after $N_{\text{re}}+1$ transmission attempts, IR-HARQ gives up and the frame is clarified to be lost.

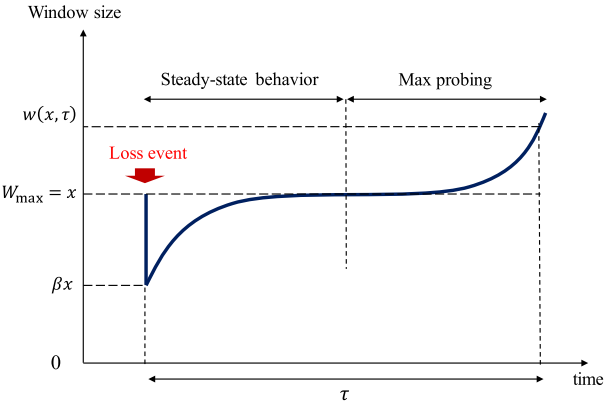

FIGURE 2. TCP Cubic congestion window behavior.

C. TCP CUBIC

The operation of TCP Cubic can be described via its congestion control mechanism of the window size as illustrated in Fig. 2. The are two states in TCP Cubic: the steady state and the max probing state. When a loss event (which might be either a congestion loss or a random loss) occurs, TCP Cubic reduces its window size by a multiplicative decrease factor denoted by β and enters the "steady state". The window size just before the reduction is set to the maximum value, $W_{\max}$. During the steady state, TCP Cubic increases its window size rapidly when the window value is far from $W_{\max}$, and then slows down when it is close to that point. This allows TCP Cubic to be very scalable in high bandwidth connectivity and

long delay networks. If the window size reaches a current maximum level, TCP cubic enters the "max probing" state. In this state, the window size increases slowly initially to find the new maximum value nearby (i.e., at a point having loss event). Then, TCP Cubic expects that the maximum point is further away and the congestion window starts growing very fast until the next loss event.

The window growth function of TCP Cubic is expressed as

$$w(x,\tau) = \alpha\,(\tau - K_{\text{cubic}})^3 + x, \tag{1}$$

where α is a Cubic factor, τ is the elapsed time from the last window reduction, β is the multiplicative decrease factor, and x is the window size just before the last window reduction [31]. $\bar{K}$ is the time period needed to increase the window size from βx to x, and can be expressed by,

$$K_{\text{cubic}} = \sqrt[3]{\frac{(1-\beta)\,x}{\alpha}}. \tag{2}$$

In addition, let $T_{x,y}$ denote the time duration that window size increases from βx to y without any loss event. Based on (1), $T_{x,y}$ can be expressed by

$$T_{x,y} = \sqrt[3]{\frac{y-x}{\alpha}} + \sqrt[3]{\frac{(1-\beta)x}{\alpha}}. \tag{3}$$

III. CHANNEL AND CHANNEL-STATE MODELS

The operation of TCP Cubic is modeled in real-time and depends on the loss events (i.e., caused by building blockages and fading phenomenon). To investigate the performance of TCP Cubic under impact of building blockages and fading phenomenon, it is necessary to capture the statistics of the building blockages and the time-varying behavior of mmWave fading channels to model the behavior of transmission loss. To do so, we first review the blockage and mmWave fading channel models. Then, the channel-state model is developed to facilitate the operation of AMC with IR-HARQ under impact of fading channels.

A. BLOCKAGE MODEL

To investigate the effect of blockages on the system performance, we use the random shape model as illustrated in Fig. 3,

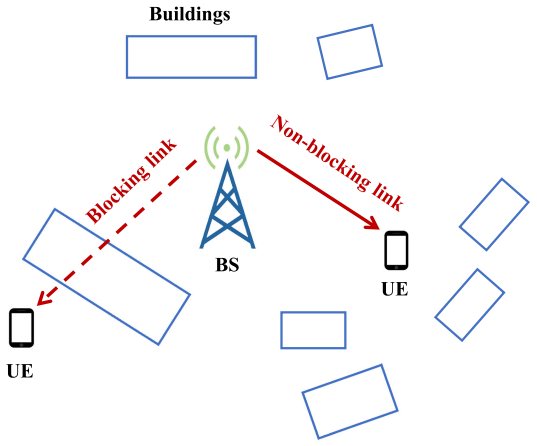

FIGURE 3. Random shape blockage model.

in which the link between the BS and UE might be blocked by the buildings. In this model, the buildings of random locations, orientations, and sizes in urban areas are modeled as a random process of rectangles, which is called a rectangle Boolean scheme in the random shape theory. The probability of a particular link being blocked is given as,

$$p_{\text{blocking}} = 1 - \exp(-\beta_{\text{B}} d_{\text{UE}}), \tag{4}$$

where d_{UE} is the distance from UE to BS and β_{B} is the blockage parameter computed using statistics of the buildings [32]. The blockage parameter β_{B} is determined by

$$\beta_{\text{B}} = \frac{\chi \ln(1-\kappa)}{A\pi}, \tag{5}$$

where χ is the average building perimeter, κ is the percentage of area covered by buildings, and A is the average area of the buildings in the considered region [33]. The blockage parameter values measured at some regions, such as Chicago, US ($\beta_{\text{B}} = 0.022$), LU, U.K. ($\beta_{\text{B}} = 0.0057$), and NC, Pakistan ($\beta_{\text{B}} = 0.0014$) are available in [33] and [34], respectively.

B. MMWAVE FADING CHANNEL MODEL

The mmWave fading channel can be well characterized by the Nakagami-m distribution [17], [35], [36]. This model represents a wide variety of realistic line-of-sight (LOS), and non-line-of-sight (NLoS) fading channels via varying the parameter m (high values of m indicates the LOS) encountered in practice [35]. The probability density function (PDF) of received signal-to-noise ratio (SNR), γ, is given as

$$f_\gamma(\gamma) = \frac{m^m \gamma^{m-1}}{\overline{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\overline{\gamma}}\right), \quad \gamma \geq 0 \tag{6}$$

where $\overline{\gamma}$ is the average SNR, $\Gamma(m) := \int_0^\infty t^{m-1} e^{-t} dt$ is the Gamma function, and m is the Nakagami factor [37]. Note that, for $m = 1$, the Nakagami distribution reduces to Rayleigh distribution, and for $m = (\mathfrak{K}+1)^2/(2\mathfrak{K}+1)$, the Nakagami distribution is approximately Rice fading with parameter $\mathfrak{K}$, then we use the values of $m < 1$ for NLoS and $m > 1$ for LoS links [36]. Futhermore, its cumulative distribution function (CDF) is also expressed as

$$F_\gamma(\gamma) = \frac{1}{\Gamma(m)} \Gamma\left(m, \frac{m\gamma}{\overline{\gamma}}\right), \tag{7}$$

where $\Gamma(m,x) := \int_x^\infty t^{m-1} e^{-t} dt$ is the complementary incomplete Gamma function [38].

In addition, the level crossing rate (LCR), which is defined as the average number of times per second that the received SNR passes the certain threshold γ_{th} in the positive direction only (or in the negative direction only), can be written as

$$\text{LCR}(\gamma_{\text{th}}) = \sqrt{2\pi \frac{m\gamma_{\text{th}}}{\overline{\gamma}}} \frac{f_d}{\Gamma(m)} \left(\frac{m\gamma_{\text{th}}}{\overline{\gamma}}\right)^{m-1} \exp\left(-\frac{m\gamma_{\text{th}}}{\overline{\gamma}}\right), \tag{8}$$

TABLE 2. Example of modulation and coding schemes (MCS).

	MCS 1	MCS 2	MCS 3	MCS 4	MCS 5
Modulation	BPSK	QPSK	QPSK	16-QAM	16-QAM
Code rate $R_c^{(n)}$	1/2	2/3	5/6	2/3	5/6
R_n (bits/symbol)	1/2	4/3	5/3	8/3	10/3
$a_{n,1}$	4447.4	2068.5	514.7	850.9	142.9
$a_{n,2}$	2298.6	1344.3	3297.8	372.3	895.7
$a_{n,3}$	5944.9	1428.4	7247.9	3567.3	1057.1
$g_{n,1}$	11.104	3.315	1.759	0.816	0.339
$g_{n,2}$	21.012	6.997	6.196	1.895	1.657
$g_{n,3}$	34.203	10.569	10.224	3.378	2.706
$\gamma_{p_{n,1}}$ (dB)	-1.212	3.623	5.502	9.172	11.660
$\gamma_{p_{n,2}}$ (dB)	-4.337	0.127	1.164	4.946	6.131
$\gamma_{p_{n,3}}$ (dB)	-5.950	-1.629	-0.608	3.841	4.105

where f_d is the maximum Doppler frequency which is the function of mmWave frequency and UE speed [39].

C. CHANNEL-STATE INDUCED BY AMC

We assume that there are N MCS modes of AMC transmission. To facilitate the data transmission using AMC, the channel is modeled by multiple states defined by a range of received signal-to-noise (SNR). Under the assumption of constant power transmission, the entire received SNR range is partitioned into $N+1$ non-overlapping consecutive intervals with boundary points denoted as $\{\gamma_n\}_{n=0}^{N+1}$. The channel is said to be in the n-th state if the received SNR falls into the interval of $[\gamma_n, \gamma_{n+1})$, where $n \in 0, 1, \cdots, N$. To avoid a high frame error rate, no transmission is allowed when $\gamma \in [\gamma_0, \gamma_1)$ (channel-state 0-*th*) while MCS n is assigned for n-th channel-state. The selection of the SNR thresholds satisfies the condition that the frame loss rate (FLR) for each channel-state is exactly the predefined target FLR, denoted by P_{loss}. Using the truncated IR-HARQ with N_{re} retransmissions, the frame is considered to be lost after $N_{\text{re}}+1$ transmission attempts, and the FLR at MCS n, then, can be calculated as

$$\text{FLR}_n = \prod_{k=1}^{N_{\text{re}}+1} \text{FER}_{n,k} := P_{\text{loss}}, \tag{9}$$

where $\text{FER}_{n,k}$ is the instantaneous FER at the output of the soft Viterbi decoder when using MCS n on the k-th transmission attempt of a frame which can be approximated as

$$\text{FER}_{n,k}(\gamma) \approx \begin{cases} 1, & 0 \le \gamma < \gamma_{n,k}, \\ a_{n,k} \exp(-g_{n,k}\gamma), & \gamma \ge \gamma_{n,k}. \end{cases} \tag{10}$$

where $a_{n,k}$, $g_{n,k}$, and $\gamma_{n,k}$ given in Table 2, are the curve-fitting parameters for MCS n and k-th transmission attempt [40]. These parameters have been obtained by least-square fitting the approximate expression of the FER at the output of the soft Viterbi decoder in [41] to the curves obtained via the simulation (i.e., as the ratio between the erroneously transmitted frames after k-th transmission attempt using MCS n and the overall number of transmitted frames. For details, refer to [42] in which the accuracy of the curve-fitting approximation is verified. In addition, we set

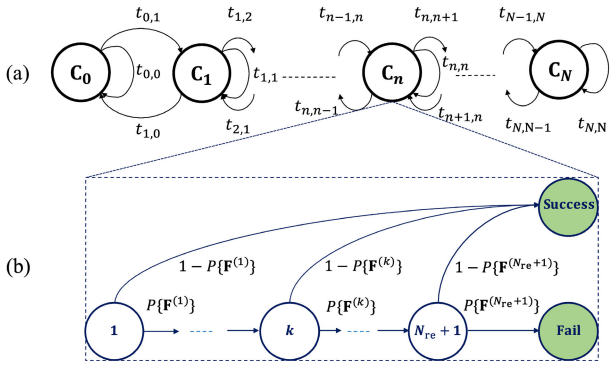

FIGURE 4. Finite-state Markov chain for (a) channel model, and (b) for the operation of IR-HARQ.

the SNR thresholds $\gamma_0 = -\infty$, $\gamma_{N+1} = +\infty$, and by incorporating $\text{FER}_{n,k}$ (in Eq. (10)) into (9), we can express the other SNR thresholds as follows

$$\gamma_n = \frac{1}{\sum_{k=1}^{N_{\text{re}}+1} g_{n,k}} \ln\left(\frac{\prod_{k=1}^{N_{\text{re}}+1} a_{n,k}}{P_{\text{loss}}}\right), \quad \text{for } 1 \le n \le N, \tag{11}$$

Given γ_n, the steady-state probability of the n-th state, denoted by Pr_n, can be expressed as

$$\text{Pr}_n = \int_{\gamma_n}^{\gamma_{n+1}} f_\gamma(\gamma)\, d\gamma = F_\gamma(\gamma_{n+1}) - F_\gamma(\gamma_n), \tag{12}$$

where $f_\gamma(\gamma)$ and $F_\gamma(\gamma)$ are the PDF and CDF of the channel SNR, as in Eqs. (6) and (7), respectively. A finite-state Markov channel (FSMC) model then can be used to describe the transition between channel states as shown in Fig. 4(a). Here, the slow-fading condition is assumed and the transition happens between adjacent channel states in which the transition probabilities from n-th channel-state to l-th channel-state denoted by $t_{n,l}$ can be determined by

$$\begin{aligned} t_{n,n+1} &= \frac{\text{LCR}(\gamma_{n+1})\, T_b}{\text{Pr}_n}, \quad \text{if } n = 0, \cdots, N, \\ t_{n,n-1} &= \frac{\text{LCR}(\gamma_n)\, T_b}{\text{Pr}_n}, \quad \text{if } n = 1, \cdots, N+1, \\ t_{n,l} &= 0, \quad \text{if } |l-n| \ge 2. \end{aligned} \tag{13}$$

The probability of staying at the same channel-state n is

$$t_{n,n} = \begin{cases} 1 - t_{n,n+1} - t_{n,n-1}, & \text{if } 0 < n < N, \\ 1 - t_{1,2}, & \text{if } n = 0, \\ 1 - t_{N+1,N}, & \text{if } n = N, \end{cases} \quad (14)$$

where T_b is the block duration and LCR is given in (8) [43].

In addition, the average frame error rate for the k-th transmission attempt in MCS n can be calculated by

$$\begin{aligned} &\overline{\text{FER}}_{n,k} \\ &= \frac{1}{\Pr_n} \int_{\gamma_n}^{\gamma_{n+1}} \text{FER}_{n,k} f_\gamma(\gamma)\, d\gamma \\ &\simeq \frac{1}{\Pr_n} \int_{\gamma_n}^{\gamma_{n+1}} a_{n,k} \exp\left(-g_{n,k}\gamma\right) f_\gamma(\gamma)\, d\gamma \\ &= \frac{1}{\Pr_n} \frac{a_{n,k}}{\Gamma(m)} \left(\frac{m}{\overline{\gamma}}\right)^m \frac{\Gamma\left(m, b_{n,k}\gamma_n\right) - \Gamma\left(m, b_{n,k}\gamma_{n+1}\right)}{\left(b_{n,k}\right)^m}, \end{aligned} \quad (15)$$

where m is the Nakagami factor and $b_{n,k} := \frac{m}{\overline{\gamma}} + g_{n,k}$ [40]. Given $\overline{\text{FER}}_{n,k}$, the average FER of the k-th transmission attempt is defined as the average number of incorrectly received frames over the total average number of transmitted frames, i.e.,

$$\overline{\text{FER}}_k = \frac{\sum_{n=1}^{N} R_n \Pr_n \overline{\text{FER}}_{n,k}}{\sum_{n=1}^{N} R_n \Pr_n}, \quad (16)$$

where R_n is the information rate of MCS n. As it is assumed that a period of channel-state can cover all possible transmission attempts of a frame, for each channel-state n-th in Fig. 4(a), we develop a simple Markov model for the operation of IR-HARQ as shown in Fig. 4(b). Let $\mathbf{F}^{(k)}$ denote the event "*decoding failure*" for k-th transmission attempt of a frame. The average number of transmission attempts for a frame can be determined by

$$\begin{aligned} \overline{K} = &\left(1 - P\{\mathbf{F}^{(1)}\}\right) + 2\, P\{\mathbf{F}^{(1)}\}\left(1 - P\{\mathbf{F}^{(2)}\}\right) \\ &+ \cdots + (N_{\text{re}} + 1)\, P\{\mathbf{F}^{(1)}\} P\{\mathbf{F}^{(2)}\} \cdots P\{\mathbf{F}^{(N_{\text{re}})}\}, \end{aligned} \quad (17)$$

where N_{re} is the maximum number retransmissions for a frame and $P\{\mathbf{F}^{(k)}\}$ is the probability of decoding failure of k-th transmission attempt in which $P\{\mathbf{F}^{(k)}\} \simeq \overline{\text{FER}}_k$. By rearranging the terms in (17), we have

$$\overline{K} = 1 + \sum_{h=2}^{N_{\text{re}}+1} \prod_{k=1}^{h-1} \overline{\text{FER}}_k. \quad (18)$$

IV. PERFORMANCE ANALYSIS

This section focuses on the throughput performance analysis of TCP Cubic with IR-HARQ and AMC over mmWave fading channels. In particular, a transmission loss model is developed from the above channel-state model due to using the truncated IR-HARQ. A loss-based TCP model, which is based on the transmission loss model, is then used to analytically derive the TCP throughput performance.

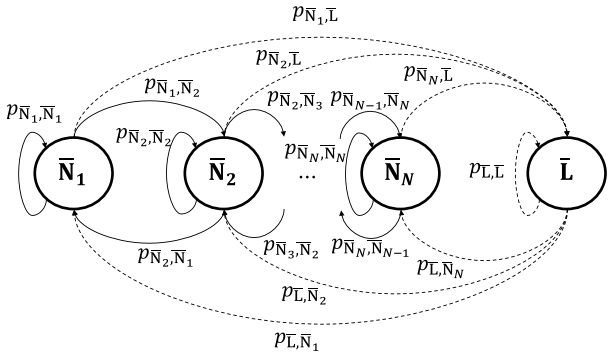

FIGURE 5. Transmission Loss model.

A. TRANSMISSION LOSS MODEL

The loss transmission behavior can be modeled with loss ($\bar{\text{L}}$) and non-loss ($\bar{\text{N}}$) states. By using the AMC, non-loss states contain N substates corresponding to N different MCS levels, i.e., $\bar{\text{N}}_n$, $n \in \{1, \ldots, N\}$ as shown in Fig. 5. In this structure, the system (in a specific MCS level) is in a loss state when a block loss happens, i.e., either one of frames in a block is uncorrectable by RCPC codes after reaching the maximum number of retransmissions N_{re} or the transmission is blocked by the buildings due to the fact that the temporal duration of blockage is order of hundreds of milliseconds [44]. Otherwise, the system is in non-loss state, i.e., no frames in block are lost. The block-loss probability at the nth channel-state, denoted as BLP_n, then can be given as

$$\text{BLP}_n = \left(1 - p_{\text{blocking}}\right)\left[1 - (1 - \overline{\text{FLR}}_n)^{n_f}\right] + p_{\text{blocking}}, \quad (19)$$

where p_{blocking} is the probability of a particular link being blocked which is determined in Eq. (3), n_f is the number of frames per block, and $\overline{\text{FLR}}_n$ is the frame loss rate at channel-state n-th in which the frame is still detected in error after N_{re} retransmissions and is determined by Eq. (16) when $k = N_{\text{re}} + 1$.

Here, we use only one state as the common loss state for all channel-state to reduce the complexity of the transmission loss model, thus simplifying the performance analysis. As a result, transmissions might occur between the loss state to any non-loss state. In addition, the transition within non-loss states is supposed to only occur between adjacent ones due to assumption of a slow-fading channel. If we denote the state of loss model by s_n, where $s_n \in \{\bar{\text{L}}, \bar{\text{N}}_n\}$, the transition probabilities of the states in transmission loss model, which

is shown in Fig. 5, can be calculated as

$$p_{s_n,s_l} \triangleq \Pr\{s_n \to s_l\}, = \frac{\Pr\{s_l \text{ at time } t+1, s_n \text{ at time } t\}}{\Pr\{s_n \text{ at time } t\}}, \quad (20)$$

Given the parameters $\Pr_n$, $t_{n,l}$, and BLP_n given in Eqs. (12), (13), and (19), the denominator in (20) can be written as

$$\Pr\{s_n \text{ at time } t\} = \begin{cases} \sum_{n=1}^{N} \Pr_n \text{BLP}_n, & \text{if } s_n = \bar{\text{L}}, \\ \sum_{n=1}^{N} \Pr_n (1-\text{BLP}_n), & \text{if } s_n = \bar{\text{N}}_n. \end{cases} \quad (21)$$

In addition, the numerator in (20) can be expressed as

$$\Pr\{s_l \text{ at time } t+1, s_n \text{ at time } t\} = \begin{cases} \sum_{n=1}^{N}\sum_{l=1}^{N} \Pr_n \text{BLP}_n t_{n,l} \text{BLP}_l, & \text{if } \{s_n, s_l\} = \{\bar{\text{L}}, \bar{\text{L}}\}, \\ \sum_{n=1}^{N} \Pr_n \text{BLP}_n t_{n,l} (1-\text{BLP}_l), & \text{if } \{s_n, s_l\} = \{\bar{\text{L}}, \bar{\text{N}}_l\}, \\ \sum_{l=1}^{N} \Pr_n (1-\text{BLP}_n) t_{n,l} \text{BLP}_l, & \text{if } \{s_n, s_l\} = \{\bar{\text{N}}_n, \bar{\text{L}}\}, \\ \Pr_n (1-\text{BLP}_n) t_{n,l} (1-\text{BLP}_l), & \text{if } \{s_n, s_l\} = \{\bar{\text{N}}_n, \bar{\text{N}}_l\}, \end{cases} \quad (22)$$

where $n, l \in \{1, 2, \cdots, N\}$. We can also re-write state transition probabilities in a matrix form of $\mathbf{P}$ as follows

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{\bar{\text{L}}\bar{\text{L}}} & \mathbf{P}_{\bar{\text{L}}\bar{\text{N}}} \\ \mathbf{P}_{\bar{\text{N}}\bar{\text{L}}} & \mathbf{P}_{\bar{\text{N}}\bar{\text{N}}} \end{bmatrix} = \begin{bmatrix} p_{\bar{\text{L}}\bar{\text{L}}} & p_{\bar{\text{L}}\bar{\text{N}}_1} & \cdots & p_{\bar{\text{L}}\bar{\text{N}}_{N-1}} & p_{\bar{\text{L}}\bar{\text{N}}_N} \\ p_{\bar{\text{N}}_1\bar{\text{L}}} & p_{\bar{\text{N}}_1\bar{\text{N}}_1} & \cdots & 0 & 0 \\ p_{\bar{\text{N}}_2\bar{\text{L}}} & p_{\bar{\text{N}}_2\bar{\text{N}}_1} & \cdots & 0 & 0 \\ p_{\bar{\text{N}}_3\bar{\text{L}}} & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{\bar{\text{N}}_{N-1}\bar{\text{L}}} & 0 & \cdots & p_{\bar{\text{N}}_{N-1}\bar{\text{N}}_{N-1}} & p_{\bar{\text{N}}_{N-1}\bar{\text{N}}_N} \\ p_{\bar{\text{N}}_N\bar{\text{L}}} & 0 & \cdots & p_{\bar{\text{N}}_N\bar{\text{N}}_{N-1}} & p_{\bar{\text{N}}_N\bar{\text{N}}_N} \end{bmatrix}. \quad (23)$$

On the other hand, we denote $p_{\bar{\text{L}}}$ and $p_{\bar{\text{N}}_n}$ to be the stationary probabilities of state $\bar{\text{L}}$ and state $\bar{\text{N}}_n$, respectively, and $\mathbf{p} = \left[p_{\bar{\text{L}}}, p_{\bar{\text{N}}_1}, p_{\bar{\text{N}}_2}, \cdots, p_{\bar{\text{N}}_N}\right]$. Then, $\mathbf{p}$ is called the vector of steady-state probabilities, and following the Markov chain theory, we have

$$\begin{cases} \mathbf{p} = \mathbf{p}.\mathbf{P}, \\ p_{\bar{\text{L}}} + \sum_{n=1}^{N} p_{\bar{\text{N}}_n} = 1. \end{cases} \quad (24)$$

By solving (24), we can easily obtain $\mathbf{p}$.

B. LOSS-BASED TCP MODELING

The operation of TCP Cubic depends on the loss events, i.e., its window size is adjusted based on the elapsed time between two consecutive loss events. To analyze the TCP performance, it is necessary to evaluate the TCP window growth between two adjacent loss events and the TCP throughput can be obtained by averaging all over the time. To do so, the range of TCP window size (0, W] is first separated into M intervals, where W is the maximum TCP window size. We assume that the average round-trip-time (RTT) is a constant, which is a common assumption in loss-based TCP analytical modeling [45], the maximum congestion window size W can be determined by

$$W = \overline{R}_b \times \text{RTT}, \quad (25)$$

where $\overline{R}_b$ (bps) is the average transmission rate, which can be calculated by

$$\overline{R}_b = \sum_{n=1}^{N} p_{\bar{\text{N}}_n} R_{b_n}, \quad (26)$$

where $p_{\bar{\text{N}}_n}$ is given in (24), and R_{b_n} is the bit rate of MCS n-th. Here, the TCP congestion window is in the i-th interval if its value belongs to $((i-1)W/M,\ iW/M]$ and we regard the window size to be the midpoint of the interval denoted by x_i, i.e., $x_i = (i-0.5)W/M$. The transition between the i-th interval to the j-th interval occurs when the loss events are in these intervals illustrated in Fig. 6(a). The transition probability between the i-th interval and the j-th interval can be determined as

$$q_{i,j} = \Pr(T_{i,j}^{\min} < \tau_{loss} \le T_{i,j}^{\max}), \quad (27)$$

where $T_{i,j}^{\min}$ and $T_{i,j}^{\max}$ are the minimum time duration (from βx_i to $(j-1)W/M$) and maximum time duration (from βx_i to jW/M) from the i-th interval to the j-th interval, respectively.

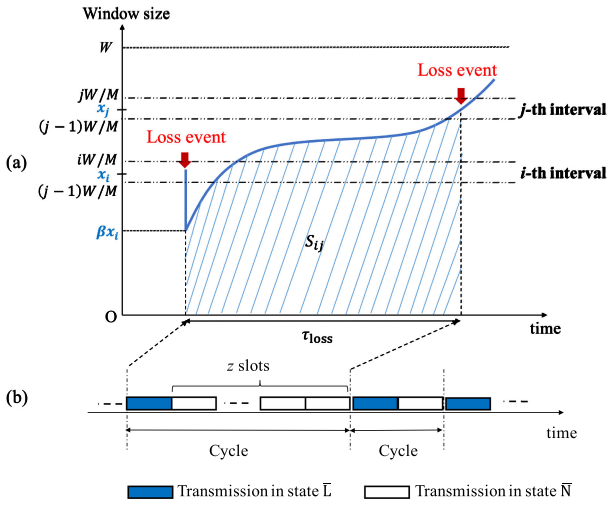

FIGURE 6. (a) TCP congestion window behavior, and (b) loss cycle of slot transmission.

Based on (3), their values can be written as

$$\begin{cases} T_{i,j}^{\min} = \sqrt[3]{\dfrac{(j-1)W/M - x_i}{\alpha}} + \sqrt[3]{\dfrac{(1-\beta)x_i}{\alpha}}, \\ T_{i,j}^{\max} = \sqrt[3]{\dfrac{jW/M - x_i}{\alpha}} + \sqrt[3]{\dfrac{(1-\beta)x_i}{\alpha}}, \end{cases} \tag{28}$$

where α and β are the Cubic factor and multiplicative decrease factor, respectively. In addition, τ_{loss} is the time duration between the i-th interval and the j-th interval (between two consecutive loss events) which is a random variable and can be assumed to follow an exponential distribution with the rate of λ as in [46]. The PDF of τ_{loss} can be written as

$$f(\tau_{loss}) = \lambda \exp(-\lambda \tau_{loss}). \ \ \tau_{loss} > 0 \tag{29}$$

Based on (29), the transition probability between loss events in (27) can be determined as

$$q_{i,j} = \exp(-\lambda T_{i,j}^{\max}) - \exp(-\lambda T_{i,j}^{\min}), \tag{30}$$

Here, $1/\lambda$ is the mean value of τ_{loss}. To compute (30), we need to determine the value of λ which is the inversion of the average time duration between two adjacent loss events (i.e., $\overline{\tau}_{loss}$). As we assume that the probability of frame loss and TCP segment loss are equal, the duration of two consecutive loss events of TCP is also cover a loss cycle of frame transmission which starts by a transmission in $\bar{\text{L}}$-state and ends by another $\bar{\text{L}}$-one as illustrated in Fig. 6(b). The value of λ can be calculated based o the average duration in a cycle, i.e.,

$$\lambda = \frac{1}{\overline{\tau}_{loss}} = \frac{1}{T_{\bar{\text{L}}} + T_{\bar{\text{N}}} \times E_{\bar{\text{N}}}}, \tag{31}$$

where $T_{\bar{\text{L}}} = (N_{\text{re}} + 1)T_b$ and $T_{\bar{\text{N}}} = \overline{K} T_b$ are the average duration of transmission in $\bar{\text{L}}$-state and $\bar{\text{N}}$-state, respectively, in which N_{re} is the maximum number of retransmission of a frame, T_b is the block duration, and $\overline{K}$ is given in (18). $E_{\bar{\text{N}}}$ is the average number of transmissions in $\bar{\text{N}}$-state in a cycle and can be calculated as

$$E_{\bar{\text{N}}} = \sum_{n=1}^{N} \overline{p}_{\bar{\text{L}},\bar{\text{N}}_n} \times \sum_{z=1}^{\infty} z \times \overline{p}_{\bar{\text{N}}_n,\bar{\text{L}}}(z) + 0 \times p_{\bar{\text{L}},\bar{\text{L}}}. \tag{32}$$

where $\overline{p}_{\bar{\text{N}}_n,\bar{\text{L}}}(z)$ is the $(n+1,1)$-th element of $\mathbf{P}(z) = \mathbf{P}^z$ given in (23).

C. TCP NORMALIZED THROUGHPUT

We now calculate the normalized TCP Cubic throughput which is defined as the ratio between the TCP throughput and the transmission rate. Given $r(t)$ to be the total amount of data transmitted during time t, the average normalized TCP throughput can be written as

$$\overline{\eta} = \frac{\lim\limits_{t\to\infty} \dfrac{r(t)}{t}}{\overline{R}_b}, \tag{33}$$

where $\overline{R}_b$ is the average transmission rate given in (26). Let $\text{cwnd}_{i,j}$ denote the total congestion window amount during the time duration $T_{i,j}$ which corresponds to the transition between the mid-point of i-th interval and the mid-point of j-th interval, the term $\lim\limits_{t\to\infty} r(t)/t$ can be determined as

$$\lim_{t\to\infty} \frac{r(t)}{t} = \frac{\sum\limits_{i=1}^{M}\sum\limits_{j=1}^{M} q_{i,j}\text{cwnd}_{i,j}}{\sum\limits_{i=1}^{M}\sum\limits_{j=1}^{M} q_{i,j}T_{i,j}}, \tag{34}$$

where M is the number of TCP window intervals, and $q_{i,j}$ is given in (30). The value of T_{ij} is the time duration between the mid-point of i-th and mid-point of j-th intervals which can be expressed by

$$T_{i,j} = \sqrt[3]{\frac{(j-0.5)W/M - x_i}{\alpha}} + \sqrt[3]{\frac{(1-\beta)x_i}{\alpha}}. \tag{35}$$

where x_i is the midpoint of i-th interval, i.e., $x_i = (i - 0.5)W/M$. In addition, the value of $\text{cwnd}_{i,j}$ can be computed as

$$\text{cwnd}_{i,j} = \int_0^{T_{i,j}} \frac{w(x_i,t)}{\text{RTT}} dt = \frac{S_{i,j}}{\text{RTT}}, \tag{36}$$

where RTT is the average round-trip time of TCP and $w(x_i, t)$ is the congestion window given in (1), while S_{ij} is the shaded area in Fig. 6(a) and can be calculated as

$$\begin{aligned} S_{i,j} &= \int_0^{T_{ij}} w(x_i,t)dt, \\ &= \int_0^{T_{ij}} \left(\alpha\left(\tau - \sqrt[3]{\frac{(1-\beta)x_i}{\alpha}}\right)^3 + x_i\right) dt, \\ &= x_i T_{i,j} + \frac{\alpha}{4}\left(\left(T_{i,j} - \sqrt[3]{\frac{(1-\beta)x_i}{\alpha}}\right)^4 - \left(\sqrt[3]{\frac{(1-\beta)x_i}{\alpha}}\right)^4\right). \end{aligned} \tag{37}$$

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present and discuss the TCP Cubic throughput performance analyzed in Section IV with different parameter settings of the 5G mmWave networks. The different puncturing patterns for each transmission of IR-HARQ and parameters related to MCS levels of AMC are shown in Tables 1 and 2, respectively, while the other system parameters are given in Table 3.

Monte Carlo simulations are also performed to validate the analytical results. Let φ_i denote the counter for the occurrence of the i-th intervals, where $i \in \{1, 2, \cdots, M\}$. The simulation at each time instant is performed as follows: at each time instant t (in i-th window interval), we first determine x_i according to the congestion window function, $w(x_i, \tau_i)$, given in (1), and update the counter (i.e., increase φ_i by 1). Then, the time duration from time instant t to the next loss event,

TABLE 3. System parameters.

Name	Symbol	Value
MmWave carrier frequency	f_c	28 GHz
UE speed	v	1.42 m/s
Frame size	N_f	1080 bits
Parameter-related block design	b	3
Persistent level of IR-HARQ	N_{re}	2
Symbol rate	R_s	50 Msps
Average RTT	RTT	10 ms
Cubic factor	α	1 Mb/s
TCP multiplicative decrease factor	β	0.7
Number of TCP window intervals	M	100
Average SNR	$\overline{\gamma}$	10 dB
Nakagami factor	m	1.5
Blockage parameter	β_{B}	0.0014
Distance from UE to BS	d_{UE}	30 m
Cell radius	r_c	200 m

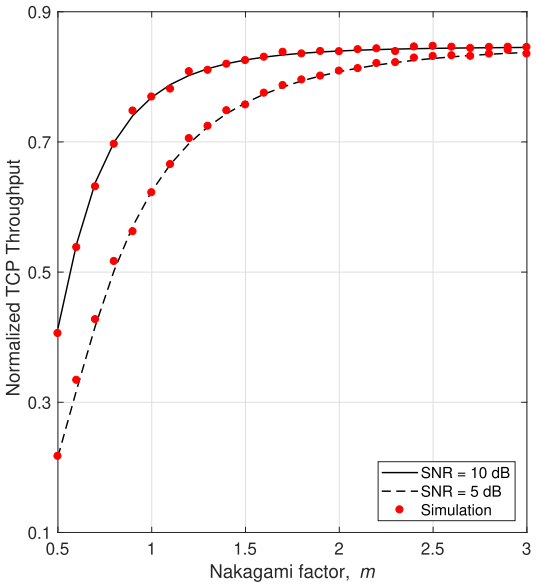

FIGURE 7. The normalized TCP throughput vs. Nakagami factor *m* for different values of received SNR.

τ_{loss} can be randomly generated according to its PDF in (29). Based on x_i and τ_{loss}, the window size at the next window reduction (in j-th window interval), i.e., $w(x_j, \tau_{\text{loss}})$ can be computed. Using the number of simulation runs of 1000, the average TCP Cubic throughput can then be determined.

First, we investigate in Fig. 7 the impact of mmWave fading channel on the TCP normalized throughput by varying the Nakagami factor m for different received SNR, when the target frame loss rate, P_{loss}, is set to 10^{-2}. From this figure, we can verify a good agreement between the analytical and simulation results, which confirm the correctness of the model and analysis. Here, it is important to note that the Nakagami shape factor m determines the severity of fading, i.e., smaller values of m specify more severity fading phenomena. Besides, LOS and NLOS links can be modeled when $m > 1$ and $0.5 \leq m \leq 1$, respectively. Therefore, a decreasing of m, which implies an increase of loss rate, results in a degradation of TCP throughput.

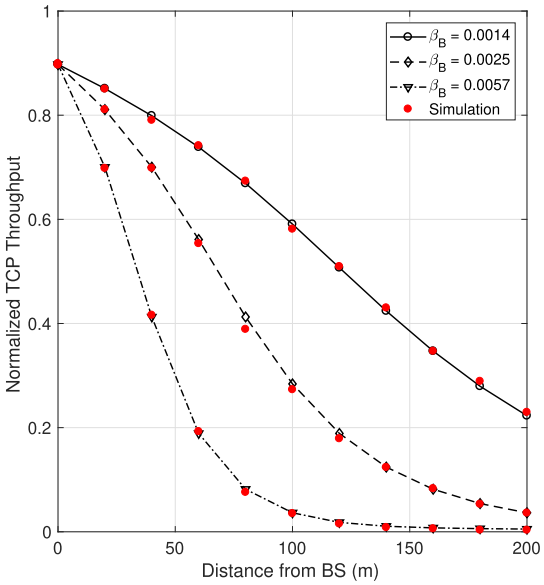

FIGURE 8. The normalized TCP throughput vs. distance from the BS for different blockage parameters of β_{B}.

In addition to the fading phenomenon, blockages caused by buildings is also one of the main challenging issues in the design of 5G mmWave networks. Figure 8 shows the impact of blockages on the TCP throughput with different transmission distances from BS, given $P_{\text{loss}} = 10^{-2}$. Also, different blockage parameters, i.e., $\beta_{\text{B}} = 0.0014$, $\beta_{\text{B}} = 0.0025$, and $\beta_{\text{B}} = 0.0057$ for low, moderate, and high density of building blockages, are considered. We can observe that TCP throughputs are degraded significantly at high values of β_{B} indicating dense blockages, especially at the edge of the cell. For example, when the UE is moving at the edge of the cell (i.e., $d_{\text{UE}} = 200\ m$), TCP Cubic can still maintain a normalized throughput of 0.2 at the low density of building blockages (i.e., $\beta_{\text{B}} = 0.0014$) while the TCP connection is basically not available at the high density of blockage (i.e., $\beta_{\text{B}} = 0.0057$).

One of the important issues in the design of the system using AMC is the optimization of MCS operation (i.e. the selection of MCS levels for channel states) so that the highest system throughput can be achieved. The operation of MCS depends on the selection of the target frame loss rate P_{loss}. This phenomenon is investigated in Fig. 9, which shows the TCP normalized throughput with different shape factor, m, versus P_{loss}. It is seen that, at the $P_{\text{loss}} = 10^{-2}$ for all Nakagami shape factor m, the TCP throughput is maximized. This result reveals the joint design problem of AMC at the PHY layer and IR-HARQ at the link layer. In particular, when the higher value of target P_{loss} is set for the AMC, the system can have more chances to select the higher MCS levels for transmission, resulting in the increase of the TCP throughput. Nevertheless, high value of target P_{loss} leads to high frame loss rate (FLR) in the system due to the fact that FLR of higher MCS level is larger than that of lower MCS level for a given value of SNR. There is no guarantee that the transmitted

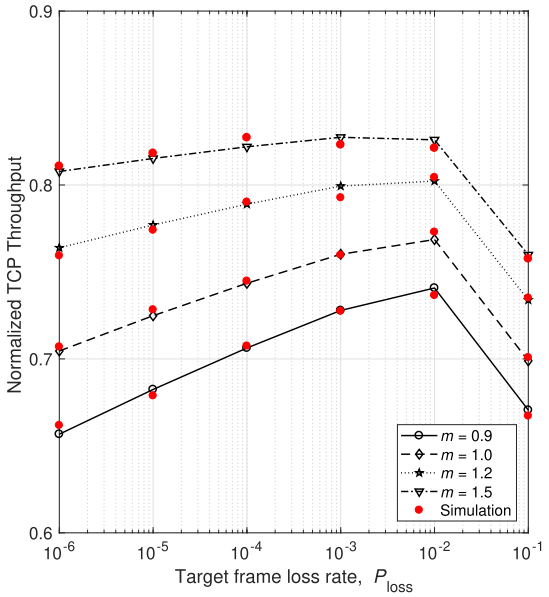

FIGURE 9. The normalized TCP throughput vs. target frame loss rate for different values of *m*.

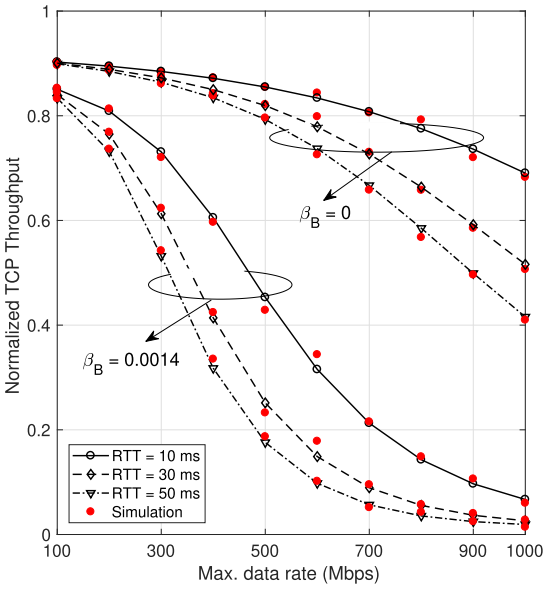

FIGURE 11. The normalized TCP throughput vs. maximum transmission rate for different average RTT in case of blockage and non-blockage.

data will be correctly received at the receiver side and many retransmissions performed by IR-HARQ may be required to cover the transmission errors, which may degrade the TCP throughput. Therefore, there exists an optimal value of target P_{loss}.

Now, using the optimal $P_{\text{loss}} = 10^{-2}$, in Fig. 10, we quantitatively highlight the effectiveness of using IR-HARQ to improve the TCP throughput performance over mmWave fading channels. Evidently, the TCP throughput increases significantly when IR-HARQ is employed. For instance, TCP connection is basically not available (the normalized TCP throughput $\overline{\eta} \simeq 0$) in the low regime of SNR (SNR $\leq$ 2 dB) when no IR-HARQ is employed. The normalized TCP

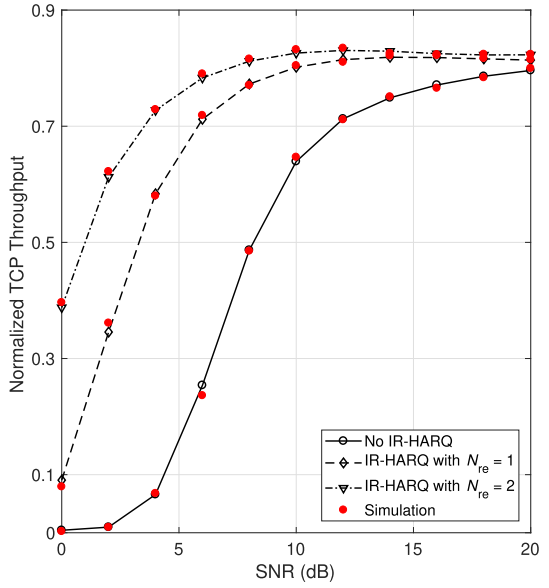

FIGURE 10. The normalized TCP throughput vs. average received SNR for different persistent levels of IR-HARQ.

throughput is considerably improved, it is in fact higher than 0.5 when IR-HARQ is employed at SNR = 2 dB. In addition, while a higher value of IR-HARQ persistent level can mitigate the effects of frame loss and increase the TCP throughput, it also increases the overall end-to-end latency. The design goals of 5G networks, nonetheless, demand both high throughput and low latency. We therefore use the persistent level of $N_{\text{re}} = 2$ as recommended [18]. With this persistent level, the normalized TCP throughput can be 0.6 when $SNR = 2$ dB.

Another essential issue of TCP in 5G networks is the performance evaluation of TCP over large bandwidth-delay product (BDP) or long fat networks (LNF). Figure 11 investigates the impact of possible maximum transmission rate on the normalized TCP throughput for different average RTT, given $P_{\text{loss}} = 10^{-2}$. We can see that the performance of TCP is severely degraded by BDP networks, especially in presence of blockages. The reason is that the performance of TCP Cubic depends on the time duration between consecutive loss events and these events occur frequently in the high error rate environment and high density of blockages which makes a large amount of data of BDP networks is lost in a given time duration. For example, as $\beta_{\text{B}} = 0.0014$, the normalized TCP throughput is nearly zero in large BDP networks, i.e., when RTT = 50 ms and maximum transmission rate of 1 Gbps (BDP = 6,250 kB).

Finally, Fig. 12 illustrates the relation between TCP Cubic parameters, i.e., multiplicative decrease factor β [Figs. 12(a) and (b)], Cubic factor α [Figs. 12(c) and (d)], and TCP throughput performance under impact of different fading channel conditions and distance from BS. Also $P_{\text{loss}} = 10^{-2}$ is used. Specifically, using Figs. 12(a) and (b) with $\alpha = 1$ Mbps, we can find the minimum value of β required to reach a maximum normalized throughput level for

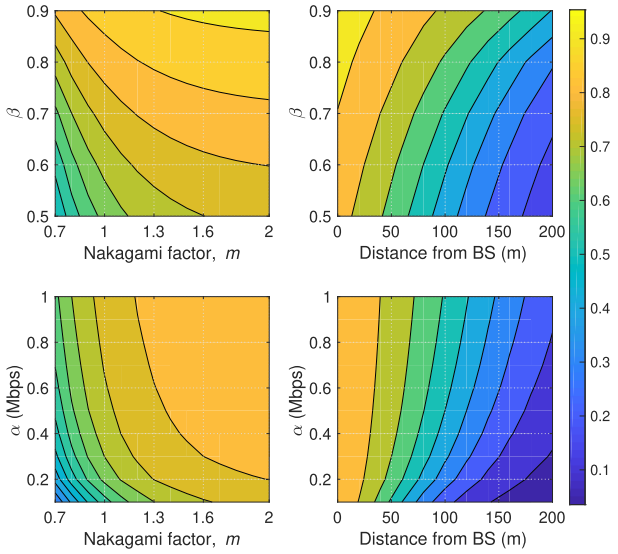

FIGURE 12. The normalized TCP Throughput versus β, given $\alpha = 1$ Mbps for different values of (a) m with $d_{\text{UE}} = 30$ m and (b) d_{UE} with $m = 1.5$, or versus α, given $\beta = 0.7$ with respect to (c) m with $d_{\text{UE}} = 30$ m and (d) d_{UE} with $m = 1.5$.

different values of m and d_{UE}. For example, when $m = 1.5$ and $d_{\text{UE}} = 100$ m, in order to reach the maximum normalized throughput, the β required is 0.82. Similarly, using $\beta = 0.7$ in Figs. 12(c) and (d), we can find the minimum value of α to achieve a maximum throughput at a particular values of m and d_{UE}. For instance, α should be used as 0.6 Mbps to guarantee the maximum normalized throughput level at $m = 1.3$ and $d_{\text{UE}} = 30$ m.

VI. CONCLUSION

In this paper, we presented an analytical framework for TCP throughput evaluation and optimization over mmWave fading channels in 5G cellular networks. The framework was constructed based on the cross-layer design approach in which AMC and truncated IR-HARQ were employed at the physical layer and link layer, respectively. Using the proposed framework, we conducted an investigation on the TCP throughput performance under effect of different parameters/settings of AMC and IR-HARQ and performed the Monte Carlo simulations to validate the analytical model. Numerical results illustrated the impact of mmWave fading channels and building blockages on the TCP throughput and supported the selection of optimal parameters.

REFERENCES

[1] M. Ree, G. Mantas, A. Radwan, S. Mumtaz, J. Rodriguez, and I. Otung, "Key management for beyond 5G mobile small cells: A survey," *IEEE Access*, vol. 7, pp. 59200–59236, 2019.

[2] L. Chiaraviglio, A. S. Cacciapuoti, G. D. Martino, M. Fiore, M. Montesano, D. Trucchi, and N. B. Melazzi, "Planning 5G networks under EMF constraints: State of the art and vision," *IEEE Access*, vol. 6, pp. 51021–51037, 2018.

[3] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-end simulation of 5G mmWave networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2237–2263, 3rd Quart., 2018.

[4] I. K. Jain, R. Kumar, and S. S. Panwar, "The impact of mobile blockers on millimeter wave cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 854–868, Apr. 2019.

[5] 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.

[6] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 6th ed. London, U.K.: Pearson, 2012.

[7] C. T. Nguyen, H. D. Le, and V. V. Mai, "A cross-layer analysis of tcp/link adaptation technologies over free-space optical links with Markov error model," *Photon. Netw. Commun.*, vol. 36, pp. 279–288, Dec. 2018.

[8] J. Gao and N. S. V. Rao, "TCP AIMD dynamics over Internet connections," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 4–6, Jan. 2005.

[9] C. T. Nguyen, V. V. Mai, and A. T. Pham, "TCP over free-space optical links with ARQ and AMC: A cross-layer performance analysis," in *Proc. Int. Conf. Adv. Technol. Commun. (ATC)*, Oct. 2016, pp. 80–84.

[10] S. Ha, I. Rhee, and L. Xu, "CUBIC: A new TCP-friendly high-speed TCP variant," *ACM SIGOPS Operat. Syst. Rev.*, vol. 42, no. 5, pp. 64–74, 2008.

[11] D. Kliazovich, F. Granelli, and D. Miorandi, "Logarithmic window increase for tcp westwood+ for improvement in high speed, long distance networks," *Comput. Netw.*, vol. 52, pp. 2395–2410, Aug. 2008.

[12] A. Baiocchi, A. P. Castellani, and F. Vacirca, "YeAH-TCP: Yet another highspeed TCP," in *Proc. 5th Int. Workshop Protocols Fast Long-Distance Netw. (PFLDnet)*, 2007, pp. 37–42.

[13] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh, and V. Jacobson, "BBR: Congestion-based congestion control," *Queue*, vol. 14, no. 5, pp. 50-20–50-53, Oct. 2016.

[14] S. Jain and G. Raina, "An experimental evaluation of cubic TCP in a small buffer regime," in *Proc. Nat. Conf. Commun. (NCC)*, Jan. 2011, pp. 1–5.

[15] 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 *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2016, pp. 730–735.

[16] M. Kim, S.-W. Ko, H. Kim, S. Kim, and S.-L. Kim, "Exploiting caching for millimeter-wave TCP networks: Gain analysis and practical design," *IEEE Access*, vol. 6, pp. 69769–69781, 2018.

[17] M. Polese and M. Zorzi, "Impact of channel models on the end-to-end performance of mmwave cellular networks," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[18] M. Polese, R. Jana, and M. Zorzi, "TCP and MP-TCP in 5G mmWave networks," *IEEE Internet Comput.*, vol. 21, no. 5, pp. 12–19, Sep. 2017.

[19] C. Lee, S. Song, H. Cho, G. Lim, and J.-M. Chung, "Optimal multipath TCP offloading over 5G NR and LTE networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 293–296, Feb. 2019.

[20] P. J. Mateo, C. Fiandrino, and J. Widmer, "Analysis of TCP performance in 5G mm-wave mobile networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–7.

[21] M. Zhang, M. Polese, M. Mezzavilla, J. Zhu, S. Rangan, S. Panwar, and M. Zorzi, "Will TCP Work in mmWave 5G cellular networks?" *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 65–71, Jan. 2019.

[22] M. Panda, H. L. Vu, M. Mandjes, and S. R. Pokhrel, "Performance analysis of TCP NewReno over a cellular last-mile: Buffer and channel losses," *IEEE Trans. Mobile Comput.*, vol. 14, no. 8, pp. 1629–1643, Aug. 2015.

[23] S. R. Pokhrel, M. Panda, and H. L. Vu, "Analytical modeling of multipath TCP over last-mile wireless," *IEEE/ACM Trans. Netw.*, vol. 25, no. 3, pp. 1876–1891, Jun. 2017.

[24] S. R. Pokhrel and M. Mandjes, "Improving multipath TCP performance over WiFi and cellular networks: An analytical approach," *IEEE Trans. Mobile Comput.*, to be published.

[25] C. Sahin, L. Liu, E. Perrins, and L. Ma, "Delay-sensitive communications over IR-HARQ: Modulation, coding latency, and reliability," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 749–764, Apr. 2019.

[26] 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 *Proc. 16th Annu. Medit. Ad Hoc Netw. Workshop (Med-Hoc-Net)*, Jun. 2017, pp. 1–6.

[27] S. H. Myint, K. Yu, and T. Sato, "Modeling and analysis of error process in 5G wireless communication using two-state Markov chain," *IEEE Access*, vol. 7, pp. 26391–26401, 2019.

[28] G. Ku and J. M. Walsh, "Resource allocation and link adaptation in LTE and LTE advanced: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1605–1633, 3rd Quart., 2015.

[29] M. A. Adedoyin and O. E. Falowo, "Qos-based radio resource management for 5G ultra-dense heterogeneous networks," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2017, pp. 1–6.
[30] V. Kumar and N. B. Mehta, "Modeling and analysis of differential CQI feedback in 4G/5G OFDM cellular systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2361–2373, Apr. 2019.
[31] M. Ahmad, S. Jabbar, A. Ahmad, F. Piccialli, and G. Jeon, "A sustainable solution to support data security in high bandwidth health care remote locations by using tcp cubic mechanism," *IEEE Trans. Sustain. Comput.*, to be published.
[32] T. Bai, R. Vaze, and R. W. Heath, Jr., "Analysis of blockage effects on urban cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, Sep. 2014.
[33] M. S. Omar, S. A. Hassan, H. Pervaiz, Q. Ni, L. Musavian, S. Mumtaz, and O. A. Dobre, "Multiobjective optimization in 5G hybrid networks," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1588–1597, Jun. 2018.
[34] M. N. Kulkarni, S. Singh, and J. G. Andrews, "Coverage and rate trends in dense urban mmWave cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 3809–3814.
[35] G. Yang, M. Xiao, and Z. Pang, "Delay analysis of traffic dispersion with Nakagami-m fading in millimeter-wave bands," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.
[36] X. Liu, J. Xu, Y. Pei, and Y.-C. Liang, "Gaussian mixture model for millimeter-wave cellular communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3174–3188, Apr. 2019.
[37] N. Bouhlel and A. Dziri, "Maximum Likelihood Parameter Estimation of Nakagami-Gamma Shadowed Fading Channels," *IEEE Commun. Lett.*, vol. 19, no. 4, pp. 685–688, Apr. 2015.
[38] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Amsterdam, The Netherlands: Elsevier, 2007.
[39] M. D. Yacoub, J. E. V. Bautista, and L. G. D. R. Guedes, "On higher order statistics of the Nakagami-m distribution," *IEEE Trans. Veh. Technol.*, vol. 48, no. 3, pp. 790–794, May 1999.
[40] P. Zhang, Y. Miao, and Y. Zhao, "Cross-layer design of AMC and truncated HARQ using dynamic switching thresholds," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 906–911.
[41] A. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Trans. Commun. Technol.*, vol. 19, no. 5, pp. 751–772, Oct. 1971.
[42] J. Ramis and G. Femenias, "Cross-layer design of adaptive multirate wireless networks using truncated HARQ," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 944–954, Mar. 2011.
[43] H. D. Le, C. T. Nguyen, V. V. Mai, N. T. Dang, and A. T. Pham, "On the performance of tcp cubic over fading channels with amc schemes," in *Proc. Int. Conf. Adv. Technol. Commun. (ATC)*, Oct. 2017, pp. 1–5.
[44] V. Raghavan, L. Akhoondzadeh-Asl, V. Podshivalov, J. Hulten, M. A. Tassoudji, O. H. Koymen, A. Sampath, and J. Li, "Statistical blockage modeling and robustness of beamforming in millimeter-wave systems," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3010–3024, Jul. 2019.
[45] J. Padhye, V. Firoiu, D. F. Towsley, and J. F. Kurose, "Modeling TCP reno performance: A simple model and its empirical validation," *IEEE/ACM Trans. Netw.*, vol. 8, no. 2, pp. 133–145, Apr. 2000.
[46] S. Hassayoun, P. Maille, and D. Ros, "On the impact of random losses on TCP performance in coded wireless mesh networks," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.

HOANG D. LE received the B.E. degree in electronics and communications from the Hanoi University of Science and Technology (HUST), Vietnam, in 2016, and the M.E. degree in computer science and engineering from The University of Aizu (UoA), Japan, in 2018, where he is currently pursuing the Ph.D. degree with the Computer Communications Laboratory. His research interest is in area of optical wireless communications for 5G-and-beyond networks, with emphasis on cross-layer protocol designs. He received the Best Paper Award from KICS/IEEE ICTC, South Korea, in 2018.

CHUYEN T. NGUYEN received the B.E. degree in electronics and telecommunications from the Hanoi University of Science and Technology (HUST), Vietnam, in 2006, the M.S. degree in communications engineering from National Tsing Hua University, Taiwan, in 2008, and the Ph.D. degree in informatics from Kyoto University, Japan, in 2013. From September to November 2014, he was a Visiting Researcher with The University of Aizu, Japan. He is currently an Assistant Professor with the School of Electronics and Telecommunications, HUST, Vietnam. His current research interests are in areas of communications theory and applications with a particular emphasis on protocol design for industrial Internet of Things applications. He received the Fellow Award from the Hitachi Global Foundation, in August 2016, First Best Paper Award, in 2019 IEEE ICT, the Best Paper Awards in 2018 KICS/IEEE ICTC and the 2019 IEEE ICC

VUONG V. MAI received the B.E. degree (Hons.) in electronic telecommunication engineering from the Posts and Telecommunications Institute of Technology (PTIT), Vietnam, in 2012, and the M.S. and Ph.D. degrees in computer science and engineering from The University of Aizu (UoA), Japan, in 2014 and 2017, respectively. After his graduation, he awarded the prestigious Japanese Government (Monbukagakusho) Scholarship for studying, in Japan. In April 2017, he joined the Korea Advanced Institute of Science and Technology (KAIST), South Korea, where he is currently a Postdoctoral Fellow with the Photonics Systems Research Laboratory, School of Electrical Engineering. At KAIST, he is involved in a national research project funded by the Agency for Defense Development (ADD). He also works in collaboration with industry partners, such as HFR, Inc., (Telecom Company) through several collaborative Research and Development projects. His general research areas include telecommunications and photonics with emphasis on wireless applications. Recent research activities include optical wireless technologies for the over 5G/aerospace/IoT systems/networks, and military applications. He has authored/coauthored about over 40 technical articles in peer-reviewed journals and international conferences. He has also served as a TPC member of various international conferences and a member of OSA and IEICE. He was a recipient/co-recipient of a number of awards, including the First Best Paper Award, in 2019 IEEE ICT, the Best Paper Award, in 2018 KICS/IEEE ICTC, the 2016 IEEE VTS Tokyo Chapter Young Researcher's Encouragement Award, the 2015 IEEE ComSoc Sendai Chapter Student Excellent Research Award, the 2015 IEEE GLOBECOM Student Travel Grant Award, and the 2013 IEEE Sendai Section Student Award. He has frequently served as a Reviewer for OSA and the IEEE JOURNALS. He received the UoA President's Award of Excellence, for his M.S. degree.

ANH T. PHAM (M'00–SM'11) received the B.E. and M.E. degrees in electronics engineering from the Hanoi University of Technology, Vietnam, in 1997 and 2000, respectively, and the Ph.D. degree in information and mathematical sciences from Saitama University, Japan, in 2005. From 1998 to 2002, he was with NTT Corporation, Vietnam. Since 2005, he has been a Faculty Member with The University of Aizu, where he is currently a Professor and the Head of the Computer Communications Laboratory, Division of Computer Engineering. His research interests are in the broad areas of communication theory and networking with a particular emphasis on modeling, design, and performance evaluation of wired/wireless communication systems and networks. He has authored/coauthored over 160 peer-reviewed articles on these topics. He is also a member of IEICE and OSA.