Improving Session Continuity with Bandwidth Reservation in mmWave Communications

Dmitri Moltchanov, Andrey Samuylov, Vitaly Petrov, Margarita Gapeyenko,
Nageen Himayat, Sergey Andreev, and Yevgeni Koucheryavy

***Abstract*—Unexpected fluctuations in radio resource demands of active sessions in millimeter-wave (mmWave) systems caused by dynamic blockage of the line-of-sight path may lead to accidental session drops. Existing solutions addressing this problem incorporate multi-connectivity operation that reroutes an ongoing session to another mmWave access point (AP) nearby. We introduce and analyze an alternative approach for improving session continuity based on the concept of bandwidth reservation. We develop an analytical model for the mmWave system that employs the proposed technique and quantify its benefits and drawbacks. We show that even a small fraction of reserved bandwidth may significantly improve session continuity.**

I. Introduction

Millimeter-wave (mmWave) radio operating in the extremely high frequency (EHF) band is a key component of the emerging 5G mobile technology. While its standardization process is almost complete and vendors are testing their early implementations, the research focus is shifting towards optimization of the system-level performance [1].

One of the effects that impacts the performance of outdoor mmWave deployments is *dynamic blockage* of the line-of-sight (LoS) path between the user equipment (UE) and the mmWave access point (AP) by human bodies of, e.g., pedestrians moving around the UE [2]. This leads to complex dynamics of the UE signal-to-noise ratio (SNR) at sub-second timescales [3], thus resulting in frequent changes of the modulation and coding scheme (MCS) and causing fluctuations in the bandwidth requirements of the ongoing sessions [4].

The described effect is detrimental to delay-sensitive non-elastic types of traffic, typical for augmented/virtual reality applications and high-quality video calls envisioned to be supported by (beyond-)5G mmWave systems. For a mmWave network handling such sessions, dynamic blockage may bring occasional drops of the ongoing sessions not only caused by the SNR outage [5], but also due to drastic spikes in the bandwidth requirements of a session leading to short-term depletion of the available radio resources at the serving AP [6]. The multipath components are much weaker than the LoS path, which requires extra resources from the AP to support a certain level of quality for the UEs [4].

Recall that from the quality of service perspective, it is often preferable to reject a session at the moment of its arrival rather than drop it during service [7]. Therefore, we formulate and study an approach to follow this thinking, which is based on the concept of *bandwidth reservation*. It suggests keeping a certain fraction of mmWave AP's radio resources unavailable for the newly arriving sessions. The AP may thus refrain from accepting new sessions and concentrate exclusively on serving the ongoing ones if its current load exceeds a certain threshold.

A similar concept named guard bandwidth was initially proposed for 2G networks to prioritize the handover procedures. Particularly, two schemes were considered: (i) static channel reservation where a certain number of channels is reserved permanently and (ii) dynamic channel reservation where the channels are only reserved when a user approaches the overlap of cells [8], [9]. In LTE systems, the latter approach has been applied for handover sessions. For mmWave technology, extra resources might be required even when the UE is stationary but the environment moves around it.

Hence, the use of bandwidth reservation at mmWave APs may improve session continuity. The natural questions are (i) how much resources should be reserved by a mmWave AP in different conditions and (ii) what are the trade-offs between the key performance indicators? To address these, we develop an analytical model for the mmWave AP service process with dynamic blockage and bandwidth reservation, which combines the methods of renewal theory, stochastic geometry, and queuing theory. We use this model to estimate the fraction of radio resources that needs to be reserved and show that even a small share of reserved bandwidth may greatly improve session continuity in crowded mmWave deployments.

II. System Model

Consider a single mmWave AP at height h_A and a number of pedestrians around it, Fig. 1. The AP has a circular coverage range of radius r. We assume that r is such that no UEs inside experience outage conditions when their LoS link is blocked.

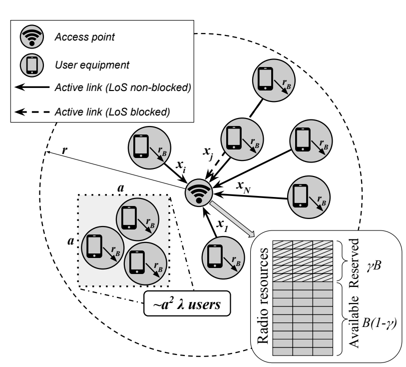


Fig. 1. An illustration of session-level service by mmWave AP.

D. Moltchanov, A. Samuylov, V. Petrov, M. Gapeyenko, S. Andreev and Y. Koucheryavy are with Tampere University of Technology, Finland (e-mail: name.lastname@tut.fi).

N. Himayat is with Intel Corporation, Santa Clara, CA, USA (e-mail: nageen.himayat@intel.com).

This work was supported by Intel Corporation and by the Academy of Finland (projects WiFiUS and PRISMA)

We use the 3GPP UMi propagation model [2] with blockage, which delivers the SNR values for a certain separation distance in LoS blocked and LoS non-blocked states.

The number of pedestrians follows a spatial Poisson process with the density of λ_B. They move according to a random direction mobility (RDM) model [10] with the speed of v m/s and an exponentially distributed run length with the mean of τ meters. Pedestrians are modeled as cylinders with height h_B and radius r_B. The height of UE carried by a pedestrian is h_U.

At an arbitrary instant of time, each of the pedestrians may initiate a session. The choice of the pedestrian who starts a new session is random, so the location of the user associated with a session is distributed uniformly within the AP's coverage area [10]. The process of session arrivals is thus Poisson with the intensity of λ sessions per second. Users remain static for the entire session duration. The session duration is distributed exponentially with the parameter μ, while the traffic is non-elastic with the constant bitrate of R Mbit/s.

The LoS path between the UE and the mmWave AP might be temporarily occluded by moving pedestrians. In this case, the UE communicates over the multipath components. Depending on the current link state (LoS blocked or LoS non-blocked) and the distance between the mmWave AP and the UE, the session employs an appropriate MCS [11]. For each link state and the set of MCSs, we divide the overall AP coverage area by concentric circles into N zones. Each of the zones is thus characterized by its specific radio resource requirements, $d_{0,i}$ and $d_{1,i}$, $i = 1,2,\dots,N$, for LoS non-blocked and LoS blocked states, respectively. These coefficients define the translation from the session bitrate in bit/s into the amount of radio resources occupied by this session at the AP in Hz/s.

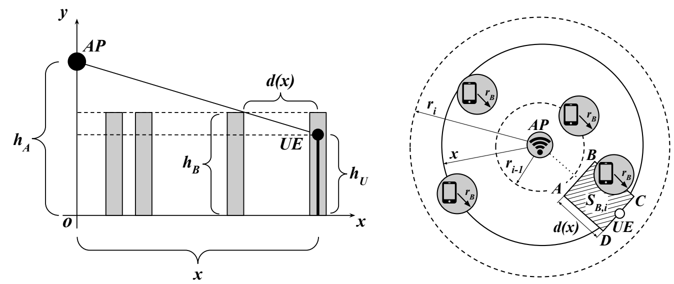
The mmWave AP is assumed to operate over the bandwidth of B Hz. Only a fraction $(1-\gamma)B$, $\gamma \in (0,1)$, of radio resources is available for new sessions. The entire pool of resources, B, is available for the ongoing sessions. Let $V(t)$ be the amount of resources occupied at time t. Upon arrival of a new session, the mmWave AP checks whether the remaining resources, $(1-\gamma)B - V(t)$, are sufficient to serve it. In case of a positive outcome, this new session is accepted. For a session changing its state from non-blocked to blocked, the decision is made based on $B - V(t)$. For a session changing its state from blocked to non-blocked the resources are always available.

We proceed by characterizing the trade-off between the drop probabilities for new and ongoing sessions, p_I and p_O, respectively. We also study the effects of γ on the resource utilization coefficient, $U = \lim_{t\to\infty} V(t)/B$.

III. Performance Modeling

A. Dynamic Blockage Model

Consider zone i bounded by the radii r_i and r_{i-1}. Session arrivals inside this zone are characterized by the same resource requirements and experience approximately the same blockage effects [6]. Consider a randomly chosen pedestrian initiating a session in zone i. Following the properties of the RDM, the position of this pedestrian's UE is distributed uniformly within the service zone [10], thus implying that the mean distance between the active user in zone i and the mmWave AP is $E[X_i] = 2(r_{i-1}^2/(r_{i-1}+r_i)+r_i)/3$ [12].



(a) Side view (b) Top view

Fig. 2. Configuration of the LoS blockage zone.

Consider the geometrical scenario in Fig. 2. There is an area between the mmWave AP and the UE, named the *LoS blockage zone*, where other pedestrians act as blockers occluding the mmWave link. This area can be approximated by a rectangle, *ABCD*, as shown in Fig. 2(b). The area of this zone is

$$S_B(x) = 2r_B d(x) = 2r_B\left(x\frac{h_B-h_U}{h_A-h_U}+r_B\right). \quad (1)$$

The average area of the LoS blockage zone i is then

$$E[S_{B,i}] = \int_{r_{i-1}}^{r_i}\frac{2x}{r_i^2-r_{i-1}^2}2r_B\left(x\frac{(h_B-h_U)}{h_A-h_U}+r_B\right)dx = \\ = \frac{4r_B(h_B-h_U)\left(r_{i-1}^2+r_{i-1}r_i+r_i^2\right)}{3(h_A-h_U)(r_{i-1}+r_i)}+2r_B^2. \quad (2)$$

To capture the dynamics of the blockage process in zone i, the intensity of pedestrian arrivals into the LoS blockage zone is required, ζ_i. Extending the result from [13], the inter-meeting time of a single pedestrian within the LoS blockage zone is approximately exponential with the parameter

$$\zeta_{i,1} = vE[S_{B,i}]\iint_{S_{U_i}} f^2(s)ds = \\ = \frac{2v(r_i-r_{i-1})\left(\frac{4r_B(h_B-h_U)\left(r_{i-1}^2+r_{i-1}r_i+r_i^2\right)}{3(h_A-h_U)(r_{i-1}+r_i)}+2r_B^2\right)}{r_i^2-r_{i-1}^2}, \quad (3)$$

where v is the speed of a moving pedestrian and $f(s) = 1/\pi(r_i^2-r_{i-1}^2)$ is the stationary distribution of the RDM [10]. Since the number of pedestrians in the mmWave AP's coverage area is Poisson with the intensity of $\pi\lambda_B(r_i^2-r_{i-1}^2)$, the process of the LoS path blockage is also Poisson with the intensity of

$$\zeta_i = \frac{2v\pi\lambda_B}{(r_i-r_{i-1})^{-1}}\left[\frac{4r_B(h_B-h_U)(r_{i-1}^2+r_{i-1}r_i+r_i^2)}{3(h_A-h_U)(r_{i-1}+r_i)}+2r_B^2\right]. \quad (4)$$

Let $\Theta_{0,i}$ and $\Theta_{1,i}$ be the times spent by the UE in LoS non-blocked and LoS blocked states inside zone i, respectively. The LoS blocked and non-blocked intervals interchange by forming an alternating renewal process, where each LoS blocked interval may comprise a number of overlapping intervals due to multiple LoS path occlusions by individual pedestrians.

The time spent by the UE in the LoS non-blocked state is distributed exponentially with the parameter ζ_i [3]. The mean LoS blocked time can be established via the fraction of time when the UE located in zone i is not blocked, $p_{L,i}$, which is related to the mean blockage time as

$$p_{L,i} = E[\Theta_{0,i}]/(E[\Theta_{0,i}]+E[\Theta_{1,i}]),\ i = 1,2,\dots,N, \quad (5)$$

thus leading to $E[\Theta_{1,i}] = E[\Theta_{0,i}]/p_{L,i} + E[\Theta_{0,i}]$.

In the stationary conditions the fraction of time that the UE is blocked coincides with the probability of non-blockage. The latter is established by using the area of the LoS blockage zone in (1) and the void probability of the Poisson process as

$$p_{L,i} = e^{-\lambda_B \left(\frac{4r_B(h_B-h_U)\left(r_{i-1}^2+r_{i-1}r_i+r_i^2\right)}{3(h_A-h_U)(r_{i-1}+r_i)} + 2r_B^2 \right)}. \quad (6)$$

B. AP Session Service Model

The mmWave AP service process can be modeled by using a multi-service queuing system with multiple arrival flows, where each flow has the intensity of $\lambda_i = \lambda\pi(r_i^2 - r_{i-1}^2)$, $i = 1,2,\dots,N$. The system state $\vec{\chi}$ is a vector having the number of users in LoS blocked and non-blocked states as its elements,

$$\vec{\chi} = (n_{0,0},\dots,n_{0,N},n_{1,0},\dots,n_{1,N}), \quad (7)$$

where $n_{0,i}$ and $n_{1,i}$ are the numbers of UEs in zone i in LoS non-blocked and blocked states. The state evolution is a multi-dimensional stochastic process, $\{\vec{\chi}(t), t > 0\}$, over

$$X = \left\{ (n_{00},\dots,n_{1N}) : \sum_{x=0}^{N} n_{0,x}d_{0,x} + \sum_{x=0}^{N} n_{1,x}d_{1,x} \le B \right\}, \quad (8)$$

where $d_{0,i}$ and $d_{1,i}$ are the radio resource requirements for each MCS in LoS non-blocked and LoS blocked states, respectively.

It can be shown that $\{\vec{\chi}(t), t > 0\}$ is a Markov process. However, the computational complexity of this model is high. Hence, below we propose a simpler approach that captures the essence of the bandwidth reservation.

The session resource requirements do not vary much in the LoS non-blocked state as compared to the difference between the LoS blocked and the LoS non-blocked states [11]. Therefore, we combine the LoS non-blocked states by weighting the resource requirements of each one of them with the corresponding zone areas. The same is done for the LoS blocked states. The evolution of a new system state, $\vec{\chi}^{\star}$, forms a two-dimensional ergodic Markov process, $\{n_0(t), n_1(t), t > 0\}$, over

$$X^{\star} = \{n_0 > 0, n_1 > 0 : n_0d_0 + n_1d_1 \le B\}, \quad (9)$$

where n_0 and n_1 are the numbers of UEs in LoS non-blocked and LoS blocked states, respectively, d_0 and d_1 are the corresponding radio resource requirements. Let also $p(n_0,n_1)$, $\{n_0,n_1\} \in X^{\star}$ be the stationary state distribution of the model, which is the solution of $\vec{p}\Lambda^{\star} = 0$, $\vec{p}\vec{1} = 1$, where $\Lambda^{\star}$ is the infinitesimal generator and $\vec{e}^T$ is the vector of ones.

Denote by $\tilde{n} = n_0d_0 + n_1d_1$ the amount of resources occupied by LoS non-blocked and LoS blocked sessions. Since there are two types of sessions in the system and sessions may change their state while in service, there are six possible transitions out of state (n_0,n_1) that happen with the following intensities:

1) $\tilde{n} - d_0 \ge 0, n_0 \ge 1$: $(n_0,n_1) \to (n_0-1,n_1) = n_0\mu$,
2) $\tilde{n} - d_1 \ge 0, n_1 \ge 1$: $(n_0,n_1) \to (n_0,n_1-1) = n_1\mu$,
3) $\tilde{n} - d_1 + d_0 \le B, n_1 \ge 1$: $(n_0,n_1) \to (n_0+1,n_1-1) = n_1\theta_1$,
4) $\tilde{n} - d_0 + d_1 \ge 0, n_0 \ge 1$: $(n_0,n_1) \to (n_0-1,n_1+1) = n_0\theta_0$,
5) $\tilde{n} + d_0 \le \gamma B$: $(n_0,n_1) \to (n_0+1,n_1) = \lambda p_0$,
6) $\tilde{n} + d_1 \le \gamma B$ $(n_0,n_1) \to (n_0,n_1+1) = \lambda p_1$,

where p_i, $i = 0,1$, is the probability that a new session is in the LoS blocked or LoS non-blocked state, λ is the arrival intensity of new sessions, $\theta_i = 1/E[\Theta_i]$, $i = 0,1$, is the intensity of transitions between states. Here, 1) and 2) reflect the departures, 3) and 4) describe the transitions between session states, while 5) and 6) correspond to session arrivals.

To obtain the new session drop probability, we define

$$\Pi_{I,0} = \{(n_0,n_1) : n_0d_0 + n_1d_1 + d_0 > \gamma B\},$$
$$\Pi_{I,1} = \{(n_0,n_1) : n_0d_0 + n_1d_1 + d_1 > \gamma B\}. \quad (10)$$

Weighting the state probabilities where new session drops may occur, while using LoS non-blockage and blockage probabilities, p_0 and p_1, the new session drop probability, p_I, is

$$p_I = \sum_{n_0=0}^{\lfloor \frac{T}{d_0} \rfloor} \sum_{n_1=\lfloor \frac{T-d_0}{d_1} \rfloor}^{\lfloor \frac{B-n_0d_0}{d_1} \rfloor} \frac{p(n_0,n_1)}{p_0^{-1}} + \sum_{n_1=0}^{\lfloor \frac{T}{d_1} \rfloor} \sum_{n_0=\lfloor \frac{T-d_1}{d_0} \rfloor}^{\lfloor \frac{B-n_1d_1}{d_0} \rfloor} \frac{p(n_0,n_1)}{p_1^{-1}}. \quad (11)$$

Deriving the drop probability for the ongoing sessions is easier, since a session can only be dropped when the LoS path is blocked, i.e., at the moment of transition from (n_0,n_1) to (n_0+1,n_1-1). The ongoing session blockage subspace is

$$\Pi_O = \{(n_0,n_1) : n_0d_0 + n_1d_1 + d_0 - d_1 > B\}, \quad (12)$$

thus leading to the following ongoing session drop probability

$$p_O = \sum_{n_0=0}^{\lfloor \frac{B}{d_0} \rfloor} \sum_{N}^{\lfloor \frac{B-n_0d_0}{d_1} \rfloor} \frac{p(n_0,n_1)n_1\theta_0}{\frac{n_0}{(\mu+\theta_0)^{-1}} + \frac{n_1}{(\mu+\theta_1)^{-1}} + \frac{\mathbf{1}_{n_0d_0+n_1d_1 \le B-d_1}}{(p_1\lambda)^{-1}}}, \quad (13)$$

where $N = \min\left(1, \left\lfloor \frac{B-d_0+d_1}{d_1} \right\rfloor + 1\right)$ and $\mathbf{1}_A$ is indicator of A.

IV. Numerical Results

We illustrate the effects of the bandwidth reservation on both user- and network-centric characteristics. The parameters employed to obtain our results are provided in Table I.

To this aim, Fig. 3(a) shows the new and the ongoing session drop probabilities as functions of γ for the session bitrate of $R = 50$ Mbps. Observe that the new session drop probability increases as γ becomes larger. A reverse effect is noted for the ongoing session drop probability. Particularly, considering the case of $\lambda = 0.2$ and reserving only 4% of the entire bandwidth, the ongoing session drop probability decreases from 0.0013 down to 10^{-4} at the expense of increased new session drop probability from 0.0091 to 0.011. By reserving only 4% of bandwidth, we decrease the ongoing session drop probability by 93%, which is crucial for applications with high reliability requirements. These figures confirm our original hypothesis

TABLE I
System parameters

Parameter	Value
Heights of AP, UE, pedestrians, h_A, h_U, h_B	$4, 1.5, 1.7$ m
Radius of pedestrians, r_B	0.4 m
Speed of pedestrians, v	1 m/s
Run length of pedestrians, τ	10 m
Coverage radius of AP, r	107 m
Target SNR	3 dB
Density of pedestrians, λ_B	0.5 bl/m^2
Frequency	28 GHz
Bandwidth, B	1 GHz

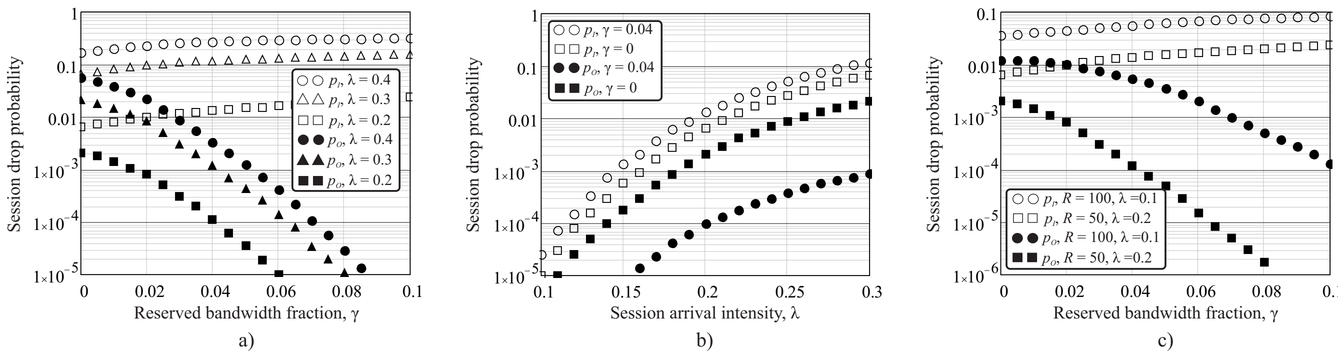


Fig. 3. Effects of reserved bandwidth fraction, γ, and intensity of sessions, λ, on drop probabilities for new and ongoing sessions.

that bandwidth reservation can control the trade-off between the two considered probabilities.

The impact of session arrival intensity, λ, is illustrated in Fig. 3(b). Observe that as λ increases the offered traffic load, $\rho = \lambda R$, grows as well, which leads to larger amounts of occupied resources and negatively affects both new and ongoing session drop probabilities. The drop probability curves for new and ongoing sessions have similar trends across all of the considered values of γ. This implies that the positive effect of bandwidth reservation is preserved over a wide range of traffic arrival rates. Further, Fig. 3(c) details the impact of session bitrate, R, on the drop probabilities for the same offered traffic load, $\rho = 10$. As one may learn, lower bitrates lead to better performance in terms of both new and ongoing session drop probabilities. This behavior is a direct consequence of the packing effect [14]. Furthermore, there is a faster decrease in the ongoing session drop probabilities for $R = 50$ Mbps as compared to $R = 100$ Mbps at the expense of a similar increase in the new session drop probabilities. The reason is that the currently available resources in the system are more likely to be occupied by a session changing its state from LoS non-blocked to LoS blocked rather than a new session.

Finally, we study the impact of bandwidth reservation on the system-level resource utilization, U. To this end, Fig. 4 displays U as a function of γ for several values of the session arrival rate λ and $R = 50$ Mbps. Clearly, the use of bandwidth reservation leads to slightly lower resource utilization. However, for the practical ranges of ongoing and new session drop probabilities, the said degradation is negligible. For instance, when $\lambda = 0.4$ the utilization drops from 0.823 to 0.817 as γ changes from 0 to 0.04. The associated decrease in the ongoing session drop probability is from 0.051 down to 0.0023.

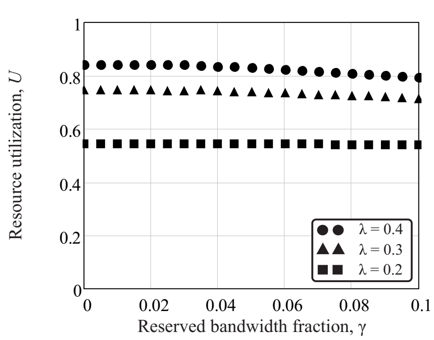


Fig. 4. Impact of bandwidth reservation on resource utilization.

V. Conclusions

To improve the ongoing session drop probability in mmWave systems under temporarily varying radio resource requirements due to dynamic blockage of the AP-to-UE LoS path, we introduced and evaluated the concept of bandwidth reservation. The main conclusions of this work are: (i) the use of bandwidth reservation allows reaching the desired trade-off between the ongoing and new session drop probabilities at a cost of negligible decrease in resource utilization, (ii) the effects of bandwidth reservation are preserved across a wide range of traffic arrival intensities, and (iii) the proposed scheme is sensitive to the mean session bitrate, i.e., the shorter the average session is, the less bandwidth is needed to achieve the desired balance between the session drop probabilities.

References

[1] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks [Editorial]," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, September 2017.

[2] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 14)," 3GPP TR 38.901 V14.3.0, December 2017.

[3] M. Gapeyenko *et al.*, "On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 124–10 138, November 2017.

[4] K. Haneda *et al.*, "5G 3GPP-like channel models for outdoor urban microcellular and macrocellular environments," in *IEEE 83rd Vehicular Technology Conference*, May 2016, pp. 1–7.

[5] J. G. Andrews *et al.*, "Modeling and analyzing millimeter wave cellular systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 403–430, January 2017.

[6] V. Petrov *et al.*, "Dynamic multi-connectivity performance in ultra-dense urban mmWave deployments," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2038–2055, September 2017.

[7] N. Seitz, "ITU-T QoS standards for IP-based networks," *IEEE Communications Magazine*, vol. 41, no. 6, pp. 82–89, June 2003.

[8] R. Guerin, "Queueing-blocking system with two arrival streams and guard channels," *IEEE Transactions on Communications*, vol. 36, no. 2, pp. 153–163, February 1988.

[9] Z. Ye *et al.*, "Predictive channel reservation for handoff prioritization in wireless cellular networks," *Computer Networks*, vol. 51, no. 3, pp. 798–822, February 2007.

[10] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 3, March 2005, pp. 1897–1907.

[11] R. Ford, M. Zhang, S. Dutta, M. Mezzavilla, S. Rangan, and M. Zorzi, "A framework for end-to-end evaluation of 5G mmwave cellular networks in ns-3," in *Workshop on NS-3*, June 2016, pp. 85–92.

[12] D. Moltchanov, "Distance distributions in random networks," *Elsevier Ad Hoc Networks*, vol. 10, no. 6, pp. 1146–1166, 2012.

[13] R. Groenevelt, "Stochastic models for mobile ad hoc networks," INRIA Sophia-Antipolis, PhD thesis, 2005.

[14] F. P. Kelly, "Loss networks," *The annals of applied probability*, pp. 319–378, 1991.