Performance Analysis of Mixture of Unicast and Multicast Sessions in 5G NR Systems

Andrey Samuylov[†], Dmitri Moltchanov[†], Alesia Krupko[‡], Roman Kovalchukov[†],
Faina Moskaleva[‡], and Yuliya Gaidamaka[‡⋆]
[†]Tampere University of Technology, Tampere, FIN-33101, Finland
[‡]Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
[⋆]Federal Research Center "Computer Science and Control" (FRC CSC RAS), Moscow, Russia

***Abstract*—3GPP New Radio (NR) air interface operating in a millimeter wave frequency band is expected to provide the main bearer service in the fifth generation (5G) mobile systems. Compensating for high propagation losses by using high gain antennas at both user equipment (UE) and access point (AP) sides these systems will greatly benefit from highly directional transmission serving unicast sessions. However, highly directional nature of NR communications may affect the conventional service procedures of multicast sessions in wireless networks as more than a single transmission may be required to serve UEs in the same multicast group. Accounting for random resource requirements induced by locations of UEs as well as human blockage phenomenon, we develop a model for performance analysis of 5G NR systems serving a mixture of unicast and multicast sessions. The main performance metrics of interest are drop probabilities of unicast and multicast sessions. The proposed model, complemented with antenna models and beam-steering procedure, can be further used to determine optimal AP intersite distance for 3GPP NR systems.**

***Index Terms*—5G, New Radio, mmWave, multicasting**

I. Introduction

3GPP New Radio (NR) wireless access technology is expected to form the basis of 5G systems providing exceptionally high data rates at the access interface. While the first standardization phase of LTE-anchored 3GPP NR is completed and the standalone NR is scheduled to be ready by July 2018, vendors and network operators promise to deploy first 3GPP NR system already by the end of 2018. Despite significant efforts spent in the past and partially due to the tight standardization schedule, the technology needs further optimizations for various communications scenarios, applications, and use-cases.

3GPP NR systems will operate in millimeter wave (mmWave) band. At these frequencies the propagation losses are extremely high calling for high gain large antenna arrays at both ends of a communications link [1]. In addition to extending coverage range, the use of large antenna arrays is expected to efficiently address the problem of interference [2]. However, highly directional transmissions may lead to inefficient use of radio resources when serving multicast sessions.

To decrease the use of radio resources, in conventional wireless access systems, e.g., LTE, multicasting is performed by organizing users in groups [3], [4]. Despite using the worst modulation and coding scheme (MCS) out of all active users in a group, the use of radio resources is still kept at minimal as there is no need to serve each user exclusively. Using highly directional transmissions this multicast scheme may not be effective as the number of users falling into the coverage of a single antenna configuration can be small and thus multiple transmissions may be required to serve all the users that belong to the same multicast group. The dynamic adaptation of the antenna radiation pattern, e.g., by decreasing the number of elements forming it when multicast sessions are served, results in significant decrease of the coverage range of NR AP requiring higher AP deployment density.

The problem of multicasting in mmWave systems has been recently addressed in several studies. In [5], the authors proposed a dynamic grouping scheme for UEs having the same multicast session based on their proximity. The associated heuristic algorithm is based on consecutive testing of different antenna half-power beamwidths (HPBW) maximizing the sum rate of the system. Among other conclusions, the authors demonstrated that the use of fixed antenna HPBW may lead to non-optimal performance. Similar approach has been proposed in [6]. In addition to the use of dynamic HPBW, the optimization framework developed in [7] accounts for non-equal power sharing between beams. A multicasting scheme based on non-orthogonal multiples access (NOMA) has been proposed in [8]. Finally, the problem of multicast transmissions in systems with directional antennas has been recently addressed in [9], where the authors proposed several transmission schemes minimizing the packets delay.

All the referenced studies exclusively concentrated on multicast sessions and did not consider the inherently contradicting requirements of unicast and multicast services. Thus, the model capable of capturing the service process at NR AP in presence of both unicast and multicast sessions is of special interest for practical dimensioning of future 5G networks.

In this paper, using the tools of queuing theory and stochastic geometry, we formulate the analytical model of NR AP simultaneously serving unicast and multicast sessions. The model explicitly captures propagation specifics of mmWave band, linear antenna arrays at both AP and UE, random resource requirements in-

duced by locations of UEs as well as human blockage phenomenon. The considered user- and system-centric metrics of interest include unicast and multicast session drop probabilities and system resource utilization. The proposed model can be used for planning of prospective NR systems. Particularly, for a given intensity of unicast and multicast sessions and the set of system parameters one may determine the optimal density of APs ensuring that a given area is fully covered and both unicast and multicast services are provided with prescribed drop probabilities.

The paper is organized as follows. In Section II we first discuss the problem of multicasting in NR systems and then introduce the service model of an NR AP. In Section III we propose the performance evaluation framework. Numerical results are provided in Section IV. The conclusions are drawn in the last section.

II. System Model for 3GPP NR Multicasting

In this section, we first discuss the trade-off involved in serving unicast and multicast users in NR AP. Then, we proceed formulating the system model by specifying its critical components including propagation, antenna and traffic models.

A. Coexistence of Unicast and Multicast Traffic

Fig. 1 illustrates the problem of coexistence of multicast and unicast sessions at a single NR AP. Both user equipment (UE), as well as AP, are expected to use antenna arrays featuring multiple elements forming directional radiation patterns by simultaneously using a part of them. As one may observe, increasing the number of elements forming radiation pattern we increase the coverage range of a single AP and thus increase inter-AP distance positively affecting the deployment costs. As long as only unicast sessions are concerned and neglecting the complexity of beam-steering procedure the lowest deployment cost is achieved by using the maximum available number of antenna elements.

Assume that in addition to unicast sessions, multicast ones need to be served. For multicast sessions, the optimal resource usage is achieved by using the least amount of antenna elements such that the radiation pattern covers the widest possible angle. However, this also results in the lowest transmit antenna gain drastically reducing the coverage of a single AP and thus increasing the deployment costs. Increasing the number of antenna elements forming a radiation pattern requires more than a single transmission to serve all multicast users even when they belong to the same multicast group. These additional transmissions are expected to affect the service process at NR APs increasing both unicast and multicast session drop probabilities.

B. System Model and Assumptions

1) Network Deployment: We consider a coverage area of a single AP. The height of AP is assumed to be

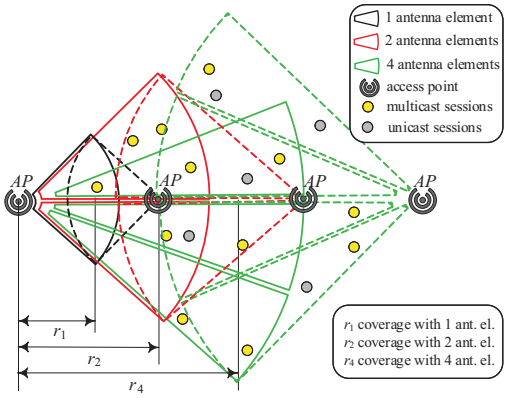


Fig. 1. Coexistence of unicast and multicast sessions at NR AP.

h_A meters. There are unicast and multicast sessions in the system. The location of a new multicast or unicast session is assumed to be uniformly distributed in the coverage area of NR AP with range r_A. The coverage range of AP, estimated in Section III-C, is the maximum separation distance between UE and AP such that UE is not in outage conditions. The height of UEs initiating any type of session is fixed at h_U.

2) NR AP Service Process: The system of interest is sketched in Fig. 2. The rates requested by multicast and unicast sessions are assumed to be constant, R_M Mbps and R_U Mbps, respectively. However, the amount of resources, measured in Hz, requested by multicast and unicast sessions, B_M and B_U, are random variables (RV) and depend on the distance between UE and AP, state of the link (blocked or non-blocked), propagation model and the set of modulation and coding schemes (MCS) used at the air interface. We use the set of MCSs specified for NR in Release 15. The cumulative probability distribution function (CDF) of B_M and B_U are determined in Section III-C and can be either discrete or continuous.

Denote by b_k, $k = 1, 2, .., K$, where $b_1 \leq b_2 \leq ... \leq b_K$, the amount of resources requested by a multicast session and by d the mean amount of resources requested by a unicast session. We also let ϕ_k, $k = 1, 2, .., K$, denote the probability that unicast session requires b_k resources. The AP is assumed to operate using the bandwidth of C, Hz. Multicast and unicast sessions arrive according to independent homogeneous Poisson processes with intensities λ_M and λ_U, respectively. The service times are assumed to follow exponential distributions with parameters μ_M and μ_U, respectively.

Let $U(t)$ be the number of resources occupied in the system at time t. The conventional service procedure is used for a unicast session, i.e., if upon arrival, $C - U(t) - B_U > 0$ is satisfied, the session is accepted to the system. Otherwise, it is dropped. For multicast sessions "transparent" service is assumed. Particularly, if upon arrival of a multicast session, there are no such

sessions in the system the session is accepted to the system given that there are sufficient resources, i.e., $C - U(t) - B_M > 0$. During the service time of a multicast session, other multicast sessions may arrive and join a multicast group. Let $B_{M,1}$ be the amount of resources requested by a session that initiated multicast service and let $B_{M,i}, i = 2, 3, ...$ be the amount of resources requested by sessions joining the multicast group during its service time. If $B_{M,2} > B_{M,1}$, then starting from arrival of the second multicast session the amount of resources increases by value $(B_{M,2}–B_{M,1})$ and the multicast service now occupies $[B_{M,1} + (B_{M,2}–B_{M,1})]$ resources. If these additional resources are not available, the second multicast session is blocked. The procedure is performed for the rest of multicast sessions arriving during the service time of a session that initiated the multicast service. The multicast service is stopped when a session that initiated it leaves the system.

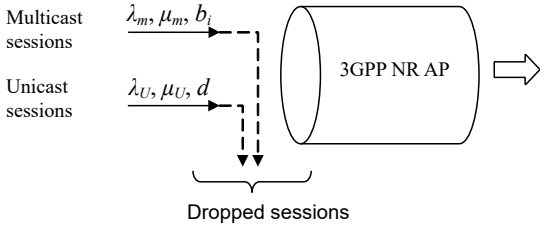


Fig. 2. Illustration of the considered system model.

3) Propagation, Blockage, and Antenna Models: To parameterize the network model, we need to relate the AP coverage range r_A and the amount of requested resources by unicast and multicast sessions to the system and environmental characteristics, taking into account, blockage, propagation, and antenna models.

MmWave propagation is known to be affected by human body blockage [10], [11]. We assume that humans (termed blockers) follow spatial Poisson process with intensity λ_B. Blockers are represented by cylinders with constant height and radius, h_B and r_B, respectively. We also utilize the standardized 3GPP UMi street-canyon model [12] with blockage enhancements, which delivers the signal-to-noise ratio (SNR) values for a specific separation distance in LoS blocked and LoS non-blocked conditions. Following measurements of human body blockage effects at mmWave frequencies [13], the LoS path occlusion by humans is assumed to result in $L_B = 20$ dB of additional degradation in the received signal strength.

We assume linear antenna arrays at AP and UEs with $N_{A,H}$ and $N_{U,H}$ elements, respectively, forming radiation patterns in horizontal direction. Following [14] the HPBW is related to the number of elements as $102/N_{A,H}$. The transmit antenna gain given by [14]

$$G = \frac{1}{\theta^+_{3db} - \theta^-_{3db}} \int_{\theta^-_{3db}}^{\theta^+_{3db}} \frac{\sin(N\pi\cos(\theta)/2)}{\sin(\pi\cos(\theta)/2)} d\theta, \quad (1)$$

where θ^-_{3dB} and θ^+_{3dB} are 3-dB points.

4) Metrics of Interest: We are interested in user- and system-centric performance metrics. The former include unicast and multicast session drop probability due to all resources occupied upon arrival. In case of multicast sessions, the drop probability accounts for those sessions initiating multicast service and those joining the ongoing multicast service. We also consider system resource utilization as a system metric defined as the limit of the ratio $U = \lim_{\to\infty} U(t)/C$.

III. Performance Evaluation Model

In this section, we develop NR AP model serving unicast and multicast sessions. We also complete parameterization of the model providing the estimates AP coverage range r_A and CDFs of the number of requested resources by multicast and unicast sessions B_M and B_U.

A. Queuing Model

Define the state of the system modeling the service process at NR AP as a vector $(n(t), r(t)), t > 0$, where $n(t)$ is number of unicast sessions in the system, $r(t)$ is the amount of resources occupied at time t. Let X be the state-space of the system defined as

$$X = \{(n, r) : 0 \le n \le \left\lfloor \frac{C}{d} \right\rfloor, 0 \le r \le C\}, \quad (3)$$

where C is the is the amount of resources in the system, d is the mean amount of resources requested by unicast sessions, $\lfloor \cdot \rfloor$ is the floor operator.

As the future evolution of the system depends on the current state and is independent of the previous

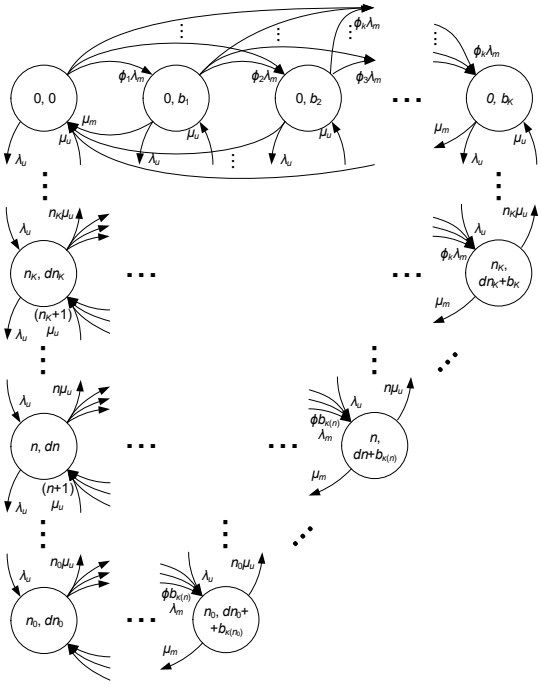


Fig. 3. State transition diagram of Markov chain.

$$
\begin{aligned}
&(\lambda_U+\lambda_M)p_{0,0}=\mu_u p_{1,d}+\mu_m\sum_{i=1}^{K}p_{0,b_i},\ n=0,\ k=0\\
&(\lambda_U+\lambda_M\sum_{i=k+1}^{K}\phi_i+\mu_M)p_{0,b_k}=\phi_k\lambda_m\sum_{i=1}^{k-1}p_{0,b_i}+\phi_k\lambda_M p_{0,0}+\mu_U p_{1,d+b_k},\ n=0,k=1,2,\ldots,K-1,\\
&(\lambda_U+\mu_M)p_{0,b_k}=\phi_K\lambda_m\sum_{i=1}^{K-1}p_{0,b_i}+\phi_K\lambda_M p_{0,0}+\mu_U p_{1,d+b_K},\ n=0,\ k=K,\\
&(\lambda_U+\lambda_M+n\mu_U)p_{n,nd}=(n+1)\mu_U p_{n+1,d(n+1)}+\mu_M\sum_{i=1}^{\kappa(n)}p_{n,nd+b_i}+\lambda_U p_{n-1,d(n-1)},\ n=1,2,...n_0-1,\ k=0,\\
&n_0\mu_U+\lambda_M p_{n_0,n_0 d}\sum_{i=1}^{\kappa(n_0)}\phi_i=\mu_M\sum_{i=1}^{\kappa(n_0)}p_{n_0,n_0d+b_i}+\lambda_U p_{n_0-1,d(n_0-1)},\ n=n_0\ k=0,\\
&(n\mu_U+\mu_M)p_{n,nd+b_{\kappa(n)}}=\phi_{\kappa(n)}\lambda_M\sum_{i=1}^{\kappa(n)-1}p_{n,nd+b_i}+\phi_{\kappa(n)}\lambda_M p_{n,nd}+\lambda_U p_{n-1,d(n-1)+b_{\kappa(n)}},\ n=n_K...n_0,k=\kappa(n),\\
&(\lambda_U+n\mu_U+\mu_M)p_{n,nd+b_{\kappa(n)}}=\phi_{\kappa(n)}\lambda_M\sum_{i=1}^{\kappa(n)-1}p_{n,nd+b_i}+\phi_{\kappa(n)}\lambda_M p_{n,nd}+(n+1)\mu_U p_{n+1,d(n+1)+b_{\kappa(n)}}+\lambda_U p_{n-1,d(n-1)+b_{\kappa(n)}},\\
&(\lambda_U+n\mu_U+\mu_M+\lambda_M\sum_{i=k+1}^{\kappa(n)}\phi_i)p_{n,nd+b_k}=\phi_k\lambda_M\sum_{i=1}^{k-1}p_{n,nd+b_i}+\phi_k\lambda_M p_{n,nd}+(n+1)\mu_U p_{n+1,d(n+1)+b_k}+\lambda_U p_{n-1,d(n-1)+b_k},\\
&(n\mu_U+\mu_M+\lambda_M\sum_{i=k+1}^{\kappa(n)}\phi_i)p_{n,nd+b_k}=\phi_k\lambda_M\sum_{i=1}^{k-1}p_{n_0,n_0d+b_i}+\phi_k\lambda_M p_{n_0,n_0d}+\lambda_U p_{n-1,d(n-1)+b_k}.
\end{aligned}
\tag{2}
$$

evolution, the process $\{(n(t),r(t)),t>0\}$ is Markov in nature. Furthermore, the state space of the system defined by (3) allows to obtain all the relevant performance characteristics including the number of unicast and multicast sessions in the system, the number of resources occupied by these sessions and unicast and multicast sessions drop probabilities.

The state transition diagram of the Markov chain $\{(n(t),r(t)),t>0\}$ is shown in Fig. 3. Using the local balance principle one may now deduce the set of linear equations describing the system at equilibrium as in (2), where the latter three equations are defined for

$$
\begin{aligned}
&n=1,\ldots,n_{K-1}-1,\ n_{K-1}+1,\ldots,n_0-1,\\
&n=1,...,n_0-1,\ k=1,\ldots,\kappa(n)-1,\\
&n=n_K,\ldots,n_0,\ k=\kappa(n+1)+1,...,\kappa(n)-1,
\end{aligned}
\tag{4}
$$

where $n_0=\lfloor\frac{C}{d}\rfloor, n_k=\lfloor\frac{C-b_k}{d}\rfloor$, and $\kappa(n)=\max_{k=0,\ldots,K}\{k:b_k\le C-nd\}, n=0,\ldots,n_0$

Define the steady-state probability vector as

$$
p^T=(p_{0,0},\ldots,p_{n_0,dn_0+b_{\kappa(n_0)}}),\ (n,r)\in X, \tag{5}
$$

and observe that the Markov process $\{(n(t),r(t)),t>0\}$ is non-reversible [15] implying that there is no closed-form solution for (2) complemented with the normalization condition. However, one may obtain p^T numerically using efficient algorithms, see, e.g., [16].

B. Performance Metrics

Having obtained the steady-state probability vector the metrics of interest immediately follows. Particularly, the mean amount of occupied resources in the system is provided by

$$
r^{(1)}=\sum_{r=0}^{C}r\sum_{n=0}^{[C/d]}p_{n,r}, \tag{6}
$$

and the resource utilization is then $U=r^{(1)}/C$.

Recall that a unicast session is dropped if, upon arrival of unicast session, there are resources in the system to serve it. The set of unicast session drop states is provided by

$$
B_U=\{(n,r)\in X: r+d>C\}, \tag{7}
$$

leading to the unicast session drop probability

$$
B_U=\sum_{(n,r)\in B_U}p_{n,r}. \tag{8}
$$

A multicast session is dropped if one of the following happens: (i) there are no multicast sessions in the system at the moment of arrival, and there are no resources to admit this session to the system, (ii) there is an ongoing multicast session in the system, the amount of resources required by a new arrival is higher than currently provided, and there are no such resources in the system. The set of multicast session drop states takes the following form

$$
B_m^k=\{(n,r)\in X: r-nd+b_k>C\}, \tag{9}
$$

leading to the multicast session drop probability

$$
B_M^k=\sum_{(n,r)\in B_M^k}p_{n,r}. \tag{10}
$$

C. AP Coverage and Resource Requests

To complete parameterization of the model we need to provide CDFs of the number of requested resources by multicast and unicast sessions, as well as the coverage range of AP. Below, we start with the latter.

To ensure full coverage of space by APs, r_A should be such that no outage happens at UE that is currently in blockage conditions and located at the distance r_A from AP. Let $S_{\min}$ be the minimum SNR a system may tolerate, i.e., in fact, $S_{\min}$ is the lower bound of the SNR range corresponding to the worst MCS. We have the following relation

$$S_{\min} = \frac{P_T G_T G_R}{N_0 L_B A}(r_A + [h_A - h_U]^2)^{-\gamma/2}, \quad (11)$$

where γ is the path loss exponent, h_A and h_U are the heights of AP and UE, P_T is the AP transmit power, G_T and G_R are the AP transmit and UE receive antenna gains, N_0 is the noise at receiver, $L_B = 20$ dB is the blockage-induced losses, A is the constant in the 3GPP path loss model converted to the form of $Ar^{-\gamma}$.

Solving (11) with respect to r_A we arrive at

$$r_A = \sqrt{(C_B/S_{min})^{\gamma/2} + (h_A - h_U)^2}. \quad (12)$$

Having obtained r_A we now proceed deriving CDFs of B_M and B_U. Since both CDFs are obtained similarly, we consider B_U as an example. We determine the sought CDF by first finding CDF of the number of requested resources in blockage and non-blockage conditions and then weighting them with corresponding probabilities. As the SNR in blockage and non-blockage conditions differ only by a constant factor, in what follows, we provide detailed derivation of CDF for blockage conditions.

Let S_{nB} be RV denoting the SNR in non-blockage conditions and $F_S(x), x > 0$, be its CDF. Recall, that the locations of new session arrivals are assumed to be uniformly distributed in the AP coverage. Thus, CDF of the distance between UE and AP is given by

$$F_D(y) = (y^2 - (h_A - h_U)^2)/r_A^2, \quad (13)$$

defined over $|h_A - h_U| < y < \sqrt{r_A^2 + (h_A - h_U)^2}$.

Since mmWave propagation model is monotonously decreasing in y, the distribution of SNR can be expressed in terms of distribution of the distance D, i.e.,

$$F_{S_{nB}}(y) = 1 - F_D([C_{nB}/y]^{\gamma/2}), \quad (14)$$

where $C_{nB} = P_T G_T G_R / N_0 A$.

The CDF of RV S_B denoting SNR in the blocked state is found similarly. To determine the overall SNR CDF we need the blockage probability. Following [10] the blockage probability at the distance x is given by

$$p_B(x) = 1 - e^{-2\lambda_B r_B (x \frac{h_B - h_U}{h_A - h_U} + r_B)}, \quad (15)$$

leading to the following weighted blockage

$$p_B = \int_0^{r_A} p_B(x)dx, \quad (16)$$

that can be evaluated in the closed-form.

The SNR CDF $F_S(y)$ can now be determined by weighting CDFs corresponding to blocked and non-blocked states with probabilities p_B and $1 - p_B$.

Let $s_k, k = 1, 2, .., K$, be SNR margins of MCS schemes, where K is a number of MCSs. Let also π_k be the probability that UE session is assigned to MCS i. Thus, we have

$$\pi_k = Pr\{s_k < s < s_{k+1}\} = F_S(s_{k+1}) - F_s(s_k). \quad (17)$$

Once π_k, $k = 1, 2, \ldots, K$, are available, the amount of resources corresponding to the rate R_U and R_M follows. Observe that the number of requested resources follow a general discrete RV.

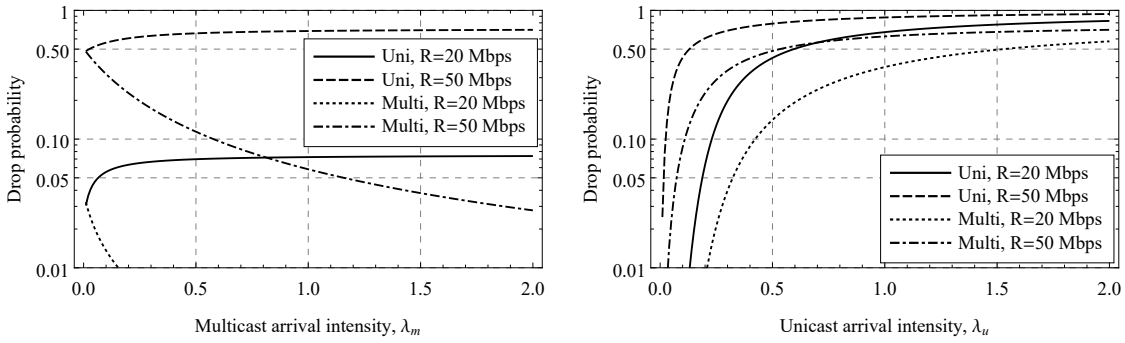
IV. Numerical Analysis

In this section, we carry out numerical investigation of the performance metrics provided to unicast and multicast sessions under different load conditions. The default parameters are summarized in Table I.

TABLE I
DEFAULT SYSTEM PARAMETERS.

Parameter	Value
Emitted power	0.2 W
Carrier frequency	28 GHz
Bandwidth	1 GHz
AP transmit antenna array	16×4 elements
UE receive antenna array	4×4 elements
Effective AP coverage	365 m
Attenuation by blockers	15 dB
Thermal noise	-84 dBm
Height of AP	4 m
Height of UE	1.5 m
Height of users (blockers)	1.7 m
Density of blockers	0.5 blockers/m^2
Default unicast arrival rate	0.1 sessions/s
Default multicast arrival rate	0.1 sessions/s
Default unicast service intensity	0.1 sessions/s
Default multicast service intensity	0.1 sessions/s
Mean bandwidth for 50Mbps rate	253.38 MHz
Mean bandwidth for 20Mbps rate	101.34 MHz

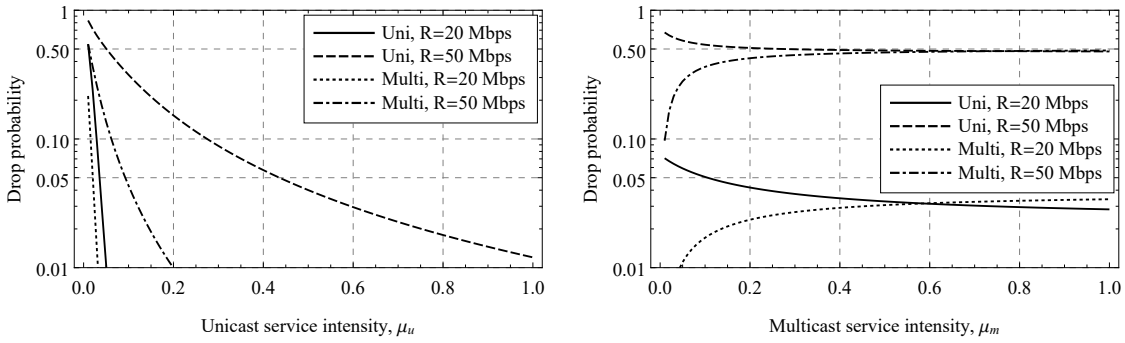
Fig. 4 demonstrates the unicast and multicast session drop probability for two coinciding rates of multicast and unicast session rates, $R_U = R_M = 20$ Mbps and $R_U = R_M = 50$ Mbps, as a function of arrival intensities of unicast and multicast arrival rates. As one may observe, the effect of the unicast arrival rate, shown in Fig. 4(a), is straightforward. Particularly, the increase in the arrival rate leads to a consistent increase in both multicast and unicast session drop probability for considered values of the session rates. The effect of multicast sessions arrival intensity, shown in Fig. 4(b), is drastically different. As the arrival rate of multicast



(a) As a function of unicast arrival intensity

(b) As a function of multicast arrival intensity

Fig. 4. Unicast and multicast session drop probabilities as a function of sessions arrival intensities.



(a) As a function of unicast service intensity

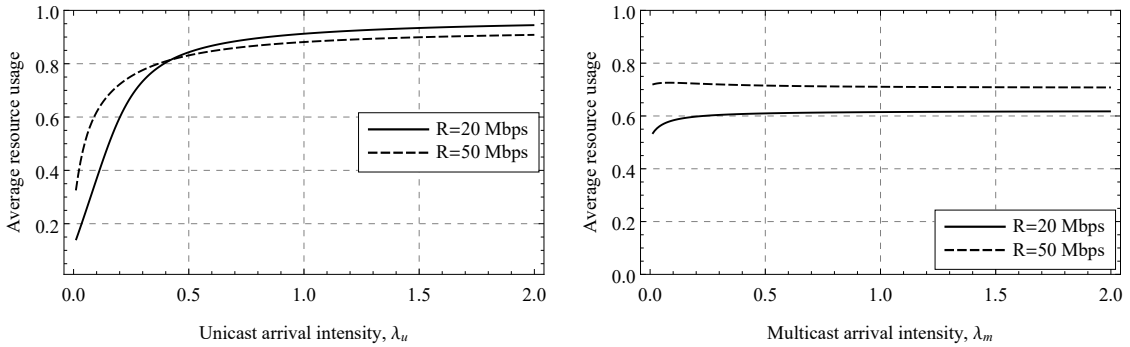
(b) As a function of multicast service intensity

Fig. 5. Unicast and multicast session drop probabilities as a function of sessions service intensity.

sessions increases the probability that a multicast session is going to be dropped decreases. This is explained by the increasing competition for the resources from the multicast session and the nature of the service process of this type of traffic. Indeed, once resources are occupied by a session that initiates multicast service additional sessions joining the group requires much fewer, "on average", resources for admission to the system, compared to the initial session. Expectedly, as multicast session drop probability decreases the unicast session drop probability increases.

Consider now the response of the system in terms of drop probabilities to different service rates of multicast and unicast sessions, μ_M and μ_U, illustrated in Fig. 5 for two values of the session rates, 20 Mbps and 50 Mbps. Recall, that the mean service time is related to the service intensity as $1/\mu$, and thus decreases along OX axis in Fig. 5. As one may observe, both multicast and unicast session drop probabilities decrease as unicast session service rate increases. Furthermore, the increased rate of sessions also negatively affects both drop probabilities. The effect of multicast service rate is more complicated. The increase in μ_M leads to the different changes in drop probabilities: unicast session drop probability decreases while the multicast session drop probability increases. This behavior is explained by the fact that increasing the session service intensity of multicast session leads to shorter, "on average", service times and more multicast sessions start to arrive in the system when there is no ongoing multicast service (group) available and thus compete for the service resources similarly to unicast sessions. Note that the described trends are inherent for any of multicast and unicast session rates. A curious reader notices that unicast and multicast curves corresponding to the same rates eventually coincide. This is explained by the fact that at extremely small mean session service times, the service process of multicast sessions becomes almost indistinguishable from that of unicast sessions while the rates of unicast and multicast sessions coincide.

Consider now the system-centric metric – mean resource utilization illustrated in Fig. 6 for different session rates as a function of unicast and multicast arrival intensities. Analyzing Fig. 6(a) one may observe that for both session rates the mean resource utilization of the system increases. Expectedly, higher utilization is achieved when smaller rates are used. This behavior is a direct consequence of the packing effect manifesting itself in higher statistical multiplexing gains for smaller session rates. Similarly to session drop probabilities, more interesting effects are observed for varying multicast session arrival rates. Particularly, starting from small values of multicast session arrival rates the system resource utilization increases along with λ_M. However, it then reaches a certain limit dictated by the system parameters and further remains



(a) As a function of unicast arrival intensity

(b) As a function of multicast arrival intensity

Fig. 6. System resource utilization as a function of sessions service intensity.

almost independent of the multicast session arrival rate. The reason behind this behavior is that, once a specific arrival rate is reached, the system almost always have a multicast service (group) running and multicast sessions joining it do not further improve resource utilization.

V. Conclusions

Conventional multicasting approach in wireless systems dramatically benefits from the use of nearly omni-directional antennas. Thus, the use of highly directional radiation patterns at both side of a communications link in prospective 3GPP NR systems may result in inefficient use of radio resources. To provide the tool for dimensioning of such systems we have developed a model of the service process at NR AP serving a mixture of unicast and multicast traffic that takes into account propagation specifics of a mmWave band, linear antenna arrays at both AP and UEs, stochastic locations of users as well as human blockage. We have derived both user- and system-centric metrics of interest including unicast and multicast session drop probabilities and system resource utilization.

Our numerical results reveal several interesting trade-offs and dependencies. Notably, in the presence of unicast service in the system, the increase in multicast session arrival rate and/or mean service time results in a decrease of the multicast session drop probability. However, the response of the system to unicast arrival process parameters is straightforward. The proposed model can be used for dimensioning of prospective NR systems. Particularly, for a given intensity of unicast and multicast sessions, one may determine the optimal density of APs ensuring that a given area is fully covered for both unicast and multicast services and prescribed session drop probabilities are not violated.

Acknowledgment

The publication has been prepared with the support of the "RUDN University Program 5-100" and funded by RFBR according to the research projects No. 18-07-00156, 18-37-00380. This work has been developed within the framework of the COST Action CA15104, Inclusive Radio Communication Networks for 5G and beyond (IRACON).

References

[1] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results," *IEEE communications magazine*, vol. 52, no. 2, pp. 106–113, 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5g be?," *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] M. Condoluci, G. Araniti, A. Molinaro, and A. Iera, "Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in lte networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2907–2921, 2016.

[4] Y. Zhou, H. Liu, Z. Pan, L. Tian, and J. Shi, "Spectral- and energy-efficient two-stage cooperative multicast for lte-advanced and beyond," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 34–41, 2014.

[5] H. Park, S. Park, T. Song, and S. Pack, "An incremental multicast grouping scheme for mmwave networks with directional antennas," *IEEE Communications Letters*, vol. 17, no. 3, pp. 616–619, 2013.

[6] W. Feng, Y. Li, Y. Niu, L. Su, and D. Jin, "Multicast spatial reuse scheduling over millimeter-wave networks," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2017 13th International*, pp. 317–322, IEEE, 2017.

[7] K. Sundaresan, K. Ramachandran, and S. Rangarajan, "Optimal beam scheduling for multicasting in wireless networks," in *Proceedings of the 15th annual international conference on Mobile computing and networking*, pp. 205–216, ACM, 2009.

[8] Z. Zhang, Z. Ma, Y. Xiao, M. Xiao, G. K. Karagiannidis, and P. Fan, "Non-orthogonal multiple access for cooperative multicast millimeter wave wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1794–1808, 2017.

[9] A. Biason and M. Zorzi, "Multicast transmissions in directional mmwave communications," in *European Wireless 2017; 23th European Wireless Conference; Proceedings of*, pp. 1–7, VDE, 2017.

[10] M. Gapeyenko, A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, E. Aryafar, S.-p. Yeh, N. Himayat, S. Andreev, and Y. Koucheryavy, "Analysis of human-body blockage in urban millimeter-wave cellular communications," in *Communications (ICC), 2016 IEEE International Conference on*, pp. 1–7, IEEE, 2016.

[11] M. Gapeyenko, A. Samuylov, M. Gerasimenko, D. Moltchanov, S. Singh, M. R. Akdeniz, E. Aryafar, N. Himayat, S. Andreev, and Y. Koucheryavy, "On the temporal effects of mobile blockers in urban millimeter-wave cellular scenarios," *IEEE Transactions on Vehicular Technology*, 2017.

[12] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 14)," 3GPP TR 38.901 V14.1.1, July 2017.
[13] K. Haneda *et al.*, "5g 3gpp-like channel models for outdoor urban microcellular and macrocellular environments," in *Vehicular Technology Conference (VTC Spring), 2016 IEEE 83rd*, pp. 1–7, IEEE, 2016.
[14] A. B. Constantine *et al.*, "Antenna theory: analysis and design," *Microstrip Antennas (third edition), John Wiley & Sons*, 2005.
[15] L. Saloff-Coste, "Lectures on finite markov chains," in *Lectures on probability theory and statistics*, pp. 301–413, Springer, 1997.
[16] W. J. Stewart, *Numerical solution of Markov chains*, vol. 8. CRC press, 1991.