Improved Session Continuity in 5G NR with Joint Use of Multi-Connectivity and Guard Bandwidth

Roman Kovalchukov†, Dmitri Moltchanov†, Vyacheslav Begishev⋆, Andrey Samuylov†,
Sergey Andreev†, Yevgeni Koucheryavy†, and Konstantin Samouylov⋆
†Tampere University of Technology, Tampere, Finland
⋆RUDN University, Moscow, Russian Federation

***Abstract*—The intermittent millimeter-wave radio links as a result of human-body blockage are an inherent feature of the 5G New Radio (NR) technology by 3GPP. To improve session continuity in these emerging systems, two mechanisms have recently been proposed, namely, multi-connectivity and guard bandwidth. The former allows to establish multiple spatially-diverse connections and switch between them dynamically, while the latter reserves a fraction of system bandwidth for sessions changing their state from non-blocked to blocked, which ensures that the ongoing sessions have priority over the new ones. In this paper, we assess the joint performance of these two schemes for the user- and system-centric metrics of interest. Our numerical results reveal that the multi-connectivity operation alone may not suffice to increase the ongoing session drop probability considerably. On the other hand, the use of guard bandwidth significantly improves session continuity by somewhat compromising new session drop probability and system resource utilization. Surprisingly, the 5G NR system implementing both these techniques inherits their drawbacks. However, complementing it with an initial AP selection procedure effectively alleviates these limitations by maximizing the system resource utilization, while still providing sufficient flexibility to enable the desired trade-off between new and ongoing session drop probabilities.**

I. Introduction

Millimeter-wave (mmWave) radio technology is expected to become the foundation of fifth-generation (5G) mobile systems by enabling unprecedentedly high data rates and low latencies at the air interface [1]. As 3GPP standardization process of the New Radio (NR) technology is nearing completion and vendors perform test trials to showcase the capabilities of this emerging system design, researchers explore the challenges related to enabling advanced options in 5G NR deployments [2].

Along these lines, 3GPP considers multi-connectivity mechanisms to support target applications with higher reliability requirements in the prospective NR systems. Accordingly, the user equipment (UE) maintains multiple spatially-diverse links with different mmWave access points (APs) [3]. Multi-connectivity is expected to mitigate the effects of line-of-sight (LoS) blockage from various obstacles in the mmWave channel by dynamically re-associating the UE with currently non-blocked APs [4]. The authors in [5] identify capacity and outage probability as the two metrics that can be improved by applying multi-connectivity operation. The corresponding capacity gains have been quantified in [6] by utilizing simulations. In [7], the signal-to-interference-plus-noise ratio (SINR) distribution in the presence of multi-connectivity has been obtained. A comparison of alternative AP switching strategies has been conducted in [8]. Offloading performance for mission-critical traffic over NR deployments in vehicular environments with multi-connectivity capabilities has been addressed in [9].

Another recently proposed option to improve session-level reliability in 3GPP NR systems is the use of guard bandwidth [10]. For user applications with strict data-rate requirements, the transition of LoS mmWave link from non-blocked to blocked state may cause session interruptions, since the NR AP may not have sufficient radio resources at the moment of such transition, which can eventually lead to session drops. From the UE quality-of-service (QoS) perspective, it is preferable to reject a session at the moment of its arrival rather than discontinue it during service [11]. In the presence of uncertainty about the channel state at each active UE, it is reasonable to reserve dedicated radio resources for the ongoing sessions by making only their certain fraction available for the arriving sessions. Once accepted, the session enjoys access to the entire pool of radio resources. In a single NR AP scenario, the authors in [10] demonstrated that reserving only 5% of system resources allows decreasing the session drop probability by several orders of magnitude.

The performance benefits brought by the multi-connectivity and guard bandwidth mechanisms are not limited to the cell-edge users but also support UEs in favorable channel conditions even when some of the mmWave signal paths are blocked. A joint utilization of multi-connectivity and guard bandwidth may offer notable session continuity improvements in 3GPP NR systems. In this paper, we analyze such joint use of guard bandwidth and multi-connectivity operation in NR-based access systems. We consider a square area of interest housing a dense crowd of people, who act as blockers to mmWave radio links. The two mechanisms in question are parametrized by the degree of multi-connectivity, M (i.e., the maximum number of simultaneously allowed UE-AP links) and the guard bandwidth fraction, γ, which is reserved at each mmWave AP. The considered performance metrics are (i) new

The publication was prepared with the support of the "RUDN University Program 5-100" and funded by RFBR according to the research projects No. 17-07-00845, 18-07-00576, 18-37-00380. The work was developed within the framework of the COST Action CA15104, Inclusive Radio Communication Networks for 5G and beyond (IRACON). This work was also supported by the Academy of Finland (projects WiFiUS and PRISMA) and by the project TAKE-5: The 5th Evolution Take of Wireless Communication Networks, funded by Tekes.

session drop (rejection) probability, (ii) ongoing session drop probability, and (iii) system (resource) utilization.

A summary of the subsequent study is as follows:

- multi-connectivity operation alone does not allow to improve the ongoing session drop probability, since the latter decreases slightly as more spatially-diverse links are allowed for the UEs;
- guard bandwidth mechanism allows to significantly decrease the ongoing session drop probability at the expense of a comparable increase in the new session drop probability and subject to lower resource utilization;
- joint use of multi-connectivity and guard bandwidth provides sufficient flexibility to reach the desired trade-off between the drop probabilities for new and ongoing sessions at the cost of decreased resource utilization; however, allowing the UE to choose its AP with sufficient resources at the session initialization phase alleviates this limitation by improving the system resource utilization.

The rest of this paper is organized as follows. First, in Section II we formulate the system model. Then, in Section III we outline the developed performance assessment framework. Numerical results and implementation guidelines follow in Section IV. Conclusions are drawn in the last section.

II. System Model

In this section, we introduce our system model and its components, which include (i) network deployment and user models, (ii) propagation and beamforming models, as well as (iii) traffic, association, and service models. Further, user- and system-centric metrics of interest are determined.

A. Network Deployment and User Models

We consider a deployment of NR APs illustrated in Fig. 1. The circular area of radius r is chosen specifically to facilitate a fair comparison of multi-connectivity strategies with different numbers of simultaneously supported links. Accordingly, N NR APs are assumed to be located on the circumference equally separated in space by $2\pi r/N$. The height of all APs is constant and equals h_A.

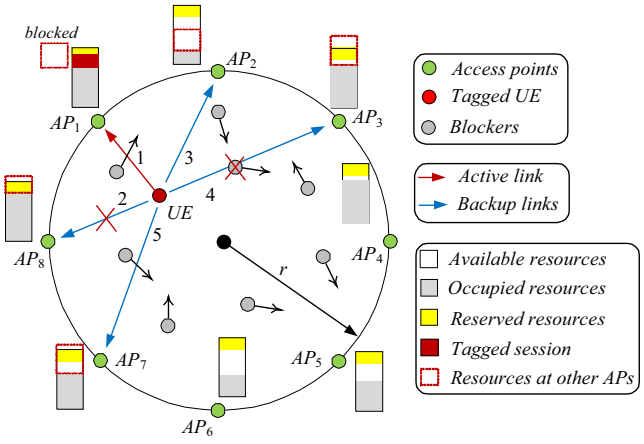


Fig. 1. Considered deployment: multi-connectivity and guard bandwidth.

The density of potential users is λ users per square meter, and they are assumed to all move within the service area according to a random walk model. The flux of users across the zone boundary is assumed to be constant. We follow random direction mobility (RDM, [12]) model with speed v m/s and exponentially distributed run length with mean τ m. Users are modeled as cylinders with height h_B and radius r_B. Each user is assumed to carry one UE. The height of UEs is h_U. Moving users act as blockers to those having active sessions.

B. Propagation and Beamforming Models

The received signal power at the UE can be written as

$$P_R(x) = P_T + G_T + G_R - L(x), \tag{1}$$

where P_T is the transmit power, G_T and G_R are the antenna gains at the transmit and receive sides, respectively, which depend on the antenna array, x is the distance between the UE and the AP, and $L(x)$ is the propagation loss.

The LoS path between the UE and the NR AP might be occluded temporarily by other pedestrians. Depending on the current link state (LoS blocked or non-blocked) as well as the distance between the NR AP and the UE, the running session employs an appropriate modulation and coding scheme (MCS) to maintain reliable data transmission. We also utilize the standard 3GPP urban micro (UMi) street canyon model [13] with blockage enhancements that provide path loss for a certain separation distance with and without blockage. Particularly, the path loss measured in dB is

$$L(x) = \begin{cases} 32.4 + 21\log(x) + 20\log f_c, & \text{non-blocked} \\ 52.4 + 21\log(x) + 20\log f_c, & \text{blocked}, \end{cases} \tag{2}$$

where x is the distance and f_c is the carrier frequency in GHz.

To complete the parametrization of our propagation model, one requires the antenna gains G_T and G_R together with the corresponding directivity angles α_T and α_R. In this paper, we assume linear antenna arrays at both transmit and receive sides. Half-power beamwidth (HPBW) of the array, α, is proportional to the number of its elements, K, in the appropriate plane and can be established by

$$\alpha = 2|\theta_m - \theta_{3db}|, \tag{3}$$

where θ_{3db} is the 3-dB point and θ_m is the location of the array maximum. The latter is computed as $\theta_m = \arccos(-\beta/\pi)$. Assuming $\beta = 0$, we have $\theta_m = \pi/2$. The upper and lower 3-dB points are thus

$$\theta^{\pm}_{3db} = \arccos[-\beta \pm 2.782/(K\pi)]. \tag{4}$$

For $\beta = 0$, the mean antenna gain over HPBW is then [14]

$$G = \frac{1}{\theta^{+}_{3db} - \theta^{-}_{3db}} \int_{\theta^{-}_{3db}}^{\theta^{+}_{3db}} \frac{\sin(K\pi\cos(\theta)/2)}{\sin(\pi\cos(\theta)/2)} d\theta. \tag{5}$$

C. Traffic, Association, and Service Models

Once a user initiates new session, it remains stationary for its entire duration. Let $p_A = \lim_{\Delta t \to 0} p(\delta t/\Delta t)$ be the probability that the UE initiates a session. Using the superposition property of point processes, we observe that the session arrival process is Poisson with the intensity of $p_A \lambda \pi r^2$ [15]. The session holding time is assumed to be exponential with the parameter μ. The session rate is constant and set to R Mbps. The traffic is assumed to be non-elastic.

Our system with multi-connectivity, guard bandwidth, and AP selection (at the session initialization phase) capabilities is assumed to operate as follows. Let M be the number of links that are simultaneously supported by the UE, named here the degree of muticonnectivity. During an active connection, other users moving across the area might occlude the LoS from UE to its current NR AP. Each AP has a dedicated fraction of its resources, termed guard bandwidth, $\gamma \in (0, 1)$, available to only its active sessions. The NR AP is assumed to operate over the bandwidth of B Hz.

Upon arrival, the UE is assumed to establish an active connection with M APs that have the strongest signal-to-noise ratio (SNR). Let $U_i(t)$ be the amount of radio resources currently occupied at the APs, which this UE is associated to. If $B(1-\gamma) - U_1(t) - W_1 > 0$, where W_1 is the amount of resources requested from the AP having the strongest SNR, the session is accepted. If not, UE attempts to associate with the AP with the second best SNR by checking $B(1-\gamma) - U_2(t) - W_2 > 0$. The session is dropped if/when no sufficient radio resources are available at all M APs. Otherwise, it is accepted by the AP having the strongest SNR and sufficient amount of resources.

Whenever UE state changes from non-blocked to blocked, it checks whether there is sufficient amount of radio resources at the current NR AP i.e., $B - U_i(t) - W_i > 0$. If true, the session continues its service at the current AP. In case of a negative outcome, the remaining APs are attempted starting from the one having the strongest SNR. If there are no APs that satisfy these conditions, the session in question is dropped.

D. Metric of Interest

The considered metrics of interest are user-centric, which includes new and ongoing session drop probabilities as well as system-centric, which includes the average resource utilization in the entire system. The latter is defined as

$$E[U] = \frac{1}{N} \lim_{t \to \infty} \sum_{i=1}^{N} U_i(t). \tag{6}$$

III. Performance Assessment Framework

In this section, we introduce the developed performance evaluation framework. First, we motivate and briefly describe our proposed methodology. Then, we proceed by characterizing stochastic properties of the dynamic blockage process. Finally, we outline the simulation framework together with statistical methods employed for data collection and analysis.

A. Methodology Description

The challenge of modeling the service process of NR APs in the presence of dynamic blockage, multi-connectivity, and guard bandwidth is extremely complex to address with either computer simulations or analytical techniques. In computer simulations of the system at hand, the bottleneck is related to modeling the dynamic blockage phenomena as discussed in [16]. Analyzing the multi-connectivity option alone leads to complex models in terms of a queuing network – with no closed-form solution – as we recently demonstrated in [8], [9]. Hence, in what follows, we apply a mixed semi-analytical framework, where we first produce stochastic characteristics of the dynamic blockage process and then apply simulation methods to capture queuing performance.

B. Blockage Process Characterization

Here, we derive stochastic characteristics of the blockage process on a link between the NR AP and the UE in the presence of a human crowd moving around the UE with an active session. Particularly, we derive the probability density functions (pdfs) of the blockage and non-blockage time intervals. These distributions are then utilized by the developed simulation framework to characterize the blockage process with respect to the NR AP.

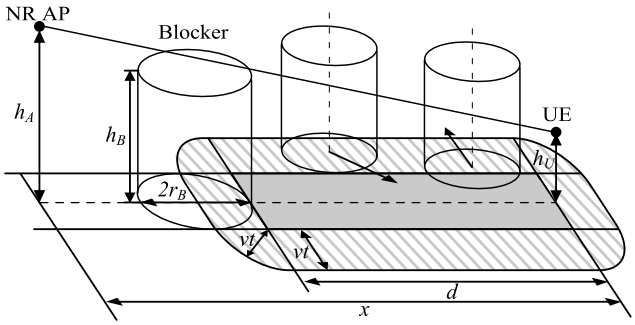


Fig. 2. Illustration of LoS blockage zone.

To this aim, Fig. 2 demonstrates the considered blockage model where AP and UE have the heights of h_T and h_U, respectively. Recalling that humans are modeled as cylinders with radius r_B, the mmWave LoS path is blocked whenever at least one center of a cylinder appears in the LoS blockage zone of a rectangular form with the sides of r_B and

$$d = \frac{x\,(h_B - h_U)}{h_A - h_U} + r_B, \tag{7}$$

where x is the two-dimensional distance between the AP and the UE, while h_B is the height of blockers.

After identifying the configuration of the LoS blockage zone, we now determine stochastic characteristics of the process of blockers entering it. Recall that blockers are assumed to move around the area of interest according to the RDM model. It has been shown in [17] that the inter-meeting times between a stationary UE and mobile UEs moving inside the specified area according to RDM approximately follow an

exponential distribution, thus resulting in a Poisson process of meetings. Using this result, we can proceed by characterizing the intensity of the Poisson process of blockers entering the LoS blockage zone.

The arrival intensity of blockers entering the LoS blockage zone can be approximated by the average intensity of blockers falling into a unit time increment area of LoS blockage zone weighted with the probability of moving towards the zone as illustrated in Fig. 2. Recalling that the speed of blockers is constant, v, the temporal intensity of blockers entering the LoS blockage zone can be approximated by

$$\lambda \approx \lambda_B \left(\pi v^2 + 2dv + 2r_B v\right) w, \tag{8}$$

where the coefficient w reflects the probability of a blocker crossing the boundary of the LoS blockage zone in a unit time. For the LoS blockage zone in Fig. 2, $w \approx 2/5$ leads to an accurate approximation.

Knowing the structure of the process of blockers entering the LoS blockage zone, we continue with characterizing the structure of the blockage process. In [16], the authors showed that under the assumption that the process of blockers entering the LoS blockage zone is Poisson, the distribution of non-blocked intervals follows an exponential distribution with the mean of $1/\lambda$. Observe that the blocked interval may comprise several overlapping intervals caused by individual blockers. Hence, its distribution coincides with the busy period distribution in the M/G/∞ queue [18], which is

$$\begin{aligned} F_B(t;x) = 1 - \Bigg(& \int_0^t (1 - F_B(t-z;x)) | d e^{-\zeta(x) F_T(z;x)} | \\ & - [1 - F_T(t;x)] \int_0^t [1 - F_B(t-z;x)] e^{-\zeta(x) F_T(z;x)} \zeta dz \\ & + [1 - F_T(t;x)] \Bigg), \end{aligned} \tag{9}$$

where the service time is the blockage time of a single blocker and x is a parameter characterizing the distance between the AP and the UE.

Let L be the distance traveled by a single blocker inside the LoS blockage zone. Assume that the run length in the RDM model is much larger than the width of the LoS blockage zone, $2r_B$. In this case, the probability that a blocker changes its direction of movement inside the LoS blockage zone is negligible. Observing that the entrance points of blockers are distributed uniformly along the perimeter of the LoS blockage zone, the distribution in question has the following cumulative distribution function (CDF)

$$F_L(x) = \begin{cases} 0, x \leq 0 \\ \frac{F_L^1(x)}{w_1} + \frac{F_L^2(x)}{w_2}, 0 < x \leq \sqrt{4r_B^2 + r^2} \\ 1, x > \sqrt{4r_B^2 + r^2}, \end{cases} \tag{10}$$

where the coefficients w_1 and w_2 as well as the CDFs $F_L^1(x)$ and $F_L^2(x)$ were provided in [16]. The blockage time interval can then be established by scaling the CDF in (10) with the speed value of v [19].

C. Simulation Framework, Data Collection, and Analysis

The developed framework is based on discrete-event simulation (DES) techniques. The software is developed by using the Java programming language with multi-thread optimizations [20]. The simulation procedure comprises two phases: DES simulations and data analysis. The developed DES framework implements the system model described in Section II. New arrival events are generated according to a spatial Poisson process where the positions are distributed uniformly within the area of interest. To decrease the computational complexity, each active session is associated with several independent stochastic processes that describe the blockage with each available NR AP. When processing the arrival events, the session completion event and the subsequent arrival event are scheduled.

To comprehensively evaluate the gains provided by multi-connectivity and guard bandwidth mechanisms, the following approach has been followed. First, for a given set of input parameters, the systems implementing multi-connectivity and guard bandwidth in isolation have been simulated. Then, the system implementing both techniques was analyzed. Finally, the system with multi-connectivity, guard bandwidth, and initial AP selection considerations, which allows choosing the AP with sufficient radio resources at the session initialization phase, was evaluated.

If there are no resources to serve a session, the respective connection request is dropped. When the UE is allowed to choose the AP at its session initialization, it first measures the signal strength from all the APs and adds M of them to the checklist in descending order. The UE then attempts the APs from the list one-by-one, and only if all of M APs do not have sufficient resources to serve this session, it is considered dropped. During the active session time, the LoS blockage state change events may occur, thus leading to resource re-allocations at NR APs. If there are not enough resources at the serving AP, the process of AP re-selection is initiated. Residence time distributions in blocked and non-blocked LoS states have been derived in subsection III-B.

A simulation campaign has also been carried out to obtain the metrics of interest by relying on the following procedure. For each considered set of input parameters, simulations were set to run for 10^6 seconds of the system time. As all of the involved processes (arrival, service, blockage) are stationary, the steady-state always exists in our system. The starting point of the steady-state period has been detected by utilizing exponentially-weighted moving average (EWMA) statistics with the weighting parameter set to 0.05 and employing the procedure in [20].

The statistical data has been collected only during the steady-state period. To remove residual correlations in the statistical data, we have used the batch means strategy. Accordingly, the entire steady-state period duration has been divided into 1000 data blocks. The metrics of interest computed for these periods became inputs to the individual statistical samples. The final values for the metrics of interest have been

TABLE I
NOTATION USED IN THIS PAPER.

Parameter	Value
Number of NR APs, N	8
Carrier frequency, f_c	28 GHz
Bandwidth at each NR AP, B	1 GHz
Transmit power, P_T	0.2 W
Number of planar antenna elements at AP, K_A	32
Number of planar antenna elements at UE, K_U	4
Radius of coverage area, r	100 m
LoS blockage loss, L_B	15 dB
NR AP height, h_A	4 m
UE height, h_U	1.5 m
Blocker height, h_B	1.7 m
Blocker radius, r_B	0.3 m
User density, λ	0.5 users/m^2
Blocker velocity, v_B	4 m/s
Session rate, R	100 Mbps
Mean session time, $1/\mu$	20 s
Session initiation probability, p_A	3.14×10^{-4}
Guard bandwidth, γ	$(0.0, 0.01 \ldots, 0.1)$
Degree of multi-connectivity, M	$(1, 2, \ldots, 8)$

estimated by processing these samples. Due to the large size of statistical samples associated with experiments, only the point estimates are shown. The interval estimates computed for the selected input parameters do not deviate by more than ± 0.001 from the point estimates under the level of significance set to $\alpha = 0.05$ and thus are not plotted in the presented graphs.

IV. NUMERICAL RESULTS

In this section, we report on the joint performance of multi-connectivity and guard bandwidth capabilities in a dense NR-based deployment introduced in Section II. First, we assess the metrics of interest when these mechanisms are applied individually. Then, we proceed by discussing the results when both mechanisms are activated simultaneously. Finally, we show the gains of allowing for AP selection at the session initialization phase. Here, the default system parameters are provided in Table I. Note that the session initiation probability, p_A, is chosen such that the intensity of new sessions is approximately 10 sessions per second. To parameterize the antenna arrays at the AP and the UE sides, we utilize the number of planar antenna array elements, K_A and K_U, respectively, and then follow (3)-(5) to determine the HPBWs, α_A and α_U, as well as the antenna gains, G_A and G_U.

A. Mechanisms in Isolation

First, consider the system performance when only the multi-connectivity operation is utilized. Fig. 3 illustrates the new and ongoing session drop probabilities as well as system utilization for different values of the degree of multi-connectivity, M. As one may observe, increasing the degree of multi-connectivity leads to better resource utilization. For the considered set of input parameters, it nearly approaches 0.8 for $M = 8$. This trend is accompanied by a decrease in the ongoing session drop probability. Indeed, as M grows, more options become available for a session that changes its state from non-blocked

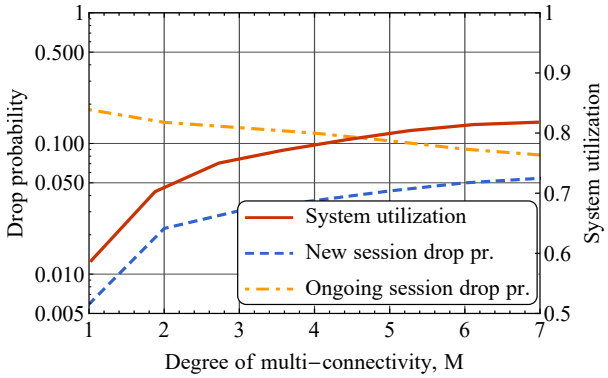


Fig. 3. System and UE performance in presence of multi-connectivity.

to blocked. However, this comes at the expense of drastically increased new session drop probability. For practical values of M, $M = 2$ or $M = 3$, the associated loss in the new session drop probability is higher than the achieved gain in the ongoing session drop probability. Hence, we conclude that multi-connectivity alone may not provide the desired flexibility in achieving the target trade-off between these two user-centric metrics.

We continue with analyzing the effects of guard bandwidth in isolation. The system- and user-centric metrics of interest for this case are offered in Fig. IV-A. First, we observe that the system resource utilization remains almost unchanged for the entire range of the considered guard bandwidth values and never reaches $U = 0.82$ achieved by using $M = 8$ in Fig. 3. Further, we notice that an increase in the amount of bandwidth reserved for the sessions already accepted for service results in a gradual and steady decrease in the ongoing session drop probability. Clearly, new session drop probability grows as we increase the value of the guard bandwidth.

It is important that the rate of changes in the considered user-centric metrics is exponential. By reserving only 2% of radio resources, one can decrease the ongoing session drop probability from 0.38 to 0.09 at the cost of an increase in the new session drop probability from 0.06 to 0.09. Recalling that the impact of multi-connectivity on the ongoing session drop probability is much milder, we may formulate a hypothesis that the use of guard bandwidth is vital for improving

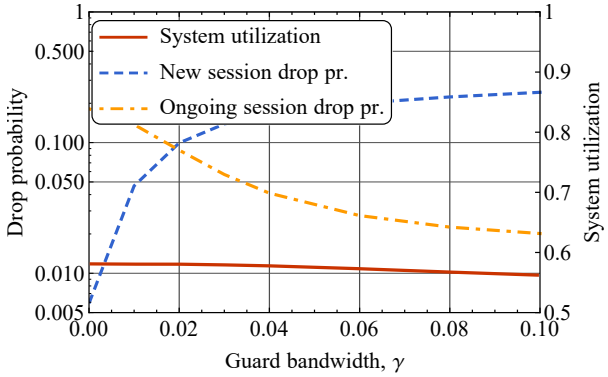
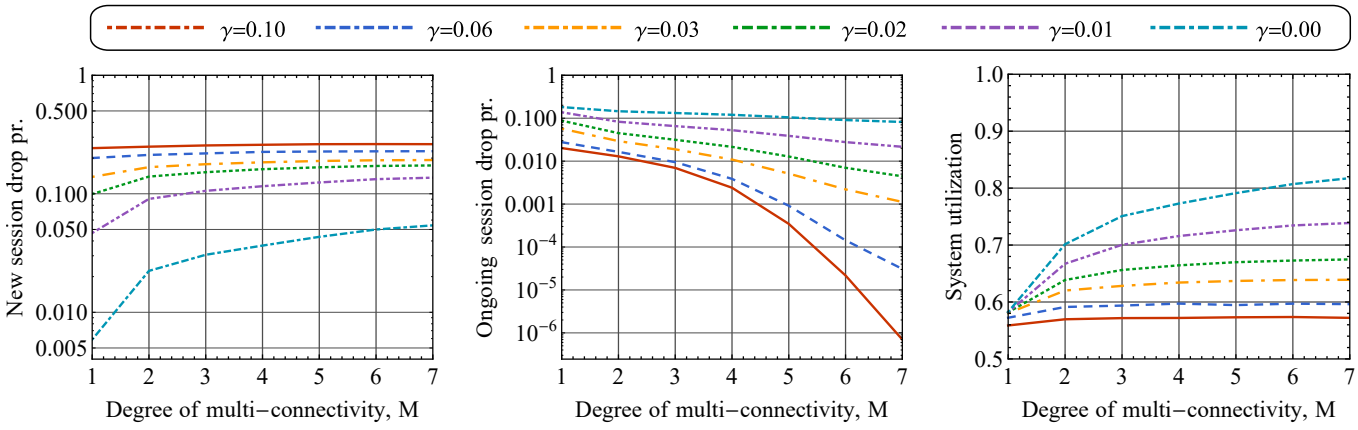
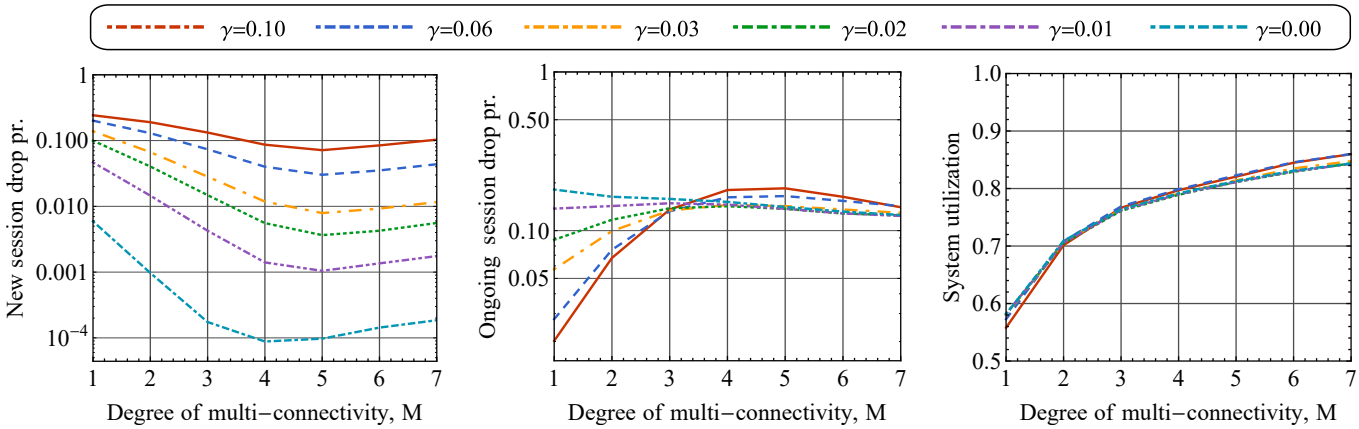


Fig. 4. System and UE performance in presence of guard bandwidth.



(a) New session drop probability (b) Ongoing session drop probability (c) System resource utilization

Fig. 5. User- and system-centric performance with multi-connectivity and guard bandwidth.



(a) New session drop probability (b) Ongoing session drop probability (c) System resource utilization

Fig. 6. User- and system-centric performance with multi-connectivity, guard bandwidth, and initial AP selection.

the session continuity in our system. By implementing both mechanisms, one may operate with two degrees of freedom to select the operating point that offers the required trade-off between ongoing and new session drop probabilities, while still maintaining the system resource utilization at acceptable levels.

B. Joint Use of Multi-Connectivity and Guard Bandwidth

Consider now the joint use of multi-connectivity and guard bandwidth. To this aim, Fig. 5 demonstrates the new and ongoing session drop probabilities as well as the system resource utilization for the parameters collected in Table I.

Analyzing the joint effect of the considered mechanisms on the new session drop probability in Fig. 5a, one may establish that the metric in question increases as we allow for additional spatially diverse links. The negative effect of guard bandwidth on this metric is also evident – even a small increase in the amount of reserved bandwidth results in a significant growth of the new session drop probability. Particularly, for $M = 1$, the parameter of interest jumps from 0.005 for $\gamma = 0$ to 0.05 for $\gamma = 0.01$. The considered performance degradation for new sessions is accompanied by a similar improvement in the service completions for the ongoing session. As observed in Fig. 6b, a higher degree of multi-connectivity leads to a steady decrease in the ongoing session drop probability. Furthermore, while this effect is almost negligible for $\gamma = 0$, it becomes much more pronounced for $\gamma = 0.1$.

Finally, consider the system resource utilization presented in Fig. 6. Here, the positive effect of multi-connectivity also applies to the system with guard bandwidth. However, its impact strongly depends on the fraction of the reserved bandwidth, γ. Mainly, for the relatively large values of γ, the system utilization is rather low and remains within the range $(0.55-0.57)$ for the considered values of M. Reserving fewer resources for the ongoing sessions allows to dramatically improve the resource utilization, which eventually reaches 0.75 for $\gamma = 0.01$. Summarizing, while the joint use of multi-connectivity and guard bandwidth offers more degrees of freedom to achieve the sought balance between new and ongoing session drop probabilities, it may come at the expense of severe system under-utilization, which is one of the key performance indicators for the network operators. To alleviate this limitation, we propose to complement the considered mechanisms with an initial AP selection mechanism.

C. Initial AP Selection Mechanism

Let us now assess the system performance when the UE is allowed to choose an AP at the session initialization phase.

Particularly, upon its arrival, the UE selects an AP based not only on the currently received SNR but also on the availability of radio resources to support its new session. The results reflecting this case are shown in Fig. 6.

The impact of initial AP selection on the new session drop probability is presented in Fig. 5a and Fig. 6a, respectively. One may observe that the behavior with and without the initial AP selection changes fundamentally. Allowing for an additional degree of freedom at the session initialization phase leads to a substantial improvement in the new session drop probability when the degree of multi-connectivity increases up to $M = 5$. This effect is observed for all the considered values of the guard bandwidth fraction γ. Adding more spatially-diverse links alters the behavior as the new session drop probability begins to increase. We specifically note that this behavior is observed for other input parameters as well. Hence, we may conclude that for any choice of γ there is a value of M that minimizes the new session drop probability.

Analyzing the effects of the initial AP selection on the ongoing session drop probability (presented in Fig. 6b), one may notice that the system is highly sensitive to the choice of the guard bandwidth fraction value γ. For small values of γ, the ongoing session drop probability still decreases. However, already starting with $\gamma \approx 0.02$, the opposite effect is observed, i.e., the ongoing session drop probability first grows up to approximately $M = 3$ and then starts to decrease. Taking into account the drastically increased signaling complexity to support a higher number of the spatially-diverse links, for the operating range of interest, i.e., $M \in \{1, 2, 3, 4\}$, both metrics (new and ongoing session drop probabilities) demonstrate a favorable behavior for non-zero values of the guard bandwidth. Notably, given a certain trade-off between these user-centric metrics, one can identify the optimal choice of (M, γ).

Finally, consider the influence of the initial AP selection on the system resource utilization. As one may establish, the negative impact of the guard bandwidth that was observed without the initial AP selection (see Fig. 5c) presently disappears and the figures corresponding to different values of γ are very close to each other. At the same time, the impact of multi-connectivity maintains: a higher degree of multi-connectivity still improves the resource utilization in the system.

V. Conclusions

Session continuity is one of the critical user-centric considerations that can be compromised by dynamic human-body blockage phenomena in 5G NR systems. The use of guard bandwidth and multi-connectivity mechanisms is considered in this work to improve the session continuity in dense NR-based deployments. We have carefully investigated the effects of these two techniques, including the benefits of their joint usage. The main conclusion of our numerical study is that utilizing both multi-connectivity and guard bandwidth may not be sufficient to achieve the desired balance between the user- and the system-centric metrics.

while offering sufficient degrees of freedom for the target

Supplying these mechanisms with an initial AP selection procedure allows maximizing the system resource utilization trade-off between the new and ongoing session drop probabilities. We specifically note that in the proposed joint implementation of the guard bandwidth and multi-connectivity, the UE does not rely on any information about the current AP loading conditions. Hence, the envisioned operation does not require any changes to the system signaling at the air interface aside from that needed to support the already standardized 3GPP multi-connectivity operation. Finally, the guard bandwidth option can be implemented with a software update, thus yielding a simple and cost-effective solution.

References

[1] "M.2376: Technical feasibility of IMT in bands above 6 GHz," ITU-R technical report, 2015.

[2] Y. Niu, Y. Li, D. Jin, L. Su, and A. Vasilakos V, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Networks*, vol. 21, pp. 2657–2676, November 2015.

[3] 3GPP, "Study on new radio (NR) access technology (Release 14)," 3GPP TR 38.912 V14.0.0, March 2017.

[4] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5G mmwave cellular networks," *IEEE Communications Magazine*, vol. 54, pp. 40–47, November 2016.

[5] D. S. Michalopoulos, I. Viering, and L. Du, "User-plane multi-connectivity aspects in 5G," in *2016 23rd International Conference on Telecommunications (ICT)*, pp. 1–5, May 2016.

[6] F. B. Tesema, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Mobility modeling and performance evaluation of multi-connectivity in 5G intra-frequency networks," in *IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, December 2015.

[7] D. Ohmann, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "Achieving high availability in wireless networks by inter-frequency multi-connectivity," in *2016 IEEE International Conference on Communications (ICC)*, pp. 1–7, May 2016.

[8] V. Petrov *et al.*, "Dynamic multi-connectivity performance in ultra-dense urban mmwave deployments," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 2038–2055, September 2017.

[9] V. Petrov, D. Lema, M. Gapeyenko, A. Antonkoglou, D. Moltchanov, F. Sardis, A. Samuylov, S. Andreev, K. Samouylov, and M. Dohler, "Achieving end-to-end reliability of mission-critical traffic in softwarized 5G network," *IEEE Journal on Selected Areas in Communications*, Apr. 2018.

[10] D. Moltchanov, A. Samuylov, V. Petrov, M. Gapeyenko, N. Himayat, S. Andreev, and Y. Koucheryavy, "Improving session continuity with bandwidth reservation in mmwave communications," *IEEE Wireless Communications Letters*, available online, 2018.

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

[12] 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, pp. 1897–1907, March 2005.

[13] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 14)," 3GPP TR 38.901 V14.1.1, July 2017.

[14] A. B. Constantine *et al.*, "Antenna theory: analysis and design," *Microstrip Antennas (third edition), John Wiley & Sons*, 2005.

[15] A. Y. Khinchin, *Mathematical foundations of information theory*. Courier Corporation, 2013.

[16] 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*, available online, 2017.

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

[18] D. J. Daley, "The busy period of the M/GI/∞ queue," *Queueing Systems*, vol. 38, no. 2, pp. 195–204, 2001.

[19] S. Ross, *Introduction to probability models*. Academic Press, 2010.

[20] H. G. Perros, "Computer simulation techniques: The definitive introduction!," 2009.