# Resource Queuing System with Preemptive Priority for Performance Analysis of 5G NR Systems[⋆]

Eduard Sopin[1][0000−0001−9082−2152],
Vyacheslav Begishev[1][0000−0002−7232−4157],
Dmitri Moltchanov[2][0000−0003−4007−7187], and
Andrey Samuylov[2][0000−0003−4007−7187]

[1] Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation
[2] Tampere University, Tampere, 33720, Finland
`{sopin-es,begishev-vo, molchanov-da, samuylov-ak}@rudn.ru`
`{dmitri.moltchanov, andrey.samuylov}@tuni.fi`

**Abstract.** One of the ways to enable smooth coexistence of ultra reliable low latency communication (URRLC) and enhances mobile broadband (eMBB) services at the air interface of perspective 5G New Radio (NR) technology is to utilize preemptive priority service. In this paper, we provide approximate analysis of the queuing system with random resource requirements, two types of customers and preemptive priority service procedure. The distinctive feature of the systems – the random resource requirements – allows to capture the essentials of 5G NR radio interface but inherently increases the complexity of analysis. We present the main performance metrics of interest including session drop probability and system resource utilization as well as assess their accuracy by comparing with computer simulations. The developed model is not inherently limited to URLLC and eMBB coexistence and can be utilized in performance evaluation of 5G NR systems with priority-based service discipline at the air interface, e.g., in context of network slicing. Among other conclusions we explicitly show that both session drop and interruption probabilities of low priority traffic heavily depend not only on the intensity of high priority traffic but on stochastic characteristics of the resource request distribution.

**Keywords:** resource queuing system, preemptive priority, blocking probability, interruption probability, URLLC traffic, network slicing.

---

[⋆] The publication has been prepared with the support of the "RUDN University Program 5-100" (recipient V. Begishev). The reported study was funded by RFBR, project numbers 18-07-00576 (recipients D. Moltchanov and A. Samuylov) and 19-07-00933 (recipient E. Sopin)

## 1 Introduction

Fifth generation (5G) cellular systems promises to bring not only extreme bandwidth at the air interface but to introduce advanced functionalities related to quality of service differentiation to mobile virtual mobile network operators (MVNO, [8, 5]) as well as traffic coexistence at the air interface [2, 6]. As the standardization of 5G New Radio (NR) interface is close to its completion the research community continue to explore the way to support these advanced functionalities.

In recent years, queuing systems with random resource requirements, where customers require not only a server but also a random volume of resources, have drawn significant attention for their ability to capture specifics of session serving process in prospective cellular systems including 5G NR technology [3, 11]. These systems are often used on conjunction with stochastic geometry models to capture the stochastic properties induced by user locations with respect to base stations (BS) as well as traffic service dynamics at BS. Recently, these models have been utilized to investigate various session continuity mechanisms in 5G NR systems including multiconnectivity [14, 15], resource reservation [3] as well as their joint functionality [9]. However, despite many research activities in the field, resource queuing systems with priorities have not been addressed yet.

The model presented in this paper has a large scope of applications in 5G NR systems. Of particular interest is provisioning of ultra reliable low latency service (URLLC) at New Radio (NR) base stations in industrial applications in context of network slicing [13, 6]. Recall that URLLC service requires extremely small delays and loss guarantees at the air interface. To ensure it when mixed with conventional enhanced mobile broadband (eMBB, [4]) service at a single NR BS, several approaches ranging from the use of intentional overlapping by using non-orthogonal multiple access (NOMA, [8]) to explicit static bandwidth reservations have been proposed in the past. In this context, explicit prioritization may provide an alternative approach to maintain extreme service characteristics of URLLC traffic.

Another application area of the queuing system with priorities is network slicing functionality of 5G systems [6]. The concept of network slicing has been originally introduced in 3GPP in Release 14 [1] in context of vehicular-to-everything (V2X) communications. Later, in 3GPP Releases 15 and 16 it has been extended to include MVNOs and bundling of services with similar QoS requirements. The use-cases are flexible and include service provisioning for third parties such as MVNOs, content providers, etc. In the context of 5G system, a network slice instance is a unification of virtual resources in end-to-end fashion wherein a set of virtual network functions are instantiated and connected via a virtual network. One of the ways to support network slices with different service requirements in 5G NR systems is to use priority-based service discipline.

The specified model allows to account for random resource requirements at the air interface of both eMBB and URLLC service induced by random locations of user equipment in the coverage area of NR BS [3, 10]. Preemptive priority discipline simultaneously accounts for efficient use of resources at the NR

BS and ensures that URLLC traffic receives absolute priority over conventional eMBB traffic reaching the prescribed loss guarantees. Supplementing the model with a certain deployment of NR BSs and UEs in the considered area one may characterize the required density of NR BSs needed to maintain the prescribed performance provided to both URLLC and eMBB traffic types.

The rest of the paper is organized as follows. The resource queuing system model is introduced and analyzed in Section 2. Applications and numerical results are discussed in Section 3. Conclusions are drawn in the last section.

## 2 Model description and analysis

In this section we first introduced the queuing model by specifying their parameters and priority mechanism. Then, we proceed analyzing the system and proposing an approximate approach to the system analysis. Metrics of interest are finally derived.

### 2.1 Description of the System

We consider a multiserver queuing system with $N$ servers and resource volume $R$. In practical context $N$ represents the number of simultaneously supported sessions and can be infinite while $R$ characterizes the amount of resources measures in, e.g., resource blocks (RB) or Hz. Two types of customers are served in the queuing system: first type are the preemptive priority customers and the second type are non-priority customers. Customers arrive according to Poisson process with intensities $\lambda_1$ and $\lambda_2$ correspondingly. These flows may represent, e.g. URLLC and eMBB traffic flows or slices with different service requirements. An arriving customer of type $l$ requires discrete random volume of resources according to probability distribution $\{p_{l,r}\}$, $l = 1, 2$, $r = 1, 2, ...$, where $\sum\limits_{r=1}^{\infty} p_{l,r} = 1$. The randomness of resource request distribution allows to capture stochastic nature of user locations in the service area of a BS. The service times are exponentially distributed with intensities $\mu_1$ and $\mu_2$, correspondingly.

Assume that there are $n_1$ customers of the first type totally occupying $r_1$ resources and $n_2$ customers of the second type occupying $r_2$ resources. Upon arrival of a second type customer that requires $j$ resources, if there is free server $(n_1 + n_2 < N)$ and total volume of unoccupied resources is greater than the required volume of the customer $(R - r_1 - r_2 \geq j)$, then the customer is accepted and the required resource volume is allocated to the customer. If arriving customer is the priority customer, then it is still accepted if $n_1 < N$ and $R - r_1 \geq j$, but the service of one or more customers of second type is interrupted. In this case, the system chooses the most "heavy" customer (the one that occupies the biggest part of resources) and terminates it. If there are still not enough resources, then the termination procedure continues until the resources can be allocated to the arriving priority customer $R - r_1 - r_2^* \geq j$.

Since the evaluation of the number of customers to be terminated upon arrival of a priority customer is the most complex part of the analysis, further we assume the following assumption. If termination of the two most "heavy" customers is not enough for the priority customer, then all non-prioruty customers are terminated. The applicability of the assumption is verified in section 3.

The behavior of preemptive priority customers is equivalent to the behavior of these customers in the same queuing system without non-priority customers. Thus, we focus on the metrics of interest associated with non-priority customers.

### 2.2 System Analysis

To decrease the complexity of the stochastic process that describes the behavior of the system, we employ the technique originally proposed in [12]. According to it, instead of keeping track of resources allocated to all the customers, we follow only a total amount of occupied resources for each type of customers. Then, the behavior of the system can be described by a simplified process, $X(t) = (\xi_1(t), \gamma_1(t), \xi_2(t), \gamma_2(t))$, where $\xi_l(t)$ is the number of $l$-type customers at time $t$ and $\gamma_l(t)$ is total resource volume occupied by $l$-type customers.

Let $q(n_1, r_1, n_2, r_2)$ be the stationary distribution of $X(t)$, $Q(n, r)$ and $P(n, r)$ – marginal stationary distributions of first and second type customers, respectively. Since preemptive priority customers are not affected by non-priority customers, then, according to [12] we have

$$Q(n, r) = Q(0, 0)\frac{\rho_1^n}{n!}p_{1,r}^{(n)}, \tag{1}$$

where $\rho_l = \frac{\lambda_l}{\mu_l}$, $p_{1,r}^{(n)}$ is the probability that $n$ first type customers totally occupy $r$ resources and $Q(0, 0)$ is calculated using the normalizing condition. Particularly, the normalizing constant and all characteristics may be evaluated using functions $G(n, r)$, which are introduced in [16] together with efficient recurrent algorithm for their calculation.

$$G(n, r) = \sum_{k=0}^{n}\sum_{j=0}^{r}\frac{\rho^k}{k!}p_{1,j}^k. \tag{2}$$

Note that probabilities $p_{1,r}^{(n)}$ are evaluated from distribution $\{p_{1,r}\}$ by $n$-fold convolution.

To derive equations for approximation of stationary probabilities $P(n, r)$, we introduce additional notation. Let $\pi(k, j)$ be the probability that arriving second type customers cannot be accepted to the system conditional to $k$ second type customers are in the system with $j$ resources occupied, i.e.,

$$\pi(k, j) = 1 - \sum_{i=1}^{R-j} p_{2,i}\frac{G(N-k-1, R-j-i)}{G(N-k, R-j)}. \tag{3}$$

The fraction inside the sum represent conditional probability that there are no more than $N-k-1$ priority customers with no more than $R-j-i$ resource occupied under condition that there is no more than $N-k$ of them occupying no more than $R-j$ resources.

Further, let $\gamma_i(k,j)$ be the probability that one nonpriority customer occupies $i$ resources under condition that $k$ of them occupy $j$ resources, $1 \le i \le j-k+1, 1 \le k \le j$,

$$\gamma_i(k,j) = \frac{p_{2,i} \cdot p_{2,j-i}^{(k-1)}}{p_{2,j}^{(k)}}. \tag{4}$$

Employing the order statistics approach, we derive the probability distribution of the most "heavy" customers. Let $\gamma_{max,i}(k,j)$ be the probability that the most "heavy" customer occupy $i$ resources under condition that there are totally $k$ nonpriority customers occupying $j$ resources:

$$\gamma_{max,i}(k,j) = \left(\mathit{\Gamma}_{k,j}(i+1)\right)^k - \left(\mathit{\Gamma}_{k,j}(i)\right)^k, 1 \le i \le j-k+1, \tag{5}$$

where $\mathit{\Gamma}_{k,j}(x)$ is the cumulative distribution function (CDF), which corresponds with distribution $\{\gamma_i(k,j)\}_{i>0}$.

One of the most challenging parts is to analyze how many nonpriority customers should be interrupted upon arrival of a priority customer. So, we introduce $\beta_i(k,j)$ - the probability that upon arrival of a priority customer, exactly $i$ resources should be released by second type customers under condition that there are $k$ of them in the system occupying totally $j$ resources.

$$\beta_i(k,j) = \sum_{r=0}^{R-j} \frac{G(N-k,r) - G(N-k,r-1)}{G(N-k,R-j)} p_{1,R+i-r-j}, \tag{6}$$
$$1 \le k \le N-1, k \le j \le R, 1 \le i \le j,$$

$$\beta_0(k,j) = \sum_{r=0}^{R-j-1} \frac{G(N-k,r) - G(N-k,r-1)}{G(N-k,R-j)} \sum_{s=1}^{R-r-j} p_{1,s}, \tag{7}$$
$$1 \le k \le N-1, 1 \le k \le R-1.$$

These probabilities will be used to analyze the number of interrupted nonpriority customers. Further we assume that upon arrival of a priority customer, there are four possibilities, namely i) no interruption, ii) interruption of one nonprioirity customer, iii) interruption of two customers and iv) interruption of all customers. Since in the considered system, the most heavy customers are interrupted in the first place, the proposed assumption is expected to give good approximation. So, let $\delta_{s,i}(k,j)$, $s = 1, 2$ be the probability that exactly $s$ customers are interrupted, which occupied totally $i$ resources under condition that there were $k$ customers with $j$ resources occupied. These probabilities are evaluated as follows:

$$\delta_{1,i}(k,j) = \gamma_{max,i}(k,j) \sum_{l=1}^{i} \beta_l(k,j), 1 \le i \le j-k+1, 1 \le k \le N-1, \tag{8}$$

$$\delta_{2,i}(k,j) = \sum_{m=1}^{i-1} \gamma_{max,m}(k,j) \cdot \gamma_{max,i-m}(k-1,j-m) \sum_{l=m+1}^{i} \beta_l(k,j), \quad 2 \le i \le j-k+2, \tag{9}$$

$$\delta_{all}(k,j) = 1 - \delta_0(k,j) - \sum_{i=1}^{j-k+1} \delta_{1,i}(k,j) - \sum_{i=2}^{j-k+2} \delta_{2,i}(k,j), \tag{10}$$

where $\delta_0(k,j)$ is the probability that no interruption occurs upon arrival of a first type customer:

$$\delta_0(k,j) = G(N-k,R-j)^{-1} \sum_{r=1}^{R-j-N+k+1} p_{1,r} \cdot G(N-k-1,R-j-r) \tag{11}$$

Finally, let $\alpha_{s,i}(j)$, $s=1,2$, be the probability that exactly $s$ non-priority customers occupying $i$ resources are interrupted upon arrival of a priority customer under condition that there are $N$ customers of second type with $j$ resources occupied, i.e.,

$$\alpha_{1,i}(j) = \gamma_{max,i}(N,j) \sum_{m=1}^{R-j+i} p_{1,m}, 1 \le i \le j-N+1 \tag{12}$$

$$\alpha_{2,i}(j) = \sum_{l=1}^{i-1} \gamma_{max,l}(N,j) \cdot \gamma_{max,i-l}(N-1,j-l) \sum_{m=R-j+l+1}^{R-j+i} p_{1,m}, \quad 2 \le i \le j-N+2 \tag{13}$$

$$\alpha_{all}(j) = 1 - \sum_{i=1}^{j-N+1} \alpha_{1,i}(j) - \sum_{i=2}^{j-N+2} \alpha_{2,i}(j), 1 \le j \le R. \tag{14}$$

Utilizing the the introduced probabilities, we can derive the system of equilibrium equations as follows:

$$\begin{aligned}\lambda_2(1-\pi(0,0))P(0,0) = \mu_2 \sum_{j=1}^{R} P(1,j) + \lambda_1 \sum_{k=1}^{N-1} \sum_{j=k}^{R} P(k,j)\delta_{all}(k,j) + \\ +\lambda_1 \sum_{j=N}^{R} P(N,j)\alpha_{all}(j) + \lambda_1 \sum_{j=1}^{R} P_1(j)\delta_j + \sum_{j=2}^{R} P_2(j)\delta_j\end{aligned} \tag{15}$$

$$
\begin{aligned}
P(k,j)\,[\lambda_2(1-\pi(k,j)) + k\mu_2 + \lambda_1(1-\pi_{b1})(1-\delta_0(k,j))] &= \\
= (k+1)\mu_2 \sum_{i=1}^{R-j} P(k+1,j+i)\cdot\gamma_i(k+1,j+i) &+ \\
+\lambda_2 \sum_{i=1}^{j-k+1} p_{2,i} P(k-1,j-i)\cdot \frac{G(N-k,R-j)}{G(N-k+1,R-j+i)} &+ \\
+\lambda_1 \sum_{i=1}^{\min(j+i-k,R-j)} P(k+1,j+i)\delta_{1,i}(k+1,j+i) &+ \\
+\lambda_1 \sum_{i=2}^{\min(R-j,j+i-k)} P(k+2,j+i)\delta_{2,i}(k+2,j+i)&, \\
1 \le k \le N-1, k \le j \le R&,
\end{aligned} \tag{16}
$$

$$
P(N,j)\,[N\mu_2 + \lambda_1] = \lambda_2 \sum_{i=1}^{j-N+1} p_{2,i} P(N-1,j-i)\cdot \frac{G(0,0)}{G(1,R-j+i)}. \tag{17}
$$

By solving the system (15) - (17) we obtain the marginal stationary probabilities $P(n,r)$.

### 2.3 Performance Metrics of Interest

Now, we proceed with performance metrics of interest. First, we analyze the blocking probabilities upon arrival of both types of customers, namely $\pi_{b1}$ and $\pi_{b2}$. For priority customers, the blocking probability is obtained in [16], while $\pi_{b2}$ is deduced using similar logic

$$
\pi_{b1} = 1 - G(N,R)^{-1} \sum_{r=1}^{R} p_{1,r} G(N-1,R-r); \tag{18}
$$

$$
\pi_{b2} = \sum_{j=0}^{R} P(N,j) + \sum_{k=0}^{N-1}\sum_{j=0}^{R} P(k,j)\pi(k,j). \tag{19}
$$

Finally, we proceed with the probability $\pi_i$ that a non-priority customer is interrupted. The intensity of customer interruption is obtained as $\lambda_2(1-\pi_{b,2}) - \bar{N}_2\mu_2$ from the equality of arrival intensity and intensity of leaving the system. Here $\bar{N}_2$ is the average number of 2-type customers in the system. Then, the interruption probability is given by the ratio of interruption and arrival intensities, that is,

$$
\pi_i = 1 - \frac{\bar{N}_2\mu}{\lambda_2(1-\pi_{b,2})}. \tag{20}
$$

## 3 Applications and Numerical Results

In this section, we first review the applications scope of developed model. then, we proceed providing numerical example characterizing the performance of low priority arrival flow in the considered system.

### 3.1 Applications Scope

The developed model can be utilized as a service model NR BSs with multiple arrival flows having different priorities in composite performance evaluation frameworks, e.g., [3, 17]. These frameworks are logically divided into two parts: (i) a queuing part and (ii) queuing parametrization. The former captures the features of radio resource allocation at NR BSs. At the input, it accepts the CDF of the amount of requested resources, $F_R(x)$, and the temporal intensity of the UE stage changes, $\alpha$, and delivers delivers the metrics of interest depending on these intermediate metrics as well as parameters of additional quality of service provisioning mechanisms.

The distinctive feature of the developed queuing model is that it simultaneously captures the randomness of resource request distribution and priority service. The former captures the effect of randomness on user locations in BS service area. Thus, it is applicable to 5G NR systems, where one has to prioritized arrivals traffic flows, e.g., systems supporting two or more types of traffic with different priorities or systems supporting network slicing capabilities at the air interface.

To utilize the proposed service model in realistic environment one needs to provide the following metrics as a function of NR BS parameters, propagation characteristics and environment: (i) coverage of NR BS, (ii) mean resource request of sessions (or slices, depending on the application). The former can be determined using the procedure offered in [3, 17] while the latter can be found as discussed in [10]. Note that we also do not explicitly specify the operational frequency of NR technology. In case of millimeter wave band one also needs to account for blockage phenomena using the models proposed in [7].

### 3.2 Accuracy Assessment

We start assessing the accuracy of the developed approximation. To this aim, we have developed a single-purpose simulation tool that models the considered resources queuing system with two types of customers and preemptive priority service discipline. Here, we assume that resource requirements of both priority and non-priority customers have geometric distribution with parameter 0.5, binomial distribution with parameter 0.25 and Poisson distribution with parameter 1. Further, we assume that $N = 60$, $R = 100$, the arrival intensity of priority customers is $\lambda_1 = 35$, and service intensities are $\mu_1 = \mu_2 = 1$. The arrival intensity of non-priority customers $\lambda_2$ varies from 10 to 20.

Figure 1 shows the comparison results. One can note that the blocking probability of the priority customers shows almost perfect match with the simulations,

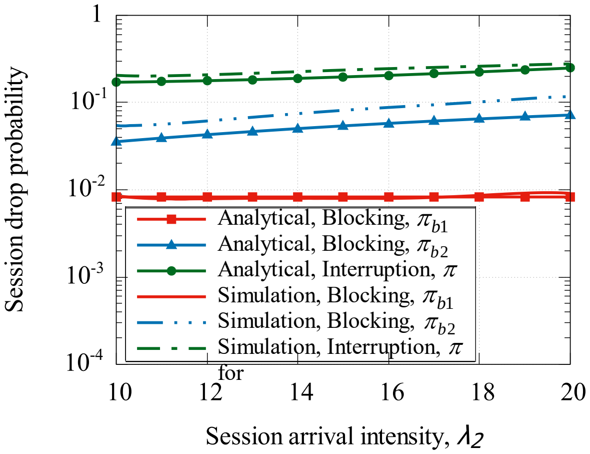

**Fig. 1.** Comparison of analytical and simulation results

while other two probabilistic metrics have bigger relative error. Notably, the relative error of all probabilities decreases with the increase of the non-priority customers intensity.

### 3.3 Performance Evaluation

We now proceed analyzing the response of the performance metrics ofd interest to the system parameters. Recall, that in the considered system we are primarily interested in service metrics related to the low priority traffic as high priority traffic receives exclusive access to the system resources.

First, we complement the dependence illustrated in Fig. 1 with Fig. 2 showing the considered performance metrics as a function of higher priority arrivals, $\lambda_1$. As one may observe, logical the increase in both, $\lambda_1$ and $\lambda_2$, results in worse performance of lower priority traffic in terms of session drop (blocking) probability upon arrival. However, the magnitude of the associated increase is different. On the other hand, the behavior of session interruption probability is drastically different. Particularly, when $\lambda_2$ increases the interruption probability always increases as well. However, the response to $\lambda_1$ is different – there is a bending point when the session interruption probability of low priority traffic starts to increase. This is explained by the fact that under these conditions more low priority sessions are blocked upon arrivals increasing the amount of resources left for those low priority sessions that are currently in the system.

The developed models accepts the CDF of resource requirements as a part of the model specifications. Depending on the application rate requirements as well as the service area of BSs, CDF may substantially differ. We thus investigate the effect of the structure of CDF of resource requirements in Fig. 3 and Fig.

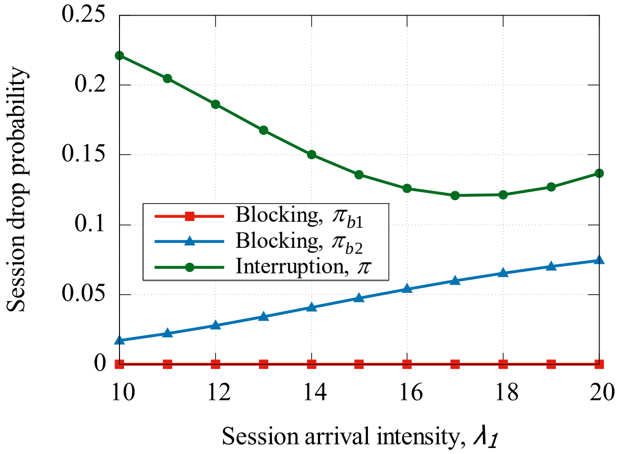

**Fig. 2.** Drop as a function of $\lambda_1$ for binomial, geometric and Poisson distribution

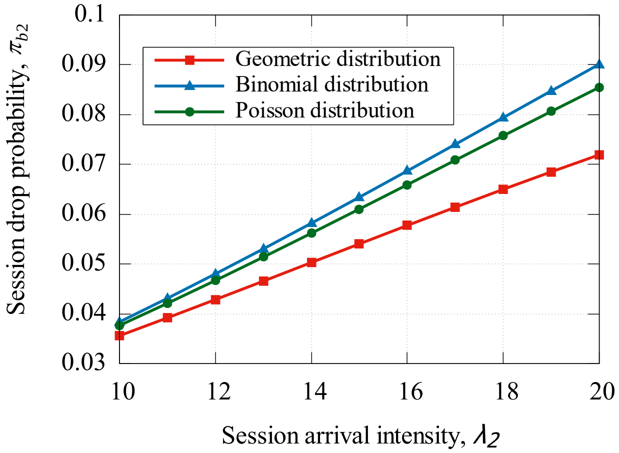

**Fig. 3.** $\pi_{b2}$ as a function of $\lambda_2$ for binomial, geometric and Poisson dis- tribution

4 by showing the metrics of interest as a function of geometric, Poisson and binomial distributions with the same mean value. As one may observe, Binomial and Poisson distributions lead to rather similar results with $3-5\%$ difference in absolute values of considered metrics of interest. One the other hand, Geometric distribution results in drastically different performance measures. Observe that the variance of the considered Binomial, Poisson and Geometric distributions are 0.75, 1 and 2, respectively. This implies that not only mean value of the

resource request distribution but higher order moments affect key performance indicators in 5G NR systems.

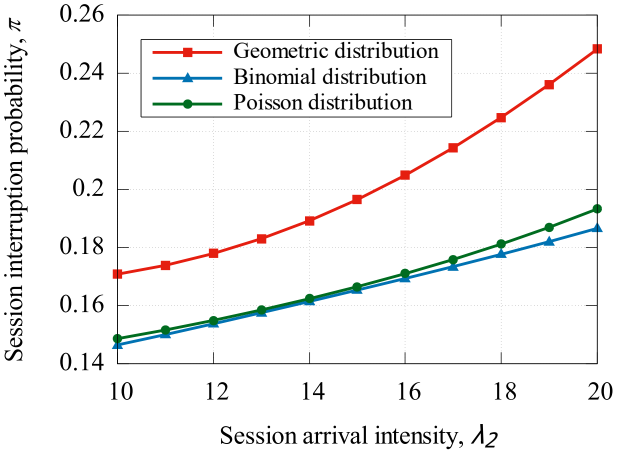

**Fig. 4.** $\pi$ as a function of $\lambda_2$ for binomial, geometric and Poisson distribution

## 4 Conclusion

Motivated by coexistence of high-priority URLLC and low priority eMBB traffic at 5G NR air interface, in the paper, we have analyzed the limited resources queuing system with two types of customers and preemptive priority service discipline. We have developed an approach for approximate analysis performance measures for both type of customers.

The application scope of the proposed system is not limited by URLLC and eMBB coexistence at the air interface. Basically, the model can be applied to 5G NR systems, where prioritization is required at the air interface, e.g., network slicing. The developed model can be utilized in composite performance evaluation frameworks, where the service process at BSs is separated from the model parameterization.

The reported results of direct comparison with computer simulations indicate that the approximation accuracy lies within $5-10\%$. Furthermore, although the complexity of the proposed approach is lower compared to the direct solution of equilibrium equations, it is still high requiring significant computational efforts. Thus, the focus of our future work is to improve the approximation accuracy and decrease the computational requirements of the procedure.

## References

1. 3GPP: Service requirements for V2X services (Release 14). 3GPP TS 22.185 V14.3.0 (March 2017)
2. Afolabi, I., Taleb, T., Samdanis, K., Ksentini, A., Flinck, H.: Network slicing and softwarization: A survey on principles, enabling technologies, and solutions. IEEE Communications Surveys & Tutorials **20**(3), 2429–2453 (2018)
3. Begishev, V., Moltchanov, D., Sopin, E., Samuylov, A., Andreev, S., Koucheryavy, Y., Samouylov, K.: Quantifying the impact of guard capacity on session continuity in 3gpp new radio systems. IEEE Transactions on Vehicular Technology **68**(12), 12345–12359 (2019)
4. Chen, Y.J., Cheng, L.Y., Wang, L.C.: Prioritized resource reservation for reducing random access delay in 5g urllc. In: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). pp. 1–5. IEEE (2017)
5. Esswie, A.A., Pedersen, K.I.: Opportunistic spatial preemptive scheduling for urllc and embb coexistence in multi-user 5g networks. Ieee Access **6**, 38451–38463 (2018)
6. Foukas, X., Patounas, G., Elmokashfi, A., Marina, M.K.: Network slicing in 5g: Survey and challenges. IEEE Communications Magazine **55**(5), 94–100 (2017)
7. Gapeyenko, M., Samuylov, A., Gerasimenko, M., Moltchanov, D., Singh, S., Aryafar, E., Yeh, S.p., Himayat, N., Andreev, S., Koucheryavy, Y.: Analysis of human-body blockage in urban millimeter-wave cellular communications. In: 2016 IEEE International Conference on Communications (ICC). pp. 1–7. IEEE (2016)
8. Kassab, R., Simeone, O., Popovski, P.: Coexistence of urllc and embb services in the c-ran uplink: an information-theoretic study. In: 2018 IEEE Global Communications Conference (GLOBECOM). pp. 1–6. IEEE (2018)
9. Kovalchukov, R., Moltchanov, D., Begishev, V., Samuylov, A., Andreev, S., Koucheryavy, Y., Samouylov, K.: Improved session continuity in 5g nr with joint use of multi-connectivity and guard bandwidth. In: 2018 IEEE Global Communications Conference (GLOBECOM). pp. 1–7. IEEE (2018)
10. Kovalchukov, R., Moltchanov, D., Gaidamaka, Y., Bobrikova, E.: An accurate approximation of resource request distributions in millimeter wave 3gpp new radio systems. In: Internet of Things, Smart Spaces, and Next Generation Networks and Systems, pp. 572–585. Springer (2019)
11. Lu, X., Sopin, E., Petrov, V., Galinina, O., Moltchanov, D., Ageev, K., Andreev, S., Koucheryavy, Y., Samouylov, K., Dohler, M.: Integrated use of licensed-and unlicensed-band mmwave radio technology in 5g and beyond. IEEE Access **7**, 24376–24391 (2019)
12. Naumov, V.A., Samuilov, K., Samuilov, A.K.: On the total amount of resources occupied by serviced customers. Automation and Remote Control **77**(8), 1419–1427 (2016)
13. Orsino, A., Kovalchukov, R., Samuylov, A., Moltchanov, D., Andreev, S., Koucheryavy, Y., Valkama, M.: Caching-aided collaborative d2d operation for predictive data dissemination in industrial iot. IEEE Wireless Communications **25**(3), 50–57 (2018)
14. Petrov, V., Lema, M.A., Gapeyenko, M., Antonakoglou, K., Moltchanov, D., Sardis, F., Samuylov, A., Andreev, S., Koucheryavy, Y., Dohler, M.: Achieving end-to-end reliability of mission-critical traffic in softwarized 5g networks. IEEE Journal on Selected Areas in Communications **36**(3), 485–501 (2018)

15. Petrov, V., Solomitckii, D., Samuylov, A., Lema, M.A., Gapeyenko, M., Moltchanov, D., Andreev, S., Naumov, V., Samouylov, K., Dohler, M., et al.: Dynamic multi-connectivity performance in ultra-dense urban mmwave deployments. IEEE Journal on Selected Areas in Communications **35**(9), 2038–2055 (2017)
16. Samouylov, K., Sopin, E., Vikhrova, O., Shorgin, S.: Convolution algorithm for normalization constant evaluation in queuing system with random requirements. In: AIP Conference Proceedings. vol. 1863, p. 090004. AIP Publishing LLC (2017)
17. Samuylov, A., Moltchanov, D., Kovalchukov, R., Pirmagomedov, R., Gaidamaka, Y., Andreev, S., Koucheryavy, Y., Samouylov, K.: Characterizing resource allocation trade-offs in 5g nr serving multicast and unicast traffic. IEEE Transactions on Wireless Communications **19**(5), 3421–3434 (2020)