

Queuing Model with Unreliable Servers for Limit Power Policy Within Licensed Shared Access Framework

Konstantin Samouylov¹, Irina Gudkova^{1,2(✉)}, Ekaterina Markova¹,
and Natalia Yarkina¹

¹ RUDN University, 6 Miklukho-Maklaya St., Moscow 117198, Russia
{samuylov_ke, gudkova_ia, markova_ev}@pfur.ru,
nat.yarkina@mail.ru

² Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences (IPI FRC CSC RAS),
44-2 Vavilova Str., Moscow 119333, Russia

Abstract. Shared access to spectrum by several parties seems to become one of the most promising approaches to solve the problem of radio spectrum shortage. The framework proposed by ETSI, licensed shared access (LSA), gives the owner absolute priority in spectrum access, to the detriment of the secondary user, LSA licensee. The latter can access the spectrum only if the owner’s QoS is not violated. If the users of both parties need continuous service without interruptions, the rules of shared access should guarantee the possibility of simultaneous access. Balancing the radio resource occupation between parties could take quite a long time compared to the dynamics of the system due to the coordination process by the national regulation authority (NRA). We examine a scheme of the simultaneous access to spectrum by the owner and the LSA licensee that minimizes the coordination activities via NRA. According to this scheme, when the owner needs the spectrum, the power of the LSA licensee’s eNB/UEs is limited. From the LSA licensee’s perspective, the scheme is described in the form of a queuing system with reliable (single-tenant band) and unreliable (multi-tenant band) servers. We show that the infinitesimal generator of the system has a block tridiagonal form. The results are illustrated numerically by estimating the average bit rate of viral videos, which varies due to aeronautical telemetry corresponding to the owner’s traffic.

Keywords: Licensed shared access · Limit power policy · Queuing system · Unreliable servers · Blocking probability · Average bit rate

The reported study was funded within the Agreement № 02.a03.21.0008 dated 24.11.2016 between the Ministry of Education and Science of the Russian Federation and RUDN University, by RFBR according to the research projects No. 15-07-03608 a and 16-37-00421 mol_a.

1 Introduction

The demand for mobile broadband services as well as the volume of traffic increases every year [1, 2]. A considerable amount of frequency resources is needed to provide to users services with a required level of quality of service (QoS) [3, 4]. The problem of resource shortage can be solved by means of the shared access to spectrum by several entities, implemented, for instance, by using the licensed shared access (LSA) framework [5, 6]. LSA framework [7] can improve the efficiency of resource usage and ensure the access to a spectrum which otherwise would be underused. The spectrum is shared between the owners (incumbents) and a limited number of LSA licensees (e.g., mobile network operators). The LSA licensee has access to both bands – the single-tenant band assigned only to it and the multi-tenant band assigned also to the incumbent. The LSA implementation is required to guarantee the QoS for all users, the strictest requirement being not to interrupt users in service due to the incumbent accessing spectrum.

For the shared access, ETSI [8] proposes to use the spectrum allocated for aeronautical and terrestrial telemetry or specific applications including cordless cameras, portable video links, and mobile video links. Various policies of interference coordination between two entities could be considered. The authors of [3] propose three of them: the so-called ignore policy [8, 9], shutdown policy, and limit power policy. The latter implies managing the user equipment (UE) power in uplink and eNodeB (eNB) power in downlink.

In the paper, we consider the case described in [3, 9, 10]. The airport (incumbent) has a frequency band for telemetry with airplanes (air traffic control, ATC). The mobile operator (LSA licensee) also has access to it, thus having its own single-tenant band and the incumbent's multi-tenant band. We assume the users of the mobile operator to watch short videos (e.g., viral) [11] in high quality. At the time when the airplane is communicating with ATC, ATC asks the mobile operator to limit the interference around the airplane. The interference threshold is achieved by reducing the downlink power of the eNB creating interference with the airplane (limit power policy). This results in a bit rate decrease [12] on the multi-tenant band so that the users continue watching video (but in a lower quality). Note that the users will continue to get service at a degraded bit rate after the release of the multi-tenant band by the airport. This is due to the fact that any changes require additional signaling procedures and potential coordination with the national regulation authority (NRA), which could lead to intolerable delays. We also assume that, on the multi-tenant band, new requests are accepted at the maximum bit rate only when all users at the degraded bit rate have finished watching videos.

The paper is organized as follows. In Sect. 2, we propose a mathematical model of the LSA framework with the limit power policy. In Sect. 3, we analyze numerically the performance measures: the blocking probability, the average bit rate, and the utilization factor. Section 4 concludes the paper.

2 Mathematical Model

2.1 General Assumptions and Parameters

We consider a single cell of mobile network with an overlaid LSA framework and one service that generates streaming traffic. We suppose that the single-tenant band has the total capacity of C_1 bandwidth units (b.u.) whereas the multi-tenant band has the total capacity of C_2 b.u. Each request processed on the single-tenant band is served at the guaranteed bit rate (GBR) d_{\max} . The number of resources allocated to the request on the multi-tenant band equals to d_{\max} or d_{\min} depending on the state of the multi-tenant band – operational or unavailable.

Let the arrival rate λ be Poisson distributed and let the service time be exponentially distributed with mean μ^{-1} . Then, we denote the corresponding offered load as $\rho = \lambda/\mu$. We assume that the multi-tenant band goes into unavailable mode with rate α and recovers into operational mode with rate β . Recovery and failure intervals follow the exponential distribution. Let us introduce the following notation:

- n_1 – the number of single-tenant band users;
- n_2^{\max} – the number of multi-tenant band users when the multi-tenant band is operational;
- n_2^{\min} – the number of multi-tenant band users when the multi-tenant band is unavailable;
- s – the state of the multi-tenant band, s equals to 1 if the band is operational and s equals to 0 if the band is unavailable;
- $N_1 = \left\lfloor \frac{C_1}{d_{\max}} \right\rfloor$ – the maximum number of single-tenant band users;
- $N_2 = \left\lfloor \frac{C_2}{d_{\max}} \right\rfloor$ – the maximum number of multi-tenant band users.

2.2 Limit Power Policy

Let us consider in more detail the policy of reducing the corresponding UE's uplink power by the eNB in order to meet the interference constraints. First of all, we determine the rules for accepting requests for service, provided that the UE's uplink power is not yet limited and is at its maximum.

Given the above considerations, when a new request arrives, four scenarios are possible:

- The request will be accepted for service on the single-tenant band, if the request finds the single-tenant band having not less than d_{\max} free b.u.
- The request will be accepted for service on the multi-tenant band, if the request finds the single-tenant band having less than d_{\max} b.u. free, the multi-tenant band is operational, i.e. $s = 1$, and having not less than d_{\max} b.u. free.
- Otherwise, the request will be blocked without any after-effect on the corresponding Poisson process.

If the power is limited due to the incumbent’s need for resources and the single-tenant band is totally occupied, then QoS on the multi-tenant band is degraded. In this case, the multi-tenant band goes into “unavailable” mode and the bit rates of all requests in service on the multi-tenant band switch from the maximum d_{\max} to the minimum d_{\min} value. When the multi-tenant band recovers, the bit rates are not switched back and all users that have been degraded continue to receive service at bit rate d_{\min} . It should be noted that the multi-tenant band has the following property: requests can be served at the maximum bit rate, i.e. the multi-tenant band goes into operational mode, only when the service of all requests at the minimum bit rate is completed.

2.3 System of Equilibrium Equations

According to the above considerations, we can describe the LSA operation by a Markov process $\mathbf{X}(t) = \left\{ (N_1(t), N_2^{\max}(t), N_2^{\min}(t), S(t)), t \geq 0 \right\}$ on the state space

$$\mathbf{X} = \left\{ n_1 = 0, \dots, N_1, n_2^{\max} = 0, \dots, N_2, n_2^{\min} = 0, s = 1 \right. \\ \left. \vee n_1 = 0, \dots, N_1, n_2^{\max} = 0, n_2^{\min} = 0, \dots, N_2, s = 0 \right\}. \tag{1}$$

State space (1) can be subdivided into two subspaces: $\{n_1 = 0, \dots, N_1, n_2^{\max} = 0, \dots, N_2, n_2^{\min} = 0, s = 1\}$ if the multi-tenant band is operational and requests can be served at the maximum bit rate, and $\{n_1 = 0, \dots, N_1, n_2^{\max} = 0, n_2^{\min} = 0, \dots, N_2, s = 0\}$ if the multi-tenant band is unavailable and requests continue their service at the minimum bit rate. Figure 1 shows the structure of the state space, considering the two subspaces.

The corresponding Markov process $\mathbf{X}(t)$, which representing the system’s states, is described by the system of equilibrium equations

$$p(n_1, n_2^{\max}, n_2^{\min}, s) [\lambda \cdot 1(n_1 < N_1) + \lambda \cdot 1(n_1 = N_1, n_2^{\max} < N_2, s = 1) \\ + \beta \cdot 1(s = 0, n_2^{\min} = 0) + n_2 \mu \cdot 1(n_2^{\min} > 0) + n_1 \mu \cdot 1(n_1 > 0) + n_2^{\max} \mu \cdot 1(n_2^{\max} > 0) \\ + \alpha \cdot 1(s = 1) = p(n_1 + 1, n_2^{\max}, n_2^{\min}, s) [(n_1 + 1) \mu \cdot 1(n_1 < N_1)] \\ + p(n_1, n_2^{\max} + 1, 0, 1) [(n_2^{\max} + 1) \mu \cdot 1(n_1 = N_1, n_2^{\max} < N_2, s = 1)] \\ + p(n_1 - 1, n_2^{\max}, n_2^{\min}, s) [\lambda \cdot 1(n_1 > 0)] \\ + p(n_1, n_2^{\max} - 1, 0, 1) [\lambda \cdot 1(n_2^{\max} > 0, n_1 = N_1)], \quad (n_1, n_2^{\max}, n_2^{\min}, s) \in \mathbf{X}, \tag{2}$$

where $\left(p(n_1, n_2^{\max}, n_2^{\min}, s) \right)_{(n_1, n_2^{\max}, n_2^{\min}, s) \in \mathbf{X}} = \mathbf{p}$ is the stationary probability distribution.

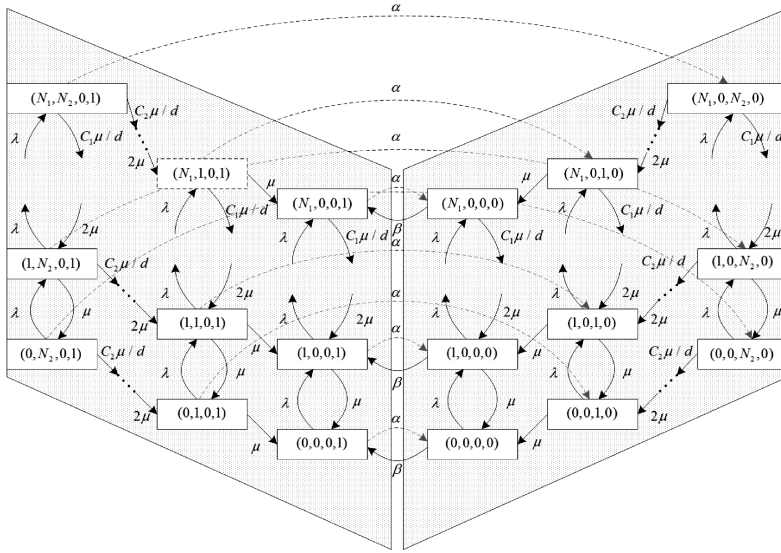


Fig. 1. The state space.

2.4 Infinitesimal Generator

The system probability distribution is numerically computed as the solution of the system of equilibrium equations $\mathbf{p} \cdot \mathbf{A} = \mathbf{0}$, $\mathbf{p} \cdot \mathbf{1}^T = 1$, where \mathbf{A} is the infinitesimal generator of Markov process $\mathbf{X}(t)$. Let us denote $n = 0, \dots, N_1 + N_2$ the number of users. If the lexicographical order on state space \mathbf{X} is defined as

$$(n_1, n_2^{max}, n_2^{min}, s) < (n'_1, n'_2^{max}, n'_2^{min}, s')$$

if and only if $n_1 + n_2^{max} + n_2^{min} < n'_1 + n'_2^{max} + n'_2^{min}$ or $n_1 + n_2 + m = n'_1 + n'_2^{max} + n'_2^{min}$ and $(n_1 > n'_1 \text{ and } n_1 = n'_1 \text{ or } s > s')$, then

(1) Infinitesimal generator \mathbf{A} is a block tridiagonal matrix and has the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{N}_0 & \mathbf{\Lambda}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{M}_1 & \mathbf{N}_1 & \ddots & \dots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \mathbf{\Lambda}_{N_1+N_2-1} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{M}_{N_1+N_2} & \mathbf{N}_{N_1+N_2} \end{bmatrix}.$$

(2) Blocks $\mathbf{\Lambda}_n$, $n = 0, \dots, N_1 + N_2 - 1$ of the upper diagonal have the sizes

$$\dim \mathbf{\Lambda}_n = \begin{cases} (2n + 2) \times (2n + 4), & n = 0, \dots, N_2 - 1, \\ (2N_2 + 2) \times (2N_2 + 2), & n = N_2, \dots, N_1 - 1, \\ (2(N_1 + N_2 - n) + 2) \times (2(N_1 + N_2 - n)), & n = N_1, \dots, N_1 + N_2 - 1, \end{cases}$$

and the following form:

$$\Lambda_n \left((n_1, n_2^{\max}, n_2^{\min}, s), (n'_1, n_2'^{\max}, n_2'^{\min}, s') \right) = \begin{cases} \lambda, & n'_1 = n_1 + 1, & n_2'^{\max} = n_2^{\max}, & n_2'^{\min} = n_2^{\min}, & s' = s & \text{or} \\ & n'_1 = n_1 = N_1, & n_2'^{\max} = n_2^{\max} + 1, & n_2'^{\min} = n_2^{\min} = 0, & s' = s = 1, \\ 0, & & & & \text{otherwise.} \end{cases}$$

(3) Blocks \mathbf{M}_n , $n = 1, \dots, N_1 + N_2$ of the lower diagonal have the sizes

$$\dim \mathbf{M}_n = \begin{cases} (2n + 2) \times 2n, & n = 1, \dots, N_2, \\ (2N_2 + 2) \times (2N_2 + 2), & n = N_2 + 1, \dots, N_1, \\ (2(N_1 + N_2 - n) + 2) \times (2(N_1 + N_2 - n) + 4), & n = N_1 + 1, \dots, N_1 + N_2, \end{cases}$$

and the following form:

$$\mathbf{M}_n \left((n_1, n_2^{\max}, n_2^{\min}, s), (n'_1, n_2'^{\max}, n_2'^{\min}, s') \right) = \begin{cases} n_1 \mu, & n'_1 = n_1 - 1, & n_2'^{\max} = n_2^{\max}, & n_2'^{\min} = n_2^{\min}, & s' = s, \\ n_2^{\max} \mu, & n'_1 = n_1, & n_2'^{\max} = n_2^{\max} - 1, & n_2'^{\min} = n_2^{\min} = 0, & s' = s = 1, \\ n_2^{\min} \mu, & n'_1 = n_1, & n_2'^{\max} = n_2^{\max} = 0, & n_2'^{\min} = n_2^{\min} - 1, & s' = s = 0, \\ 0, & & & & \text{otherwise.} \end{cases}$$

(4) Blocks \mathbf{N}_n , $n = 0, \dots, N_1 + N_2$ of the main diagonal have the sizes:

$$\dim \mathbf{N}_n = \begin{cases} (2n + 2) \times (2n + 2), & n = 0, \dots, N_2 - 1, \\ (2N_2 + 2) \times (2N_2 + 2), & n = N_2, \dots, N_1, \\ (2(N_1 + N_2 - n) + 2) \times (2(N_1 + N_2 - n) + 2), & n = N_1 + 1, \dots, N_1 + N_2. \end{cases}$$

and the following form:

$$\mathbf{N}_n \left((n_1, n_2^{\max}, n_2^{\min}, s), (n'_1, n_2'^{\max}, n_2'^{\min}, s') \right) = \begin{cases} \alpha, & n'_1 = n_1, & n_2'^{\max} = n_2^{\max} = 0, & n_2'^{\min} = n_2^{\min} = 0, & s' = s - 1 & \text{or} \\ & n'_1 = n_1, & n_2'^{\max} = 0, & n_2'^{\min} = n_2^{\max}, & s' = s - 1, \\ \beta, & n'_1 = n_1, & n_2'^{\max} = n_2^{\max} = 0, & n_2'^{\min} = n_2^{\min}, & s' = s + 1, \\ *, & n'_1 = n_1, & n_2'^{\max} = n_2^{\max}, & n_2'^{\min} = n_2^{\min}, & s' = s, \\ 0, & & & & \text{otherwise,} \end{cases}$$

where $*$ = $-(\lambda \cdot 1\{n_1 < N_1\} + n_1\mu \cdot 1\{n_1 > 0\}) + \lambda \cdot 1\{n_1 = N_1, n_2^{\max} < N_2, s = 1\} + n_2^{\max}\mu \cdot 1\{n_2^{\max} > 0\} + n_2^{\min}\mu \cdot 1\{n_2^{\min} > 0\} + s\alpha + (1 - s)\beta$.

3 Numerical Analysis

3.1 Performance Measures

Having found the probability distribution $p(n_1, n_2^{\max}, n_2^{\min}, s)$, $(n_1, n_2^{\max}, n_2^{\min}, s) \in X$, one can compute the performance measures of the considered scheme: the probability B that a request is blocked, the average bit rate \bar{d} , the average bit rate $\bar{d}(C_2)$ on the multi-tenant band, and the utilization factor UTIL of the bands:

$$B = \sum_{i=0}^{N_2} p(N_1, 0, i, 0) + p(N_1, N_2, i, 0), \quad (3)$$

$$\bar{d} = \frac{\sum_{(n_1, n_2^{\max}, n_2^{\min}, s) \in X / (0,0,0,0), (0,0,0,1)} \frac{n_1 d_{\max} + n_2^{\max} d_{\max} + n_2^{\min} n_1 d_{\min}}{n_1 + n_2^{\max} + n_2^{\min}} \cdot p(n_1, n_2^{\max}, n_2^{\min}, s)}{\sum_{(n_1, n_2^{\max}, n_2^{\min}, s) \in X / (0,0,0,0), (0,0,0,1)} p(n_1, n_2^{\max}, n_2^{\min}, s)}, \quad (4)$$

$$\bar{d}(C_2) = \frac{\sum_{(n_1, n_2^{\max}, n_2^{\min}, s) \in X - n_2^{\max} \neq 0 \vee n_2^{\min} \neq 0} d_{\max} \cdot p(n_1, n_2^{\max}, 0, 1) + d_{\min} \cdot p(n_1, 0, n_2^{\min}, 1)}{\sum_{(n_1, n_2^{\max}, n_2^{\min}, s) \in X - n_2^{\max} \neq 0 \vee n_2^{\min} \neq 0} p(n_1, n_2^{\max}, n_2^{\min}, s)}, \quad (5)$$

$$\begin{aligned} \text{UTIL} \cdot C = & \sum_{(n_1, n_2^{\max}, n_2^{\min}, s) \in X: n_2^{\min}=0, s=1} (n_1 + n_2^{\max}) d_{\max} \cdot p(n_1, n_2^{\max}, 0, 1) + \\ & + \sum_{(n_1, n_2^{\max}, n_2^{\min}, s) \in X: n_2^{\max}=0, s=0} (n_1 d_{\max} + n_2^{\min} d_{\min}) \cdot p(n_1, 0, n_2^{\min}, 1). \end{aligned} \quad (6)$$

3.2 Numerical Example

Let us assume that users view short video clips, e.g. viral video, the length of which is about 20–30 s. The video is in high quality at bit rate $d_{\max} = 2$ Mbps. If a part of the frequency band has to be returned, the mobile operator reduces the corresponding eNB uplink power, whereby the bit rate decreases to $d_{\min} = 0.7$ Mbps. This bit rate d_{\min} also allows users to browse video, but in lower quality. Finally, let us assume that the multi-tenant band goes into unavailable mode every hour (3600 s) or every four hours

(14400 s) on average and the recovery takes around one minute. Table 1 summarizes the initial data of the example. Note that 1 b.u. for the example under consideration equals to 1 Mbps.

Table 1. System parameters

Parameter description	Notation	Value
Peak bit rate for single-tenant band	C_1	20 Mbps [14]
Peak bit rate for multi-tenant band	C_2	20 Mbps [14]
Average service time of one user	μ^{-1}	30 s
Average time when multi-tenant band is available	α^{-1}	3540 s, 14340 s
Average time when multi-tenant band is unavailable	β^{-1}	60 s [3]
Maximum bit rate	d_{\max}	2 Mbps [11]
Minimum bit rate	d_{\min}	0.7 Mbps [11]
Offered load	ρ	$0 \div 30$

The figures below show the behavior of each performance measure under examination – blocking probability B (Fig. 2), average bit rates \bar{d} and $\bar{d}(C_2)$ serving requests on both bands or on multi-tenant band respectively (Fig. 3), and utilization factor UTIL (Fig. 4) – for different values of α^{-1} (the average time when the multi-tenant band is available). All three figures show that the less multi-tenant band goes into “unavailable” mode, the better the performance metrics that characterize the impact of LSA on the QoS, namely, the blocking probability is lower, whereas the average bit rate and the utilization factor are higher.

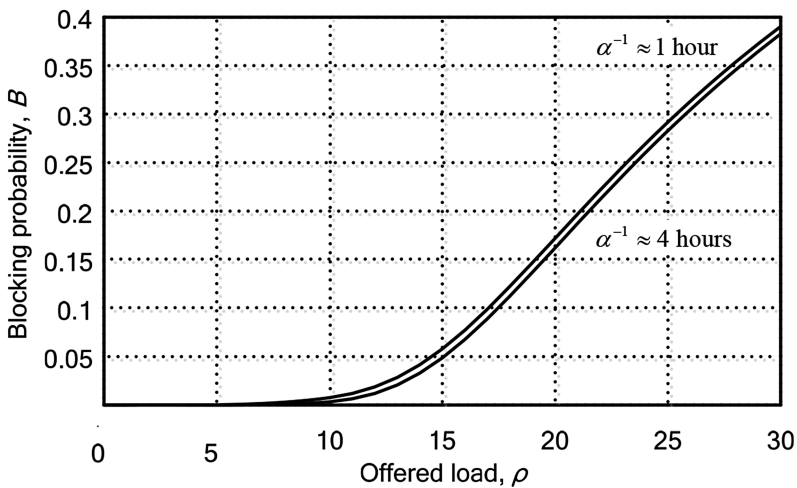


Fig. 2. Blocking probability B for different α^{-1}

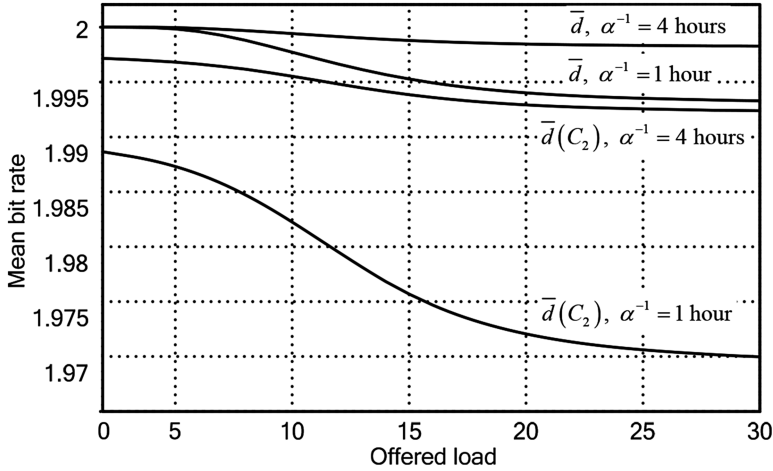


Fig. 3. Average bit rates \bar{d} and $\bar{d}(C_2)$ for different α^{-1}

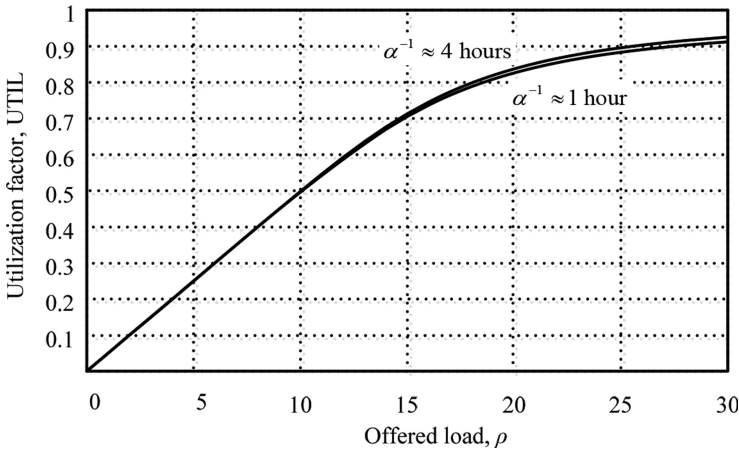


Fig. 4. Utilization factor UTIL for different α^{-1}

4 Conclusion

We have presented a queuing system for analyzing the simultaneous access to spectrum under the limit power policy. The selected policy is based on reducing the eNBs' power and consequently on degrading the service quality from high to standard definition. We have obtained the infinitesimal generator as a block tridiagonal matrix. This form is required for the numerical solution of the system of equilibrium equations and the calculation of the performance metrics that characterize the impact of LSA on the QoS – the blocking probability, the average bit rate, and the utilization factor.

Acknowledgment. The authors are grateful to the students of the Applied Probability and Informatics Department of RUDN University Anastasia Feduro and Dmitry Polouektov for computing the numerical example.

References

1. Cisco Visual Networking Index: Forecast and Methodology, 2015–2020 (2016)
2. Andrews, J., Buzzi, S., Choi, W., Hanly, S.V., Lozano, A., Soong, A.C.K., Zhang, J.C.: What will 5G be? *IEEE J. Sel. Areas Commun.* **32**, 1065–1082 (2014)
3. Ponomarenko-Timofeev, A., Pyattaev, A., Andreev, S., Koucheryavy, Y., Mueck, M., Karls, I.: Highly dynamic spectrum management within licensed shared access regulatory framework. *IEEE Commun. Mag.* **54**(3), 100–109 (2015)
4. Shorgin, S.Y., Samouylov, K.E., Gudkova, I.A., Galinina, O.S., Andreev, S.D.: On the benefits of 5G wireless technology for future mobile cloud computing. In: 1st International Science and Technology Conference Modern Networking Technologies (MoNeTec): SDN & NFV, pp. 151–154 (2014)
5. Buckwitz, K., Engelberg, J., Rausch, G.: Licensed shared access (LSA) – regulatory background and view of administrations. In: CROWNCOM (invited paper), pp. 413–416 (2014)
6. Ahokangas, P., Matinmikko, M., Yrjola, S., Mustonen, M., Luttinen, E., Kivimäki, A., Kemppainen, J.: Business models for mobile network operators in licensed shared access (LSA). In: DYSPAN, pp. 407–412 (2014)
7. Gomez-Migueluez, I., Avdic, E., Marchetti, N., Macaluso, I., Doyle, L.E.: Cloud-RAN platform for LSA in 5G networks – tradeoff within the infrastructure. In: Communications, Control and Signal Processing, pp. 522–525 (2014)
8. ETSI TS 103 113 Mobile broadband services in the 2300 MHz 2400 MHz band under Licensed Shared Access regime (2013)
9. Borodakiy, V.Y., Samouylov, K.E., Gudkova, I.A., Ostrikova, D.Y., Ponomarenko A.A., Turlikov, A.M., Andreev, S.D.: Modeling unreliable LSA operation in 3GPP LTE cellular networks. In: 6th International Congress on Ultra Modern Telecommunications and Control Systems ICUMT-2014, pp. 490–496 (2014)
10. Gudkova, I.A., Samouylov, K.E., Ostrikova, D.Y., Mokrov, E.V., Ponomarenko-Timofeev, A.A., Andreev, S.D., Koucheryavy, Y.A.: Service failure and interruption probability analysis for licensed shared access regulatory framework. In: 7th International Congress on Ultra Modern Telecommunications and Control Systems ICUMT-2015, pp. 123–131 (2015)
11. Live encoder settings, bitrates and resolutions. YouTube Help (2016). <https://support.google.com/youtube/answer/2853702?hl=en>
12. Borodakiy, V., Gudkova, I., Markova, E., Samouylov, K.: Modelling and performance analysis of pre-emption based radio admission control scheme for video conferencing over LTE. In: ITU Kaleidoscope Academic Conference, pp. 53–59 (2014)
13. Neuts, M.F.: *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The John Hopkins University Press, Baltimore (1981). 332 p.
14. 3GPP TS 36.300 Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2: Release 13 (2015)