Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (online), 2658-4670 (print)

2021, 29 (3) 242–250
http://journals.rudn.ru/miph

Research article

UDC 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-3-242-250

Towards the analysis of the performance measures of heterogeneous networks by means of two-phase queuing systems

Tatiana V. Rykova

Peoples' Friendship University of Russia (RUDN University)
6, Miklukho-Maklaya St., Moscow, 117198, Russian Federation

(received: August 16, 2021; accepted: September 9, 2021)

Due to a multistage nature of transmission processes in heterogeneous 4G, 5G mobile networks, multiphase queuing systems become one of the most suitable ways for the resource allocation algorithms analysis and network investigation. In this paper, a few scientific papers that approached heterogeneous networks modelling by means of multiphase queuing systems are reviewed, mentioning the difficulties that arise with this type of analytical analysis. Moreover, several previously investigated models are introduced briefly as an example of two-phase systems of finite capacity and a special structure in discrete time that can be used for analysing resource allocation schemes based on the main performance measures obtained for wireless heterogeneous networks. One of the model presents a two-phase tandem queue with a group arrival flow of requests and a second phase of the complex structure that consists of parallel finite queues. The second model is a two-phase tandem queue with Markov modulated geometric arrival and service processes at the first phase and exhaustive service process at the second phase, which solves a cross-layer adaption problem in a heterogeneous network.

Key words and phrases: two-phase model, queuing system, Markov chain, resource allocation, heterogeneous networks

1. Introduction

The Fifth Generation (5G) mobile networks are characterized by advanced algorithms for time-frequency resource allocation schemes in a heterogeneous cell between Base Station (BS) and User Equipment (UE) [1], [2]. Due to a multistage nature of transmission processes in the heterogeneous environment, multiphase queuing systems become one of the most suitable ways for the resource allocation algorithms analysis and network investigation. In [3], researchers have proposed to use single-phase queuing systems for modelling local networks, by giving the necessary physical meaning to the stages of service process using a phase-type service distribution. However, in the case of the Next Generation Mobile Networks (NGMN) the given assumptions are not able to take into account the complex structure of a network with intermediate storage of transmitted information. A large number of publications [4]–[6] are devoted to the analysis of multiphase queuing systems that consider various variants of structural parameters: capacitance of the buffers at the phases, the number of servers at phases, an ordinary or non-ordinary arrival flow, blocking of service at a phase or loss of a request given that the buffer of the next phase is fully occupied, the possibility of the retransmission at the phase or in the system in general, and various arrival and service distributions of requests. In the given publications, the number of phases is usually limited to two, and they are considered mainly in continuous time.

Only a few works approached to investigate heterogeneous networks by means of multiphase (two-phase) queuing systems in discrete time, see, for example, [7], [8]. However, the models in [7], [8] cannot be used because they do not take into account the complex phase structure when modelling transmission processes in a cell and, therefore, do not fully correspond to solving a resource allocation problem in a context of a NGMN cell. It should be noted that most of the foreign publications when using "discrete" and "tandem queue" terms in their papers cover, in fact, mean cyclical service systems in discrete time, but not multiphase systems.

In most of the cases, the number of phases in a multiphase queuing system that is taken as a mathematical model for analysis of the performance measures in a NGMN cell should be taken equal to two. This is due to the fact that each phase itself is a structurally complex queuing system with complex rules of functioning, and a further increase of phases severely complicates formalization of the entire system, leads to multidimensional processes that describe its behaviour and a difficult practical use. The analysis in this case becomes extremely bulky with high risks of obtaining inaccurate results. The decomposition of such a system with the analysis of individual phases or groups of phases is most often not applicable due to the significant mutual influence of phases, in contrast to almost completely decomposable systems [9], [10], and can lead in most cases to significant modelling errors. Cases of independence of the functioning of a phase from the previous phase and, accordingly, an admissible decomposition are rare and arise when conditions [11], are met, for example, when using exponential distributions and buffers of unlimited capacity [12], or under assumptions about specific conditions for the functioning of phases [13]. Summarizing all of that mentioned above, in this paper we briefly overview several two-phase systems of finite capacity in discrete time of a special structure that can be used for analysing resource allocation schemes based on the main performance measures obtained for wireless heterogeneous networks.

2. Two-phase model in discrete time for resource allocation analysis in heterogeneous networks

Heterogeneous networks with the utilization of lower power levels Relay Nodes (RN) improve the capacity of the system, coverage due to the availability of the alternative paths to users, located in shadow areas, and lower deployments costs. Moreover, relay nodes are characterized by wireless backbone access. However, to achieve its potential, the heterogeneous networks are to utilize an efficient cooperative resource allocation procedure on various paths, e.g. from the base station (gNB, gNodeB) to the RN, and from the RN to the User Equipment (UE), in order to avoid data shortage or overflow of the data at relay nodes. An analytical model of heterogeneous network in terms of a two-phase model in discrete time is further introduced, that presents an efficient tool to study resource allocation procedures by means of the found stationary probability distribution and derived performance measures.

2.1. Model's description

Let us consider downlink transmission in a heterogeneous network with K RNs and a single gNB with a subframe that is divided into S channels, which are distributed between the heterogeneous nodes in a centralized manner. The figure 1 demonstrates the structure of the given model, and the main parameters are shown in the table 1. As can be seen from the table, the arrival rate follows a $(K+1)$-dimensional group geometrical distribution. In its turn, the service time in both phases is selected to follow deterministic law equal to one time slot, which corresponds to the transmission of one packet. After the request is being serviced it occupies one space in the buffer.

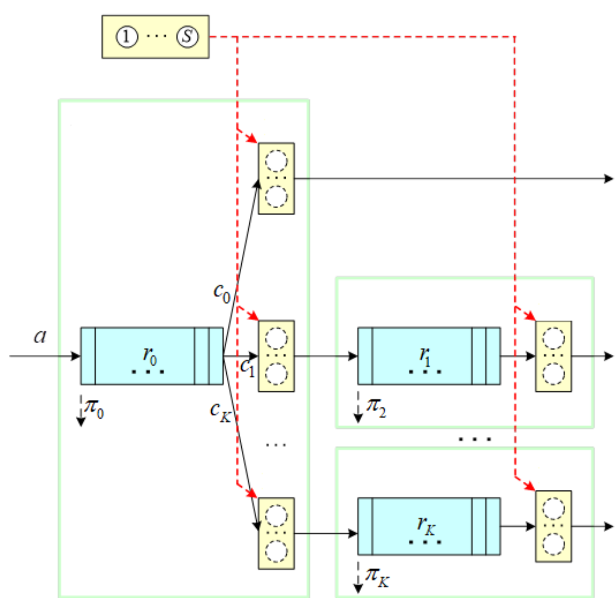

Figure 1. Structural representation of the two-phase model with S channels

The functioning of the given model is described by the homogeneous Markov chain ξ_n at time moments $nh+0, n \geqslant 0$, with the following state space:

$$X = \left\{ \vec{x} = (x_0, x_1, \dots, x_K)^T : \; x_k = 0, \dots, r_k, \; k = 0, \dots, K \right\},$$

$$|X| = \prod_{k=0}^{K} (r_k + 1),$$

Table 1

Definition of the main parameters used in the model.

Parameter	Description
0-request	0-type request to be send to the UE in the coverage area of gNB
k-request	k-type request to be send to the UE in the coverage area of k-RN, $k = 1, 2, \dots, K$
r_0	buffer capacity of the gNB, $r_0 < \infty$
r_k	buffer capacity of the k-RN, $r_k < \infty$, $k = 1, 2, \dots, K$
$h = 1$	constant length of a time slot, in which the system functioning is measured and is equal to LTE downlink data subframe
a	arrival rate that follows a group geometric distribution Geom^G
c_k	probability of a request from the arrival group belonging to type k, k-request, $k = 0, 1 \dots, K$
$c_.$	full sum of the variable c_k, $k = 0, 1 \dots, K$

In the table 1, x_k is a number of k-requests stored in the buffer of corresponding heterogeneous node: gNB or k-RN. Please refer to [14] for more details on derivation of stationary probability distribution and the main performance measures. One of the advantages of the given analytical model is the ability to study various resource allocation procedures by utilizing the following vector in the balance equations:

$$s^n = (s_0^n, s_1^n, \dots, s_K^n)^T = (f_0^n(\vec{x}), f_1^n(\vec{x}), \dots, f_K^n(\vec{x})),$$

where $f^n(\vec{x})$ is a function that introduces the resource allocation strategy. The definition of the different resource allocation procedures and experimental analysis can be found in [14]. All in all, the given model allows analysing various resource assignment schemes, including dynamic schemes, e.g. proportional fair [15] and with fixed allocation.

3. Two-phase model in discrete time for cross-layer optimization in heterogeneous networks

Video transmission comes along with huge demands on resources and low delay, which can be provided by means of Cross-Layer Adaptation (CLA) principle. The given principle is responsible for optimizing the selected metric by adapting the parameters of different layers of open systems interconnection model. The common assumption is to locate CLA mechanism at the gNB, which brings certain shortcomings in terms of the achieved performance. In this paper, we introduce a two-phase analytical model in discrete time that evaluates the behaviour of a downlink video transmission system using CLA principle. We assume Dynamic Adaptive Streaming over HTTP (DASH) in our modelling, the details of which can be found in [16]. The model introduced below covers both the video delivery from gNB to UE at the first phase and video processing at UE at the second phase. Moreover, the CLA is achieved by varying the arrival rate based on the received DASH message, and service probability based on the Channel Quality Indicator (CQI) sent from the UE.

3.1. Model's description

We assume a DASH-based video transmission from the gNB to a single UE in a heterogeneous network. The figure 2 demonstrates the structure of the given model, and the main parameters are shown in the table 2. The functioning of the given model is described by the homogeneous Markov chain ξ_n at time moments $nh+0, n \geqslant 0$, with the following state space:

$$X = \{\vec{x} = (q_1, q_2, s) : q_1 = 0, 1, \ldots, r_1,\ q_2 = 0, 1, \ldots, r_2 - 1,\ s = 1, 2, \ldots, S\},$$

where q_1 and q_2 are the numbers of requests at the first and second phase, respectively, and s is a value of the CQI in the current state.

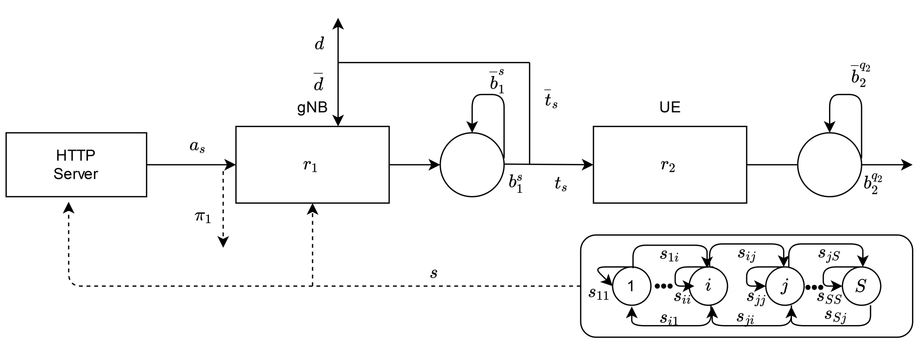

Figure 2. Structural representation of the two-phase model

It should be noted that the stationary probability distribution can be found in a matrix recursive form [13]. However, due to the fact that the functioning of the first phase along with variation of the CQI are independent from the second phase, conditions [11] fulfilled, it is possible to decompose the system to analyse the systems separately, which allows reducing the computational complexity. The conducted experimental analysis of the main performance measures derived from the stationary distribution can be found in [13]. The given two-phase model presents an efficient tool that covers video transmission process from gNB to UE at the first phase and the video decoding process at the UE at the second phase. It takes into account CLA principle, along with the losses due to fading and retransmission.

Table 2

Definition of the main parameters used in the model

Parameter	Description
r_1	buffer capacity of the gNB, $r_1 < \infty$
r_2	buffer capacity of the UE, $r_2 < \infty$
h	constant length of a time slot, in which the system functioning is measured and is equal to LTE downlink data subframe, 1 ms
s	CQI report, which is available both at gNB and UE every time slot $s = 1, 2 \ldots, S$, where S is an overall number of its values
s_{ij}	transition probability of s from state i to state j
a_s	arrival rate that follows geometric distribution
b_1^s	service time at the first phase that follows geometric distribution
$\bar{t}_s \bar{d}$	retransmission probability due to wireless channel errors
$\bar{t}_s d$	probability that the packet lifetime is expired and cannot be used for the video playback
$b_2^{q_2}$	service time at the second phase that follows geometric distribution in exhaustive manner

4. Conclusions

This paper reviews a few scientific papers that approached heterogeneous networks modelling by means of multiphase queuing systems and stressed the difficulties that arise with this type of analytical modelling. Two efficient two-phase models were briefly introduced that can be used for analysing resource allocation schemes based on the main performance measures obtained for wireless heterogeneous networks. One of the model presents a two-phase tandem queue with a group arrival flow of requests and a second phase of the complex structure that consists of parallel finite queues. The second model is a two-phase tandem queue with Markov modulated geometric arrival and service processes at the first phase and exhaustive service process at the second phase, which solves a cross-layer adaption problem in a heterogeneous network.

References

[1] "ITU-R M.2134. Requirements Related to Technical Performance for IMT-Advanced Radio Interface(s)," 2018.

[2] E. Medvedeva, A. Gorbunova, Y. Gaidamaka, and K. Samuylov, "A Discrete Queueing Model for Performance Analysis of Scheduling Schemes in Multi-User MIMO Systems," IEEE, 2019, pp. 1–5. DOI: 10.1109/ICUMT48472.2019.8970876.

[3] G. P. Basharin and V. A. Efimushkin, *Graph-matrix models of local area networks [Grafomatrichnyye modeli lokal'noy seti].* Moscow: UDN, 1986, in Russian.

[4] T. V. Efimushkina, "Performance evaluation of a tandem queue with common for phases servers," in *18-th International Conference on Distributed Computer and Communication Networks (DCCN-2015): Control, Computation, Communications*, ICS RAS, Moscow, Russia: Technosphere, 2015, pp. 44–51.

[5] V. Klimenok and A. Dudin, "Dual tandem queueing system with multiserver stations and retrials," in *International Conference on Distributed computer and communication networks: control, computation, communications (DCCN-2013)*, ICS RAS, Moscow, Russia, 7–10 October 2013: Technosphere, 2013, pp. 394–401.

[6] P. P. Bocharov, A. V. Pechinkin, and S. Sanchez, "Stationary state probabilities of a two-phase queueing system with a markov arrival process and internal losses," in *Proc. of the Fourth Int. Workshop on Queueing Networks with Finite Capacity*, Ilkley, 2000, pp. 06/1–10.

[7] G. P. Basharin and V. A. Efimushkin, "Algorithmic Analysis of Structurally Complex Systems of Finite Capacity with a Two-Dimensional State Space," in *Teletraffic Theory Models in Communication Systems and Computer Technology.* Nauka, 1985, pp. 28–41.

[8] E. V. Viskova, *Two-phase queuing system with Markov flow and discrete-time service [Dvukhfaznaya sistema massovogo obsluzhivaniya s markovskimi potokom i obsluzhivaniyem v diskretnom vremeni]*, 3. Information processes, 2005, vol. 5, pp. 247–257, in Russian.

[9] G. P. Basharin, P. P. Bocharov, and Y. A. Kogan, *Analysis of queues in computer networks. Theory and methods of calculation [Analiz ocheredey v vychislitel'nykh setyakh. Teoriya i metody rascheta].* Moscow: Nauka, 1989, p. 336, in Russian.

[10] P. J. Courtois, *Queueing and computer system application.* New York: Academic Press, 1977.

[11] V. A. Naumov, "On the independent operation of subsystems of a complex system," Russian, in *Proceedings of the 3rd All-Union School-Meeting on the Theory of Queuing*, in Russian, Moscow, Russia: MSU, 1976, pp. 169–177.

[12] J. R. Jackson, *Networks of Waiting Lines*, 4. Operations Research, 1957, vol. 5, pp. 518–521.

[13] T. V. Efimushkina, M. Gabbouj, and K. E. Samuylov, "Analytical model in discrete time for cross-layer video communication over LTE," *Automatic Control and Computer Sciences*, vol. 48, no. 6, pp. 345–357, 2014. DOI: 10.3103/S0146411614060029.

[14] T. V. Efimushkina and K. E. Samuylov, "Analysis of the Resource Distribution Schemes in LTE-Advanced Relay-Enhanced Networks," *Communications in Computer and Information Science*, vol. 279, pp. 43–57, 2014. DOI: 10.1007/978-3-319-05209-0_4.

[15] D. Avidor, S. Mukherjee, J. Ling, and C. Papadias, "On some properties of the proportional fair scheduling policy," in *2004 IEEE 15th International Symposium on Personal, Indoor and Mobile Radio Communications (IEEE Cat. No.04TH8754)*, vol. 2, 2004, pp. 853–858. DOI: 10.1109/PIMRC.2004.1373820.

[16] "ISO/IEC 23009-1. Dynamic adaptive streaming over HTTP (DASH)-Part 1: Media presentation description and segment formats. Draft International Standard," 2011.

For citation:

T. V. Rykova, Towards the analysis of the performance measures of heterogeneous networks by means of two-phase queuing systems, Discrete and Continuous Models and Applied Computational Science 29 (3) (2021) 242–250. DOI: 10.22363/2658-4670-2021-29-3-242-250.

Information about the authors:

Rykova, Tatiana V. — Master of Science in Applied Mathematics and Informatics (PFUR), Master of Science in Information Technology (Tampere University of Technology), researcher at Fraunhofer Heinrich Hertz Institute (Berlin, Germany) (e-mail: tatiana.rykova@hhi.fraunhofer.de, ORCID: https://orcid.org/0000-0002-8561-7514)

УДК 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-3-242-250

К анализу показателей эффективности гетерогенных сетей с помощью двухфазных систем массового обслуживания

Т. В. Рыкова

Российский университет дружбы народов
ул. Миклухо-Маклая, д. 6, Москва, 117198, Россия

Благодаря многоступенчатому характеру процессов передачи в гетерогенных мобильных сетях 4G, 5G, многофазные системы массового обслуживания становятся одним из наиболее подходящих способов анализа алгоритмов распределения ресурсов и исследования сетей. В этой статье приводится обзор нескольких научных работ, посвящённых моделированию гетерогенных сетей с помощью многофазных систем массового обслуживания, и упоминаются трудности, возникающие при этом типе аналитического анализа. Более того, несколько ранее исследованных моделей кратко представлены в качестве примера двухфазных систем конечной ёмкости и специальной структуры в дискретном времени, которые можно использовать для анализа схем распределения ресурсов на базе основных показателей производительности, полученных для беспроводных гетерогенных сетей. Одна из моделей представлена двухфазной тандемной очередью с групповым потоком поступающих запросов, а вторая — фазой сложной структуры, состоящей из параллельных конечных очередей. Вторая модель представляет собой двухфазную тандемную очередь с марковскими модулированными геометрическими процессами поступления и обслуживания на первом этапе и полным процессом обслуживания на втором этапе, что решает проблему межуровневой адаптации в гетерогенной сети.

Ключевые слова: двухфазная модель, система массового обслуживания, цепь Маркова, распределение ресурсов, гетерогенная сеть