BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 68, No. 2, 2020
DOI: 10.24425/bpasts.2020.133111

An analytical model of a system with priorities servicing a mixture of different elastic traffic streams

B. NOWAK[2], M. PIECHOWIAK[1], M. STASIAK[2], and P. ZWIERZYKOWSKI[2]*

[1] Kazimierz Wielki University, ul. Kopernika 1, 85-074 Bydgoszcz, Poland
[2] Poznań University of Technology, ul. Polanka 3, Poland

Abstract. This article proposes an analytical model of a system with priorities servicing a mixture of different elastic traffic streams. The model presented in the article was developed as the extension of earlier works published by the authors. It utilizes the concept of equivalent bandwidth and then, following bandwidth discretization, uses the dependencies introduced on the basis of the assumptions adopted for the generalized Kaufman-Roberts formula and for the model of a full-availability group with traffic compression. The article presents a possibility of using the proposed model to model the radio interface in a multi-service mobile network and provides an example of the above with the interface of an LTE network. Since the proposed model is an approximate one, the results of the calculations are compared with the results of simulations. A comparison of the results confirms an acceptable level of accuracy of the model. The model can be successfully used in the analysis and design of links and nodes of telecommunication and computer networks.

Key words: teletraffic engineering, analytical model, priorities.

1. Introduction

Since early 1990s, the development of mobile telephony digital networks has intensified noticeably [1]. The pace at which changes are being introduced is prompted by general technological advancement on the one hand, and by the evolution of new types of services and the subsequent interconnection between different operators and individual service providers (different telecommunication and computer networks) on the other. It is sufficient to say that the present-day mobile networks, or more broadly wireless networks, are in the vanguard of the changes that are to take place in the services offered to users of telecommunication networks [2].

Along with the development of mobile networks and their more and more advanced technology, a need has appeared for a development of appropriate methods that would make effective estimation of the network traffic capacity possible. These methods allow network designers to develop new mobile networks and optimize those already existing. The methods can also be used to predict an economically viable and technically feasible extension of a network. The issues of the evaluation of traffic capacity become even more important and topical with multiservice networks, in which a provision of specific and welldefined quality of service parameters involves a large number of call classes. The quality of service in the present-day mobile networks is frequently defined on the basis of the so-called system parameters [3]. Typically, the values of the KP indicators (Key Performance Indicator) [3] are used for their determination. The way KP indicators are defined, as well as their terminology and values, rests upon the producers and manufacturers of devices used in a given mobile network [3] and are set up on the initial stage of the running process of new device in the mobile network. In addition, system parameters can also be used to determine the GoS parameters (*grade of service*) and QoS parameters [2].

To evaluate traffic effectiveness of modern mobile network analytical models of multiservice systems, so-called *multirate* models are used. Although they were initially developed for narrowband and broadband integrated services digital networks [4–14], multi-rate models are currently also used to model wireless multiservice telecommunications networks, i.e. UMTS (Universal Mobile Telecommunications System) and LTE (Long Term Evolution) [2, 15–23].

In the case of multi-service models the assumption is that the resources required for a call of a given traffic class to be serviced are the multiple number of the so-called allocation unit (AU), that can be defined as the highest common divisor of the resources that are required (demanded) by calls that belong to all traffic classes offered to the system. The method for determining the basic allocation unit depends on the relevant type of a system with multiservice traffic to a description of which it is used. In the case of multiservice mobile networks, the allocation unit can refer to the real bitrate (in a 2G network) [24, 25] or the equivalent bitrate (in a 3G network) [6, 26], and additionally can also refer to the so-called noise-limited resources (radio interface of 3G/4G networks), where it is expressed in the percentage of the capacity of the radio interface [27, 28].

---

*e-mail: piotr.zwierzykowski@put.poznan.pl

Manuscript submitted 2019-11-28, revised 2020-01-22, initially accepted for publication 2020-02-11, published in April 2020

On account of a relatively high capacity of the backbone part in the present-day mobile networks, it is primarily radio interfaces and other interfaces of the access part that are decisive in traffic effectiveness. The radio interface is characterized by a relatively low capacity (as compared to the remaining interfaces in a mobile network) that is additionally limited by interference coming from neighboring cells [29]. The traffic capacity of a radio interface is also hindered by the remaining interfaces of the access part of a mobile network, which results from their limited capacity or the way they are organized or structured [28, 30–32]. The influence of the radio interface on traffic effectiveness in a network is reported by operators and manufacturers of network devices alike. A minimization of this phenomena is then a guiding principle in the development of new technological solutions that aim to significantly increase the capacity of the radio interface, e.g. HSPA (high speed packet access) [33]. In addition, an introduction of appropriate traffic management mechanisms will allow a significant increase in effectiveness of a network to be introduced. This article proposes a model of a new mechanism for prioritization of traffic that, when applied, can positively influence an increase in performance effectiveness of the radio interface in LTE (Long Term Evolution) networks [30, 32]. The proposed model is an extension of the models proposed in earlier works of the authors [34–38] and makes it possible to apply priorities also in the case of elastic and adaptive traffic. To the best knowledge of the authors, these issues have not been addressed so far.

The article is structured as follows. Section 1 contains basic information on the issues discussed in the article. Section 2 describes the discretization of resources in cellular networks. Section 3 presents the analytical model of the radio interface proposed in the article. Section 4 shows an example illustrating the possibilities of application of the analytical model proposed in the article. Section 5 summarizes the article.

2. Discretization of resources in mobile networks

In modern mobile networks, the packet is the basic carrier of information. The structure and size of packets and the way they are transmitted depend on the technology involved in a network. One of the first multiservice 3G mobile networks was UMTS [39]. In this network, as well as in the networks of the fourth generation (LTE and Advanced LTE), the IP protocol is used in the backbone part [33, 40].

The widespread application of packet transmission imposes the necessity of hierarchical approach to the traffic representation in networks. Any application that can be used by an owner of a mobile phone generates a packet stream. Such a string of packets is considered to be a call. Therefore two levels of traffic representation can be distinguished: the packet level and the call level (Fig. 1).

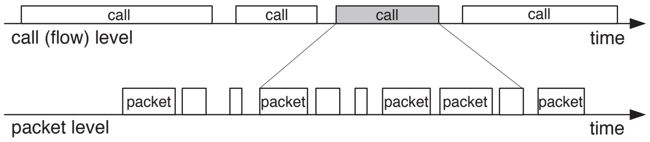

Fig. 1. The two basic levels of traffic analysis

A traffic stream observed at the packet level is typically characterized by a complex internal structure. The literature of the subject proposes a large number of models with different degree of complexity that can be used to analyze these streams. It is worthwhile to observe at this point that the relatively simple ON/OFF models [28] and more complicated models of self-similar traffic [41–43] are particularly fit for the purpose. The latter group of models represents the streams that are typically used in packet networks [44–47]. It is however characterized by high complexity, which significantly influences the description and analysis of those systems to which streams of self-similar traffic are offered. The description of such models is complex to such a degree that has to be limited to approximate or simulation solutions [48]. Traffic at the call level can be then treated in a Poisson-like way [49, 50], which means that call streams and service streams have exponential nature. It is these properties that allow us to apply multiservice Markov processes to model and dimension network system at the call level.

The basic obstacle in the application of Erlang models to analyze a network is variable bitrate of packet streams that are generated by individual specific applications used by the user. A solution to this problem can be an application of the so-called equivalent bandwidth [51]. The equivalent bandwidth is defined as the fixed bitrate that describes the amount (size) of the resources allocated to a given call in the network. The equivalent bandwidth is determined for each service class as the function of the total bitrate of the system, the maximum and the average bitrate of the packet stream, bitrate variance and the acceptable packet delay. The definition of the equivalent bandwidth can also include other parameters that are characteristic for a given type of a stream and the applied service system [6]. Currently in most works authors assume the maximum bitrate for a given class of calls as the equivalent bandwidth [50].

A determination of fixed bitrates demanded by calls of individual classes makes it possible to determine the allocation unit for the interface under investigation [6]. The maximum value of the so-called allocation unit is defined as the highest common divisor of all bitrates allocated to calls of individual classes [28, 30, 32]. The process of a determination of appropriate parameters for calls of individual classes and the capacity of particular interfaces in allocation units is called bandwidth discretization [25]. In this way, the resources required by calls of individual services and the capacities of the interfaces in integer numbers can be determined, which significantly makes any analysis and modelling of the interfaces of a multiservice mobile network easier.

3. Analytical model of the radio interface

To model present-day mobile networks, analytical models of multiservice systems are used. Multiservice models were initially developed for the purpose of an appropriate description of narrow and broadband networks with multirate services [4–14]. Currently, they are also used to model wireless telecommunications and computer networks, i.e. UMTS or LTE networks [15–22].

The basic analytical model of this interface can be a full-availability group with elastic traffic. The full-availability group (FAG) is the model of the system in which only free resources of the system limit the admission of a new call. The basic FAG model assumes that the incoming calls are described by the Poisson stream characteristics and that the occupancy distribution is given by Erlang formula for the single-service system [52], or e.g. for the Kaufmann-Roberts recursion for multi-service system [53, 54]. When the FAG services elastic traffic, the assumption is that the absence of free resources required for a new call of a given class is followed by a compression of all serviced calls in the system, i.e. a decrease in their bitrates to the value that makes services of this new call possible (also in the compressed form). The accompanying assumption is that service time of all calls is prolonged to enable all data to be transmitted.

The general assumption in the model is that calls are compressed until the number of occupied AUs in the group, determined on the basis of the sum of uncompressed demands of calls of all classes, exceeds a certain virtual capacity of the server C_v, where $C_v > C_r$ (Fig. 2) and C_r is the real capacity of the system. If this capacity is exceeded, then a new call will be lost. The occupancy states of the virtual capacity of the group (i.e. such states n that $C_r \le n \le C_v$) determine the compression area for elastic traffic. A choice as to the value C_v is an indicator of the "compression depth", that is equal to the ratio between the virtual capacity of the group and the real capacity C_v/C_r, and determines how many times the total bitrate of serviced calls in the system can be decreased at the maximum. The occupancy distribution in a full-availability group with compression for k traffic classes can be expressed by the following dependence [55, 56]:

$$\begin{cases} P_n = \dfrac{1}{\min(n, C_r)} \displaystyle\sum_{i=1}^{k} A_{i,k} c_{i,k} P_{n-c_{i,k}} & \text{for } 0 \le n \le C_v \\ P_n = 0 & \text{for } C_v < n < 0 \\ \displaystyle\sum_{n=0}^{C_v} P_n = 1 \end{cases} . \tag{1}$$

In Eq. (1), the parameter $A_{i,k}$ is the traffic intensity and $c_{i,k}$ is the number of AUs required by a call of class i in the group servicing k traffic classes.

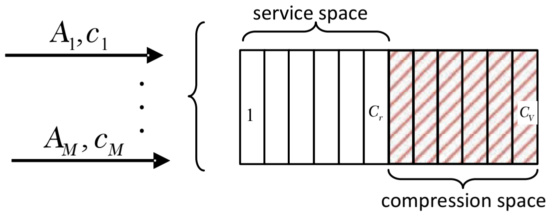

Fig. 2. Schematic diagram of the group with compression

The total carried traffic of class i in the group servicing k traffic classes $Y_{i,k}$ can be written as follows:

$$Y_{i,k} = \sum_{i=1}^{k} A_{i,k} c_{i,k} \left(1 - E_{i,k}\right), \tag{2}$$

where the blocking probability $E_{i,k}$ of class i calls can be determined as [57]:

$$E_{i,k} = \sum_{n=C_r-c_{i,k}+1}^{C_r} P_n . \tag{3}$$

The modelled interface employs the prioritization mechanism. The operation of the prioritization mechanism is based on a determination of the sequence of resource allocation. The introduction of priorities to the system can be followed by a suppression (limitation) of access to resources for calls with lower priorities, while in the case of an absence of free resources, calls with lower priorities can be pushed out. A decision as to prioritization of a call is taken by the operator that defines the weight of particular services in the backbone network. This means that in the access part of the network the operator identifies the priority given (or allocated) to given services and then applies an appropriate traffic control mechanism or packet scheduling mechanism. The operation of the compression mechanism leads to a decrease in the bitrate allocated to serviced connections with lower priority than that of a call that is currently admitted for service (further on called a new call). In the case where the maximum compression of calls with lower priorities doest not provide the amount of resources sufficient to service the demands of a new call, a call with a higher priority forces calls with lower priorities to be terminated until service of a new call with higher priority is possible. If all lower priority calls are held or "stacked" to be handled when time and resources allow and the system still does not have sufficient amount of free resources to service a new call with higher priority, then a compression of the calls with the priority equal to the new call follows. If, even in the maximum compression of calls with the priority equal to the priority of the new call, the system still has no sufficient amount of free resources to service this new call, then this call is rejected [3, 36–38, 41].

The assumption in the interface model under consideration is that the operator has allocated two priorities to the classes of elastic traffic. Further assumption is that the class with the highest priority is marked with index "1", whereas the other class has the lowest priority. Yet another assumption is that the following principles for the mode of operation of the priority mechanism apply in the model – service of calls with the lowest priority does not influence in any way the blocking probability for calls with higher priority.

To model the system with priorities, the model of a full-availability group with compression [55] will be used.

Consider now a system that services two classes of calls, of which one has higher priority. Let us designate the blocking probabilities and carried traffic in the group with priorities with an additional upper index "P".

Table 1
Total carried traffic in the system without and with priorities

Traffic intensity (a)	$a = 0.5$ [Erl/AU]		$a = 1.0$ [Erl/AU]	
System	without priorities	with priorities	without priorities	with priorities
Class 1 – number of class	$13242 +/- 143$	$13257 +/- 136$	$22963 +/- 178$	$26218 +/- 188$
Class 2 – number of class	$11867 +/- 218$	$11762 +/- 126$	$19681 +/- 104$	$16181 +/- 127$
Carried traffic (Y)	$262962 +/- 3825$	$261960 +/- 2745$	$446101 +/- 2918$	$440165,5 +/- 3273$
Absolute error [%]	1.2–1.4%		0.4–0.8%	

In line with the adopted assumptions, the handling of calls with lower priorities in the group with priorities does not influence in any way the handling of calls with higher priorities. This means that the blocking probability and serviced traffic for the first class will be the same as in the system that services one class of calls only:

$$E^P_{1,2} = E_{1,1}\,, \tag{4}$$

and

$$Y^P_{1,2} = Y_{1,1}\,. \tag{5}$$

The probability $E_{1,1}$ and total carried traffic $Y_{1,1}$, with the assumption that the group services the first class of calls only, can be determined on the basis of Equations (1)–(3) for $k = 1$, respectively.

The operation of the system with priorities is based on the assumption that in the case of the absence of free resources, calls with higher priorities push out calls with lower priorities – that is, impose their compression or (as in the case of a lack of the possibility of further compression) prompt a termination of service of calls with lower priorities, and then start to occupy the resources which has just been released by these calls. After considering various requirements of plausible traffic behavior, notably conservation traffic law [58], it can be assumed that the total traffic serviced in the system with priorities is exactly the same as traffic carried in the system without priorities [3][1]. It is possible then to write:

$$Y^P_{1,2} + Y^P_{2,2} = Y_{1,2} + Y_{2,2}\,. \tag{6}$$

Formula (6), taking into account Equation (2) (for $k = 2$), can be rewritten as follows:

$$Y^P_{1,2} + Y^P_{2,2} = \sum_{i=1}^{2} A_{i,2}c_{i,2}\left(1 - E_{i,2}\right). \tag{7}$$

In Formula (7), traffic of class one $Y^P_{1,2}$ serviced in the system with priorities is described by Equation (5). Therefore, traffic of the second class in the system with priorities can be determined, on the basis of (5) and (7), for $k = 2$:

$$Y^P_{2,2} = \sum_{i=1}^{2} A_{i,2}c_{i,2}\left(1 - E_{i,2}\right) - Y_{1,1}\,. \tag{8}$$

Observe that traffic $Y^P_{2,2}$ is determined on the basis of the difference between the total traffic serviced in the system without priorities and the total traffic serviced in the single-service system with the highest priority traffic only. By taking into consideration the general dependence between offered traffic and serviced traffic, we get:

$$Y^P_{2,2} = A_{2,2}c_{2,2}\left(1 - E^P_{2,2}\right). \tag{9}$$

After substituting Formula (9) to Formula (8), it is possible to determine the blocking probability for calls of the second class in the system with priorities:

$$E^P_{2,2} = \frac{A_{2,2}c_{2,2} - \sum\limits_{i=1}^{2} A_{i,2}c_{i,2}\left(1 - E_{i,2}\right) + Y_{1,1}}{A_{2,2}c_{2,2}}\,. \tag{10}$$

Traffic characteristics for offered traffic in both systems are identical ($A_{1,1} = A_{1,2}$, $c_{1,1} = c_{1,2}$ oraz $A_{2,2} = A^P_{2,2}$, $c_{2,2} = c^P_{2,2}$). Therefore, after taking into account the dependence (2), Formula (10) can be ultimately transformed to the following form:

$$E^P_{2,2} = \frac{A_{1,2}c_{1,2}(E_{1,2} - E_{1,1}) + A_{2,2}c_{2,2}E_{2,2}}{A_{2,2}c_{2,2}}\,. \tag{11}$$

On the basis of (11) it is possible to show that the blocking probability of calls of the second class (with lower priorities) can be successfully estimated on the basis of the blocking probabilities in the first and the second systems.

On the basis of Formulas (4) and (11) it is possible then to determine the blocking probability for both classes of calls offered to the radio interface in the LTE network.

4. Numerical example

The access part in the LTE network is known as the EUTRAN (Evolved Universal Terrestrial Radio Access Network). EUTRAN is composed of eNodeBs only that perform tasks that

---

[1] An exemplary simulation results confirming the correctness of the assumptions made are presented in Table 1

are similar to the NodeBs and RNC (radio network controller) elements that can be found in the UTRAN network (UMTS Terrestrial Radio Access Network). The basic reason behind the combination of the functions of both devices is to achieve a reduction in a delay of all operations performed in the radio interface of the network. eNodeBs are connected to one another by X2 interfaces and connected with the packet backbone network through the S1 interface. The present model under consideration refers to the radio interface of an LTE network. This interface makes a component part of the radio access network of the LTE network and is shown in Fig. 3.

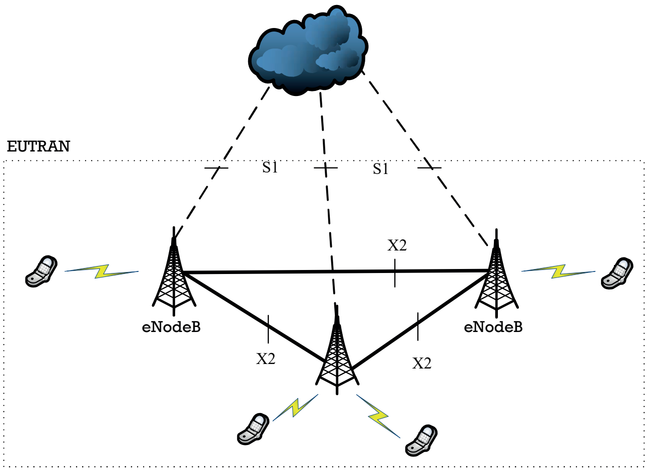

Fig. 3. LTE access part diagram

The capacity of the radio interface depends on the channel width and on the applied transmission techniques used in the radio channel (Table 2). Let us assume that the operator in question offers a channel with the width of 5 Mhz and the terminals utilize the MIMO 2×2 technique, and in consequence the radio link offers the throughput $C_r = 36.7$ AUs (1 AU = 1 kbps)Mbps. Let us then assume that the radio interface services two classes of downlink calls: the first class (higher priority) requires the bitrate from 5 to 10 Aus ($c_{1,2}$) and the second class (lower priority) requires the bitrate from 5.5 to 11 AUs ($c_{2,2}$). Thus, the compression in the system is equal to 50% for both traffic classes and $C_v = 2C_r$.

The blocking probabilities obtained as a result of calculations and simulations for both traffic classes depending on the traffic volume a are shown in Fig. 4 and Fig. 5, where a is calculated as follows:

$$a = \frac{\sum_{i=1}^{2} A_{i,2} c_{i,1}}{C_r}. \tag{12}$$

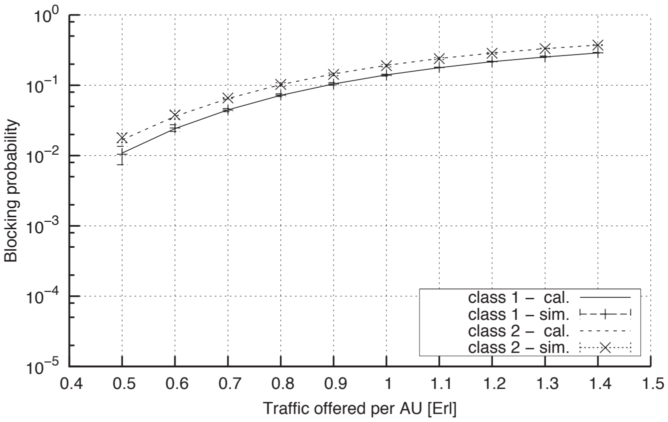

Fig. 4. Blocking probability in the LTE radio interface carrying two classes of calls without priorities in relation to traffic offered per AU

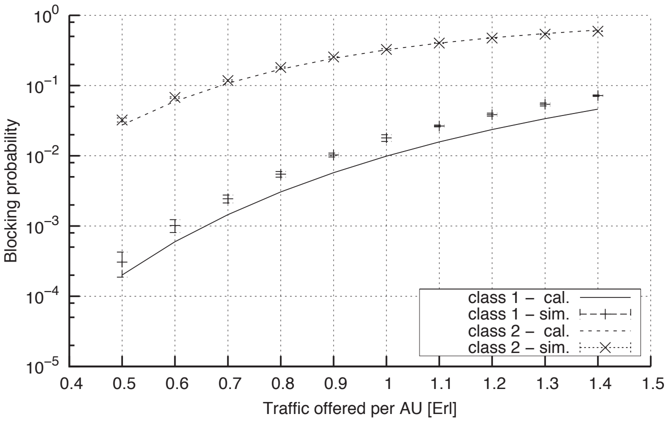

Fig. 5. Blocking probability in the LTE radio interface servicing two classes of calls with priorities in relation to traffic offered per AU

Table 2
Capacity of the radio link in the function between the channel with and transmission techniques

Transmission Mode/ System Bandwidth	SISO Downlink Peak (Mbps)	Transmit Downlink Peak (Mbps)	MIMO 2×2 Downlink Peak (Mbps)	All Transmission Modes Uplink Peak (Mbps) 16 QAM	All Transmission Modes Uplink Peak (Mbps) 64 QAM
1.4 MHz	4.4	4.4	8.8	3	4.4
3 MHz	11.1	11.1	22.1	7.5	11.1
5 MHz	18.3	18.3	36.7	12.6	18.3
10 MHz	36.7	36.7	75	25.5	36.7
15 MHz	55.1	55.1	110	37.9	55.1
20 MHz	75	75	150	51	51

In both figures the results of analytical calculations are presented as a solid line, and the results are expressed by symbols.

While comparing the results presented in Fig. 4 and Fig. 5 one can observe that the introduction of priorities has been followed by a change in the blocking values for calls of individual traffic classes. It should also be highlighted that both classes require a similar amount of resources for service provisioning. Therefore, the blocking probabilities obtained for both traffic classes in the system without priorities presented in Fig. 4 are similar, whereas in the system with priorities the results for the traffic class with higher priority are significantly lower than the results obtained in the system without priorities (Fig. 5). The introduction of priorities was also followed by an increase in the blocking probability for calls with lower priorities. The results obtained for both classes in the system with priorities differ significantly from one another. The blocking probabilities for calls of the first traffic class (higher priority class) are much lower than the blocking probabilities for calls of the second traffic class (with lower priority). Given that the calls of both classes require similar amount of resources to service a call, we can assume that the observed difference results directly from the prioritisation phenomenon. A comparison of the obtained analytical results with the results of simulation experiments shows that the obtained calculation results can be considered to be a fairly satisfactory approximation. The simulation experiments were conducted in a simulation environment prepared by the authors.

The simulation results presented in this study are the average values obtained from five series of simulation experiments (at least 1 million calls of each class were simulated in each series). During simulations the confidence intervals based on the t-Student's distribution were calculated. The values of the confidence intervals are also shown in Fig. 4 and Fig. 5. In some cases, Fig. 4 and Fig. 5 do not show confidence intervals because they were lower than the symbol representing the result of a simulation.

Let us consider the traffic carried by the system with and without priories. The results presented in Table 2 show the average number of serviced calls of the first and the second traffic class (rows 1 and 2) serviced in the system and the average traffic occupied by calls of both traffic classes in the system (row 3). Each of these results is presented for the system with priorities and the system without priorities for two selected values of traffic intensity per AUs ($a = 0{:}5$ and $a = 1{:}0$) in the system. The presented results were calculated as the average of the results obtained in five simulation series. Each result of the simulations is also complemented with the value of the confidence interval. While comparing the obtained values of the average traffic serviced in the system with priorities and the system without priorities one can observe only very slight differences between the obtained traffic values – ranging from 0.4 to 1.4%. It is correct to claim then that the adoption of the assumption of traffic behavior for traffic serviced in the system with priorities and the system without priorities has been confirmed. The results confirm the assumption that in the system with priorities the conservation traffic law is met.

5. Conclusions

This article proposes a new analytical model that can be used for modeling multiservice systems servicing a mixture of elastic traffic classes with priorities. In the article, possible application of the model to evaluate traffic efficiency of the radio interface in the LTE network is presented. The calculation and simulation results confirm the possibility of using the presented analytical model to approximate the capacity of cellular network interfaces.

References

[1] K. Wesołowski, *Mobile Communication Systems*, John Wiley & Sons, 2002.

[2] P. Zwierzykowski, *Traffic menagement mechanisms in multiservice cellular networks*, Poznan University of Technology, Poznań, 2014.

[3] D. Parmenter, *Key Performance Indicators: Developing, Implementing, and Using Winning KPIs*, John Wiley & Sons, 2007.

[4] M. Pióro, J. Lubacz, and U. Korner, "Traffic engineering problems in multiservice circuit switched networks", in *Proc. Traffic Theories for Telecommun. Services*, 1990, 127–136.

[5] H. Akimuru and K. Kawashima, *Teletraffic: Theory and Application*, Berlin-Heidelberg-New York, 1999.

[6] J. Roberts, V. Mocci, and I. Virtamo, *Broadband Network Teletraffic, Commission of the European Communities*, Springer, Berlin, 1996.

[7] K. Ross, *Multiservice Loss Models for Broadband Telecommunication Network*, Springer, London, 1995.

[8] W. Bziuk, "Approximate state probabilities in large shared multirate loss systems with an application to trunk reservation", *Eur. Trans. Telecommun.* 16(3), 145–152 (2002).

[9] I. Moscholios, M. Logothetis and G. Kokkinakis, "Connection-dependent threshold model: a generalization of the Erlang multiple rate loss model", *Perform. Eval.* 48(1-4), 177–200 (2002).

[10] M. Janaszka and J. Lubacz, "Multirate traffic in dynamic synchronous transfer mode networks", in *Proc. 2nd Polish-German Teletraffic Symp.* 2002, 133–146.

[11] V. Iversen, *Teletraffic Engineering Handbook*, ITU-D, Study Group 2, Question 16/2, Geneva, 2005.

[12] H.L. Hartmann and M. Knoke, "The one-level functional equation of multi-rate loss systems", *Eur. Trans. Telecommun.* 14(2), 107–118 (2003).

[13] V.G. Vassilakis, I.D. Moscholios, and M.D. Logothetis, "Call-level performance modelling of elastic and adaptive service-classes with finite population", *IEICE Trans. Commun.* E91-B(1), 151–163 (2008).

[14] Z. Dziong and J.W. Roberts, "Congestion probabilities in a circuit-switched integrated services network", *Perform. Eval.* 7(4), 267–284 (1987).

[15] D. Staehle and A. Mäder, "An analytic approximation of the uplink capacity in a UMTS network with heterogeneous traffic", in *Proc. 18th Int. Teletraffic Cong.* 2003, 81–91.

[16] V. Iversen, V. Bentis, N. Ha, and S. Stepanov, "Evaluation of multi-service CDMA networks with soft blocking", in *Proc. 16th. ITC Specialist Sem. on Perform. Eval. of Wireless and Mobile Syst.* Antwerp, 2004.

[17] A. Viterbi and A. Viterbi, "Erlang capacity of a power controlled CDMA system", *IEEE J. Sel. Areas Commun.* 11(6), 892–900 (1993).

[18] M. Głąbowski, M. Stasiak, A. Wiśniewski and P. Zwierzykowski, "Blocking probability calculation for cellular systems with WCDMA radio interface servicing PCT1 and PCT2 multirate traffic", *IEICE Trans. Commun.* 92-B(4), 1156–1165 (2009).

[19] G.A. Kallos, V.G. Vassilakis, I.D. Moscholios, and M.D. Logothetis, "Performance modelling of W-CDMA networks supporting elastic and adaptive traffic", in *Proc. 4th Int. Working Conf. on Perform. Modell. and Eval. of Heterogen. Net.*, Ilkley, 2006.

[20] J. Juarez, R. Paulo, E. San Jose, and F. Velez, "Modelling and simulation of multi-rate multi-service traffic in the presence of mobility", in *Proc. Veh. Technol. Conf.*, 2006, 1–6.

[21] V.G. Vassilakis, G.A. Kallos, I.D. Moscholios, and M.D. Logothetis, "The wireless engset multi-rate loss model for the call-level analysis of W-CDMS networks", in *Proc. 18th IEEE Ann. Int. Symp. on Pers. Indoor and Mobile Radio Commun.*, Athens, 2007, 1–5.

[22] M. Sobieraj, M. Stasiak, J. Weissenberg, and P. Zwierzykowski, "Analytical model of the single threshold mechanism with hysteresis for multi-service networks", *IEICE Trans. Commun.* 95-B(1), 120–132 (2012).

[23] I. Moscholios and M. Logothetis, *Efficient Multirate Teletraffic Loss Models Beyond Erlang*, Wiley and IEEE Press, 2019.

[24] J. Roberts, "Teletraffic models for the Telcom 1 integrated services network", in *Int. Teletraffic Cong.* 1983, Session 1.1, paper 2, 1–7.

[25] J. Roberts, *Performance Evaluation and Design of Multiservice Networks*, Commission of the European Communities, Brussels, 1992.

[26] F. Kelly, "Effective bandwith at multi-class queues", *Queueing Syst.* 9(1), 5–15 (1991).

[27] D. Staehle, *Analytic Methods for UMTS Radio Network Planning*, PhD thesis, Bayerische Julius-Maximilians-Universitat Wurzburg, 2004.

[28] M. Stasiak, M. Głąbowski, A. Wiśniewski, and P. Zwierzykowski, *Modeling and dimensioning of mobile networks: from GSM to LTE*, John Wiley & Sons, Chichester, West Sussex, 2011.

[29] H. Holma and A. Toskala, *WCDMA for UMTS. Radio Access For Third Generation Mobile Communications*, John Wiley & Sons, Chichester, West Sussex, 2004.

[30] M. Stasiak, P. Zwierzykowski, and M. Głąbowski, "High-Speed Downlink Packet Access (HSDPA) and High-Speed Uplink Packet Access (HSUPA) Cells: Teletraffic Engineering", in *Encyclopedia of Wireless and Mobile Commun.*, Taylor & Francis, London–New York, 2012, 552–568.

[31] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, John Wiley & Sons, Chichester, West Sussex, 2011.

[32] E. Dahlman, S. Parkvall and J. Skold, *4G, LTE-Advanced Pro and The Road to 5G*, Academic Press, Inc., Orlando, FL, USA, 2016.

[33] H. Holma and A. Toskala, *HSDPA/HSUPA for UMTS: High Speed Radio Access for Mobile Communications*, John Wiley & Sons, Chichester, West Sussex, 2006.

[34] S. Hanczewski, M. Stasiak, and P. Zwierzykowski, "Modelling of the access part of a multi-service mobile network with service priorities", *J. Wireless Com. Network.* 194 (2015).

[35] M. Stasiak, P. Zwierzykowski, J. Wiewióra, and D. Parniewicz, "An Approximate Model of the WCDMA Interface Servicing a Mixture of Multi-rate Traffic Streams with Priorities", in *Comp. Perform. Eng.*, Vol. 5261, Springer, Spain, 2008, 168–180.

[36] M. Stasiak, J. Wiewióra, and P. Zwierzykowski, "The analytical model of the WCDMA interface with priorities in the UMTS network", in *Proc. Int. Symp. Inf. Theory and its App.*, Auckland, New Zealand, 2008.

[37] M. Stasiak, J. Wiewióra, and P. Zwierzykowski, "WCDMA interface in UMTS network carrying a mixture of multi-rate traffic with different priorities", in *Proc. Int. Conf. Telecomun.*, Marrakech, Maroco, 2009, 265–270.

[38] M. Stasiak, J. Wiewióra, and P. Zwierzykowski, "Analytical model of the WCDMA radio interface in UMTS network with user differentiation", in *Proc. of 5th Adv. Int. Conf. Telecomun.*, Venice, Italy, 2009, 83–88.

[39] J. Kołakowski and J. Cichocki, *UMTS System telefonii komórkowej trzeciej generacji*, Wydawnictwo Komunikacji i Łączności, Warszawa, 2003.

[40] H. Holma and A. Toskala, *WCDMA for UMTS: HSPA Evolution and LTE*, John Wiley & Sons, New York, London, 2010.

[41] M.S. Taqqu, W. Willinger, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling", *ACM Comp. Commun. Rev.* 27(2), 5–23 (1997).

[42] W. Willinger, V. Paxson, R.H. Riedi, and M.S. Taqqu, "Long-Range Dependence and Data Network Traffic", in *Theory and applications of long-range dependence*, Brikhauser Boston and Springer Vertlag, Boston, USA, 2003, 373–406.

[43] A. Chydzinski, "Transient analysis of the MMPP/G/1/K queue", *Telecommun. Syst.* 32(4), 247–262 (2006).

[44] B. Filipowicz and J. Kwiecień, "Queueing systems and networks. Models and applications", *Bull. Pol. Ac.: Tech.* 56(4), 379–390 (2008).

[45] W. Oniszczuk, "Loss tandem networks with blocking – a semi-markov approach", *Bull. Pol. Ac.: Tech.* 58(4), 673–681 (2010).

[46] O. Tikhonenko and M. Ziółkowski, "Single-server queueing system with external and internal customers", *Bull. Pol. Ac.: Tech.* 66(4), 539–551 (2018).

[47] O. Tikhonenko, M. Ziółkowski, and M. Kurkowski, "M/G/N/(0,V) Erlang queueing system with non-homogeneous customers non-identical servers and limited memory space", *Bull. Pol. Ac.: Tech.* 67(3), 489–500 (2019).

[48] B. Ryu and S. Lowen, "Fractal traffic models for internet simulation", in *Proc. of the 5th IEEE Symp.Comp.Commun.*, Washington, DC, USA, IEEE Computer Society, 2000, 200–206.

[49] T. Bonald and J. Virtamo, "A recursive formula for multirate systems with elastic traffic", *IEEE Commun. Lett.* 9(8), 753–755 (2005).

[50] T. Bonald and J. Roberts, "Internet and the Erlang formula", *SIGCOMM Comput. Commun. Rev.* 42(1), 23–30 (2012).

[51] F. Kelly, *Notes on effective bandwidth*, Technical report, University of Cambridge, 1996.

[52] A.K. Erlang, "Løsning af nogle Problemer fra Sandsynlighedsregningen af Betydning for de automatiske Telefoncentraler", *Elektroteknikeren* 13, 138-–155 (1917).

[53] J. Kaufman, "Blocking in a shared resource environment", *IEEE Trans. Commun.* 29(10), 1474–1481 (1981).

[54] J. Roberts, "A service system with heterogeneous user requirements – application to multi-service telecommunications systems", in *Proc. of Perform. of Data Commun. Syst. and their App.*, Amsterdam, Elsevier Science Ltd., 1981, 423–431.

[55] G. Stamatelos and V. Koukoulidis, "Reservation-based bandwidth allocation in a radio ATM network", *IEEE/ACM Trans. Netw.* 5(3), 420–428 (1997).

[56] M. Stasiak, "Queuing systems for the Internet", *IEICE Trans. Commun.* 99-B(6), 1234–1242 (2016).

[57] S. Hanczewski, M. Stasiak, and J. Weissenberg, "A queueing model of a multi-service system with state-dependent distribution of resources for each class of calls", *IEICE Trans. Commun.* 97-B(8), 1592–1605 (2014).

[58] L. Katzschner, "Loss sytstems with displacing priorities", in *Proc. 6th Int. Teletraffic Cong.*, Munchen, Germany (1970), paper 224II–224.