Modified Direct Method for Point-to-Point Blocking Probability in Multi-service Switching Networks with Resource Allocation Control

Mariusz Głąbowski(✉), Maciej Sobieraj, and Maciej Stasiak

Faculty of Electronics and Telecommunications, Poznan University of Technology, Poznań, Poland
mariusz.glabowski@put.poznan.pl

Abstract. This article proposes a simplified approach to the internal blocking probability calculation in switching networks with mechanisms controlling resource allocation for offered multi-service traffic streams. This resource allocation control can be executed using the so-called threshold and resource reservation mechanisms, according to which the volume of resources admitted depends on a traffic class and on the occupancy state of the interstage and outgoing links of the switching network. The developed method is of generic nature and allows one to model switching systems regardless of the implemented resource allocation control mechanism. However, despite its generic character, the method provides better accuracy as compared to the methods worked out earlier.

Keywords: Switching networks · Teletraffic · Performance evaluation · Elastic optical networks

1 Introduction

One of the basic elements that influences the efficiency and effectiveness of telecommunications networks is, besides their structure and traffic routing rules, network nodes and network devices like routers and switches that are used to connect services in these networks. All high-performance network devices contain a switching structure, implemented in the form of a switching network. Available studies and investigations [2,12,13] clearly show that the multi-stage Clos switching network is one of the most effective structures that provide high scalability. The universality of the Clos switching network also allows it to be applied to large data centres and elastic optical networks [10]. Multi-stage switching networks, both electronic and optical, even though they have a number of substantial differences, share many similarities, which makes it possible to model them at the level of streams (flows, calls) using the so-called *multi-rate* models [1,9].

T. Q. Duong et al. (Eds.): Qshine 2018, LNICST 272, pp. 109–118, 2019.
https://doi.org/10.1007/978-3-030-14413-5_9

Currently, among a large number of methods for modeling switching networks, the so-called *effective availability methods* are decidedly dominant. According to the effective-availability concept, blocking probability in a multi-stage switching network with multi-rate (multi-service) traffic are determined using equivalent single-stage single-service model, i.e., Ideal Erlang's Grading [3,11]. Initially, these methods were used to model switching networks that did not differentiate the quality of service for individual call streams [11]. The analytical studies that made it possible to take into account appropriate resource reservation algorithms for individual traffic streams in inter-stage links and outgoing links of a switching network are proposed, among others, in [8]. The works that followed deal with switching networks with elastic and adaptive services (implemented using threshold mechanisms) have been proposed in [7]. A generalized recurrent method for analytical calculation of blocking probability in multi-stage switching networks with implemented mechanisms for controlling resource allocation is proposed in [4]. The unification of the modeling patterns for different mechanisms for controlling resource allocation for particular traffic streams made it then possible, in the process of a determination of blocking probability in switching networks, to take into consideration traffic streams that are generated according to the following distributions: Erlang, Engset and Pascal. The recurrent method for modelling switching networks used in [4] is dedicated to multi-stage networks with a greater number of stages (greater than 3). This method was initially applied to networks without resource allocation control [5]. A modification to the recurrent method, introduced in [4], that provided better accuracy of calculations of the internal blocking, can be also applied in the case of the remaining effective availability methods that allow the point-to-point blocking probability to be determined, i.e. the PPBMT method and the direct method [8] that are dedicated to three-stage networks.

The present article discusses the modified direct method for determining point-to-point blocking probability in switching networks with introduced mechanisms for controlling resource allocation for multi-service traffic flows. The remaining part of the article is divided as follows. In Sect. 2 a basic model of the switching network is presented. Section 3 presents the method for determining the external blocking probability and the modified method for a determination of the internal blocking probability. Section 4 includes the numerical results that allow the accuracy of the introduced modifications to be evaluated. Section 5 sums up the most important results presented in the paper.

2 General Traffic Model of the Switching Network

Let us consider the multi-service ($z = 3$)-stage blocking switching networks [4] with the Clos structure (Fig. 1). Each of z stages is composed of υ square switches with $\upsilon \times \upsilon$ links. The capacity of inter-stage links is limited to f allocation units (AUs). The output links of the switching network are grouped in the so-called outgoing directions that lead to the neighboring nodes. In the case of a typical execution of output directions, the direction r is composed of υ links outgoing from the switches of the z stage.

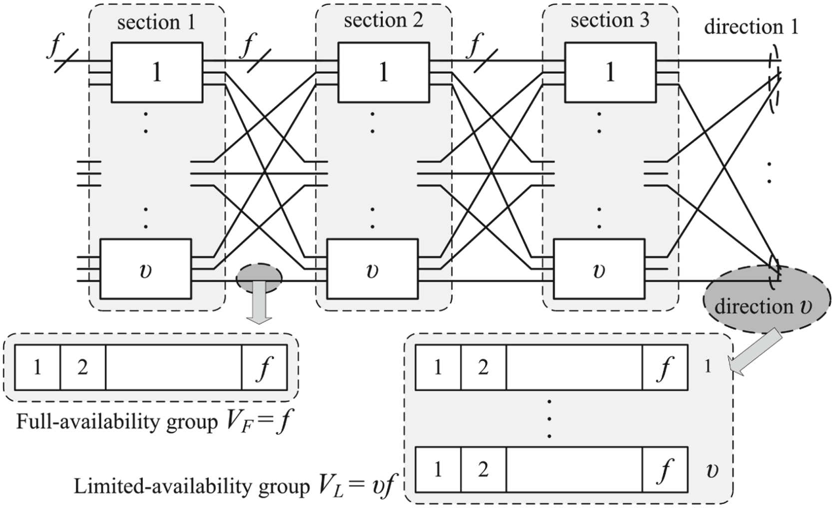

Fig. 1. 3-stage switching network

This network is offered traffic streams generated by three types of traffic sources, defined in traffic theory, i.e. Erlang, Engset and Pascal sources. Each of these types is related to different call streams, described by the flow intensity (call arrival) and the number of required AUs. The network under consideration is offered m_I Erlang traffic streams, m_J Engset traffic streams and m_K Pascal traffic streams. A call of any class c ($1 \le c \le m$, where $m = m_I + m_J + m_K$) requires t_c AUs for a service. Service times for calls of class c follows an exponential distribution with the parameter μ_c. Traffic intensity A_i of class i ($1 \le i \le m_I$) Erlang stream, traffic intensity $A_j(n)$ of class j ($1 \le j \le m_J$) Engset stream and traffic intensity of class k ($1 \le k \le m_K$) Pascal stream $A_k(n)$ are equal:

$$A_i = \lambda_i / \mu_i, \tag{1}$$

$$A_j(n) = (S_j - n_j(n)) \Lambda_j / \mu_j, \tag{2}$$

$$A_k(n) = (S_k + n_k(n)) \gamma_k / \mu_k, \tag{3}$$

where: n – the number of currently occupied allocation units – network state; λ_i – call arrival intensity for Erlang calls (according to Poisson distribution); S_j – the cardinality of Engset sources of class j; $n_j(n)$ – the number of serviced calls of class j in the state of n AU being busy; Λ_j – call arrival intensity for a free Engset source of class j; S_k – the cardinality of Pascal sources of class k; $n_k(n)$ – the number of serviced calls of class k in the state of n AU being busy; γ_k – call arrival intensity for a free Pascal source of class k.

Switching networks can operate in different modes for output link selection of links that lead to a successive node over a connection path. One of the most frequently used selection modes considered in traffic engineering is the point-to-point selection and point-to-group selection. According to the point-to-point selection, which is the subject of the present article, the process of finding a path for a new call (stream) of class c starts with a determination of a switch of the last stage that has a free outgoing link (having not less than t_c free AUs) in the demanded direction. Then, an attempt is made at setting up a connection inside the network, between the first stage switch (where a class c call has arrived) and the selected last stage switch. If a free connection path is available, then the connection is successfully executed. Otherwise, the call is lost (blocked) due to the occurrence of internal blocking phenomenon. The call can be also lost due to the occurrence of external blocking phenomenon, i.e., when there is no last stage switch having at least one outgoing link having no less than t_c AUs. In order to calculate the total blocking probability $E_{\mathrm{Tot}}(c)$ for class c calls offered to the switching network, the internal and external blocking phenomena should be taken into account, according to the following formula:

$$E_{\mathrm{Tot}}(c) = E_{\mathrm{Ex}}(c) + E_{\mathrm{In}}(c)\left[1 - E_{\mathrm{Ex}}(c)\right], \quad (4)$$

where $E_{\mathrm{Ex}}(c)$ denotes the external blocking probability, while $E_{\mathrm{In}}(c)$ denotes the internal blocking. The form of the Formula (4) results from the operation of the algorithm for the selection of a connection path in the switching network with the point-to-point selection.

3 Total Blocking Probability

3.1 External Blocking Probability

The blocking probability in outgoing directions is usually determined on the basis of a model of the limited-availability group (LAG). This probability depends on the structure of LAG (the number of outgoing links, the capacity of outgoing links) and on the introduced mechanisms to control the volume of allocated resources, i.e. on threshold mechanisms and reservation mechanisms.

The Structure of LAG. The LAG model is a model of the distributed system composed of υ separated links having the capacity of f AU. The total capacity of the outgoing direction (multi-service system) $V_L = \upsilon f$ AU. A new call can be admitted for service only when it can be serviced by the resources of one of available links. The influence of the structure on traffic characteristics is taken into consideration in analytical models by an introduction of the conditional transition coefficients $\sigma_{c,S_L}(n)$:

$$\sigma_{c,S_L}(n) = [F(V_L - n, \upsilon, f, 0) - F(V_L - n, \upsilon, t_c - 1, 0)]/F(V_L - n, \upsilon, f, 0), \quad (5)$$

where $F(x, \upsilon, f, t)$ allows us to calculate the number of possible arrangements of x of free AUs in υ links, each of which having the capacity equal to f AUs, with an additional assumption that there are at least t free (unoccupied) AUs in each link:

$$F(x, \upsilon, f, t) = \sum_{r=0}^{\left\lfloor \frac{x-\upsilon t}{f-t+1} \right\rfloor} (-1)^r \binom{\upsilon}{r} \binom{x - \upsilon(t-1) - 1 - r(f-t+1)}{\upsilon - 1}. \quad (6)$$

Threshold Mechanism in Output Directions. Threshold mechanisms make it possible to dynamically change the amount of resources admitted to particular service classes on the basis of the number of AUs being occupied in the outgoing directions of a switching network [7]. When a certain predefined limit of the occupancy of the outgoing direction in a switching network is exceeded, then a decrease in the resources allocated to calls of a given service class follows. As a result, this leads to an extension of the service time (for elastic services, in which data have to be transmitted in their entity), or to a decreasing of bitrate of transmitted multi-media data (for adaptive services, in which a prolongation of service time is not allowed) [7].

In analytical models of threshold systems it is assumed that each class c has a defined set of thresholds: $\{Q_{c,1}, Q_{c,2}, ..., Q_{c,q_c}\}$, where $Q_{c,1} \leq Q_{c,2} \leq ... \leq Q_{c,q_c}$. The further assumption is that the threshold area u of class c, limited by the threshold $Q_{c,u}$ and threshold $Q_{c,u+1}$, is defined by the given set of parameters $\{t_{c,u}, \mu_{c,u}\}$, where $t_{c,0} > t_{c,1} > ... > t_{c,u} > ...t_{c,q_c}$ and $\mu_{c,0}^{-1} \leq \mu_{c,1}^{-1} \leq ... \leq \mu_{c,u}^{-1} \leq ... \leq \mu_{c,q_c}^{-1}$. The parameter $\sigma_{c,\mathrm{T},u}(n)$ determines the occupancy states of threshold area u in which offered traffic is defined by the given parameters $t_{c,u}$ and $\mu_{c,u}$:

$$\sigma_{c,\mathrm{T},u}(n) = \begin{cases} 1 & \text{for } Q_{c,u} < n \leq Q_{c,u+1}, \\ 0 & \text{for remaining } n. \end{cases} \tag{7}$$

Let us notice that after applying threshold mechanisms in outgoing links (modeled by the LAG) we have introduced another type of dependency (besides the one imposed by the group structure) to the service process occurring in the switching networks under consideration. Since the threshold mechanism does not depend on the structure of the outgoing direction (only on its occupancy), it provides an opportunity for a product-form determination of the transition coefficient in LAG $\sigma_{c,S,u}(n)$:

$$\sigma_{c,S,u}(n) = \sigma_{c,S_L}(n) \cdot \sigma_{c,\mathrm{T},u}(n). \tag{8}$$

By taking into consideration the conditional transition coefficient it is possible to determine the occupancy state for each area u [7]:

$$\begin{aligned} n\,[Q_n]_{V_L} = \sum\nolimits_{i=1}^{m_I} \sum\nolimits_{u=0}^{q_i} A_i t_i \sigma_{i,S,u}(n-t_i)\,[Q_{n-t_i}]_{V_L} \\ + \sum\nolimits_{j=1}^{m_J} \sum\nolimits_{u=0}^{q_j} A_j(n)\sigma_{j,S,u}(n-t_j) t_j \left[Q_{n-t_j}\right]_{V_L} \\ + \sum\nolimits_{k=1}^{m_K} \sum\nolimits_{u=0}^{q_k} A_k(n)\sigma_{k,S,u}(n-t_k) t_k\,[Q_{n-t_k}]_{V_L}. \end{aligned} \tag{9}$$

Bandwidth Reservation Mechanism in Output Directions. The reservation mechanism can be treated as a particular case of a threshold mechanism. According to the reservation mechanism, calls of particular classes cannot be admitted for service in a certain space of occupancy states of groups. The reservation mechanism is usually used to enforce an appropriate access to the resources of a system for all traffic classes. In a particular case, the reservation mechanism ensure equal blocking probability for all traffic classes, regardless their resource requirements. In analytical models of systems with reservation, the so-called *reservation limit/boundary* R_c for each traffic class is defined. The parameter R_c defines such a particular state of the system in which it can still admit a class c call. Note that an analysis of thus considered system can be brought in its essence to an analysis of a threshold system in which, after the last threshold is exceeded, the volume of allocated resources is equal to zero.

Determination of the External Blocking Probability. The external blocking probability $E_{\mathrm{Ex}}(c)$ can be calculated using LAG model of the outgoing directions in the switching network [7]. To determine the external blocking probability it is necessary to first calculate the occupancy distribution in LAG using the generalized Kaufman-Roberts distribution (9). After the occupancy distribution $[Q_n]_V$ has been determined, the blocking probability of calls of particular traffic classes in LAG, regardless of the applied resource allocation control mechanism, can be expressed by the following formula:

$$E_{\mathrm{Ex}}(c) = \sum\nolimits_{n=V_L-\upsilon(t_{c,q_c}-1)}^{V_L} [Q_n]_{V_L}. \tag{10}$$

3.2 Internal Blocking Probability

Determination of the Effective Availability Parameter. Internal blocking phenomenon decreases the number of last stage switches available to the first stage switch. According to the approach of the effective availability methods, the availability of last stage switches for calls of class c (appearing at the first stage switches) in multi-service switching networks can be calculated using the concept of the so-called equivalent network [11], determined individually for each of the classes of offered traffic. The assumption is that the equivalent single-service network for calls of class c services only the considered type of calls. The equivalent switching network has the same physical topology as the real multi-service network. The only difference is that the load (fictitious) of each inter-stage links have the load $y_{c,l}$ equal to the blocking probability calculated for calls of class c in the real, between stages l and $l+1$. This probability can be determined on the basis of appropriate models of inter-stage groups that will be presented further on in the article. To determine the effective availability $d_{c,3}$ in networks, the universal formula [11] for 3-stage switching networks with multi-service traffic can be used:

$$d_{c,3} = [1-\pi_{c,3}]\upsilon + \pi_{c,3}y_{c,1} + \pi_{c,3}[\upsilon - y_{c,1}]y_{c,3}[1-\pi_{c,3}], \tag{11}$$

where $\pi_{c,3}$ – probability of direct unavailability of a given (specified) switch in stage 3 for a call of class c; $\pi_{c,3}$ that determines the probability of an event in which a connection of class c cannot be set up between specified (given) switches of the first and the last stage [4]; $y_{c,l}$ – fictitious load of inter-stage link between stages l and $l+1$ in the equivalent switching network for calls of class c, equal in terms of the value to the blocking probability $[e_{c,l}]_f$ for calls of class c in the link of a real network [4].

Blocking Probability in Inter-stage Links. In the models that have been developed and used thus far, the influence of threshold mechanisms used in the outgoing directions on the blocking probability of the inter-stage links was determined by mapping the thresholds from outgoing directions (with higher capacity) to the thresholds in inter-stage links (with lower capacity) [7]. In the method described in this article, this influence is reflected by increasing the number of traffic classes that are offered to inter-stage links. The increase in the number of traffic classes results directly from the threshold mechanism applied to a given outgoing direction since the calls of class c, offered to outgoing directions, can be allocated $(q_c + 1)$ different $t_{c,u}$ values of resources (AUs), depending on the occupancy state of the direction. Consequently, the simple full-availability model (FAG), without threshold mechanisms, can be used to model the inter-stage links with the increased number of traffic classes offered [6]. In order to avoid duplication of indexes, denoting particular traffic classes offered to the inter-stage links, the following formulas should be used:

$$\forall_{1\le c\le m}\forall_{0\le u\le q_c} t_{c+u+\sum_{z=1}^{c-1} q_z} = t_{c,u},\ \forall_{1\le c\le m}\forall_{0\le u\le q_c} \mu_{c+u+\sum_{z=1}^{c-1} q_z} = \mu_{c,u}. \quad (12)$$

In order to calculate the occupancy distribution in full-availability groups (modelling inter-stage links with the increased number of traffic classes) it is necessary to calculate traffic intensity in particular threshold areas of the outgoing direction, which is offered subsequently to the interstage links. Let us assume that in the pre-threshold area of the outgoing direction the traffic intensity of the class c traffic stream that requires $t_{c,0}$ AUs is equal to $A_{c,0}$. Consequently, an inter-stage link will be offered a class c call that demands $t_{c+0} = t_{c,0}$ AUs to set up a connection, but the traffic intensity for this traffic stream will be equal to $A_{c,0}(1 - e_{c,1})/\upsilon$, where $e_{c,1}$ is the blocking probability for class c calls with the original demands (in the pre-threshold area). Subsequently, the traffic blocked in the pre-threshold area is offered to the first threshold area with the intensity $A_{c,0}e_{c,1}$. As a result, this traffic leads to the appearance of "new" streams that require $t_{c+1} = t_{c,1}$ in the inter-stage link. Continuing the above reasoning, the part of the traffic that is not blocked in the area u creates "new" traffic of class c with the value:

$$(1 - b_{c,u+1})\frac{A_{c,0}}{\upsilon}\prod_{q=1}^{u} b_{c,q}, \quad (13)$$

where the parameter $b_{c,u}$ determines the blocking probability of calls of class c to which a given number of AUs is allocated, proper for the threshold area $u-1$:

$$b_{c,u} = \sum_{n=Q_{c,u}^{L}+1}^{V_L} [Q_n]_{V_L}. \quad (14)$$

Note that in Formula (13) the traffic intensity, that demands $t_{c,u}$ AUs, is divided by υ in order to take into account the difference in the capacity of an outgoing direction and an interstage links (an outgoing direction is υ times higher than an inter-stage link).

With the values of Erlang, Enget and Pascal traffic offered to inter-stage links determined, on the basis of the above reasoning, we are in position to determine the occupancy distribution $[Q_n]_{V_F}$ in the inter-stage link with the capacity $V_F = f$ on the basis of Formula (9), assuming the value of the conditional transition coefficient to be equal to 1 and taking into account the increased number of traffic classes and capacity of the inter-stage links. This distribution allows the blocking probability for all the traffic classes offered to the inter-stage groups to be determined (Formula (10)).

The values of the blocking probability in the inter-stage links determined on the basis of the method presented in Sect. 3.2 make it possible to determine the effective availability parameter. Thus modified value of the effective availability provides then the basis for the modification of the two methods for a determination of the internal point-to-point blocking probability, i.e. the direct MDu method.

Direct Method. The internal point-to-point blocking probability in the universal direct method (MDu), based on the direct method (MD) [8], is defined as the ratio between the free links (for calls of class c) in the group of switches unavailable for a switch and all free (unoccupied) links that belong to a given direction. With the assumption that the occupancy probabilities of any links in the group are equal and independent of the occupancy of other links, the average value of the internal blocking probability can be expressed by the following formula:

$$E_{\mathrm{In}}(c) = (\upsilon - d_{c,3})/\upsilon, \tag{15}$$

where υ is the capacity of the output direction, expressed in the number of links, whereas $d_{c,3}$ is the average value for availability (the number of available switches of the last stage) for a call of class c. This parameter can be determined on the basis of Formula (11), in which the parameters of the fictitious load $y_{c,l}$ are determined on the basis of the full-availability group model with an increased number of traffic classes.

4 Numerical Results

The generalized MDu method for determining the blocking probability in switching networks with resource allocation control mechanisms is an approximate method. To evaluate its accuracy and the adopted assumptions for the method, the results of the analytical calculations were compared with the data obtained in simulation experiments. The simulation study was carried out for 3-stage Clos networks. The switching network under investigation was composed of square switches $\upsilon \times \upsilon$ links, each with the capacity of f AUs. The data obtained on the basis of the simulation study are presented in Fig. 2 as points with the confidence intervals calculated after the t-Student distribution (with 95-percent confidence level) for 5 series with 1,000,000 calls each (the classes with the lowest call intensity). In a large number of instances, the value of the confidence interval is lower that the height of the symbol representing the simulation result.

A comparison was also made for the changes in the blocking probability for individual call classes obtained on the basis of the proposed generalized MDu method and on the basis of its original version. This is illustrated in Fig. 3 that show the changes in the values of errors introduced by the modified MDu method and the values of errors from before the introduction of modifications. Due to a limited length of the present article, the results of the study are limited only to one network with the following structure: $v = 4$, $f = 42$ AUs, $V = 168$ AUs. This network was offered traffic with the following parameters: Traffic classes: $m = 4$, $t_{1,0} = 1$ AU, $\mu_{1,0}^{-1} = 1$, $t_{2,0} = 6$ AUs, $\mu_{2,0}^{-1} = 1$, $t_{3,0} = 8$ AUs, $\mu_{3,0}^{-1} = 1$, $t_{4,0} = 12$ AUs, $\mu_{4,0}^{-1} = 1$, $t_{4,1} = 8$ AUs, $\mu_{4,1}^{-1} = 1.25$, $t_{5,0} = 12$ AUs, $\mu_{5,0}^{-1} = 1$, $t_{5,1} = 8$ AUs, $\mu_{5,1}^{-1} = 1.5$; Threshold mechanism: $q_4 = 1$, $q_5 = 1$, $Q_{4,1} = Q_{5,1} = 126$ AUs.

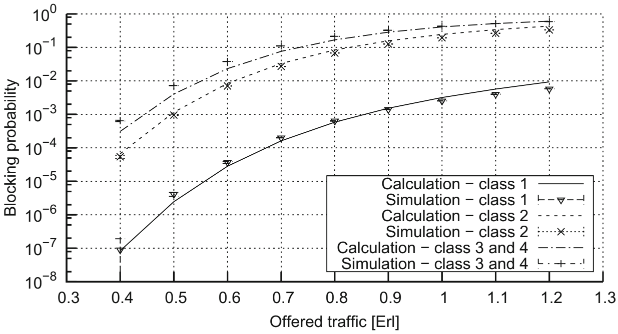

Fig. 2. Total blocking probability according to direct MBu method

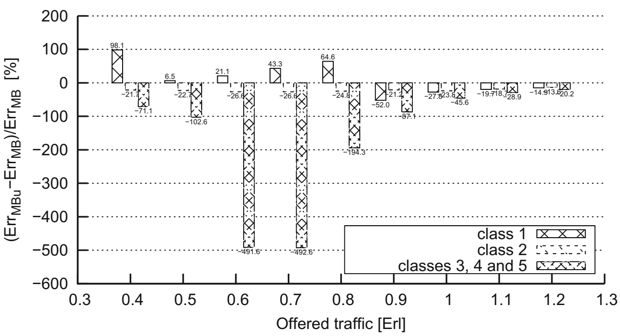

Fig. 3. Change in the blocking probability error (Err) – methods MD and MDu

5 Conclusions

This article proposes a modification to available methods for determining the point-to-point blocking probability that is based on the concept of a substitution of the system of inter-stage links with threshold mechanisms for a system of full-availability groups with appropriately increased number of traffic classes, which in consequence leads to considerably higher accuracy of the latter methods. In the case of the MDu method, the increase in the accuracy is particularly distinct for those traffic classes that demand larger amounts of resources[1].

References

1. Bonald, T.: A recursive formula for estimating the packet loss rate in IP networks. In: Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools, pp. 56:1–56:2 (2009)
2. Chrysos, N., Minkenberg, C., Rudquist, M., Basso, C., Vanderpool, B.: SCOC: high-radix switches made of bufferless clos networks. In: 2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA), pp. 402–414 (2015)
3. Ershov, V.: Some further studies on effective accessibility: fundamentals of teletraffic theory. In: Proceedings of 3rd International Seminar on Teletraffic Theory, pp. 193–196. Moscow (1984)
4. Głąbowski, M., Sobieraj, M.: Analytical modelling of multiservice switching networks with multiservice sources and resource management mechanisms. Telecommun. Syst. **66**(3), 559–578 (2017)
5. Głąbowski, M.: Recurrent calculation of blocking probability in multiservice switching networks. In: Proceedings of Asia-Pacific Conference on Communications, pp. 1–5, Busan (2006)
6. Głąbowski, M., Kaliszan, A., Stasiak, M.: Modeling product-form state-dependent systems with BPP traffic. Perform. Eval. **67**, 174–197 (2010)
7. Głąbowski, M., Sobieraj, M.: Modelling of network nodes with threshold mechanisms and multi-service sources. In: 2014 16th International Telecommunications Network Strategy and Planning Symposium (Networks), pp. 1–7 (2014)
8. Głąbowski, M., Stasiak, M.: Point-to-point blocking probability in switching networks with reservation. Ann. Telecommun. **57**(7–8), 798–831 (2002)
9. Pras, A., Nieuwenhuis, L., van de, R.M., Mandjes, M.: Dimensioning network links: a new look at equivalent bandwidth. IEEE Netw. 23(2), 5–10 (2009). http://doc.utwente.nl/65443/
10. Sehery, W., Clancy, T.: Load balancing in data center networks with folded-Clos architectures. In: 2015 1st IEEE Conference on Network Softwarization (NetSoft), pp. 1–6, April 2015
11. Stasiak, M.: Combinatorial considerations for switching systems carrying multichannel traffic streams. Anna. Des Télécomm. **51**(11–12), 611–625 (1996)
12. Xia, Y., Hamdi, M., Chao, H.: A practical large-capacity three-stage buffered Clos-network switch architecture. IEEE Trans. Parallel Distrib. Syst. **27**(2), 317–328 (2016)
13. Ye, T., Lee, T., Hu, W.: AWG-based non-blocking Clos networks. IEEE/ACM Trans. Netw. **23**(2), 491–504 (2015)

[1] The work is supported by Polish Ministry of Science and Higher Education 08/82/DSPB/8216.