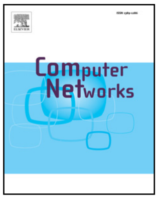
Computer Networks 108 (2016) 171–183

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



Modelling overflow systems with distributed secondary resources

Mariusz Głąbowski*, Adam Kaliszan, Maciej Stasiak

Faculty of Electronics and Telecommunications, Poznan University of Technology, ul. Polanka 3, Poznan 60-965, Poland

ARTICLE INFO

Article history:
Received 30 November 2015
Revised 19 July 2016
Accepted 15 August 2016
Available online 17 August 2016

Keywords:
Multi-service
Data centres
MPLS/GMPLS
Multi-networking
Overflow system
Distributed resources
Blocking probability
Elastic traffic

ABSTRACT

This article presents, in short, the state of the art of modelling overflow systems and proposes a new method for a determination of characteristics of multi-service overflow systems with elastic traffic and with distributed secondary resources. A particular attention is given to the method for a determination of the parameters of multi-service traffic that overflows to secondary resources and the method for modelling of these resources that takes into consideration both the structure of the distributed resources, elastic traffic management-handling mechanisms and the level of peakedness degeneration of overflow traffic. The results of the analytical calculations are compared with the results of the simulation experiments of a number of selected structures of overflow systems with distributed secondary resources. The results of the study have confirmed and verified the correctness of all the theoretical assumptions adopted throughout the article.

1. Introduction

Traffic overflow is one of the oldest techniques for optimization of traffic distribution in network and computer systems. Traffic overflow occurs when resources of a given system, called primary resources, are completely occupied and, as a result of this situation, new calls that are offered to these resources cannot be immediately serviced. Such calls can be redirected (or can overflow) to other systems, called secondary resources, that concurrently have free resources necessary to service new calls [1–3].

Initially, traffic overflow was primarily used in hierarchical telecommunications networks with alternate routing [4]. Over time, the technique of traffic overflow has been also applied to packet networks [5–8], where the main reason behind a decision to apply traffic overflow is not routing any more, but rather primarily load balancing [2]. In recent years, the traffic overflow technique has started to be used to increase the performance and capability of networked cloud data centres [2], content delivery network [9] and multi-service communication networks [1,3,10].

In the case of data centres (or cloud data centres), the basic rationale in favour of using traffic overflow includes: load balancing [9,11] and power saving [2,12]. Load balancing can be implemented either within the same Data Centre or between distributed Data Centres interconnected with one another using new appropriate network technologies, distributed data centres. Tasks that cannot be run in a given Data Centre due to heavy workloads can overflow to a server (virtual machine) within the Data Centre, or to another Data Centre of the same Cloud Data Centre. Furthermore, the inclusion of time and energy needed for activating successive machines (a given Data Centre is composed of) in the consideration, as well as the energy necessary to keep individual machines in stand-by readiness, results in an optimal management of energy used for the purpose. The applied traffic overflow mechanism can thus lead to a better operation of an already existing IT infrastructure, maintaining at the same time the Service Level Agreement parameters stipulated with users and minimizing the consumption of electrical energy.

Moreover, the advancing integration of IT and telecommunications solutions results in a situation where modern communication networks must carry large amounts of traffic and provide service to traffic streams with very differentiated demands, both in terms of bitrate, service time and demanded parameters of the Quality of Service [13]. Within the context of industrial applications, one of the most rapidly developing application of communication networks in recent times is a smart grid technology [14]. The advancing diversification in sources of energy imposes a necessity of using tools that would make remote data collection possible and in real time, most notably on the state of devices in electric power systems [15]. Currently, one of the most commonly used solutions that would guarantee reliable and effective measurement data delivery is – in the access part – wireless network technologies (GSM,

---

* Corresponding author. Fax: +48616653922.
E-mail addresses: mariusz.glabowski@put.poznan.pl (M. Głąbowski), adam.kaliszan@put.poznan.pl (A. Kaliszan), maciej.stasiak@put.poznan.pl (M. Stasiak).

http://dx.doi.org/10.1016/j.comnet.2016.08.015

UMTS, LTE) in which traffic overflow between networks operating in different technologies is commonly used (e.g. between 3G macrocells and 4G microcells) [1,16,58]. Nowadays, in order to effectively service traffic streams with demands from the range of kilobits per second to gigabits per second, forwarded from access to backbone networks, the so-called hybrid optical networks, e.g. Optical Burst Transport Network (OBTN) [17] are most widely used. In lower layers, networks of this type use the technology of switching networks: Dense Wavelength Division Multiplexing (DWDM) that assures – in its most practical implementations – the bitrate of 10Gb/s with the use of a single wavelength. To execute channels with lower bitrate, the Multi-Protocol Label Switching (MPLS) packet system is commonly used, which makes it possible to divide resources offered by a single wavelength. Due to availabilities of DWDM channels, the capacity of the MPLS network can be appropriately adjusted depending on needs and requirements. In the case of present-day telecommunications networks, both access and backbone networks, the traffic overflow technique increases utilization of network resources and optimization of traffic distribution [1,17,18].

The analytical models of systems with traffic overflow proposed in the literature of the subject deal with both single-service and multiservice systems. Single-service traffic overflow systems have been the subject of numerous analyses presented in the literature, e.g. in [4,19–22]. Basic mathematical models for overflow systems with single-service traffic was developed as early as around the 1950s. In [4], the Equivalent Random Traffic method (ERT) is proposed. The proposed method is based on the Riordan formulas [4] that determine the average value and variance of overflow traffic depending on the capacity of primary resources and traffic offered to these resources. On the basis of traffic parameters for overflow traffic, the ERT method defines parameters for fictitious resources, called equivalent resources, that, in turn, form the basis for a determination of the blocking probability and other characteristics of the overflow system. Fredericks [19] proposes a more simpler method for a determination of the blocking probability in secondary resources that is based on a modification of Erlang formula, called Hayward formula, in which the capacity of resources and overflow traffic are divided by the peakedness coefficient for this traffic. Such an approach significantly simplifies the method for dimensioning of systems with traffic overflow.

The problem of traffic overflow in multiservice systems [23] has been addressed in a number of works. Chung and Lee [24] propose a concept for modelling overflow traffic on the basis of the analysis of the Markov-Modulated Poisson Process, while [25,26] on the basis of the Batched Poisson Process. A number of methods for modelling secondary resources with multiservice overflow traffic with required value of the peakedness coefficient has been presented, for example, in [25,27]. A coherent methodology for modelling and dimensioning of multiservice overflow systems is proposed in [18,28,29]. In [18,28,30], primary resources are modelled by a model of the full-availability group – FAG. FAG is a model of a single link with complete sharing policy, i.e. a model of a state-independent system in which a call will be always admitted for service, provided that the system has enough free resources to service this call. If FAG is offered a mixture of Erlang traffic[1], then the system can be modelled by a multi-dimensional Markov process, which, in consequence, leads to a determination of the occupancy distribution expressed by a simple recurrence formula [32,33]. In addition, there are also recurrence models of FAG for a mixture of Erlang, Engset and Pascal traffic [34]. The solutions presented in [18,28,30] make it possible to determine in a simple way the peakedness and intensity of traffic that overflows to alternative resources. To determine the characteristics of alternative resources that service overflow traffic, in [18,28,30], a modified FAG model – according to the method adopted by Hayward [19] for single-service systems – is used. This means that both the capacity of FAG and offered value of the traffic intensity for individual traffic classes are divided by peakedness coefficients of corresponding classes. In [29], an efficient analytical model is developed. The model is based on an appropriately defined two-dimensional convolution operation that provides a possibility to model overflow systems with any type of calls streams offered to primary resources. Wang et al. [3,10] discuss overflow systems in which the primary and secondary resources are characterized by other service parameters, such as service time and bitrate.

Traffic management mechanisms in the TCP/IP network allow bit rates of serviced calls to be changed (modified) thanks to the possibility of compression of serviced packet streams. In the case of a heavy load in network systems, the control functions decrease bitrates for serviced calls and, simultaneously, their service times are prolonged accordingly. Traffic that can undergo compression is often termed as elastic traffic (e.g. TCP traffic [35]). In [36], a FAG model that services elastic traffic is proposed. This model is used in the present article to model primary resources.

In all hitherto considered models presented in the literature the assumption was that secondary resources formed a single resource with complete sharing policy (even in the case of multi-tier networks, the resources at each tier were considered as the resources with complete sharing policy). This assumption made it possible to model secondary resources by a model of FAG to which overflow traffic streams with the peakedness coefficient higher than unity were offered. The present article considers for the first time a system in which secondary resources are not uniform (have distributed nature), i.e., are composed of a number of separated component resources. The term “separation” means that a call that overflows will be serviced only when at least one of the secondary resources has the appropriate amount of resources required for a call to be serviced. A definition such as the one above excludes a possibility of a division of the call between component resources. A limited availability group model (LAG model) is used in this article to model secondary resources. LAG is a good example of the so-called state dependent system in which the admittance of a new call for service depends on the structure of a system or on the call admission function [1,37]. LAG models are discussed and analysed in a number of works, including [38–41]. Stasiak [41] proposes a model in which all separated resources have identical capacities. In this model, a multi-dimensional Markov process is used to approximate the real service process in LAG. Such an approach makes it possible to determine the occupancy distribution in the system on the basis of simple recurrence equations. In [42], this model is generalized for a system that is composed of separated resources with different capacities, whereas in [37,43] models which also take into account Erlang, Engset and Pascal traffic are proposed. This article proposes a generalization of the LAG model to include a possibility of elastic traffic service.

As yet, no generalized model of an overflow system with elastic traffic that would provide a possibility to distribute secondary resources has been developed. In the present article, a new model of a multiservice overflow system that supports elastic traffic, both in primary and in distributed secondary resources, is presented. This means that in the case of a lack of free bitrates in the primary resources, currently serviced calls will be compressed, i.e. bitrates for serviced calls will be decreased and the service times for these calls will be prolonged accordingly. If the required compression boundary is exceeded, then a call in the primary resources will overflow to the secondary resources, while a call in the sec-

---

[1] We use the term “Erlang traffic” as the synonym of PCT1 traffic (Pure Chance Traffic of Type 1), i.e. traffic created by calls generated according to a Poisson distribution, offered to the system with a finite capacity and serviced according to an exponential distribution [31]

ondary resources will be lost. The accuracy of the proposed model will be evaluated on the basis of a comparison of the relevant analytical calculations with the results of simulation experiments for a number of selected structures of the dispersion of the secondary resources.

The article is structured as follows. Section 2 presents a description of a multiservice overflow system with elastic traffic and with distributed secondary resources. Section 3 includes a description of the FAG with elastic traffic as a model of primary resources. This section also includes a description of a method for the evaluation of the parameters for traffic that overflows from primary resources. Section 4 provides a description of a model of distributed secondary resources with elastic traffic and presents an algorithm for a determination of the blocking probability in an overflow system with distributed secondary resources. Section 5 presents a comparison of the results of analytical calculations with the results of simulation experiments for some selected structures of systems with distributed secondary resources. Section 6 sums up the article.

2. Elastic traffic overflow system with distributed secondary resources

2.1. Traffic representation in the system

Data delivery in modern networks is done in packets. The structure of packets and the method for their delivery to the receiver depend on the adopted technology of transmission. The use of packet techniques imposes a hierarchical view upon traffic representation in the network. Packets that belong to a given service being carried out at a certain moment in time create streams that can be considered and treated as calls (sessions, flows). At the packet level, streams usually have a very complex structure [44–47]. A description and analysis of systems that are offered packet streams are very complex and in most cases do not go beyond solutions that are only approximate and with little accuracy, or are just simulation solutions [48].

Research studies carried out in recent years have proved that at the call level traffic has the “Poissonian” nature [49]. This means that in the analysis of systems at the call level it is acceptable to adopt the exponential character of the call stream and service stream. This is a very favourable circumstance that allows us to approach dimensioning of complex network systems in the “Erlang” way, i.e., on the basis of Markov processes. A problem that appears in the application of “Erlang-type” models to the analysis of multiservice packet systems is variable bitrate of the packet stream. A solution to this problem, widely used in traffic engineering, is a change of variable bitrates of the packet stream into constant bitrates. They are selected on the basis of the maximum bitrates for individual calls or [49], alternatively, on the basis of the so-called equivalent bandwidth [50,51] determined for each of the call streams. The equivalent bandwidth is then a constant bitrate that determines the amount of the resources that are to be assigned to a given call by the network in order to provide successfully appropriate quality of service parameters (QoS). The equivalent bandwidth is most frequently determined according to the heuristic method [50,51] as a function of the total bitrate of the system, the maximum and average bitrate of the packet stream, the variance of the bitrate, acceptable delay of packets and other parameters characteristic for a given system. The method for a determination of constant bitrates (maximum bitrates or equivalent bandwidth) has no influence on the analytical structure of the proposed model. A choice as to the method for a determination of constant bitrates depends, in turn, on the network structure and is subject to arrangements between a network operator and those units that are responsible for designing and optimization of a network.

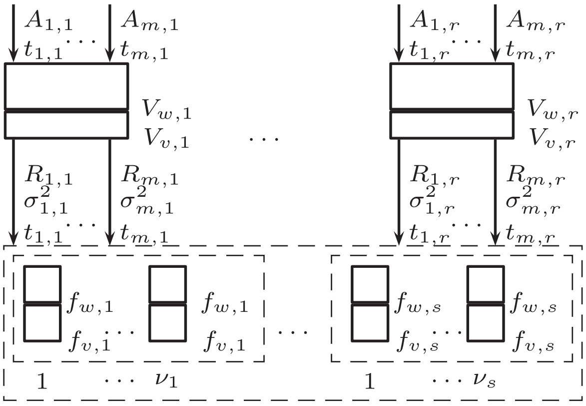


Fig. 1. Elastic traffic overflow system with distributed secondary resources.

After a determination of constant bitrates for calls of individual classes serviced in the system it is possible to proceed with the bitrate discretization process [52], which is based on a determination of the allocation unit for a given system. The allocation unit, the so-called Basic Bandwidth Units (BBU), is defined as the Greatest Common Divisor of all bitrates (maximum bit rates or equivalent bandwidths) allocated to calls of individual classes:

$$c_{\mathrm{BBU}} = \mathrm{GCD}(c_1, c_2, \ldots, c_m), \tag{1}$$

where m is the number of traffic classes offered to the system (i.e., the number of available services), c_i is the bitrate (the amount of resources) that corresponds to the maximum bit rates or equivalent bandwidth of class i, whereas c_{BBU} is the bitrate of the allocation unit. In the bitrate discretization process both the capacity of the system V and the amount of resources t_i necessary for a call of a given class i to be serviced, are expressed in BBUs:

$$V = C/c_{\mathrm{BBU}}, \tag{2}$$

$$t_i = c_i/c_{\mathrm{BBU}}. \tag{3}$$

2.2. Primary resources system

Fig. 1 shows a general diagram of an overflow system with distributed secondary resources.

The system of primary resources is composed of r primary resources. Each resource j $(0 < j \le r)$ has the real (working) capacity $V_{w,j}$ and virtual capacity $V_{v,j}$, expressed in BBUs, where $V_{v,j} > V_{w,j}$. Calls undergo compression until the number of occupied BBUs in the system, measured by the total sum of uncompressed demands of calls of all classes, exceeds the virtual capacity $V_{v,j}$. If the virtual capacity is exceeded, then new calls will be directed (will overflow) to secondary resources. The occupancy states n BBUs, such that $V_{w,j} < n \le V_{v,j}$, determine the compression area for elastic traffic. The value $V_{v,j}$ defines the maximum possible level of compression that is equal to the ratio between the virtual capacity of the resource and the real capacity $V_{v,j}/V_{w,j}$, and determines how many times the total bitrate for serviced calls in a given primary resource can be maximally decreased. In state n $(V_{w,j} < n \le V_{v,j})$ the level of compression for all serviced calls is identical and is equal to $n/V_{w,j}$.

Each resource is offered a set of m Erlang traffic streams. In Fig. 1, the following notation is adopted for offered traffic:

- $A_{i,j}$ – intensity of traffic of class i $(0 < i \le m)$ offered to resource j $(0 < j \le r)$ of a system of primary resources,
- $t_{i,j}$ – the number of BBUs that is necessary to service a call of class i in resource j of the primary resources system with no

compression involved; further on in the article this parameter will be called a demand of class *i* in resources *j*.

The article adopts the approach that, at the call level, each traffic stream can be considered as Erlang traffic [49]. This means that in the analysis of the primary resources system one can assume the exponential character of each call stream and service stream and assume that the value of the variance of offered traffic is equal to its average value.

Additionally, it is assumed that an overflow call from the primary system can access the entire secondary system (the call admission control in the secondary system is described in the following subsection).

2.3. Secondary resources system

The system of secondary resources is composed of *s* types of distributed secondary resources. Each type k $(0 < k \leq s)$ is unequivocally determined by the number ν_k of resources of a given type. Each of the ν_k resources has the real (working) capacity of $f_{w,k}$ BBUs and the virtual capacity $f_{v,k}$ BBUs. The total real capacity $V_{w,0}$ and the total virtual capacity $V_{v,0}$ of the system of secondary resources, expressed in BBUs, is then:

$$V_{w,0} = \sum\nolimits_{k=1}^{s} \nu_k f_{w,k}, \quad V_{v,0} = \sum\nolimits_{k=1}^{s} \nu_k f_{v,k}. \tag{4}$$

The secondary resources system can admit a call of a given class for service only when it can be entirely serviced by BBUs of one of the distributed component resources. This means that *t* BBUs required to set up a given connection, cannot be "divided" between different component resources and must be serviced exclusively by the BBUs that belong to one component resource. Additionally, we assume a pseudo-random choice algorithm of one of the component resources that is to service a call of a given class.

The secondary resources system is offered traffic streams that overflow from the primary resources system. The following notation for traffic offered to the secondary resources system is adopted in the article:

- $R_{i,j}$ – the average value of the intensity of traffic of class i $(0 < i \leq m)$ that overflows from resource j $(0 < j \leq r)$ that belongs to the system of primary resources,
- $\sigma^2_{i,j}$ – variance of the intensity of traffic of class *i* that overflows from resource *j* of the primary resources system,
- $t_{i,j}$ – uncompressed call demand of a call of class *i* that overflows from resource *j* of the primary resources system.

Unlike in Erlang-type traffic, the variance of overflow traffic is characterized by values that exceed the average value of this traffic. This particular fact is further used in constructing mathematical models that describe overflow systems.

3. Modelling of primary resources

3.1. Model of the system of primary resources

The model of a system with multiservice elastic traffic overflow to distributed secondary resources is composed of a model of the system of primary resources, a model for a determination of the parameters of overflow traffic and a model of the system of secondary resources. These three elements make it possible to determine all important characteristics of an overflow system, in particular the blocking probability for call streams of all classes offered to the system. To model each of the primary resources, the FAG model for elastic traffic [36] was used, whereas to determine the parameters of overflow traffic from the primary resources the method developed in [18] was used.

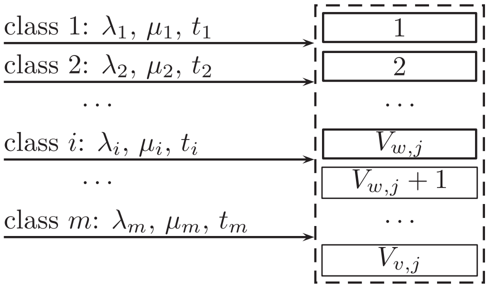


Fig. 2. Full-availability group (FAG) model.

The approach adopted in the article assumes that in order to model primary resources the recurrence FAG model with elastic traffic [36] can be used. FAG, presented in Fig. 2, is a model of a single resource with complete sharing policy (Fig. 2 shows in detail a structure of a single primary resource *j* of the overflow system presented in Fig. 1). According to this call service policy, a call is always admitted for service provided free resources for its service are available. In the FAG model, no constraints are introduced concerning resource allocation for incoming calls. It is thus a model of a system with state-independent call admission process for calls generated according to the Poisson distribution. The considered model assumes that calls of individual traffic classes that arrive to the system with the intensity: $\lambda_1, \lambda_2, \ldots, \lambda_m$, demand respectively $t_1, t_2, \ldots, t_m$ BBUs for service, while service times of calls of individual classes are in conformity with the exponential distribution with the parameters: $\mu_1, \mu_2, \ldots, \mu_m$. These assumptions allow us to determine the average traffic intensity offered by a stream of class *i* with the application of the following formula:

$$A_i = \lambda_i / \mu_i. \tag{5}$$

Let us notice that the class *i* traffic load offered to the system in the state of *n* busy (occupied) BBUs can be expressed as follows:

$$A_i(n) = \frac{\lambda_i}{\mu_i(n)} t_i(n), \tag{6}$$

where $1/\mu_i(n)$ denotes the service time of class *i* calls in the state of *n* busy BBUs, while $t_i(n)$ – the number of BBUs admitted to class *i* calls in the state *n*. Since in the case of elastic traffic a decrease in the number of resources allocated to particular demands is directly proportional to the prolongation of the service time, according to a coefficient $\tau(n)$:

$$t_i(n) = \frac{t_i}{\tau(n)}, \quad \frac{1}{\mu_i(n)} = \frac{1}{\mu_i}\tau(n), \tag{7}$$

we can re-written Formula (6) in the following form:

$$A_i(n) = \lambda_i \frac{\tau(n)}{\mu_i} \cdot \frac{t_i}{\tau(n)} = \frac{\lambda_i}{\mu_i} t_i = A_i t_i. \tag{8}$$

On the basis of this model, the occupancy distribution and the blocking probability for traffic of different classes in a given primary resource *j* with the real capacity $V_{w,j}$ and the virtual capacity $V_{v,j}$ can be written in the following way:

$$[P_n]_{V_{v,j}} = \frac{1}{\min(n, V_{w,j})} \sum\nolimits_{i=1}^{m} A_{i,j} t_{i,j} [P_{n-t_{i,j}}]_{V_{v,j}}, \tag{9}$$

$$[E_{i,j}]_{V_{v,j}} = \sum\nolimits_{n=V_{v,j}-t_{i,j}+1}^{V_{v,j}} [P_n]_{V_{v,j}}, \tag{10}$$

where n $(0 \leq n \leq V_j)$ denotes the number of busy (occupied) BBUs in resource *j*. The parameter $[P_n]_{V_{v,j}}$ is the occupancy probability of *n* BBUs in a system with the virtual capacity $V_{v,j}$ BBUs. The model assumes that for all $n < 0$ the probability $[P_n]_{V_{v,j}} = 0$. The symbol $[E_{i,j}]_{V_{v,j}}$ in (10) denotes the blocking probability for a call stream of class i in the resource j. The parameter $V_{w,j}$ in the expression $\min(n, V_{w,j})$ determines the maximum possible service stream, expressed in the total number of occupied BBUs. The number of busy BBUs will never exceed the real capacity of the multiservice server.

3.2. Parameters of overflow traffic

Overflow traffic of class i that overflows from the primary resource j will be characterised by the three following parameters: the average value $R_{i,j}$, variance $\sigma^2_{i,j}$ and the volume of demands $t_{i,j}$ (Fig. 1). The average value of this traffic results from the rejection of calls by resources that are in blocking state. The blocking probability of calls can be determined on the basis of Formulas (9) and (10). Thus, the average value $R_{i,j}$ for traffic of class i that overflows from the resource j will be determined by the following formula:

$$R_{i,j} = A_{i,j} \cdot [E_{i,j}]_{V_{\nu,j}}. \tag{11}$$

To determine the variance of overflow traffic we use the approximate method proposed in [30]. According to this method, a decomposition of each of the primary resources $V_{\nu,j}$ into m fictitious resources with the capacity $V^*_{i,j}$ is to be carried out first. The assumption is that each fictitious group will service exclusively calls of one class, which provides an opportunity to apply the Riordan formula [4] to determine the variance of traffic of class i that overflows from the resource j. The capacity of the fictitious resource $V^*_{i,j}$, servicing class i calls only, is determined as such a capacity that leads to the obtaining the blocking probability $E_{i,j}$ for offered traffic $A_{i,j}$:

$$E_{V^*_{i,j}}(A_{i,j}) = E_{i,j}. \tag{12}$$

To accurately determine the value of $V^*_{i,j}$, that can be a non-integer number, an integral form of Erlang formula can be applied [18,53,54].

Now, having the parameters $A_{i,j}$, $R_{i,j}$ and $V^*_{i,j}$ we can determine, on the basis of the Riordan formula, the variance $\sigma^2_{i,j}$ for individual call streams that overflow from the primary resource j to the system of secondary resources:

$$\sigma^2_{i,j} = R_{i,j}\left(\frac{A_{i,j}}{V^*_{i,j} + 1 - A_{i,j} + R_{i,j}} + 1 - R_{i,j}\right). \tag{13}$$

At this point we already have all the parameters that characterize traffic streams offered to the system of secondary resources and we are in position to determine their peakedness coefficients:

$$Z_{i,j} = \sigma^2_{i,j}/R_{i,j}. \tag{14}$$

4. Modelling of secondary resources

4.1. General model

In the present article, a model of the limited-availability group (LAG) will be used to model secondary resources. The system under consideration admits a call for service only when the following two conditions are concurrently satisfied: the system has free resources to service a call of a given class and this call can be entirely serviced by BBUs of one of the resources. In order to present the operation of this system, let us consider a simple limited-availability group composed of two resources. The assumption is that each of the resources has two free BBUs left. The total number of free BBUs is then four. Despite four free BBUs being free, the admission of a call that demands 3 BBUs for service is not possible because there is only two free BBUs in each of the two resources. The presented example shows that the limited-availability group is a model of a system with state-dependent call admission process in which state-dependence results from the structure of the group [1,37].

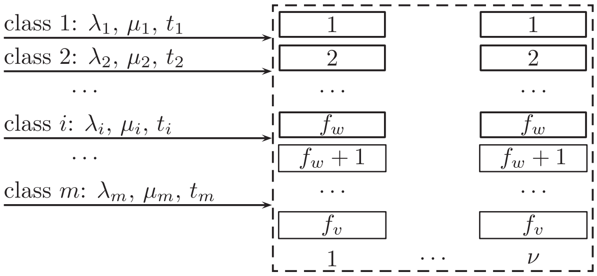


Fig. 3. Basic model of limited-availability group.

LAG models have been analysed in a number of works, notably in [39,41]. Stasiak [41] proposes a model in which all separated resources have identical capacities. This model makes use of a one-dimensional Markov chain to approximate the real service process in LAG. Such an approach provides an opportunity to determine the occupancy distribution in the system on the basis of simple recurrence equations. In [42] this model is generalised for a system composed of separated resources with different capacities, while in [37,43] models which take into account Erlang, Engset and Pascal traffic are proposed.

The basis for the construction of a model of secondary resources will be provided by a modification to the LAG model. Accordingly, in the following two sections (Sections 4.2 and 4.3), first an appropriate basic LAG model with equal (identical) capacities of distributed resources and elastic traffic will be presented, and then a generalised LAG model with different capacities of distributed resources and elastic traffic will be presented.

Thereafter, in Section 4.4, a modified LAG model to which multiservice overflow traffic streams are offered will be discussed. In the construction of this model, elastic traffic and Ershov's remark that refers to single-service systems [55] are used. According to the remark, the Hayward modification [19] can be also applied to state-dependent systems. According to the best knowledge of the authors, the present model that makes it possible to effectively model multiservice state-dependent systems, in particular to model a distributed system of secondary resources of systems with elastic traffic overflow, is here proposed for the first time.

4.2. LAG model with identical capacities of distributed resources

Let us consider now a LAG that is composed of ν identical distributed component resources, each with the real capacity of f_w BBUs and with the virtual capacity of f_v BBUs. The total capacity of the system is $V_{\nu,0} = \nu f_v$ BBUs (Fig. 3).

The occupancy distribution in LAG with the capacity $V_{\nu,0}$ to which a mixture of m Erlang-type traffic is offered can be approximated by the generalised Kaufman-Roberts distribution [41]. Assuming that there is a possibility for LAG to service elastic traffic, i.e. taking into account the virtual resources of the system (Section 2.3), the occupancy distribution will be modified in the same way as for the FAG model modified in [36] for elastic traffic. Thus:

$$[P_n]_{V_{\nu,0}} = \frac{1}{\min(n, V_{w,0})} \sum_{i=1}^{m} A_i t_i \xi_i(n - t_i)[P_{n-t_i}]_{V_{\nu,0}}, \tag{15}$$

where $\xi_i(n)$ is the so-called conditional transition probability that determines which part of the stream of offered traffic of class i is transferred from state n to state $n + t_i$. This coefficient allows the dependence between the call admission stream and the occupancy state of the process to be taken into consideration. The complement of this probability, i.e., the expression $(1 - \xi_i(n))$, is the con-

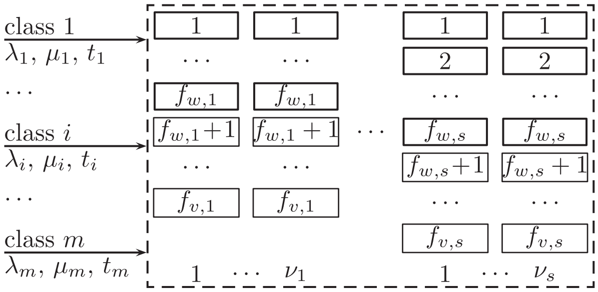


Fig. 4. Generalized model of limited-availability group.

ditional blocking probability of the system in a given occupancy state of n BBUs.

In the case of the basic LAG model (with equal capacities of resources), the conditional transition probability $\xi_i(n)$ is determined in a combinatorial way as follows [41]:

$$\xi_i(n) = 1 - \frac{F(V_{\nu,0} - n, \nu, t_i - 1)}{F(V_{\nu,0} - n, \nu, f_\nu)}. \tag{16}$$

In Formula (16) the combinatorial function $F(x, \nu, f)$ determines the number of arrangements of x free BBUs in ν resources with the capacity of f BBUs each [41]:

$$F(x, \nu, f) = \sum_{u=0}^{\lfloor \frac{x}{f+1} \rfloor} (-1)^u \binom{\nu}{u} \binom{x + \nu - u(f+1) - 1}{\nu - 1}. \tag{17}$$

The construction of Formula (16) is as follows. The numerator in the quotient (16) determines the number of arrangements that are unfavourable for a call of class i in state n to be admitted for service, i.e., determines the number of such arrangements that make finding t_i BBUs in at least one component resource impossible. The denominator (16) determines, in turn, the number of all possible arrangements of free BBUs in state n.

By having the occupancy distribution (15) defined, we are in position to determine the blocking probability for individual multiservice traffic classes [41]:

$$[E_i]_{V_{\nu,0}} = \sum\nolimits_{n=0}^{V_{\nu,0}} \{1 - \xi_i(n)\} [P_n]_{V_{\nu,0}}. \tag{18}$$

Note that in a state-independent system ($\nu = 1,\ V_{\nu,0} = f_\nu$) the conditional transition probability (16) is equal to unity and Formulas (15) and (18) are in their essence Formulas (9) and (10).

4.3. LAG model with differentiated capacities of distributed resources

Let us assume that a LAG is composed of s types of component resources (Fig. 4). Each type k is unequivocally defined by the number ν_k of component resources of a given type and the capacity of each of the component resource: real $f_{w,k}$ and virtual $f_{\nu,k}$ expressed in BBUs. The total real capacity $V_{w,0}$ and virtual capacity $V_{\nu,0}$ of a LAG with elastic traffic and differentiated capacities of distributed component resources can be determined by Formula (4). The occupancy distribution in thus defined system can be approximated by the distribution (15) and (18) in which the conditional transition probability $\xi_i(n)$ is determined by the following formula [1,42]:

$$\xi_i(n) = 1 - \frac{F\{V_{\nu,0} - n, (\nu_1, \ldots, \nu_s), (t_i - 1, \ldots, t_i - 1)\}}{F\{V_{\nu,0} - n, (\nu_1, \ldots, \nu_s), (f_{\nu,1}, \ldots, f_{\nu,s})\}} \tag{19}$$

where the combinatorial function $F\{x, (\nu_1, \ldots, \nu_s), (f_{\nu,1}, \ldots, f_{\nu,s})\}$ determines the number of possible arrangements of x free BBUs in a LAG that is composed of s types of resources that are characterised by the number of component resources ν_k and the virtual capacity of a single component resource $f_{\nu,k}$:

$$F\{x, (\nu_1, \ldots, \nu_s), (f_{\nu,1}, \ldots, f_{\nu,s})\} = \sum_{x_1=0}^{x} \cdots \sum_{x_{s-1}=0}^{x - \sum_{q=1}^{s-2} x_q} \times \left\{ \left[\prod_{k=1}^{s-1} F(x_k, \nu_k, f_{\nu,k}) \right] F\left(x - \sum_{l=1}^{s-1} x_l, \nu_s, f_{\nu,s}\right) \right\}, \tag{20}$$

where the function $F(x,\ \nu,\ f)$ is determined on the basis of the dependence (17). The idea of the construction of Formula (19) is analogous to Formula (16). The numerator in the quotient of Formula (19) determines all unfavourable combinations of the occupancy of the system, i.e., such combinations that do not make it possible to find t_i BBUs in at least one component resource. The denominator determines all possible combinations.

4.4. Model of the system of secondary resources

In order to model the secondary resources system described in Section 2.3 we apply the Hayward approach [19] based on a division of the parameters of the system by the peakedness factor of offered traffic. Such an approach is used in [18] to model multi-service state-independent non-separated secondary resources (i.e. one that consists of a single resource). A separation of secondary resources results in the occurrence of state-dependence. The Hayward approach to model state-dependent systems was proposed for the first time in [55], where a single-service system was considered. Such an approach will be used in the present section to a state-dependent multiservice system, i.e., to a LAG with elastic traffic and with differentiated capacity of separated secondary resources. As a result, applying the notation from Section 3.2, the occupancy distribution (15) and the blocking probability (18), will be written in the following way:

$$[P_n]_{V_{\nu,0}/Z_0} = \frac{1}{\min(n, V_{w,0})} \sum_{j=1}^{r} \sum_{i=1}^{m} \frac{R_{i,j}}{Z_{i,j}} t_{i,j} \xi_{i,j}(n - t_{i,j}) [P_{n-t_{i,j}}]_{V_{\nu,0}/Z_0}, \tag{21}$$

$$\left[E_{i,j}\right]_{V_{\nu,0}/Z_0} = \sum_{n=0}^{V_{\nu,0}/Z_0} \left\{1 - \xi_{i,j}(n)\right\} [P_n]_{V_{\nu,0}/Z_0}, \tag{22}$$

where the transition probability $\xi_{i,\,j}(n)$ is determined on the basis of the modified (19):

$$\xi_{i,j}(n) = 1 - \frac{F\{\frac{V_{\nu,0}}{Z_0} - n, (\nu_1, \ldots, \nu_s), (t_{i,j} - 1, \ldots, t_{i,j} - 1)\}}{F\{\frac{V_{\nu,0}}{Z_0} - n, (\nu_1, \ldots, \nu_s), (\frac{f_{\nu,1}}{Z_0}, \ldots, \frac{f_{\nu,s}}{Z_0})\}}. \tag{23}$$

In Formulas (21)–(23), the peakedness factors $Z_{i,\,j}$ are determined on the basis of (14). The parameter Z_0 is the so-called aggregated peakedness factor. The latter results from the necessity of normalization of the total virtual capacity of the system of secondary resources $V_{\nu,0}/Z_0$ which is offered a mixture of m call streams that overflow from each of r primary resources.

In the case of non-integer values of $f_{\nu,s}/Z_0$ in Eq. (23), calculation of the transition probability $\xi_{i,\,j}(n)$ are performed for the two integer values of the capacity of the secondary resources: $V_{\nu,0,\min}/Z_0 = \sum_{k=1}^{s} \nu_k \lfloor f_{\nu,k} Z_0 \rfloor$ and $V_{\nu,0,\max}/Z_0 = \sum_{k=1}^{s} \nu_k \lceil f_{\nu,k} Z_0 \rceil$. Subsequently, the occupancy distribution in the secondary resources for two integer values of $V_{\nu,0,\min}/Z_0$ and $V_{\nu,0,\max}/Z_0$ can be determined based on Eq. (21). Then, for each of the integer values $V_{\nu,0,\min}/Z_0$ and $V_{\nu,0,\max}/Z_0$, the values of the blocking probability are determined (Formula (22)). When we have the values for blocking probabilities obtained for $V_{\nu,0,\min}/Z_0$ and $V_{\nu,0,\max}/Z_0$, we can apply linear interpolation to determine the value of blocking for the non-integer value $V_{\nu,0}/Z_0$.

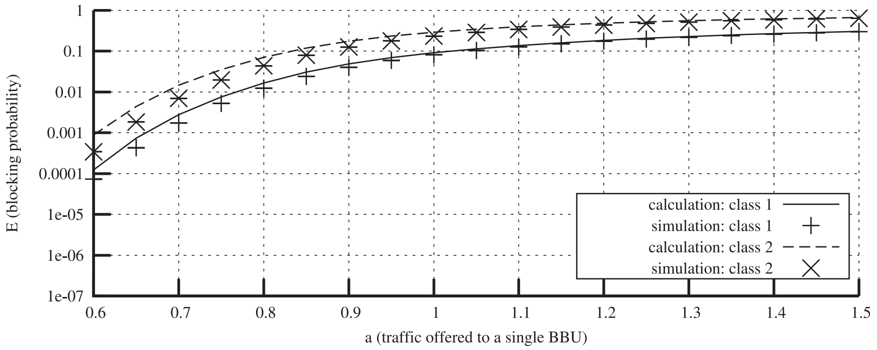


Fig. 5. Blocking probability in System No. 1.

Each of the call streams that overflow from the primary resources is characterised by its own coefficient $Z_{i,\,j}$. In the article – according to the idea proposed in [18] – the aggregate coefficient Z_0 will be determined on the basis of the weighted average of the peakedness coefficients of individual traffic classes $Z_{i,\,j}$. The accompanying assumption is that the weight related to the participation of each of the coefficients $Z_{i,\,j}$ in the aggregated peakedness coefficient Z_0 is directly proportional to the number of BBUs necessary for offered traffic $R_{i,\,j}$ to be serviced:

$$Z_0 = \sum\nolimits_{j=1}^{r} \sum\nolimits_{i=1}^{m} Z_{i,j} \frac{R_{i,j} t_{i,j}}{\sum\nolimits_{u=1}^{r} \sum\nolimits_{l=1}^{m} R_{l,u} t_{l,u}}. \tag{24}$$

Formulas (21)–(23) make it possible to determine all important characteristics of the overflow system with elastic traffic and distribution of secondary resources.

4.5. Algorithm for a determination of traffic characteristics of the overflow system

The models presented in Sections 3.1, 3.2 and 4.4 make it possible to formulate the algorithm for calculation of the blocking probability distribution and other important traffic characteristics of multiservice elastic traffic streams offered to multiservice systems with distributed secondary resources. The algorithm can be presented in the following steps:

- determination of the occupancy distribution (Formula (9)) and the blocking probability (Formula (10)) for call streams of all traffic classes offered to primary resources,
- determination, for all traffic classes, of the parameters of overflow traffic (that overflows from the system of primary resources): the average value $R_{i,\,j}$ (Formula (11)), variance $\sigma^2_{i,j}$ (Formula (13)) and the peakedness coefficient $Z_{i,\,j}$ (Formula (14)),
- determination of the conditional transition probabilities $\xi_{i,\,j}(n)$ for each traffic class of traffic offered to the system of secondary resources (Formulas (19), (20) and (23)),
- determination of the occupancy distribution $[P]_{V_{v,0}/Z_0}$ (Formula (21)), blocking probabilities $[E_{i,j}]_{V_{v,0}/Z_0}$ (Formula (22)) and other characteristics of traffic streams of all classes in the system of distributed secondary resources.

The presented algorithm is characterized by a simple and clear construction that makes its easy implementation possible. A numerical example of modelling an exemplary system according to the presented algorithm is given in Appendix A.

4.6. Comments

The proposed method for the analysis of systems with traffic overflow and distributed resources allows effective modelling at the so-called call (session) level to be introduced. The basic characteristics, obtained on the basis of the developed method, is a determination of the occupancy distribution in analysed systems, i.e. the probability of occupancy of a given amount of resources (Formula (21)). By knowing the determined distribution, and on the basis of Formula (22), we can determine the blocking probabilities $[E_{i,j}]_{V_0/Z_0}$. In addition, it is also possible to determine the loss probability for calls of appropriate classes as a ratio of the number of calls that have entered the system when blocking applies to all calls that have arrived within the observation time of the system. The determined distribution in the system with traffic overflow allows us to also determine the value of the traffic loss probability for individual classes. Because the system is analysed at the call (session) level, direct determination of traffic parameters at the packet level is not possible. However, it should be noted that the possibility of expressing the amount of resources demanded at the packet level as the maximum bit rate or the equivalent bandwidth at the call level allows us to dimension the system (by determining its capacity) in such a way as to include, for example, acceptable packet delay, acceptable packet loss ratio, etc.

The considered model of the systems with overflow traffic directed to distributed secondary resources is an approximate model. The process of dimensioning and determination of the blocking probability in overflow systems is based on the approximation of the Markov process in these systems by an appropriate analytical model composed of accurate, or more frequently approximate, component models of individual elements of the system. Such an approach was already adopted in classical methods for modelling single-service overflow systems (Equivalent Random Method [4] and Hayward Method [19] are approximated methods). In the considered model, the following elements are determined in an approximate way:

1. Conditional transition probabilities for a limited-availability group are determined on the basis of combinatorial dependencies (Formula (20)); this probability is determined in an approximate manner for both traffic with the parameter $Z_0 = 1$, and for $Z_0 \neq 1$;
2. The aggregate coefficient Z_0 is determined on the basis of the weighted average of the peakedness coefficients of individual traffic classes (Formula (24));
3. Variance of overflow traffic, due to the method of decomposition of each of the primary resources $V_{v,j}$ into m fictitious resources with the capacity $V^*_{i,j}$ (Formula (12));
4. The blocking probability calculation in the overflow system, due to the application of linear approximation for non-integer values of V_0/Z_0 (Formula 22).
5. Conditional transition probabilities for a limited-availability group are determined for two nearest integer values of $f_{v,s}/Z_0$, i.e., $\lfloor f_{v,s}/Z_0 \rfloor$ and $\lceil f_{v,s}/Z_0 \rceil$ (Formula 23).

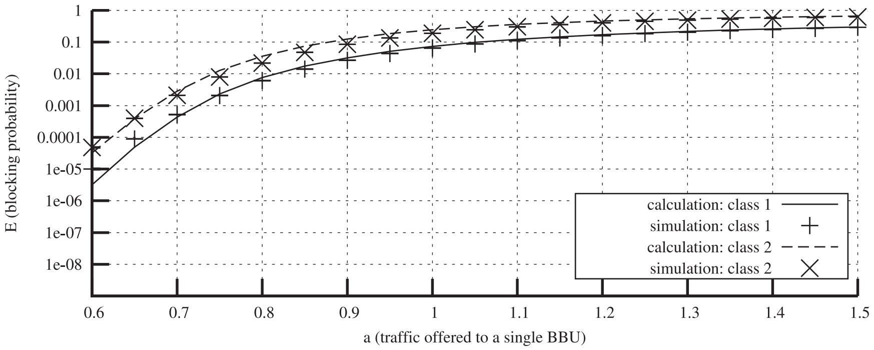


Fig. 6. Blocking probability in System No. 2.

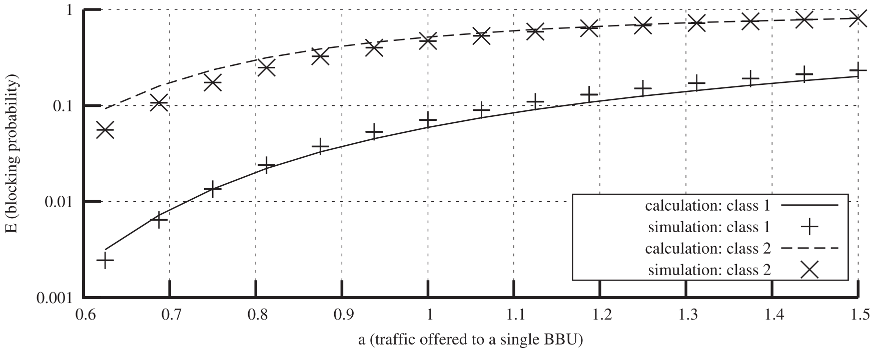


Fig. 7. Blocking probability in System No. 3.

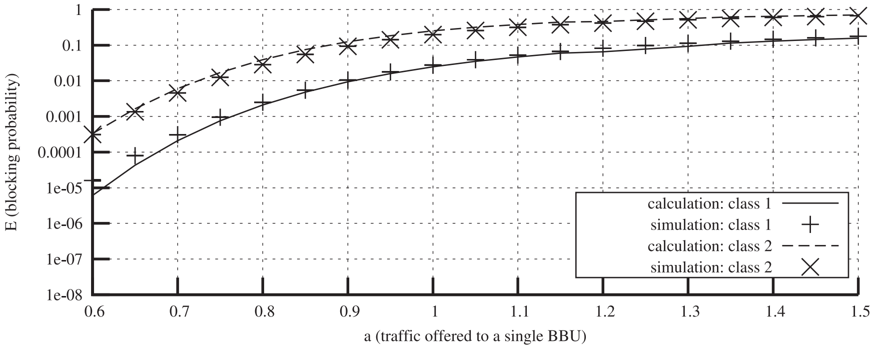


Fig. 8. Blocking probability in System No. 4.

5. A study on overflow systems with distributed secondary resources

To evaluate the accuracy of the proposed model the results of the analytical calculations were compared with the data provided by the simulation experiments. For this particular purpose, a dedicated simulator of the considered networks at the call level was constructed [56]. The approach to modelling of multiservice systems with traffic overflow at the call level adopted in this article makes it possible to use it to determine traffic characteristics, e.g. hierarchical wireless networks [1], distributed data centres [2], hybrid optical networks [17], or the so-called parallel Internet [57].

The study involved systems consisting of primary resources (modelled by FAGs) and a distributed secondary resource. The calculations and simulation results presented in the article were obtained for the systems that are summarized in Table 1. Table 1 also presents the structure and nature of offered traffic for each of the primary resources. The assumption was that the proportion of offered traffic $A_{1,1}t_{1,1} : A_{2,1}t_{2,1} : \ldots : A_{m,r}t_{m,r}$ was equal to $1 : 1 : \ldots : 1$. If the virtual capacity is the same as the real capacity ($V_{w,j} = V_{v,j}$), the virtual capacity is denoted as "——" in Table 1.

The results of the calculations and simulations obtained for the systems with the parameters presented in Table 1 are shown in Figs. 5–10. The results are expressed in dependence on the value of traffic a offered to one BBU of the entire overflow system:

$$a = \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{r} A_{i,j}t_{i,j} \Big/ \left(\sum\nolimits_{k=1}^{s} \nu_k f_{w,k} + \sum\nolimits_{j=1}^{r} V_{w,j} \right). \tag{25}$$

The results of the analytical calculations were compared with the data obtained in digital simulation that took into account a random algorithm for secondary resources selection for call service. The results of the simulation are presented in Figs. 5–10 in the form of appropriately marked points with 99% confidence interval calculated after the t-Student distribution for the number of ten series. The results presented in Figs. 5–10 allow us to determine the influence of various parameters of the considered systems on accuracy of the analytical model, i.e., the structure of primary resources, the structure of secondary resources, the number of offered traffic classes, the value of required resources for calls of particular classes and the influence of elastic traffic (compres-

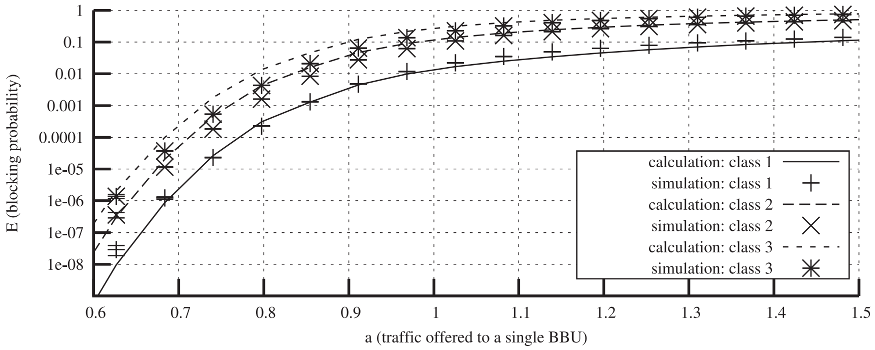


Fig. 9. Blocking probability in System No. 5.

Table 1
Analysed overflow systems.

	Primary			Offered traffic						Secondary			
No.	j	$V_{w,j}$	$V_{v,j}$	$t_{1,j}$	$\mu_{1,j}$	$t_{2,j}$	$\mu_{2,j}$	$t_{3,j}$	$\mu_{3,j}$	k	ν_k	$f_{w,k}$	$f_{v,k}$
1	1	16	−−	1	1	2	1			1	3	16	−−
	2	16	−−	1	1	2	1						
	3	16	−−	1	1	2	1						
2	1	16	24	1	1	2	1			1	3	16	−−
	2	16	24	1	1	2	1						
	3	16	24	1	1	2	1						
3	1	10	−−	1	1	3	1			1	2	10	−−
	2	10	−−	1	1	3	1			2	2	5	−−
	3	10	−−	1	1	3	1						
	4	10	−−	1	1	3	1						
	5	10	−−	1	1	3	1						
4	1	10	15	1	1	3	1			1	2	10	15
	2	10	15	1	1	3	1			2	2	5	8
	3	10	15	1	1	3	1						
	4	10	15	1	1	3	1						
	5	10	15	1	1	3	1						
5	1	20	−−	1	1	2	1	3	1	1	4	8	12
	2	30	−−	1	1	2	1	3	1	2	3	12	18
	3	40	−−	1	1	2	1	3	1				
6	1	30	60	1	1	3	1	8	1	1	4	20	30
	2	30	60	1	1	3	1	8	1	2	3	15	20
	3	30	60	1	1	3	1	8	1				

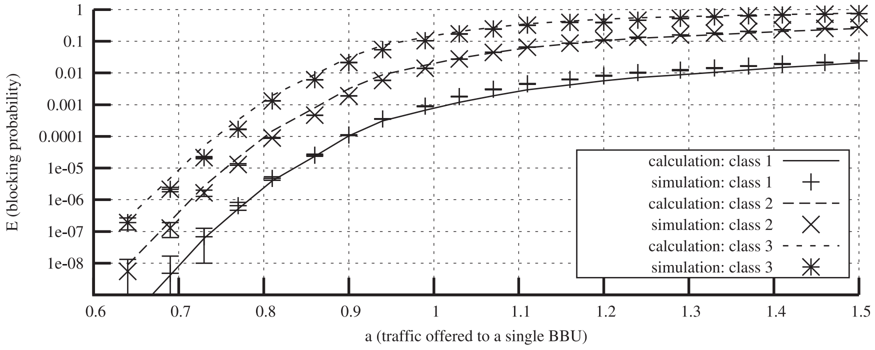


Fig. 10. Blocking probability in System No. 6.

sion). On the basis of the obtained results and taken into account the results presented in [18] and [42] it can be stated that the accuracy of the model does not deviate much from: (1) the accuracy achieved by models without distributed secondary resources and without elastic traffic, and (2) is comparable with the accuracy of the LAG model for Erlang traffic.

The presented method for a determination of traffic characteristics for multiservice overflow systems with distributed resources is an approximated method. However, it should be stressed that the characteristics of even the most simple single-service overflow systems are determined on the basis of approximate methods in which the accuracy is comparable to that obtained by the model proposed in the article. According to authors' best knowledge, the model proposed in the article is the only model that describes such a complex overflow system. Therefore, the accuracy obtained in the study, comparable to the accuracy of overflow models without distributed resources and without elastic traffic, affirms the virtue and viability of the proposed solution.

6. Summary

This article proposes a model of a multiservice overflow system with distributed secondary resources servicing overflow traffic. The model proposes a new approach to the analysis of complex state-dependent multiservice systems that are offered a mixture of elastic overflow traffic with peakedness factors higher than unity. The approach consists in taking into account a possibility of servicing elastic traffic both in primary and secondary resources and is based on the application of the Hayward approach to modelling of state-dependent systems, in particular on a division of the value of the traffic intensity and the structural parameters of the system by the peakedness factors of corresponding classes of offered traffic. Such an approach is used in the article to model a system of distributed secondary resources. The results of the comparison of the analytical calculations with the results of the simulation experiments indicate high accuracy of the proposed model. The proposed model provides an opportunity for further research in the field, in particular one that is focused on modelling of other, complex, state-dependent multiservice systems to which elastic overflow traffic is offered. The focus of our future studies will be to develop an analytical model of an overflow system with distributed secondary resources that services independent traffic streams with any kind of distribution.

Appendix A. Numerical example

The section contains the numerical results of the calculation process according to the proposed method of analytical modelling of overflow systems, for System No. 4. The presented results were obtained for the value of traffic $a = 0.8$ Erlang offered per single BBU of the entire overflow system. The system consists of three primary resources with the same structure. Consequently, the following results are presented for one primary resource, denoted by x, where $x = 1, 2, 3$. The primary resource x with the real capacity $f_{w,x} = 10$ BBUs and the virtual capacity $f_{v,x} = 15$ BBUs is offered two traffic classes with the following parameters: $A_{1,x} = 8$ Erlang, $t_{1,x} = 1$ BBU, $A_{2,x} = 2.66(6)$ Erlang, $t_{2,x} = 3$ BBUs.

The calculation process is as follows.

1. Calculation of the occupancy distribution in the primary resource x (Formula (9)):

$$\{[P_n]_{15}\} = \{3.290966769729E-05,\ 0.000263277345,\ 0.001053109397,\ 0.002896050884,\ 0.006318656553,\ 0.011794825697,\ 0.019587835734,\ 0.029607420199,\ 0.041402246513,\ 0.054213407250,\ 0.067056662959,\ 0.086767128871,\ 0.112784430578,\ 0.143872876973,\ 0.184512007425,\ 0.237837153947\}$$

2. Calculation of the blocking probability E (Formula (18)) and the value of overflow traffic R (Formula (11)) in primary group x:
   Class 1: $E_{1,x} = 0.238,\ R_{1,x} = 1.903$
   Class 2: $E_{2,x} = 0.144 + 0.184 + 0.238 = 0.566,\ R_{2,x} = 1.510$
3. Calculation of the capacity V^* of the fictitious resources (Formula (12)):
   $V^*_{1,x} = 7.966$
   $V^*_{2,x} = 1.663$
4. Calculation of the variance of the overflow traffic (Formula (13)):
   $\sigma^2_{1,x} = 3.588$
   $\sigma^2_{2,x} = 1.902$
5. Determination of the peakedness coefficients (Formula (14)):
   $Z_{1,x} = 1.886$
   $Z_{2,x} = 1.260$
6. Calculation of the aggregated peakedness coefficient Z_0 (Formula (24)):
   $Z_0 = 0.099 \cdot 1.886 + 0,235 \cdot 1.260 + 0.099 \cdot 1.886 + 0,235 \cdot 1.260 + 0.099 \cdot 1.886 + 0,235 \cdot 1.260 = 1.445$
7. Determination of the parameters of the secondary resources:
   $f_{v,1,\min} = \left\lfloor \frac{f_{v,1}}{Z_0} \right\rfloor = 10,\ f_{v,2,\min} = \left\lfloor \frac{f_{v,2}}{Z_0} \right\rfloor = 5$
   $V_{v,0,\min} = \sum_{j=1}^{2} \nu_j f_{v,j,\min} = 30,\ V_{w,0,\min} = \sum_{j=1}^{2} \nu_j f_{w,j,\min} = 18$
   $f_{v,1,\max} = \left\lceil \frac{f_{v,1}}{Z_0} \right\rceil = 11,\ f_{v,2,\max} = \left\lceil \frac{f_{v,2}}{Z_0} \right\rceil = 6$
   $V_{v,0,\max} = \sum_{j=1}^{2} \nu_j \left\lceil \frac{f_{v,j}}{Z_0} \right\rceil = 34,\ V_{w,0,\max} = \sum_{j=1}^{2} \nu_j \left\lceil \frac{f_{w,j}}{Z_0} \right\rceil = 20$
8. Calculation of the conditional transition probability $\xi_i(n)$ for the secondary resources withe following parameters: $V_{v,0,\min} = 30$, $V_{w,0,\min} = 18,\ \nu_1 = 2,\ f_{v,1,\min} = 10,\ \nu_2 = 2,\ f_{v,2,\min} = 5$:

   $$\forall_{0 \le n \le 21} \xi_i(n) = 1$$

n	$\xi_1(n)$	$\xi_2(n)$	n	$\xi_1(n)$	$\xi_2(n)$	n	$\xi_1(n)$	$\xi_2(n)$
22	1	0.9931	25	1	0.7143	28	1	0
23	1	0.9643	26	1	0.4571	29	1	0
24	1	0.878	27	1	0.2	30	0	0

9. Calculation of the occupancy distribution in the secondary resources with the capacity $V_{v,0,\min} = 30$ (Formula (21)):

$$\{[P_n]_{30}\} = \{\ 0.001240,\ 0.003754,\ 0.005682,\ 0.010192,\ 0.017836,\ 0.023056,\ 0.029955,\ 0.040439,\ 0.046389,\ 0.051505,\ 0.059212,\ 0.061785,\ 0.061884,\ 0.063542,\ 0.061344,\ 0.056882,\ 0.053600,\ 0.048468,\ 0.042239,\ 0.039225,\ 0.035642,\ 0.031306,\ 0.028771,\ 0.026198,\ 0.023167,\ 0.021019,\ 0.018674,\ 0.015331,\ 0.011575,\ 0.007062,\ 0.003025\}$$

10. Calculation of the conditional transition probability $\xi_i(n)$ for the secondary resources withe following parameters: $V_{v,0,\max} = 34$, $V_{w,0,\max} = 20,\ \nu_1 = 2,\ f_{v,1,\max} = 11,\ \nu_2 = 2,\ f_{v,2,\max} = 6$:

    $$\forall_{0 \le n \le 25} \xi_i(n) = 1$$

n	$\xi_1(n)$	$\xi_2(n)$	n	$\xi_1(n)$	$\xi_2(n)$	n	$\xi_1(n)$	$\xi_2(n)$
26	1	0.9936	29	1	0.7143	32	1	0
27	1	0.9661	30	1	0.4571	33	1	0
28	1	0.881	31	1	0.2	34	0	0

11. Calculation of the occupancy distribution in the secondary resources with the capacity $V_{v,0,\max} = 34$ (Formula (21)):

$$\{[P_n]_{34}\} = \{\ 0.001265,\ 0.003830,\ 0.005796,\ 0.010397,\ 0.018196,\ 0.023520,\ 0.030559,\ 0.041254,\ 0.047323,\ 0.052543,\ 0.060405,\ 0.063030,\ 0.063131,\ 0.064822,\ 0.062580,\ 0.058028,\ 0.054680,\ 0.049445,\ 0.043090,\ 0.037909,\ 0.032406,\ 0.028145,\ 0.024706,\ 0.021217,\ 0.018391,\ 0.016108,\ 0.013881,\ 0.012020,\ 0.010507,\ 0.009030,\ 0.007630,\ 0.006147,\ 0.004409,\ 0.002548,\ 0.001049\}$$

12. Calculation of the blocking probability in the secondary resources (Formula (22)):
    - Blocking probability in the secondary resources with the capacity $\frac{V_{v,0,\min}}{Z_0}$:

      $E_{1,x,\min} = 0.003025,\ E_{2,x,\min} = 0.054029$

- Blocking probability in the secondary resources with the capacity $\frac{V_{v,0,\max}}{Z_0}$:

  $$E_{1,x,\max} = 0.001049, \; E_{2,x,\max} = 0.021393$$

13. Calculation of the final blocking probability using linear approximation for non-integer values of V_0/Z_0:

$$E_{i,x} = E_{i,x,\min} + \frac{(E_{i,x,\max} - E_{i,x,\min})(V_{v,0}/Z_0 - V_{v,0,\min})}{V_{v,0,\max} - V_{v,0,\min}}$$

$$E_{1,x} = 0.003025 + (0.001049 - 0.003025)(31.83524 - 30)/(34 - 30) = 0.002118$$

$$E_{2,x} = 0.054029 + (0.021393 - 0.054029)(31.83524 - 30)/(34 - 30) = 0.039055$$

14. Simulation results:

$$E_{1,x} = 0.002399 \pm 0.000022$$

$$E_{2,x} = 0.0033964 \pm 0.000130$$

References

[1] M. Stasiak, M. Głąbowski, A. Wiśniewski, P. Zwierzykowski, Modeling and Dimensioning of Mobile Networks, Wiley, 2011.

[2] P. Kühn, M.E. Mashaly, M. Fiedler, Multi-server, finite capacity queuing system with mutual overflow, in: Proceedings of 2nd European Teletraffic Seminar, Karlskrona, 2013. N/a–n/a

[3] M. Wang, S. Li, E. Wong, M. Zukerman, Performance analysis of circuit switched multi-service multi-rate networks with alternative routing, J. Lightwave Technol. 32 (2) (2014) 179–200, doi:10.1109/JLT.2013.2289925.

[4] R. Wilkinson, Theories of toll traffic engineering in the USA, Bell Syst. Tech. J. 40 (1956) 421–514.

[5] M. Głąbowski, S. Hanczewski, M. Stasiak, Erlang's ideal grading in diffserv modelling, in: Proceedings of IEEE Africon 2011, IEEE, Livingstone, Zambia, 2011, pp. 1–6, doi:10.1109/AFRCON.2011.6072139.

[6] E.W. Wöng, J. Guö, B. Möran, M. Zukerman, Information exchange surrogates for approximation of blocking probabilities in overflow loss systems, in: Proceedings of 25th International Teletraffic Congress, 2013, pp. 1–9. filead-min/ITCBibDatabase/2013/wong2013information.pdf

[7] J. Matsumoto, Y. Watanabe, Theoretical method for the analysis of queueing system with overflow traffic, Electron. Commun. Japan (Part I: Commun.) 64 (6) (1981) 74–83, doi:10.1002/ecja.4410640610.

[8] Y.C. Chan, J. Guo, E. Wong, M. Zukerman, Performance analysis for overflow loss systems of processor-sharing queues, in: 2015 IEEE Conference on Computer Communications (INFOCOM), 2015, pp. 1409–1417, doi:10.1109/INFOCOM.2015.7218518.

[9] M. Mashaly, P.J. Kühn, Load balancing in cloud-based content delivery networks using adaptive server activation/deactivation, in: Proceedings of the 24th International Teletraffic Congress, ITC '12, International Teletraffic Congress, 2012. 21:1–21:3. http://dl.acm.org/citation.cfm?id=2414276.2414302.

[10] B.M. Bakmaz, M.R. Bakmaz, Solving some overflow traffic models with changed serving intensities, AEU - Int. J. Electron.Commun. 66 (1) (2012) 80–85, doi:10.1016/j.aeue.2011.05.007.

[11] G. Soni, M. Kalra, A novel approach for load balancing in cloud data center, in: Advance Computing Conference (IACC), 2014 IEEE International, 2014, pp. 807–812, doi:10.1109/IAdCC.2014.6779427.

[12] M. Yoshino, N. Nishibe, M. Oba, N. Komoda, Classification of energy-saving operations from the perspective of system management, in: 8th IEEE International Conference on Industrial Informatics (INDIN), 2010, pp. 651–656, doi:10.1109/INDIN.2010.5549663.

[13] I. Moscholios, M. Logothetis, A. Boucouvalas, Blocking probabilities of elastic and adaptive calls in the Erlang multirate loss model under the threshold policy, Telecommun. Syst. (2016), doi:10.1007/s11235-015-0056-z.

[14] R. Deng, Z. Yang, M. Chow, J. Chen, A survey on demand response in smart grids: mathematical models and approaches, IEEE Trans. Ind. Inf. 11 (3) (2015) 570–582, doi:10.1109/TII.2015.2414719.

[15] A.I. Sabbah, A. El-Mougy, M. Ibnkahla, A survey of networking challenges and routing protocols in smart grids, IEEE Trans. Ind. Inf. 10 (1) (2014) 210–221, doi:10.1109/TII.2013.2258930.

[16] S. Fernandes, A. Karmouch, Vertical mobility management architectures in wireless networks: a comprehensive survey and future directions, Commun. Surv. Tut. IEEE 14 (1) (2012) 45–63, doi:10.1109/SURV.2011.082010.00099.

[17] C. Gauger, P. Kühn, E. Breusegem, M. Pickavet, P. Demeester, Hybrid optical network architectures: bringing packets and circuits together, IEEE Commun. Mag. 44 (8) (2006) 36–42, doi:10.1109/MCOM.2006.1678107.

[18] M. Głąbowski, K. Kubasik, M. Stasiak, Modeling of systems with overflow multi-rate traffic, Telecommun. Syst. 37 (1–3) (2008) 85–96, doi:10.1007/s11235-008-9070-8.

[19] A. Fredericks, Congestion in blocking systems – a simple approximation technique, Bell Syst. Tech. J. 59 (6) (1980) 805–827. http://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1980.tb03034.x/abstract.

[20] R. Schehrer, On the calculation of overflow systems with a finite number of sources and full availiable groups, IEEE Trans. Commun. 26 (1) (1978) 75–82.

[21] E. Wong, A. Zalesky, M. Zukerman, On generalizations of the Engset model, IEEE Commun. Lett. 11 (4) (2007) 360–362, doi:10.1109/LCOM.2007.348301.

[22] C. McArdle, D. Tafani, L. Barry, Overflow traffic moments in channel groups with bernoulli-poisson-pascal (BPP) load, in: 2013 IEEE International Conference on Communications (ICC), 2013, pp. 2403–2408, doi:10.1109/ICC.2013.6654891.

[23] I.D. Moscholios, M.D. Logothetis, J.S. Vardakas, A.C. Boucouvalas, Congestion probabilities of elastic and adaptive calls in Erlang-engset multirate loss models under the threshold and bandwidth reservation policies, Comput. Netw. 92 (Part 1) (2015) 1–23, doi:10.1016/j.comnet.2015.09.010.

[24] S.P. Chung, J.C. Lee, Performance analysis and overflowed traffic characterization in multiservice hierarchical wireless networks, IEEE Trans. Wireless Commun. 4 (3) (2005) 904–918, doi:10.1109/TWC.2005.847031.

[25] J.S. Kaufman, K.M. Rege, Blocking in a shared resource environment with batched poisson arrival processes, J. Perform. Eval. 24 (4) (1996) 249–263, doi:10.1016/0166-5316(94)00029-8.

[26] I.D. Moscholios, J.S. Vardakas, M.D. Logothetis, A.C. Boucouvalas, Congestion probabilities in a batched poisson multirate loss model supporting elastic and adaptive traffic, Ann. Telecommun. 68 (5) (2013) 327–344.

[27] L. Delbrouck, On the steady-state distribution in a service facility carrying mixtures of traffic with different peakedness factors and capacity requirements, IEEE Trans. Commun. 31 (11) (1983) 1209–1211.

[28] M. Głąbowski, Modeling systems with multi-service overflow Erlang and engset traffic streams, Int. J. Adv.Telecommun. 1 (1) (2008) 14–26. issn 1942–2601

[29] M. Głąbowski, A. Kaliszan, M. Stasiak, Two-dimensional convolution algorithm for modelling multiservice networks with overflow traffic, Math. Prob. Eng. 2013 (18) (2013), doi:10.1155/2013/852082. Article ID 852082.

[30] Q. Huang, K.T. Ko, V.B. Iversen, An approximation method for multiservice loss performance in hierarchical networks, in: L. Mason, T. Drwiega, J. Yan (Eds.), Managing Traffic Performance in Converged Networks, 20th International Teletraffic Congress, ITC20 2007, Lecture Notes in Computer Science, vol. 4516, Springer, 2007, pp. 901–912, doi:10.1007/978-3-540-72990-7_78.

[31] V. Iversen, Teletraffic engineering handbook, ITU-D SG 2/16 and ITC Draft, 2001.

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

[33] J. Roberts, G. Pujolle, A service system with heterogeneous user requirements—application to multi-service telecommunications systems, in: Proceedings of Performance of Data Communications Systems and their Applications, North Holland, Amsterdam, 1981, pp. 423–431.

[34] M. Głąbowski, M. Stasiak, J. Weissenberg, Properties of recurrent equations for the full-availability group with BPP traffic, Math. Prob. Eng. 2012 (17) (2012), doi:10.1155/2012/547909. Article ID 547909

[35] J. Postel, Transmission control protocol. RFC 793 (INTERNET STANDARD), updated by RFCs 1122, 3168, 6093, 6528 (Sep. 1981). http://www.ietf.org/rfc/rfc793.txt.

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

[37] M. Głąbowski, A. Kaliszan, M. Stasiak, Modeling product-form state-dependent systems with BPP traffic, Perform. Eval. 67 (2010) 174–197, doi:10.1016/j.peva.2009.10.002.

[38] L. Katzschner, R. Scheller, Probability of loss of data traffics with different bit rates hunting one common PCM-channel, in: Proceedings of 8th International Teletraffic Congress, Melbourne, 1976. 525/1–8

[39] J. Conradt, A. Buchheister, Considerations on loss probability of multi-slot connections, in: Proceedings of 11th International Teletraffic Congress, Kyoto, 1985. 4.4B–2.1

[40] J. Karlsson, Loss performance in trunk groups with different capacity demands, in: Proceedings of 13th International Teletraffic Congress, Vol. Discussion Circles, Copenhagen, 1991, pp. 201–212.

[41] M. Stasiak, Blocking probability in a limited-availability group carrying mixture of different multichannel traffic streams, Annales des Télécommunications 48 (1–2) (1993) 71–76.

[42] M. Głąbowski, M. Stasiak, Multi-rate model of the group of separated transmission links of various capacities, in: J. de Souza, P. Dini, P. Lorenz (Eds.), Telecommunications and Networking - ICT 2004, Lecture Notes in Computer Science, vol. 3124, Springer Berlin Heidelberg, 2004, pp. 1101–1106, doi:10.1007/978-3-540-27824-5_143.

[43] M. Głąbowski, Modelling of state-dependent multi-rate systems carrying BPP traffic, Ann. Telecommun. 63 (7–8) (2008) 393–407, doi:10.1007/s12243-008-0034-5.

[44] P. Mrozowski, A. Chydzinski, Performance evaluation of routers with the dropping-function queueing, in: A. Kwiecien, P. Gaj, P. Stera (Eds.), Computer Networks, Communications in Computer and Information Science, vol. 431, Springer International Publishing, 2014, pp. 243–252, doi:10.1007/978-3-319-07941-7_25.

[45] D. Parniewicz, M. Stasiak, P. Zwierzykowski, Traffic engineering for multicast connections in multiservice cellular networks, IEEE Trans. Ind. Inf. 9 (1) (2013) 262–270, doi:10.1109/TII.2012.2188902.

[46] A. Chydzinski, L. Chrost, Analysis of AQM queues with queue size based packet dropping, Appl. Math. Comput. Sci. 21 (3) (2011) 567–577, doi:10.2478/v10006-011-0045-7.

[47] D. Parniewicz, M. Stasiak, J. Wiewióra, P. Zwierzykowski, An approximate model of the WCDMA interface servicing a mixture of multi-rate traffic streams with priorities, in: N. Thomas, C. Juiz (Eds.), Computer Performance Engineering, Lecture Notes in Computer Science, vol. 5261, Springer Berlin Heidelberg, 2008, pp. 168–180.

[48] S. Floyd, V. Paxson, Difficulties in simulating the internet, in: IEEE/ACM Trans. Netw., 9, 2001, pp. 392–403, doi:10.1109/90.944338.

[49] T. Bonald, J. Roberts, Internet and the Erlang formula, in: ACM Computer Communications Review, vol. 42, 2012, pp. 23–30.

[50] J.Y. Hui, Resource allocation in broadband networks, J. Sel. Areas Commun. 6 (9) (1988) 1598–1608.

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

[52] J. Roberts, Performance evaluation and design of multiservice networks, Final Report COST 224, Commission of the European Communities, Brussels, 1992.

[53] R. Fortet, B. Canceill, Probalitès de perte en selection conjuguée, in: Teleteknik 1, 1957.

[54] R. Syski, Introduction to congestion theory in telephone systems, in: Studies in Telecommunication, North Holland, 1986.

[55] E. Ershova, V. Ershov, Digital systems for information distribution, Radio and Communications, Moscow in Russian, 1983.

[56] M. Głąbowski, A. Kaliszan, Simulator of full-availability group with bandwidth reservation and multi-rate Bernoulli-Poisson-Pascal traffic streams, in: Proceedings of Eurocon 2007, Warszawa, 2007, pp. 2271–2277, doi:10.1109/EURCON.2007.4400605.

[57] A. Kaliszan, S. Hanczewski, M. Głąbowski, M. Stasiak, P. Zwierzykowski, Routing and control plane in the parallel internet IPv6 qos, in: Proceedings of the 8th IEEE, IET International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP 2012), Poznañ, Poland, 2012, doi:10.1109/CSNDSP.2012.6292770.

[58] M. Głąbowski, S. Hanczewski, M. Stasiak, Modelling of cellular networks with traffic overflow, Math. Prob. Eng. 2015 (2015). Article ID 286490. [Online]. doi: 10.1155/2015/286490.

Mariusz Głąbowski received the M. Sc., Ph. D. and D. Sc. (Habilitation) degrees in telecommunication from the Poznan University of Technology, Poland, in 1997, 2001, and 2010, respectively. Since 1997 he has been working in the Department of Electronics and Telecommunications, Poznan University of Technology. He is engaged in research and teaching in the area of performance analysis and modeling of multiservice networks and switching systems. Prof. Mariusz Głąbowski is the author/co-author of 4 books, 7 book chapters and of over 100 papers which have been published in communication journals and presented at national and international conferences. He has refereed articles for many international conferences and magazines, including: IEEE Globecom, IEEE ICC, IEEE HPRS, IEEE Transactions on Communications, IEEE Communications Magazine, Computer Networks, IEEE Communications Letters, IEEE Transactions on Wireless Communications, Performance Evaluation, European Transactions on Telecommunications.

Adam Kaliszan received the M. Sc., Ph. D. degrees in telecommunication from the Poznan University of Technology (PUT), Poland, in 2005, 2010, respectively. Since 2008 Adam Kaliszan has been working in the Department of Electronics and Telecommunications, PUT. He is engaged in research and teaching in the area of programming and operating systems. He works also as developer in research projects. Adam Kaliszan is the author/co-author of 30 papers which have been published in communication journals and presented at national and international conferences. Adam Kaliszan is teaching in Poznan University of Technology largely focuses on the M.Sc.in Telecommunication where he gives lectures and conducts laboratory courses on Operating Systems, Network Embedded Systems, Programming in C ++ and C#. Adam Kaliszan also participates in industrial education acting as a lecturer in courses on programming.

Maciej Stasiak received M.Sc. and Ph.D. degrees in electrical engineering from the Institute of Communications Engineering, Moscow, Russia in 1979 and 1984, respectively. In 1996 he received D.Sc. degree in electrical engineering from Poznan University of Technology. In 2006 he was nominated as full professor. Between 1983–1992 he worked in Polish industry as a designer of electronic and microprocessor systems. In 1992, he joined Poznan University of Technology, where he is currently Head of the Chair of Communications and Computer Networks at the Faculty of Electronics and Telecommunications.

He is the author or co-author of more than 250 scientific papers and five books. He is engaged in research and teaching in the area of performance analysis and modelling of multiservice networks and switching systems, in particular resource allocation, call admission control, quality of service, multi-rate systems and queuing systems. Since 2004, he has been actively carrying out research on modeling and dimensioning cellular networks.