Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7870164, 17 pages
https://doi.org/10.1155/2018/7870164

Research Article

Modelling of Multiservice Networks with Separated Resources and Overflow of Adaptive Traffic

Mariusz Głąbowski , Damian Kmiecik , and Maciej Stasiak

Poznań University of Technology, Polanka 3, 60-965 Poznań, Poland

Correspondence should be addressed to Mariusz Głąbowski; mariusz.glabowski@put.poznan.pl

Received 30 March 2018; Revised 29 June 2018; Accepted 10 July 2018; Published 14 August 2018

Academic Editor: Dejan Vukobratovic

The article proposes a new method of determining traffic characteristics of multiservice overflow systems that carry adaptive traffic. When the total offered load in primary resources exceeds a certain value, this type of traffic is admitted for service with lower bitrate. A particular attention is given in the article to a method for a determination of the parameters of traffic that overflows to secondary resources as well as to the way adaptive traffic is serviced. The method takes into consideration three possible types of traffic: Erlang, Engset, and Pascal traffic. It is based on a generalization of Hayward's concept and its application to model systems with adaptive traffic with threshold compression. The method can be used for optimal dimensioning of logical networks (slices) in modern mobile systems due to possibility of analytical determination of grade of service parameters (blocking probability, carried traffic, and network load). To verify the accuracy of the proposed model the results of analytical calculations, obtained on the basis of the proposed model, are then compared with the results of simulation experiments for a number of selected structures of overflow systems that service adaptive traffic. The results of the study demonstrate high accuracy of the proposed theoretical model.

1. Introduction

The mechanism of traffic overflow has been used in telecommunications networks for over seventy years as one of the oldest optimization techniques for traffic distribution. The overflow mechanism is initiated when the resources of a given system (the so-called primary resources (PR)) are occupied and, in consequence, unavailable for new calls. These calls, however, can be admitted for service by other systems that have free resources (the so-called secondary resources (SR)) available to handle new calls [1]. In the 1950s, the introduction of cross-connect switching systems to telecommunications networks allowed overflow mechanisms to be applied in network engineering. These mechanisms for single-service networks have been thoroughly described and analyzed in such classic works as [2–12]. In [2, 3], the Equivalent Random Traffic (ERT) method is developed for a determination of the blocking probability in single-service overflow systems. This method takes advantage of the first two moments of traffic that overflows from PR to SR, i.e., the mean value and variance. This, in turn, provides a basis for the so-called equivalent resources to be defined. Equivalent resources allow then the blocking probability in an overflow system to be directly determined. An interesting method for a description of single-service overflow systems is proposed in [10]. The method is based on a modification of Erlang's formula, in which the capacity of secondary resources and the value of traffic offered to the overflow group are divided by the peakedness factor of overflow traffic, i.e., by the ratio of the variance of overflow traffic to its mean value. Such an approach has significantly simplified the way the blocking probability in overflow systems can be determined as compared to the ERT method. The works [5, 13, 14] point out the possibility of an approximation of the call stream that overflows from the primary resources in a call stream of Pascal type. In [15, 16] systems with mutual traffic overflow are considered and discussed.

A commercial application of the first multiservice Integrated Services Digital Network (ISDN) in continuously carried out until now within the context of new technologies and network standards that are being successively introduced. The works [17–19] propose a description of overflow traffic with Markov-Modulated Poisson Process (MMPP processes), whereas in [20, 21] the Batched Poisson Process (BPP) is used. The authors of [20, 22, 23] consider a possibility of an application of traffic stream with an assumed value of the peakedness factor. The papers [24–26] propose a number of engineering methods for dimensioning multi-service systems with overflow traffic. These methods are based on an application of the approach [10] with regard to multiservice full-availability resources (FAR full availability resources), the so-called full-availability groups (FAG) [1, 27]. The full-availability approach means that a new call will be admitted for service if the system has sufficient capacity (resources) necessary for this call to be serviced. FAR with offered mixture of Erlang, Engset, and Pascal traffic streams with varied bitrates can be described on the basis of multiservice Markov processes that, when solved, provide simple recurrence formulas [28, 29] (for a mixture of Erlang traffic) and [30, 31] (for a mixture of Erlang, Engset and Pascal traffic). Dividing the values of traffic of individual classes offered to the secondary resources by the corresponding peakedness factors makes it possible to model multiservice SR and, in consequence, the blocking probability of an overflow system. In [32], to model overflow systems a two-dimensional convolution model is proposed, while in [33] the overflow system is approximated by ideal nonfull availability resources (called Erlang Ideal Grading (EIG) in the literature of the subject [34]). In models [35–37] overflow systems are considered in which overflow traffic changes the service parameters in the secondary resources, such as the service time and bitrate. In the study, a change in the parameters is not, however, related to network mechanisms for traffic shaping.

Present-day multiservice networks, including 4G and 5G mobile networks, are packet networks (based on IP (Internet protocol) at Internet Layer and TCP (Transmission Control Protocol) at transportation layer in which packet streams undergo different traffic shaping mechanisms. One of the most widely used mechanisms used in TCP/IP networks, both wired and mobile, are the mechanisms of threshold and thresholdless traffic compression. Traffic that is subject to the operation of these mechanisms is called elastic [38] (based on TCP) or adaptive traffic [39] (based on RTP/UDP (Real Time Protocol/User Datagram Protocol)), respectively. Compression mechanisms lead to a decrease in the bitrate of new or currently serviced packet streams and, in consequence, make it possible for a larger number of streams to be serviced. A decrease in the bitrate can be accompanied with an extension of service time, which occurs, for example, in the case of an execution of a service in which all data has to be transferred in its entirety (elastic traffic). In other cases, the service time is fixed (adaptive traffic). Elastic and adaptive traffic can be executed by threshold and thresholdless compression mechanisms. Thresholdless compression mechanisms influence calls that are being serviced, whereas threshold compression mechanisms reduce bitrate of streams in the admission stage of new calls with the help of the Call Admission Control (CAC) function. When this service has been initiated, the bitrate determined by the CAC function is not changed any more. In practice, elastic traffic is executed in a thresholdless manner, since it is characteristic for all services that employ the TCP protocol. These are typically nonreal time services, in which all data has to be transmitted. Hence, a decrease in the bitrate will be accompanied by an extension of the service time. Adaptive traffic, in turn, is typically executed in the threshold manner and is characteristic for real time services that employ the UDP protocol. Since the UDP protocol itself does not have an ability of influencing a packet source, it was necessary to introduce the RTP (Real Time Protocol) and RTSP (Real Time Streaming Protocol) protocols in upper layers. The protocols allow rapid changes in the bitrate speed of generated packet streams to be executed without any extension of the service time.

The work [40] proposes a multiservice FAR model that services thresholdless elastic traffic with finite compression. The notion of finite compression means that the allocated bitrate for a given flow (call) can be decreased within certain boundaries. In [41], this model is generalized to include infinite thresholdless compression in which flows always decrease their bitrates as soon as resources are occupied. In such a system, the phenomenon of call blocking will never occur even though the bitrates of serviced calls can, with large loads of the system, tend towards zero. The analytical base for the models [40, 41] is multidimensional reversible Markov processes with strictly specified values for service streams in individual states. They directly determine the distribution of FAR resources between serviced call classes. This distribution, resulting from a Markov process, is compliant with the so-called balanced fairness algorithm [42–44]. The algorithm leads to a state-dependent Markov process service of all traffic classes offered to the system. In [45], the model [40] is expanded to include service of thresholdless elastic and adaptive traffic.

The first FAR model with threshold compression is proposed in [46, 47]. In this model, developed for Erlang elastic traffic, one threshold is used (the so-called Single Threshold System (STS)). When the threshold is crossed by input signal, the bitrates of new calls that arrive are decreased. In [40], STS for elastic Engset traffic is proposed. The work [48] discusses systems with multiple thresholds (the so-called Multi Threshold System (MTS)) and for elastic Erlang traffic, while works [49, 50] discuss systems for Engset traffic and ON/OFF traffic, respectively. The authors of [1, 51] propose MTS for elastic and adaptive Erlang, Engset, and Pascal traffic. In [52], the threshold compression mechanism is replaced with a more complex hysteresis mechanism, in which thresholds are dependent on the direction of changes in the load of the system, i.e., two thresholds are introduced: one for the direction of changes: low loads-high loads; the other for the direction of changes: high loads-low loads. The paper [53] describes a model with double hysteresis and Erlang, Engset, and Pascal traffic.

The problem of determining traffic parameters of multiservice networks and optimal resource allocation in multiservice networks became especially important in the case of 5G mobile networks that rely on virtualization (slicing) of resources [54–57]. One of the fundamental difficulties arises from the necessity of servicing different classes of traffic streams by a network. The next important element influencing the complexity of the resource management mechanisms in virtual networks (slices) is the fact that resources can be executed both with the utilization of a single physical resource and with the utilization of many physical resources. The initial analysis of the problem related to designing feasible strategies of resource management in multiservice network indicates that the ones of the most effective strategies can be reservation mechanisms of resources, both dynamic (executed online and securing well-balanced access to network resources) and static (executed appropriately ahead of time with time advancement) [1], as well as threshold mechanisms [58, 59] and priority mechanisms [60].

The mechanisms adopted by telecommunications operators and network protocols (e.g., reservation mechanism and compression) require performing appropriate traffic analysis of the operating network systems and their optimal dimensioning. Until recently, the main emphasis has been put on the blocking probability calculation in systems without virtualization. Nowadays, in order to fully determine the influence of resource management mechanisms on the effectiveness of telecommunications networks as well as in order to determine the optimal share of virtual operators in physical resources, it is necessary to work out analytical methods that would enable us to model traffic characteristics of virtual networks with various resource management mechanisms and network protocols. Additionally, in the case of the mobile networks implementing slicing of resources, the very promising technique of optimization of network resources utilization is the overflow technique. The possibility of having dedicated resources (treated as primary resources) for particular traffic streams (flows) allows mobile network operator also to reserve a common resources (treated as secondary resources) for streams that—even after compression—could not be carried by the primary resources.

The first results on modelling networks with traffic overflow and elastic traffic were published in [45, 61]. The model [61] is generalized in [62] to include systems that service thresholdless elastic traffic in which secondary resources have a distributed nature, i.e., are composed of a number of separated resources with identical capacities [63] or differentiated capacities [64]. A model in which both primary resources and secondary resources have distributed nature is proposed in [65].

Until now, no models of overflow systems with threshold traffic compression have been developed, i.e., no model of an overflow system with adaptive traffic in which calls can undergo threshold compression in both primary resources and secondary resources has been developed. This means that, in the case where certain loads of the resources, determined earlier by an adopted set of thresholds, are exceeded, new calls will be admitted for service with decreased bitrates and unchanged service time. The absence of free resources in primary resources results in a situation where a call overflows to secondary resources, whereas lack of free resources in the secondary resources leads to a loss of the call. This article proposes a new model of an overflow system with adaptive Erlang, Engset, and Pascal traffic, executed by threshold mechanisms. The basis for the analytical description of primary and secondary resources will be the resource models proposed in [1, 51]. The accuracy of the proposed model will be verified on the basis of a comparison of the results of the analytical calculations with the results of the simulation experiments carried out for a network with dedicated primary resources and a common secondary resources.

The present article is structured as follows. Section 2 presents a description of the multiservice overflow system for adaptive traffic in which calls undergo the threshold compression mechanism. Section 3 includes a description of primary resources with streaming and adaptive traffic that is compressed in the threshold manner. This section also provides a description of a method for an evaluation of the traffic parameters for traffic that overflows from primary resources. Section 4 provides a method for modeling secondary resources as well as the algorithm for a determination of the blocking probability in an overflow adaptive traffic system. Section 5 provides a comparison of the results of the analytical calculation with the results of the simulation experiments for selected structures of overflow systems. Section 6 sums up the article.

2. Overflow of Adaptive Traffic in Mobile Networks

Literature on modeling networks (systems) with traffic overflow mostly considers systems in which primary resources are composed of a number of independent resources, called primary groups or direct groups, and secondary resources called secondary or alternative groups [66]. The approach adopted for this article employs the following terminology for primary resources (PR (primary resources)) and secondary resources (SR (secondary resources)) [62, 65] that better express the substantial idea of the overflow phenomenon described in the article, i.e., the authors believe that the notions of "virtual resources", "cell resources", or "radio interface resources" will be more understandable by the reader than the notion of "group".

The assumption in the article is that in both primary resources (slices) and secondary resources (slices) certain classes of traffic—related to the RTP/UDP protocols—can independently undergo threshold compression. This means that when given resources exceed the assumed occupancy threshold (load), new calls will be admitted with decreased bitrate that, during the service, will remain unchanged. Since the model aims at a description of adaptive traffic, the call service time is always the same and is independent of the compression process. Absence of appropriate bitrates for calls that are compressed in primary resources will cause, in turn, these calls to overflow to secondary resources. Since the threshold compression mechanism is assigned to given resources, the assumption is that a given flow overflows in noncompressed form and, while in secondary resources, can undergo independent (of primary resources) threshold compression. Absence of free resources in secondary resources leads then to losses in flows (calls) of particular classes.

2.1. Traffic Representation in the System. In modern mobile networks information is transmitted in IP packets. Packets that belong to a given service that is just being executed form streams that can be considered as calls or flows, e.g., [42, 45, 67]. A mathematical analysis of systems in which the internal structure of packet streams is taken into consideration is very complex [68, 69] and frequently leads to solutions that are approximate and of a very limited practical applicability or to simulation solutions [70, 71]. In dimensioning and optimization of network systems, however, the approach based on an analysis at the packet level cannot be applied to an all-encompassing and broad-based network analysis. The analysis at packet level only allows the aggregated parameters of a packet stream to be evaluated, whereas a determination of the characteristics of streams related to a given service in such a mixture of different streams with different characteristics is impossible. Therefore, the packet approach is not workable and inadequate for network operators. In traffic theory of multiservice systems a dominant approach is then the "call level" approach. It is only this approach that makes it possible to develop coherent models for dimensioning and optimization of networks that would make evaluation of the characteristics of individual services with their relevant GoS (Grade of Service), QoS (Quality of Service), and QoE (Quality of Experience) parameters taken into consideration possible. The results of the studies [58, 67] carried out in recent years indicate that streams at the call level can be described or approximated by streams that have "Poisson" nature. The adoption of such an approach makes it possible to discretize the system and to analyze it on the basis of multidimensional Markov processes. Discretization is based on an exchange of the variable bit rate (VBR) of a packet stream that forms a call by a certain constant bit rate (CBR), called equivalent bandwidth (EB) [72, 73]. EB describes then a constant bitrate that is logically allocated to a given call to provide appropriate (accurate) service parameters for this call in the network. Equivalent bandwidth is typically determined heuristically [74, 75], with regard to such parameters as the total capacity of the system, maximum and average bitrate of the packet stream, bit rate variance, maximum packet delay (latency), jitter, and other parameters characteristics for a given network technology [75–78]. Nowadays, in the description of multiservice systems related to modern packet networks (TCP/IP networks in particular), the assumption is that EB of relevant call classes is determined on the basis of the maximum bitrates of packet streams that correspond to calls. Such an approach is compliant with the system dimensioning principle for the highest network load conditions. It should be pointed out at this point, however, that the method for EB determination has no influence on a mathematical model of a system under consideration, while a choice as to the most appropriate method should be subjected to relevant arrangements between the network operator and involved entities or stakeholders that are to execute all the tasks concerning the design, development, and optimization of a network.

The next stage in the discretization process is a determination of the allocation unit (AU) for a given system, also called in literature Basic Bandwidth Unit (BBU) [58, 79, 80]. Most frequently this unit is defined as such a bitrate that a demanded EB of calls of individual classes offered to the system (called demands) is the multiple of the allocation unit. If an assumption is that a demanded EB for calls of individual classes that are offered to the system is determined on the basis of the maximum bitrates of calls of individual classes, then the maximum value of AU can be determined as the Greatest Common Divisor (GCD) of all maximum bitrates of calls offered to the system:

$$\max c_{\mathrm{AU}} = \mathrm{GCD}\left(c_{1,\max}, c_{2,\max}, \ldots, c_{M,\max}\right), \tag{1}$$

where $c_{i,\max}$ is the maximum bitrate of a call of class i, whereas c_{AU} is the bitrate of AU. In the last stage of the discretization process, both the demands of calls of individual classes and the capacity of the system (real or virtual) are expressed in AUs:

$$\begin{aligned} t_i &= \frac{c_{i,\max}}{c_{\mathrm{AU}}}, \\ V &= \left\lfloor \frac{C}{c_{\mathrm{AU}}} \right\rfloor, \end{aligned} \tag{2}$$

where C is the capacity of the system.

Since the value of AU determined on the basis of (1) is the maximum value, then if at a certain stage of considerations noninteger values of demands t_i for calls of particular classes appear (e.g. as a result of threshold traffic compression), then the system can be rescaled by an appropriate decrease in the value of AUs, which will provide integer values for demands of all call classes that are serviced in a system with threshold compression. For example, if the assumption is that as a result of the operation of a certain traffic shaping mechanism in the system, calls of class i with altered bitrate $c_{i,\max,\mathrm{new}}$ will appear, then the value of AUs can be selected on the basis of (1), with new bitrates for calls of individual classes taken into consideration:

$$\begin{aligned} \max c_{\mathrm{AU}} = \mathrm{GCD}\big(&c_{1,\max}, c_{2,\max}, \ldots, c_{M,\max}, c_{1,\max,\mathrm{new}}, \\ &c_{2,\max,\mathrm{new}}, \ldots, c_{M,\max,\mathrm{new}}\big). \end{aligned} \tag{3}$$

In modelling network systems related to TCP/IP networks the frequent assumption is that one AU has the value of 1 bps (or 1 kbps). Such an approach allows modeling of network systems to be greatly simplified [58]. The assumption in this article is that a call of each traffic class is determined by the number of demanded allocation units, calculated earlier on the basis of (2) and (3), that will always be integer numbers. Another assumption in the article is that both primary resources and secondary resources (real or virtual) have a defined capacity, expressed in AUs. Demands of individual call classes, also expressed in AUs, are known.

2.2. Resource Model with Threshold Compression for Adaptive Erlang Traffic. As an appropriate resource model with threshold compression for Erlang traffic, the Multi Threshold System (MTS), discussed in [1] can be used. Full-availability resources (FARs), with the capacity V AUs, are offered m call classes from the set M. For each call of class i ($0 < i \leq$

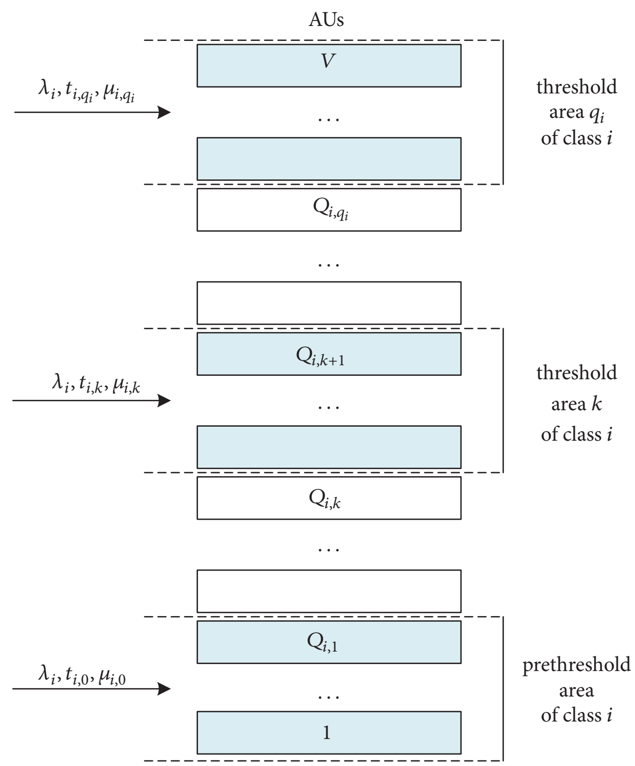

Figure 1: Resource model with threshold compression for adaptive Erlang traffic for class i calls.

m), a set q_i of thresholds $\{Q_{i,1}, Q_{i,2}, \ldots, Q_{i,q_i}\}$ is introduced individually, while $\{Q_{i,1} \le Q_{i,2} \le \ldots \le Q_{i,q_i}\}$. A threshold is understood to be a certain occupancy state in FAR, expressed in the number of occupied AUs. The occupancy area, such that the number of occupied n AUs satisfies the condition $0 \le n \le Q_{i,1}$, is called the prethreshold area. If the number of occupied AUs belongs to the postthreshold area k, then the condition $Q_{i,k} < n \le Q_{i,k+1}$ is fulfilled, where $0 < k \le q_i$. The postthreshold area q_i satisfies the condition $Q_{i,q_i} < n \le Q_{i,V}$ (Figure 1). In each postthreshold area k, the number of AUs allocated to service a call of class i is $t_{i,k}$, whereas this parameter in the prethreshold area is $t_{i,0}$. The assumption is that the parameters $t_{i,k}$ satisfy the inequality $t_{i,0} \le t_{i,1} \le \ldots \le t_{i,q_i}$. In the prethreshold area and in each postthreshold area k, offered traffic of class i is described by its own set of parameters $\{\lambda_{i,k}, \mu_{i,k}, t_{i,k}\}$. The parameter $\lambda_{i,k}$ is the intensity of calls of a Poisson stream (Erlang type traffic) and is independent of the FAR occupancy state. Since adaptive traffic is considered in MTS in which a change in the demanded AUs is not accompanied by any extension of the service time, then the average service intensity $\mu_{i,k}$ in each area, pre- and postthreshold, is identical. Therefore

$$\forall_{0\le k\le q_i} \lambda_{i,k} = \lambda_i, \qquad \mu_{i,k} = \mu_i. \tag{4}$$

Following (4), the intensity of traffic of class i in the prethreshold area and in each postthreshold area is identical:

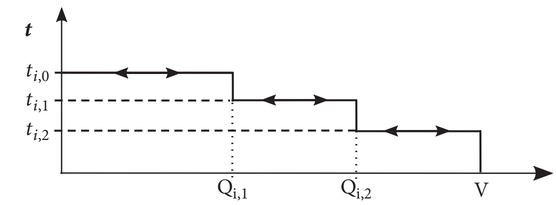

Figure 2: MTS for a single class and for $q_i = 2$.

$$\forall_{0\le k\le q_i} A_{i,Er,k} = \frac{\lambda_i}{\mu_i} = A_{i,Er}. \tag{5}$$

The symbol Er in the lower index indicates that traffic under consideration is Erlang traffic.

The operation of MTS with the example of one call class for two thresholds is shown in Figure 2. In the prethreshold area ($0 \le n \le Q_{i,1}$), the CAC function allocates to each call (of class i) $t_{i,0}$ AUs. An increase in the load of the system, i.e., the crossing of the threshold $Q_{i,1}$ results in a transition to the first postthreshold area ($Q_{i,1} < n \le Q_{i,2}$), in which the number of allocated AUs for each call of class i is $t_{i,1}$. A transition to the second postthreshold area ($Q_{i,2} < n \le V$) is followed by a decrease in the number of allocated AUs to the value $t_{i,2}$. The occupancy distribution in MTS can be determined by the following recurrence formula:

$$n\left[P_n\right]_V = \sum_{i=1}^{m}\sum_{k=0}^{q_i} A_{i,Er} t_{i,k} \delta_{i,k}\left(n - t_{i,k}\right)\left[P_{n-t_{i,k}}\right]_V, \tag{6}$$

where $[P_n]_V$ is occupancy probability n AUs in FAR with the capacity V AUs and $\delta_{i,k}(n)$ is conditional probability of transition for a stream of class i that can be determined in the following way:

$$\delta_{i,k}(n) = \begin{cases} 1 & \text{for } Q_{i,k} < n \le Q_{i,k+1}, \\ 0 & \text{for the remaining } n. \end{cases} \tag{7}$$

The assumption in Formula (7) is that $Q_{i,0} = 0$. Notice that the conditional transitional probability $\delta_{i,k}(n)$ is equal to unity only in such a load area in which the number of allocated AUs is equal to $t_{i,k}$.

The blocking probability in MTS can be determined on the basis of Formula (8):

$$E_i = \sum_{k=0}^{q_i} E_{i,k}, \tag{8}$$

where $E_{i,k}$ is the blocking probability of class i in the postthreshold area k:

$$E_{i,k} = \begin{cases} 0 & \text{for } \begin{cases} V - t_{i,k} \ge Q_{i,k+1}, \\ V - t_{i,k} > Q_{i,k}, \end{cases} \\ \displaystyle\sum_{n=V-t_{i,k}+1}^{Q_{i,k+1}} \left[P_n\right]_V & \text{for } \begin{cases} V - t_{i,k} < Q_{i,k+1}, \\ V - t_{i,k} > Q_{i,k}, \end{cases} \\ \displaystyle\sum_{n=Q_k+1}^{Q_{i,k+1}} \left[P_n\right]_V & \text{for } \begin{cases} V - t_{i,k} < Q_{i,k+1}, \\ V - t_{i,k} \le Q_{i,k}. \end{cases} \end{cases} \tag{9}$$

On the basis of (9) it is possible to verify that the blocking probability in a given postthreshold area k depends on the position of a given reference state $V - t_{i,k}$ in relation to the pair of thresholds $Q_{i,k}$ and $Q_{i,k+1}$ that limit the threshold area k. The assumption in our further considerations is that blocking of the system can occur only in the oldest post-threshold area q_i. This means that all reference states $V - t_{i,k}$, for $k \neq q_i$, are located above the threshold $Q_{i,k+1}$, whereas the threshold Q_{i,q_i} is selected in such a way that the conditions $Q_{i,q} < V - t_{i,q_i} < V$ are fulfilled. In this particular case the blocking probability in the threshold system is defined by the following formula:

$$E_i = E_{i,q_i} = \sum_{n=V-t_{i,q_i}+1}^{V} [P_n]_V . \qquad (10)$$

It should be stressed that such a choice as to the thresholds, where blocking state is possible only in the oldest postthreshold area, is in compliance with the very idea of the introduction of threshold mechanism; i.e., to avoid the blocking phenomenon, the system decreases bitrates of new calls after successive thresholds are crossed and it is only in the oldest postthreshold area (where the lowest bitrate is allocated to new calls) that a certain number of calls are blocked. Therefore, a selection of thresholds that allows blocking in different postthreshold areas to occur is not a workable choice from the engineering point of view.

If a given traffic class i is a class of streaming traffic that does not undergo threshold compression, then, in line with the adopted notation, one threshold is introduced to this traffic with the value equal to the capacity of FAR, i.e., $Q_{i,1} = V$. This means that the system in all the areas of possible occupancies can be treated as a prethreshold area in which the number of allocated AUs is always constant and is $t_{i,0}$. Hence, streaming traffic of class i can be described by the following conditions:

$$\delta_{i,k}(n) = \begin{cases} 1 & \text{for } k = 0 \text{ and } 0 \leq n \leq V, \\ 0 & \text{for } k = 1, \end{cases} \qquad (11)$$

$$t_{i,k} = \begin{cases} t_{i,0} & \text{for } k = 0, \\ 0 & \text{for } k = 1. \end{cases} \qquad (12)$$

If all traffic classes are of streaming nature, then (6) comes in its essence to the well-known recurrence [28, 29]:

$$n [P_n]_V = \sum_{i=1}^{m} A_{i,Er} t_{i,0} \left[P_{n-t_{i,0}}\right]_V . \qquad (13)$$

2.3. Model of Resources with Threshold Compression for Adaptive Erlang, Engset, and Pascal Traffic. Consider a FAR in which the capacity is V AUs and to which m traffic classes from the set M are offered. The assumption is that m_{Er} classes that belong to the set M_{Er} are of Erlang type, m_{En} classes that belong to the set M_{En} are Engset traffic, and m_{Pa} classes that belong to the set M_{Pa} are Pascal traffic, while $m_{Er} + m_{En} + m_{Pa} = m$ and $M_{Er} \cup M_{En} \cup M_{Pa} = M$. Engset and Pascal traffic are state-dependent and depend on the state of the system, in particular on the number of Engset and Pascal calls that are already being serviced [31]:

$$A_{j,En}(n) = \alpha_{j,En}\left[S_{j,En} - y_{j,En}(n)\right], \quad \text{where } j \in M_{En}, \qquad (14)$$

$$A_{l,Pa}(n) = \alpha_{l,Pa}\left[S_{l,Pa} + y_{l,Pa}(n)\right], \quad \text{where } l \in M_{Pa}, \qquad (15)$$

where

(i) $A_{j,En}(n)$ is average intensity of Engset traffic of class j in FAR that is in the occupancy state n AUs,

(ii) $A_{l,Pa}(n)$ is average intensity of Pascal traffic of class l in FAR that is in the occupancy state n AUs,

(iii) $\alpha_{j,En}$ is average intensity of Engset traffic of class j generated by one free source:

$$\alpha_{j,En} = \frac{\gamma_{j,n}}{\mu_{j,n}}, \qquad (16)$$

(iv) $\gamma_{j,n}$ is average intensity of Engset calls of class j, generated by one free source,

(v) $\mu_{j,n}$ is average service intensity for Engset calls of class j,

(vi) $\alpha_{l,Pa}$ is average intensity of Pascal traffic of class l generated by one free source:

$$\alpha_{l,Pa} = \frac{\gamma_{l,Pa}}{\mu_{l,Pa}}, \qquad (17)$$

(vii) $\gamma_{l,Pa}$ is average intensity of Pascal calls of class l generated by one free source,

(viii) $\mu_{l,Pa}$ is average service intensity for Pascal calls of class l,

(ix) $S_{j,En}$ is the number of Engset traffic sources of class j,

(x) $S_{l,Pa}$ is the number of Pascal traffic sources of class l.

The parameters $y_{j,En}(n)$ and $y_{l,Pa}(n)$ are the average values of the number of serviced Engset calls of class j and Pascal calls of class l in FAR that are in the occupancy state n AUs. The method for a determination of these parameters will be given further on in the section. Let us assume now that each class of Engset and Pascal traffic will be assigned its own set of thresholds and that traffic is of adaptive nature. The occupancy distribution in MTS to which a mixture of Erlang, Engset, and Pascal traffic is offered can be determined on the basis of the following recurrence formula [51]:

$$n\left[P_n\right]_V = \sum_{i\in M_{Er}} \sum_{k=0}^{q_j} A_{i,Er} t_{i,k} \delta_{i,k}\left(n-t_{i,k}\right)\left[P_{n-t_{i,k}}\right]_V + \sum_{j\in M_{En}} \sum_{k=0}^{q_j} \alpha_{j,En}\left[S_{j,En} - y_{j,En,k}\left(n-t_{j,k}\right)\right] \cdot t_{j,k}\delta_{j,k}\left(n-t_{j,k}\right)\left[P_{n-t_{j,k}}\right]_V + \sum_{l\in M_{Pa}} \sum_{k=0}^{q_l} \alpha_{l,Pa} \cdot \left[S_{l,Pa} + y_{l,Pa,k}\left(n-t_{l,k}\right)\right] t_{l,k}\delta_{l,k}\left(n-t_{l,k}\right) \cdot \left[P_{n-t_{l,k}}\right]_V, \quad (18)$$

where the parameters $y_{j,En,k}$ and $y_{l,Pa,k}(n)$ are the average numbers of serviced Engset calls of class j and Pascal calls of class l in FAR that are in the occupancy state n AUs, whereas n belongs to the prethreshold area ($k = 0$) or postthreshold area ($0 < k \le q_j, 0 < k \le q_l$). These parameters can be determined on the basis of the following formula:

$$y_{j,En,k}(n) = \frac{\alpha_{j,En}\left[S_{j,En} - y_{j,En,k}\left(n-t_{j,k}\right)\right]\left[P_{n-t_{j,k}}\right]_V}{\left[P_n\right]_V}, \quad (19)$$

$$y_{l,Pa,k}(n) = \frac{\alpha_{l,Pa}\left[S_{l,Pa} + y_{l,Pa,k}\left(n-t_{l,k}\right)\right]\left[P_{n-t_{l,k}}\right]_V}{\left[P_n\right]_V}. \quad (20)$$

Formulas (19) and (20) define, respectively, the parameters $y_{j,En,k}(n)$ and $y_{l,Pa,k}(n)$ for the following ranges of the variable n: $Q_{j,k} + t_{j,k} < n \le Q_{j,k+1}$; $Q_{l,k} + t_{l,k} < n \le Q_{l,k+1}$ [51].

Assuming, exactly as in the case of MTS with Erlang traffic, that blocking of the system can occur only in the oldest postthreshold area, the blocking probability for calls of individual types can be expressed by Formula (10) that in the adopted notation for particular traffic types can be written as follows:

$$E_{i,Er} = \sum_{n=V-t_{i,q_i}+1}^{V} \left[P_n\right]_V \quad \text{for } i \in M_{Er}, \quad (21)$$

$$E_{j,En} = \sum_{n=V-t_{j,q_j}+1}^{V} \left[P_n\right]_V \quad \text{for } j \in M_{En}, \quad (22)$$

$$E_{l,Pa} = \sum_{n=V-t_{l,q_l}+1}^{V} \left[P_n\right]_V \quad \text{for } l \in M_{Pa}. \quad (23)$$

In the occupancy distribution FAR with adaptive traffic executed in the threshold manner (18), there are the parameters $y_{j,En,k}(n)$ and $y_{l,Pa,k}(n)$ that can, in turn, be determined on the basis of the occupancy distribution (Formulas (19) and (20)). Therefore, to determine the distribution (18) it is necessary to construct an iterative algorithm in which in each iteration step the approximate occupancy distribution can be determined on the basis of the values of the average number of serviced calls of Engset and Pascal classes, determined in the preceding iteration step. A general method for a construction of these algorithms is proposed in [31, 34].

3. Modelling of Primary Resources with Adaptive Traffic and Threshold Compression

Figure 3 shows a general diagram of a multiservice traffic overflow system. The system of primary resources (PRs) is composed of r FARs, from which each can be considered as MTS with Erlang, Engset, and Pascal adaptive traffic.

Each PR s ($0 < s \le r$) has the capacity $V^{(s)}$ expressed in AUs. The primary resources s are offered $m^{(s)}$ traffic classes from the set $M^{(s)}$, where $m^{(s)}_{Er}$ classes that belong to the set $M^{(s)}_{Er}$ are Erlang, $m^{(s)}_{Er}$ classes that belong to the set $M^{(s)}_{En}$ are Engset, and $m^{(s)}_{Pa}$ classes that belong to the set $M^{(s)}_{Pa}$ are Pascal traffic sources, while $m^{(s)}_{Er} + m^{(s)}_{En} + m^{(s)}_{Pa} = m^{(s)}$ and $M^{(s)}_{Er} \cup M^{(s)}_{En} \cup M^{(s)}_{Pa} = M^{(s)}$. In Figure 3, the following notation for the respective traffic intensities is adopted:

(i) $A^{(s)}_{c,X}$ is average intensity of traffic of class c ($c \in M^{(s)}$) of type X ($X = Er \mid En \mid Pa$) offered to PR no. s,

(ii) $R^{(s)}_{c,X}$ is average intensity of traffic of class c ($c \in M^{(s)}$) that overflows from PR no. s, with the assumption that traffic of class c offered to primary resources s is of type X ($X = Er \mid En \mid Pa$).

In each PR s, for each call class c, an individual set $q^{(s)}_c$ of thresholds $\{Q^{(s)}_{c,1}, Q^{(s)}_{c,1}, \ldots, Q^{(s)}_{c,1}\}$ and an individual set $q^{(s)}_c + 1$ of demands allocated in appropriate load areas $\{t^{(s)}_{c,0}, t^{(s)}_{c,1}, \ldots, t^{(s)}_{c,q_c},\}$ are defined. With these assumptions and on the basis of the model presented in Section 2.3, Formulas (18)–(23) can be applied to determine blocking probabilities for individual call classes in each primary resource s. The blocking probability for calls of class c ($c \in M^{(s)}$) of type X ($X = Er \mid En \mid Pa$) can be written on the basis of (21)–(23) as follows:

$$E^{(s)}_{c,X} = \sum_{n=V^{(s)}-t^{(s)}_{c,q_c}+1}^{V^{(s)}} \left[P_n\right]_{V^{(s)}} \quad \text{for } c \in M^{(s)}_X. \quad (24)$$

3.1. Decomposition of Primary Resources. In multiservice models that are based on a description of overflow traffic with two moments, the average value and variance, each PR s undergoes decomposition into $m^{(s)}$ fictitious primary resources (FPR s_c) [24], each with the capacity $v^{(s)}_{c,X}$. Then, each FPR s_c is replaced by equivalent fictitious primary resources (EFPR s_c). Each EFPR s_c can be characterized by equivalent capacity $v^{*(s)}_{c,X}$ and offered, equivalent, Erlang traffic with the intensity $A^{*(s)}_{c,X}$, while these parameters are selected in such a way that traffic that overflows from EFPR s_c has exactly the same parameters (average value and variance) as traffic that overflows from FPR s_c.

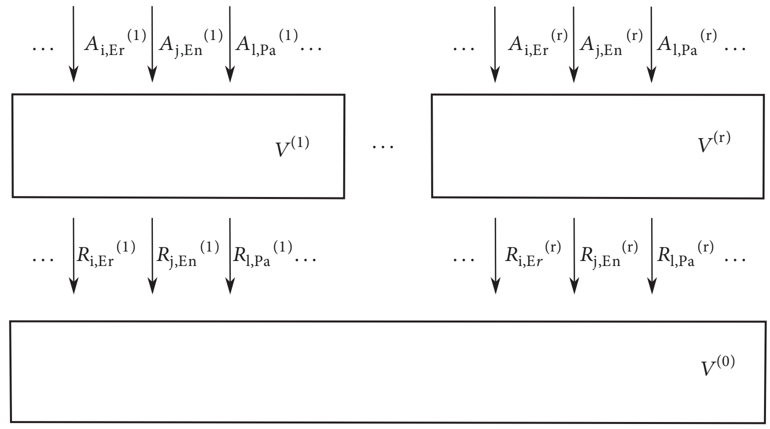

Figure 3: Multiservice overflow traffic system.

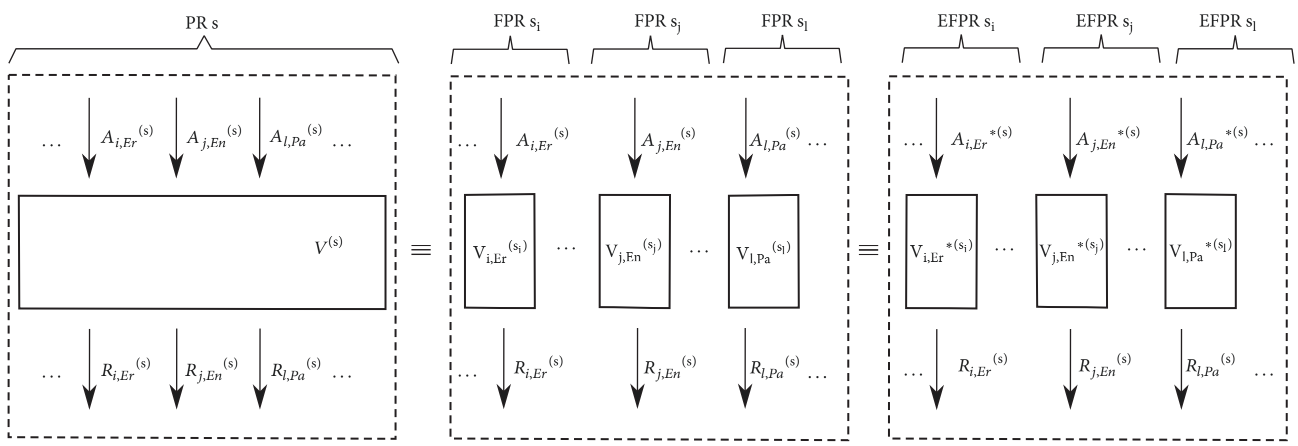

Figure 4: Decomposition of PR s.

3.2. Determination of the Parameters of FPR s_c. The need for decomposition results from the fact that PR s carries multiservice traffic, and in consequence a direct determination of variance for traffic of individual classes is not possible. Each FPR s_c services exclusively calls of just one class c, which, further on in the article, will allow Riordan formulas [3] to be applied to determine variance of traffic of class c that overflows from FPR s_c. The following assumptions are made to determine the capacity of FPRs:

(1) Blocking probability $E_{c,X}^{(s_c)}$ of calls of class c type X in FPR s with the capacity $v_{c,X}^{(s)}$ is exactly the same as the blocking probability $E_{c,X}^{(s)}$ of calls of this class in PR s with the capacity $V^{(s)}$:

$$E_{c,X}^{(s_c)} = E_{c,X}^{(s)} \quad \text{for } c \in M_X^{(s)}. \tag{25}$$

(2) Traffic in FPR s_c does not undergo the threshold compression mechanism.

Figure 4 shows the way PR s is decomposed into FPR s_c and EFPR s_c.

The adopted assumptions define the method for a decomposition of the resources. As it is mentioned earlier, on the basis of the model presented in Section 2.3 (Formulas (18)–(23), it is possible to determine blocking probabilities for individual call classes in each of the PRs s. It results from the adoption of Assumption (1) (Formula (25)) that the obtained probabilities are exactly the same as the probabilities in the corresponding FPR. Assumption (2) indicates that there is a possibility to determine the capacity $v_{c,X}^{(s)}$ of decomposed resources on the basis of single-service models FAR for Erlang, Engset, and Pascal traffic, in which it is possible to determine the blocking probability as the occupancy probability of all AUs:

$$E_{c,X}^{(s_c)} = E_{c,X}^{(s)} = \left[P_{v_{c,X}^{(s)}}\right]_{v_{c,X}^{(s)}} \quad \text{for } c \in M_X^{(s)}. \tag{26}$$

Thus, in the case of Erlang traffic, the capacity of the fictitious primary resources can be determined on the basis of Erlang B Formula:

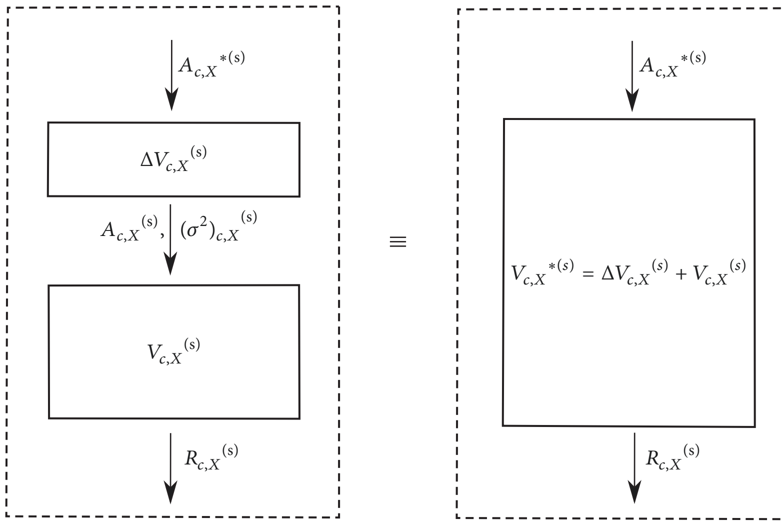

Figure 5: Diagram of the replacement of FPR s_c by EFPR s_c.

$$E_{i,Er}^{(s_i)} = \left[P_{\nu_{i,Er}^{(s)}}\right]_{\nu_{i,Er}^{(s)}} = E_{\nu_{i,Er}^{(s)}}\left(A_{i,Er}^{(s)}\right) = \frac{\left(A_{i,Er}^{(s)}\right)^{\nu_{i,Er}^{(s)}} / \left(\nu_{i,Er}^{(s)}\right)!}{\sum_{n=0}^{\nu_{i,Er}^{(s)}}\left(\left(A_{i,Er}^{(s)}\right)^n / n!\right)}. \quad (27)$$

In Formula (27), the parameters $E_{i,Er}^{(s_i)}$ and $A_{i,Er}^{(s)}$ are known; hence, on their basis, it is possible to determine the parameter $\nu_{i,Er}^{(s)}$. In a similar way, the capacities of fictitious resources to which single-service Engset and Pascal traffic is offered can be determined on the basis of appropriate formulas that determine the blocking probability in the Engset and Pascal model:

$$E_{j,En}^{(s_j)} = \left[P_{\nu_{j,En}^{(s)}}\right]_{\nu_{j,En}^{(s)}} = \frac{\binom{S_{j,En}^{(s)}}{\nu_{j,En}^{(s)}}\left(\alpha_{j,En}^{(s)}\right)^{\nu_{j,En}^{(s)}}}{\sum_{n=0}^{\nu_{j,En}^{(s)}}\binom{S_{j,En}^{(s)}}{n}\left(\alpha_{j,En}^{(s)}\right)^{\nu_{j,En}^{(s)}}}, \quad (28)$$

$$E_{l,Pa}^{(s_l)} = \left[P_{\nu_{l,Pa}^{(s)}}\right]_{\nu_{l,Pa}^{(s)}} = \frac{\binom{-S_{l,Pa}^{(s)}}{\nu_{l,Pa}^{(s)}}\left(-\alpha_{l,Pa}^{(s)}\right)^{\nu_{l,Pa}^{(s)}}}{\sum_{n=0}^{\nu_{l,Pa}^{(s)}}\binom{-S_{l,Pa}^{(s)}}{n}\left(-\alpha_{l,Pa}^{(s)}\right)^{\nu_{l,Pa}^{(s)}}}. \quad (29)$$

Note that Formulas (27)–(29) apply to single-service systems in which admission of a new call means the occupation of one AU.

3.3. Determination of the Parameters of EFPR s_c. The next element in the decomposition of PR s is the replacement of FPR s_c by EFPR s_c. The need for such a replacement is related to a possibility to determine, further on in this article, the variance of traffic that overflows from EFPR s_c with the help of Riordan formulas [3]. Since these formulas make it possible to determine variance exclusively in the case where offered traffic is Erlang traffic, then a replacement of FPR s_c by EFPR s_c demands a replacement of Engset and Pascal traffic by equivalent Erlang traffic. Let us denote the variance of traffic of class c type X by $(\sigma_\Delta^2)_{c,X}^{(s)}$. Equivalent Erlang traffic of $A_{j,En}^{*(s)}$ and $A_{l,Pa}^{*(s)}$ is such traffic offered to certain fictitious, additional resources with the capacities $\Delta\nu_{j,En}^{(s)}$ and $\Delta\nu_{l,Pa}^{(s)}$ that traffic that overflows from these resources is equal, with respect to the average value and variance, to Engset $(A_{j,En}^{(s)}, (\sigma_\Delta^2)_{j,En}^{(s)})$ and Pascal traffic $(A_{l,Pa}^{(s)}, (\sigma_\Delta^2)_{l,Pa}^{(s)})$, respectively. Note that, in the notation of equivalent traffic, e.g., in $A_{j,En}^{*(s)}$, the symbol *En* that indicates the primary nature of this traffic has been retained in the lower index. The "asterisk" introduced to the upper index means that this traffic is already equivalent Erlang traffic. In the case where primary traffic is Erlang traffic, we have $A_{i,Er}^{*(s)} = A_{i,Er}^{(s)}$, $(\sigma_\Delta^2)_{i,Er}^{(s)} = A_{i,Er}^{(s)}$, and $\Delta\nu_{i,Er}^{(s)} = 0$.

The EFPR parameters s_c can be determined on the basis of the equivalent random theory method (ERT) [2, 3] that, for Engset traffic, is described in [8]. The method will be generalized to include the particular case of Pascal traffic. The diagram of traffic replacement from Engset and Pascal traffic to equivalent Erlang traffic is presented in Figure 5. Average values and variances for Engset and Pascal traffic can be determined on the basis of the following dependencies:

(i) Erlang traffic:

$$A_{i,Er}^{(s)}, \left(\sigma_\Delta^2\right)_{i,\mathrm{Er}}^{(s)} = A_{i,Er}^{(s)}, \quad (30)$$

(ii) Engset traffic:

$$A_{j,En}^{(s)} = S_{j,En}^{(s)}\frac{\alpha_{j,En}^{(s)}}{1+\alpha_{j,En}^{(s)}},$$
$$\left(\sigma_\Delta^2\right)_{j,En}^{(s)} = S_{j,En}^{(s)}\frac{\alpha_{j,En}^{(s)}}{\left(1+\alpha_{j,En}^{(s)}\right)^2}, \quad (31)$$

(iii) Pascal traffic:

$$A_{l,Pa}^{(s)} = S_{l,Pa}^{(s)} \frac{\alpha_{l,Pa}^{(s)}}{1-\alpha_{l,Pa}^{(s)}},$$
$$\left(\sigma_{\Delta}^{2}\right)_{l,Pa}^{(s)} = S_{l,Pa}^{(s)} \frac{\alpha_{l,Pa}^{(s)}}{\left(1-\alpha_{l,Pa}^{(s)}\right)^{2}}. \tag{32}$$

According to the ERT method, the average value R and variance σ^2 of traffic that overflows from PR with the capacity V to which Erlang traffic A is offered can be determined on the basis of Riordan formulas [3]:

$$R = AE_V(A), \tag{33}$$
$$\sigma^2 = R\left(\frac{A}{V+1-A+R}+1-R\right). \tag{34}$$

In the considered traffic replacement diagram (Figure 5), the resources with the capacity $V = \Delta v_{c,X}^{(s)}$ is offered equivalent Erlang traffic with the intensity $A = A_{c,X}^{*(s)}$. Traffic that overflows from these resources has the average value $R = A_{c,X}^{(s)}$ and variance $(\sigma_{\Delta}^{2})_{c,X}^{(s)}$. Therefore, in the notation adopted in the article, Formulas (33) and (34) can be written in the following way:

$$A_{c,X}^{(s)} = A_{c,X}^{*(s)} E_{\Delta v_{c,X}^{(s)}}\left(A_{c,X}^{*(s)}\right), \tag{35}$$

$$\left(\sigma_{\Delta}^{2}\right)_{c,X}^{(s)} = A_{c,X}^{(s)}\left(\frac{A_{c,X}^{*(s)}}{\Delta v_{c,X}^{(s)}+1-A_{c,X}^{*(s)}+A_{c,X}^{(s)}}+1-A_{c,X}^{(s)}\right). \tag{36}$$

Formulas (35) and (36) allow the pair of parameters ($A_{c,X}^{*(s)}$, $\Delta v_{c,X}^{(s)}$) to be determined on the basis of known values of the pair of parameters ($A_{c,X}^{(s)}$, $(\sigma_{\Delta}^{2})_{c,X}^{(s)}$) that, depending on considered type of traffic, are described by Formulas (30)–(32). At this particular point it is worthwhile to note the fact that the diagram of equivalent replacement of FPR s_c by EFPR s_c for Engset traffic offers a solution exclusively for negative values of the capacity $\Delta v_{c,X}^{(s)}$ in Formulas (35) and (36).

Having determined the value of the parameters ($A_{c,X}^{*(s)}$, $\Delta v_{c,X}^{(s)}$), it is possible to evaluate the capacity of EFPR s_c, i.e., the parameter $v_{c,X}^{*(s)}$ [3]:

$$v_{c,X}^{*(s)} = v_{c,X}^{(s)} + \Delta v_{c,X}^{(s)}. \tag{37}$$

Note that the pair of parameters ($A_{c,X}^{*(s)}, v_{c,X}^{*(s)}$) for each EFPR s_c determines the Erlang model for FAG [81].

3.4. Determination of Parameters of Traffic That Overflows from EFPR s_c. Traffic of class c that overflows from PR s is equivalent to traffic that overflows from EFPR s_c and will be characterized by two parameters: the average value of traffic intensity $R_{c,X}^{(s)}$ and variance $(\sigma^2)_{c,X}^{(s)}$. Since EFPR s_c are determined by a single-service Erlang model, then the parameters $R_{c,X}^{(s)}$ and $(\sigma^2)_{c,X}^{(s)}$ can be determined on the basis of Riordan formulas (33) and (34) that, in the adopted notation, will be written in the following way:

$$R_{c,X}^{(s)} = A_{c,X}^{*(s)} E_{v_{c,X}^{*(s)}}\left(A_{c,X}^{*(s)}\right), \tag{38}$$

$$\left(\sigma^{2}\right)_{c,X}^{(s)} = R_{c,X}^{(s)}\left(\frac{A_{c,X}^{*(s)}}{v_{c,X}^{*(s)}+1-A_{c,X}^{*(s)}+R_{c,X}^{(s)}}+1-R_{c,X}^{(s)}\right). \tag{39}$$

The peakedness factor $Z_{c,X}^{(s)}$ for traffic of class c that overflows from EFPR s_c is defined as the ratio of variance to the average value:

$$Z_{c,X}^{(s)} = \frac{\left(\sigma^{2}\right)_{c,X}^{(s)}}{R_{c,X}^{(s)}}. \tag{40}$$

Further on in the article, traffic that overflows from EFPR s_c will be characterized by the pair of parameters ($R_{c,X}^{(s)}$ i $(\sigma^2)_{c,X}^{(s)}$).

4. Modeling of Secondary Resources with Adaptive Traffic and Threshold Compression

The parameters $R_{c,X}^{(s)}$ and $(\sigma^2)_{c,X}^{(s)}$ of traffic that overflows from individual EFPR s_c are simultaneously the parameters for traffic offered to SR with the capacity $V^{(0)}$ AUs. In SR, traffic can also undergo the threshold compression mechanism. This means that, in exactly the same way as in the case of PR, for each call class c ($0 < c \le m$) the set $q_c^{(0)}$ of thresholds $\{Q_{c,1}, Q_{c,2}, \ldots, Q_{c,q_c}\}$ is introduced individually, where the occupancy area that satisfies the condition $0 < n \le Q_{c,1}$ is the prethreshold area, whereas the occupancy area that satisfies the condition $Q_{c,k} \le Q_{c,k+1}$ is called the postthreshold area k, where $1 \le k \le q_c^{(0)}$. In each postthreshold area k, the number of AUs allocated to service calls of class c is $t_{c,k}^{(0)}$, while the value of this parameter in the prethreshold area is $t_{c,0}^{(0)}$.

4.1. Determination of the Occupancy Distribution in SR. To model the system of secondary resources in this article, Hayward approach [10] is used. The approach is based on a division of SR parameters (traffic intensity and capacity) by the peakedness factor of offered traffic. Such an approach is used in [24] to model secondary resources composed of a single FAR and servicing a mixture of different classes of overflow traffic. In this section, Hayward approach is used to describe the characteristics of SR that is composed of a single MTS. As a result, the occupancy distribution and blocking probability in SR can be written, using the adopted notation, as follows:

$$n\left[P_n\right]_{V^{(0)}/Z^{(0)}} = \sum_{s=1}^{r}\sum_{c=1}^{m^{(s)}}\sum_{k=0}^{q_c^{(0)}} \frac{R_{c,X}^{(s)}}{Z_{c,X}^{(s)}} t_{c,k}^{(0)} \delta_{c,k}^{(0)}\left(n-t_{c,k}^{(0)}\right)\left[P_{n-t_{c,k}^{(0)}}\right]_{V^{(0)}/Z^{(0)}}, \tag{41}$$

where $\delta_{c,k}^{(0)}(n)$ is the conditional transition probability for a stream of class c that in the case of the distribution (41) can be determined in the following way:

$$\delta_{c,k}^{(0)}(n) = \begin{cases} 1 & \text{for } \dfrac{Q_{i,k}}{Z^{(0)}} < n \leq \dfrac{Q_{i,k+1}}{Z^{(0)}}, \\ 0 & \text{for the remaining } n. \end{cases} \quad (42)$$

In Formula (41), the peakedness factors $Z_{c,X}^{(s)}$ of traffic that overflows from PRs are defined by Formulas (38)–(40). The parameter $Z^{(0)}$ is the so-called aggregate peakedness factor. The introduction of this approach results from a necessity, in compliance with Hayward's concept, to normalize the total capacity of the system of secondary resources $V^{(0)}/Z^{(0)}$ to which an appropriate mixture of traffic, which overflows from each of the r of primary resources, is offered. The aggregate peakedness factor can be determined approximately on the basis of the weighted mean of the peakedness factor of an individual traffic classes offered to SR [24]:

$$Z^{(0)} = \frac{\sum_{s=1}^{r} \sum_{c=1}^{m^{(s)}} \left(\sigma^2\right)_{c,X}^{(s)}}{\sum_{s=1}^{r} \sum_{c=1}^{m^{(s)}} R_{c,X}^{(s)}}. \quad (43)$$

After a determination of the occupancy distribution in SR, the blocking probability for each class of calls in secondary resources can be determined:

$$\left(E_{c,X}^{(s)}\right)^{(0)} = \sum_{n=V^{(0)}/Z^{(0)}-t_{c,q_c}^{(0)}+1}^{V^{(0)}} \left[P_n\right]_{V^{(0)}/Z^{(0)}}. \quad (44)$$

4.2. Method of Modelling SR. The models presented in Sections 3 and 4 allow a method for a calculation of the blocking probability and other characteristics in a multiservice overflow system with adaptive traffic in primary and secondary resources to be defined. The method can be presented in the following steps:

(1) Determination of occupancy distributions $[P_n]_{V^{(s)}}$ (Formula (18)) and the blocking probability $E_{c,X}^{(s)}$ (Formula (24)) for traffic streams of all classes in each PR s, $1 \leq s \leq r$.

(2) Determination of the capacity $v_{c,X}^{(s)}$ for each FPR s_c (Formulas (27)–(29)).

(3) Determination of the variance $(\sigma_\Delta^2)_{c,X}^{(s)}$ of each Erlang, Engset, and Pascal traffic offered to PR s, $1 \leq s \leq r$ (Formulas (30)–(32)).

(4) Determination of the parameters of each of EFPR s_c, i.e., the equivalent intensity of Erlang traffic $A_{c,X}^{*(s)}$ and the equivalent capacity $v_{c,X}^{*(s)}$ (Formulas (35)–(37)).

(5) Determination of the parameters of overflow traffic: mean value of traffic intensity $R_{c,X}^{(s)}$ and variance $(\sigma^2)_{c,X}^{(s)}$ (Formulas (38)–(39)) for traffic of all classes that overflows from the system of primary resources.

(6) Determination of the occupancy distribution $[P_n]_{V^{(0)}/Z^{(0)}}$ (Formula (41)) and the aggregate peakedness factor $Z^{(0)}$ in SR (Formula (43)).

(7) Determination of the blocking probability for traffic streams of all classes offered to SR (Formula (44)).

4.3. Comment. It is possible to determine on the basis of the blocking probability other important QoS characteristics, such as the call loss probability. The loss probability in an overflow system for Erlang traffic is equal to the blocking probability. It has been proved in traffic theory, e.g., [13], that in a single-service Engset (Pascal) model the call loss probability is equal to the blocking probability in a system with identical capacity in which the number of traffic sources has been decreased (increased) by one source. This approach can be then used to approximately determine the call loss probability for Engset and Pascal traffic classes in multiservice systems. Therefore, the loss probability for Engset (Pascal) traffic classes can be determined on the basis of the blocking probability, assuming that the number of traffic sources of each Engset (Pascal) traffic class has been decreased (increased) by one source.

5. Numerical Examples

The proposed model of multiservice overflow system with adaptive traffic is an approximate one. In order to determine its accuracy, the results of analytical modelling have been compared with the simulation data obtained for the system described in Table 1.

The system under consideration is characterised in Table 1 by specifying the number of primary resources and their capacity, the capacity of the secondary resources, and the number of AUs required by particular traffic classes in prethreshold and postthreshold areas. It was assumed that each of primary resource was offered traffic of various classess in the following proportions: $A_{1,0,s}t_{1,0,s} : A_{2,0,s}t_{2,0,s} : \cdots = 1 : 1 : \ldots$. We can notice that the PR no. 1 was offered three traffic classes (flows, services), with the demands equal to 1, 6, and 12 AUs, respectively. In the case of the PR no. 2, only two traffic classes were offered, demanding 3 and 11 AUs. The PR no. 3 and PR no. 4 were dedicated for servicing only a single traffic class. Such distribution of demands for particular resources can be treated as an example of allocating of separated resources (virtual resources, slices) to particular demands. The flows that cannot be serviced by dedicated primary resources are offered to the common secondary resources. In both types of resources the flows undergo threshold compression, according to the parameters specified in Table 1. The volume of resources admitted for flows in particular load areas were different in primary and secondary resources.

The results of blocking probabilities for all traffic classes in both primary and secondary resources are presented in Figures 6–9 and 10, respectively. In the case of the secondary resources, the results of blocking probability for traffic classes nos. 3, 5, and 6 were the same since all these classes were admitted the same amount of AUs in the last threshold

Table 1: Description of the system with threshold compression in primary and secondary resources. The secondary resources are denoted by $s = 0$.

s	$V^{(s)}$	k	c	$Q^{(s)}_{c,k}$	$t^{(s)}_{c,k}$	typ	$S^{(s)}_{c,k}$
		0	1	0	1	Er	–
		0	2	0	6	En	10
		0	3	0	12	Pa	10
		1	2	36	3	En	10
		1	3	32	8	Pa	10
1	48	2	3	42	6	Pa	10
		0	4	0	3	Er	–
		0	5	0	11	En	12
		1	4	30	2	Er	–
2	50	1	5	24	6	En	12
		0	6	0	15	En	20
		1	6	30	12	En	20
3	100	2	6	50	10	En	20
		0	7	0	17	Pa	10
4	120	1	7	50	12	Pa	10
		0	1	0	1	Er	–
		0	2	0	6	En	10
		0	3	0	12	Pa	10
		0	4	0	3	Er	–
		0	5	0	11	En	12
		0	6	0	15	En	20
		0	7	0	17	Pa	10
		1	2	20	3	En	10
		1	3	20	10	Pa	10
		1	4	30	2	Er	–
		1	5	30	5	En	12
		1	6	20	10	En	20
		1	7	20	12	Pa	10
		2	3	30	5	Pa	10
		2	6	30	5	En	20
0	100	2	7	30	10	Pa	10

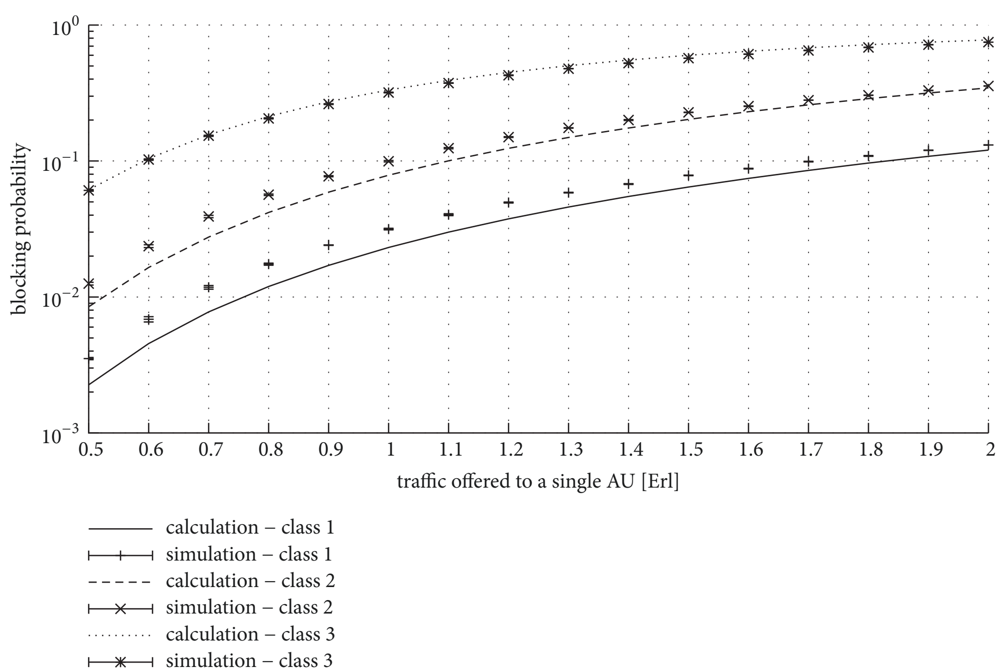

Figure 6: Blocking probability in primary resources (slice) no. 1 for traffic classes nos. 1, 2, and 3 from Table 1. Traffic offered per single AU calculated for initial value of demanded AUs (before compression).

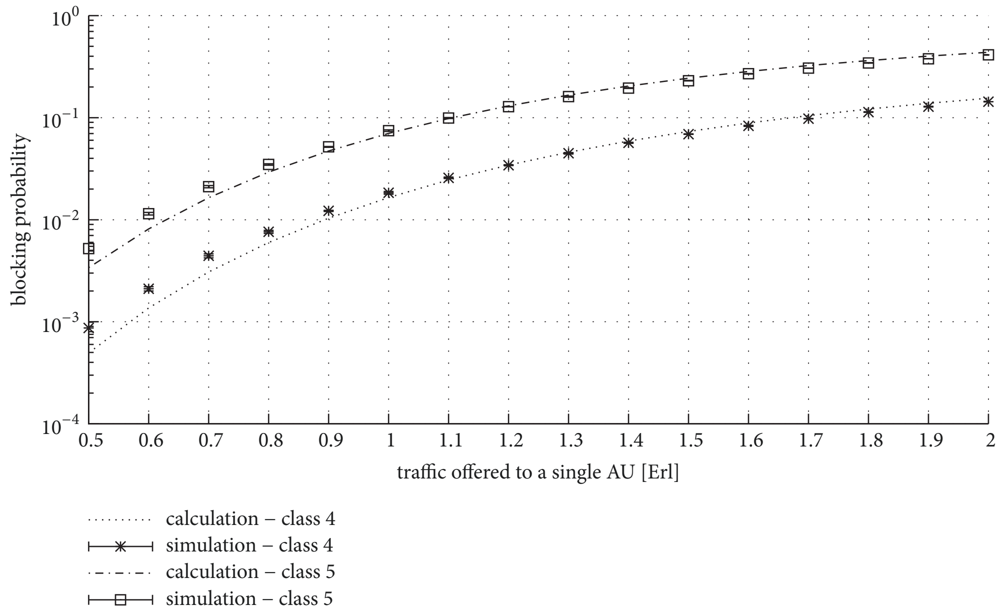

FIGURE 7: Blocking probability in primary resources (slice) no. 2 for traffic classes nos. 4 and 5 from Table 1. Traffic offered per single AU calculated for initial value of demanded AUs (before compression).

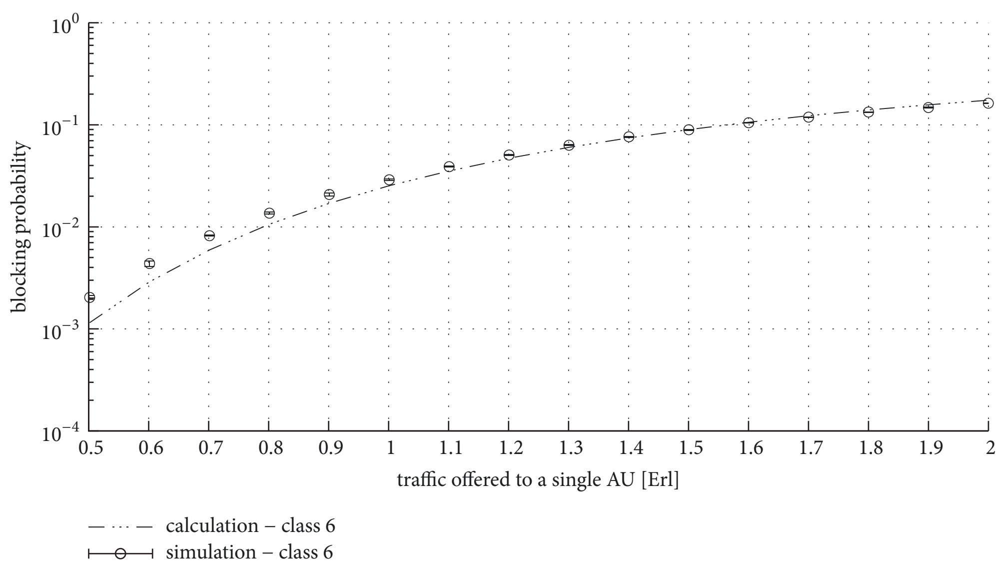

FIGURE 8: Blocking probability in primary resources (slice) no. 3 for traffic class no. 6 from Table 1. Traffic offered per single AU calculated for initial value of demanded AUs (before compression).

(in the area of the highest load). In order to increase the readability of the Figure 10, the results for class 4 were omitted.

The results presented in Figures 6–10 indicate good accuracy of the proposed method, for both primary and secondary resources. In the proposed method, the errors within the area of lower loads mainly result from the fact that in Formula (41) the coefficient $V^{(0)}/Z^{(0)}$ can take on noninteger values. In such a case, linear interpolation between the values $\lfloor V^{(0)}/Z^{(0)} \rfloor$ and $\lceil V^{(0)}/Z^{(0)} \rceil$ is applied in the model and this approximation is largely responsible for the decrease in accuracy of the proposed method in the area of low losses. The problem of noninteger values of the coefficients $V^{(0)}/Z^{(0)}$ in generalized Hayward's formula is discussed in [62].

6. Conclusions

This article proposes a model of multiservice traffic overflow system with adaptive traffic that undergoes the threshold compression mechanism in both primary and secondary resources. Both primary and secondary resources can be physical resources as well as virtual resources (e.g., slices).

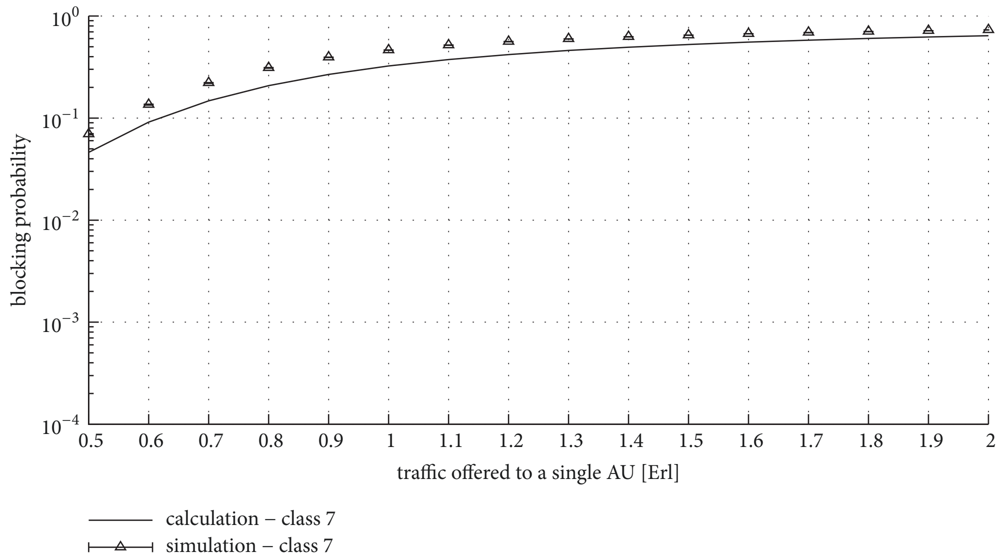

FIGURE 9: Blocking probability in primary resources (slice) no. 4 for traffic class no. 7 from Table 1. Traffic offered per single AU calculated for initial value of demanded AUs (before compression).

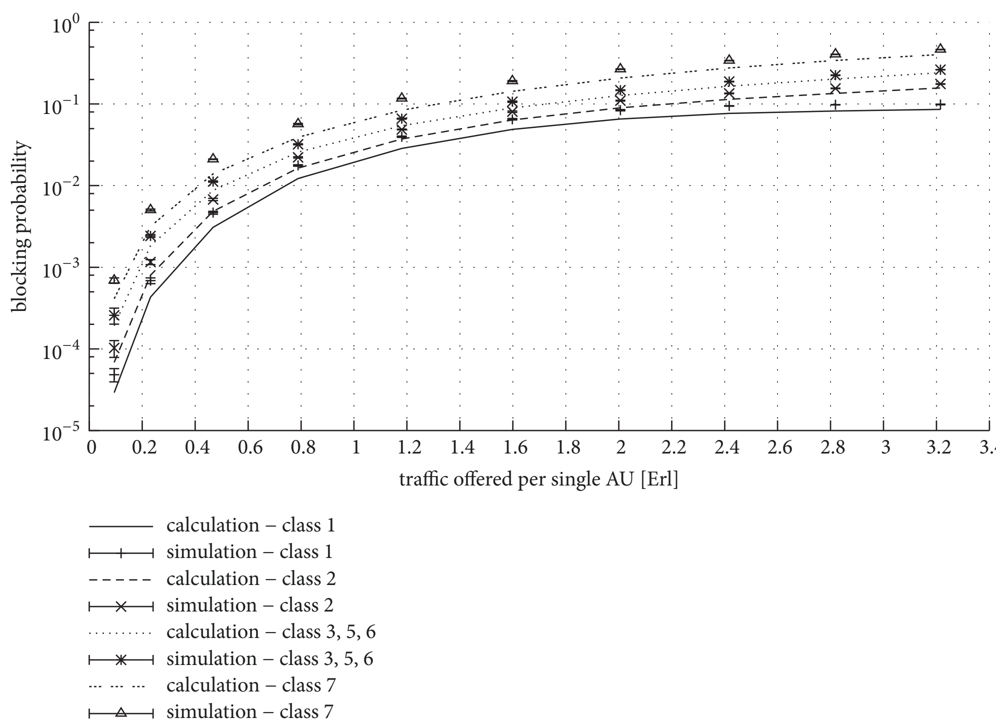

FIGURE 10: Blocking probability in secondary resources (slice) no. 0 for traffic classes nos. 1, 2, 3, 5, 6, and 7 from Table 1. Traffic offered per single AU calculated for initial value of demanded AUs (before compression).

The model takes into consideration three possible types of traffic: Erlang, Engset, and Pascal traffic. The model is based on a generalization of Hayward's concept and its application to model systems with adaptive traffic with threshold compression. The proposed model is an approximate model and therefore the results of the analytical modeling are compared with the results of the simulation experiments. The results obtained from the comparison prove good accuracy of the proposed model that is independent of both the number and values of introduced thresholds and of the number of classes and the type of offered traffic.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work is supported by the Polish Ministry of Science and Higher Education, Grant 08/82/DSPB/8224.

References

[1] M. Stasiak, M. Głąbowski, A. Wis'niewski, and P. Zwierzykowski, *Modeling and Dimensioning of Mobile Networks*, Wiley, 2011.

[2] G. Bretschneider, "Die berechnung von leitungsgruppen für überfliessenden verkehr in fernsprechwählanlagen," *Nachrichtentechnische Zeitung*, no. 11, pp. 533–540, 1956.

[3] R. I. Wilkinson, "Theories of toll traffic engineering in the USA," *Bell System Technical Journal*, vol. 40, pp. 421–514, 1956.

[4] Y. Rapp, "Planning of junction network in a multi-exchange area," in *Proceedings of the 4th International Teletraffic Congress*, vol. 1964, London, UK.

[5] B. Wallström, "A distribution model for telephone traffic with varying call intensity, including overflow traffic," *Ericsson Technics*, vol. 20, no. 2, pp. 183–202, 1964.

[6] U. Herzog, "Die exakte Berechnung des Streuwertes von überlaufverkehrhinter Koppelanordnungen beliebiger Stufenzahl mit vollkommener bzw. unvollkommener Erreichbarkeit," *International Journal of Electronics And Communications*, vol. 20, no. 3, 1966.

[7] R. Schehrer, "On the exact calculation of overflow systems," in *Proceedings of the in Proceedings of 6th International Teletraffic Congress*, Munich, Germany, 1970, 1970, https://itc-conference.org/_Resources/Persistent/21594392c496c4c0193b4864ec860502ebdec2bc/schehrer70.pdf.

[8] G. Bretschneider, "Extension of the equivalent random method to smooth traffics," in *Proceedings of the in Proceedings of 7th International Teletraffic Congress*, Stockholm, 1973.

[9] R. Schehrer, "On the calculation of overflow systems with a finite number of sources and full availiable groups," *IEEE Transactions on Communications*, vol. 26, no. 1, pp. 75–82, 1978.

[10] A. A. Fredericks, "Congestion in blocking systems—a simple approximation technique," *Bell Labs Technical Journal*, vol. 59, no. 6, pp. 805–828, 1980.

[11] J. F. Shortle, "An equivalent random method with hyper-exponential service," *Performance Evaluation*, vol. 57, no. 3, pp. 409–422, 2004.

[12] E. W. M. Wong, A. Zalesky, Z. Rosberg, and M. Zukerman, "A new method for approximating blocking probability in overflow loss networks," *Computer Networks*, vol. 51, no. 11, pp. 2958–2975, 2007.

[13] V. Iversen, "Teletraffic engineering handbook," Tech. Rep., Technical University of Denmark, 2010.

[14] C. G. Park and D. H. Han, "Comparisons of loss formulas for a circuit group with overflow traffic," *Journal of Applied Mathematics & Informatics*, vol. 30, no. 1-2, pp. 135–145, 2012.

[15] P. Kühn and M. E. Mashaly, "Multi-server, finite capacity queuing system with mutual overflow," in *Proceedings of the 2nd European Teletraffic Seminar*, M. Fiedler, Ed., Karlskrona, Sweden, 2013.

[16] Y. Chan, J. Guo, E. W. Wong, and M. Zukerman, "Surrogate models for performance evaluation of multi-skill multi-layer overflow loss systems," *Performance Evaluation*, vol. 104, pp. 1–22, 2016.

[17] L.-R. Hu and S. Rappaport, "Personal communications systems using multiple hierarchical cellular overlays," in *Proceedings of the 1994 3rd IEEE International Conference on Universal Personal Communications*, pp. 397–401, San Diego, CA, USA.

[18] X. Lagrange and P. Godlewski, "Performance of a hierarchical cellular network with mobility-dependent hand-over strategies," in *Proceedings of the IEEE 46th Vehicular Technology Conference*, vol. 3, pp. 1868–1872, May 1996.

[19] S.-P. Chung and J.-C. Lee, "Performance analysis and overflowed traffic characterization in multiservice hierarchical wireless networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 904–918, 2005.

[20] J. S. Kaufman and K. M. Rege, "Blocking in a shared resource environment with batched Poisson arrival processes," *Performance Evaluation*, vol. 24, no. 4, pp. 249–263, 1996.

[21] I. Moscholios, J. Vardakas, M. Logothetis, and A. Boucouvalas, "Congestion probabilities in a batched Poisson multirate loss model supporting elastic and adaptive traffic," *annals of telecommunications - annales des télécommunications*, vol. 68, no. 5-6, pp. 327–344, 2013.

[22] L. E. N. Delbrouck, "On the steady-state distribution in a service facility carrying mixtures of traffic with different peakedness factors and capacity requirements," *IEEE Transactions on Communications*, vol. 31, no. 11, pp. 1209–1211, 1983.

[23] E. A. van Doorn and F. J. M. Panken, "Blocking probabilities in a loss system with arrivals in geometrically distributed batches and heterogenous service requirements," *IEEE/ACM Transactions on Networking*, vol. 1, no. 6, pp. 664–667, 1993.

[24] M. Głąbowski, K. Kubasik, and M. Stasiak, "Modeling of systems with overflow multi-rate traffic," *Telecommunication Systems*, vol. 37, no. 1–3, pp. 85–96, 2008.

[25] M. Głąbowski, K. Kubasik, and M. Stasiak, "Modelling of systems with overflow multi-rate traffic and finite number of traffic sources," in *Proceedings of the 6th International Symposium Communication Systems, Networks and Digital Signal Processing (CSNDSP '08)*, pp. 196–199, Graz, Austria, July 2008.

[26] Q. Huang, K.-T. Ko, and V. B. Iversen, "Approximation of loss calculation for hierarchical networks with multiservice overflows," *IEEE Transactions on Communications*, vol. 56, no. 3, pp. 466–473, 2008.

[27] A. Lotze, "History and development of grading theory," in *Proceedings of the in Proceedings of 5th International Teletraffic Congress*, pp. 148–161, New York, NY, USA, 1967.

[28] J. S. Kaufman, "Blocking in a shared resource environment," *IEEE Transactions on Communications*, vol. 29, no. 10, pp. 1474–1481, 1981.

[29] J. Roberts, "A service system with heterogeneous user requirements — application to multi-service telecommunications systems," in *Proceedings of the Performance of Data Communications Systems and their Applications*, G. Pujolle, Ed., pp. 423–431, Amsterdam, Netherlands, 1981.

[30] M. Głąbowski, "Modelling of state-dependent multirate systems carrying BPP traffic," *Annals of Telecommunications-Annales des Télécommunications*, vol. 63, no. 7-8, pp. 393–407, 2008.

[31] M. Głąbowski, M. Stasiak, and J. Weissenberg, "Properties of Recurrent Equations for the Full-Availability Group with BPP Traffic," *Mathematical Problems in Engineering*, vol. 2012, Article ID 547909, 17 pages, 2012.

[32] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Two-dimensional convolution algorithm for modelling multiservice networks with overflow traffic," *Mathematical Problems in Engineering*, vol. 2013, Article ID 852082, 18 pages, 2013.

[33] M. Głąbowski, S. Hanczewski, and M. Stasiak, "Modelling of Cellular Networks with Traffic Overflow," *Mathematical Problems in Engineering*, vol. 2015, Article ID 286490, 15 pages, 2015.

[34] M. Głąbowski, S. Hanczewski, M. Stasiak, and J. Weissenberg, "Modeling Erlang's Ideal Grading with multirate BPP traffic," *Mathematical Problems in Engineering*, vol. 2012, Article ID 456910, 35 pages, 2012.

[35] Q. Huang, Y.-C. Huang, K.-T. Ko, and V. B. Iversen, "Loss performance modeling for hierarchical heterogeneous wireless networks with speed-sensitive call admission control," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2209–2223, 2011.

[36] B. M. Bakmaz and M. R. Bakmaz, "Solving some overflow traffic models with changed serving intensities," *AEÜ - International Journal of Electronics and Communications*, vol. 66, no. 1, pp. 80–85, 2012.

[37] M. Wang, S. Li, E. W. M. Wong, and M. Zukerman, "Performance analysis of circuit switched multi-service multi-rate networks with alternative routing," *Journal of Lightwave Technology*, vol. 32, no. 2, Article ID 6658839, pp. 179–200, 2014.

[38] J. Postel, "Transmission Control Protocol," Tech. Rep. RFC 793 (INTERNET STANDARD), Internet Engineering Task Force, 1981, 1981, http://www.ietf.org/rfc/rfc793.txt.

[39] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," Tech. Rep. RFC3550 (INTERNET STANDARD), Internet Engineering Task Force, 2003, 2003, http://www.ietf.org/rfc/rfc3550.txt.

[40] G. M. Stamatelos and V. N. Koukoulidis, "Reservation-based bandwidth allocation in a radio ATM network," *IEEE/ACM Transactions on Networking*, vol. 5, no. 3, pp. 420–428, 1997.

[41] T. Bonald and J. Virtamo, "A recursive formula for multirate systems with elastic traffic," *IEEE Communications Letters*, vol. 9, no. 8, pp. 753–755, 2005.

[42] T. Bonald, A. Proutière, and J. Virtamo, "A queueing analysis of max-min fairness, proportional fairness and balanced fairness," *Queueing Systems*, vol. 53, no. 1, pp. 65–84, 2006.

[43] J.-P. Haddad and R. R. Mazumdar, "Congestion in large balanced multirate networks," *Queueing Systems*, vol. 74, no. 2-3, pp. 333–368, 2013.

[44] T. Bonald and A. Proutiere, "Insensitive bandwidth sharing," in *Proceedings of the GLOBECOM 2002 - IEEE Global Communications Conference*, pp. 2659–2663, Taipei, Taiwan.

[45] S. Rácz, B. P. Gerö, and G. Fodor, "Flow level performance analysis of a multi-service system supporting elastic and adaptive services," *Performance Evaluation*, vol. 49, no. 1-4, pp. 451–469, 2002.

[46] J. S. Kaufman, "Blocking with retrials in a completely shared resource environment," *Performance Evaluation*, vol. 15, no. 2, pp. 99–116, 1992.

[47] J. Kaufman, "Blocking in a completely shared resource environment with state dependent resource and residency requirements," in *Proceedings of the IEEE INFOCOM '92: The Conference on Computer Communications*, vol. 3, pp. 2224–2232, IEEE Computer Society Press, Los Alamitos, CA, USA, May 1992.

[48] I. D. Moscholios, M. D. Logothetis, and G. K. Kokkinakis, "Connection-dependent threshold model: A generalization of the Erlang multiple rate loss model," *Performance Evaluation*, vol. 48, no. 1-4, pp. 177–200, 2002.

[49] I. D. Moscholios, M. D. Logothetis, and P. I. Nikolaropoulos, "Engset multi-rate state-dependent loss models," *Performance Evaluation*, vol. 59, no. 2-3, pp. 247–277, 2005.

[50] I. D. Moscholios, M. D. Logothetis, and G. K. Kokkinakis, "Call-burst blocking of ON-OFF traffic sources with retrials under the complete sharing policy," *Performance Evaluation*, vol. 59, no. 4, pp. 279–312, 2005.

[51] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Modeling product-form state-dependent systems with BPP traffic," *Performance Evaluation*, vol. 67, no. 3, pp. 174–197, 2010.

[52] M. Sobieraj, M. Stasiak, J. Weissenberg, and P. Zwierzykowski, "Analytical model of the single threshold mechanism with hysteresis for multi-service networks," *IEICE Transactions on Communications*, vol. 95, no. 1, pp. 120–132, 2012.

[53] M. Sobieraj, M. Stasiak, and P. Zwierzykowski, "Model of the Threshold Mechanism with Double Hysteresis for Multi-service Networks," in *Computer Networks*, vol. 291 of *Communications in Computer and Information Science*, pp. 299–313, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[54] X. Zhou, R. Li, T. Chen, and H. Zhang, "Network slicing as a service: Enabling enterprises' own software-defined cellular networks," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146–153, 2016.

[55] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network Slicing in 5G: Survey and Challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[56] P. Rost, A. Banchs, I. Berberana et al., "Mobile network architecture evolution toward 5G," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 84–91, 2016.

[57] X. Wu and Z. Ma, "Modeling and Performance Analysis of Cellular and Device-to-Device Heterogeneous Networks," in *Proceedings of the 2017 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Singapore, Singapore, December 2017.

[58] M. Stasiak, "Queuing systems for the internet," *IEICE Transactions on Communications*, vol. 99, no. 6, pp. 1234–1242, 2016.

[59] M. Głąbowski, S. Hanczewski, and M. Stasiak, "Modelling load balancing mechanisms in self-optimising 4G mobile networks with elastic and adaptive traffic," *IEICE Transactions on Communications*, vol. E99B, no. 8, pp. 1718–1726, 2016.

[60] S. Hanczewski, M. Stasiak, and P. Zwierzykowski, "Modelling of the access part of a multi-service mobile network with service priorities," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, 2015.

[61] M. Głąbowski, D. Kmiecik, and M. Stasiak, "Overflow of elastic traffic," in *Proceedings of the 1st International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications, CoBCom 2016*, Austria, September 2016.

[62] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Modelling overflow systems with distributed secondary resources," *Computer Networks*, vol. 108, pp. 171–183, 2016.

[63] M. Stasiak, "Blocking probability in a limited-availability group carrying mixture of different multichannel traffic streams," *Annales des Télécommunications*, vol. 48, no. 1-2, pp. 71–76, 1993.

[64] M. Głąbowski and M. Stasiak, "Multi-rate model of the group of separated transmission links of various capacities," in *Telecommunications and Networking—ICT 2004, 11th International Conference on Telecommunications, Fortaleza, Brazil, August 1–6, 2004, Proceedings*, J. N. de Souza, P. Dini, and P. Lorenz,

Eds., vol. 3124 of *Lecture Notes in Computer Science*, pp. 1101–1106, Springer, Berlin, Germany, 2004.

[65] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Analytical Modelling of Multi-tier Cellular Networks with Traffic Overflow," in *Computer Networks*, vol. 718 of *Communications in Computer and Information Science*, pp. 256–268, Springer International Publishing, Cham, Switzerland, 2017.

[66] ITU-T, "Calculation of the number of circuits in a group carrying overflow traffic," Recommendation E.521, 1988.

[67] T. Bonald and J. Roberts, "Internet and the Erlang formula," *ACM Computer Communications Review*, vol. 42, pp. 24–30, 2012.

[68] H. Akimaru and K. Kawashima, *Teletraffic: Theory and Application*, Springer, Berlin, Germany, 1999.

[69] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 16, no. 3-4, pp. 387–396, 1994.

[70] V. Paxson and S. Floyd, "Wide-area traffic: the failure of traffic modeling," in *Proceedings of the 1994 SIGCOMM Conference*, pp. 257–268, 1994.

[71] S. Floyd and V. Paxson, "Difficulties in simulating the Internet," *IEEE/ACM Transactions on Networking*, vol. 9, no. 4, pp. 392–403.

[72] J. Y. Hui, "Resource allocation in broadband networks," *Journal on Selected Areas in Communications*, vol. 6, no. 9, pp. 1598–1608, 1988.

[73] F. Kelly, *Notes on Effective Bandwidth*, University of Cambridge, 1996.

[74] J. Roberts, *Performance Evaluation and Design of Multiservice Networks, Final Report COST 224*, Commission of the European Communities, Brussels, Belgium, 1992.

[75] J. Roberts, V. Mocci, and I. Virtamo, *Broadband Network Teletraffic, Final Report of Action COST 242*, Commission of the European Communities, Springer, Berlin, Germany, 1996.

[76] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, pp. 968–981.

[77] I. Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 953–962, 1995.

[78] A. Pras, L. Nieuwenhuis, R. van de Meent, and M. Mandjes, "Dimensioning network links: a new look at equivalent bandwidth," *IEEE Network*, vol. 23, no. 2, pp. 5–10, 2009.

[79] I. D. Moscholios, M. D. Logothetis, and A. C. Boucouvalas, "Blocking probabilities of elastic and adaptive calls in the Erlang multirate loss model under the threshold policy," *Telecommunication Systems*, vol. 62, no. 1, pp. 245–262, 2016.

[80] I. D. Moscholios, M. D. Logothetis, J. S. Vardakas, and A. C. Boucouvalas, "Congestion probabilities of elastic and adaptive calls in Erlang-Engset multirate loss models under the threshold and bandwidth reservation policies," *Computer Networks*, vol. 92, part 1, pp. 1–23, 2015.

[81] E. Brockmeyer, H. Halstrom, and A. Jensen, "The life and works of A.K. Erlang," *Acta Polytechnica Scandinavica*, vol. 6, no. 287, 1960.