Modeling of Complex Non-Full-Availability Systems

Sławomir Hanczewski, Maciej Stasiak, and Michał Weissenberg

Faculty of Computing and Telecommunications, Poznan University of Technology, Poznan, Poland

Abstract—This article presents an analytical model of complex non-full-availability telecommunications systems. The high degree of accuracy of the model is demonstrated by a comparison with the results of simulation experiments. Due to the introduction of an availability parameter, this model may be used in the future for analyzing real systems, such as, for example, cloud computing infrastructure. This will be possible provided that the function combining the physical structure of the system and the availability parameter is specified. This problem will be addressed in our future work and will constitute the next stage of research undertaken by the authors.

Keywords—*analytical model, availability, non-full-availability systems.*

1. Introduction

From the traffic engineering perspective, telecommunications and computer systems that are currently under development are characterized by a peculiar but rarely noticed property, namely they constitute non-full-available systems. Non-full-availability is based on the fact that calls arriving at the input end of a system have no access to all of the system's resources. Instead, they only have access to a certain subset of those resources. In the literature, this subset is referred to as availability [1], [2]. Basically, systems of this type have existed in telecommunications from the very inception of the domain. Over time, however, the source of non-full-availability has changed. Originally, non-full-availability resulted from the limited nature of telecommunication networks' resources, e.g. outputs of communications networks and telephone exchange systems, where the number of such outputs exceeded the capacities of output groups. Today, non-full-availability of a system results primarily from the individual traffic management mechanisms that are used by networks operators to optimize the use of the available resources (e.g. resource reservation, compression, priorities and queuing) and, very frequently, from the wide area or distributed nature of the physical infrastructure. Non-full-availability of a system may be illustrated with an example of a dynamic resource reservation system [3], [4]. Its operating principle is based on the system's occupancy limit (boundary) adopted which, when exceeded, causes the resources of the system to be available only to certain pre-defined calls, at the expense of the remaining calls.

Concurrent use of a number of mechanisms means that the degree of dependence between the amount of resourcesavailable to a given call and the system occupation level becomes even higher. Physical infrastructure of a cloud computing system may serve as the best example of a modern non-full-availability solution. A high number of devices, the distributed nature of the system, the mechanisms used to optimize of the utilization of resources in cloud computing, the mechanisms that guarantee safe provision of the services, the algorithms employed to reduce power consumption, as well as the nature of the services offered [5]–[8] all make it possible for new calls to be supported and serviced by the resources in the cloud.

There is a number of models available in the literature concerned with analyzing current telecommunications systems. As a rule, however, these models deal with a single traffic management mechanism, while a real system is approximated by a model of a full-availability system (historically known as a full-availability group model), e.g. shown in [9]–[15]. Models of non-full-availability systems, such as the limited-availability group model (LAG) or Erlang's ideal grading (EIG), are used for that purpose less frequently [16]–[19].

Non-full-availability systems belong to a group of the so-called state-dependent systems [20], [11]. This dependence may result from the existence of the function that relates the number of busy resources in a system to the calls that arrive at the input, or, alternatively, dependence between the number of occupied system resources and the decision to admit a new call for service may exist. In the former case, we deal with systems with a state-dependent call stream. In the latter case, we deal with systems with a state-dependent service process. [20] presents general rules applicable to modeling this type of systems with the use of a one-dimensional Markov process, after the so-called state transition coefficient has been taken into account. This coherent methodology makes it possible to model very complicated systems as well, providing that the relevant transition coefficients may be determined.

This article presents an analytical model of a complex non-full-availability system. The structure of the system shows some non-full-availability features, i.e. it is composed of a group of separated resources that are non-full-availability systems themselves. It is the authors' opinion that the model may be relied upon, in the future, to model the

physical infrastructure of cloud computing solutions. The authors of the article are currently conducting research on the conditions under which this may become feasible.

The remaining part of the article is structured as follows. Section 2 presents the main assumptions for modeling state-dependent systems. Section 3 provides a description of the structure of this complex non-full-availability system and the corresponding analytical model proposed. On account of the approximate nature of the model, Section 4 presents a comparison between the results obtained with the analytical model and the results obtained during simulation experiments. The article is summarized in Section 5.

2. State-dependent Systems

2.1. Resources and Traffic in Telecommunications Networks

Today's telecommunications networks are primarily packet-switched networks based on the IP protocol. The analysis of a network at the packet level is complicated and time consuming. Papers [9], [21], [22] prove that it is possible to analyze packet networks at the call level, where calls correspond to packet streams (with constant or variable transmission rates) that are sent between a given source and the receiver. Thus, the network traffic represented has the properties of Poisson traffic flow. Thanks to this operation, it is possible to express the amount of resources of the system in dimensionless units, i.e. in resource allocation units (AU). This transition from dimensional units to dimensionless units is known as discretization. The said operation increases versatility of models, since the scope of their operation is not limited to the analysis of the link bit rate (in bps), but is expanded to include other system resources, such as storage memory, RAM memory or processors [23].

The above-mentioned discretization process will be illustrated with an example of a network resource (e.g. a link between routers) with the capacity of C bps. This resource services constant bit rate (CBR) streams and variable bit rate (VBR) streams. In the discretization process, the VBR streams are represented by streams with a constant bit rate, with the transmission speed equal to the so-called equivalent bandwidth (EB). EB for individual call classes may be determined either on the basis of the maximum bit rates or on the basis of heuristic algorithms, in line with the principle that the result of a VBR call service in the network is the same as the result of a CBR call service with the bit rate equal to EB [24]. Most frequently, these algorithms take into account such traffic parameters and networks system parameters as the maximum and the average bit rate of a call, bit rate variance, acceptable delay, jitter and quality of service (QoS) parameters adopted for a given service, etc. It is worth emphasizing at this point that the determination of the EB value on the basis of the maximum bit rates makes it possible to analyze the system under consideration for the worst conditions, i.e. for the highest offered traffic. The knowledge of EB values (the manner in which

EB is determined for individual packet streams is beyond the scope of the present article) for all traffic classes offered to the system makes it possible to determine the resource allocation unit of the system, as the traffic class is composed of all packet streams with similar parameters. The general assumption is that 1 AU is equal to the greatest common divisor of the equivalent bandwidths of all traffic classes that are offered to the system:

$$1 \text{ AU} \equiv c_{AU} = \text{GCD}(c_1, \ldots, c_i, \ldots, c_m) , \qquad (1)$$

where c_i (bps) is the equivalent bandwidth of class i, whereas m is the number of traffic classes offered to the system. The parameter c_{AU} determines the bit rate per a single AU.

After the determination of c_{AU} it is possible to express the capacity of resource V and the demands of individual traffic classes t_i in the AUs, where C is the resource capacity:

$$V = \left\lfloor \frac{C}{c_{AU}} \right\rfloor , \qquad (2)$$

$$t_i = \left\lceil \frac{c_i}{c_{AU}} \right\rceil . \qquad (3)$$

Call classes are described by:

- λ_i intensities of call stream of class i ($1 \le i \le m$); this parameter characterizes the frequency at which new calls arrive,

- μ_i intensities of service stream of class i ($1 \le i \le m$); this parameter characterizes call service time,

- a_i average intensity of traffic of class i (in relation to calls):

$$a_i = \lambda_i / \mu_i. \qquad (4)$$

To simplify the analysis of modern networks, an additional assumption may be made that 1 AU corresponds to the bit rate of 1 bps (or 1 kbps):

$$c_{AU} = 1 \text{ bps.} \qquad (5)$$

This choice of the AU value causes the demands of individual classes – Eq. (3) and the capacity of the system – Eq. (2) to be equal, in terms of their value, with the demanded bit rates and the total capacity of the system:

$$c_i = t_i , \quad C = V. \qquad (6)$$

2.2. General Model of State-dependent Systems

The general method for modeling state-dependent systems by a one-dimensional Markov process is presented in [20]. The dependence of a system on its state (i.e. resource occupancy) may be a result of the following:

- limited number of traffic sources – in this case, we deal with systems with a state-dependent call stream – a single source can generate a new call only when it is unoccupied. As a consequence, the ration the number of calls that arrive at the input of the sys-

between tem within a given time unit and the number of calls admitted earlier is changed. For instance, for Engset traffic, this number decreases along with an increase in the number of occupied traffic sources;

- system structure – in this case, we deal with a system with a state-dependent service stream. The structure of the system and the call service process may cause only a part of the resources to be available, just as it is the case with switching networks, for example. The structure of the connections within the network and the calls that are currently serviced may then affect the setting up of a connection between a free input and a free output of the network, rendering it impossible;

- algorithms operating the call admission control function – similarly as in the former case, we deal with a state-dependent service stream. Call admission control algorithms may prevent new calls from being serviced under certain operating conditions, e.g. based on a dynamic resource reservation in relation to the total number of busy AUs in the system.

In state-dependent systems, the transition between states depends not only on the parameters λ_i and μ_i, but additionally on the so-called conditional transition coefficient $\sigma_i(n)$, where i is the number of a call classes, and n is the number of busy AUs in the system. This is presented, in a simplified form, in Fig. 1.
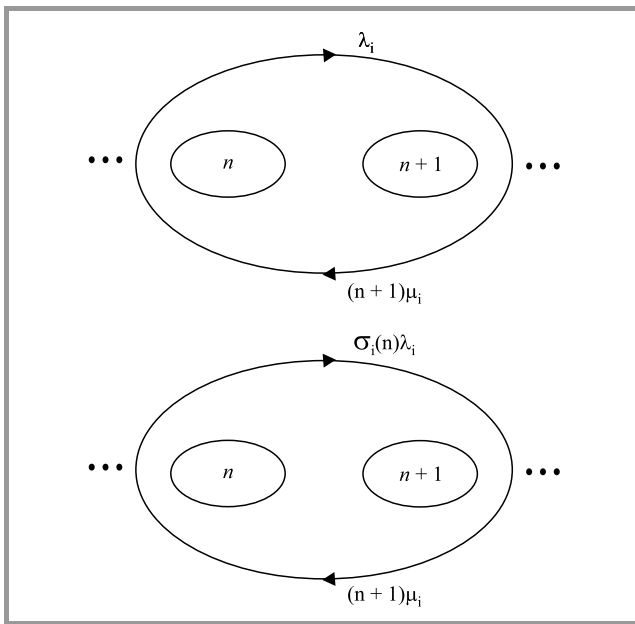

Fig. 1. Transition between states of Markov process: (a) state-independent system, (b) state-dependent system.

It is proved in [20] that in the case of systems in which state-dependence is the result of a number of different and independent sources, the total conditional transition coefficient may be determined in the following way:

$$\sigma_i(n) = [\sigma_i(n)]^R [\sigma_i(n)]^A [\sigma_i(n)]^P, \tag{7}$$

where:

- $[\sigma_i(n)]^R$ is the conditional transition coefficient of class i in occupancy state n, resulting from the properties of a call stream of this class,

- $[\sigma_i(n)]^A$ is the conditional transition coefficient of class i in occupancy state n, resulting from the structure of the system,

- $[\sigma_i(n)]^P$ is the conditional transition coefficient of class i in occupancy state n, resulting from the function of new call admission in the system.

The knowledge of the values of all $\sigma_i(n)$ parameters makes it possible to determine occupancy distribution within the system with the use of the following formula [20]:

$$n[P_n]_V = \sum_{i=1}^{m} a_i t_i [\sigma_i(n)]_V [P_{(n-t_i)}]_V, \tag{8}$$

where $[P_n]_V$ is the probability of occupancy n AUs in the system.

3. Model of a Complex Non-full-availability System

3.1. Structure of Complex Non-full-availability Systems

The analytical model proposed in the article was developed for a system that is presented, in its simplified form, in Fig. 2. The assumption was that the system was composed of k separated resources (subgroups) with the capacity of f AUs. The total capacity of the system is:

$$V = kf. \tag{9}$$

The division of all AUs within the system into separated full-availability resources renders the system a non-full-available solution. The literature of the subject describes its analytical model, known as the limited-availability group model [25]. In this paper, we assumed that each of k separated resources shows the properties of a non-full-availability system. Non-full-availability of a single separated resource may stem from its internal structure or from the call admission control algorithm. The literature presents a small number of analytical models pertaining to non-full-availability systems, e.g. Palm-Jacobeus formula, modified Palm-Jacobeus formula or the EIG model [1].
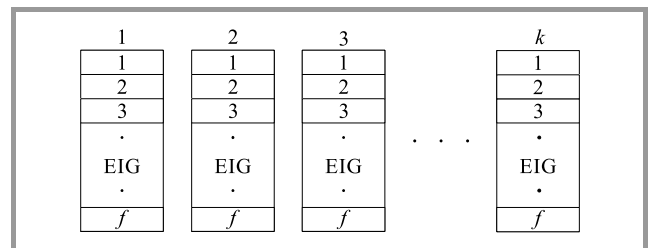

Fig. 2. Simplified diagram of the structure of a complex non-full-availability system.

In analytical models, non-full-availability of resources is characterized by a parameter knowns as availability d_i ($1 \leq i \leq m$) [1], [26]. This parameter determines the number of AUs in the system to which calls of a given class have access. It is unique for each class of calls (and it may assume different values for each of those classes), while its value is based on the structure of the system and on the call admission control mechanism. It should be stressed that if calls of class i have availability equal to d_i AUs of the system, this does not necessarily mean that all of them have access to the same AUs. Typically, they have access to different AUs of the system. Calls or, to be more precise, their sources that have access to the same AUs in the system are called a load group. In EIG, the number of load groups (for calls of class i) is equal to the number of choices of d AUs from V:

$$g_i = \binom{V}{d_i}. \tag{10}$$

In real world systems, this number is typically lower. Figure 3 shows a simple single-service system with the capacity of 3 AUs and availability of $d = 2$ AUs that illustrates the idea of availability and presents an example of a blocking state.
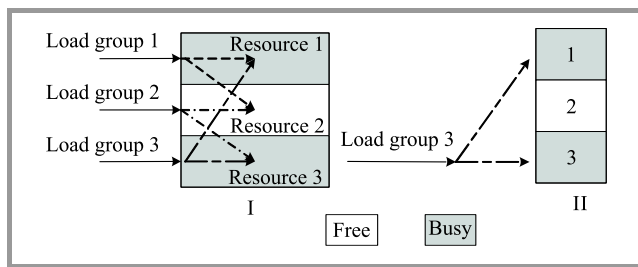

Fig. 3. Non-full-availability system: (a) the availability concept, (b) occupied load group no. 3.

The assumption is that in the complex non-full-availability system under consideration, new calls are admitted according to the following algorithm. Assume that a call of class i that demands t_i AUs for service arrives at the input to the system. In the first instance, the number of unoccupied AUs in the entire system is checked. If the number of free AUs in the entire system is lower than t_i AUs, then the call is lost. When this is not the case, a further check is performed to verify whether in a single separated resource there exist at least t_i unoccupied AUs. This condition is particularly important, as calls may be supported only by AUs that belong to one separated resource. If such resources are not available, then the call is rejected. In the case where the system has separated resources that have at least t_i free AUs available, one of the available resources is chosen randomly and the system checks whether the resource may admit the call under consideration for service. The absence of a possibility to admit a new call may result from the fact that this new call has no access to (i.e. cannot occupy) free AUs in the resource. If the selected separated resource does not offer such a possibility, then the next one

is selected. In the worst case, all k separated resources will be checked. If none of k resources can service the new call, the call is lost again.

3.2. Analytical Model

To determine occupancy distribution in the system under consideration, as given by Eq. (8), and, in consequence the blocking probability, it is necessary to determine the values of the conditional transition coefficient $\sigma_i(n)$ for all call classes offered and for all system occupancy states. The state-dependence of the service stream and, consequently, the non-full-availability in the system under consideration, results from two factors:

- all AUs in the system are divided into k identical separated resources (subsystems), while a call may be completely serviced by one separated resource only,

- each separated resource has limited access to its own AUs which leads, in consequence, to a situation in which - despite a sufficient number of free AUs – the call cannot be admitted for service due to the lack of access to free AUs.

On the basis of Eq. (7), we can write:

$$\sigma_i(n) = [\sigma_i(n)]^{SZ}[\sigma_i(n)]^S, \tag{11}$$

where $[\sigma_i(n)]^{SZ}$ is the transition coefficient that results from the division of the AUs of the system into the separated resources, whereas $[\sigma_i(n)]^S$ results from the non-full-availability of a single separated resource.

Now, let us consider the manner in which parameters $[\sigma_i(n)]^{SZ}$ and $[\sigma_i(n)]^S$ may be determined.
The values of the coefficient $[\sigma_i(n)]^{SZ}$ may be determined as follows. Since calls may be completely serviced only by a single separated resource, this resource has to have at least t_i unoccupied AUs. To satisfy this condition, an appropriate distribution of free AUs between all separated resources is required. An unfavorable distribution of free AUs within the system may be followed by new calls being rejected, as presented in Fig. 4. The said figure shows the distribution of three free AUs in a system that is composed of two separated resources. If, at the input to this system, a call that demands 3 AUs arrives, then it may be admitted for service only in the case of 2 from 4 potential distributions of unoccupied AUs within the entire system.
To determine the conditional transition coefficient, the dependence proposed in [25] can be used:

$$[\sigma_i(n)]_V^{SZ} = \frac{F(V-n,k,f) - F(V-n,k,t_i-1)}{F(V-n,k,f)} , \tag{12}$$

where the function $F(x,k,f)$ is given by:

$$F(x,k,f) = \sum_{i=0}^{\lfloor \frac{x}{f+1} \rfloor} (-1)^i \binom{k}{i} \binom{x+k-1-i(f+1)}{k-1} . \tag{13}$$
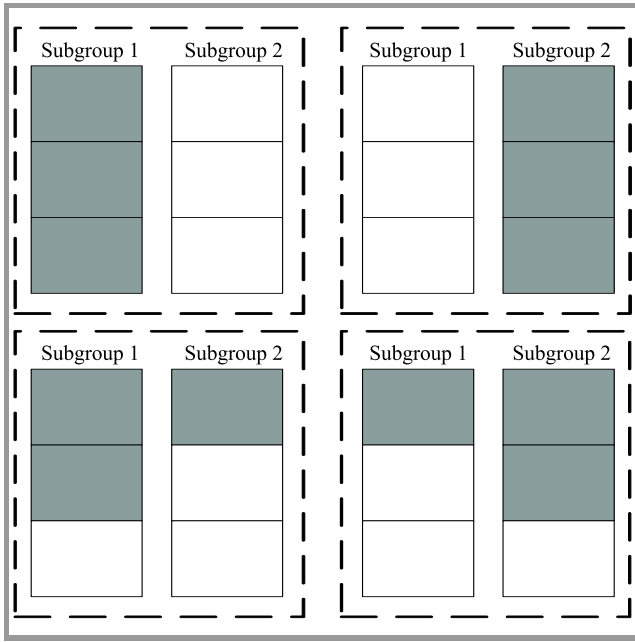
Fig. 4. Distribution of free AUs in separated resources.

Parameter $[\sigma_i(n)]_V^{SZ}$ is calculated based on the assumption that separated resources are of the full-availability variety. This assumption is based on the new connection setup algorithm implemented. According to this algorithm, the separated resources are checked for free AUs first, without checking whether there is access to them.

Each of the separated resources may be the source for non-full-availability as well. Systems in which calls have limited access to their own AUs may be modeled by EIG [26], [1]. In this model, the conditional transition coefficient is determined with the use of the following formula:

$$[\sigma_i(n)]_f^{\text{EIG}} = 1 - \sum_{d_i-t_i+1}^{z} \frac{\binom{d_i}{x}\binom{f-d_i}{n-x}}{\binom{f}{n}} , \qquad (14)$$

where: $z = n - t_i$, if $(d_i - t_i + 1) \le (n - t_i) < d_i$, $z = d_i$, if $(n - t_i) \ge d_i$.

In order to substitute Eqs. (12) and (14) into Eq. (11), two problems have to be solved first. Firstly, the occupancy of a single separated resource changes from 1 to f, while the occupancy of the whole of the system changes from 1 to V. Secondly, the call admission algorithm allows all separated resources that have at least t_i free AUs to be checked. In the proposed analytical model, the first problem was solved in the following way: since the assumption was that the choice as to the separated resources was random, then it could be adopted that approximately (rough estimate) each of them had equal loads and, what followed, the average number of occupied AUs in each resource was equal to $\frac{n}{k}$, where n was the occupancy state in the entire system. The other problem, in turn, was solved thanks to the adoption of the assumption that since service admission of a new call depends on the state of each of the separated resources, then

product dependence must exist between the transition coefficients in each of the separated resources and the resultant transition coefficient $[\sigma_i(n)]^S$. The proposed Eq. (15) was defined as follows: since the call admission algorithm for new calls assumes the possibility of checking all separated resources, then the admission of a new call is possible if such a possibility exists in just one of them:

$$[\sigma_i(n)]^S = 1 - \left(1 - \sigma_i\left(\frac{n}{k}\right)\right)^k . \qquad (15)$$

By taking into consideration Equations (12), (14) and (15), we can ultimately write:

$$\sigma_i(n) = [\sigma_i(n)]^{SZ} \sigma_i(n)^S =$$
$$= [\sigma_i(n)]^{SZ}\left[1 - \left(1 - \sigma_i\left(\frac{n}{k}\right)\right)^k\right] =$$
$$= \frac{F(V-n,k,f) - F(V-n,k,t_i-1)}{F(V-n,k,f)} \times$$
$$\left[1 - \left[\frac{\sum\limits_{d_i-t_i+1}^{z}\binom{d_i}{x}\binom{f-d_i}{\lfloor\frac{n}{k}\rfloor-x}}{\binom{f}{\lfloor\frac{n}{k}\rfloor}}\right]^k\right] . \qquad (16)$$

4. Results

To verify the proposed analytical model, the results obtained with the use of the model were compared with the results of the simulation. For this particular purpose, a simulation model of the non-full-availability system under consideration was devised and implemented in the C++ language. The simulator used the event scheduling methodology. To determine a single measurement, 5 series of simulations with 1,000,000 calls of the class that demanded the highest number of AUs for series each, were performed. The results obtained are presented in the form of a function of traffic offered to a single AU of the system:

$$a = \frac{\sum\limits_{i=1}^{m} a_i t_i}{V} . \qquad (17)$$

Another assumption was that the total traffic offered was divided between individual call classes in the following manner: $a_1 t_1 : a_2 t_2 : ... : a_m t_m = 1 : 1 : ... : 1$.

Figure 5 shows the result for the following system: $k = 3$, $f = 25$ AUs. The system was offered three classes of calls ($m = 4$) that demanded 1, 2, 3 and 4 AUs for service, respectively, while the availabilities were 8, 10, 12 and 15 AUs (Case 1). Figure 6 shows the result for the following system: $k = 3$, $f = 15$ AUs. The system was offered three classes of calls ($m = 3$) that demanded 1, 2 and 3 AUs for service, respectively, while the availabilities were 6, 8 and 10 AUs (Case 2). Figure 8 shows the result for the following system: $k = 3$, $f = 21$ AUs. The system was offered three classes of calls ($m = 3$) that demanded 1, 2
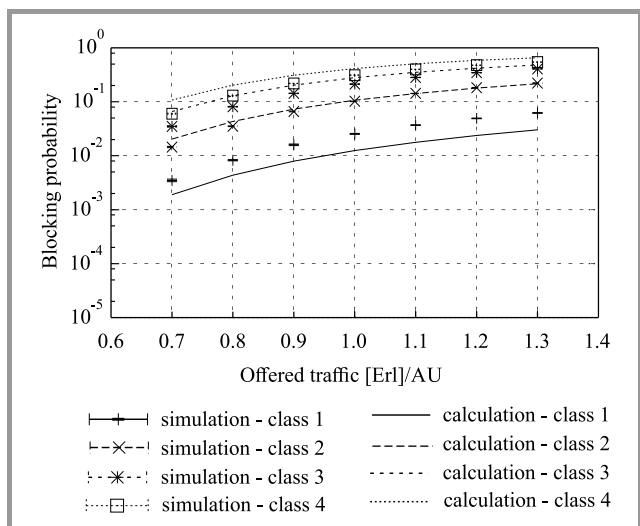
Fig. 5. Blocking probability in a complex non-full-availability system (Case 1).
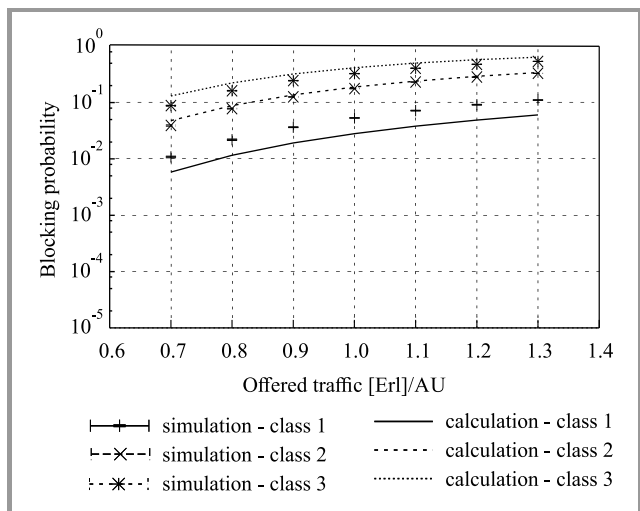

Fig. 6. Blocking probability in a complex non-full-availability system (Case 2).
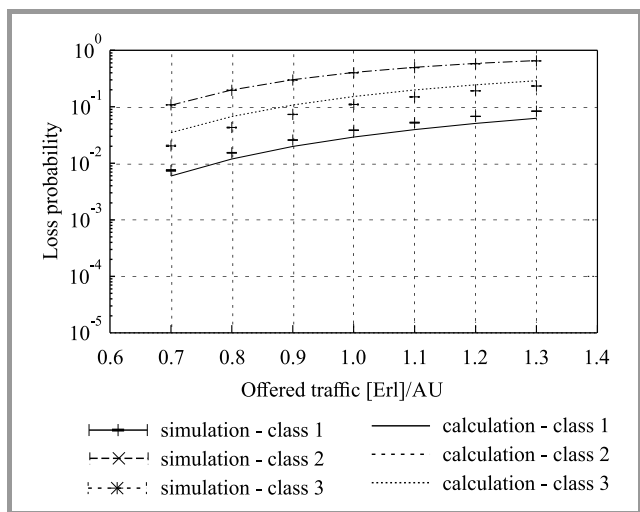

Fig. 7. Blocking probability in a complex non-full-availability system (Case 3).

and 5 AUs for service, respectively, while the availabilities were 7, 10 and 18 AUs (Case 3).

The model presented is an extension of the research described in article [27]. The previous model, in the scenario with low availability for the class requesting the least number of resources, showed some significant inaccuracies. Figure 5 shows the results obtained using models described in [27], [4], and the model described in the current article in connection with the following system: $k = 3$, $f = 24$ AUs. The system was offered three classes of calls ($m = 3$) that demanded 1, 3 and 4 AUs for service, respectively, while the availabilities were 5, 10 and 13 AUs (Case II). Thanks to the modification, the problem of inaccuracies in the youngest class has been eliminated.
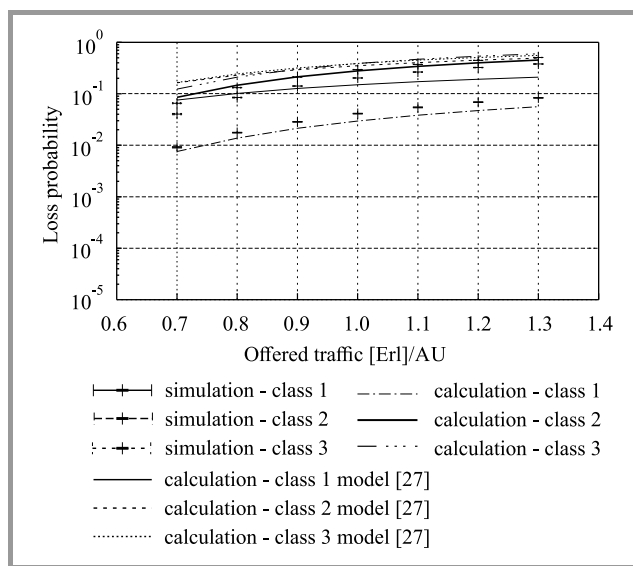

Fig. 8. Blocking probability in a complex non-full-availability system (Case 4) calculate by 2 models.

The analytical model proposed in the article is of the approximate variety. However, the presented results of the study show that the model is efficient at approximating the non-full- availability system under consideration. The highest error rate under investigation occurs for the class that demands the lowest number of AUs. However, this inaccuracy is much lower than that in the model presented in [27]. This error is based on a certain underestimation of the blocking probability, but the repeatability of this error makes it possible to introduce a relevant adjustment to the model. The accuracy of the model, as confirmed in the study, makes it possible rely on the model for modeling non-full-availability systems, such as those used in cloud computing. This will be feasible, however, only when appropriate equations that make it possible to determine the availability in such systems are proposed. This will be the subject of further research the authors intend to conduct.

5. Summary

This article presents the architecture of a complex non-full-availability system and its analytical model developed by

the authors, enabling to determine loss coefficients and the blocking probability in this system. Though the proposed model is of the approximate nature, the results obtained in the study clearly indicate that is its highly accurate. The approximation errors that occur in the results, and are reported in the investigation, are of the fixed nature, which makes it possible to introduce simple corrections and adjustments that would nullify these differences during the further development stages. The authors are of the opinion that further research into complex non-full-availability systems is of utmost importance due to the dynamic growth of IT services and increasingly more sophisticated systems that are offered to users. The model presented in the article may become, in the future, a useful tool in modeling real telecommunications systems, such as data centers or cloud computing solutions. The fact that it requires a small amount of computer resources to perform calculations and is characterized by a relatively short time of operation is an additional advantage of the model presented.

Acknowledgements

References

[1] M. Głąbowski, S. Hanczewski, M. Stasiak, and J. Weissenberg, "Modeling Erlang's Ideal Grading with multi-rate BPP traffic", *Mathem. Problems in Engin.*, vol. 2012, no. 2, 2012 (DOI: 10.1155/2012/456910).

[2] A. Lotze, "History and development of grading theory", in *Proc. 5th International Teletraffic Congress ITC5*, New York, USA, 1967, pp. 148–161, 1967.

[3] J. Kaufman, "Blocking with retrials in a completely shared resource environment", *J. of Perform. Eval.*, vol. 15, no. 2, pp. 99–116, 1992 (DOI: 10.1016/0166-5316(92)90058-O).

[4] M. Stasiak and M. Głąbowski, "A simple approximation of the link model with reservation by a one-dimensional Markov chain", *Perform. Eval.*, vol. 41, no. 2–3, pp. 195–208, 2000 (DOI: 10.1016/S0166-5316(00)00008-0).

[5] "Vmware and your cloud. Maximize IT Agility to Drive Business Agility", Tech. Rep., VMware, 2011 [Online]. Available: https://www.vmware.com/content/dam/digitalmarketing/vmware/en/files/images/microsite/whiteboard/vmcloud.pdf

[6] "Future of cloud computing", Tech. Rep., Google, 2018 [Online]. Available: https://services.google.com/fh/files/misc/futurecloudcomputing.pdf

[7] D. Kapil, P. Tyagi, S. Kumar, and V. P. Tamta, "Cloud computing: Overview and research issues", in *Proc. Int. Conf. on Green Inform. ICGI 2017*, Fuzhou, China, 2017, pp. 71–76 (DOI: 10.1109/ICGI.2017.18).

[8] F. Shakeel and S. Sharma, "Green cloud computing: A review on efficiency of data centres and virtualization of servers", in *Proc. Int. Conf. on Comput., Commun. and Autom. ICCCA 2017*, Greater Noida, India, 2017, pp. 1264–1267 (DOI: 10.1109/CCAA.2017.8230012).

[9] T. Bonald and J. Roberts, "Internet and the Erlang formula", *ACM Comp. Commun. Rev.*, vol. 42, pp. 23–30, 2012 [Online]. Available: https://hal-imt.archives-ouvertes.fr/hal-00941783/PDF/BR11.pdf

[10] G. Fodor, S. Rácz, and M. Telek, "On providing blocking probability and throughput guarantees in a multi-service environment", *Int. J. of Commun. Syst.*, vol. 15, no. 4, 257–285, 2002 (DOI: 10.1002/dac.532).

[11] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Modelling overflow systems with distributed secondary resources", *Comp. Networks*, vol. 108, pp. 171–183, 2016 (DOI: 10.1016/j.comnet.2016.08.015).

[12] S. Hanczewski, M. Stasiak, and J. Weissenberg, "Queueing model of a multi-service system with elastic and adaptive traffic", *Comp. Networks*, vol. 147, pp. 146–161, 2018 (DOI: 10.1016/j.comnet.2018.09.023).

[13] I. D. Moscholios, J. S. Vardakas, M. D. Logothetis, and A. C. Boucouvalas, "Congestion probabilities in a batched Poisson multirate loss model supporting elastic and adaptive traffic", *Annals of Telecommun.*, vol. 68, no. 5, pp. 327–344, 2013 (DOI: 10.1007/s12243-012-0326-7).

[14] D. Parniewicz, M. Stasiak, J. Wiewióra, and P. Zwierzykowski, "An approximate model of the WCDMA interface servicing a mixture of multi-rate traffic streams with priorities", in *Computer Performance Engineering. 5th European Performance Engineering Workshop, EPEW 2008, Palma de Mallorca, Spain, September 24-25, 2008, Proceedings*, N. Thomas and C. Juiz, Eds., *LNCS*, vol. 5261, pp. 168–180. Berlin Heidelberg: Springer, 2008 (DOI: 10.1007/978-3-540-87412-6_13).

[15] G. M. Stamatelos and V. N. Koukoulidis, "Reservation-based bandwidth allocation in a radio ATM network", *IEEE/ACM Trans. on Network.*, vol. 5, no. 3, pp. 420–428, 1997 (DOI: 10.1109/90.611106).

[16] M. Głąbowski, S. Hanczewski, and M. Stasiak, "Modelling load balancing mechanisms in self-optimising 4G mobile networks with elastic and adaptive traffic", *IEICE Trans. on Commun.*, vol. E99.B, no. 8, pp. 1718–1726, 2016 (DOI: 10.1587/transcom.2015CCP0016).

[17] M. Głąbowski, "Recurrent calculation of blocking probability in multiservice switching networks", in *Proc. of Asia-Pacific Conf. on Commun.*, Busan, South Korea, 2006 (DOI: 10.1109/APCC.2006.255964).

[18] M. Głąbowski, S. Hanczewski, and M. Stasiak, "Modelling of cellular networks with traffic overflow", *Mathem. Problems in Engin.*, vol. 2015, Article ID 286490, 2015 (DOI: 10.1155/2015/286490).

[19] S. Hanczewski, M. Sobieraj, and M. D. Stasiak, "The direct method of effective availability for switching networks with multiservice traffic", *IEICE Trans. on Commun.*, vol. E99-B, no. 6, pp. 1291–1301, 2016 (DOI: 10.1587/transcom.2015EUP0009).

[20] M. Głąbowski, A. Kaliszan, and M. Stasiak, "Modeling product-form state-dependent systems with BPP traffic", *Perform. Eval.*, vol. 67, no. 3, pp 174–197, 2010 (DOI: 10.1016/j.peva.2009.10.002).

[21] V. Paxson and S. Floyd, "Wide area traffic: the failure of poisson modeling", *IEEE/ACM Trans. on Network.*, vol. 3, no. 3, 226–244, 1995 (DOI: 10.1109/90.392383).

[22] J. Roberts, V. Mocci, and I. Virtamo, Eds., *Broadband Network Teletraffic: Performance Evaluation and Design of Broadband Multiservice Networks, Final Report of Action COST 242*, LNCS, vol. 1155. Berlin: Springer, 1996 (ISBN: 978-3540618157).

[23] S. Hanczewski and M. Weissenberg, "Concept of an analytical model for cloud computing infrastructure", in *Proc. 11th Int. Symp. on Commun. Syst., Netw. & Digit. Sig. Process. CSNDSP 2018*, Budapest, Hungary, 2018 (DOI: 10.1109/CSNDSP.2018.8471814).

[24] F. Kelly, "Notes on effective bandwidth", in *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary, and I. Ziedins, Eds. Clarendon Press, 1996 (ISBN: 978-0198523994).

[25] M. Stasiak, "Blocking probability in a limited-availability group carrying mixture of different multichannel traffic streams", *Annales des Télécommun..* vol. 48, no. 1–2, pp. 71–76, 1993 (DOI: 10.1007/BF03005233).

[26] E. Brockmeyer, H. Halstrøm, and A. Jensen, *The life and works of A. K. Erlang*. Acta polytechnica scandinavica. Mathematics and computing machinery series, no. 6. Danish Academy of Technical Sciences, 1960.

[27] S. Hanczewski, M. Stasiak, and M. Weissenberg, "The analytical model of complex non-full-availability system", in *Image Processing and Communications*, M. Choraś and R. S. Choraś, Eds. *AISC*, vol. 1062, pp. 279–286. Springer, 2019 (DOI: 10.1007/978-3-030-31254-1_33).

Sławomir Hanczewski received his M.Sc. and Ph.D. degrees in Telecommunications from Poznan University of Technology, Poland, in 2001 and 2006, respectively. In 2020, he received a D.Sc. degree from Poznan University of Technology in ICT. Since 2001, he has been working at Poznan University of Technology (currently at the Faculty of Computing and Telecommunications). He is an Assistant Professor at the Institute of Communications and Computer Networks. He is the author and co-author of more than 60 scientific papers, mostly related to analytical modeling of telecommunications systems.

https://orcid.org/0000-0001-8524-7851
E-mail: slawomir.hanczewski@put.poznan.pl
Faculty of Computing and Telecommunications
Poznan University of Technology
Polanka 3
60-965 Poznan, Poland

Maciej Stasiak received his 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 a D.Sc. degree in Electrical Engineering from Poznan University of Technology. In 2006, he was appointed Full Professor. Between 1983–1992, he worked in the Polish industry sector as a designer of electronic and microprocessor systems. In 1992, he joined Poznan University of Technology, where he is currently the Head of the Institute of Communications and Computer Networks at the Faculty of Computing and Telecommunications. He is the author and co-author of over 250 scientific papers and five books. He is engaged in research and teaching in the area of performance analysis and modeling of queuing systems, multiservice networks and switching systems. Since 2004, he has been actively involved in research concerning modeling and dimensioning of 2G/3G/4G cellular networks.
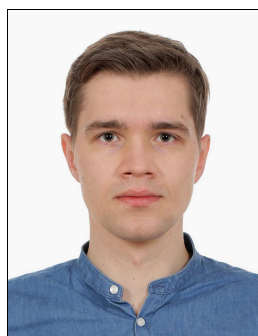
https://orcid.org/0000-0002-6572-6246
E-mail: maciej.stasiak@put.poznan.pl
Faculty of Computing and Telecommunication
Poznan University of Technology
Polanka 3
60-965 Poznan, Poland

Michał Weissenberg received his M.Sc. degree in Electronics and Telecommunications from the Faculty of Electronics and Telecommunications at Poznan University of Technology, Poland, in 2017, and is currently pursuing his Ph.D. at the Institute of Communications and Computer Networks, Faculty of Computing and Telecommunications, at the same University. His research interests include analytical modeling of telecommunications and cloud systems.

https://orcid.org/0000-0003-4830-433X
E-mail: michal.weissenberg@put.poznan.pl
Faculty of Computing and Telecommunication
Poznan University of Technology
Polanka 3
60-965 Poznan, Poland